

Joint Bilateral Upsampling

Johannes Kopf
University of Konstanz

Michael F. Cohen
Microsoft Research

Dani Lischinski
The Hebrew University

Matt Uyttendaele
Microsoft Research

Abstract

Image analysis and enhancement tasks such as tone mapping, colorization, stereo depth, and photomontage, often require computing a solution (e.g., for exposure, chromaticity, disparity, labels) over the pixel grid. Computational and memory costs often require that a smaller solution be run over a downsampled image. Although general purpose upsampling methods can be used to interpolate the low resolution solution to the full resolution, these methods generally assume a smoothness prior for the interpolation.

We demonstrate that in cases, such as those above, the available high resolution input image may be leveraged as a prior in the context of a joint bilateral upsampling procedure to produce a better high resolution solution. We show results for each of the applications above and compare them to traditional upsampling methods.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture

Keywords: bilateral filter, upsampling

1 Introduction

A variety of new image analysis and image processing methods, both automatic and user guided, have recently been demonstrated in the computer graphics and computer vision literature. These include stereo depth computations [Scharstein and Szeliski 2002], image colorization [Levin et al. 2004; Yatziv and Sapiro 2006], tone mapping of high dynamic range (HDR) images [Reinhard et al. 2005], and applications of minimal graph cuts to image composition [Agarwala et al. 2004]. All of these methodologies share a common problem of finding a global *solution*: a piecewise smooth function describing some value of interest (depth, chromaticity, exposure, label, etc.) over the pixel grid of the input image.

Digital images continue to grow in size from one quarter million pixel video frames to multi-Megapixel digital photos, to recent Gigapixel images arising from specialized cameras [Flint 2007] and from stitching multiple images into a panorama [Kopf et al. 2007]. Such high resolutions pose a difficult challenge for the methods cited above, which typically require at least linear time and, more importantly, linear space to compute a global solution. Thus, in order to operate on such high resolution images, they must first be downsampled to a lower resolution to make the computation tractable. This is particularly essential for interactive applications.

Once a solution is available for the smaller downsampled image, the question then becomes how to upsample the solution to the full

original resolution of the input image. Upsampling is a fundamental image processing operation, typically achieved by convolving the low-resolution image with an interpolation kernel, and resampling the result on a new (high-resolution) grid. Wolberg [1990] provides a good survey of common interpolation kernels. Images upsampled in this manner typically suffer from blurring of sharp edges, because of the smoothness prior inherent in the linear interpolation filters.

However, for the applications cited above, additional information is available in the form of the original high-resolution input image. Ignoring this information and relying on the smoothness prior alone is clearly not the best strategy. We propose to leverage the fact that we have a high-resolution image in addition to the low-resolution solution. In particular, we demonstrate that a *joint bilateral upsampling* (JBU) operation can produce very good full resolution results from solutions computed at very low resolutions. We show results for stereo depth, image colorization, adaptive tone mapping, and graph-cut based image composition.

2 Bilateral Filters

The bilateral filter is an edge-preserving filter, originally introduced by Tomasi and Manduchi [1998]. It is related to broader class of non-linear filters such as anisotropic diffusion and robust estimation [Barash 2002; Durand and Dorsey 2002; Elad 2002]. The bilateral filter uses both a *spatial* (or *domain*) filter kernel and a *range* filter kernel evaluated on the data values themselves. More formally, for some position p , the filtered result is:

$$J_p = \frac{1}{k_p} \sum_{q \in \Omega} I_q f(||p - q||) g(||I_p - I_q||), \quad (1)$$

where f is the spatial filter kernel, such as a Gaussian centered over p , and g is the range filter kernel, centered at the image value at p . Ω is the spatial support of the kernel f , and k_p is a normalizing factor, the sum of the $f \cdot g$ filter weights. Edges are preserved since the bilateral filter $f \cdot g$ takes on smaller values as the range distance and/or the spatial distance increase.

Recently we have seen the introduction of *joint* (or *cross*) bilateral filters in which the range filter is applied to a second *guidance image*, \tilde{I} , for example, when trying to combine the high frequencies from one image and the low frequencies from another [Petschnigg et al. 2004; Eisemann and Durand 2004]. Thus,

$$J_p = \frac{1}{k_p} \sum_{q \in \Omega} I_q f(||p - q||) g(||\tilde{I}_p - \tilde{I}_q||). \quad (2)$$

The only difference to (1) is that the range filter uses \tilde{I} instead of I .

2.1 Previous Work

The bilateral filter has been used previously for various image processing tasks. Durand and Dorsey [2002] applied the bilateral filter to HDR tone mapping and also described a fast approximation, which was recently improved upon [Paris and Durand 2006; Weiss 2006].

Ramanath and Snyder [2003] used the bilateral filter in the context of demosaicking to improve edge sensitivity. Their method is restricted to Bayer patterns with a fixed small upsampling factor, and does not use a guidance image as we do.

Durand *et al.* [2005] mention using a bilateral filter to up-sample the shading results of a ray tracer. However, no details are given in the paper and no other applications are explored.

Sawhney *et al.* [2001] upsample stereoscopic images where one view has higher resolution than the other. Their method estimates an alignment mapping, and then uses warping and fill-in from neighboring movie frames to upsample the low-resolution image.

3 Joint Bilateral Upsampling

In contrast to general purpose image upsampling, in the problems that we are interested in, additional information is available to us in the form of the original high-resolution input image. Given a high resolution image, \tilde{I} , and a low resolution solution, S , computed for a downsampled version of the image, we propose a simple method that applies a joint bilateral filter to upsample the solution.

The idea is to apply a spatial filter (typically a truncated Gaussian) to the low resolution solution S , while a similar range filter is jointly applied on the full resolution image \tilde{I} . Let p and q denote (integer) coordinates of pixels in \tilde{I} , and p_{\downarrow} and q_{\downarrow} denote the corresponding (possibly fractional) coordinates in the low resolution solution S . The upsampled solution \tilde{S} is then obtained as:

$$\tilde{S}_p = \frac{1}{k_p} \sum_{q_{\downarrow} \in \Omega} S_{q_{\downarrow}} f(||p_{\downarrow} - q_{\downarrow}||) g(||\tilde{I}_p - \tilde{I}_{q_{\downarrow}}||) \quad (3)$$

This is almost identical to eq. (2) with the exceptions that we are constructing a high resolution *solution* as opposed to an image, and operate at two different resolutions simultaneously.

Note, that q_{\downarrow} takes only integer coordinates in the low resolution solution. Therefore the guidance image is only sparsely sampled, and the performance does not depend on the upsampling factor (see Section 5).

4 Applications

In this section we demonstrate the usefulness of the joint bilateral upsampling operation for a variety of applications.

Tone Mapping: With the increasing popularity and utility of High Dynamic Range (HDR) imaging [Reinhard *et al.* 2005], there is a need for *tone mapping* methods to display HDR images on ordinary devices. A variety of such methods have been proposed over the years (see [Reinhard *et al.* 2005] for an extensive survey). Some of these methods produce high-quality results, but require solving a very large system of linear equations [Fattal *et al.* 2002; Lischinski *et al.* 2006]. Although these systems are sparse and may be solved efficiently using multi-resolution solvers [Szeliski 2006], handling today's multi-megapixel images remains a challenge: once the data exceeds the available physical memory, iteratively sweeping over the data results in thrashing.

We apply the joint bilateral upsampling filter as follows. Let I be the low-resolution HDR image, and $T(I)$ the tone mapped image produced by some tone mapping operator. The corresponding low-resolution solution is then defined as the pixelwise quotient $S = T(I)/I$. In other words, the solution is an *exposure map*, which states the amount of exposure correction to be applied at each pixel. Such exposure maps are generally smooth but may have discontinuities along significant image edges [Lischinski *et al.* 2006]. Thus, they are ideal candidates for our upsampling technique. Note that the exposure map may have a single channel (if only the luminance has been adjusted), or multiple channels (to support *arbitrary* tonal manipulations). Figure 2 shows how applying an exposure map upsampled using our technique compares with a number of standard

upsampling methods. The joint bilateral upsampling yields results that are visually and numerically closer to the ground truth.

Colorization: A similar linear system to those in the tone mapping methods cited above arises in the colorization and recoloring method of Levin *et al.* [2004]. Thus, again, processing of very large images is not tractable due to thrashing. This is also true for the more recent colorization method of Yatziv and Sapiro [2006], which does not solve a linear system, but nevertheless iteratively sweeps over the data.

To upsample a low-resolution colorization result, we first convert it into the YIQ color space (or to any other color space separating luminance from chrominance), and then apply our upsampling technique to each of the two chrominance channels. Figure 3 shows the result. As in the tone mapping example, one can see that the JBU avoids having the chromaticity spill over edges in the image.

Stereo Depth: Stereo matching is a fundamental task in image analysis, whose goal is to determine the disparities between pairs of corresponding pixels in two or more images. Many different approaches to stereo matching have been explored over the years (for a comprehensive overview see [Scharstein and Szeliski 2002]). In many of these methods an optimization problem of some sort is solved, yielding a piecewise continuous disparity field over the entire image.

Our technique can be used to upsample low resolution depth maps with guidance from the high resolution photos. Depth maps also have ideal properties for our technique. They are rather smooth, and the discontinuities typically correspond with edges in the image. Figure 4 shows the advantages of our technique in action.

Graph-cut based image operations: Several recent interactive image editing techniques involve finding minimal cuts in graphs. For example, the interactive digital photomontage [Agarwala *et al.* 2004] system uses graph-cut optimization [Boykov *et al.* 2001] to compute the least objectionable seams when fusing together several pre-aligned photographs. The result of the optimization is a label map, indicating for each pixel in the composite which photograph it originates from.

We tested our joint bilateral upsampling technique with an image stitching application. Here, the user constrains a number of pixels to come from a certain input image. The stitching algorithm then computes a label map, which assigns a label to each of the remaining unconstrained pixels, such that the resulting seams are least conspicuous.

This application differs fundamentally from the previous ones, because here we have a *quantized* solution (a discrete number of labels), rather than a continuous one. Furthermore, in this case there are *multiple* full resolution images.

We apply our technique in the following way: suppose we want to compute the label for a pixel. Each low-resolution solution pixel with a non-zero bilateral weight *votes* for its label. The winning label is the one that has aggregated the highest total weight. Figure 5 demonstrates our technique for this application.

5 Performance and Accuracy

The complexity of the joint bilateral upsampling operation is $O(Nr^2)$ where N is the output image size and r is the domain filter radius. The performance is proportional to the output size and not to the upsampling factor, because the domain filter is always applied to the low resolution solution. For all results we have used a 5×5 Gaussian, which is very fast but still has enough spatial support to pull solution values from some distance. Our implementation takes approximately 2 seconds per megapixel of output.

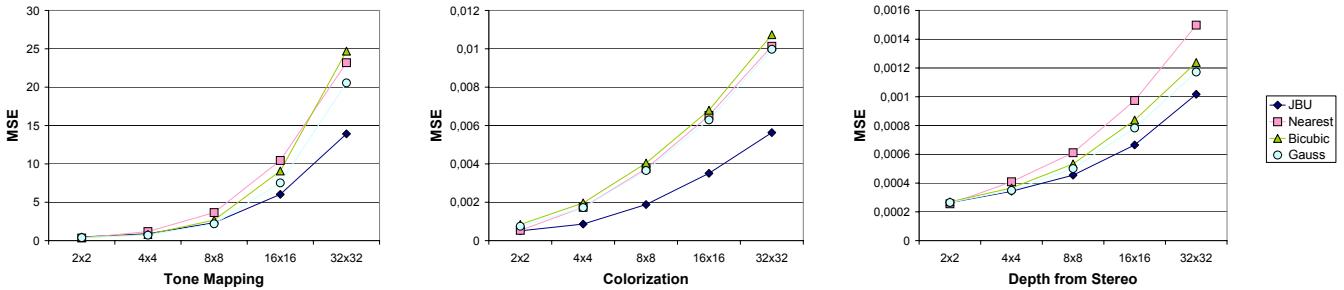


Figure 1: MSE error profiles for various applications and upsampling methods.

This is significantly faster than running the original algorithms on the full resolution images. For example, the tone mapper took 80 seconds for a 3.1 megapixel image, while our upsampling took only 6 seconds to upsample a smaller solution which was computed much faster. The colorization solver of Levin *et al.* [2004] was even slower, and needed several minutes for a megapixel sized image. As noted above, due to the memory issue we cannot run a very high resolution solution so upsampling a low resolution solution is our only way to approach such large images.

The JBU is strictly local with a very small memory footprint. Large images can be computed in a single sweep, where only parts are paged in at any time. We have successfully applied our method to upsample tone mapping solutions for multi-gigapixel images [Kopf *et al.* 2007].

In our experiments, we generally set the domain filter’s Gaussian σ_d to 0.5 with 5×5 support. The range filter Gaussian σ_r is strongly application dependent. The following default values worked well for the images we tried: colorization, stereo depth, and graph-cut labelings used images with color values normalized to the $[0, 1]$ interval. $\sigma_r = 0.1$ worked well on most images. The tone mapping application works with unbounded luminance values. We found that setting σ_r to the standard deviation of the values has always given good results.

Figure 1 shows MSE error profiles for the JBU compared to other upsampling methods. To compute the errors, we computed a full resolution solution (or simply used the full resolution color image or depth map for colorization and stereo depth) as ground truth. We then downsampled by factors of 2, 4, 8, 16, and 32 in each direction. Then, we performed upsampling using various methods and plotted the difference from the ground truth. Our filter performed well at all downsampling levels, and, as expected, the relative improvement increased with each additional level of down sampling.

Not surprisingly, the MSE error increases with the upscaling factor. But in practice it often turns out that the application limits how much one can downsample the problem. The results we show are for solutions on quite tiny downsampled images. Since some of the applications require some UI, you need enough image left to, for example, scribble on the hints for tone mapping or colorization.

6 Conclusion

We have demonstrated the benefits of a joint bilateral upsampling strategy when a high resolution prior is available to guide the interpolation from low to high resolution. The four applications we have shown all improve relative to previous “blind” upsampling methods. We believe this strategy is applicable to a number of other domains within and beyond image processing. For example, a global illumination solution computed over a coarse simplified mesh can be upsampled to a finer mesh. The domain filter’s kernel might

be measured in geodesic distance, while the range kernel would be over the Gaussian Sphere (differences in normal). We look forward to trying the joint bilateral upsampling on this and other problems of interest in computer graphics.

References

- AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. *ACM Transactions on Graphics* 23, 3 (Proc. SIGGRAPH 2004), 294–302.
- BARASH, D. 2002. A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 6, 844–847.
- BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 2001. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 11, 1222–1239.
- DURAND, F., AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics* 21, 3 (Proc. SIGGRAPH 2002), 257–266.
- DURAND, F., HOLZSCHUCH, N., SOLER, C., CHAN, E., AND SILLION, F. X. 2005. A frequency analysis of light transport. *ACM Transactions on Graphics* 24, 3 (Proc. SIGGRAPH 2005), 1115–1126.
- EISEMANN, E., AND DURAND, F. 2004. Flash photography enhancement via intrinsic relighting. *ACM Transactions on Graphics* 23, 3 (Proc. SIGGRAPH 2004), 673–678.
- ELAD, M. 2002. On the bilateral filter and ways to improve it. *IEEE Transactions on Image Processing* 11, 10, 1141–1151.
- FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. *ACM Transactions on Graphics* 21, 3 (Proc. SIGGRAPH 2002), 249–256.
- FLINT, G. 2007. *The Gigapxl Project*.
- KOPF, J., UYTENDAELE, M., DEUSSEN, O., AND COHEN, M. F. 2007. Capturing and viewing gigapixel images. *ACM Transactions on Graphics* 26, 3 (Proc. SIGGRAPH 2007).
- LEVIN, A., LISCHINSKI, D., AND WEISS, Y. 2004. Colorization using optimization. *ACM Transactions on Graphics* 23, 3 (Proc. SIGGRAPH 2004), 689–694.
- LISCHINSKI, D., FARBMAN, Z., UYTENDAELE, M., AND SZELISKI, R. 2006. Interactive local adjustment of tonal values. *ACM Transactions on Graphics* 25, 3 (Proc. SIGGRAPH 2006), 646–653.

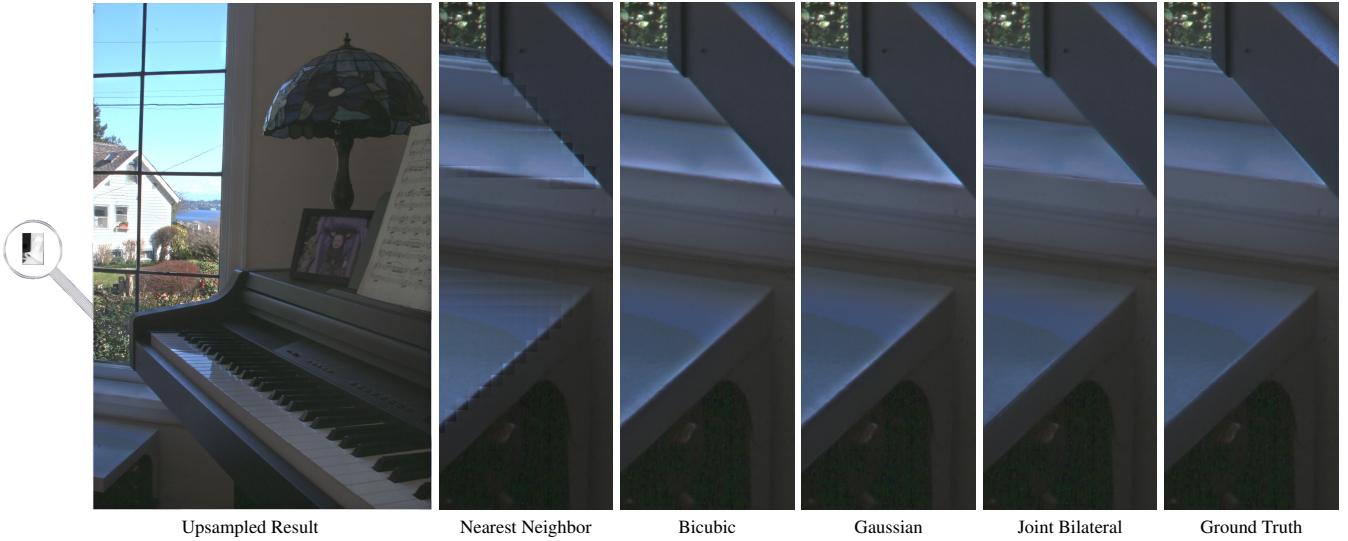


Figure 2: Tone Mapping: The low resolution exposure map solution at left is at scale relative to the final upsampled result next to it. Detail insets to the right show results from Nearest Neighbor, Gaussian Interpolation, Bicubic Interpolation, Joint Bilateral Upsampling, and ground truth based on a full resolution solution. Note that JBU does not exhibit the blocking and halo artifacts of the other upsampling methods.

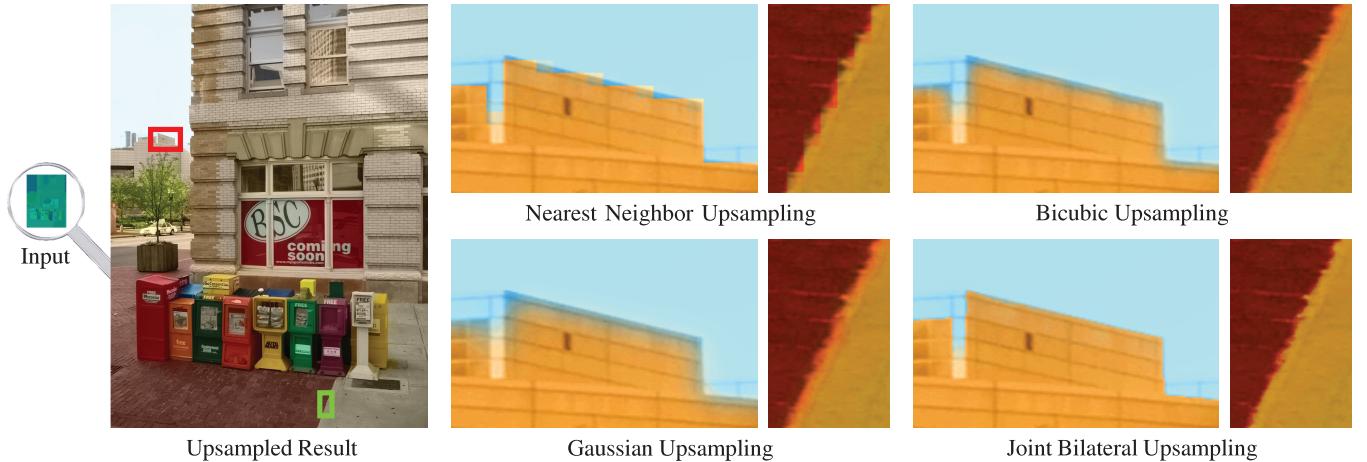


Figure 3: Colorization: The low resolution pseudo-colored chrominance solution is at left. Detail insets corresponding to four different upsampling methods are shown on the right (with increased saturation). Note that the JBU result does not exhibit color spills.

PARIS, S., AND DURAND, F. 2006. A fast approximation of the bilateral filter using a signal processing approach. *Proceedings of ECCV 2006*.

PETSCHNIGG, G., SZELISKI, R., AGRAWALA, M., COHEN, M., HOPPE, H., AND TOYAMA, K. 2004. Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics* 23, 3 (Proc. SIGGRAPH 2004), 664–672.

RAMANATH, R., AND SNYDER, W. 2003. Adaptive demosaicking. *Journal of Electronic Imaging* 12, 4, 633–642.

REINHARD, E., WARD, G., PATTANAIK, S., AND DEBEVEC, P. 2005. *High Dynamic Range Imaging*. Morgan Kaufmann.

SAWHNEY, H. S., GUO, Y., HANNA, K., KUMAR, R., ADKINS, S., AND ZHOU, S. 2001. Hybrid stereo camera: an ibr approach for synthesis of very high resolution stereoscopic image sequences. *Proceedings of SIGGRAPH 2001*, 451–460.

SCHARSTEIN, D., AND SZELISKI, R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision* 47, 1–3, 7–42.

SZELISKI, R. 2006. Locally adapted hierarchical basis preconditioning. *ACM Transactions on Graphics* 25, 3 (Proc. SIGGRAPH 2006), 1135–1143.

TOMASI, C., AND MANDUCHI, R. 1998. Bilateral filtering for gray and color images. *Proceedings of ICCV 1998*, 839–846.

WEISS, B. 2006. Fast median and bilateral filtering. *ACM Transactions on Graphics* 25, 3 (Proc. SIGGRAPH 2006), 519–526.

WOLBERG, G. 1990. *Digital Image Warping*. IEEE Computer Society Press, Los Alamitos, CA.

YATZIV, L., AND SAPIRO, G. 2006. Fast image and video colorization using chrominance blending. *IEEE Transactions on Image Processing* 15, 5, 1120–1129.

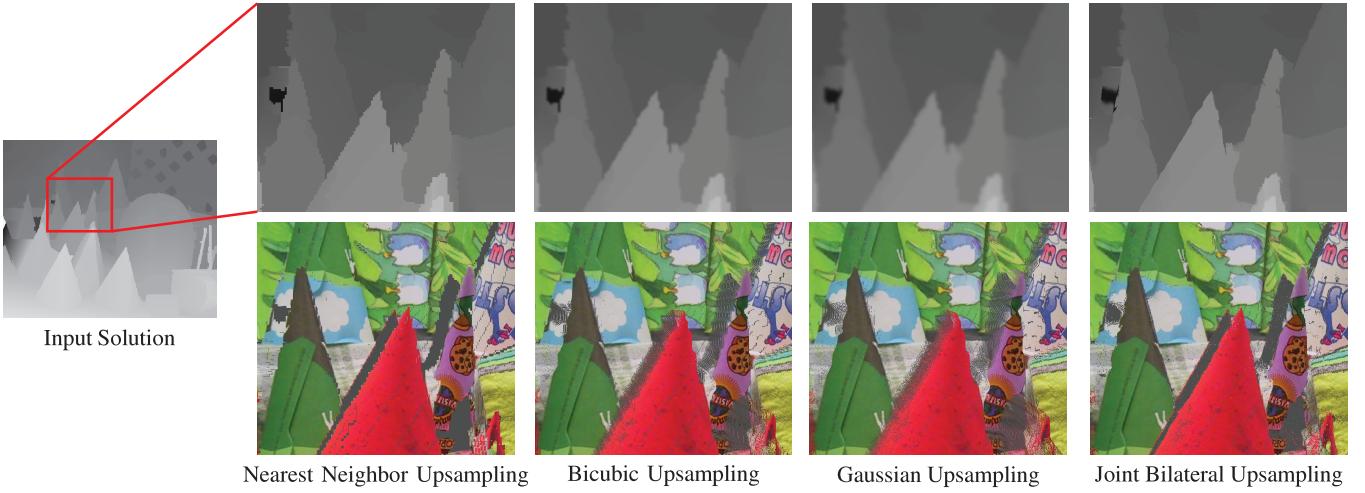


Figure 4: Stereo Depth: The low resolution depth map is shown at left. The top right row shows details from the upsampled maps using different methods. Below each detail image is a corresponding 3d view from an offset camera using the upsampled depth map.



Figure 5: Graph Cut based Photomontage. Upper left: two original aligned images. Upper center: the downsampled images and the resulting labeling solution. Upper right: a composite generated using the JBU-upsampled labeling solution. The label colors are superimposed over the image. The bottom row shows a detail that highlights the advantages of using JBU over standard upsampling methods.