Blind Super Resolution of Real-Life Video Sequences

Esmaeil Faramarzi, *Member, IEEE*, Dinesh Rajan, *Senior Member, IEEE*, Felix C. A. Fernandes, *Member, IEEE*, and Marc P. Christensen, *Senior Member, IEEE*

***Abstract*— Super resolution (SR) for real-life video sequences is a challenging problem due to complex nature of the motion fields. In this paper, a novel blind SR method is proposed to improve the spatial resolution of video sequences, while the overall point spread function of the imaging system, motion fields, and noise statistics are unknown. To estimate the blur(s), first, a nonuniform interpolation SR method is utilized to upsample the frames, and then, the blur(s) is(are) estimated through a multi-scale process. The blur estimation process is initially performed on a few emphasized edges and gradually on more edges as the iterations continue. Also for faster convergence, the blur is estimated in the filter domain rather than the pixel domain. The high-resolution frames are estimated using a cost function that has the fidelity and regularization terms of type Huber–Markov random field to preserve edges and fine details. The fidelity term is adaptively weighted at each iteration using a masking operation to suppress artifacts due to inaccurate motions. Very promising results are obtained for real-life videos containing detailed structures, complex motions, fast-moving objects, deformable regions, or severe brightness changes. The proposed method outperforms the state of the art in all performed experiments through both subjective and objective evaluations. The results are available online at http://lyle.smu.edu/~rajand/Video_SR/.**

***Index Terms*— Video super resolution, blur deconvolution, blind estimation, Huber Markov random field (HMRF).**

I. Introduction

MULTI-IMAGE super resolution (SR) is the process of estimating a high resolution (HR) image by fusing a series of low-resolution (LR) images degraded by various artifacts such as aliasing, blurring, and noise. Video super resolution, by contrast, is the process of estimating a HR video from one or multiple LR videos in order to increase the spatial and/or temporal resolution(s). The spatial resolution of an imaging system depends on the spatial density of the detector (sensor) array and the point spread function (PSF) of the induced detector blur. The temporal resolution, on the other hand, is influenced by the frame rate and exposure time of the camera [1], [2]. Spatial aliasing appears in images or video frames when the cut-off frequency of the detector is lower than that of the lens. Temporal aliasing arises in video sequences when the frame rate of the camera is not high enough to capture high frequencies caused by fast moving objects. The blur in the captured images and videos is the overall effect of different factors such as defocus, motion blur, optical blur, and detector's blur resulting from light integration within the active area of each detector in the array. The references [3]–[5] provide overviews of different SR approaches.

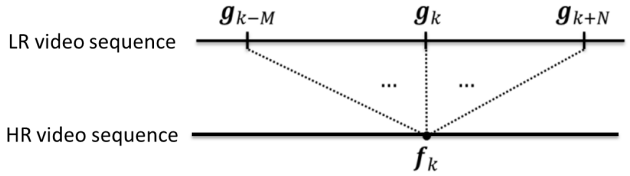

Fig. 1. A sliding window of size $M + N + 1$ is defined around each LR frame $\mathbf{g}_k$. The corresponding HR frame $\mathbf{f}_k$ is reconstructed using the SR operation by fusing the LR frames inside the window. Frames near the two sides of the video sequence may have shorter lengths.

One way to increase the resolution of a video is by overlaying a sliding window upon each frame and combining all frames falling inside the window to build the corresponding HR frame (Fig. 1) [6]. Then the window slides to the location of the other frames and the process repeats. For this system to work, usually a local registration method (such as optical flow, block-based, pel-recursive, or Bayesian [7]) is required to accurately estimate the displacement vector of each pixel or block within the frames. However, local registration may not be reliable in some cases, especially when there are complex dynamic changes (e.g. complex 3D motions), nonrigid deformations (e.g. flowing water, flickering fire), or changes in illumination [8].

Another class of single-video SR techniques is the one known as learning-based, patch-based or example-based video SR [9], [10]. The basic idea is that small space-time patches within a video are repeated many times inside the same video or other videos, at multiple spatio-temporal scales. Therefore, by replacing LR patches in the input video with equivalent HR patches from internal/external sources, the resolution can be improved. The major advantage of patch-based image/video SR methods is that motion estimation and object segmentation are not required. However, techniques of this group often have high computational complexity and most of them need offline database training. Furthermore, it is necessary that LR patches are generated from HR patches by a known PSF.

Manuscript received October 7, 2013; revised June 11, 2014 and March 30, 2015; accepted January 14, 2016. Date of publication January 28, 2016; date of current version February 23, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sergio Goma.

E. Faramarzi and F. C. A. Fernandes are with Samsung Research America, Richardson, TX 75082 USA (e-mail: e.faramarzi@samsung.com; felix.f@samsung.com).

D. Rajan and M. P. Christensen are with Southern Methodist University, Dallas, TX 75205 USA (e-mail: rajand@lyle.smu.edu; mpc@lyle.smu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2523344

Most works on the image and video SR are non-blind, i.e. they do not consider blur identification during the SR reconstruction. These methods assume that the PSFs are either known *a priori* or negligible, both of which are simplistic assumptions for realistic applications. For images, there has been a significant amount of publications on blind deconvolution, e.g. [11]–[18], and a few on blind SR, e.g. [2], [19]. However, to the best of our knowledge, there are only two independent works on blind SR for videos, which are discussed next.

In [20] a method is proposed which uses a total variation (TV) prior for frames and a combination of quadratic and TV priors for blurs. The motion is estimated globally either through a phase correlation method or by using an 8-parameter perspective (homography) model. Pixels that are detected to have inaccurate motion parameters are filtered out from the reconstruction by applying a masking operation. A limitation of the work in [20] is that since motion is estimated globally, all regions having local motions need to be masked out. Therefore, the reconstructed video would be of low quality for all locally-moving objects in the scene which could be the most prominent regions to reconstruct. Moreover, results are only presented for PSFs with small spatial support.

In [21] and [22] a Bayesian approach is proposed for simultaneously estimating the HR frames, motion, blur and noise parameter. The regularization terms for all unknowns are of type l_1 norm. An estimated noise parameter is used to update the weight of the fidelity term at each iteration of the optimization procedure. The noise level is updated at each iteration, but assumed to be identical for all pixels. The blur kernel is assumed to be separable ($h = h_x * h_y$) and results are only provided with Gaussian blurs. Promising results are shown for $4\times$ upscaling of real-life videos.

In [2] we proposed a method for blind deconvolution and super resolution of still images. Similar to most literature on SR, we assumed the motion fields between the input images to be global and translational. In this paper, we extend [2] to the case of video sequences with complex motion fields. Errors in the estimated motions make the frame and blur reconstructions more challenging, so a careful estimation process is required to achieve accurate results.

For blur estimation, the input video is first upsampled (in case of SR) using a nonuniform interpolation (NUI) SR method, then an iterative procedure is applied using the following considerations: 1) during the initial iterations, the blur is estimated exclusively using a few emphasized edges while weak structures are smoothed out, 2) the number of contributing edges gradually increases as iterations proceed, 3) structures finer than the blur support are omitted from estimation, 4) the estimation is done in the filter domain rather than pixel domain, and finally 5) the estimation is performed at multiple scales to avoid getting trapped in local minima.

The cost function used for frame debluring during the blur estimation process has fidelity and regularization terms both of type Huber-Markov Random Field (HMRF). A fidelity term of this type diminishes outliers caused by inaccurate motion estimation and preserve edges. By contrast, a HMRF prior exploits the piecewise smoothness nature of the HR frames to suppress noise while preserving the edges. Unlike in [2], we discard the output of frame debluring process after the blur estimation is accomplished and perform a non-blind SR reconstruction to obtain the final estimates of the video frames. To improve the performance of this final frame estimation, the fidelity term is weighted adaptively at each iteration pixel by pixel. We prove the performance of our proposed method with different experiments and comparisons with the state of the art methods.

We assume that noise is additive white Gaussian with similar statistics for all frames and color channels. The blurs in the input LR video are space-invariant (SI) (identical for all pixels within the frames), spatial (no temporal extension), equal for all color channels (by ignoring the chromatic aberration of the lens), and either identical or with gradual variations over time. However, no prior knowledge about the type and size of the PSF is required.

In summary, the major differences of this work compared to our previous one [2] are as follows: 1) processing videos with arbitrary local motions rather than images with global and translational motion differences, 2) discussions on YCbCr/RGB color spaces and sequential/central motion estimations, 3) adding a final non-blind frame reconstruction after blur estimation, 4) removing structures finer than the blur support during motion estimation, 5) using a fidelity term of type HMRF rather than quadratic for frame reconstruction to improve the performance, and 6) using a masking operation during frame reconstruction to suppress artifacts.

The usual way to model motion blur in video sequences is to define it as a 2D spatial PSF. Using this model, the PSF would be space-variant (SV) when the scene contains objects that move fast during the exposure time of the camera. To reconstruct in such a case, segmentation techniques are required to separate these objects from the rest of the scene, estimate their motions, deblur them, and then place them back to the scene in a way that the reconstructed frames seem consistent and artifact-free. This process may be difficult or even impossible when the motion blur is so severe that the shape of the objects is distorted. However, motion blur has a temporal nature [23], [24], so by separating this blur from other spatial blurring sources and modeling it as a rectangular temporal PSF with a length equivalent to the exposure time, the overall 3D spatio-temporal PSF would be SI (if spatial blurs are all SI). In this paper, we assume that either the motion blur is global, or the camera's exposure time is high enough so that no local motion blur appears in the captured videos.

This paper is organized as follows: Section II discusses the SR observation (forward) model, different color spaces for SR processing, and two general approaches for motion estimation. The blur estimation procedure is introduced in Section III. A non-blind SR process to estimate the final HR frames is discussed in Section IV. Experimental results are presented in Section V, and finally Section VI concludes the paper.

II. Model Definition

A. Observation Model

As shown in Fig. 1, a sliding window (temporal) of length $M + N + 1$ (with M frames backward and N frames forward)

is overlaid around each LR frame g_k of size $N_x^g \times N_y^g \times C$, and all LR frames inside the window are combined through the SR process to generate the HR reference frame f_k of size $N_x^f \times N_y^f \times C$. Here, N_x and N_y are frame dimensions in two spatial directions and C is the number of color channels. The linear forward imaging model that illustrates the process of generating a LR frame g_i inside the window from the HR frame f_k is given by:

$$\begin{aligned} g_i\left(x_\downarrow, y_\downarrow; c\right) = & \left[m_{k,i}\left(f_k(x, y; c)\right) * h(x, y)\right]_{\downarrow L} \\ & + n_{k,i}\left(x_\downarrow, y_\downarrow; c\right), \quad c = 1, \ldots, C, \\ & k = 1, \ldots, P, \quad i = k - M, \ldots, k + N \end{aligned} \tag{1}$$

where P is the total number of frames, $(x_\downarrow, y_\downarrow)$ and (x, y) indicate the pixel coordinates in LR and HR image planes respectively, L is the downsampling factor or SR upscaling ratio (so that $N_x^f = LN_x^g$ and $N_y^f = LN_y^g$), and $*$ is the two-dimensional convolution operator. According to this model, the HR frame f_k is warped with the warping function $m_{k,i}$, blurred by the overall system PSF h, downsampled by factor L, and finally corrupted by the additive noise $n_{k,i}$.

It is more convenient to express this linear process in the vector-matrix notion:

$$\mathbf{g}_i = \mathbf{DHM}_{k,i}\mathbf{f}_k + \mathbf{n}_i \tag{2}$$

In (2) $\mathbf{f}_k$ is the kth HR frame in lexicographical notation indicating a vector of size $N_x^f N_y^f C \times 1$, matrices $\mathbf{M}_{k,i}$ and $\mathbf{H}$ are the motion (warping) and convolution operators of size $N_x^f N_y^f C \times N_x^f N_y^f C$, $\mathbf{D}$ is the downsampling matrix of size $N_x^g N_y^g C \times N_x^f N_y^f C$, and $\mathbf{g}_i$ and $\mathbf{n}_i$ are vectors of the ith LR frame and noise respectively, both of size $N_x^g N_y^g C \times 1$. The matrix $\mathbf{M}_{k,i}$ registers (or motion compensates) the reference frame $\mathbf{f}_k$ to match the frame $\mathbf{f}_i$. As a result, $\mathbf{M}_{kk}$ is an identity (unit) matrix since no motion compensation is required between a HR frame and its coincident LR frame. For a blur deconvolution (BD) problem (i.e. $L = 1$), $\mathbf{D}$ is the identity matrix and so the input and output videos are of the same size. Hence BD can be considered as a special case of SR. The objective in SR and BD is to estimate the HR frames $\mathbf{f}_k$ and the blur $\mathbf{H}$ given the LR frames $\mathbf{g}_i$ while the motion $\mathbf{M}_{k,i}$ and the noise $\mathbf{n}_i$ are unknown as well.

B. Color Space

The human visual system (HVS) is less sensitive to chrominance (color) than to luminance (light intensity). In the RGB (red, green, blue) color space, the three color components have equal importance and so all are usually stored or processed at the same resolution. But a more efficient way to take the HVS perception into account is by separating the luminance from the color information and representing luma with higher resolution than chroma [25]. A popular way to achieve this separation is to use the YCbCr color space where Y is the luma component (computed as a weighted average of R, G, and B) and Cb and Cr are the blue-difference and red-difference chroma components. The YUV video format is commonly used by video processing algorithms to describe video sequences encoded using YCbCr.

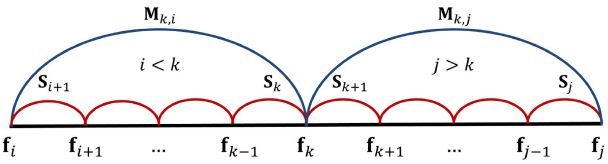

Fig. 2. Central motion (blue) versus sequential motion (red).

In our implementation, video sequences can be processed in either RGB or YUV formats. In the former case, SR is used to increase the resolution of all R, G, and B channels. However in the latter one, only the Y channel is processed by SR for faster computation while the Cb and Cr channels are simply upscaled to the resolution of the super-resolved Y channel using a single-frame upsampling method such as bilinear or bicubic interpolation. The obtained results related to these two cases are comparable using a subjective quality assessment.

C. Motion Estimation

Accurate motion estimation (registration) with subpixel precision is crucial for video SR to achieve a good performance. Two different approaches can be considered for registration in video SR: central and sequential (Fig. 2). In the former, motion is directly computed between each reference frame and all LR frames inside its sliding window (Fig. 1). By contrast, in the latter, each frame is registered against its previous frame; then to use with SR, sequential motion fields must be converted to central fields for registration as follows: if $\mathbf{S}_i = \left[\mathbf{S}_{x_i}, \mathbf{S}_{y_i}\right]$ is the sequential motion field for the ith frame (w.r.t. the $(i-1)$th frame), then $\mathbf{M}_{k,i} = \left[\mathbf{M}_{x_{k,i}}, \mathbf{M}_{y_{k,i}}\right]$, the central motion field for the ith frame when the central frame is the kth frame is obtained as:

$$\begin{aligned} \mathbf{M}_{k,i} &= -\sum_{n=i+1}^{k} \mathbf{S}_n = -\mathbf{S}_{i+1} + \mathbf{M}_{k,i+1}, \quad k - M \le i < k \\ \mathbf{M}_{k,k} &= \mathbf{I} \\ \mathbf{M}_{k,j} &= \sum_{n=k+1}^{j} \mathbf{S}_n = \mathbf{S}_j + \mathbf{M}_{k,j-1}, \quad k < j \le k + N \end{aligned} \tag{3}$$

where $\mathbf{I}$ is the identity matrix.

With the sequential approach in SR, each frame needs to be registered only against the previous frame, whereas with the central approach each frame is registered against all neighboring frames within its reconstruction window. Therefore, the computational complexity and the storage size of the motion fields in the central approach is higher than that of using the sequential approach.

The lower storage size of the motion fields in the sequential approach is important in applications where the ground-truth video is available (e.g. when the video to be transmitted is downsampled intentionally to cope with the bandwidth limitations of the communication channel). In this situation, the sequential motion fields estimated from the original video can be used by a SR processing unit at the receiver side to improve the SR performance and also reduce the computational cost

(specially in real-time applications). To transmit the motion fields as metadata, they can be either embedded into the bitstream of the encoded downsampled video (e.g. via the SEI[1] message of AVC[2] or HEVC[3] video compression standards) or sent out separately over the channel along with the video.

Contrary to the sequential approach that should be computed between successive frames of the same resolution (e.g. between bicubically upsampled successive LR frames), with the central approach a reference HR frame $\mathbf{f}_k$ can be directly registered against its LR neighboring frame $\mathbf{g}_i$, as performed in [21] and [22]. This method may result in a more accurate estimation for the central approach, although at the cost of more computational complexity for each registration.

With both sequential and central approach, the motion fields can be reestimated after several iterations of the blur estimation procedure (Section III) to refine the accuracy of motion estimates.

In our work, we use the sequential approach to estimate the motion fields between the successive frames by the use of dense optical flow method described in [26].

III. Blur Estimation

In a multi-channel BD problem, the blurs could be estimated accurately along with the HR images [27]. However in a blind SR problem with a possibly different blur for each frame, some ambiguity in the blur estimation is inevitable due to the downsampling operation [19]. By contrast, in a blind SR problem in which all blurs are supposed to be identical or have gradual changes over time, such an ambiguity can be avoided [2]. Moreover, as discussed in Section III-A, the assumption of identical (or gradually changing) blurs makes it possible to separate the registration and upsampling procedures from the deblurring process which significantly decreases the blur estimation complexity.

In Section III-A, the NUI method to reconstruct the upsampled frame is explained. This upsampled yet-blurry frame is used to estimate the PSF(s) and the deblurred frames through an iterative alternative minimization (AM) process. The blur and frame estimation procedures are discussed in Sections III-B and III-C, respectively. The estimated frames are used only for the deblurring process and so omitted thereafter. Finally, the overall AM optimization process is described in Section III-D.

A. Frame Upsampling

In [2] we discuss the situations in which the warping and blurring operations in (2) are commutable. Although for videos with arbitrary local motions this commutability does not hold exactly for all pixels, however we assume here that this is approximately satisfied. The ultimate appropriateness of the approximation is validated by the eventual performance of the algorithm that is derived based on this model. With this assumption, (2) can be rewritten as:

$$\mathbf{g}_i = \mathbf{D}\mathbf{M}_{k,i}\mathbf{H}\mathbf{f}_k + \mathbf{n}_i = \mathbf{D}\mathbf{M}_{k,i}\mathbf{z}_k + \mathbf{n}_i \tag{4}$$

[1]Supplemental enhancement information.
[2]Advanced Video Coding.
[3]High Efficiency Video Coding.

where $\mathbf{z}_k = \mathbf{H}\mathbf{f}_k$ is the upsampled but still blurry frame. Equation (4) suggests that we can first construct the upsampled frames $\mathbf{z}_k$ using an appropriate fusion method and then apply a deblurring method to $\mathbf{z}_k$ to estimate $\mathbf{f}_k$ and $\mathbf{h}$.

If noise characteristics are also the same for all frames, an appropriate way to estimate $\mathbf{z}_k$ is using the NUI method [3]–[5]. In NUI, the pixels of all LR frames are projected on to the HR image grid according to their motion fields, and then the intensities of the true locations on the grid are computed via interpolation [2]. Our experiments show that using NUI for upsampling the frames leads to better estimates of $\mathbf{f}$ and $\mathbf{h}$ (Sections III-B and III-C) compared to when $\mathbf{z}_k$ is estimated iteratively from the LR frames $\mathbf{g}$ using a MAP (Maximum A Posteriori) or ML (Maximum Likelihood) method such as [29] and [30].

B. Frame Deblurring

After upsampling the frames, we use the following cost function, J, to estimate the HR frames $\mathbf{f}_k$ having an estimate of the blur $\mathbf{h}$ (or $\mathbf{H}$):

$$J(\mathbf{f}_k) = \|\boldsymbol{\rho}\left(\mathbf{H}\mathbf{f}_k - \mathbf{z}_k\right)\|_1 + \lambda^n \sum_{j=1}^{4} \left\|\boldsymbol{\rho}\left(\nabla_j \mathbf{f}_k\right)\right\|_1 \tag{5}$$

where $\|\cdot\|_1$ denotes the l_1 norm (defined for a sample vector $\mathbf{x}$ with elements x_i as $\|\mathbf{x}\| = \sum_i |x_i|$), λ^n is the regularization coefficient, $\boldsymbol{\rho}(\cdot)$ is the vector Huber function, $\|\boldsymbol{\rho}(\cdot)\|_1$ is called the Huber norm, and ∇_j $(i = 1, \ldots, 4)$ are the gradient operators in 0°, 45°, 90° and 135° spatial directions [2]. The first term in (5) is called the fidelity term which is the Huber-norm of error between the observed and simulated LR frames. While in most works the l_2-norm is used for the fidelity term, we use the robust Huber norm to better suppress the outliers resulting from inaccurate registration. The next two terms in (5) are the regularization terms which apply spatio-temporal smoothness to the HR video frames while preserving the edges.

Each element of the vector function $\boldsymbol{\rho}(\cdot)$ is the Huber function defined as:

$$\rho(x) = \begin{cases} x^2 & \text{if } |x| \le T \\ 2T|x| - T^2 & \text{if } |x| > T, \end{cases} \tag{6}$$

The Huber function $\rho(x)$ is a convex function that has a quadratic form for values less than or equal to a threshold T and a linear growth for values greater than T. The Gibbs PDF of the Huber function is heavier in the tails than a Gaussian. Consequently, edges in the frames are less penalized with this prior than with a Gaussian (quadratic) prior.

To minimize the cost function in (5), we use the conjugate gradient (CG) iterative method [30] because of its simplicity and efficiency. Compared to some other iterative methods such as Gauss-Seidel (GS) or SOR that need explicit derivation of matrix $\mathbf{A}$ when solving a linear equation $\mathbf{A}\mathbf{x} = \mathbf{b}$, CG can decompose the matrix $\mathbf{A}$ to concatenation of filtering and weighting operations. However, CG can only be used with linear equation sets, whereas the cost function in (5) is non-quadratic and so its derivative is nonlinear. To overcome this limitation, we use lagged diffusivity fixed-point (FP) iterative

method [31] to lag the diffusive term by one iteration [15]. Using this method for a sample vector $\mathbf{x}$, at the nth iteration the non-quadratic Huber-norm $\|\boldsymbol{\rho}(\mathbf{x}^n)\|_1$ is replaced by the following quadratic form:

$$\left\|\boldsymbol{\rho}(\mathbf{x}^n)\right\|_1 = (\mathbf{x}^n)^T \mathbf{V}^n(\mathbf{x}^n) = \left\|\mathbf{x}^n\right\|^2_{\mathbf{V}^n}, \tag{7}$$

where $\mathbf{V}^n$ is the following diagonal matrix:

$$\mathbf{V}^n = \text{diag}\left(\begin{cases} 1 & \mathbf{x}^{n-1} \dot{\leq} T \\ T \dot{/} \mathbf{x}^{n-1} & \mathbf{x}^{n-1} \dot{>} T \end{cases}\right) \tag{8}$$

In (8) the dots above the division and comparison operators indicate element-wise operations. Applying the FP method to (5) and setting the derivative of the cost function with respect to $\mathbf{f}_k$ to zero results in the following linear equation set:

$$\mathbf{H}^{n\,T}\mathbf{V}^n\mathbf{H}^n + \lambda^n \sum_{j=1}^{4} \nabla_j^T \mathbf{W}_j^n \nabla_j = \mathbf{H}^{n\,T}\mathbf{V}^n\mathbf{z}_k, \tag{9}$$

where:

$$\begin{aligned} \mathbf{V}^n &= \text{diag}\left(\boldsymbol{\rho}(\mathbf{H}\mathbf{f}_k^{n-1} - \mathbf{z}_k)\right), \\ \mathbf{W}_j^n &= \text{diag}\left(\boldsymbol{\rho}(\nabla_j \mathbf{f}_k^{n-1})\right) \end{aligned} \tag{10}$$

We discuss how to update the regularization parameter λ^n at each iteration in Section III-D.

C. Blur Estimation

Within an image or video frame, non-edge regions and weak structures are not appropriate for blur estimation. Hence, more accurate results would be obtained if the estimation is not performed in such regions. For this reason, in [11] and [33] the user should first manually select a region with rich edge structure, whereas in [2], [13], [14], and [34] the most salient edges are automatically chosen. Moreover, sharpening salient edges would also improve the accuracy of blur estimation. The authors of [34] leveraged these two strategies by preprocessing blurred images with the shock filtering method proposed in [35]. Shock filtering is an edge preserving smoothing operation by which soft edges gradually approach step edges within a few iterations while non-edge regions are smoothed. Since shock filtering is sensitive to noise, sometimes a pre-filtering operation is applied to first suppress noise. For example, in [13] bilateral filtering (proposed by [36]) is used and in [14] and [34] a lowpass Gaussian filtering is utilized before shock filtering. A similar concept for the blur estimation is exploited in [37] in which the image is first sharpened by redistributing the pixels along the edge profiles in such a way that antialiased step edges are produced. Having the sharpened image and the blurry input image, the blur is then estimated using a maximum a posteriori (MAP) framework.

In our work, we employ the edge-preserving smoothing method of [40] in which the number of surviving edges after smoothing is globally controlled by the regularization coefficient. This feature is helpful when one desires to limit the number of salient edges at each iteration. This smoothing method aims to keep an intended number of non-zero gradients through l_0 gradient minimization using the following cost function:

$$J\left(\mathbf{f}'^n_k\right) = \left\|\mathbf{f}'^n_k - \mathbf{f}^n_k\right\|_2^2 + \beta^n \left(\left\|\nabla_x \mathbf{f}'\right\|_0 + \left\|\nabla_y \mathbf{f}'\right\|_0\right), \tag{11}$$

where $\mathbf{f}'^n_k$ is the output of the edge-preserving smoothing algorithm and the l_0 norm is defined as $\|\mathbf{x}\|_0 = \#\left(i \,|\, x_i \neq 0\right)$.

Algorithm 1 Blur Estimation Procedure

Require: $\mathbf{g}_1, ..., \mathbf{g}_p, \lambda_{min}, \gamma_{min}$ and initials $\mathbf{h}^0, \lambda^0, \gamma^0, \beta^0, T_1^0, T_2^0$
Set $n := 0$ **% AM loop iteration number**
$S :=$ # of Scales

Use luma or one color channel of $\mathbf{g}_1, ..., \mathbf{g}_{p_1}$
for $k := 1$ **to** P_1 **do** **% Loop on P_1 reference frames**

  if $L > 1$ **then** **% For SR reconstruction**
    $\mathbf{z}_k$ = NUI$(\mathbf{g}_{k-M}, ..., \mathbf{g}_{k+N})$
  else **% For BD reconstruction**
    $\mathbf{z}_k = \mathbf{g}_k$
  end if
  $\mathbf{f}_0^k = \mathbf{z}_k$

  % HR frame and blur estimation
  for $s := 1$ **to** S **do** **% Multi-scale approach**
    Rescale $\mathbf{z}_k, \mathbf{f}_k^n$ and $\mathbf{h}^n$

    % AM loop iteration
    while "AM stopping criterion" is not satisfied **do**
      $n = n + 1$

      % Updating procedure for f
      Compute $\mathbf{V^n}$ and $\mathbf{W}_j^n$ using (10)
      Update λ^n
      while $\mathbf{f}^n$ does not satisfy "CG stopping criterion" **do**
        $\mathbf{f}_k^n :=$ CG iteration for system in (9); starting at $\mathbf{f}_k^{n-1}$
      end while
      Apply constraints on $\mathbf{f}_k^n$

      % Updating procedure for h
      Update γ^n, β^n, T_1^n and T_2^n
      Compute the smoothed frame $\mathbf{f}'^n_k$ from (11)
      Compute $\nabla \mathbf{f}''^n_k$ from (15)
      Edge tapping of $\nabla \mathbf{f}''^n_k$
      Compute $h_k^n(x, y)$ from (17)
      Apply constraints on $\mathbf{h}^n$

    end while
  end for
end for

Algorithm 2 Final Frame Estimation Procedure

Require: $\mathbf{g}_1, ..., \mathbf{g}_p$ and λ
Set $n := 0$ **% FP loop iteration number**
for $k := 1$ **to** P **do** **% Loop on P reference frames**
  Estimate sequential motion fields $\mathbf{S}_1, ..., \mathbf{S}_p$
  Compute central motion fields $\mathbf{M}_1, ..., \mathbf{M}_p$ using (**??**)
  Estimate the blur $\mathbf{h}$ using Algorithm 1

  % Estimate HR frames using FP loops
  while "FP stopping criterion" is not satisfied **do**
    $n = n + 1$
    Compute $\mathbf{O}_{kj}^n$ using (22)
    Compute $\mathbf{V^n}$ and $\mathbf{W}_j^n$ using (21)
    while $\mathbf{f}^n$ does not satisfy "CG stopping criterion" **do**
      $\mathbf{f}_k^n :=$ CG iteration for system in (20); starting at $\mathbf{f}_k^{n-1}$
    end while
    Apply constraints on $\mathbf{f}_k^n$

  end while
end for

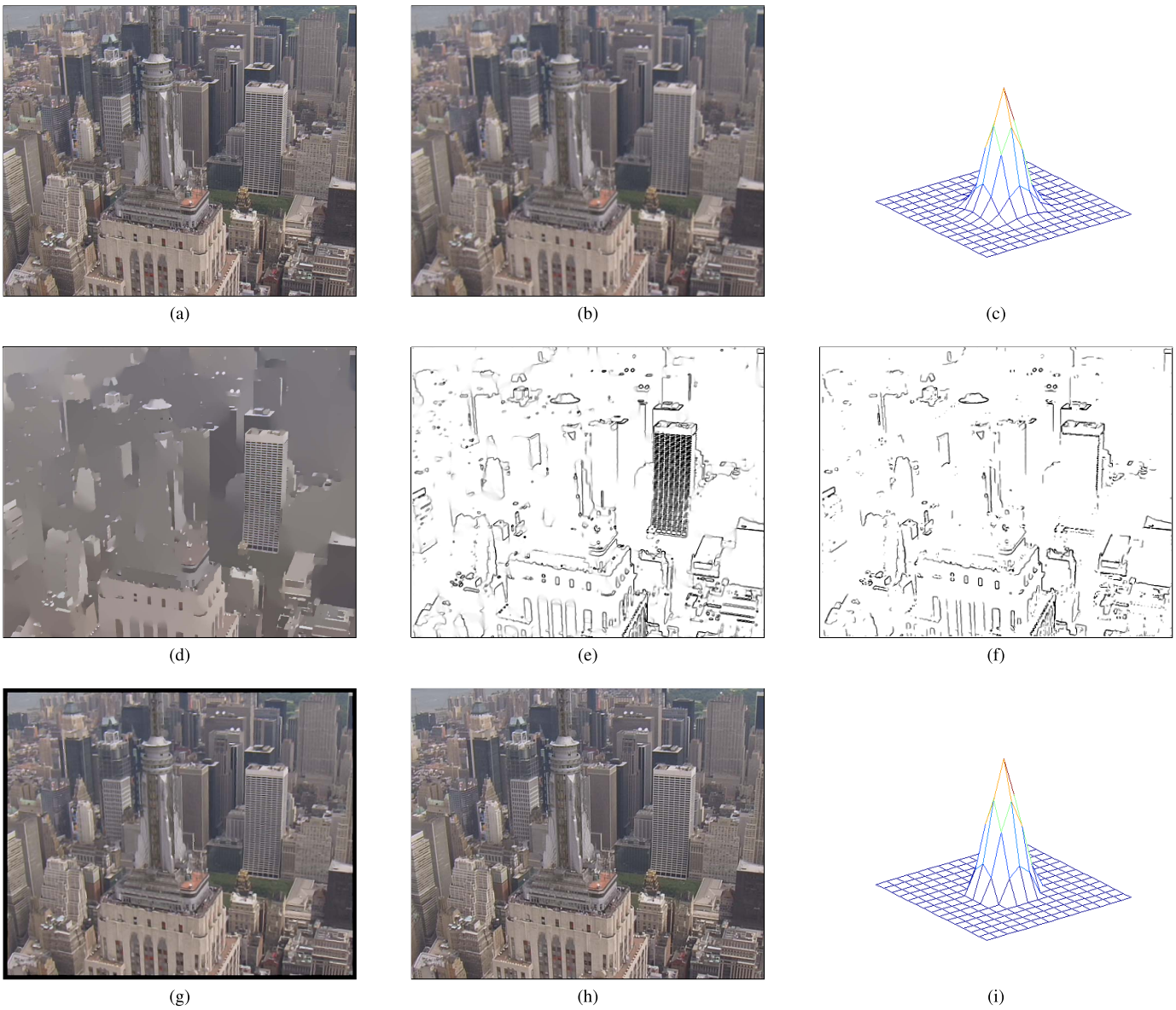

Fig. 3. Reconstruction result for the *City* video sequence (please zoom into the figure on screen; see the videos at **http://lyle.smu.edu/~rajand/Video_SR/**). (a) Ground-truth frame; (b) LR frame by applying a Gaussian PSF with $\sigma = 1.2$ having size of 15×15, downsampling ratio of 2, and Gaussian noise with SNR of $30dB$, then upsampled to the original resolution using Bicubic; (c) Original blur; (d) Smoothed frame; (e) Negative of gradient magnitude of (d); (f) Salient edges not narrower than the kernel support; (g) Result of *3D-ISKR* [38] deblurred by [39] with PSNR of $28.9dB$ (after border removal); (h) Result of our proposed method with PSNR of 32.4 dB; (i) Estimated blur with NMSE of 0.1.

Unlike shock filtering, this smoothing method does not need pre-filtering of noise.

Though sufficient edge pixels are required for accurate blur estimation, it is shown in [14] that structures with scales smaller than the PSF support could harm blur estimation. Inspired by that work, we define $\mathbf{R}_k^n$ in (12) to measure the usefulness of each pixel for blur estimation:

$$\mathbf{R}_k^n = \left|\mathbf{A}\mathbf{B}\mathbf{f}'^n_k\right|, \tag{12}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the convolution operators for the spatial filters a and b, respectively, as defined below:

$$a = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \tag{13}$$

$$b = \nabla_x + \nabla_y = \begin{bmatrix} 2 & -1 \\ -1 & 0 \end{bmatrix} \tag{14}$$

In (13) and (14), a is the all-ones filter of size 11×11 and b is the sum-of-gradients filter. According to (12)-(14), to compute $\mathbf{R}_k^n$, the sum of gradient components of $\mathbf{f}'^n_k$ is computed first, then at each pixel it is summed up with the values of all neighboring pixels, and finally its absolute value is obtained. For pixels on narrow structures, the sum of gradient values cancels out each other. Therefore, $\mathbf{R}_k^n$ usually has a small value at the location of narrow edges and smooth regions. Then $\mathbf{f}'^n_k$ is refined by only retaining strong and non-spike edges:

$$\nabla\mathbf{f}''^n_k = \begin{cases} \nabla\mathbf{f}'^n_k & \text{if } \left|\nabla\mathbf{f}'^n_k\right| \dot{>} T_1^n \text{ and } \mathbf{R}_k^n \dot{>} T_2^n \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

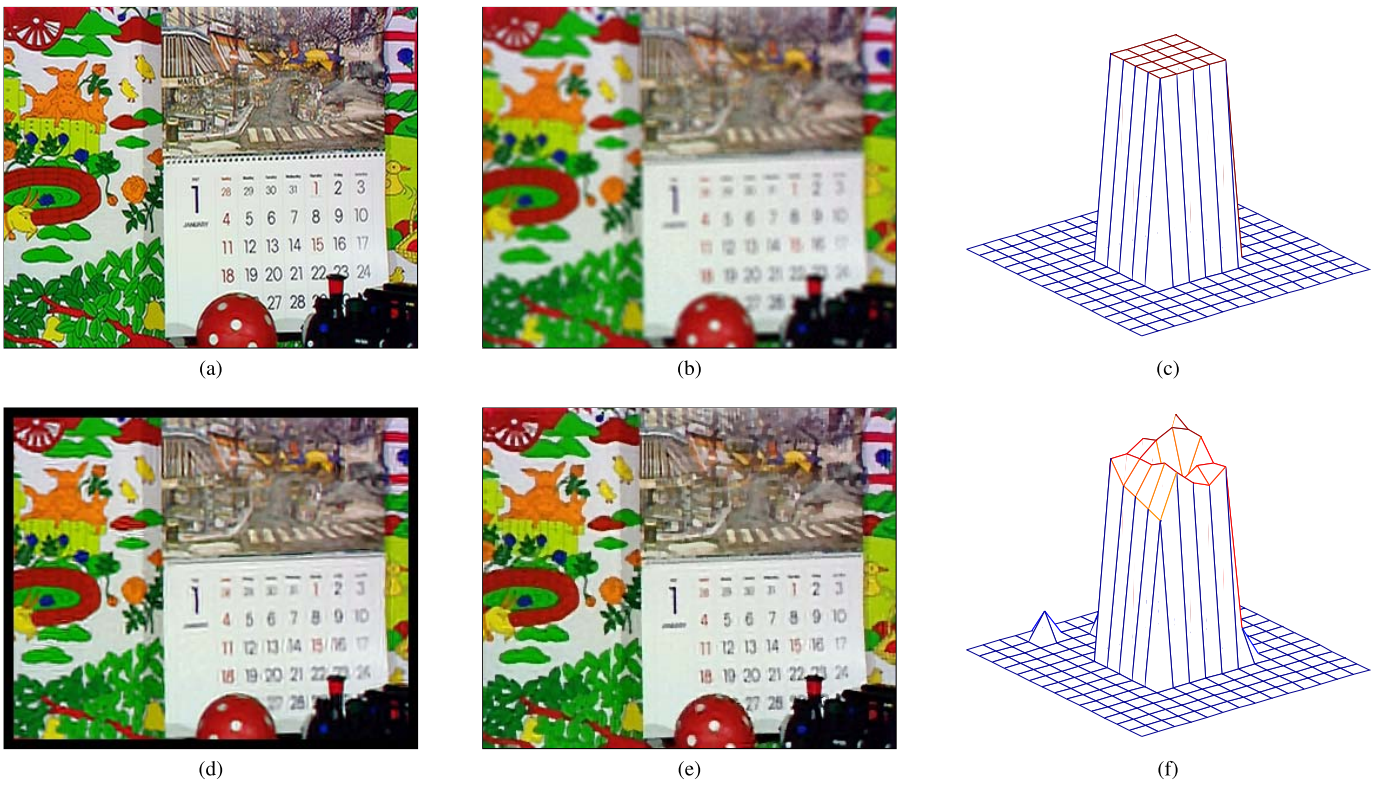

Fig. 4. Experimental results for the *Mobile* video sequence (please zoom into the figure on screen; see the videos at **http://lyle.smu.edu/~rajand/Video_SR/**). (a) Ground-truth video frame. (b) One LR frame generated by applying a 5×5 out-of-focus blur having size of 15×15, spatial downsampling of 2, and Gaussian noise with SNR of $30dB$, then upsampled to the original resolution using Bicubic; (c) Original blur; (d) Reconstruction result of *3D-ISKR* [38] deblurred by [39] with PSNR of $20.8dB$; (e) Reconstruction result of our proposed method with PSNR of $22.7dB$; (f) The estimated PSF with NMSE of 0.15.

where T_1^n and T_2^n are threshold parameters which decrease at each iteration.

In our work, the blur is estimated from the gradients of $\mathbf{z}_k$ and $\mathbf{f}''_k$ instead of their pixel values since estimation in the filter domain converges faster than the pixel domain. The reason for faster convergence is that in a linear equation set $\mathbf{Ah} = \mathbf{b}$ (derived from a quadratic cost function defined for $\mathbf{h}$), matrix $\mathbf{A}$ would be better conditioned when the gradients of images are used [13].

To avoid ringing artifact, we apply the MATLAB function *edgetaper()* to $\nabla\mathbf{f}''^n_k$. Then we estimate each blur $\mathbf{h}_k$ using the cost function $J(\mathbf{h})$ below:

$$J(\mathbf{h}) = \sum_{k=1}^{P_1} \left\| \nabla\mathbf{z}_k - \nabla\mathbf{F}''_k\mathbf{h} \right\|_2^2 + \gamma^n \left\| \nabla\mathbf{h} \right\|_2^2, \qquad (16)$$

where $P_1 \leq M + N$ and $\mathbf{F}''_k$ is the convolution matrix of $\mathbf{f}''_k$. Since $J(\mathbf{h})$ in (16) is quadratic, it can be easily minimized by pixel-wise division in the frequency domain [41] as:

$$h_k^n(x,y) = \mathcal{F}^{-1}\left(\sum_{k=1}^{P_1} \sum_{i=1}^{2} \left\{ \left[\overline{\left(\mathcal{F}(\nabla_i) \dot{\times} \mathcal{F}(f''^n_k)\right)} \dot{\times} \left(\mathcal{F}(\nabla_i) \dot{\times} \mathcal{F}(z_k)\right) \right] \dot{/} \left[\left|\mathcal{F}(\nabla_i) \dot{\times} \nabla(f''^n_k)\right|^{\dot{2}} + \gamma^n \left|\mathcal{F}(\nabla_i)\right|^{\dot{2}} \right] \right\} \right) \qquad (17)$$

where $\nabla_i (i = 1, 2)$ is ∇_x or ∇_y, $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ are FFT and inverse-FFT operations, and $\overline{(\cdot)}$ is the complex conjugate operator. We then apply the following constraints to the estimated PSF: its negative values are set to zero, then the PSF is normalized to the range [0, 1], and centered in its support window.

D. Overall Optimization for Blur Estimation

The overall optimization procedure for estimating the PSF is shown in Algorithm 1. The HR frames and the PSF are sequentially updated within the AM iterations. We use a multi-scale approach to avoid trapping in local minima. The regularization coefficients λ^n in (9) and γ^n in (17) decrease at each AM (alternating minimization) iteration up to some minimum values λ_{min} and γ_{min}, respectively (see [2] for a discussion). The variation of these coefficients is given by:

$$\lambda^n = \max\left(r\lambda^{n-1}, \lambda_{min}\right), \qquad \gamma^n = \max\left(r\gamma^{n-1}, \gamma_{min}\right) \qquad (18)$$

where r is a scalar less than 1. Also the values of β^n in (11) and T_1^n and T_2^n in (15) fall at each AM iteration which increases the number of contributing pixels to blur estimation as the optimization proceeds.

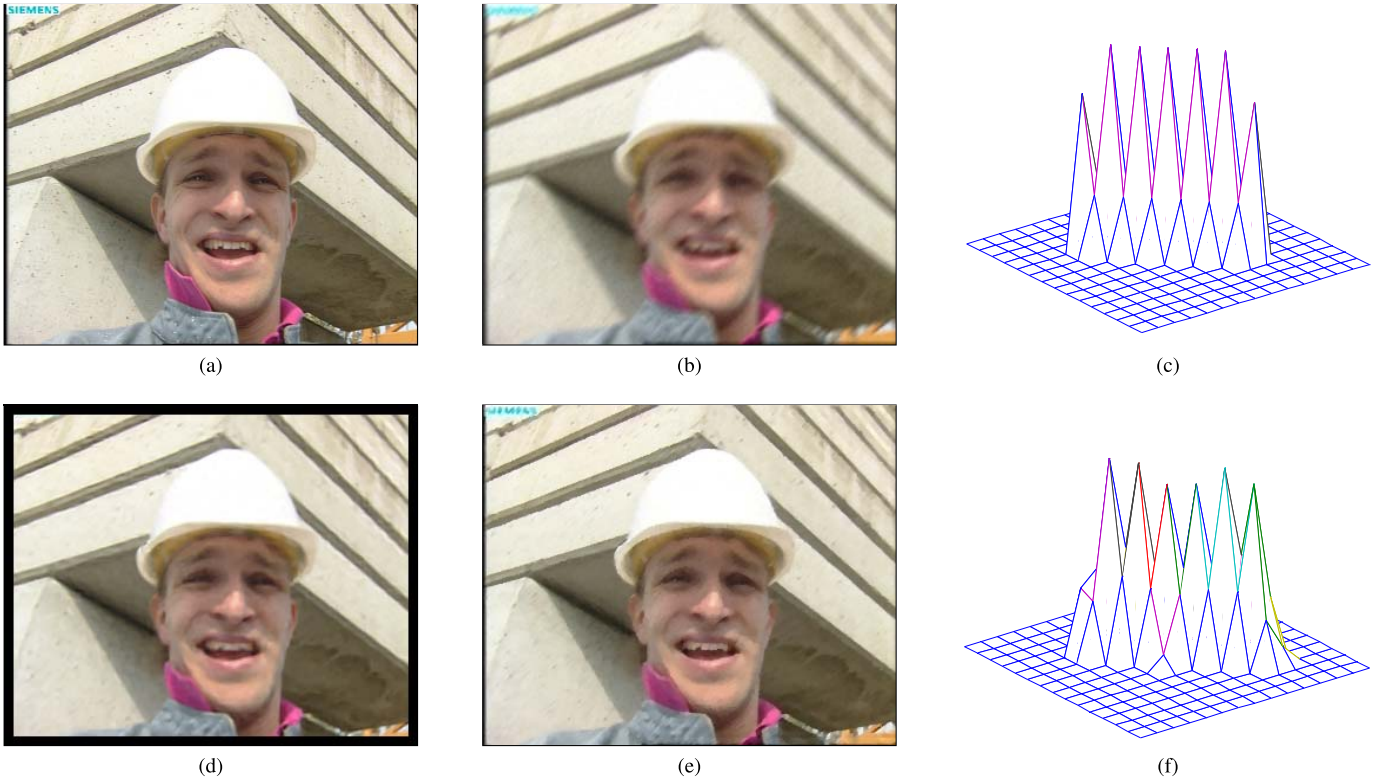

Fig. 5. Experimental results for the *Mobile* video sequence (please zoom into the figure on screen; see the videos at **http://lyle.smu.edu/~rajand/Video_SR/**). (a) Ground-truth video frame. (b) One LR frame generated by applying a 9×9 motion blur with size of 15×15, spatial downsampling of 2, and Gaussian noise with SNR of $30dB$, then upsampled to the original resolution using Bicubic; (c) Original blur; (d) The reconstruction result of *3D-ISKR* [38] deblurred by [39] with PSNR of $31dB$; (e) The reconstruction result of our proposed method with PSNR of $32.2dB$; (f) The estimated PSF with NMSE of 0.015.

IV. Final HR Frame Estimation

After the PSF estimation is completed, the final HR frames are reconstructed through minimizing the following cost function:

$$J\left(\mathbf{f}_1,\ldots,\mathbf{f}_P\right)=\sum_{k=1}^{P}\left(\sum_{i=k-M}^{k+N}\left\|\boldsymbol{\rho}\left(\mathbf{O}_{k,i}\left(\mathbf{DHM}_{k,i}\mathbf{f}_k-\mathbf{g}_i\right)\right)\right\|_1+\lambda\sum_{j=1}^{4}\left\|\boldsymbol{\rho}\left(\nabla_j\mathbf{f}_k\right)\right\|_1\right) \tag{19}$$

where $\mathbf{O}_{k,i}$ is a diagonal weighting matrix that assigns less weights to the outliers. Minimizing this cost function with respect to $\mathbf{f}_k$ yields:

$$\left(\sum_{i=k-M}^{k+N}\mathbf{M}_{k,i}^T\mathbf{H}^T\mathbf{D}^T\mathbf{O}_{k,i}^n\mathbf{V}^n\mathbf{DHM}_{k,i}+\lambda\sum_{j=1}^{4}\nabla_j^T\mathbf{W}^n\nabla_j\right)\mathbf{f}_k^n=\mathbf{M}_{k,i}^T\mathbf{H}^T\mathbf{D}^T\mathbf{O}_{k,i}^n\mathbf{V}^n\mathbf{g}_i \tag{20}$$

where:

$$\mathbf{V}^n=\operatorname{diag}\left(\boldsymbol{\rho}(\mathbf{DHM}_{k,i}\mathbf{f}_k^{n-1}-\mathbf{g}_i)\right),\qquad \mathbf{W}_j^n=\operatorname{diag}\left(\boldsymbol{\rho}(\nabla_j\mathbf{f}_k^{n-1})\right) \tag{21}$$

and the m-th diagonal element of $\mathbf{O}_{k,i}^n$ is computed according to:

$$o_{k,i}[m]=exp\left\{-\frac{\left\|\mathbf{R}_m\left(\boldsymbol{\rho}(\mathbf{DHM}_{k,i}\mathbf{f}_k^{n-1}-\mathbf{g}_i)\right)\right\|}{2\sigma^2}\right\} \tag{22}$$

where in (22) R_m is a patch operator which extracts a patch of size $q \times q$ centered at the m-th pixel of $\mathbf{f}_{k,i}$.

The final frame estimation procedure is demonstrated in Algorithm 2.

V. Experimental Results

In this section, the performance of our method is evaluated and compared with the state-of-the-art video SR methods *3D-ISKR*[4] [38] and *Fast Upsampler* [42] which are available for public evaluation, and also with the commercial software *Video Enhancer* [43]. Among these three, we only display the results from *3D-ISKR* [38]. This non-blind SR method does not include a debluring step, so we post process its outputs with the debluring method of [39]. Different parameters for debluring were tried out in each experiment to get the best possible outcomes from *3D-ISKR*. Furthermore, since *3D-ISKR* implementation does not estimate pixels near frame boundaries, we remove the boundaries from the reconstructed frame before an objective evaluation. As the outputs of *Fast Upsampler* and *Video Enhancer* have always a small global misalignment with the ground-truth frames, we use *Keren* method [44]

[4]Iterative Steering Kernel Regression.

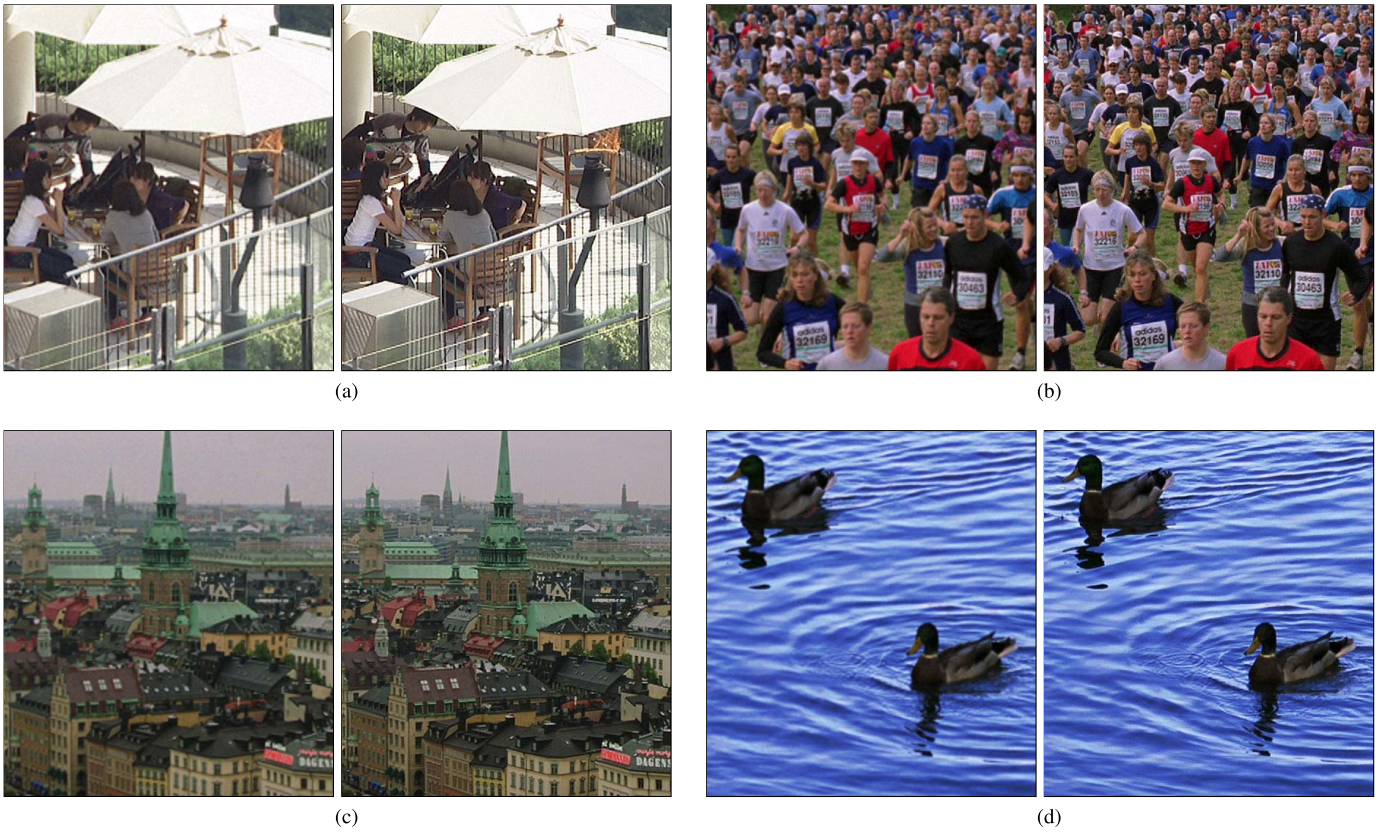

Fig. 6. Experimental results for some popular HD (1080p) video sequences with complicated motions; the frames are cropped for better visibility (please zoom into the figure on screen; see the videos at **http://lyle.smu.edu/~rajand/Video_SR/**).The SR upsampling ratio and temporal window size are 2 and 5, respectively. (a) BQ Terrace. (b) Crowd Run. (c) Old Town Cross. (d) Ducks Take Off.

to estimate and compensate the global misalignments. **The reconstructed videos of our proposed method are available at http://lyle.smu.edu/~rajand/Video_SR/.**

To measure the accuracy of our proposed blind method for different blur types, we synthetically generate LR video sequences from four popular videos commonly used in video processing experiments: *City*, *Mobile*, *Foreman*, and *Bus*. All these videos have 4:2:0 chroma subsampling format which means the chrominance channels have half of the horizontal and vertical resolutions of the luminance channel [25]. The PSFs in the experiments are generated using the MATLAB function *fspecial()*. Also we demonstrate the results of upscaling some other popular videos to the FHD[5] or 1080p (1920×1080 progressive) resolution. Furthermore, to see the performance for an actual video (i.e. not downsampled by our method), we use *Highway* video sequence to upscale it from CIF (352×288) to 4CIF (704×576) resolution.

For objective evaluation of the frame reconstruction performance of the proposed method, we use peak signal-to-noise ratio (PSNR) which is defined for the pixel intensity range of $[0, 255]$ as:

$$\mathrm{PSNR}(\hat{\mathbf{f}}) = 10\log_{10}\left(\frac{255^2 N_x^f N_y^f C}{\|\mathbf{f}-\hat{\mathbf{f}}\|^2}\right), \tag{23}$$

where $\mathbf{f}$ and $\hat{\hat{\mathbf{f}}}$, are the ground-truth and reconstructed frames, respectively. Also the accuracy of blur estimation is evaluated by normalized mean square error (NMSE) defined as:

$$\mathrm{NMSE}(\hat{\mathbf{h}}) = \frac{\|\mathbf{h}-\hat{\mathbf{h}}\|^2}{\|\mathbf{h}\|^2}, \tag{24}$$

where $\mathbf{h}$ and $\hat{\hat{\mathbf{h}}}$, are the original and estimated blurs, respectively. The goal is to obtain high frame-PSNR and low PSF-NMSE values.

For the first experiment, the *City* video sequence (one frame of which is shown in Fig. 3(a)) in 4CIF resolution is used. It contains many structures at different scales, some of which are smaller than the support of applied blur. This sequence is blurred by applying a Gaussian PSF of size 15×15 with the standard deviation of 1.2, downsampled by a factor of 2, and corrupted by Gaussian noise with SNR of $30dB$. The Bicubically-upsampled LR frame and the applied PSF are shown in Figs. 3(b) and (c), respectively. The size of SR temporal window is 5 (with 2 frames forward and 2 frames backward). To estimate the blur(s), the frames are first upsampled using the NUI method, then the luma channel of each frame is smoothed out (Fig. 3(d)), its gradient magnitude is calculated (Fig. 3(e)), and its salient edges not belonging to structures finer than the kernel support are extracted (Fig. 3(f)). The reconstructed frame using *3D-ISKR* [38] deblurred by [39] is shown in Fig. 3(g) with PSNR of $28.9dB$ (after border removal). The estimated HR frame and PSF using our method are shown in Figs. 3(h) and (i) with frame-PSNR of $32.4dB$ and PSF-NMSE of 0.01. Both subjective and object

[5]Full High Definition.

(a) (b)

Fig. 7. (a) One frame of Highway video sequence; (b) The result of proposed method. The resolution is improved (e.g. see the green sign) and the noise is reduced. Please see the video at **http://lyle.smu.edu/~rajand/Video_SR/**.

TABLE I
PSNR COMPARISON BETWEEN THE PROPOSED METHOD AND THE STATE OF THE ART

SR Methods	City	Mobile	Foreman	Bus
Proposed	35.7	26.6	35.5	28.7
3D-ISKR [39] & Debluring [46]	29.1	22.5	34.5	error
Video Enhancer [44]	30.4	22.8	34.1	26.2
Fast Upsampler [43]	29.6	22.5	34.7	26.1
Bicubic	28.1	21.1	32.6	24.6

comparisons confirm superiority of the proposed method over *3D-ISKR*.

For the second experiment, the *Mobile* sequence in CIF resolution is chosen (Fig. 4(a)). This sequence is blurred by a 5×5 out-of focus PSF with size of 15×15, downsampled by applying a SR factor of 2 and contaminated by additive Gaussian noise with SNR of $40dB$. One Bicubically-upsampled LR frame and the original PSF are shown in Figs. 4(b) and (c), respectively. To reconstruct each HR frame, we use a window of length 5. Fig. 4(d) shows the reconstruction result of *3D-ISKR* [38] deblurred by [39] with PSNR of $21.2dB$. Also, Figs. 4(e) and (f) demonstrate the estimated frame and PSF using our method with frame-PSNR of $22.7dB$ and PSF-NMSE of 0.015.

As the third experiment, the *Foreman* video sequence (Fig. 5(a)) is blurred with a 45-degree motion blur of size 15×15 with the support size of 9×9, downsampled by a SR factor of 2, and contaminated by 30 dB noise. One Bicubically-upsampled LR frame and the PSF are presented in Figs. 5(b) and (c), respectively. The reconstruction result of [38] is demonstrated in Fig. 5(d) with PSNR of $32dB$. The reconstructed image and the estimated PSF obtained by our proposed method are shown in Figs. 5(e) and (f), respectively with PSNR of $33.1dB$ and NMSE of 0.08.

Table I summarizes the PSNR values from our method compared to those from *3D-ISKR* [38] (deblurred by [45]), *Video Enhancer* [43], *Fast Upsampler* [42], and Bicubic for different video sequences. In all experiments, the proposed method outperforms other methods with significant PSNR differences.

Fig. 6 demonstrates the performance of our SR method versus Bicubic for some popular 1080p video sequences having complicated motions. The resolution improvement is clearly observable in all cases. The masking operation has successfully suppressed motion artifacts in occluded regions (e.g. around the runners) and deformable area (e.g. torch flame, stream of water).

Now we evaluate the proposed method using a real-life low quality and noisy video sequence. Fig. (a) displays one frame of the Highway sequence in CIF resolution. This is a fast moving scene and so is challenging for SR processing. The resulting 4CIF video using our proposed method is shown in (b). The resolution is visibly improved as can be seen for instance from the green signboard. Also the noise level is significantly suppressed.

VI. CONCLUSION

A method for blind deconvolution and super resolution from one low-resolution video is introduced in this paper. The complicated nature of motion fields in real-life videos make the frame and blur estimations a challenging problem. To estimate the blur(s), the input frames are first upsampled using non-uniform interpolation (NUI) SR method assuming that the blurs are either identical or have slow variations over time. Then the blurs are determined iteratively from some enhanced edges in the upsampled frames. After completion of blur estimation, the reconstructed frames are discarded and a non-blind iterative SR process is performed to obtain the final reconstructed frames using the estimated blur(s). A masking operation is applied during each iteration of the final frame reconstruction to successively suppress artifacts resulted by inaccurate motion estimation. Comparison is made with the state of the art and the superior performance of our proposed method is confirmed through different experiments.

REFERENCES

[1] E. Faramarzi, V. R. Bhakta, D. Rajan, and M. P. Christensen, "Super resolution results in PANOPTES, an adaptive multi-aperture folded architecture," in *Proc. 17th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 2833–2836.

[2] E. Faramarzi, D. Rajan, and M. P. Christensen, "Unified blind method for multi-image super-resolution and single/multi-image blur deconvolution," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2101–2114, Jun. 2013.
[3] S. Borman and R. L. Stevenson, "Spatial resolution enhancement of low-resolution image sequences: A comprehensive review with directions for future research," Dept. Elect. Eng., Univ. Notre Dame, Notre Dame, IN, USA, Tech. Rep., Jul. 1998.
[4] S. Borman and R. L. Stevenson, "Super-resolution from image sequences—A review," in *Proc. Midwest Symp. Circuits Syst.*, Notre Dame, IN, USA, Aug. 1998, pp. 374–378.
[5] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.
[6] R. R. Schultz, L. Meng, and R. L. Stevenson, "Subpixel motion estimation for super-resolution image sequence enhancement," *J. Vis. Commun. Image Represent.*, vol. 9, no. 1, pp. 38–50, Mar. 1998.
[7] A. M. Tekalp, *Digital Video Processing* (Prentice Hall Signal Processing Series). Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.
[8] Y. Caspi and M. Irani, "Spatio-temporal alignment of sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 11, pp. 1409–1424, Nov. 2002.
[9] O. Shahar, A. Faktor, and M. Irani, "Space-time super-resolution from a single video," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 3353–3360.
[10] V. Cheung, B. J. Frey, and N. Jojic, "Video epitomes," *Int. J. Comput. Vis.*, vol. 76, no. 2, pp. 141–152, 2008.
[11] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, 2006.
[12] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph.*, vol. 27, no. 3, p. 73, 2008.
[13] S. Cho and S. Lee, "Fast motion deblurring," *ACM Trans. Graph.*, vol. 28, no. 5, 2009, Art. ID 145.
[14] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *Proc. 11th Eur. Conf. Comput. Vis.*, 2010, pp. 157–170.
[15] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 370–375, Mar. 1998.
[16] Y.-L. You and M. Kaveh, "A regularization approach to joint blur identification and image restoration," *IEEE Trans. Image Process.*, vol. 5, no. 3, pp. 416–428, Mar. 1996.
[17] Y.-L. You and M. Kaveh, "Blind image restoration by anisotropic regularization," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 396–407, Mar. 1999.
[18] F. Šroubek and J. Flusser, "Multichannel blind deconvolution of spatially misaligned images," *IEEE Trans. Image Process.*, vol. 14, no. 7, pp. 874–883, Jul. 2005.
[19] F. Šroubek, G. Cristóbal, and J. Flusser, "A unified approach to super-resolution and multichannel blind deconvolution," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2322–2332, Sep. 2007.
[20] F. Šroubek, J. Flusser, and M. Šorel, "Superresolution and blind deconvolution of video," in *Proc. 19th Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2008, pp. 1–4.
[21] C. Liu and D. Sun, "A Bayesian approach to adaptive video super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 209–216.
[22] C. Liu and D. Sun, "On Bayesian adaptive video super resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 346–360, Feb. 2014.
[23] E. Shechtman, Y. Caspi, and M. Irani, "Space-time super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 531–545, Apr. 2005.
[24] H. Takeda and P. Milanfar, "Removing motion blur with space–time processing," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2990–3000, Oct. 2011.
[25] I. E. Richardson, *The H.264 Advanced Video Compression Standard*. New York, NY, USA: Wiley, 2011.
[26] C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.
[27] F. Šroubek and P. Milanfar, "Robust multichannel blind deconvolution via fast alternating minimization," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1687–1700, Apr. 2012.
[28] R. C. Hardie, K. J. Barnard, J. G. Bognar, E. E. Armstrong, and E. A. Watson, "High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system," *Opt. Eng.*, vol. 37, no. 1, pp. 247–260, 1998.
[29] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1187–1193, Aug. 2001.
[30] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-94-125, 1994.
[31] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, no. 1, pp. 227–238, 1996.
[32] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. IEEE CVPR*, Jun. 2011, pp. 233–240.
[33] C. Wang, L. Sun, P. Cui, J. Zhang, and S. Yang, "Analyzing image deblurring through three paradigms," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 115–129, Jan. 2012.
[34] J. H. Money and S. H. Kang, "Total variation minimizing blind deconvolution with shock filter reference," *Image Vis. Comput.*, vol. 26, no. 2, pp. 302–314, 2008.
[35] S. Osher and L. I. Rudin, "Feature-oriented image enhancement using shock filters," *SIAM J. Numer. Anal.*, vol. 27, no. 4, pp. 919–940, 1990.
[36] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Comput. Vis. (ICCV)*, Washington, DC, USA, Jan. 1998, pp. 839–846.
[37] N. Joshi, R. Szeliski, and D. Kriegman, "PSF estimation using sharp edge prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.
[38] H. Takeda, P. Milanfar, M. Protter, and M. Elad, "Super-resolution without explicit subpixel motion estimation," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 1958–1975, Sep. 2009.
[39] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, 2007, Art. ID 70.
[40] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L_0 gradient minimization," *ACM Trans. Graph.*, vol. 30, no. 6, 2011, Art. ID 174.
[41] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 2657–2664.
[42] Q. Shan, Z. Li, J. Jia, and C.-K. Tang, "Fast image/video upsampling," *ACM Trans. Graph.*, vol. 27, no. 5, 2008, Art. ID 153.
[43] *Video Enhancer v1.9.10*. [Online]. Available: http://www.infognition.com/VideoEnhancer/
[44] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using sub-pixel displacements," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 1988, pp. 742–746.
[45] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1964–1971.

Esmaeil Faramarzi received the B.S. and M.S. degrees from the Amirkabir University of Technology, Tehran, Iran, in 2000 and 2003, respectively, and the Ph.D. degree from Southern Methodist University, Dallas, TX, USA, in 2012. From 2003 to 2008, he was a Research Faculty Member with the Iranian Research Institute for Information Science and Technology, Tehran, where he managed several research projects and developed automation software for image analysis and text recognition of scanned university dissertations and books written in Farsi and Latin languages. He is currently a Staff Engineer with Samsung Research America, Richardson, TX. He made some contributions to the Display Adaptation part of ISO/IEC Green MPEG Standard for energy-efficient video consumption.

Dinesh Rajan received the B.Tech. degree in electrical engineering from IIT IIT, Madras, and the M.S. and Ph.D. degrees in electrical and computer engineering from Rice University, Houston, TX. He joined the Electrical Engineering Department, Southern Methodist University (SMU), Dallas, TX, in 2002, as an Assistant Professor, where he is currently the Department Chair and Cecil and Ida Green Professor with the Electrical Engineering Department. His current research interests include communications theory, wireless networks, information theory, and computational imaging. He received the NSF CAREER Award for his work on applying information theory to the design of mobile wireless networks. He is a recipient of the Golden Mustang Outstanding Faculty Award and the Senior Ford Research Fellowship from SMU.

Felix C. A. Fernandes received the M.S. (Hons.) degree in computer science from the University of Kansas, Lawrence, in 1997, and the Ph.D. degree in electrical engineering from Rice University, in 2001. His dissertation was on directional, shift-insensitive, and complex wavelet transforms with controllable redundancy. From 2001 to 2005, he was a member of the Technical Staff with the Video and Image Processing Branch, R&D Center, Texas Instruments, where he was the official ISO/IEC delegate and contributed to the standardization of MPEG-4. His research spanned image/video codecs, transcoding, error resilience, and digital camera pipelines. From 2005 to 2009, he was with WiQuest Communications, an ultrawideband startup, where he architected a proprietary video codec that was productized as a wireless docking station by Toshiba in their Portege R400 laptop. In 2009, he designed and implemented image/video algorithms with Ambrado, a startup targeting HD codecs on a parallel processor in Toshiba/Ikegami studio cameras. Since 2010, he has been a Director of media-processing research and standardization with Samsung Research America, Dallas. His team has developed and standardized new technology for video coding and fingerprinting, image search, multimedia transport, and energy-efficient processing. He holds 14 issued patents, several pending, and over 40 publications in peer-reviewed journals and conferences. He is the Co-Chair of the ISO/IEC Green MPEG AdHoc Group, an Editor of the MPEG Green Metadata Standard, and a member of the Eta Kappa Nu Engineering Honor Society.

Marc P. Christensen received the B.S. degree in engineering physics from Cornell University, in 1993, the M.S. degree in electrical engineering from George Mason University, in 1998, and the Ph.D. degree in electrical and computer engineering from George Mason University, in 2001. From 1991 to 1998, he was a Staff Member and Technical Leader with the Sensors and Photonics Group (now part of Northrop Grumman Mission Systems), BDM. His work ranged from developing optical signal processing and VCSEL-based optical interconnection architectures, to infrared sensor modeling, simulation, and analysis. In 1997, he co-founded Applied Photonics, a free-space optical interconnection module company. His responsibilities included hardware demonstration for the DARPA MTO FAST-Net, VIVACE, and ACTIVE-EYES programs, each of which incorporated precision optics, microoptoelectronic arrays, and micromechanical arrays into large system level demonstrations. In 2002, he joined Southern Methodist University. He has contributed to large industry/university consortia centered on integrated photonics, such as the DARPA PhASER and CIPhER programs. He has co-authored over 100 journal and conference papers. He holds two patents in the field of free space optical interconnections, one pending in the field of integrated photonics, and four pending in the field of computational imaging. In 2010, he was selected as the inaugural Bobby B. Lyle Professor of Engineering Innovation and serves as the Dean Ad Interim of the Lyle School of Engineering. In 2008, he was recognized for outstanding research with the Gerald J. Ford Research Fellowship.