4D Median Filtering for the Multi-Viewable 3D Moving Images of Surgery Generatable by a Computer-Aided Stereoscopic Endoscope

Takeshi Ohtsuka*, Shuichi Utsugi†, Kujira Suzuki‡, Hisashi Suzuki* and Hitoshi Katai§
*Department of Information and System Engineering, Chuo University, Tokyo, Japan
†Department of Computer Science and Technology, Salesian Polytechnic, Tokyo, Japan
‡Graduate School of Social Governance, Hosei University, Tokyo, Japan
§Department of Gastric Surgery Division, National Cancer Center Hospital, Tokyo, Japan

Abstract—A stereoscopic endoscope in a usual architecture transmits simply a left/right image into the left/right eye of a user to provide a 3D perception. In contrast to the usual architecture, a computer-aided stereoscopic endoscope in a sophisticated architecture reconstructs a sequence of 3D spatial data in realtime of approximately 30 fps from a HVGA-resolution stereo moving image; processes the sequence of 3D spatial data in space and time dimensions; and renders a multi-viewable 3D moving image so that we can display on any multi-viewable lenticular displays without glasses.

A decisive advantage of once reconstructing 3D spatial data, in addition to eliminating the need for glasses, is that we become possible to introduce various CG techniques for 3D enhancement. For example, users can look into a same multi-viewable 3D display from respective viewing directions independently of the actual direction of endoscope, which will resolve VR sickness. For another example, a user can arbitrarily emphasize the concavity and convexity in 3D perception, which will make visual detection of cancer easy. Moreover, a user can attract any local space in 3D perception as if a "gravitational lens" were applied, which will assist surgeries.

Although the core of the computer-aided stereoscopic endoscope is a dynamic programming (DP) algorithm for high-precision stereo matching that the authors newly developed, an inherent problem is that the 3D spatial data reconstructed by using any DP algorithms include occasional errors in stereo matching for depth estimation that are caused by lack of moderate textures on original left and right images. (This problem corresponds to so-called "illusory perception" in terms on human body.) This paper reports success of generating some drastically enhanced multi-viewable 3D moving images of surgery by applying a time-space, that is, 4D median filter for efficiently reducing such inherent errors in depth estimation.

Keywords—Computer-aided, stereoscopic, endoscope, 3D-perception, glasses-free, multi-viewable, 3D-enhancement, lenticular.

I. INTRODUCTION

A stereoscopic endoscope in a usual architecture transmits simply a left/right image into the left/right eye of a user to provide a 3D perception. In contrast to the usual architecture, a computer-aided stereoscopic endoscope in a sophisticated architecture reconstructs a sequence of 3D spatial data in realtime of approximately 30 fps from a HVGA-resolution stereo moving image; processes the sequence of 3D spatial data in space and time dimensions; and renders a multi-viewable 3D moving image so that we can display on any multi-viewable lenticular displays without glasses[1].

Although the core of the computer-aided stereoscopic endoscope is a dynamic programming (DP) algorithm for high-precision stereo matching that the authors newly developed[2], an inherent problem is that the 3D spatial data reconstructed by using any DP algorithms include occasional errors in stereo matching for depth estimation that are caused by lack of moderate textures on original left and right images.

A multi-viewable 3D moving image is generated by sequential processing of the stereo matching algorithm, which include temporal fluctuation of the errors. This is caused by independency between the frames. Therefore, a error problem caused by stereo matching errors and frame independency is a imperative duty to be solved in application of the computer-aided stereoscopic endoscope. This paper reports success of generating some drastically enhanced multi-viewable 3D moving images of surgery by applying a time-space, that is, 4D median filter for efficiently reducing such inherent errors in depth estimation.

II. COMPUTER-AIDED STEREOSCOPIC ENDOSCOPE

A computer-aided stereoscopic endoscope consists of three interchangeable modules. Fig. 1 shows the architecture of computer-aided stereoscopic endoscope. The corresponding project is under a joint research between Chuo University and National Cancer Center in Japan with a close cooperation by a medical equipment manufacturer in Japan.

A. Stereo Endoscope module

Development of a fully-digitalized stereo endoscope module is at the finalizing phase (cf. Fig. 2). A newly developed DP algorithm for high-precision stereo matching can reconstruct a sequence of 3D spatial data in realtime of approximately 30 fps from an HVGA-resolution stereo moving image. The performance of the algorithm is that it can speed up the DP matching by FPGA and process each frame of 3840×2160-pixel (4K UHDTV) stereo images consisting of a pair of left & right 1920×2160-pixel images at 30 fps with 54 processors. The rate of correct matching achieves 94.2% in average for the typical images called Tsukuba, Venus, Teddy, Map, and Sawtooth in Middlebury Stereo Datasets (cf. Fig. 3).
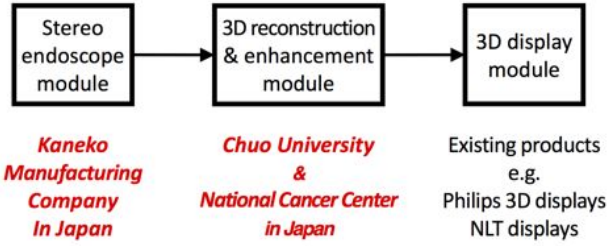
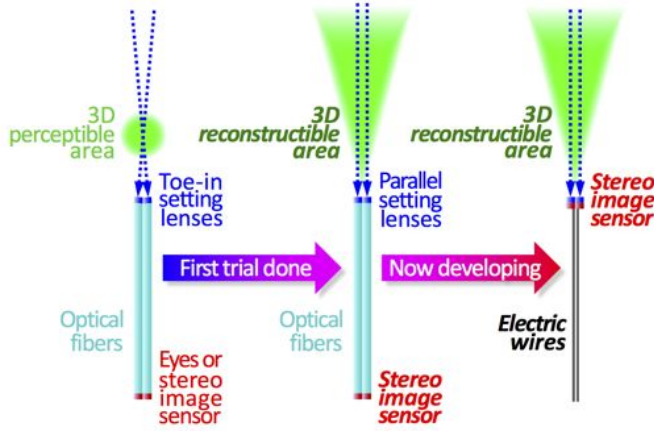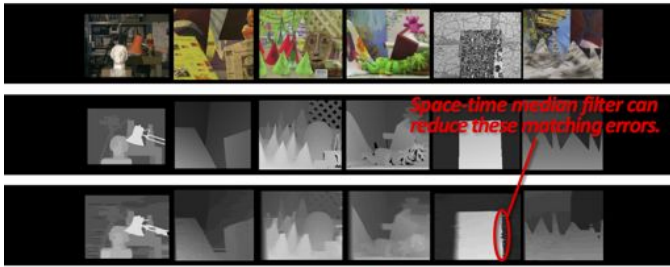Fig. 1: Architecture of a computer-aided stereoscopic endoscope.

Fig. 2: Phases of development.

The rate of correct matching achieves **94.2%** in average for the typical images called Tsukuba, Venus, Cones, Teddy, Map, and Sawtooth in Middlebury Stereo Datasets.

Fig. 3: Precision of stereo matching.

Fig. 4: Stereo matching.

Fig. 5: Composition of 3D spatial data.

B. 3D Reconstruction and enhancement

The core of the computer-aided stereoscopic endoscope is a dynamic programming (DP) algorithm for high-precision stereo matching (cf. Fig. 4) that the authors developed. This algorithm reconstructs in realtime a sequence of the 3D spatial data composed of depth data and surface data (Fig. 5) from a stereo moving image of live body; processes the sequence of 3D spatial data in space and time dimensions; and renders a multi viewable 3D moving image so that we can display on any multi-viewable 3D displays without glasses.

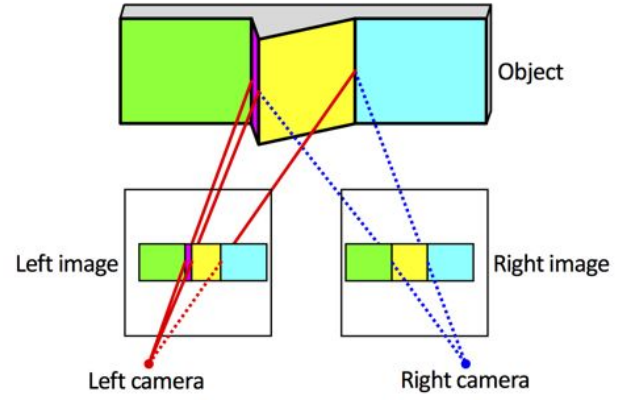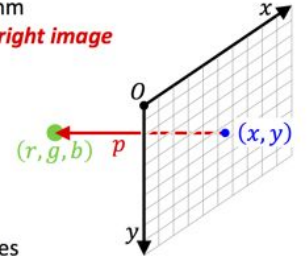In addition to eliminating the need for glasses, a decisive advantage of once reconstruction 3D spatial data is that we become possible to introduce various CG techniques for 3D enhancement. For example, each user can look into a same multi-viewable 3D display from arbitrary directions independently of endoscope (in other words, multi-views inherently can absorb individualities of binocular disparity for lenticular stereo-viewing), which resolves VR sickness that is a risk factor in surgeries. For another example, a user can arbitrarily emphasis the concavity and convexity in 3D perception, which will make visual detection of cancer easy.

Moreover, a user can attract any local space in 3D perception as if a "gravitational lens" were applied, which will assist surgeries significantly. Thus there is an enormous potential of introducing CG techniques for 3D enhancement.

III. 4D MEDIAN FILTER

An inherent problem of 3D spatial data reconstruction is that any DP algorithms include occasional errors in stereo matching for depth estimation. This problem corresponds to so-called "Illusory perception" in terms on human body. 3D moving images are generated by sequential stereo matching, which include temporal fluctuation caused by stereo matching errors and frame independency. Therefore, the errors degrade the quality of the 3D movie.

4D filtering is inspired by the researches in noise removal of image sequences [3], [4] introducing time-space filtering to

remove both spatial and temporal noise, which are observed as intensive changes of the pixel values in the same location on different frames. It seems that the errors of the 3D moving images are also depressed by the time-space filtering. Because 3D moving images are consisted of a sequence of 3D spatial data reconstructed by stereo matching, there are 4 directions to be considered. (3 spatial directions plus 1 temporal direction equals 4 directions.) Therefore, 4D filters take advantages of the correlations between spatial and temporal dimensions.

4D median filter is the most basic and important filters, which is easily constructed by extending the 2D median filter to the temporal dimension. The kernel size is $x \times y \times t$, where x and y are spatial dimensions, and t is the time dimension among frames. (the pixels that store estimated depth value represent z dimension.) Pixel values are sampled the length of $x \times y \times t$ and slide across the images. The filter output at each position, the modified z dimension value, is given as the median value of the samples in the kernel.

A. Properties of a Median Filter

Median filter is the most often used on 2D imagery to remove impulsive noise because of its characteristics [5]. Two main properties of median filter are:

1) Impulsive noise removal.
2) Edge preserving.

An output of the median filter is a median of the sorted signals in the kernel. Impulsive noise will not be adopted to the filter because the noise is moved to head or tail when sorted. As a consequence, median filter has a property to remove impulsive noise.

Edge is an area that two neighborhoods with different constant values are connected. When the filter is approached to edge, median filter selects either neighborhood having more elements, which means that it works as a majority function in edge. On the other hand, in case of simple linear averaging, edge is not preserved.

B. Extension to 4D filtering

Median filter in 2D imagery removes impulsive noise in spatial field. In other words, it removes spatial impulsive noise. On the other hand, temporal filter having $1 \times 1 \times t$ kernel size can avoid the temporal fluctuation by modeling the moving images as a time series of 1D pixel.

Let apply time-space filters for the 3D moving images including the errors of the stereo matching which is generated independently with respect to time. The errors considered here are different from meaning of noise in the general image processing. Therefore, it is important to examine how the time-space image processing works to the 3D movies. The reason that named 4D filtering is because the target is different from the noise reduction in the general image processing.

Fig. 6 shows the concept of the 4D median filter. The images in each frame have 3D information and include spatial errors expressed in a red line. In addition, the depth information fluctuates about time, and it is thought that they include temporal errors expressed in blue line. Spatial and temporal errors are orthogonally generated. Spatial noise is caused by
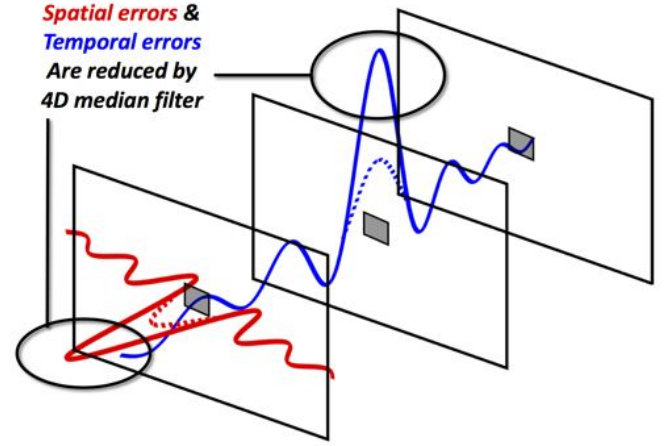

Fig. 6: The concept of 4D filtering.

stereo matching errors. Meanwhile, temporal noise is seen, for the first time, when connecting all frames into a image sequence.

4D median filter has variety in configurations of the 3D kernel, that is, we can arbitrary change the weight of spatial and temporal dimensions. Peters and Nichols [6] tested the performances of several configurations to image sequences of rocket plumes, which shows that long duration median filters with moderate spatial support appear to work best in terms of signal to noise ratio (SNR).

C. Apply to 3D Moving Images of Surgery

For the first step of the realization of computer-aided stereoscopic endoscope, let generate a 3D moving image of biological tissues of surgical scenes. The high-precision DP algorithm is applied to a stereo moving image of actual gastric surgery. Fig. 7 shows a frame of the movie and its depth information. Because depth estimation is applied to the stereo moving images on a frame-by-frame basis, depth information of each frame include impulsive noise of stereo matching. Hence, 4D median filter is necessary for the 3D moving image to remove the errors. Fig. 8 shows that 3D reconstructed data and its 4D median filtered images are visually compared by rendering at 3D virtual space, which shows the obvious removal of its impulsive noise. Frame intervals are connected smoothly by the filtering. Accordingly, the multi-viewable 3D moving image of surgery is drastically enhanced by applying the 4D medial filter for efficiently reducing such inherent errors in depth estimation.

The DP algorithm of the stereo matching works on the surgical images because the biological tissues have rich textures, that is, high frequency components of the images. Whereas, they include occasional errors in stereo matching. This is caused by the lack of moderate textures on the objects such as the hook and the tweezers which include low frequency components. In addition, depth estimation does not work for the objects which move fast.
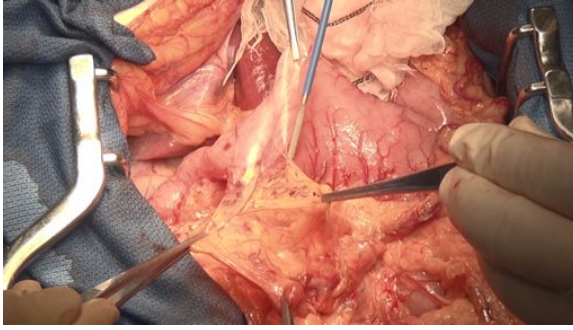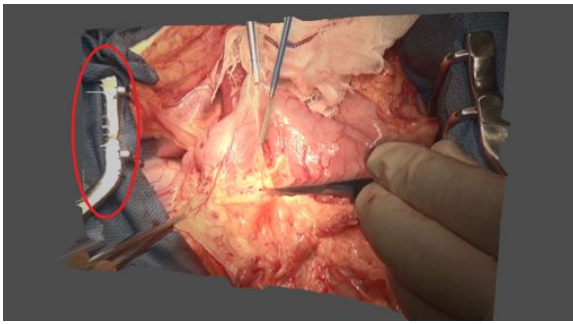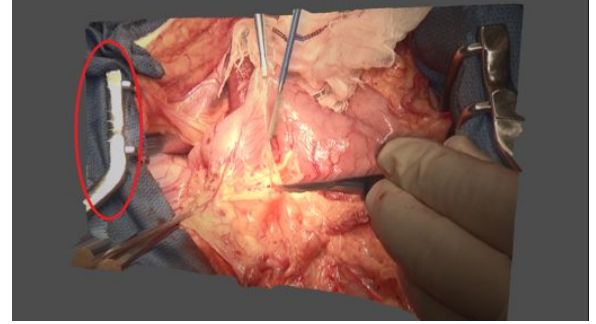
(a) Surface data.

(b) Depth data.

Fig. 7: A frame of the 3D moving image of surgery.

(a) Original reconstructed.

(b) 4D median filtered.

Fig. 8: Comparison of the reconstructed data.

D. Consideration

The 4D median filter, the simplest implementation, is effective for improvement of visual quality of the 3D moving images of surgery. There are two reasons: impulsive noise reduction due to median filtering; and reconnection of frame intervals by temporal extension of the kernel. As a consequence, time-space filtering works on improvement of visual quality because the 3D moving images are frame-individually generated by stereo matching.

4D filters have rich variations in the configuration. For example, typical filters such as Gaussian filter is easily extended to temporal direction. In addition, motion-compensated filtering will be effective for error reduction. More effective filtering will be enabled by following a lot of techniques for noise reduction of moving images.

A problem is how to quantitatively evaluate error reduction because it is not appropriate to approximate errors of stereo matching as Gaussian noise. Synthetic data for measuring SNR, the performance of filtering, should be prepared. Therefore, it is necessary to construct the generative model of the stereo matching errors.

IV. CONCLUSION

Development of A computer-aided stereoscopic endoscope introduces the generation of multi-viewable 3D moving images of surgery. The 3D moving images include occasional stereo matching errors, which degrades the visual quality. It includes temporal errors caused by frame independency. To solve these problems, this paper introduces time-space, that is, 4D median filter for efficiently reducing stereo matching errors in depth estimation. The concept of 4D filtering is to process errors of stereo matching at the same time while keeping relations between frames. A problem is to quantitatively evaluate the 4D filtering, and therefore the modeling of the stereo matching error is necessary.

REFERENCES

[1] H. Suzuki, S. Utsugi, and H. Katai, "Multi-view auto-stereoscopic endoscopy system," Patent JP 2012-065 851 A, Apr., 2012.

[2] S. Utsugi and H. Suzuki, "Stereo camera range finder with a slit-like window for visual aid," *Abstracts of Collaborative Conference on 3D and Materials Research (CC3DMR) 2013*, pp. 7–8, Jun. 2013.

[3] M. Alp and T. Neuvo, "3-dimensional median filters for image sequence processing," in *Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference on*, vol. 4, Apr. 1991, pp. 2917–2920.

[4] J. Brailean, R. Kleihorst, S. Efstratiadis, A. Katsaggelos, and R. Lagendijk, "Noise reduction filters for dynamic image sequences: a review," *Proceedings of the IEEE*, vol. 83, no. 9, pp. 1272–1292, Sep. 1995.

[5] J. Gallagher, N.C. and G. Wise, "A theoretical analysis of the properties of median filters," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 29, no. 6, pp. 1136–1141, Dec. 1981.

[6] I. Peters, R.A and J. Nichols, "Rocket plume image sequence enhancement using 3d operators," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 33, no. 2, pp. 485–498, Apr. 1997.