THE UNIVERSITY of WISCONSIN MADISON

Piecewise Planar Super-Resolution for 3D Scene

Gaoang Wang, Yu Hen Hu, Hongrui Jiang

University of Wisconsin – Madison

Dept. Electrical and Computer Engineering

Madison, WI 53706

hu@engr.wisc.edu

GlobalSIP 2015, Orlando, FL

Outline

- Background of super-resolution
- Related work
- Model introduction
- Proposed method
- Experiments and results
- Conclusion

Background of super-resolution

- Reconstruct a high resolution image with one or multiple low resolution images.
- Compensation for the limit of the camera device.
- Use for image or video enhancement.

Related work

- Single-view image resolution

  J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation."

  Disadvantage: cannot take advantages from multiviews.

- Multi-view image resolution

  S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust super-resolution."

  A. V. Bhavsar and A. Rajagopalan, "Resolution enhancement in multi-image stereo."

  Disadvantage: some algorithms only solve for 2D images; or largely depends on the accurate depth estimation to reconstruct for 3D scene.

Model introduction

$$Y_k = DHW_k I_{X_0}, k = 0, 1, 2, \cdots, K-1.$$

- I_{x0} is the reference image frame; W_k is some warp transform (the transform may be different for different parts of the image since it is a 3D scene image); H is the blur kernel; D is the down-sampling operator; Y_k is the low resolution image of the k-th frame.
- The question is how to reconstruct the high resolution image from multiview low resolution images.
- **Solution**: Segment images into piecewise planar parts and estimate the W_k transform for each part!

Proposed method

- Depth estimation by graph cut

  Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts.”

  The depth information is only used for plane segmentation.

- Plane segmentation

  We label each pixel point by minimizing the following object function,

$$l_{p_0} = \arg\min_{l} \left(|1 - n_l^T p_0|^2 + \lambda \sum_{p \in C(p_0)} |l - l_p|^2 \right)$$

Proposed method

- Homography and warp transform estimation

Here we update the pixel label by minimizing the intensity difference for the same point from different frames,

$$l_{p_0} = \arg \min_{l=1,2,\ldots,m_{opt}} \left(\sum_{k=1,2,\ldots,K-1} \sum_{p_l \in C(p_0)} d_k(p_l) \right)$$

where the intensity difference is defined below,

$$d_k(p_\pi) = |I(p_\pi) - I(H_{\pi,k} p_\pi)|^2 V_{p_\pi} / N_{C(p_\pi)}$$

Finally after label updated, we estimate the transform by labeled feature points.

Proposed method

- Image reconstruction

  Project the reconstructed image into image subspace by using Tikhonov regularization,

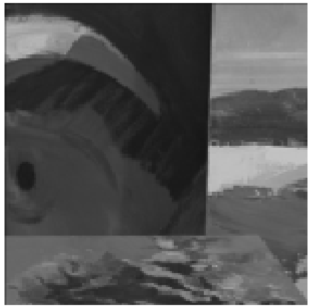
$$\hat{I}_{X_0} = \arg\min_{I_{X_0}}\{\sum_{k=1}^{K} ||QM_k I_{X_0} - QY_k||_2^2 + \alpha||LI_{X_0}||_2^2\}$$

  Here, we use a diagnal matrix Q to remove the occlusion and artifacts between plane boundaries. We set the diagnal elements equal to zero if they are invisible.

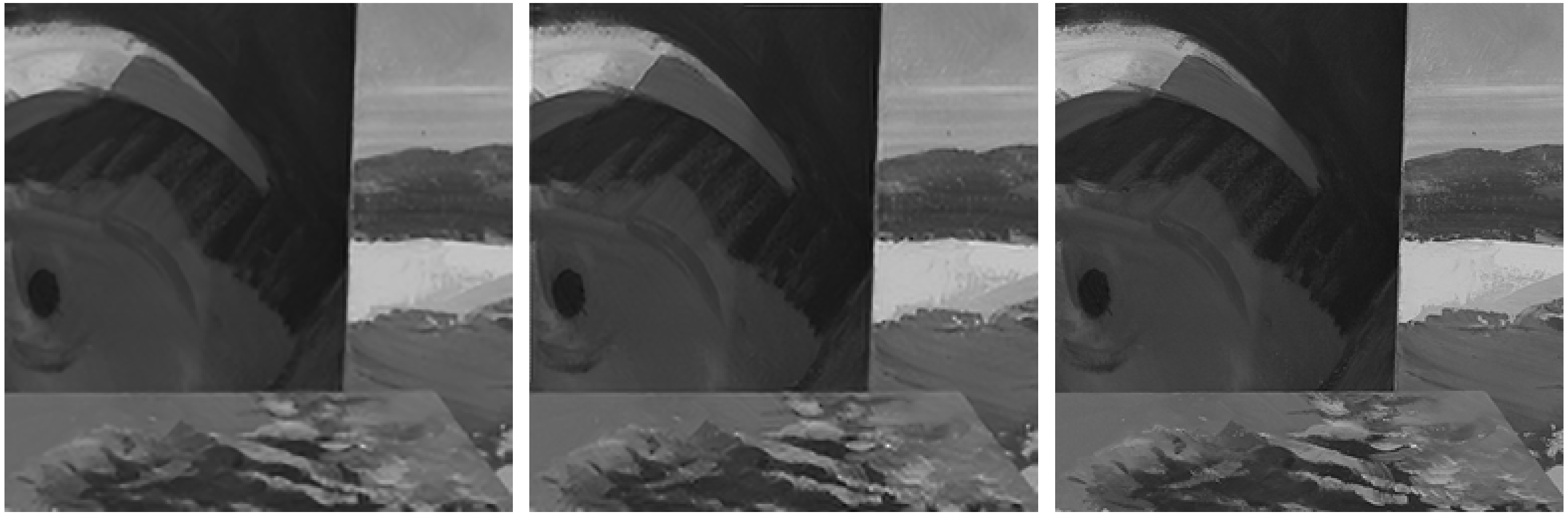
Experiments and results

- "Bull" images from Middlebury dataset

  Four low resolution images are shown below.

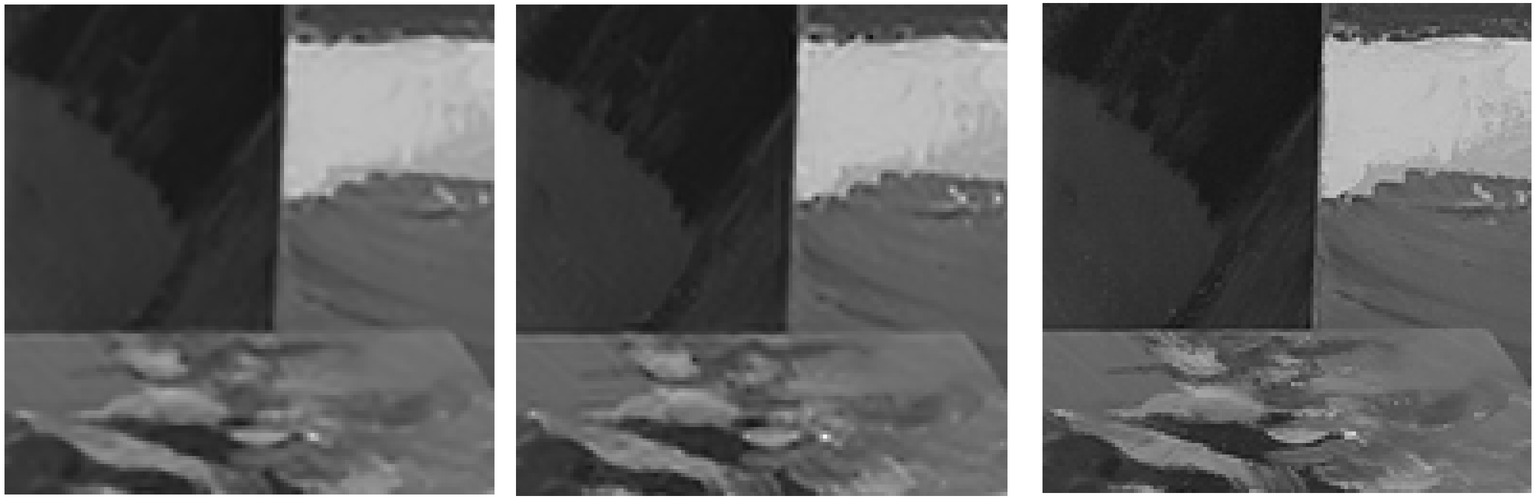



Experiments and results



From left to right: Single-view SR (PSNR = 35.1667), Proposed method (PSNR = 35.3245), Ground truth.

Experiments and results



Cropped image results. From left to right: Single-view SR, Proposed method, Ground truth.

Experiments and results

- Kitchen Scene

  Four low resolution images are shown below.




Experiments and results

From left to right: Single-view SR (PSNR = 28.0398), Proposed method (PSNR = 29.3387), Ground truth.

Experiments and results

Cropped image results. From left to right: Single-view SR, Proposed method, Ground truth.

Conclusion and future work

- With combined information from multiview low resolution images, we can get a better performance than single view reconstruction.

- We will look for solutions that combines the advantages from both single-view and multiview super-resolution.