SMART SAMPLING AND TRANSDUCING 3D SCENES FOR THE VISUALLY IMPAIRED

Hao Tang[1,2], Tony Ro[3] and Zhigang Zhu[2]
[1] Department of Computer Information System, CUNY Borough of Manhattan Community College
[2] Department of Computer Science, CUNY City College and Graduate Center
[3] Department of Psychology and CUNY Program in Cognitive Neuroscience
Convent Avenue and 138th Street, New York, NY 10031
htang@bmcc.cuny.edu, tro@ccny.cuny.edu, zhu@cs.ccny.cuny.edu

ABSTRACT

Current visual prosthetic devices provide only very limited restoration or substitution of vision for the visually impaired, in part due to their low resolution and simple scene transducing approaches with uniform sampling and quantization. In this paper, a novel smart sampling method using a color-patch-based stereo reconstruction approach is described to automatically select, sample and transduce the most useful scene information to end users using visual substitution devices. The proposed method first constructs a patch-based 3D model of the scene using the color-patch-based stereovision algorithm given a pair of video frames captured by a stereo camera head. Then, the patch-based 3D model is analyzed using the smart sampling algorithm and further transduced into various alternative perception choices, using both color and depth information. Some preliminary experimental results are shown to validate the proposed method.

Index Terms— image sampling, 3D reconstruction, assistive technology, visual prosthesis

1. INTRODUCTION

Various visual prostheses have been developed in the past decade. The most prevalent visual prosthesis, the retinal prosthesis, is an experimental visual device aimed at restoring vision functions of the visually impaired [2, 15]. It can provide low resolution images to a blind user with these retinal prostheses, which electrically stimulate his/her retinal cells. Currently, the state-of-the-art retinal prosthesis has very limited resolution (60 – 100 channels, in the form of a 6x10 or 10x10 array) [15]. Another example of a visual prosthesis that does not require surgery is tongue stimulation: the Brainport technique [1] of Wicab Inc. captures an image and processes the image by converting it into impulses which are sent via an electrode array on the tongue. The tongue simulator currently has 400 channels (a 20x20 resolution).

Both methods face a very serious problem: low resolution. If we simply sample an original image into a 20x20 or lower resolution array to drive the retinal or tongue stimulation, it would be hard to preserve small objects that are close to the user. Another problem is when a scene is cluttered, it is difficult to represent the complex scene in a low-resolution display. It is challenging to convey and enhance the most useful and important information without a comprehensive analysis of the scene.

To meet this challenge, we propose a smart-sampling method. The basic idea is as follows. Using 3D computer vision techniques, we can first pre-process a scene and obtain its 3D model. Then, using the 3D model, only the important objects will be selected, sampled and conveyed to the low-resolution display for an end user who is visually impaired. The output of the system can be easily encoded into the input device of any kind of visual prosthetic that has low resolution.

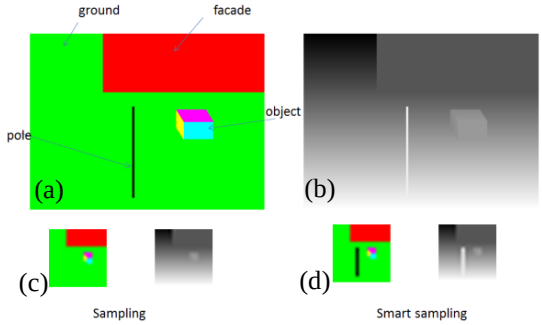


Fig. 1. Smart sampling based on 3D scene segmentation. (a) The simulated scene with a number of objects (e.g., a pole is in a close range). (b) 3D depth map of the simulated scene. (c) Sampling results of the 2D image (left) and 3D depth map (right) using a regular sampling method. Note the pole is missing after the regular sampling. (d) Sampling results of the 2D image (left) and 3D depth map (right) using our smart sampling method. The pole is still preserved after sampling.

Fig. 1 illustrates the idea using a simulated scene. Fig. 1a is a simulated image with a green background (ground plane) and three objects: a building façade, a cubic object, and a thin, vertical pole. Using the patch-based stereo approach [14], the 3D information of all regions is obtained (Fig. 1b). Fig. 1c shows the results of a regular uniform sampling of the image into 20x20 pixels: the thin long pole disappeared. However, using our smart sampling method, the thin pole is preserved in the final 20x20 sampled image.

This paper will focus on a novel 3D scene transducing method under very low image resolution: the smart sampling algorithm using both color and depth segmentation. During the writing of this paper, we are performing more experiments that send the transducing results to the tongue stimulator of the Brainport device. We hope the new sampling approach will increase the capability of alternative perception techniques for the visually impaired.

The paper is organized as follows. Section 2 discusses some closely related work. In Section 3, we present the smart sampling and enhancement method. Section 4 provides some experimental results. Finally we conclude our work in Section 5.

2. RELATED WORK

Computer vision techniques are playing an increasingly important role in the development of visual prostheses [3]. Computer vision can be used to help restore some specific visual abilities, such as light perception and object recognition in retinal prostheses. Image segmentation has been applied in a visual prosthesis to enhance object recognition [5] and face detection and tracking methods are used to assist with recognizing faces [4]. Another challenge that visually impaired people encounter is navigation, and many different vision technologies have been applied in the development of electronic travel aids for the visually impaired. Coughlan et al. [7, 8] propose systems for helping the visually impaired find a path to a machine-readable sign using a cellphone camera. Using stereo cameras [9, 10], depth maps are produced to aid navigation. Staircases [11 and 12] and zebra-crossings [13] are detected using stereo cameras to help blind users identify and climb stairs and cross streets.

The work most related to ours is the method proposed by McCarthy et al. [6], which is a vision algorithm for retinal prostheses to support visual navigation. With stereo vision techniques, the system classifies a scene into ground and non-ground surfaces and renders a depth image in a low resolution version. But it might miss small/thin objects, such as a pole, a horizontal bar, or a thin tree branch in front of the user due to the use of uniform sampling.

3. SMART SAMPLING

The smart sampling approach consists of two steps. First, a patch-based stereovision method [14] is applied to a pair of stereo images captured by a stereo camera head. The outcome of the patch-based method is not just an array of individual 3D points that are usually produced by a typical stereovision system. Instead, it is a geometric representation of plane parameters, with geometric relations among neighboring planar surfaces. Second, a smart sampling algorithm is applied, using the patch-based 3D and color segmentation results. This paper will focus on the second step.

In order to make full use of the limited resolution of alternative perception devices such as retinal prostheses and tongue stimulators, starting with the patched-based 3D representation, our smart sampling algorithm provides enhancements to the alternative perception devices of end users.

Sampling needs to be conducted to reduce a 2D/3D map from an original high resolution image (Ro) to a low resolution sampling (Rs) for visual prosthetic or tongue stimulation. Regular uniform sampling methods sample one pixel every N pixels (N=Ro/Rs). For some thin objects, for example, a lamp pole in front of the blind user, it is impossible to preserve the object after sampling if the width of the pole in the image is smaller than N. It will be very dangerous because the user may bump into the objects right in front of him/her.

The smart sampling we have proposed can preserve such thin objects, which are determined to be significant by a number of measurements: the distance from the user, the confidence in the 3D measurements, and the shapes of the objects. Currently, we consider thin but long objects that could be vertical poles and horizontal bars. From the patch based stereo vision method, a 3D map consists of many planar patches with known geometric relations. The goal of smart sampling is to not lose any important information when the sampling is performed based on the patch-based 3D representation.

In order to reduce the computational cost for a portable device implementation, we propose a very efficient algorithm based on the patch-based stereovision result. The following shows the details for the smart sampling of the depth image.

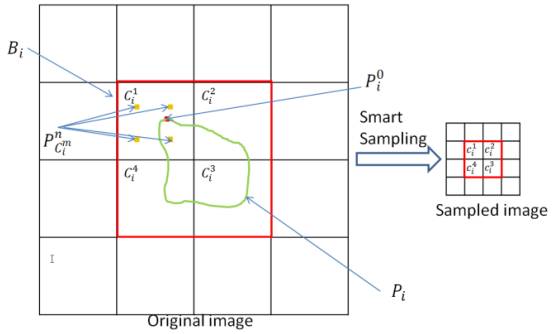


Fig. 2. Illustration of smart sampling of a patch P_i. The original image is divided into multiple rectangle cells, each cell corresponds to a pixel in the sampled image. The bounding box B_i of the patch P_i occupies a number of cells $\{C_i^m\}$, each corresponding to a pixel c_i^m in the sampled image (m=1, 2, 3 4 in the figure). P_i^0 is the first point in P_i, and $P_{C_i^m}^n, n = 1..4$, are the four sampled pixels in the cell C_i^m.

Smart Sampling Algorithm for the Depth Image

Input:

Z: the original depth image, of the size W × H

Output:

Z_s: the sampled depth image, of the size w × h

Notations Used in the Algorithm (Fig. 2)**:**

$P_i = \{P_i^j\}, i = 1 \ldots K$ where

P_i: the ith patch obtained in the image segmentation step (Note: the depth in a single patch may be different);

P_i^j: the jth pixel in the patch P_i;
K: the total number of patches in Z;

(s_x, s_y) : the ratios between the resolutions of Z and Z_s, i.e., $s_x = W/w$ and $s_y = H/h$.
B_i : the bounding box of patch P_i, which includes a number of cells $\{C_i^m\}$ that are rectangular regions in Z.
(Note: $\cup_{m=1}^{M} C_i^m = B_i$, but $C_i^m \cap P_i$might be $\emptyset$. M is the total number of the Cells in B_i)
$\{c_i^m\}$: the set of the sampled pixels in Z_s, corresponding to $\{C_i^m\}$

---

1. Initialization of Z_s using a uniform subsampling method
For $q = 1$ to $w * h$
$x_l = x_q^s * s_x, y_l = y_q^s * s_y$ // $(x_l, y_l) \in Z, (x_q^s, y_q^s) \in Z_s$
$Z_s(x_q^s, y_q^s) = Z(x_l, y_l)$ //initial sampled image
End for
2. Smart sampling
For $i = 1$ to K // loop for the patches
a) Sample the first pixel $P_i^0 = (x_i^0, y_i^0)$ of the patch P_i in the original depth image Z (the red pixel in Fig. 2) to determine if pixel $(x_0^s, y_0^s) = (x_i^0/s_x, y_i^0/s_y)$ in the sampled image will be replaced:
if $Z(x_i^0, y_i^0) < Z_s(x_0^s, y_0^s)$
$Z_s(x_0^s, y_0^s) = Z(x_i^0, y_i^0)$ //update depth value
b) For $m = 1$ to M //do regular sampling in the $\{C_i^m\}$,
Sample four pixels $P_{C_i^m}^n, n = 1..4$ (uniformly distributed, orange points in Fig. 2) in the cell C_i^m,
if $P_{C_i^m}^n \in P_i$
$x_n^s = x_{C_i^m}^n/s_x, y_n^s = y_{C_i^m}^n/s_y$
if $Z(x_{C_i^m}^n, y_{C_i^m}^n) < Z_s(x_n^s, y_n^s)$
$Z_s(x_n^s, y_n^s) = Z\left(x_{C_i^m}^n, y_{C_i^m}^n\right)$ //update depth value
End for
End for

The smart sampling is performed on each patch and it is guaranteed that at least one pixel can be sampled regardless of the size of the patch. Initially the sub-sampled image is filled with the regular sampling method. During the smart subsample process, a sampled 3D value is filled in by comparing the new 3D values with the existing 3D value and the value of the closer value to the user is kept. In this way, thin objects in close range can still be preserved during sampling. With the same method (based on 3D information), the original color image can be subsampled and any important information can be kept as well.

The computation complexity of initialization is O(w×h), where both w and h are 20 in our experiments. The smart sampling is processed on each patch, and then we do the regular sampling in each patch so the time complexity is $\sum_K(T_k + 1)$, where T_k represents the computation of the regular sampling in patch k, which depends on the shape of each patch but may not be proportional to the size of patch. The worst case happens when each patch is a diagonal line after image segmentation, but the average computation should still be proportional to the total size of the subsampled image, thus O(w×h). Therefore the overall computation of smart sampling is O(w×h). In addition, because the process of each patch could be performed separately, it can be easily programmed with a parallel processing method, which is extremely useful for real-time processing, such as during navigation by blind people.

In order to provide blind people with stronger input when observing obstacles and other types of objects of interest, the objects of interest (OIs) can be highlighted by using or combining following methods: background removal, highlighting OIs using motion parallax simulation, or dynamic object highlighting. Details of these methods and corresponding experiments can be found in an accompanying paper [16].

4. EXPERIMENTAL AND RESULTS

Experiments have been performed to test our approach. Image sequences were captured by the stereovision head Bumblebee, which is fixed on a mobile platform. For a pair of stereo images, the left camera serves as the reference camera. Fig. 3a shows a stereo pair of color images captured in an office, and Fig. 3b shows the rendered depth map from the estimated planar representations using the patch-based stereo matching algorithm. The plane parameters in the form of "no. (a, b, c) d, n", where no. represents the index of the plane, (a, b, c) represents the planar surface normal vector, d represents the average distance of the plane to the viewer and n represents the uncertainty measurement, are marked for a number of large patches. Note that patches with large uncertainties are highlighted in green.

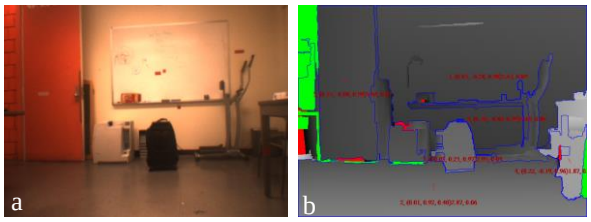
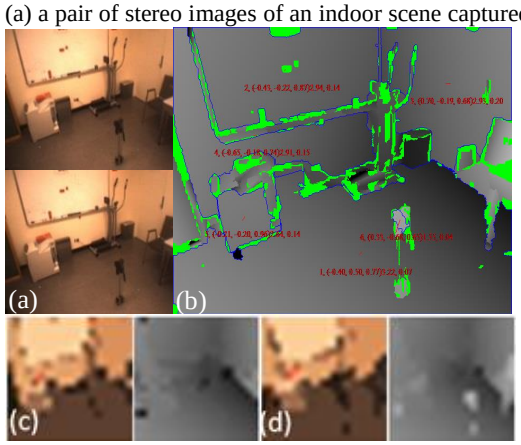


Fig. 3. (a) A stereo pair of color images (b) 3D depth map generated by patch-based method (the brighter, the closer). For several large regions indexed, the boundaries of regions are marked by closed curves (blue) and planar parameters are drawn on the regions.

In Fig. 4a, a pair of stereo images of an indoor scene is shown, including a table, a chair, a printer and a tripod, which are about 1 to 4 meters away from the stereovision head. A depth map (the brighter, the closer) is shown in Fig. 4b. For several large surfaces, the plane parameters in the form of "no. (a, b, c) d, n" are also shown, with their boundaries highlighted in blue. These plane estimation results are consistent with the results measured by hand. The parametric representation can be transduced to a blind user easier than an array of depth points with a uniform sampling meth-

od. Fig. 4 shows results after applying a uniform (c) and the smart sampling approach (d). Fig. 4c shows that the tripod, which is about 1.55 meters from the user, is missing after uniform sampling, but it is preserved using the proposed smart sampling approach (Fig 4d). The geometric representations enable safe and efficient navigation for the visually impaired.

Fig. 4. (a) a pair of stereo images of an indoor scene captured in an



office with a number of objects (note: a tripod is in a close range); (b) 3d depth map of the indoor scene; pixels with large uncertainty are marked in green; (c) sampling results of 2D image (left) and 3D depth map (right) using a uniform sampling method: the tripod is missing after regular sampling; (d) sampling results of 2D image (left) and 3D depth map (right) using our smart sampling method: the tripod is kept after sampling.

5. SUMMARY AND DISCUSSIONS

Transducing digital video images into displays of a low resolution device is required for state-of-the-art visual prosthetics. The proposed smart sampling method can preserve close range objects that are significant by a number of measurements: the distance from the user, the confidence in the 3D measurements, and the shapes of the objects. Using this smart sampling approach, a number of practical sampling and enhancement methods can be applied to transduce important information and highlight objects of interest in different ways in order to allow an end user to easily understand the environment. We are currently working on applying the proposed method to existing visual prosthetic systems, such as the Wicab tongue simulator; some of the results will be reported in [16].

6. ACKNOWLEDGMENTS

This work is supported by NSF under Awards #EFRI-1137172 and #BCS-0843148, and a CCNY City SEEDs 2011 Grant. The first author is also supported by a PSC-CUNY Research Award (Round 44).

7. REFERENCES

[1] BrainPort Vision Technology. http://vision.wicab.com/technology, last visited November 2012

[2] J.D. Loudin, D.M. Simanovskii, K. Vijayraghavan, C.K. Sramek, A.F. Butterwick, P. Huie, G.Y. McLean, and D.V. Palanker. Optoelectronic retinal prosthesis: system design and performance. *Journal Neural Engineering*, 4 (1): S72–S84. 2007

[3] N Barnes. The role of computer vision in prosthetic vision, *Image and Vision Computing*, 20, 478-439, 2012.

[4] X. He, C. Shen, N. Barnes. Face detection and tracking in video to facilitate face recognition in a visual prosthesis. *Annual Meeting of the Association for Research in Vision and Ophthalmology*, Florida, 2011

[5] L. Horne, N. Barnes, C. McCarthy, X. He. Image segmentation for enhancing symbol recognition in prosthetic vision. *Annual International IEEE Engineering in Medicine and Biology Society Conference*, 2012.

[6] C. McCarth, N. Barnes and P. Lieby. Ground surface segmentation for navigation with a low resolution visual prosthesis. *Annual International IEEE Engineering in Medicine and Biology Society Conference*, 2011.

[7] J. Coughlan, R. Manduchi, H. Shen, Cell phone-based wayfinding for the visually impaired. *1st International Workshop on Mobile Vision*, 2006.

[8] R. Manduchi, J. Coughlan and V. Ivanchenko. Search strategies of visually impaired persons using a camera phone wayfinding system. *ICCHP* 2008.

[9] R. Audette, J. Balthazaar, C. Dunk, and J. Zelek, A stereo-vision system for the visually impaired. *Tech. Rep, Sch. Eng., Univ. Guelph, Guelph, ON, Canada.* 2000-41x-1, 2000.

[10] L. Gonz´alez-Mora, A. Rodr´ıguez-Hern´andez, L. F. Rodr´ıguez-Ramos,L. D´ıaz-Saco, and N. Sosa. Development of a new space perception system for blind people, based on the creation of a virtual acoustic space. *Technical Report*, May 8 2009

[11] X. Lu and R. Manduchi. Detection and localization of curbs and stairways using stereo vision. *IEEE International Conference on Robotics and Automation*, 2005

[12] S. Se, B. Michael. Vision-based Detection of Staircases, *Asian Conference on Computer Vision*, 2000

[13] S. Se. Zebra-crossing detection for the partially sighted. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2000

[14] H. Tang, Z. Zhu, J. Xiao, Stereovision-based 3D planar surface estimation for wall-climbing robots. *International Conference on Intelligent Robots and Systems*. 2009.

[15] Second Sight, http://2-sight.eu/en/home-en, last visited November 2012.

[16] H. Tang, M. Vincent, T. Ro and Z. Zhu. From RGB-D to low-resolution tactile: smart sampling and early testing, *IEEE ICME Workshop on Multimodal and Alternative Perception for Visually Impaired People*, July 15, 2013.