# Quality Improvement for Intermediate Views Using Example-Based Super-Resolution

*Hidenori Sato†, William T. Freeman, and Akira Onozawa*

**Abstract**

We propose an algorithm that improves the quality of intermediate-view images obtained by mapping the average color of multiple-view images onto a visual hull. The algorithm estimates a real image from an intermediate-view image using example-based super-resolution. Estimation is performed using relationships between two kinds of training images: an ideal image whose quality is as good as a real image and an intermediate view image viewed from the same viewpoint. High frequencies of the real image of an intermediate-view are estimated using the relationship between high frequencies of an ideal image and those of a corresponding intermediate-view image. Experimental results show that the quality of the resulting images is almost as good as that of ideal images.

## 1. Introduction

Photorealistic three-dimensional (3-D) data will be important content for broadband and ubiquitous communication services. Such data about real objects will let us richly communicate with friends and relatives over a distance. You will be able to view and manipulate the data from arbitrary viewpoints and feel as if you are actually touching a real object during the communication.

Recently, 3-D data generation algorithms that combine both correct geometric models reconstructed using laser digitizers and multiple-view images have been proposed. On the other hand, we have developed an algorithm that generates 3-D data automatically from only multiple-view images of real objects [2]. However, the rendered results for intermediate viewpoints are blurred.

This paper describes an algorithm that we have developed to improve the quality of the intermediate-views. The algorithm uses the learning-based approach proposed by Freeman et al. [1]. It estimates high frequencies of a real image from an intermediate view using a training set that describes the relationships between high frequencies of an ideal view and those of an intermediate view from the same viewpoint.

† NTT Microsystem Integration Laboratories
Atsugi-shi, 243-0198 Japan
E-mail: hide@aecl.ntt.co.jp

## 2. 3-D data generation algorithm

Figure 1 illustrates our 3-D data generation algo-

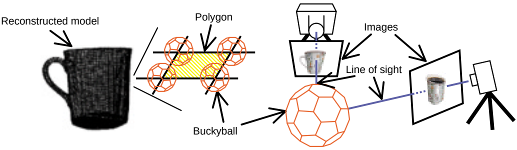


Fig. 1. 3-D data generation algorithm.

rithm [2]. The algorithm first reconstructs a visual hull[*1] as a geometric model of a target object using the shape-from-silhouette approach [3]. Colors in the images are then assigned to all vertices of the reconstructed surface. For each vertex, a polygonized buckyball is virtually assigned and only the colors of input images visible from that vertex are mapped onto the surfaces of the buckyball corresponding to each camera direction. Interpolation for all uncolored polygons is performed by the weighted-average technique using colors of adjacent polygons. Although a buckyball is a truncated icosahedron and has 32 surfaces, we can subdivide the buckyball polygons hierarchically to assign colors of more than 32 images.

Actual rendered views corresponding to input images are photorealistic. However, views from intermediate viewpoints are blurred because of the limits on image resolution, errors in camera calibration, and so on. Rendered views of a coffee mug are shown in Fig. 2, where (a) is the rendered view from an intermediate viewpoint and (b) the rendered view mapping the real image from the same viewpoint. In this paper, we call these the *intermediate-view image* and *ideal image*, respectively. The quality of the intermediate-view images is one of the problems of the algorithm.

*1 A visual hull is a geometric shape obtained using silhouettes of an object as seen from a number of views. (http://graphics.lcs.mit.edu/~wojciech/vh/vhoverview.html)


Fig. 2. (a) Intermediate view-image of a mug. (b) Ideal image of (a) from the same viewpoint.

## 3. Example-based super-resolution

The algorithm we developed to improve the quality of the intermediate-view images is based on example-based super-resolution. This is a learning-based approach that estimates a high-resolution image of an input image using a training set generated from training images, and uses the training set to predict high frequencies of high-resolution images from low-resolution images [1]. An example of an enlargement is shown in Fig. 3. The resulting image is successfully enlarged without loss of quality.

The training set is generated as follows. First, a high-pass filtered image and a mid-pass filtered image are generated for each training image. Then, the filtered images are divided into rectangular patches so that corresponding pairs of patches have the same center coordinates, and a low-resolution patch will be bigger than its corresponding high-resolution patch. The pairs of patches form the training set. We call patches generated from high- and mid-pass filtered images *high-resolution* and *low-resolution patches*, respectively.

When a high-resolution image is estimated, a high-pass filtered image of the input image is enlarged first by a conventional algorithm, such as a cubic spline[*2]. This is the mid-pass filtered image. Then, the image is divided into patches, and the high frequencies of each patch are predicted by searching for the low-resolution patches of the training set that best match the patches of the mid-pass filtered image, and for which the borders of the corresponding high-resolution patch best match the high-resolution detail synthesized so far in the image. Finally, the sum of the resulting high frequencies and input image is generated as an output.

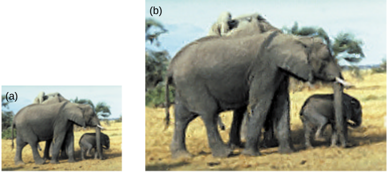


Fig. 3. (a) Input image. (b) Result of example-based super-resolution.

## 4. Estimation of ideal images

The goal of the algorithm is to suc-

*2 A cubic spline is a spline constructed of piecewise third-order polynomials which pass through a set of *m* control points. (http://mathworld.wolfram.com/CubicSpline.html) cessfully obtain the ideal image from an intermediate-view image automatically. Although we believe a good quality intermediate-view image is obtained by correcting errors in calibration and segmentation for each input image iteratively, it is actually difficult to do this automatically. Therefore, we decided that the best way to achieve the goal would be to have the algorithm learn relationships between an ideal image and an intermediate-view image for the same viewpoint as a training set and predict high frequencies of other intermediate-view images using the training set. This is performed under the assumption that the relationship between an ideal image and a corresponding intermediate-view image is the same as that for another viewpoint for the generated 3-D data. Note that example-based super-resolution can generate a training set from only one image. Our algorithm needs a pair of images as input: an ideal image and an intermediate-view image viewed from the same viewpoint.

Training images are obtained using two kinds of 3-D data: data generated using all input images and data generated using all input images except for the one from the target viewpoint. For both sets of data, rendered views from the target viewpoint are captured to obtain an ideal image and an intermediate-view image. A high-pass filtered image for each image is then generated and divided into patches. Then, a training set is generated as a collection of pairs of patches as in example-based super-resolution. Here, we can regard a patch generated from intermediate-view images as a low-resolution patch (Fig. 4(a)) and one generated from ideal images as a high-resolution patch (Fig. 4(b)).

To estimate an ideal image from an intermediate-view image, a high-pass filtered image is generated from an input image and divided into patches. Then, high frequencies of each patch are predicted based on a search for the best-matching patches in a pair in the same manner as example-based super-resolution. The obtained high-resolution patches are also used for predicting the high frequencies of each patch. In this case, the obtained high-resolution patches construct the same-resolution high-frequency image directly. Finally, the sum of the resulting high frequencies and input image are generated as an output.

## 5. Results

We applied the algorithm to a mug. Five pairs of images were used for training images. We obtained about 155,000 patches. One of the pairs of training images is shown in Fig. 2.

Figure 5 shows (a) the intermediate-view image, (b) the result for the proposed algorithm, and (c) the ideal image. For comparison, the image sharpened by Laplacian filtering is shown in (d). Figure 6 shows the same regions in the original resolution. The results show that the quality of the obtained image is as good as that of the ideal image. On the other hand, the Laplacian filtered image is jagged, although the contrast is improved.

## 6. Conclusion

To develop a fully automatic algorithm to generate photorealistic 3-D data from real objects, we have developed an algorithm that improves the quality of intermediate views. It learns relationships between ideal images and corresponding intermediate views and uses them as a training set once. Ideal images are then estimated from intermediate-view images from other viewpoints using the training set. Experimental results for real data show that the quality of obtained images is almost as good as that of ideal images.

In the future, we plan to investigate in more detail which parameter settings provide adequate results and integrate our 3-D data generation algorithm with

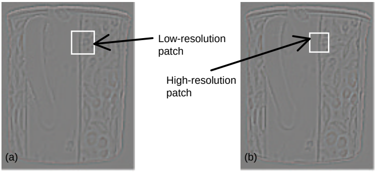


Fig. 4. (a) High-pass filtered image of image in Fig. 2(a). (b) High-pass filtered image of image in Fig. 2(b).


Fig. 5. Results for mug. (a) Input image. (b) Proposed algorithm. (c) Ideal image. (d) Laplacian filtered.

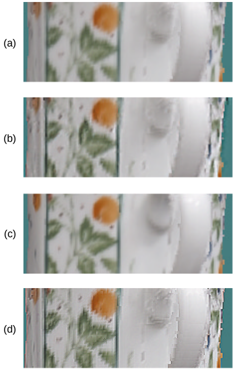


Fig. 6. Results in original resolution of Fig. 5. (a) Input image. (b) Proposed algorithm. (c) Ideal image. (d) Laplacian filtered.

a view-interpolation algorithm.

## References

[1] W. T. Freeman, T. R. Jones, and E. C.Pasztor, “Example-Based Super-Resolution,” IEEE CG & A, Mar./Apr. 2002, pp. 56-65, 2002.

[2] H. Sato, H. Matsuoka, A. Onozawa, and H. Kitazawa, “Hexagonal Image Representation for 3-D Photorealistic Reconstruction,” Proc. of ICPR 2002, Vol. 3, pp. 672-676, 2002.

[3] R. Szeliski, “Rapid octree reconstruction from image sequences,” CVGIP: Image Understanding, 58(1), pp. 23-32, 1993.

**Hidenori Sato**

Senior Research Engineer, Ubiquitous Interface Laboratory, NTT Microsystem Integration Laboratories.

He received the B.S. and M.S. degrees in physics from Tohoku University, Sendai, Miyagi in 1987 and 1989, respectively. In 1989, he joined NTT LSI Laboratories, Kanagawa, Japan, where he worked on LSI CAD. His current research interests are in computer graphics and human computer interaction system using computer vision. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), the Information Processing Society of Japan (IPSJ), and IEEE.

**William T. Freeman**

Associate Professor of electrical engineering and computer science, Massachusetts Institute of Technology.

He received a Ph.D. from MIT in 1992 for his studies on computer vision. From 1992–2001, he worked at Mitsubishi Electric Research Labs (MERL), in Cambridge, MA, most recently as a Senior Research Scientist and Associate Director. His research interests include machine learning applied to computer vision, Bayesian models of visual perception, and interactive applications of computer vision. In 1997, he received the Outstanding Paper prize at the Conference on Computer Vision and Pattern Recognition for work on applying bilinear models to “separating style and content”.

**Akira Onozawa**

Senior Research Engineer, Supervisor, Ubiquitous Interface Laboratory, NTT Microsystem Integration Laboratories.

He received the B.E. and M.E. degrees in electronic communication engineering in 1983 and 1985, respectively and Ph.D. in information science in 2002 from Waseda University, Tokyo. He joined NTT in 1985 and has been working on the research and development of LSI CAD algorithms and systems. His current interests include human-computer interfaces, computer vision, and computer graphics. He is a member of the Association for Computing Machinery, IEEE, IEICE, and IPSJ.