

Binocular Endoscopic 3-D Scene Reconstruction Using Color and Gradient-Boosted Aggregation Stereo Matching for Robotic Surgery

Xiongbiao Luo¹(✉), Uditha L. Jayarathne¹,
Stephen E. Pautler², and Terry M. Peters¹

¹ Robarts Research Institute, Western University, London, Canada
{xluo,ujayarat,tpeters}@robarst.ca

² Department of Urology, St. Josephs Hospital, London, Canada
stephen.pautler@sjhc.london.on.ca

Abstract. This paper seeks to develop fast and accurate endoscopic stereo 3-D scene reconstruction for image-guided robotic surgery. Although stereo 3-D reconstruction techniques have been widely discussed over the last few decades, they still remain challenging for endoscopic stereo images with photometric variations, noise, and specularities. To address these limitations, we propose a robust stereo matching framework that constructs cost function on the basis of image gradient and three-channel color information for endoscopic stereo scene 3-D reconstruction. Color information is powerful for textureless stereo pairs and gradient is robust to texture structures under noise and illumination change. We evaluate our stereo matching framework on clinical patient stereoscopic endoscopic sequence data. Experimental results demonstrate that our approach significantly outperforms current available methods. In particular, our framework provided 99.5 % reconstructed density of stereo images compared to other available matching strategies which achieved at the most an 87.6 % reconstruction of the scene.

Keywords: Stereo matching · Binocular endoscope · Laparoscopy · Endoscopic intervention · 3D Scene reconstruction · Cost construction · Cost aggregation · Robotic surgery · Prostatectomy · da Vinci surgical system

1 Introduction

Fast, accurate and robust dense stereo matching is required for many practical applications, e.g., scene surveillance and urban terrain 3-D reconstruction. Most of the reconstruction challenges often result from image sensor noise, specularity and reflection, illumination changes, and occlusion [2]. To tackle these challenges, numerous dense stereo matching algorithms have been published in the literature and mainly consist of *local* and *global* correspondence methods.

Konolige implemented a bock-matching method for stereo reconstruction [10]. He estimated by matching small patches of one image to another one

on the basis of the cross-correlation similarity. Local gradient-based methods, particularly optical flow [3], calculate the disparity map in terms of inter-frame motion and image brightness. Hirschmuller proposed a semi-global matching algorithm to improve local correspondence methods [7]. Computational stereo also can be formulated as a multi-labeling problem to compute the disparity map on the basis of the pixel-wise correspondence between the left and the right camera images. To solve such a disparity labeling problem, many global optimization approaches, e.g., intrinsic curves [15], graph cuts [1,9], dynamic programming [11,17], and belief propagation [14] or their convex counterparts, are used for reducing sensitivity to stereo image variations. Unfortunately, global optimization is usually a time-consuming procedure that can potentially be hampered by premature convergence.

As one of many minimally invasive surgery approaches, robotic surgery, e.g., robotic laparoscopic prostatectomy, usually uses binocular or stereo endoscopes to provide the surgeon with direct vision close visual proximity of the surface structures in the operating room. However, the binocular endoscopic cameras are a small with quite narrow field of view, which limits the surgeon's sense of accurate geometrical appreciation of the surgical site.

To expand the surgeon's view of the surgical site and enhance his ability to control the surgical instruments, stereoscopic endoscopic scene 3-D reconstruction, as one of interactive and augmented visualization strategies, has been discussed recently by the medical robotic community. To achieve accurate scene visualization, dense stereo correspondence plays a critical role in binocular endoscopic video reconstruction. Various methods have been proposed in recent literature to address such a stereo correspondence problem. Stoyanov et al. reported a quasi-dense correspondence method for laparoscopic stereo images without compensating tissue deformation [13], by first detecting and matching reliable features between stereo pairs. Using these matched features as landmark structures, they propagated them using zero-mean normalized cross correlation (ZNCC) as the similarity metric to determine the disparity map. However, their method fails to reconstruct image regions with uniform or homogeneous texture and only can obtain semi-dense reconstruction. Chang et al. [5] recently proposed a stereo matching approach that constructs a cost volume in terms of intensity information and aggregates such a volume in terms of ZNCC and convex optimization [4,8]. Although their proposed method outperforms the state-of-the-art, they did not use sufficient color information and also neglected gradient information on stereo pairs. Totz et al. [16] reported a similar matching method as the work of Stoyanov et al. [13] and obtained a semi-dense reconstruction.

Even though the methods discussed above work well in binocular endoscopic video reconstruction, they still remain challenging in the presence of photometric variations, specular reflections, noise, uniform texture, and occlusion. This work aims to develop fast and accurate binocular endoscopic 3-D scene reconstruction for robotic surgery. Being motivated by the fast cost-volume filtering method [8,12], we introduce a color and gradient-driven stereo correspondence framework to boost stereoscopic endoscopic scene reconstruction. Such a local

filtering method can avoid some of the drawbacks of the aforementioned schemes and yields a high-quality disparity map that is robust to illumination changes making it applicable in stereoscopic endoscopic scenarios.

The main highlight or contribution of this work lies in successfully constructing the correspondence cost aggregation on the basis of three-channel color and gradient image information for matching stereoscopic endoscopic pairs and achieving fully automatic dense 3-D reconstructions of surgical scenes. Recent robotic surgery systems, e.g., the clinically used *da Vinci* surgical system, provide high-definition (HD) binocular endoscopic images. These HD images provide sufficient color information to allow accurate scene reconstruction even if lacks texture and is noisy. Image gradient information can provide the means to overcome difficulties provided by illumination changes and noise.

The rest of this paper is organized as follows. Section 2 describes the various steps of our color and gradient-boosted aggregation stereo matching method. Experimental setups and results are shown in Sect. 3. Section 4 discusses different experimental results obtained from different stereo matching methods before concluding this work in Sect. 5.

2 Approaches

This section describes our color and gradient-boosted aggregation matching method that basically compares color and gradient differences between stereo pairs with left and right images. Currently available robotic surgical procedures usually provide HD endoscopic stereo videos. These HD stereo videos contain sufficient color information that can characterize the difference between stereo pairs more powerfully than intensity information directly used for the disparity computation in 3-D reconstruction. On the other hand, gradient information, as a simple image descriptor, is much more robust to illumination and exposure changes. Based on the work of Honsi et al. [8], our stereo matching framework first defines a cost volume function with respect to color and gradient information. We then aggregate the cost volume with color and gradient structures by the guided image filtering method [6]. An optimization strategy of *winner-takes-all* (WTA) is then employed for disparity selection. Finally, two postprocessing steps including occlusion removal and smoothness are performed to further improve the disparity accuracy. Each step is discussed below.

2.1 Cost Construction

Suppose \mathbf{I}_u and \mathbf{I}_v denote the left and right images of a stereo pair. Image \mathbf{I}_u with size of $W \times H$ pixels (similarly \mathbf{I}_v) has three color channels of \mathbf{I}_u^r , \mathbf{I}_u^g , and \mathbf{I}_u^b in RGB model space. For each pixel \mathbf{p} on \mathbf{I}_u and each pixel \mathbf{q} on \mathbf{I}_v , a color matching cost function $\mathcal{F}_\alpha(\mathbf{p}, \mathbf{q})$ minimizes average absolute color difference $d_C(\mathbf{p}, \mathbf{q})$ in each color channel:

$$\mathcal{F}_\alpha(\mathbf{p}, \mathbf{q}) = \min(d_C(\mathbf{p}, \mathbf{q}), \alpha), \quad (1)$$

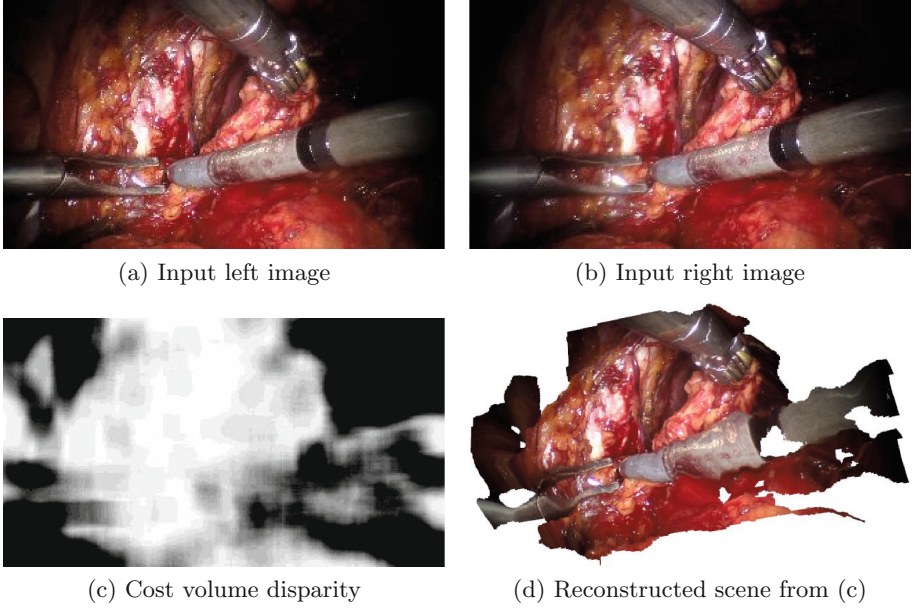


Fig. 1. Input stereo images and the reconstructed scene on the basis of the constructed cost volume disparity map

where α is a threshold and $d_C(\mathbf{p}, \mathbf{q})$ is calculated by

$$d_C(\mathbf{p}, \mathbf{q}) = \frac{1}{3} \sum_{i \in r, g, b} |\mathbf{I}_u^i(\mathbf{p}) - \mathbf{I}_v^i(\mathbf{q})|. \quad (2)$$

Moreover, we define a gradient matching cost function $\mathcal{F}_\beta(\mathbf{p}, \mathbf{q})$ to measure absolute gradient difference $d_G(\mathbf{p}, \mathbf{q})$ between the left and right images of the stereo pair:

$$\mathcal{F}_\beta(\mathbf{p}, \mathbf{q}) = \min(d_G(\mathbf{p}, \mathbf{q}), \beta), \quad (3)$$

where β is also a predefined threshold and $d_G(\mathbf{p}, \mathbf{q})$ computes derivative $\nabla_x \hat{\mathbf{I}}_u$ along the x -direction of stereo pairs:

$$d_G(\mathbf{p}, \mathbf{q}) = |\nabla_x \hat{\mathbf{I}}_u - \nabla_x \hat{\mathbf{I}}_v|, \quad (4)$$

$$\forall \mathbf{p} = (x, y), \nabla_x \hat{\mathbf{I}}_u = \frac{\hat{\mathbf{I}}_u(x+1, y) - \hat{\mathbf{I}}_u(x-1, y)}{2}, \quad (5)$$

where $\hat{\mathbf{I}}_u$ and $\hat{\mathbf{I}}_v$ are grayscale images converted from color images \mathbf{I}_u and \mathbf{I}_v . $\nabla_x \hat{\mathbf{I}}_v$ is similarly defined as $\nabla_x \hat{\mathbf{I}}_u$.

Based on Eqs. 1 and 3, a stereo image color and gradient-boosted matching cost function \mathcal{F}_μ can be defined as:

$$\mathcal{F}_\mu(\mathbf{p}, \mathbf{q}) = (1 - \mu)\mathcal{F}_\alpha(\mathbf{p}, \mathbf{q}) + \mu\mathcal{F}_\beta(\mathbf{p}, \mathbf{q}), \quad (6)$$

where the prefixed constant parameter μ is introduced to balance the color and gradient costs.

Finally, for pixel $\mathbf{p} = (x, y)$ on \mathbf{I}_u and its potential correspondence $\mathbf{q} = (x + \lambda, y)$ with disparity λ on \mathbf{I}_v , a cost volume $\mathcal{V}(\mathbf{p}, \lambda)$ is constructed in a 3-D space $\mathfrak{R}^3 = \{x \in [1, W], y \in [1, H], \lambda \in [\lambda_{min}, \lambda_{max}]\}$ (λ_{min} and λ_{max} are predefined minimal and maximal disparity values):

$$\forall (x, y, \lambda) \in \mathfrak{R}^3, \mathcal{V}(\mathbf{p}, \lambda) = \mathcal{F}_\mu(\mathbf{p}(x, y), \mathbf{q}(x + \lambda, y)). \tag{7}$$

Figure 1 shows the reconstructed surface (Fig. 1(d)) after constructing the cost volume disparity map (Fig. 1(c)) in terms of input left and right images (Fig. 1(a) and (b)).

2.2 Cost Aggregation

The cost construction is a generally ambiguous process. The correct matches might easily have a higher cost than the incorrect ones because of image noise. To deal with mismatches in the stereo pair, we use a guided filtering to weight each cost volume $\mathcal{V}(\mathbf{p}, \lambda)$ built above and obtain a filtered cost volume $\tilde{\mathcal{V}}(\mathbf{p}, \lambda)$ [6]:

$$\tilde{\mathcal{V}}(\mathbf{p}, \lambda) = \sum_{\mathbf{k}} \Psi_{\mathbf{p}, \mathbf{k}}(\mathbf{I}_u) \mathcal{V}(\mathbf{k}, \lambda), \tag{8}$$

which employs the left color image \mathbf{I}_u as the *guidance* image.

Let $\Omega_{\mathbf{o}}$ be a squared region, of size $a \times a$, centered at pixel \mathbf{o} on image \mathbf{I}_u . Weight $\Psi_{\mathbf{p}, \mathbf{k}}(\mathbf{I}_u)$ is computed by

$$\Psi_{\mathbf{p}, \mathbf{k}} = \frac{1}{N_{\mathbf{o}}} \sum_{\mathbf{o} \in \Omega_{\mathbf{o}}} (1 + (\mathbf{I}_u(\mathbf{p}) - \delta_{\mathbf{o}}) \Lambda (\mathbf{I}_u(\mathbf{k}) - \delta_{\mathbf{o}})), \tag{9}$$

where $N_{\mathbf{o}} = N_{\mathbf{p}} N_{\mathbf{k}}$, $N_{\mathbf{p}}$ and $N_{\mathbf{k}}$ are the number of pixels in squared regions $\Omega_{\mathbf{p}}$ and $\Omega_{\mathbf{k}}$, $\Omega_{\mathbf{o}} = \Omega_{\mathbf{p}} \cap \Omega_{\mathbf{k}}$, color image $\mathbf{I}_u = (\mathbf{I}_u^r, \mathbf{I}_u^g, \mathbf{I}_u^b)^T$, average color image $\delta_{\mathbf{o}} = (\delta_{\mathbf{o}}^r, \delta_{\mathbf{o}}^g, \delta_{\mathbf{o}}^b)^T$, and each average color channel of \mathbf{I}_u is calculated by

$$\delta_{\mathbf{o}}^i = \frac{1}{N_{\mathbf{o}}} \sum_{\mathbf{p} \in \Omega_{\mathbf{o}}} \mathbf{I}_u^i(\mathbf{p}), \quad i \in \{r, g, b\}. \tag{10}$$

Term $\Lambda = (\Sigma_{\mathbf{o}} + \epsilon \mathbf{A})^{-1}$ ($\mathbf{A} \mapsto 3 \times 3$ identity matrix) with smoothness factor ϵ and covariance matrix $\Sigma_{\mathbf{o}}$ of squared region $\Omega_{\mathbf{o}}$ from image \mathbf{I}_u . Matrix $\Sigma_{\mathbf{o}}$ is determined by $\Sigma_{\mathbf{o}}$, which is in accordance with variance $\sigma_{\mathbf{o}}^{ij}$ on $\Omega_{\mathbf{o}}$:

$$\Sigma_{\mathbf{o}} = \begin{pmatrix} \sigma_{\mathbf{o}}^{rr} & \sigma_{\mathbf{o}}^{rg} & \sigma_{\mathbf{o}}^{rb} \\ \sigma_{\mathbf{o}}^{rg} & \sigma_{\mathbf{o}}^{gg} & \sigma_{\mathbf{o}}^{gb} \\ \sigma_{\mathbf{o}}^{rb} & \sigma_{\mathbf{o}}^{gb} & \sigma_{\mathbf{o}}^{bb} \end{pmatrix}_{3 \times 3}, \quad i, j \in \{r, g, b\}. \tag{11}$$

$$\sigma_{\mathbf{o}}^{ij} = \frac{1}{N_{\mathbf{o}}} \sum_{\mathbf{p} \in \Omega_{\mathbf{o}}} \mathbf{I}_u^i(\mathbf{p}) \mathbf{I}_u^j(\mathbf{p}) - \delta_{\mathbf{o}}^i \delta_{\mathbf{o}}^j, \quad i, j \in \{r, g, b\}. \tag{12}$$

After aggregating the cost volume, the disparity and reconstruction become more accurate (Fig. 2(a) and (b)).

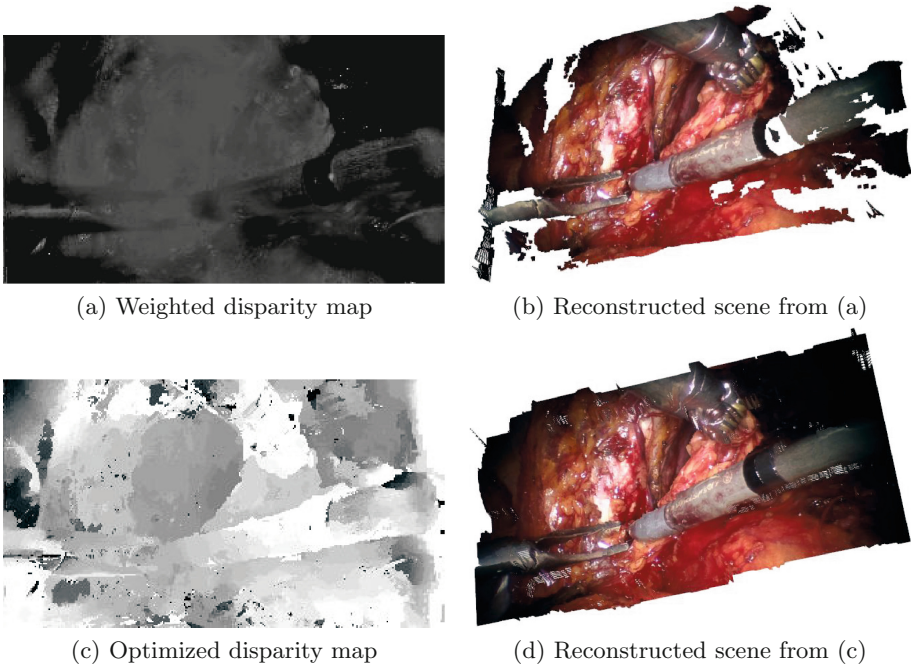


Fig. 2. Weighted disparity map (a) from Fig. 1(a), optimized disparity map (c) from (a), reconstructed scenes (b) and (d)

2.3 Disparity Determination

Based on the aggregated cost volume $\tilde{V}(\mathbf{p}, \lambda)$, we use WTA to select disparity $\tilde{\lambda}(\mathbf{p})$ for minimizing the matching cost relative to pixel \mathbf{p} within range $\mathcal{D} = [\lambda_{min}, \lambda_{max}]$:

$$\tilde{\lambda}(\mathbf{p}) = \arg \min_{\lambda \in \mathcal{D}} \tilde{V}(\mathbf{p}, \lambda). \quad (13)$$

Figure 2(c) displays the chosen disparity map and its corresponding reconstructed scene (Fig. 2(d)).

2.4 Occlusion Removal

Occlusion is often unavoidable in stereo matching procedures. Detection of occlusion involves determining whether the calculated disparity in the left image is consistent with that of the right image. An occluded pixel will be removed and assigned the minimal disparity value if $\tilde{\lambda}(\mathbf{p}) \neq -\lambda(\mathbf{p} + \tilde{\lambda}(\mathbf{p}))$. To consider true or subpixel disparity, we employ a more general constraint for removing occluded pixels:

$$|\tilde{\lambda}(\mathbf{p}) + \lambda(\mathbf{p} + \tilde{\lambda}(\mathbf{p}))| > 1. \quad (14)$$

Figure 3(a) shows the detected occlusion using Eq. 14. Figure 3(b) gives the result after occlusion rejection.

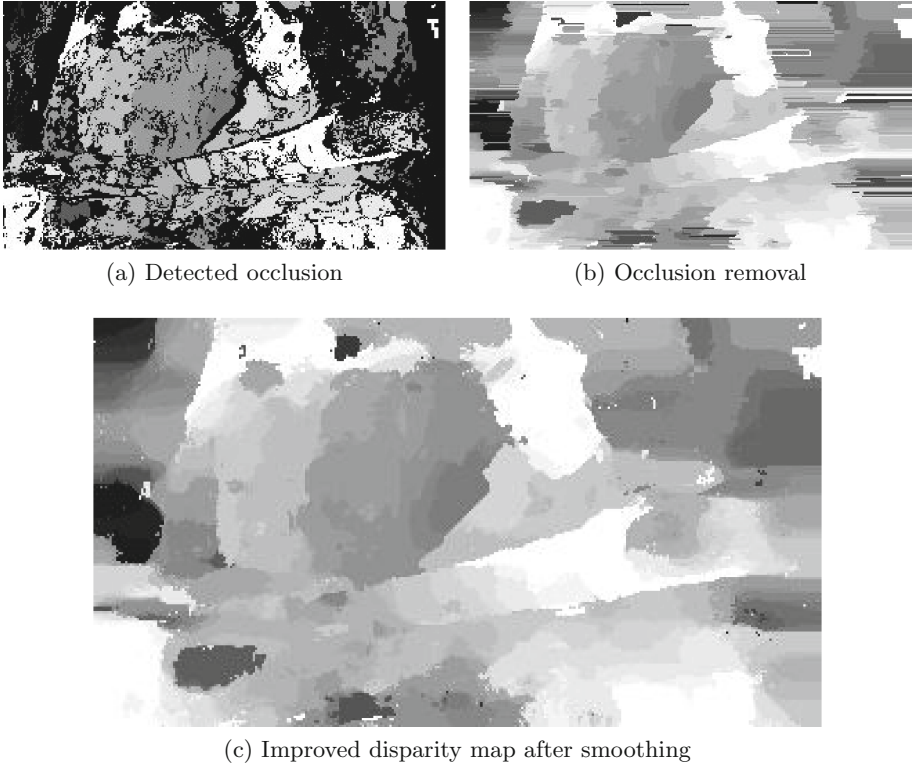


Fig. 3. Accuracy-enhanced disparity by occlusion rejection and smoothing from the optimized disparity in Fig. 2(c)

2.5 Disparity Smoothness

The occlusion removal step usually introduces edge blurring in the disparity map (Fig. 3(b)). To preserve the edge in the disparity map, we use bilateral filtering to weight each disparity after the occlusion rejection step:

$$\varphi_{st} = \exp\left(-\frac{\|\mathbf{s} - \mathbf{t}\|^2}{\gamma_1^2} - \frac{\|\mathbf{I}_u^i(\mathbf{s}) - \mathbf{I}_u^i(\mathbf{t})\|^2}{\gamma_2^2}\right), \tag{15}$$

where pixel \mathbf{t} is located at squared region Ω_s with size of $s \times s$ centered at pixel \mathbf{s} , γ_1 and γ_2 indicate pixel spatial and color similarities, and $\|\cdot\|$ denotes the Euclidean distance ($i \in r, g, b$). We then calculate the cumulative histogram $\mathcal{H}(\tilde{\lambda})$ for every disparity value $\tilde{\lambda}$:

$$\forall \tilde{\lambda} \in \mathcal{D} = [\lambda_{min}, \lambda_{max}], \mathcal{H}(\tilde{\lambda}) = \sum_{\mathbf{t} \in \Omega_{\mathbf{s}} | \tilde{\lambda}(\mathbf{t}) \leq \tilde{\lambda}} \varphi_{st}. \tag{16}$$

Finally, the optimal disparity $\tilde{\lambda}_*(\mathbf{p})$ at pixel \mathbf{p} is decided by

$$\tilde{\lambda}_*(\mathbf{p}) = \{\tilde{\lambda} | \mathcal{H}(\tilde{\lambda}) \geq 0.5\mathcal{H}(\lambda_{max})\}. \tag{17}$$

Algorithm 1. Cost volume stereo matching

Input: Stereo pairs with left and right color images, \mathbf{I}_u and \mathbf{I}_v , disparity range $\mathcal{D} = [\lambda_{min}, \lambda_{max}]$

Output: Disparity map $\tilde{\lambda}_*(\mathbf{p})$ at each pixel \mathbf{p}

❶ Calculate Σ_o and δ_o (Eqs. 10~12), $\tilde{\lambda}_*(\mathbf{p}) \mapsto \infty$;

for $\lambda = \lambda_{min}$ **to** λ_{max} ($\lambda \in \mathcal{D}$) **do**

❷ Cost volume construction (Sect. 2.1);

for each pixel \mathbf{p} on left image \mathbf{I}_u **do**

Compute $d_C(\cdot)$ and $d_G(\cdot)$ (Eqs. 2 and 4);

Compute cost \mathcal{F}_μ w.r.t. \mathbf{p} for λ (Eq. 6);

Set cost $\mathcal{V}(\mathbf{p}, \lambda)$ for pixel \mathbf{p} (Eq. 7);

end

❸ Cost volume aggregation (Sect. 2.2);

Calculate Σ_o and δ_o (Eqs. 10~12);

for each pixel \mathbf{p} on left image \mathbf{I}_u **do**

Use the guided filtering to weight $\mathcal{V}(\mathbf{p}, \lambda)$ for edge-preserving and obtain $\hat{\mathcal{V}}(\mathbf{p}, \lambda)$ (Eq. 8);

end

❹ Disparity determination (Sect. 2.3);

for each pixel \mathbf{p} on left image \mathbf{I}_u **do**

Select disparity by WTA and obtain $\tilde{\lambda}(\mathbf{p})$ (Eq. 13);

end

end

for each $\tilde{\lambda}(\mathbf{p})$ on disparity map $\{\tilde{\lambda}(\mathbf{p})\}_{W \times H}$ **do**

❺ Occlusion removal (Sect. 2.4);

❻ Disparity smoothing (Sect. 2.5);

Return $\tilde{\lambda}_*(\mathbf{p})$;

end

Figure 3(b) shows the disparity map with edge blurring after the occlusion is removed. Figure 3(c) gives the final disparity map after the smoothness step using the bilateral filtering.

In general, the color and gradient-driven matching method for 3-D reconstruction is implemented in Algorithm 1.

3 Experimental Results

Clinical stereoscopic endoscopic video sequences were recorded by the *da Vinci* Si surgical system (Intuitive Surgical Inc., Sunnyvale, CA, USA) during a robotic prostatectomy procedure. The HD (1920×1080) video frames were downsampled to 320×180 pixels. We investigate several methods as follows: (1) stereo block-matching (SBM) [10], (2) semi-global matching (SGM) [7], (3) quasi-dense stereo matching (QSM) [13], and (4) our color and gradient-boosted cost aggregation matching (CAM), as discussed in Sect. 2. All these methods were tested

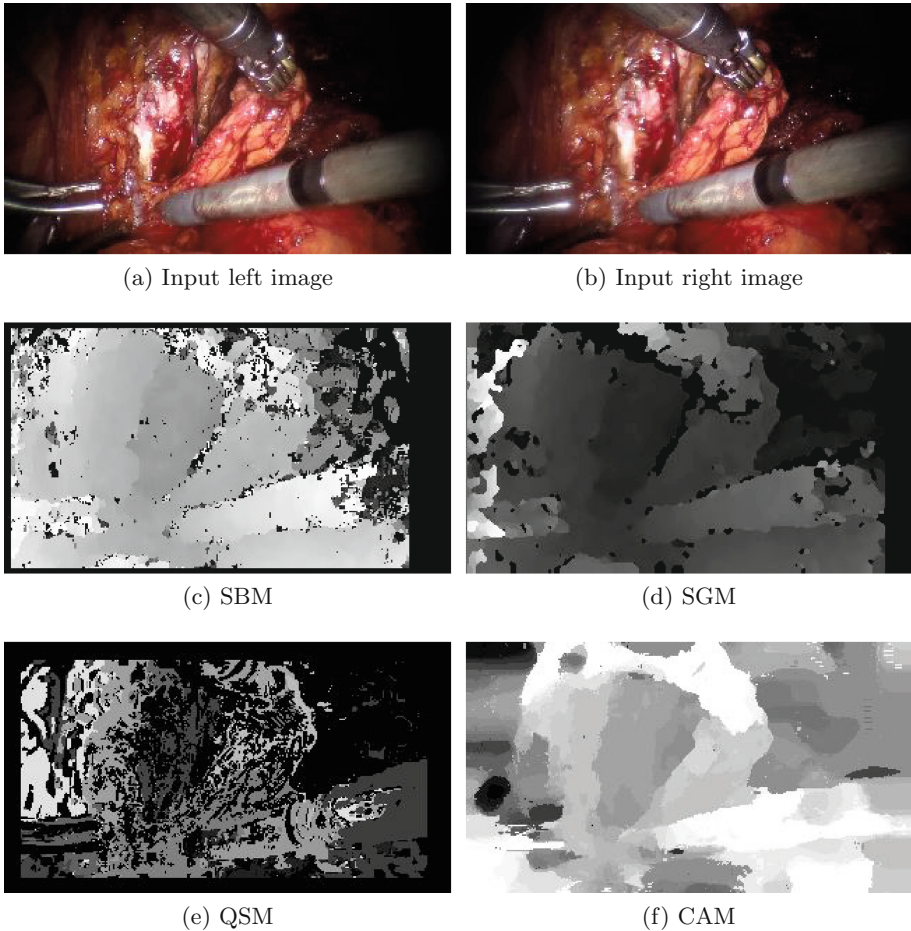


Fig. 4. Visual comparison of disparity maps of using matching methods of SBM [10], SGM [7], QSM [13] and CAM.

on a laptop installed with Windows 8.1 Professional 64-Bit System, 16.0-GB Memory, and Processor Intel(R) Core(TM) i7 CPU×8 and were implemented on Microsoft Visual C++.

Figure 4 compares the disparity maps estimated from different stereo matching methods. Figure 5 visualizes the reconstructed scenes in terms of the disparity maps from Fig. 4. Figure 6(a) compares the reconstructed density of using different correspondence methods. The percentage of the reconstructed density was 81.7%, 87.6%, 60.8%, and 99.5% when using SBM [10], SGM [7], QSM [13] and CAM, respectively. The proposed method almost reconstructed all the pixels from the stereo color images. Figure 6 investigates the processing time of the disparity computation of each compared method. The processing time of the dis-

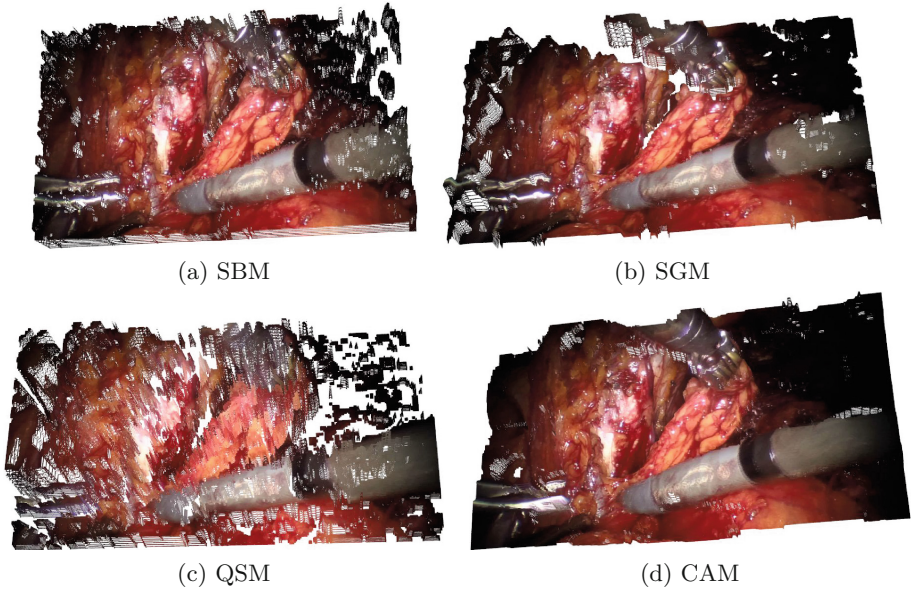


Fig. 5. Visual comparison of reconstructed scenes corresponding to disparity maps from Fig. 4

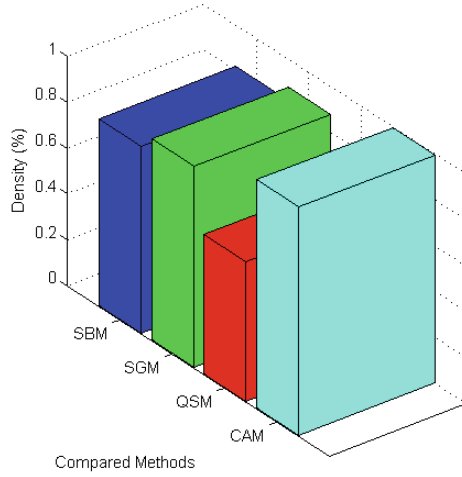
parity calculation was 143, 146, 532, and 428 milliseconds (ms) using SBM [10], SGM [7], QSM [13] and CAM.

4 Discussion

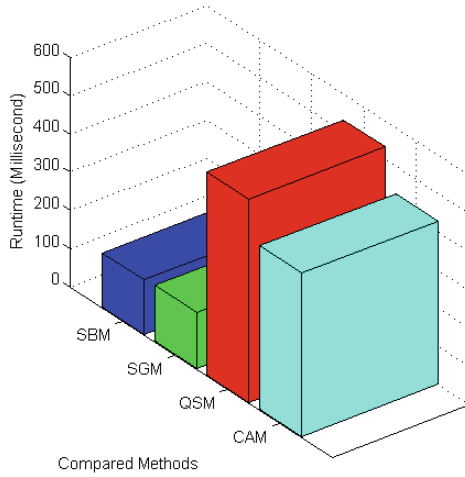
The objective of this work is to explore fast and robust 3-D scene reconstruction of binocular endoscopic video sequences with image variations including noise, occlusion, specular reflection, and motion blurring. In general, we achieved fast and accurate binocular endoscopic stereo scene reconstruction for augmenting image-guided robotic surgery.

The effectiveness of our proposed stereo matching approach lies in two aspects of the matching cost aggregation function. First, sufficient color information was propagated in the disparity map estimation. Compared to intensity-based local matching methods (e.g., SBM and SGM) [7, 10], color is a more powerful characteristic to tackle textureless noisy stereo images. Furthermore, gradient-based cost aggregation can preserve the disparity edge and be robust to illumination variation, occlusion, and motion blurring. In contrast to feature-driven semi-dense matching (e.g.), the gradient cost does not suffer from the featureless problem. The color and gradient-boosted correspondence strategies balance each other to construct the matching cost enabling it to outperform other current available stereo matching methods.

Although our proposed stereo matching method performs much better than others, it still suffers from image sensor noise and edge blurring problems, as



(a) Reconstruction density



(b) Processing time

Fig. 6. Comparison of reconstruction density and processing time of disparity computation of using different methods

shown in Fig. 7. To address these problems, edge-preserving filtering methods, which are robust to image variations, can be used to improve the quality of the computed disparity map during binocular endoscopic scene reconstruction. On the other hand, we currently employ simple optimization method, i.e., the WTA strategy, which is sensitive to image noise. Despite the fact that global optimization methods for dense stereo matching are somewhat time-consuming, they

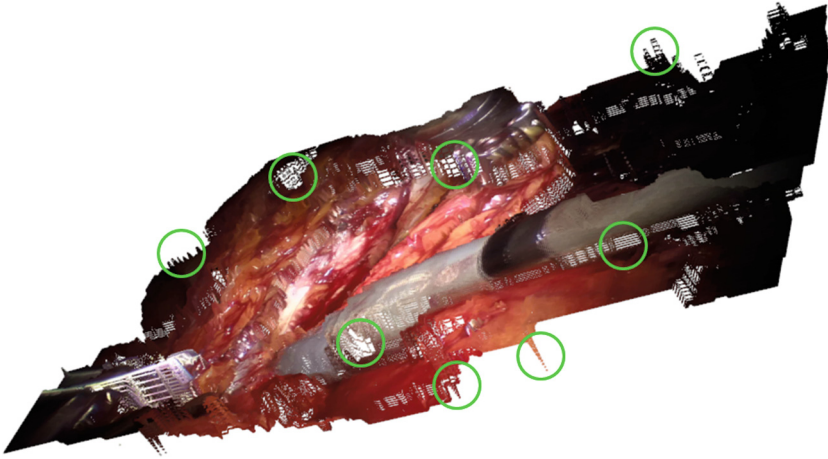


Fig. 7. An example of collapsed reconstruction regions (*green circles*) due to image sensor noise and edge blurring

are nevertheless another option to further deal with image noise. Additionally, we still need to improve the processing time of the disparity map calculation. Graphics processing unit (GPU) and multi-threading programming techniques can be used to accelerate our stereo matching and reconstruction procedures to meet real-time requirements in clinical applications.

5 Conclusions

This work proposed a color and gradient-boosted cost aggregation stereo matching framework to reconstruct binocular endoscopic video scene. We integrate color and gradient information in HD images to construct the matching cost function making it robust to stereoscopic endoscopic image variations. We then employ guided image filtering procedure to weight each disparity value in the disparity map. Experimental results demonstrate that our proposed stereo correspondence framework significantly outperforms other matching methods. The reconstruction density improved from 87.6% to 99.5%. Future work includes reduction of the computational time for stereo matching, improvement of noise insensitivity, edge-preserving filtering procedure, and combination of global optimization methods.

Acknowledgements. This work was supported in part by the Canadian Institutes for Health Research (CIHR) Team Grant on Prostate Imaging, the Canadian Foundation for Innovation, and the Ontario Research Fund.

References

1. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(9), 1124–1137 (2004)
2. Brow, M.Z., Burschka, D., Hager, G.D.: Advances in computational stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(8), 993–1008 (2003)
3. Brox, T., Malik, J.: Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(3), 500–513 (2011)
4. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**(1), 120–145 (2011)
5. Chang, P.-L., Stoyanov, D., Davison, A.J., Edwards, P.E.: Real-time dense stereo reconstruction using convex optimisation with a cost-volume for image-guided robotic surgery. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) *MICCAI 2013, Part I. LNCS*, vol. 8149, pp. 42–49. Springer, Heidelberg (2013)
6. He, K., Sun, J., Tang, X.: Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(6), 1397–1409 (2013)
7. Hirschmuller, H.: Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(2), 328–341 (2008)
8. Hosni, A., Rhemann, C., Bleyer, M., Rother, C., Gelautz, M.: Fast cost-volume filtering for visual correspondence and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(2), 504–511 (2013)
9. Kolmogorov, V., Zabih, R.: Computing visual correspondence with occlusions using graph cuts. In: *IEEE International Conference on Computer Vision*, vol. 2, pp. 508–515 (2001)
10. Konolige, K.: Small vision systems: hardware and implementation. In: *International Symposium on Robotic Research, Part 6*, pp. 203–212 (1998)
11. Meerbergen, G.V., Vergauwen, M., Pollefeys, M., Gool, L.V.: A hierarchical symmetric stereo algorithm using dynamic programming. *Int. J. Comput. Vis.* **47**(1–3), 275–285 (2002)
12. Rhemann, C., Hosni, A., Bleyer, M., Rother, C., Gelautz, M.: Fast cost-volume filtering for visual correspondence and beyond. In: *IEEE Conference on Computer Vision and Pattern Recognition - CVPR 2011*, pp. 3017–3024 (2011)
13. Stoyanov, D., Scarzanella, M.V., Pratt, P., Yang, G.-Z.: Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010, Part I. LNCS*, vol. 6361, pp. 275–282. Springer, Heidelberg (2010)
14. Sun, J., Zheng, N.N., Shum, H.Y.: Stereo matching using belief propagation. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(7), 787–800 (2003)
15. Tomasi, C., Manduch, R.: Stereo matching as a nearest-neighbor problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(3), 333–340 (1998)
16. Totz, J., Thompson, S., Stoyanov, D., Gurusamy, K., Davidson, B.R., Hawkes, D.J., Clarkson, M.J.: Fast semi-dense surface reconstruction from stereoscopic video in laparoscopic surgery. In: Stoyanov, D., Collins, D.L., Sakuma, I., Abolmaesumi, P., Jannin, P. (eds.) *IPCAI 2014. LNCS*, vol. 8498, pp. 206–215. Springer, Heidelberg (2014)
17. Veksler, O.: Stereo correspondence by dynamic programming on a tree. In: *IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 384–390 (2005)