Rolling Shutter Super-Resolution

Abhijith Punnappurath A.P. Vijay Rengarajan A.N. Rajagopalan
Department of Electrical Engineering, Indian Institute of Technology Madras, India

Introduction

- CMOS sensors are prevalent nowadays, especially in mobile phones, due to their lower cost and power consumption
- Sequential exposure of rows of sensors in CMOS cameras leads to rolling shutter (RS) effect
- Super-resolution (SR) from such images is a challenging task
- First attempt for the task of SR in CMOS cameras
- An RS-SR observation model that explains the image formation process in CMOS cameras is proposed
- Given multiple low-resolution (LR) images that are RS affected, a unified framework is developed to obtain an undistorted and super-resolved image by alternating between solving for the underlying high-resolution (HR) image and the row-wise motion
- Assumption: The first LR image is free from RS effect and has only undergone a downsampling operation with respect to the HR image

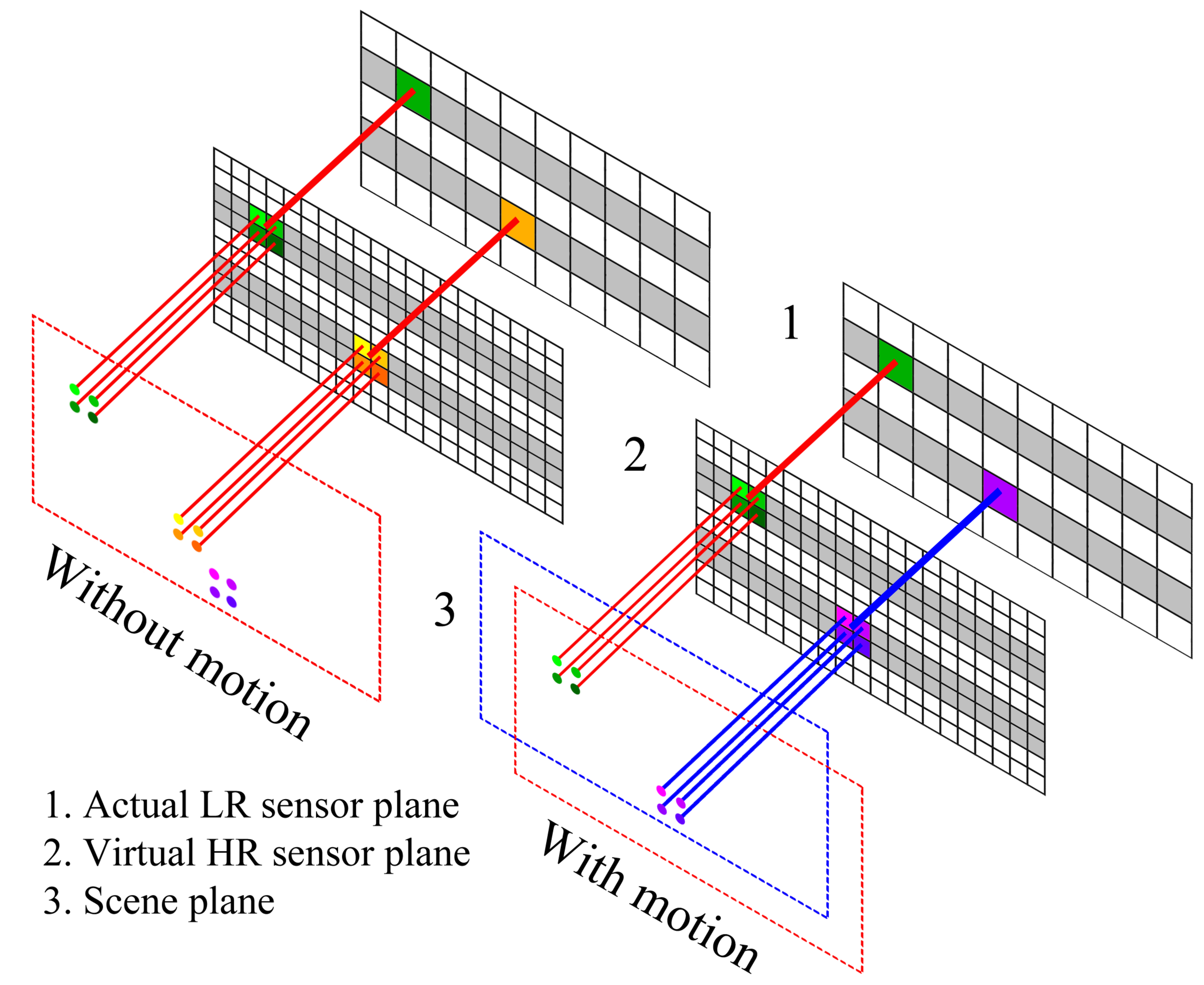
RS-SR Image Formation Model

- The classical SR equation for a CCD camera

$$\mathrm{g} = \mathrm{D}_\epsilon \mathrm{W} \mathrm{f} \qquad (1)$$

$\epsilon \quad (>1)$	: super-resolution factor
$\mathrm{g} \in \mathbb{R}^{MN\times 1}$	: LR image of size $M \times N$ lexicographically ordered
$\mathrm{f} \in \mathbb{R}^{\epsilon^2 MN\times 1}$	: HR image of size $\epsilon M \times \epsilon N$ lexicographically ordered
$\mathrm{W} \in \mathbb{R}^{\epsilon^2 MN\times \epsilon^2 MN}$	: warping matrix that multiplies f to produce its warped instance
$\mathrm{D}_\epsilon \in \mathbb{R}^{MN\times \epsilon^2 MN}$	: decimation matrix which averages ϵ^2 neighboring pixels in the HR image

- **Image formation model for an RS camera - static versus moving**



- The *virtual* HR sensor plane is the HR representation of the scene that an HR camera would have captured
- It is this HR image that is to be recovered
- For an SR factor of 2, a *pair* of rows in the HR plane experience the same motion
- For an SR factor of ϵ, this corresponds to a block of ϵ rows in the virtual HR sensor plane having the same motion associated with them
- Unlike in a GS camera where all rows of W are associated with a *single* camera motion, in RS cameras, the motion varies depending on which particular block of rows in the HR image the pixel belongs to
- (1) can be rewritten for a CMOS camera as

$$\mathrm{g} = \mathrm{D}_\epsilon \mathcal{W} \mathrm{f} \qquad (2)$$

where $\mathcal{W}$ is the warping matrix that multiplies f to produce an RS image
- There are M warps associated with $\mathcal{W}$ as against a single warp for W

Optimization Problem

- **Aim:** Recover f given K LR images $\{\mathrm{g_k}\}$, where $\mathrm{g_k} = \mathrm{D}_\epsilon \mathcal{W}_\mathrm{k} \mathrm{f}$, for $k = 1$ to K
- Alternating minimization scheme to solve for the two unknowns f and $\mathcal{W}_\mathrm{k}$
- The minimization sequence $(\mathrm{f}_p, \mathcal{W}_{\mathrm{k}_p})$, where p indicates the iteration number, can be built by alternating between two minimization subproblems
- Starting with an initial estimate f_0 (obtained by upsampling the first LR image), the two alternating steps are: step 1) estimate $\mathcal{W}_{\mathrm{k}_p}$ using the previous iterate f_{p-1}, step 2) use the current estimate $\mathcal{W}_{\mathrm{k}_p}$ to compute f_p
- **Warp Estimation**
  - Estimate a single camera pose/warp from a discrete camera pose space S for every row i, where $1 \leq i \leq M$, in the LR images $\{\mathrm{g_k}\}_{k=2}^{K}$
  - The cost function is formulated such that a few camera poses around the actual pose are selected from the search space for each row, and the centroid of these poses yields the true motion for that row

$$\widehat{w}_{\mathrm{k}_p}^{(i)} = \underset{w_\mathrm{k}^{(i)}}{\mathrm{argmin}}\{||\mathrm{g_k}^{(i)} - \mathrm{D}_\epsilon^{(i)} \mathcal{F}_{p-1}^{(\epsilon i)} w_\mathrm{k}^{(i)}||_2^2 + \lambda ||w_\mathrm{k}^{(i)}||_1\} \qquad (3)$$

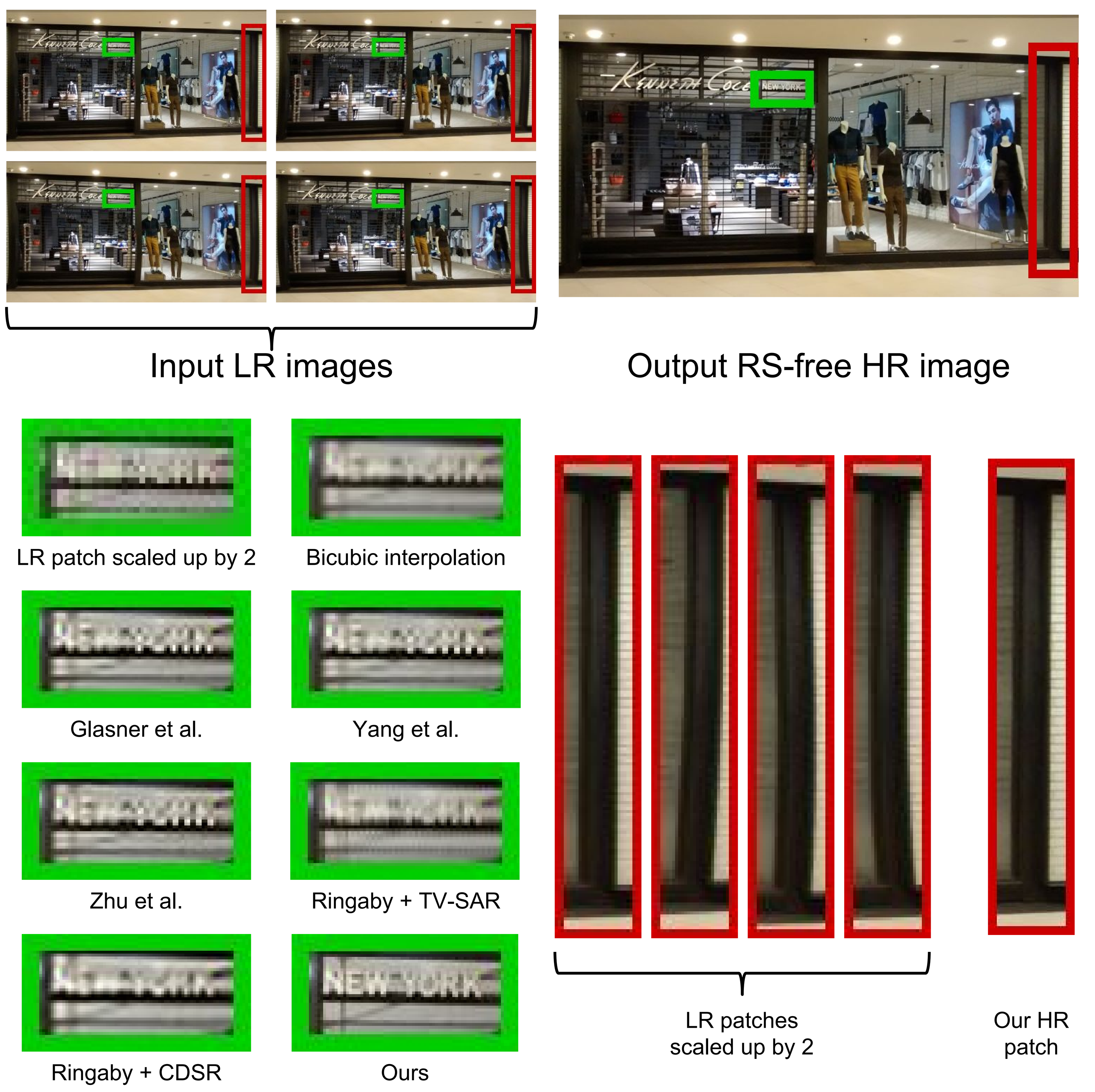
$$\text{subject to} \quad w_\mathrm{k}^{(i)} \geq 0$$

  - $\mathrm{g_k}^{(i)}$ denotes the ith row of the LR image $\mathrm{g_k}$ and $w_\mathrm{k}^{(i)}$ is its corresponding weight vector of size $|\mathrm{S}| \times 1$ which chooses the required set of poses from the search space S
  - Since $w_\mathrm{k}^{(i)}$ is sparse, l_1-norm with non-negativity is imposed so as to choose a sparse set of camera poses with corresponding weights to calculate the centroid
  - The weighted average of the rotations and translations in the search space is found to give the centroid pose; $\mathrm{R}_c = \widehat{w}_{\mathrm{k}_p}^{(i)} \circ \{\mathrm{R}_j\}_{j=1}^{|S|}$ and $\mathrm{T}_c = \widehat{w}_{\mathrm{k}_p}^{(i)} \circ \{\mathrm{T}_j\}_{j=1}^{|S|}$, where $\circ$ represents element-wise multiplication
- **HR Image Estimation**

$$\widehat{\mathrm{f}}_p = \underset{\mathrm{f}}{\mathrm{argmin}}\{\sum_{k=1}^{K} ||\mathrm{D}_\epsilon \mathcal{W}_{\mathrm{k}_p} \mathrm{f} - \mathrm{g_k}||_2^2 + \alpha \mathrm{f}^T \mathrm{L} \mathrm{f}\} \qquad (4)$$

  - L is the discrete form of the variational prior

Experiments



http://www.ee.iitm.ac.in/∼ee10d038/RSSR.html ee10d038@ee.iitm.ac.in