

# Two-frame stereo photography in low-light settings: A preliminary study

Kartic Subr  
University College London

Gwyneth Bradbury  
University College London

Jan Kautz  
University College London

## ABSTRACT

Image-pairs captured from a rig of two, carefully arranged cameras are increasingly used to reconstruct partial 3D information. A crucial step in this reconstruction is the matching of points in the two images that are projections of the same 3D point through each camera. Despite receiving much attention, algorithms to match corresponding points in two-frame stereo images are both slow, as well as surprisingly fragile. The problem is exacerbated by noise or blur in the input images because of the potential ambiguities they introduce in the matching process.

For scenes that are poorly illuminated, it is necessary to make a combination of three adjustments: To increase the size of the aperture to allow more light; to increase the duration of exposure; and to increase the sensor-gain (ISO). These adjustments potentially introduce defocus, motion blur and noise — all of which adversely affect reconstruction. We present an exploratory study of how they relatively affect stereo-correspondence algorithms by comparing the accuracy and precision of three reconstruction algorithms over the space of exposures.

## 1. INTRODUCTION

As 3D films have become increasingly commonplace, capturing digital 3D footage has also become mainstream technology. The 3D production pipeline often entails reconstructing scene depth, e.g., in order to embed virtual objects or in order to perform depth grading [17]. While there are commercial tools that perform depth reconstruction, such as Nuke/Ocula [16], it is notoriously challenging [14, 20, 21, 22], especially in low light situations where the signal to noise ratio decreases.

One of the hurdles is to robustly, and automatically, identify pixels in a left-right image pair that correspond to the same point in 3D. Typically this step of finding correspondences relies on a comparison of features across the image pair. While this is already a difficult and potentially ambiguous problem, it is even more difficult when the images are blurred or noisy. Stereo matching methods are not inherently designed for low signal to noise ratio (SNR) reconstruction, often failing when exposure time, aperture diameter or sensor gain are increased or become sub-optimal. Essentially,

in low-light situations, it is impossible to take well-exposed, sharp images with a wide depth of field such that a good stereo reconstruction can be achieved.

Sensor gain, exposure time and aperture are the obvious parameters to tune in order to obtain a better exposure but are also distinct sources of image noise. In the case of low light stereo capture, a compromise must be made: either increasing the aperture, leading to a shallow depth of field and defocus blur; increasing sensor gain, and therefore increasing the quantization error and reducing the SNR; or increasing the exposure duration, and increasing noise due to motion blur. In this paper, we explore the effect of these settings on stereo reconstruction quality. To this end, we have captured stereoscopic images of several scenes under many different aperture, exposure, and ISO settings. We then quantify the reconstruction error of three different stereo algorithms [3, 7]. Whilst there have been many quantitative comparisons of stereo matching algorithms [23, 20], these techniques have not been compared on light-limited scenes.

### 1.1 Related Work

Low SNR is a significant challenge for stereo matching and reconstruction algorithms. The resulting image noise causes two matching pixels which should have the same intensity (under Lambertian assumption) to differ. This becomes very apparent in footage captured under poor lighting conditions. The problem is amplified further when two different cameras are used in stereo capture. Depending on the matching technique being employed, some processing is often necessary to correct the issue of false matches due to noise.

Increasing sensor gain is clearly one solution to the problem of low-light, but this leads to a notable increase in quantisation noise, causing the uncertainty of a given stereo match to increase further [1].

Alter et al [1] create a new noise measure which compares well against a Euclidean Norm (which tends to break down under the uncertainty of low SNR) and note that increasing sensor gain means that the quantisation error becomes the dominant source of noise due to the lack of intensity resolution and the lack of dynamic range.

Another solution in low lighting conditions is to increase the exposure time. This of course can lead to an increase in image blur due to camera shake or subject motion. Xu and Jia [24] produce a coarse stereo reconstruction based on feature points in the original blurred images, use this to estimate the point spread function and deconvolve the image. They then produce a fine stereo reconstruction. Structured blur, or coded aperture, can also help to estimate better disparity maps, provided the blur kernel is known [15].

Heo et al [12] present an algorithm which simultaneously solves the image de-noising and stereo reconstruction problems. The so-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CVMP '12, December 5 – 6, 2012, London, United Kingdom  
Copyright 2012 ACM 978-1-4503-1311-7/12/11 ...\$15.00.

lution presented avoids using the L1 and L2 norm distance metrics which are known to fail under changes in illumination and instead proposes a new metric based on restored intensity and non-local pixel distribution dissimilarity around matched pixels.

Heo et al [13] begin by applying synthetic noise to traditional data sets, showing that noisy imagery can result in serious inaccuracies. Using NCC and belief propagation results in lower error but a very ‘blocky’ result and methods solely using pixel-based intensity are also very sensitive.

However, the above are all post-processing techniques which give valuable improvements on degraded data sources. Another option is to preprocess the scene before capturing any data such that the resulting images that are obtained lead to a much better reconstruction.

Combining a range of exposures (exposure stacking) is one such technique commonly used in noise reduction and the capture of high dynamic range images. Hasinoff and Kutulakos [8] assess the optimal set of images needed to capture the full dynamic range of a scene, with the aim of minimising noise. They find that contrary to the normal practice of selecting a low ISO, using high ISO settings can enable significant improvements in the SNR.

Further, the authors later show that a dense sequence of wider apertures can be used as a faster alternative to taking a single long-exposure shot for the desired depth of field. Hasinoff and Kutulakos [11, 10] also attempt to determine the optimum number of photos to be taken given a fixed time constraint, addressing the compromise between defocus blur and sensor related noise. Results are assessed based on the uncertainty in resolving scene depth.

Confocal stereo [9] also addresses the problems of capturing low light scenes, noting that by controlling the focus and aperture (focus stacking), the intensity of a given visible scene pixel changes independently of the scene. The technique relies on a prior radiometric lens calibration.

We identify the following three leading stereo algorithms which are either known to produce good results on the Middlebury dataset [14, 20, 21, 22], common implementations of well-know algorithms or adhere to the time-restriction of real-world stereo applications.

The Patch Match stereo algorithm [2, 3], currently a high performer on the Middlebury dataset, targets planar, slanted surfaces which are notoriously hard to reconstruct, relying on a local algorithm.

Efficient Large-scale Stereo Matching(ELAS) [7], also performs well on the Middlebury dataset, relying on a generative probabilistic model for stereo matching and producing good, dense matches without the need for global optimisation. ELAS has a very quick implementation and can produce results in close to real time. Resulting features are also inherently robust to certain illumination changes.

The third algorithm, Graph Cuts [4, 19], is chosen because of its easy-to-access OpenCV implementation and frequent application in stereo vision tasks. Graph cuts applies an energy minimisation solution to the correspondence problem but treats the two images asymmetrically, and does not make fully exploit the information in both images. Freedman and Turek [5] propose an illumination-invariant extension to the Graph Cuts algorithm.

## 2. LOW-LIGHT STEREO PHOTOGRAPHY EXPERIMENTS

We conducted our experiments in three main settings: One static scene and camera with predominantly fronto-parallel surfaces (Fig. 2a); the same scene with camera motion (Fig. 2b) and an out-

door scene with a background that is slightly out of focus (Fig. 2c).

The stereo image pairs were captured using a Canon EOS 7D along with a Loro stereo lens (9005a) and camera motion was achieved by introducing a camera shake and, as a stereo lens, pairs of images underwent identical camera shake.

For each of these scenes, we acquired stereo image pairs and used them to estimate disparities with three different algorithms: A fast and robust, but less accurate, reconstruction algorithm as implemented in the library libELAS [7]; a slower graph-cuts optimisation [4, 19] as implemented in openCV; and finally, Patchmatch stereo [2, 3]. Example disparity maps for different combinations of scene and algorithm are visualized in Figures 1 and 9.

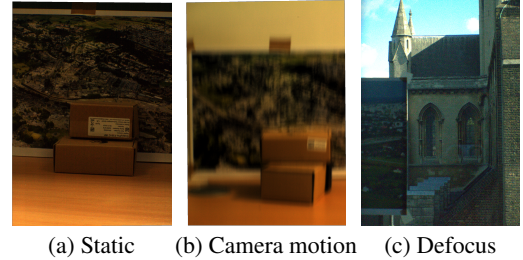


Figure 2: Left-images from our input dataset.

Ideally, the estimated disparities would be constant across pixels on a fronto-parallel surface (as indicated by the user). We reconstruct the disparities over these surfaces, to assess the accuracy and dispersion of the methods under different exposure settings. We use the relative error of the mean of the disparities over each fronto-parallel surface as a measure of accuracy and the normalized variance within the region as a measure of dispersion (see Fig. 3). To understand the importance of each, imagine a fronto-parallel textured plane as foreground with a differently textured background. An *accurate algorithm* is one that provides a reliable disparity estimate, on average, when the pixels covered by the foreground plane are known. A *precise algorithm* (with low variance), on the other hand, will be more useful to detect the boundary of the foreground plane.

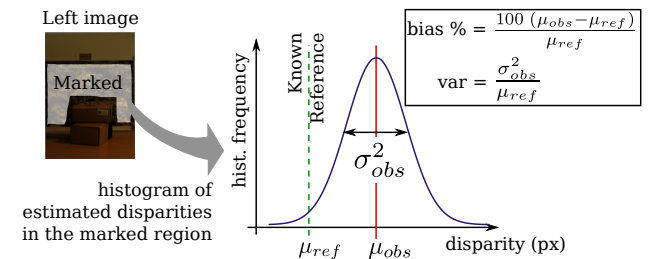
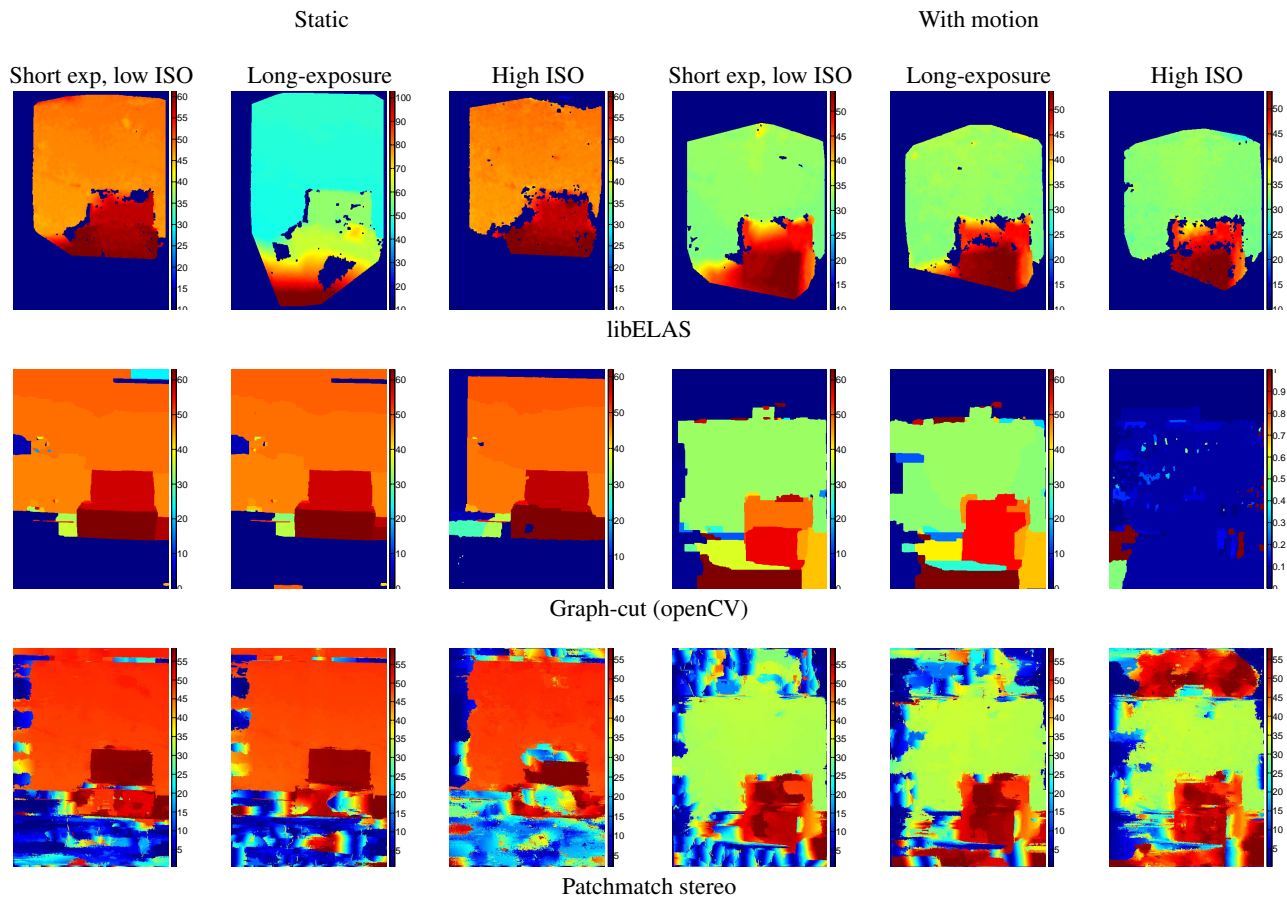


Figure 3: Our measures for accuracy and dispersion given a marked image and estimated disparities. We term as bias, as the relative error of the mean estimated value with respect to the ground truth reference value  $\mu_{ref}$  within the marked region. We will refer to the variance normalized by  $\mu_{ref}$  as simply the variance. This normalization simplifies quantitative comparison across images and different depths in an image.

In our experiments, we vary two different parameters related to exposure — exposure time and camera sensor gain (ISO) — with constant illumination, and tabulate the bias and variance of each of the three reconstruction algorithms. We then visualize this tabulated data to assess the effects of increased noise due to long ex-



**Figure 1: Disparity maps computed using three different algorithms (rows). libELAS (top row) is fast and robust but approximate. openCV’s method (second row) is reasonably robust for the static case but performs poorly with motion blur. Patchmatch stereo is sensitive to noise as well as motion blur.**

posures or high gain and increased blur due to motion or a wide aperture.

The effect of motion-blur versus noise due to high-gain is summarised in Figure 4, where image-pairs used in the reconstructions were acquired using a camera in motion. The figure shows two user-indicated fronto-parallel regions (top row) and the errors (mean within scribbled region - ground truth value) and variances of the reconstructed disparities in those regions, using the three different algorithms (rows). Each graph visualizes the errors (or variances) of the disparities at the different ISO (Y-axis) settings and exposure times (X-axis). The experiment is conducted on two different surfaces (indicated with white scribbles), one with a subtle texture (left) and another with a high-contrast texture (right). For the former, Patchmatch stereo performs better with noise to high-ISO than blur due to long exposures while the other two algorithms perform better with long exposures and low ISO. For the textured region, on the other hand, increasing the ISO worsens the performance of Patchmatch stereo. This could potentially be attributed to spurious matches in the repetitive pattern, due to the noise.

When the above experiment was repeated without camera motion, indeed, the algorithms performed better with longer exposures than high-ISO (Figure 5). A similar experiment was also performed on an outdoor scene with two user-indicated regions, one in focus and the other slightly out of focus (Figure 6). As expected, high-ISO was again a bigger problem than longer exposure time. In the

outdoor scene, with more available light in the scene, a combination of high-ISO and long exposures deteriorated reconstruction results possibly due to over-exposure.

Figure 8 summarises the performance of the different algorithms (columns) for each input image (rows). Each graph visualizes variances (dashed curves) and bias (solid curves) for different exposure values for two scribbled regions in each image. Splines were fit with a stiffness coefficient of 0.05. The exposure values (X-axis) were computed as [18]

$$e = \log_2 \frac{N^2}{t} + \log_2 \frac{S}{100} \quad (1)$$

where  $N$  is the relative aperture (f-number),  $t$  is the exposure time and  $S$  is the ISO. A high exposure value may be realized by either decreasing exposure time, gain or aperture width (larger f-number  $N$ ). Thus, exposure settings mapping to a high exposure value are chosen for brightly illuminated scenes.

## 2.1 Implementation details

The stereo image pairs were captured using a Canon EOS 7D along with a Loreo stereo lens (9005). Given the fixed camera intrinsics for each view, calibration was not necessary. Vertical rectification of the image pairs was done using uncalibrated epipolar rectification [6].

The three algorithms used were libELAS [7], openCV’s graph-

cut-based reconstruction and an implementation of Patchmatch stereo.

LibELAS works by finding matching points on a regular grid using the  $l_1$  distance between vectors composed of the horizontal and vertical Sobel filter responses in a 9 by 9 window. Patchmatch stereo works by optimising a Nearest Neighbour field as a function of offsets defined over all of the possible patch centres in the first image, for some distance function between the two patches.

LibELAS performed the reconstruction in the order of milliseconds while the latter two took typically 4 to 10 minutes. All reconstructions were run on a laptop with an Intel i7 quad-core processor and 8GB RAM, on  $640 \times 480$  pixel images.

### 3. DISCUSSION AND CONCLUSION

**High-ISO** We observe that both accuracy and precision are poor when the ISO is boosted to extremely high values keeping exposure short. This is expected, since the signal to noise ratio is known to be low in this setting. However, in the presence of large blur due to motion in textured areas with repetitive patterns, we observed that noise is preferable to blur. Despite the decreased accuracy, the variance of the reconstructed disparities is low. That is, while increasing the gain may yield an incorrect depth estimate, depth discontinuities (boundaries) are more easily identified.

**Long exposure time** Longer exposures result in low accuracy and precision unless the ISO is boosted. While this is intuitive for dynamic scenes, where motion blur plays spoilsport to finding correspondences, surprisingly, we notice this is the case even for static scenes. A possible cause is dark-current noise in longer exposures. Further study is required to ascertain the cause of this observation. Since short exposures have a low SNR and long exposures either introduce blur or a low SNR, identifying a suitable choice of exposure time is non-trivial.

**Wide aperture** Our experiment on the effect of defocus was limited by the Loreo lens, which does not provide very wide aperture settings. In our experiment at  $f/11$  (see bottom two rows of Fig. 8), defocus did not pose a significant problem to the correspondence algorithms in our set. We tested individually with the foreground and background in focus. Our observation is that the relative error is always higher for objects that are closer. This is, perhaps, due to a combination of the need for searching for correspondences within a larger neighbourhood as well as the higher absolute values of disparity for points that are closer. The Patchmatch correspondence algorithm exhibited notably higher bias relative to the other two algorithms.

**Accuracy-precision trade-off** In the static scene, we observe qualitatively that accuracy and precision are positively correlated (see first row of Fig. 8). However, in the presence of motion or defocus, their relationship is complex and dependent on the image content. We notice in the plots of Fig. 4 that for long exposures, with low ISO, the variance is low in places where the bias is high.

**Accuracy and precision against exposure-values** Fig. 8 shows that, for a static scene (row 1) in dim illumination, choosing a lower exposure value generally results in lower bias as well as variance. This is also the case for regions that are out of focus (red curves in row 3). However in the presence of motion or defocus, the curves are less predictable. The behaviour of these curves stresses the need for a deeper study of the effects of motion blur and defocus on stereo-correspondence algorithms. The absolute values of the normalized biases and variances suggest that, of the three algorithms compared, Patchmatch stereo works best for our scene with

motion, the GraphCut algorithm works best for static scenes and libElas works best for our scene with defocus. By 'best', we mean the algorithm with lowest bias as well as variance.

### 3.1 Conclusion

We presented a preliminary study of the effects of changing aperture settings, sensor gain and exposure time on reconstruction for binocular stereo images shot in low-light settings. Our primary observation is that the choice of exposure settings has a statistically significant effect on the accuracy and precision of stereo-correspondence algorithms. We observed that, under low-light settings, attaining high accuracy as well as high precision requires careful identification of a small region of exposure space. We compared three stereo-correspondence algorithms and found that Patchmatch stereo works best for our scene with motion (lowest bias and variance), the GraphCut algorithm works best for static scenes and libElas works best for our scene with defocus. Based on our experiments, we also propose the following guidelines on adjusting camera exposure.

1. **Static scenes:** Short exposures with high gain are detrimental to both accuracy as well as precision, regardless of the image content. For a fixed exposure time, choose the highest gain possible without overexposing for better accuracy as well as gain. For a fixed gain value, choose medium exposures over longer exposures.
2. **Dynamic scenes:** Long exposures with low gain and short exposures with high gain are both detrimental to accuracy, with the latter being preferable. For greater precision, in the presence of high-contrast texture, short exposures with high gain results is preferable. For greater precision, in the absence of high-contrast texture, medium gain with longer exposures is preferable.

We conclude that exposure settings for attaining a compromise on accuracy and precision are non-trivial to identify since they depend on the content of the image-pairs. We believe that automatically identifying these settings within the space of exposures is an exciting area for future research.

### 4. ACKNOWLEDGEMENTS

We thank Neill Campbell for his image-rectification program. We thank the anonymous reviewers for their suggestions. Kartic Subr acknowledges funding, in the form of the Newton Fellowship, from the Royal Academy of Engineering. Gwyneth Bradbury acknowledges funding from the EPSRC and Disney Research.

### 5. REFERENCES

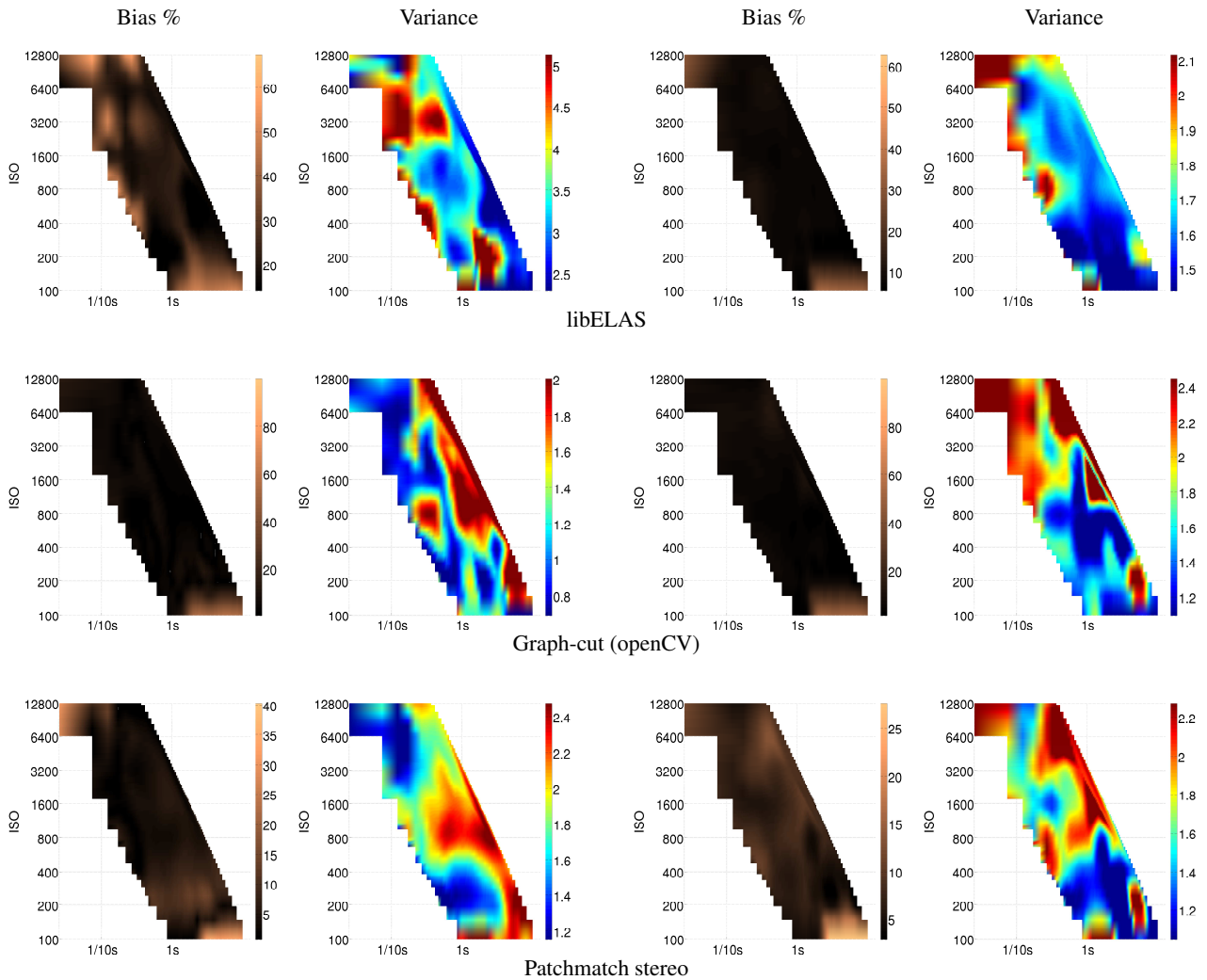
- [1] F. Alter, Y. Matsushita, and X. Tang. An Intensity Similarity Measure in Low-Light Conditions. In *European conference on Computer Vision*, pages 267–280, 2006.
- [2] D. Barnes, Connelly, Schechtman, Eli, Finkelstein, Adam, Goldman. PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing, 2009.
- [3] M. Bleyer and C. Rother. PatchMatch Stereo - Stereo Matching with Slanted Support Windows. In *British Machine Vision Conference*, 2011.
- [4] Y. Boykov. Markov Random Fields with Efficient Approximations. *IEEE Computer Vision and Pattern Recognition*, 1998.
- [5] D. Freedman and M. W. Turek. Illumination-Invariant Tracking via Graph Cuts. *IEEE Computer Vision and Pattern Recognition*, 2:10–17, 2005.



Ground truth disparity (selection) = 47.67



Ground truth disparity (selection) = 32.33



**Figure 4: A comparison of stereo reconstruction algorithms over the space of ISO and exposure time when the camera is hand-held. While high-ISO settings allow quick exposures, without motion blur, they introduce noise that potentially affects reconstruction. Figure shows accuracies and variances of the disparity maps within the user-indicated regions, obtained using three different stereo reconstruction algorithms (rows), as a function of exposure time and sensor ISO. The errors are large because the mean is swayed by some outlier pixels in the scribbled regions that might have a disparity value far from the mean. The indications are overlaid (semi-transparent white) on the input left-images.**



Ground truth disparity (selection) = 58.0



Ground truth disparity (selection) = 49.33

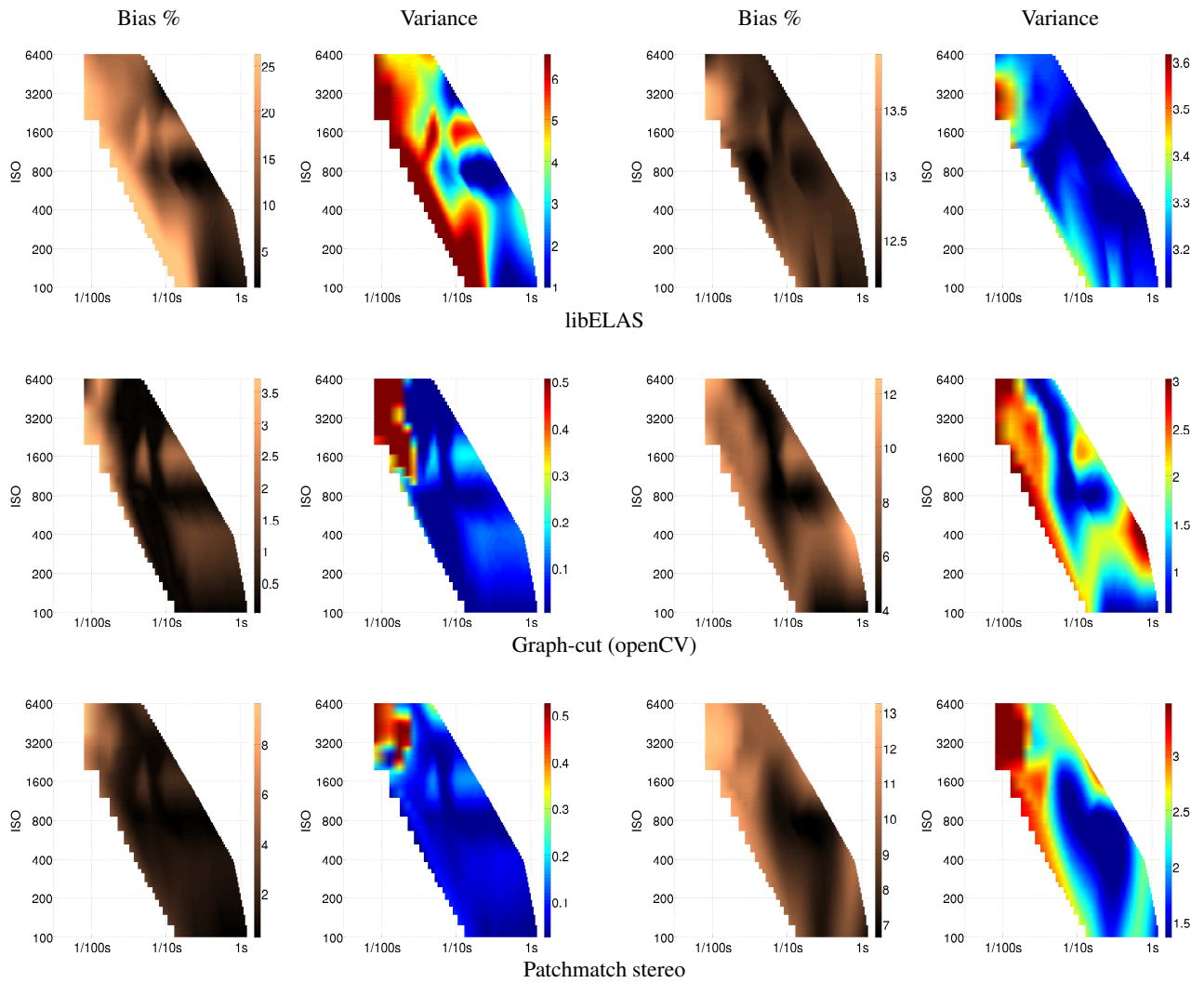
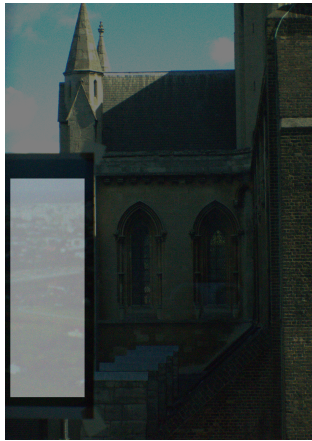
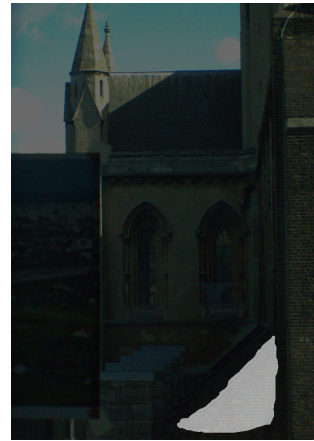


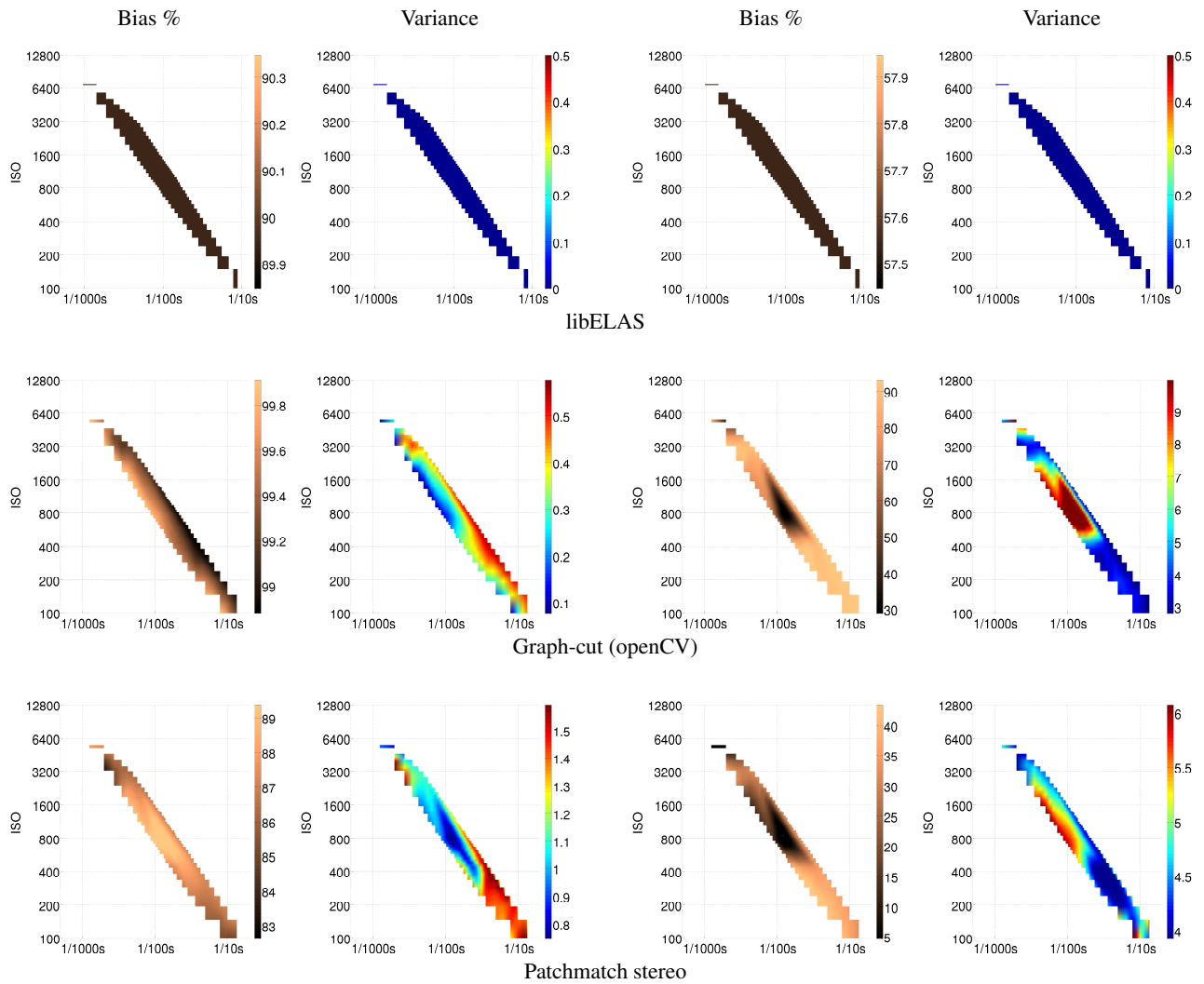
Figure 5: A comparison of stereo reconstruction algorithms over the space of ISO and exposure time with a static camera. Figure shows accuracies and variances of disparity maps, obtained using three different stereo reconstruction algorithms (rows), as a function of sensor ISO and exposure time. The scribbles are shown (semi-transparent white) overlaid on the input left-images. Not surprisingly, in most settings, long exposures are preferable to high-ISO.



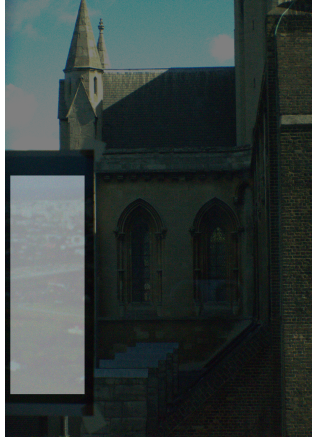
Ground truth disparity (selection) = 98.5



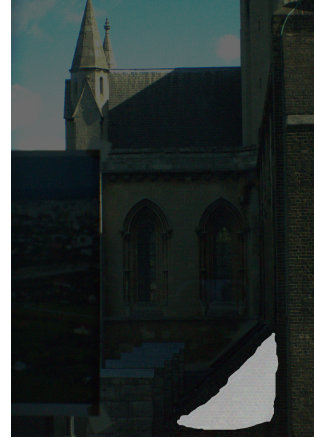
Ground truth disparity (selection) = 23.5



**Figure 6: A comparison of stereo reconstruction algorithms over the space of ISO and exposure time for an outdoor scene, with a background that is in focus. Figure shows accuracies and variances of disparity maps, obtained using two different stereo reconstruction algorithms (rows), as a function of exposure time and sensor ISO. The scribbles are shown (semi-transparent white) overlaid on the input left-images. For the static, out-of-focus region, the reconstructions are better with long exposures than with high-ISO.**



Ground truth disparity (selection) = 98.5



Ground truth disparity (selection) = 23.5

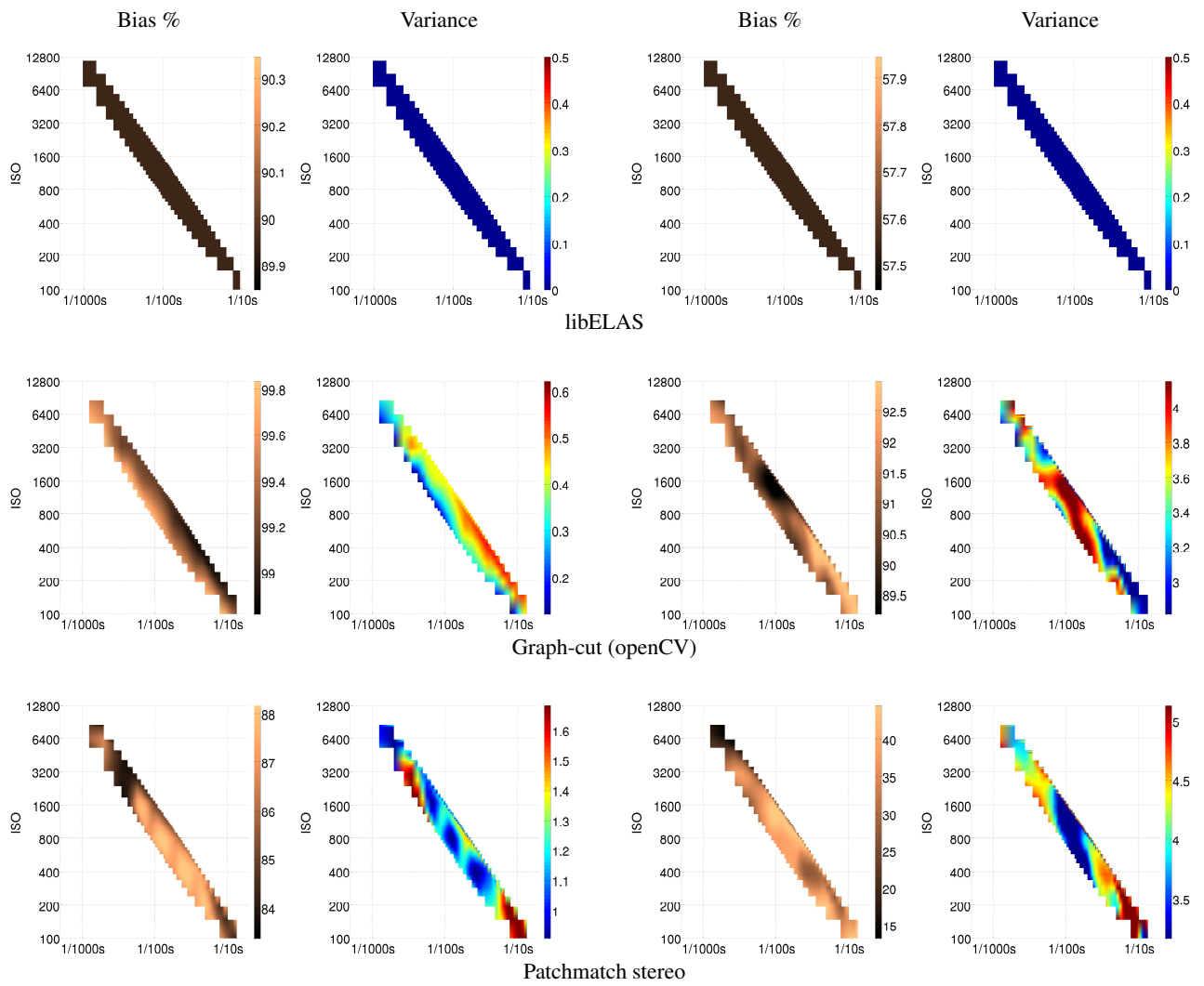


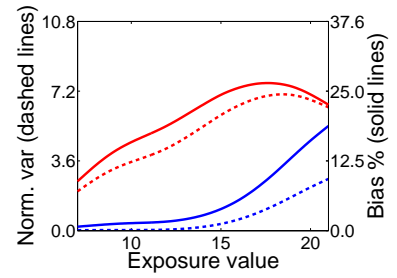
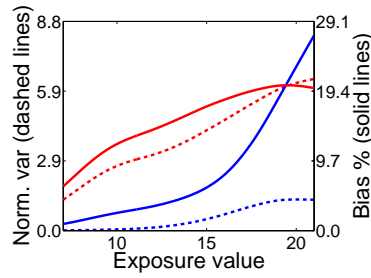
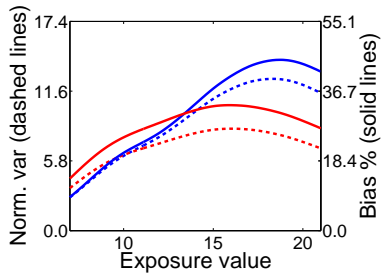
Figure 7: Results of the same experiment as shown in Fig. 6, but this time with the foreground in focus.



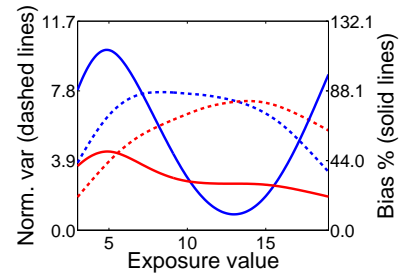
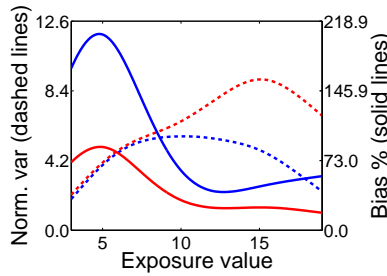
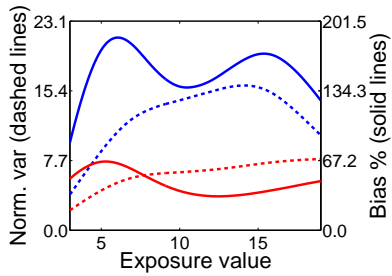
libELAS

OpenCV graph-cut

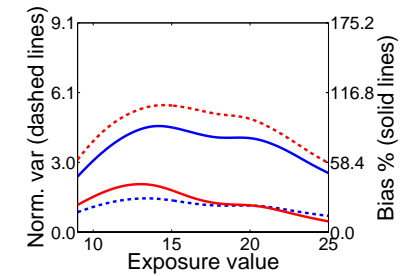
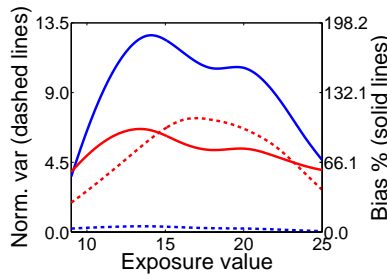
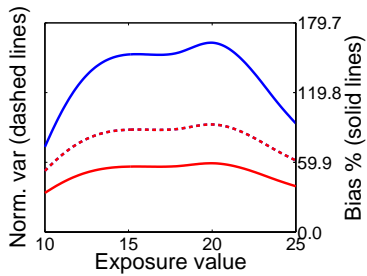
Patchmatch stereo



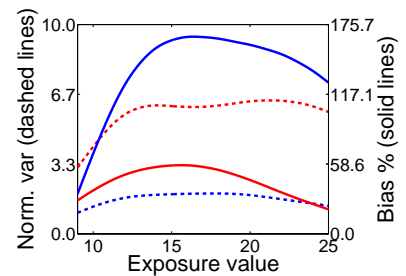
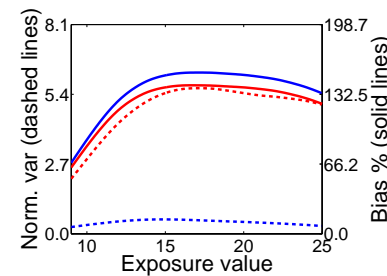
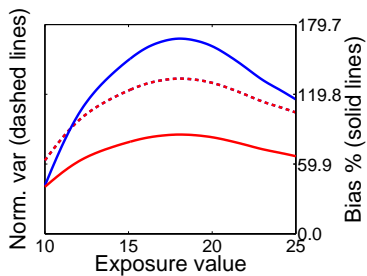
Static: Cardboard box region (blue) and background pattern region (red)



With motion: Cardboard box region (blue) and background pattern region (red)

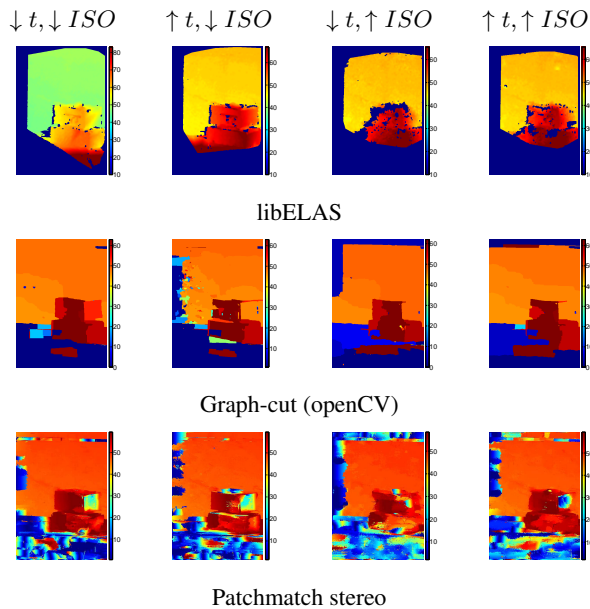


With shallow depth of field: Out-of-focus region (blue) and in-focus region (red)



With shallow depth of field: in-focus region (blue) and out-of-focus region (red)

Figure 8: Variance and bias in the reconstructed disparities for three scenes (rows) using each of the three algorithms (columns). Each graph plots the variance (dashed lines, left y-scale) and bias % (solid lines, right y-scale) against exposure values (computed using Eq. 1). Different colors (blue and red) correspond to the two scribbled regions in each image. Splines were fit to the points to depict trends. The blue curves in columns 1, 2 and 3 correspond to the scribbles on the left in figures 4, 5, 6 and 7 respectively. The absence of axis labeling (variance of libELAS last two rows) suggests very low values, in the order of  $1e-6$ .



**Figure 9: Disparity maps for a scene with surfaces that are not fronto-parallel. libELAS is fast, but again less accurate in this setting. Patchmatch stereo is better than openCV’s graph cut optimisation at capturing the gradual changes in depth, but is more sensitive to high-ISO.**

[6] A. Fusiello and L. Irsara. Quasi-Euclidean Uncalibrated Epipolar Rectification. In *International Conference on Pattern Recognition*, 2008.

[7] A. Geiger, M. Roser, and R. Urtasun. Efficient Large-Scale Stereo Matching. In *Proceedings of the 10th Asian Conference on Computer Vision*, pages 25–38, 2010.

[8] S. W. Hasinoff and W. T. Freeman. Noise-Optimal Capture for High Dynamic Range Photography. *Proc. 23rd IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010*, pages 553–560, 2010.

[9] S. W. Hasinoff and K. N. Kutulakos. Confocal Stereo. *International Journal of Computer Vision*, 81(1):82–104, Sept. 2008.

[10] S. W. Hasinoff and K. N. Kutulakos. Light-efficient photography. *IEEE transactions on pattern analysis and machine intelligence*, 33(11):2203–14, Nov. 2011.

[11] S. W. Hasinoff, K. N. Kutulakos, and W. T. Freeman. Time-Constrained Photography. *Proc. 12th IEEE International Conference on Computer Vision, ICCV 2009*, pages 333–340, 2009.

[12] Y. S. Heo, K. M. Lee, and S. U. Lee. Simultaneous Depth Reconstruction and Restoration of Noisy Stereo Images using Non-local Pixel Distribution. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[13] Y. S. Heo, K. M. Lee, and S. U. Lee. Robust stereo matching using adaptive normalized cross-correlation. *IEEE transactions on pattern analysis and machine intelligence*, 33(4):807–22, Apr. 2011.

[14] H. Hirschm and D. Scharstein. Evaluation of Cost Functions for Stereo Matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[15] A. Levin and W. T. Freeman. Image and Depth from a Conventional Camera with a Coded Aperture. *Image (Rochester, N.Y.)*, 26(3), 2007.

[16] T. F. V. Ltd. Nuke/ocula. <http://www.thefoundry.co.uk/products/ocula/>.

[17] B. Mendiburu. *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Focal Press, 2009.

[18] S. F. Ray. Camera exposure determination. *The Manual of Photography*, 9:318, 2000.

[19] S. Roy. Stereo without epipolar lines: A maximum-flow formulation. *Int. J. Comput. Vision*, 34(2-3):147–161, Oct. 1999.

[20] D. Scharstein. A Taxonomy and Evaluation of Dense Two-Frame Stereo. *International Journal of Computer Vision*, 47(1):7–42, 2002.

[21] D. Scharstein and C. Pal. Learning Conditional Random Fields for Stereo. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[22] D. Scharstein and R. Szeliski. High-Accuracy Stereo Depth Maps Using Structured Light. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 195–202, 2003.

[23] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR’06)*, pages 519–528, 2006.

[24] J. Xu, Li, Jia. Depth-Aware Motion Deblurring. In *IEEE Conference on Computational Photography (ICCP)*, pages 1–8, 2012.