

Performance of Stereo Methods in Cluttered Scenes

Fahim Mannan* and Michael S. Langer†

School of Computer Science

McGill University

Montreal, Quebec H3A 2A7, Canada

*{*fmannan, †langer}@cim.mcgill.ca*

Abstract—This paper evaluates the performance of different stereo formulations in the context of cluttered scenes [1] with large number of binocular-monocular boundaries (i.e. occlusion boundaries). Three stereo methods employing three different constraints are considered. These are basic [2], [3](Basic), uniqueness [4] (KZ-uni), and visibility [5](KZ-vis). Scenes for the experiments are synthetically generated and some are shown to have significantly more occlusion boundaries than the Middlebury scenes. This allows evaluating the methods with different types of scenes to understand the efficacy of different constraints for cluttered scenes. The evaluation considers mislabeled pixels of different types (binocular/monocular) in different regions (on or away from occlusion boundary). We have found that for sparse scenes (fewer occlusion boundaries) all three methods have similar performance. For dense scenes the performance is dominated by pixels on the boundary. For binocular pixels Basic always does better but for monocular pixels KZ-vis has the lowest error. If binary occlusion labeling is considered then the cross-checked version of basic constraint Basic-cc performs best followed by KZ-uni.

Keywords—Stereo; Performance Analysis; Visibility; Occlusion; Occlusion Boundary; Markov Random Field

I. INTRODUCTION

An important step in stereo reconstruction is finding corresponding pixels from a pair of images. Over the years a wide range of techniques for stereo correspondence problem has been proposed. Many of the top performing methods use Markov Random Field formulation of the problem and try to find the Maximum-a-Posteriori estimate (MAP-MRF) using some global optimization algorithm. Different formulations can be characterized by the type of constraints that are enforced on the solution. The primary set of constraints that are used in many vision problems are color consistency and smoothness constraint. In the context of stereo there are also methods that consider visibility and uniqueness constraints. The main objective of this paper is to analyze how these three types of approach perform for monocular pixels (also known as semi-occluded pixels) and occlusion boundary pixels (or binocular-monocular discontinuity pixels) in the cluttered scene setting. It was previously reported that visibility reasoning improves the performance of stereo methods near discontinuity in [4]–[6]. In general discontinuity can mean both occlusion boundary (i.e. binocular-monocular boundary) and depth discontinuity. More specifically we consider pixels on the occlusion boundary and away from

the boundary. It should be noted that pixels on the occlusion boundary can either be on the same surface or on a different surface. We look at where this performance improvement occurs and by how much. We also investigate if this improvement has any dependence on the type of cluttered scene.

We define cluttered scenes to consist of objects of some size distribution placed uniform randomly in a 3D space. Some common examples of such scenes in nature are tree canopies, bushes, etc. Stereo reconstruction for such cluttered scenes can potentially be used in forest measurement, where the problem is to determine the growth of a forest by making measurements such as leaf area index (LAI), canopy size, etc. [7]–[9]. Existing methods rely on either making measurements using range finders or by estimating the amount of sunlight falling in an area using tedious measurement techniques. Cluttered scene stereo can significantly simplify the process of making such measurements.

The paper is organized as follows. Section II gives a general overview of the types of methods that have been proposed over the years as well as performance evaluation of those methods. The stereo formulation that is used in this paper is discussed in Section III. In Section IV, the cluttered scene model and scene generation technique is discussed. The experiments are presented in Sec. V and finally Sec. VI concludes the paper.

II. PREVIOUS WORK

Very broadly, stereo formulations can be categorized into local and global approaches. In the local approach some type of window based correlation method such as sum of squared difference or absolute difference is used. There are also variations that uses adaptive window size or weight to improve the result. In the global approach the problem is formulated as a global energy minimization problem. In this setting the problem is usually formulated using MRFs where each pixel or a group of pixels is considered as a random variable that can take on values from a set of disparity labels [10]. The problem then is to solve the MAP assignment problem. The solution can be obtained by minimizing an energy function that maps an assignment of labels to real numbers. Global approaches usually produce better results than the local approaches [3], [11]. The three methods that

are considered in this paper employ global optimization techniques. An overview of the three techniques is given in Sec. III.

In the following we look at some of the evaluation methods that have been used for comparing different stereo methods and discuss their strengths and weaknesses. The scope of our review is limited to global approaches.

One of the early works that looked at different stereo methods was [11]. The authors categorized and compared different local and global techniques for stereo reconstruction. Compared to their work we only consider MRF-based stereo formulations and their performance for a special class of scenes. Tappen et al. in [12] compares Graph Cuts and Belief Propagation using the same energy formulation. The authors found that both algorithms yield optimal solutions whose energy is lower than that of the true scene. This indicates that the energy formulation did not sufficiently model the problem. The authors restricted their evaluation to the basic formulation, but we also use additional constraints, but restrict the analysis to cluttered 3D scenes. More recently, in [3], Szeliski et al. studies the performance of different optimization algorithms for the basic constraint on a set of benchmark problems. While their objective was to find the algorithm with the lowest energy configuration and fastest convergence rate, our objective is to evaluate different consistency constraints and observe their performance in specific regions of a scene and for different types of scenes.

In [13] we looked at the performance for basic formulation. The objective was to find the optimal parameter settings for different types of cluttered scene. We restricted the analysis to binocular pixels. In this paper, both monocular and binocular pixels are considered. We also look at how the error depends on the amount of occlusion and occlusion boundary.

III. STEREO FORMULATION

In this section we give an overview of the methods that we used in our evaluation. These methods model the stereo correspondence problem in the MAP-MRF framework and find a solution using global optimization techniques. In all cases, we use the α -Expansion algorithm [2] for optimization. In the MRF-formulation the stereo problem is formulated as a labeling problem. The objective is to find a configuration or label assignment X such that the following energy is minimized.

$$E(X) = E_{data}(X) + E_{smooth}(X). \quad (1)$$

In other words, we want to solve $X^* = \arg \min_X E(X)$. In the above equation, E_{data} encodes the data consistency constraint and E_{smooth} encodes the smoothness constraint which enforces that neighboring pixels to have similar labels assigned. These two terms form the fundamental constraint that is used in many vision problems. We refer to this basic

formulation as **Basic**. In the context of stereo the above formulation can be expressed as follows.

For stereo pair images I_l and I_r with disparity values f_p and f_q for pixels p and q respectively, the energy function that we optimize is of the following form [3]:

$$E = \sum_p E_p(f_p) + \lambda \sum_{p,q} E_{p,q}(f_p, f_q) \quad (2)$$

$$E_p(f_p) = d(I_l, I_r, p, q)^{k_d} \quad (3)$$

$$E_{p,q}(f_p, f_q) = w_{pq} \min(|f_p - f_q|^{k_s}, V_{max}) \quad (4)$$

Equation 3 is the *data term* and 4 the *smoothness term*. The function d in the data term is the dissimilarity measure between corresponding pixels (i.e. pixels p and q in I_l and I_r respectively). It can be absolute difference ($k_d = 1$), squared difference ($k_d = 2$), or the more sophisticated Birchfield-Tomasi measure [14]. We use the latter, with $k_d = 2$. In the smoothness term, w_{pq} is a scalar weight defined for every neighboring pixel pair p and q , and depends on the color gradient. The smoothness term is a function of label difference. We use $k_s = 1$. The parameter λ specifies how much weight should be given to the smoothness term.

Besides these basic constraints, additional constraints such as uniqueness and visibility can also be considered. In [4] the authors propose a model that encodes this uniqueness constraint. This method which is referred to as **KZ-uni** in the rest of the paper, considers pairs of pixels or voxels as the nodes in the MRF. The objective is to label each pair with a binary label denoting correspondence or mismatch. In this way a pixel can only be matched with at most one other pixel. As a result the approach is enforcing a uniqueness constraint. If a pixel is not matched it is considered as an occluded pixel, i.e. not visible to the other camera. In this paper we refer to such occluded pixels as monocular pixels.

In [5] the authors proposed an alternative approach which we refer to as **KZ-vis**. Like **KZ-uni** (but unlike the basic approach), **KZ-vis** labels pixels in both left and right images. A pixel p in the left image can be assigned disparity label f_p if and only if a visibility constraint is satisfied, namely the scene point defined by (p, f_p) does not occlude the scene point (q, f_q) in the right image where $q = p + f_p$. Equivalently, the scene point (p, f_p) does not lie on the line segment joining the right camera to the scene point (q, f_q) . The algorithm starts by assigning a disparity value of 0 to every pixel and maintains the visibility constraint in each step of the algorithm.

IV. CLUTTERED SCENE MODEL AND SCENE GENERATION

The experiments that we will present use synthetic scenes with a range of parameters. These scenes provide us some control over the number of occlusion boundary pixels and monocular pixels, and this allows us to examine how well an algorithm performs as a function of these percentages.

Table II gives a summary of the percentages of monocular pixels and occlusion boundary pixels in our synthetic scenes versus typical Middlebury images.

Each of our cluttered scenes is defined by squares of a certain size and distributed uniformly over a 3D volume. Let the average number of objects per unit volume be ρ , the halfwidth of each square be r , and the z -range of the bounded volume be $[z_{min}z_{max}]$. Using these parameters one can formulate a model that represents the probability of disparity, visibility and other properties of cluttered scenes [1], though such a model is not used here.

Our scene generation process is similar to that of the dead-leaves model [15], with the exception that we choose the image size of each object to be inversely related to the depth of the object according to the laws of perspective. Object centers were chosen randomly in the 3D volume. In preliminary experiments, the objects were projected to two different viewpoints using perspective (central) projection. However, because of pixel binning this perspective projection method often produced inconsistent size and disparity values in the two images. To avoid these artifacts, we have instead generated the images using orthographic projection, as follows. The size of each object in the left image was inversely proportional to its depth. We chose the disparity label of the object based on its depth, and instantiated a corresponding object at the disparity-defined distance in the right image. The OpenGL depth buffer handled the visibility for the orthographic projection in each image.

For the ground truth labels, for each pixel on an object in the left image, we used its disparity to find the corresponding pixel in the right image and we checked its depth value. If the depth value was less than that of the original pixel in the left image, the left pixel image was marked as monocular.

A texture map for each square object was generated by choosing random RGB values on a 4×4 grid spanning the object and then interpolating. A background square was used at the far depth plane. Its texture was chosen using a 16×16 grid of random RGB values. Independent noise was then added to the RGB pixel values.

The parameters we use for our scenes are shown in Table I. The scenes are defined by two levels of each of three categories: density, depth range, and object size. The (a) and (b) are the large radius and small radius, respectively. Note that these object radii r interact with the z range, such that we are effectively varying the baseline of the camera, rather than the images sizes of the objects. For example, the denser and larger object near scenes (1a) produce single images that have the same properties as the denser, larger object far scenes (2a). A few examples of the scenes are shown in Figure 1.

The crucial point is that the eight classes of cluttered scenes all have different visibility and occlusion characteristics. To give an idea of the differences, and how they compare to scenes in the Middlebury dataset, we list the

No.	baseline	density ρ	range z	radius r
1a	0.2	0.54	2-8	0.1
1b				0.025
2a			8-32	0.4
2b				0.1
3a		0.1	2-8	0.1
3b				0.025
4a			8-32	0.4
4b				0.1

Table I
CLUTTERED SCENE PARAMETERS

scene	% monocular	% occ. boundary
map	10%	5%
sawtooth	7%	3%
cones	15%	2%
teddy	13%	1%
rocks2	17%	4%
flowerpots	28%	5%
1a	30%	4%
1b	48%	12%
2a	8%	2%
2b	15%	8%
3a	32%	4%
3b	39%	15%
4a	12%	4%
4b	26%	13%

Table II
PERCENTAGE OF MONOCULAR PIXELS AND OCCLUSION BOUNDARY (I.E. BINOCULAR-MONOCULAR BOUNDARY). STATISTICS COMPUTED FROM 100 IMAGES OF EACH SCENE.

percentage of monocular pixels occlusion boundary edges in Table II. For the case of *cones* and *teddy* we used the provided occlusion map. For the rest of the images from the Middlebury dataset we computed the occlusion map by doing cross-checking. It should be noted that slanted surfaces produce false positives for the occlusion detection and so the true percentage of occlusions for the Middlebury dataset should be less. Note that many of our scenes contain a greater number of monocular pixels and most of our scenes contain a much greater number of occlusion discontinuities than the Middlebury scenes. Furthermore the eight different classes of scenes have variations in these percentages. This allows us to examine how different stereo methods perform under different types of clutter.

V. EXPERIMENTS

This section starts by giving an overview of the types of errors that are considered and the motivation behind them. Then we present the results and summarize the observations. Using the parameters specified in Table II, 10 instances of each scene were generated. The software provided in [6] was used for running the experiments on those scenes.

1. Types of pixel (Fig. 3): Binocular and monocular labeling errors are considered separately in the experiments. In each case it is the percentage of mislabeled pixels (i.e. non-zero difference with the groundtruth) out of all the pixels

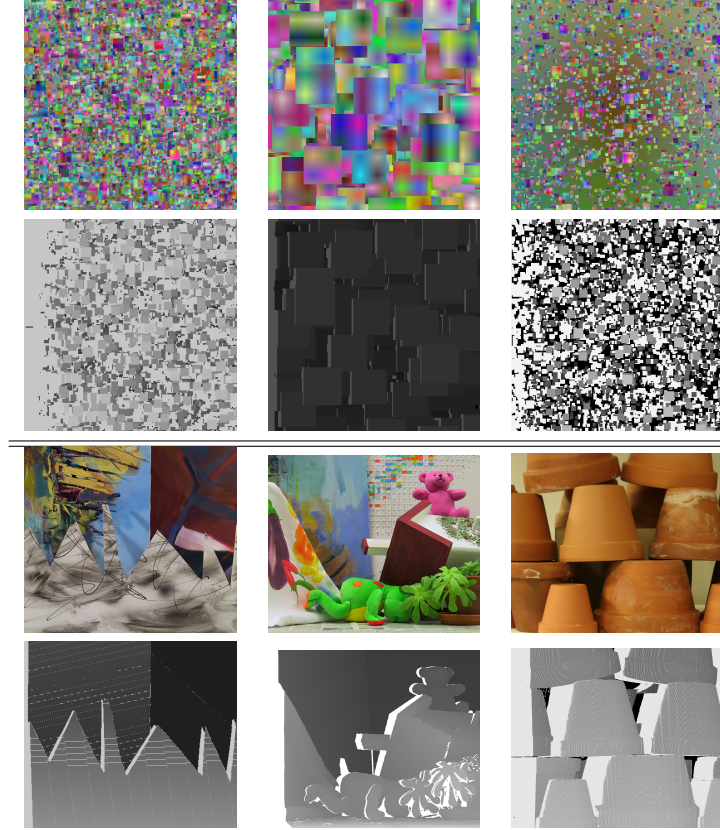


Figure 1. Sample scenes with disparity and occlusion (brightest) from our test dataset (top) and Middlebury (bottom). Top part (left–right) showing scenes 1b, 2a, and 3b. Bottom part showing (left–right) sawtooth, teddy and flowerpots.

of a given type. Binocular errors are expected to be less than the monocular errors. Furthermore, the methods do not explicitly handle monocular pixel labeling. For scenes with fewer monocular pixels these errors do not have a big impact on the overall error. However these errors become significant for cluttered scenes with large number of monocular pixels (e.g. scenes 1a-b, 3a-b Table II). In the experiments two types of monocular labels are considered. The first is a disparity label and the second is a binary monocular (i.e. an invalid disparity value) and non-monocular (i.e. a valid disparity value) labeling. **Basic** and **KZ-vis** do not make any distinction between monocular and binocular pixels and as a result both label monocular pixels with a disparity value. **KZ-uni** however has occlusion label which is assigned to any pixel that does not have a corresponding pixel. To make the comparison consistent, the output of **Basic** and **KZ-vis** are assigned monocular labels in an additional pass. This is done by performing cross-checking with the left and right disparity maps. If a pixel and its corresponding pixel do not have the same disparity value then that pixel is marked as monocular. These results are referred to as **Basic-cc** and **KZ-vis-cc**. To summarize, two types of comparisons are made for monocular pixels. A comparison between **Basic** and **KZ-**

vis (i.e. for disparity labels), and comparison between **Basic-cc**, **KZ-vis-cc** and **KZ-uni** (i.e. binary monocular label). Fig. 3 shows this error for a subset of the scenes.

2. Region (Fig. 4): This categorization is based on whether a pixel is on the occlusion boundary or away from the boundary. The accuracy of the border pixels is more sensitive to the choice of prior and the additional constraints (i.e. visibility and uniqueness) that are used to enforce correct labeling of these pixels. Naturally for scenes with large number of occlusion boundaries this type of error becomes significant. These regions are further subdivided based on binocular and monocular pixels and analyzed as before. The plots are shown in Fig. 4.

3. Joint ground-truth and output (Fig. 2): This is the joint histogram of disparity assignment and ground-truth. Here too, the pixels are categorized into binocular and monocular. The objective of such plots is to highlight the type of errors occurring at each disparity. For the plots, the λ parameter is fixed to the value giving the lowest error. It should be noted that ideally the monocular joint plot for **KZ-uni** should be empty since the algorithm labels those pixels as occluded (and hence no disparity label). Therefore the monocular joint plots for **KZ-uni** basically shows the types of mislabeled

Sparse	Binocular	Monocular
B+NB	Basic \approx KZ-vis \approx KZ-uni	disp: KZ-vis < Basic occ: Basic-cc \leq KZ-uni \leq KZ-vis-cc
NB	same as above	same as above
B	Basic \approx KZ-vis < KZ-uni	disp: same as above occ: Basic-cc < KZ-uni < KZ-vis-cc

(a) Sparse Binocular-Monocular Boundary

Dense	Binocular	Monocular
B+NB	Basic < KZ-uni < KZ-vis	disp: KZ-vis < Basic occ: Basic-cc < KZ-uni < KZ-vis-cc
NB	Basic \approx KZ-vis \approx KZ-uni	same as above
B	Basic < KZ-uni < KZ-vis	same as above

(b) Dense Binocular-Monocular Boundary

Table III

SUMMARY OF RESULTS FOR DIFFERENT SCENES (I.E. A) SPARSE AND B) DENSE), TYPES OF PIXELS (BINOCULAR OR MONOCULAR) AND REGIONS (B: OCCLUSION BOUNDARY, NB: NOT BOUNDARY AND B+NB: ALL REGION). THE COMPARISON IS BASED ON THE LOWEST ERROR I.E. $a < b$ IMPLIES METHOD a 'S LOWEST ERROR IS LESS THAN b 'S LOWEST ERROR.

pixels. Fig. 2 shows the joint error for both binocular and monocular pixels for scenes 1b, 3b and 2b.

Table III(a) and III(b) summarizes the results. The analysis is primarily focused on how the three methods perform relative to each other for different types of pixels (i.e. binocular or monocular) in different regions (i.e. near occlusion or not near). Scenes are categorized into sparse and dense scenes based on the amount of occlusion boundary (i.e. last column in Table II) present in them. For each type of scene the binocular, monocular, boundary and joint errors are considered.

Sparse Scenes: For binocular pixels in sparse scenes all three methods have similar performance (i.e. lowest error is the same). For binocular pixels on the border the ordering between the methods is different but since the scenes are sparse the border pixels make a very small contribution to the overall error. For sparse scenes, if monocular disparity is considered then **KZ-vis** always perform better than **Basic**. For binary occlusion labeling, the cross-checking based methods and **KZ-uni** have almost similar error.

Dense Scenes: For binocular pixels in dense scenes, the order of performance is usually **Basic**, **KZ-uni**, and **KZ-vis**. This is also the order for binocular pixels on the boundary. However all three methods have equal performance for pixels that are away from the boundary. Since the amount of occlusion boundary is large, the mislabeled pixels on the boundary have more impact on the overall error. Note that scene 3b is an exception. In this case **KZ-vis** does significantly better than the other methods. This scene has a lot of visible background and from the second row of Figure 2 it can be seen that **KZ-vis** (middle column) is doing better than the other two at labeling those pixels

in the background (i.e. the (0,0) bin has more weight for **KZ-vis**). The joint histogram also shows that for 1b (top row), **KZ-vis** mostly mislabels pixels in the mid disparity range. This is mostly because most of the binocular pixels on the occlusion boundary are at this disparity. For this scene the other methods (left and right column) have most pixels concentrated on the diagonal. Finally for monocular pixels with disparity labels, **KZ-vis** does better than **Basic** independent of pixel region. From the joint histogram of monocular pixels on the third row of Figure 2, it can be seen that **KZ-vis** has a strong diagonal. On the other hand, the monocular labels for **Basic** is almost uniformly distributed over all the bins. For occlusion labeling the ordering is always **Basic-cc**, **KZ-uni**, and **KZ-vis-cc**. The joint plot of 2b shows the mislabeled monocular pixels.

VI. CONCLUSION

In this paper, we investigated how different constraints in the MAP-MRF formulation for stereo works for cluttered scenes. We have shown that our definition of cluttered scenes can incorporate different amounts of monocular pixels and discontinuity (both depth discontinuity in general and occlusion edges). We generated cluttered scenes with different parameter settings so as to obtain scenes with different properties. Furthermore we have the advantage of obtaining a large number of scenes with the same property. This allowed us to test the different methods under different condition and statistically estimate the error rate.

We observed that when all pixels are considered (i.e. both binocular and monocular error) then **KZ-vis** outperforms the other two methods. This is not because **KZ-vis** does better for binocular pixels near occlusion boundaries but mainly because it does significantly better for monocular pixels near occlusion boundary. In fact **KZ-vis** can often (except in 3b) perform poorly for binocular pixels for dense scenes. Since **KZ-vis** outperforms all other methods for monocular pixels, it should perform better for general cluttered scenes. For binary occlusion labeling the basic constraint with cross-checking i.e. **Basic-cc** has the best performance. But it is a two pass process and **KZ-uni** can achieve comparable result in a single pass. For binocular pixels **KZ-uni** performs somewhere in between the other two methods.

To summarize, we generated a range of scenes with different properties and showed how different methods perform for these scenes. In this work we restricted our attention to methods that enforce pixel-wise constraints. However, most of the current top performing methods use region-based constraints. In future work we will investigate how those methods perform for cluttered scenes.

ACKNOWLEDGMENT

The authors would like to thank Vladimir Kolmogorov for making his code available.

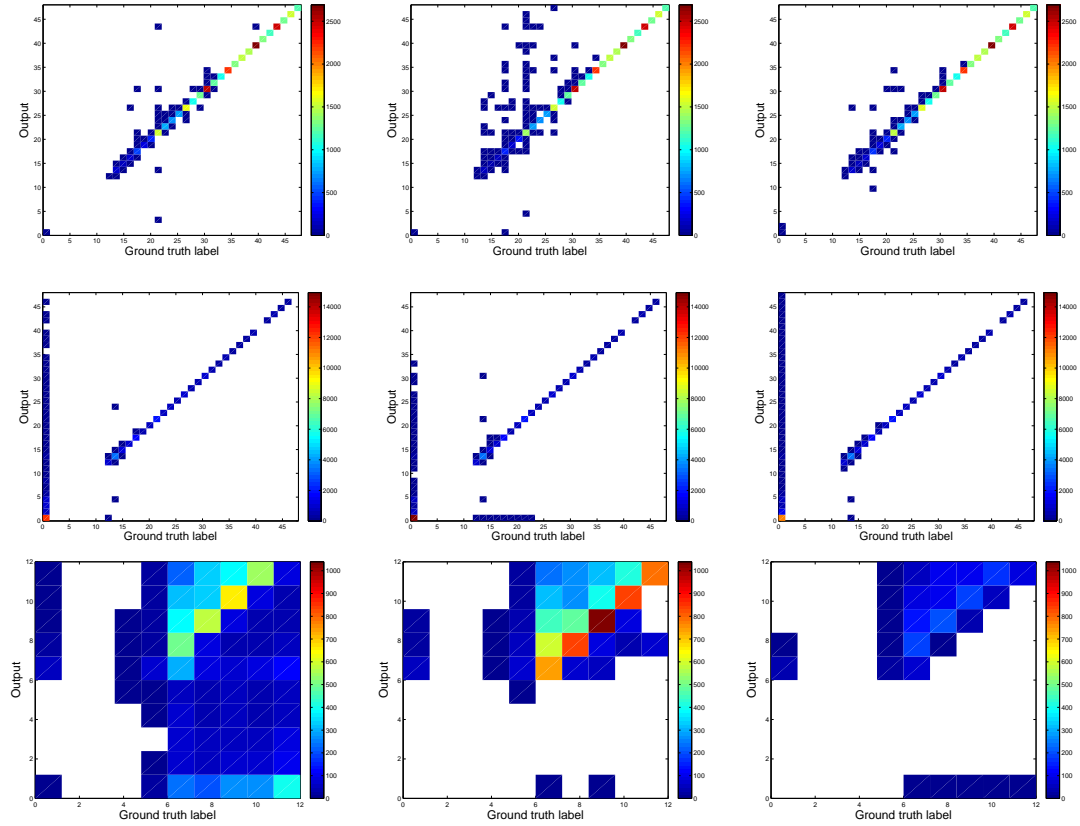


Figure 2. Joint disparity error (top-bottom) for binocular pixels of 1b, 3b, and monocular pixels of 2b. The columns from left to right represent **Basic**, **KZ-vis** and **KZ-uni**.

REFERENCES

- [1] M. S. Langer, "Surface visibility probabilities in 3d cluttered scenes," in *European Conference on Computer Vision*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 401–412.
- [2] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, Nov 2001.
- [3] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1068–1080, June 2008.
- [4] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions via graph cuts," in *International Conference on Computer Vision*, 2001, pp. 508–515.
- [5] —, "Multi-camera scene reconstruction via graph cuts," in *European Conference on Computer Vision (3)*, ser. Lecture Notes in Computer Science, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds., vol. 2352. Springer, 2002, pp. 82–96.
- [6] —, "What energy functions can be minimized via graph cuts?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, Feb. 2004.
- [7] J. Chen and J. Cihlar, "Plant canopy gap-size analysis theory for improving optical measurements of leaf-area index," *Applied Optics*, vol. 34, no. 27, pp. 6211–6222, 1995.
- [8] P. R. van Gardingen, G. E. Jackson, S. Hernandez-Daumas, G. Russell, and L. Sharp, "Leaf area index estimates obtained for clumped canopies using hemispherical photography," *Agricultural and Forest Meteorology*, vol. 94, no. 3-4, pp. 243 – 257, 1999.
- [9] D. Riao, F. Valladares, S. Conds, and E. Chuvieco, "Estimation of leaf area index and covered ground from airborne laser scanner (lidar) in two contrasting forests," *Agricultural and Forest Meteorology*, vol. 124, no. 3-4, pp. 269 – 275, 2004.
- [10] S. Li, *Markov Random Field Modeling in Image Analysis*. Springer Publishing Company, Incorporated, 2009.
- [11] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

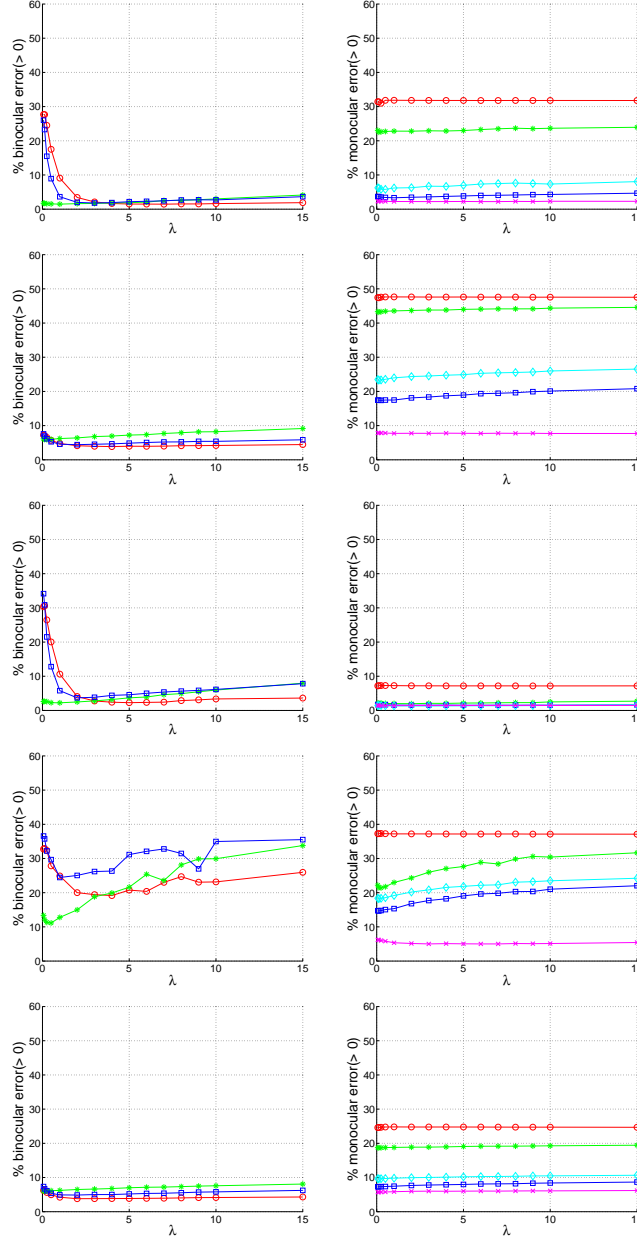


Figure 3. Error Statistics for scenes 1a-b, 2a, 3b and 4b(top-bottom). The left column represents the error for binocular and the right columns for monocular pixels. **Basic**, **KZ-vis**, **KZ-Uni**, **KZ-vis-cc** and **Basic-cc** are represented using \circ , \diamond , \square , \diamond , and \times . **KZ-vis-cc** and **Basic-cc** are not shown for binocular pixels.

- [12] M. F. Tappen and W. T. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters," in *IEEE International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 2003, p. 900.
- [13] F. Mannan and M. Langer, "Performance of mrf-based stereo algorithms for cluttered scenes," in *Brain, Body and Machine*, ser. Advances in Soft Computing, J. Angeles, B. Boulet, J. Clark, J. Kvecses, and K. Siddiqi, Eds. Springer Berlin / Heidelberg, 2010, vol. 83, pp. 125–136.
- [14] S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 401–406, 1998.
- [15] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imaging Vis.*, vol. 18, no. 1, pp. 17–33, 2003.

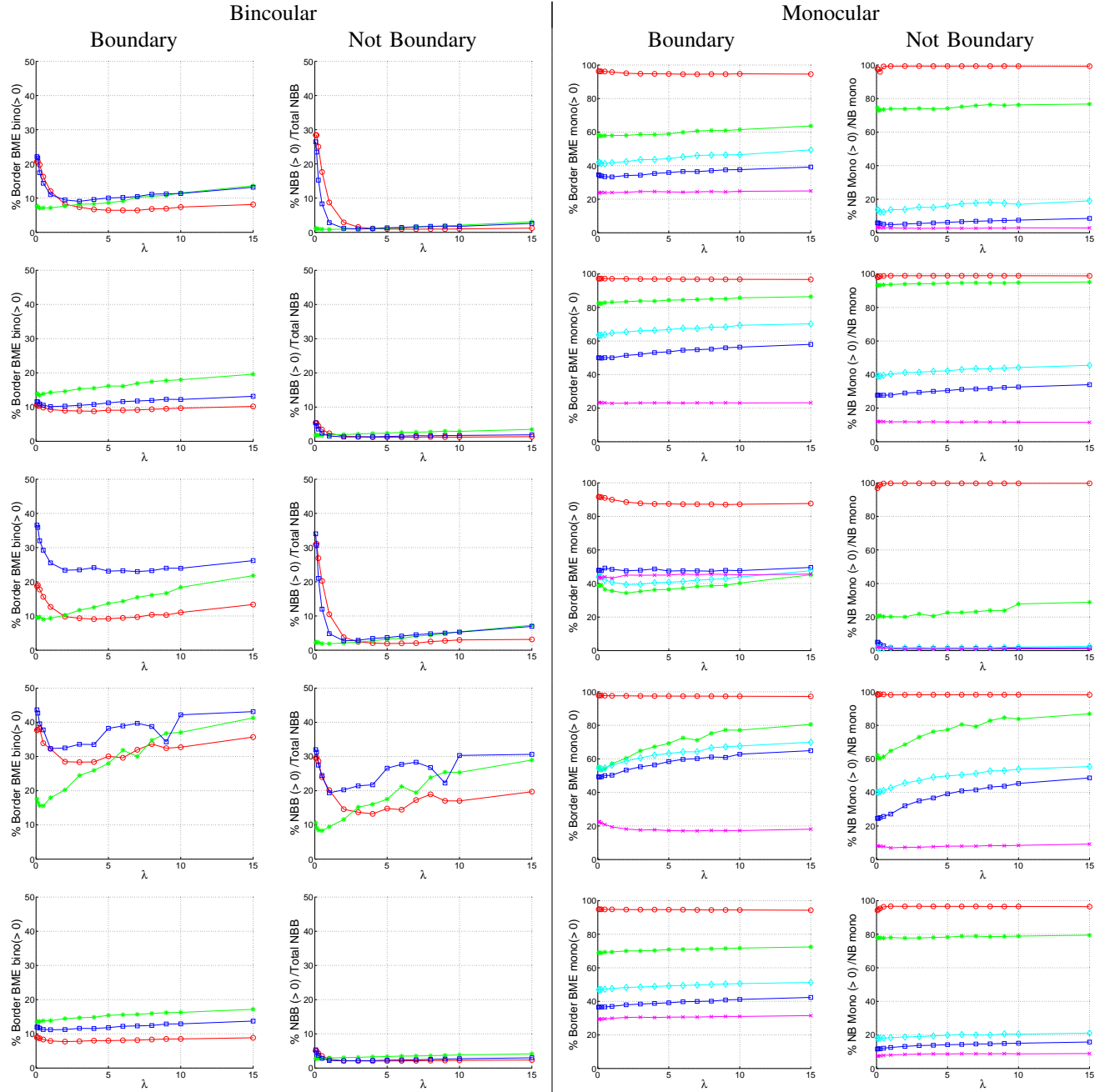


Figure 4. Error occurring for binocular (left) and monocular (right) pixels on the boundary for scenes 1a-b, 2a, 3b and 4b(top-bottom). As before **Basic**, **KZ-vis**, **KZ-Uni**, **KZ-vis-cc** and **Basic-cc** are represented using \circ , \star , \square , \diamond and \times . The binocular plots for **KZ-vis-cc** and **Basic-cc** are not shown.