An Analysis of Depth Estimation within Interaction Range

Cem Karaoguz, Andrew Dankers, Tobias Rodemann, Mark Dunn

2010

Preprint:

This is an accepted article published in IEEE-RSJ International Conference on Intelligent Robot and Systems (IROS 2010). The final authenticated version is available online at: https://doi.org/[DOI not available]

Accepted for IROS 2010, October 18 – 22, 2010 in Taipei, Taiwan

An Analysis of Depth Estimation within Interaction Range

Cem Karaoguz[1,2], Andrew Dankers[2], Tobias Rodemann[2] and Mark Dunn[2]

***Abstract*— Interactions between humans or humanoids and their environment through tasks like grasping or manipulation typically require accurate depth information. The human vision system integrates various monocular and binocular depth estimation mechanisms in order to achieve robust and reliable depth perception. Such an integrated approach can be applied to humanoid depth perception. Integration requires a knowledge of the characteristics of the methods being combined. Three different methods incorporating active vision (stereo disparity, vergence and familiar size) were statistically examined and combinations of these methods based on this statistical examination were investigated. We found evidence that active vision provides better depth estimations than the standard static-parallel stereo methods examined within interaction range and therefore is better suited for tasks like reaching, grasping and manipulation. We also demonstrate that a combination of methods have the potential to increase the accuracy of estimations.**

I. INTRODUCTION

Complex tasks like reaching, grasping or driving involve depth perception and vision is a sufficient and reliable depth information source for such kinds of tasks. However, depth estimation is an ill-posed problem since the optical process of visual information gathering projects the 3D information onto a 2D retina [14]. Various mechanisms contribute to depth estimation for the human visual system, each with their own characteristics (accuracy, constraints, neural substrate, etc.) and one may perform dominantly for certain tasks. For humans, combining these different mechanisms provides robust and reliable depth perception. As in the human vision system, depth estimation methods can be combined to achieve more reliable depth estimations for humanoids. However, the characteristics of these methods must be understood in order to determine the optimal combination.

We focus on depth estimation in the near visual field to support tasks like grasping for humanoids. Such tasks benefit from accurate estimations of depth in the near workspace. In this instance, the use of active vision may yield benefit over traditional static stereo vision. A visual system able to adjust its visual parameters to aid task oriented behavior is an approach called active vision [1]. As we shall see, using active vision produces accurate depth information at close ranges. Additionally an active vision approach better copes with dynamic scenes (e.g. eliminating motion blur) and allows sub-pixel analysis via small camera movements.

1 Research Institute for Cognition and Robotics (CoR-Lab), Bielefeld University, 33594 Bielefeld, Germany ckaraogu@cor-lab.uni-bielefeld.de

2 Honda Research Institute Europe GmbH, Carl-Legien-Str. 30, 63073 Offenbach, Germany {andrew.dankers, tobias.rodemann, mark.dunn}@honda-ri.de

A limited comparison of vergence and photogrammetry methods is done in [10]. In [5] and [15] reviews on different stereo disparity computation methods were presented. However these were restricted to a static-parallel stereo camera setup. We examined three depth estimation methods (stereo disparity, vergence and familiar size) with an active vision setup in an extensive test setting. These methods are common and well studied in the vision community [2], [5], [10], [15], [16]. The goal of this work is to determine how accurately depth can be estimated, which method is more accurate in what depth region, what the main error sources are and how the estimations be improved by combining the methods. The evaluation was conducted based on raw sensory data. The results may be improved by applying post-processing methods (e.g. [3]). Based on the statistical data obtained from the experiments, combinations of different depth estimation methods using maximum-likelihood estimation is investigated. A comparison between the active and static-parallel stereo vision approaches for depth estimation was also conducted.

The depth estimation methods are explained in Sec. II. The hardware system and the experimental procedure is described in Sec. III. The results are shown in Sec. IV. The findings and discussions are presented in Sec. V. Finally, the summary and conclusion are presented in Sec. VI .

II. DEPTH ESTIMATION METHODS

A. Vergence

Stereo fixation on an object is achieved when both eyes are positioned such that optical axes intersect on the surface of the object, allowing the projection of the object to fall on the foveae of both retinae. This type of eye movement is called vergence and is an important source of information about depth in the human visual system [14]. For an artificial system, depth estimation by vergence triangulation using a pinhole camera model is shown in Fig. 1. The distance to the stereo fixation point can be derived from the vergence angle as:

$$z = \frac{b}{2 \cdot tan(\frac{\Theta_v}{2})}, \tag{1}$$

where Θ_v is the vergence angle and b is the baseline (Fig. 1). The vergence angle is computed from left and right camera angles as $\Theta_v = \Theta_{left} + \Theta_{right}$. In our application symmetric vergence ($|\Theta_{left}| = |\Theta_{right}| = \Theta$) is applied. We used a correlation based vergence control algorithm [2] to achieve fixation on the object surface. The vergence angle to depth correspondence was determined by measuring the vergence

angles used to stereo fixate an object for different points in depth. Fig. 2 plots depth vs vergence angle (left vertical axis). The curve is non-linear with steep slopes in the close range and flat slopes in the far range. From this, more precise estimations in the near ranges are expected with vergence.

B. Stereo Disparity (SD)

An object that is not at stereo fixation projects to different locations on the left and right retinae according to scene depth and the horizontal baseline separating the eyes [14]. The difference between these two locations is referred to as stereo disparity. Stereo disparity is a common depth cue used in artificial vision systems [5], [15]. In active vision estimations, depth computed from stereo disparity is relative to the fixation point. In order to obtain absolute depth information (i.e. the distance from the baseline to the stereo fixated object) an active rectification process [6] is used. As shown in Fig. 1 this process epipolar rectifies images from an active stereo camera configuration (cameras with solid lines) to virtual image planes of a parallel stereo camera configuration (cameras with red dotted lines). The computation of depth in this framework is done as:

$$z = \frac{bf}{d} + r + f, \tag{2}$$

where d is the disparity (defined as $d = x_{VL} - x_{VR}$, x_{VL} and x_{VR} being the projections of the object on the virtual left and right image planes), f is the focal length of the cameras, r is the distance from the center of rotation of the cameras to the image planes (Fig. 1). Disparity maps are computed using the block matching algorithm from OpenCV Version 2.0 [4]. For comparison, disparity maps were also computed in another experimental session with a static-parallel stereo camera setup. Two algorithms were used in this case: the block matching algorithm from OpenCV Version 2.0, which was used in the active vision case, and the SRI Small Vision System (SVS) Version 4.4d (2007) [13]. The SVS uses the same block matching algorithm as OpenCV, however disparities are refined via post processing (sub-pixel interpolation and post-filtering). The disparity search range was 32 pixels (-16 to +15) for the active vision case and 96 pixels for the static-parallel stereo case. All other parameters were the same. A simple color based segmentation process was used to distinguish the disparities corresponding to the object in the disparity maps and the average of these disparities was taken for depth estimation each frame. It should be noted that more advanced segmentation methods [7], [8] may be applied. Since depth estimation from stereo disparity varies with the vergence angle in the active vision case, estimations in the near range are expected to be more precise than in the far range, as in the vergence case.

C. Familiar Size (FS)

The depth of an object can be estimated from the size of its projection on the camera images if the real size of the object is known. Various approaches exist for this operation [16].

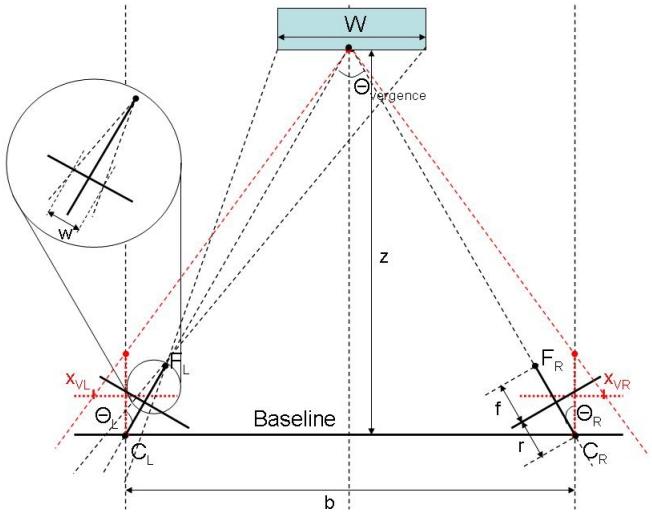


Fig. 1. Analytical model of the active vision system and depth estimation methods. Depth z is defined by the distance from the baseline to the object. F_L and F_R denote focal points, C_L and C_R denote center of rotations of the left and right cameras respectively. The static parameters of our system are as follows: r = 18.75 mm, f = 5.4 mm, b = 65 mm.

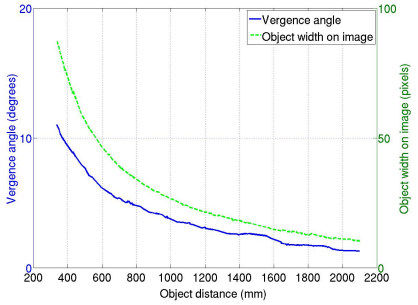


Fig. 2. Change of vergence angle with the object distance (left vertical axis, avg. over 10 objects). Change of object width on the image plane with the object distance (right vertical axis, avg. over 10 objects).

We used simple analytical relations derived from a pinhole camera model (Fig. 1). Object depth can be derived as:

$$z = \left(\frac{fW}{w} + r + f\right) cos(\Theta), \tag{3}$$

where Θ is the camera angle and $cos(\Theta) \approx 1$ due to small baseline. The physical size W was measured beforehand for all objects used in the experiments, the retinal size w was computed using a simple color based segmentation process (the same used with the stereo disparity method). The width of the objects was used for estimations since it showed better overall accuracy than the height. Fig. 2 shows depth versus object size on the image plane (right vertical axis) which is very similar to the vergence versus depth curve. Hence the estimations from familiar size are expected to exhibit similar characteristics to the estimations via vergence.

D. Combination of Methods

Bayesian cue integration [9] was used to combine the estimations from different methods. This is formulated as:

$$z = \sum_i w_i z_i. \tag{4}$$

where z is the maximum-likelihood estimate and z_i is the estimate from the i^{th} method. The weight w is expressed as:

$$w_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2}, \tag{5}$$

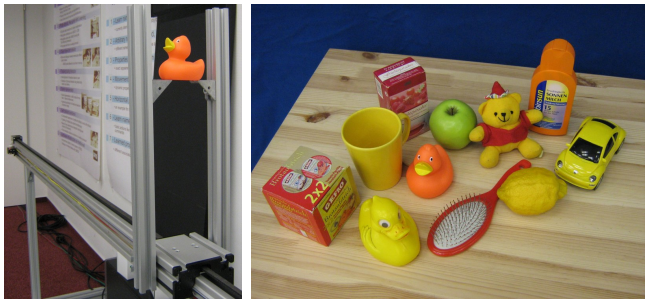
where σ_i^2 is the variance of the i^{th} method. Variance of each method was calculated using the statistics provided by the experiments. The combinations of the methods are computed on the experimental data using leave-one-out cross validation.

III. EXPERIMENTS

A. Hardware

An experimental stereo vision head with 4 DoF (2 DoF for head and 1 DoF for each camera) and a baseline of 65 mm was used as a platform for the examined depth estimation methods. All experiments were performed using images with a resolution of 400x300 pixels. This was the standard resolution for most of our vision applications [11].

A linear unit (Fig. 3(a)) that moves a small object platform on a linear axis is utilized to rapidly and autonomously generate data for depth estimation algorithms and acquire *ground-truth* depth information. The object platform is moved via a stepper motor within an error of 0.1 mm/100 mm.



(a) (b)

Fig. 3. (a) The linear unit with a rubber duck object on its carrier platform (upper right corner of the image). (b) The objects used in the experiments.

B. Experimental Setup

The linear unit is positioned in front of the stereo camera head and aligned to the center of the baseline (Fig. 4). The object platform is elevated to the height of the cameras. From the HRI150 database [12] 11 objects were selected that are suitable for the analysis (Fig. 3(b)). One of the objects was used for calibration purposes as explained in Section II-A. Care has been taken for the placement of the objects on the carrier. However there could still be misplacements of objects from the center of the platform, which could introduce a bias between the actual distance of the object and measured distance of the platform. Based on the average depth of the objects used in the experiments, this bias is expected to be in the range of ± 3 cm.

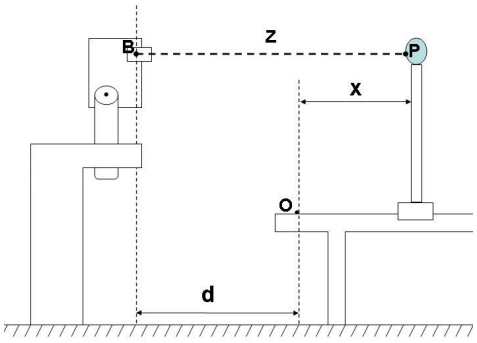


Fig. 4. Experimental setup. The depth (z) is defined as the distance from the mid-point between the center of rotation of the cameras (B) to the object (P). The object is placed on a moving platform on the linear unit. The zero position of the moving platform is denoted as O. Distance d is the horizontal gap between point B and point O. Depth of an object is calculated as $z = d + x$.

C. Experiment Routine

100 different positions of the linear unit in the range from 300 to 1500 mm were generated randomly beforehand. For each object the following routine was followed: the linear unit platform was moved to a perscribed position, relevant data was saved and the routine iterates. An experiment was concluded when all 100 positions had been completed. Repeating the experiment for all 10 objects provided 1000 data samples in total. The same experiments were conducted with a static-parallel stereo camera setup.

IV. RESULTS

The methods were expected to give more precise estimations in the near ranges than in the far. Results show this is indeed the case. The accuracies of the three estimation methods are shown in Fig. 5 as mean estimation errors[1] of the methods. The error is defined as the absolute value of the difference between the mean estimated depth and actual depth. The accuracies of the methods were also examined in different ranges. The close, middle and far ranges were defined as 300-700 mm, 700-1100 mm and 1100-1500 mm respectively. The ranges were determined such that each comprises of an equal number of measurements. These distinctions are also relevant for humanoid applications such as reaching tasks where object detection typically occurs in the far range, approaching takes place in the middle range and grasping is executed in the close range. Table I shows the mean and standard deviation of the average errors over these ranges.

The theoretical maximum errors have been calculated for all methods (see Appendix) and are listed in Table II. The maximum inaccuracies that are likely to occur in our experimental setting were set as follows: Δb = Δr = ΔW = 1 mm, $\Delta\Theta_v$ = 0.18 degree, Δf = 0.1 mm, Δd = Δw = 0.5 pixel. For the familiar size method W was set at 90 mm which is the average size of the objects used.

[1]The terms mean estimated depth and mean estimation error imply given information is averaged over all 10 objects.

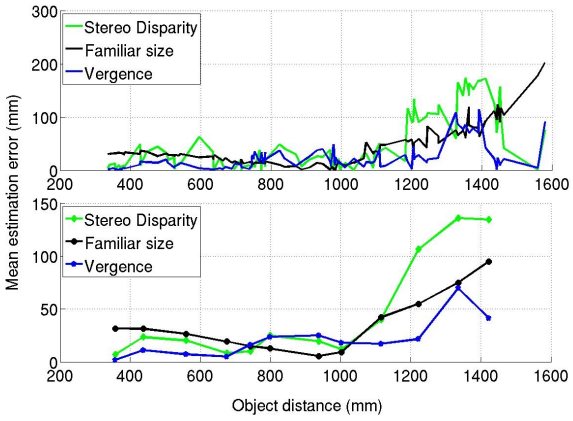


Fig. 5. Mean estimation errors of the methods. Plots above show pure data; plots below show running average of the data over a window size of eight data points.

TABLE I

MEAN (AND STANDARD DEVIATION) OF ESTIMATION ERRORS (IN MM) FOR ALL OBJECTS AT DIFFERENT RANGES.

Methods	Near	Middle	Far
Vergence	6.37 (5.51)	20.75 (11.79)	38.85 (32.92)
FS	27.22 (6.12)	11.82 (9.46)	76.68 (38.47)
SD	14.91 (15.32)	16.78 (12.84)	100.92 (51.56)
Combination of methods			
Vergence+FS	8.26 (5.40)	16.81 (10.41)	80.83 (54.67)
SD+FS	17.46 (11.60)	13.05 (10.84)	106.15 (41.03)
Vergence+SD	8.86 (7.36)	17.46 (10.54)	72.75 (41.97)
Vergence+FS+SD	10.33 (6.79)	15.23 (8.95)	84.74 (38.80)
Static-parallel stereo setup			
OpenCV	207.89 (45.73)	38.29 (34.63)	97.14 (75.53)
SVS	137.35 (81.05)	20.56 (10.42)	25.1 (9.49)

The overall accuracy of each method is shown in Fig. 6 where the estimation errors are averaged over 10 objects and the whole range. The mean errors and the standard deviations were approximately 22±24, 39±36 and 45±52 mm for the vergence, FS and SD methods respectively.

Results of the combinations of methods using maximum-likelihood estimators are shown in Fig. 7. Table I shows the mean and standard deviations of the estimation errors for the combined methods.

A comparison between the stereo disparity calculation for the active vision case and the static-parallel stereo case was also conducted (Fig. 8). Mean estimation errors of each of the two stereo algorithms used in the static-parallel stereo case is shown in Table I. Besides depth estimation performance, quality of disparity information across the objects was also examined for both active and static-parallel stereo cases (Fig. 9 and Fig. 10).

TABLE II

AVERAGE THEORETICAL ERROR VALUES (IN MM) FOR DIFFERENT RANGES.

Methods	Near	Middle	Far
Vergence	20.47	53.69	102.38
FS	19.54	41.18	68.62
SD	23.24	50.36	85.5

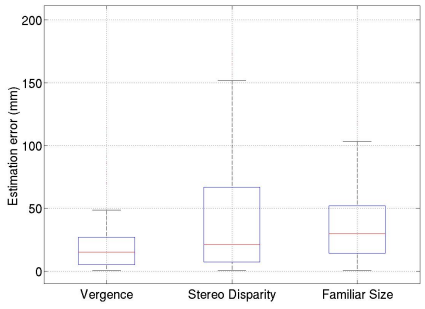


Fig. 6. Estimation error statistics (avg. over the objects and the whole range). Boxplot function of the Matlab Statistics Toolbox is used. The central mark is the median, the edges of the boxes are the 25th and 75th percentiles, the whiskers extend to the most extreme datapoints.

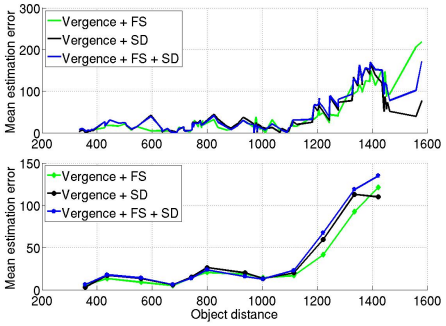


Fig. 7. Mean estimation errors of the combination of methods. Plots above show pure data; plots below show running average of the data over a window size of eight data points.

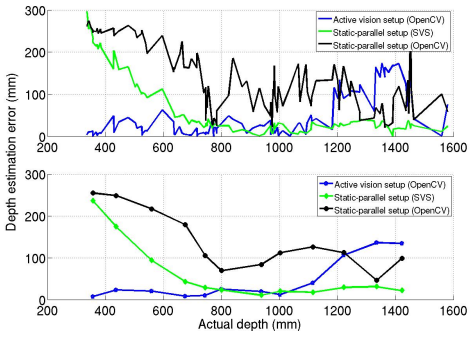
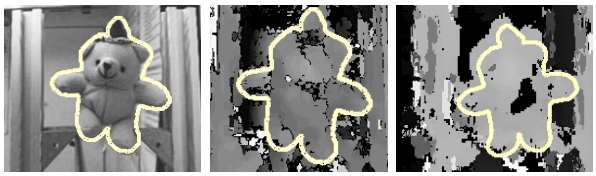
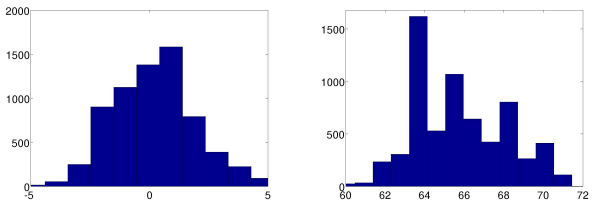


Fig. 8. (a) Mean estimation results of stereo disparity algorithms with active and static-parallel cases. (b) Mean estimation errors of stereo disparity algorithms with active and static-parallel cases. Plots above show pure data; plots below show running average of the data over a window size of eight data points.



(a) (b) (c)

Fig. 9. Disparity maps computed for an object at 340 mm distance. (a) Left rectified camera image (b) Disparity map computed with active stereo setup (c) Disparity map computed with static-parallel stereo setup. The object boundaries were highlighted in all three images. The images are cropped to show only relevant information.



(a) (b)

Fig. 10. Histograms of the disparity values across the object from the disparity maps shown in Fig. 9. (a) Active stereo case. (b) Static-parallel stereo case. Outliers are left out and only the relevant disparity ranges are displayed.

V. DISCUSSION

A. Error Sources

Depth estimation accuracy may be reduced by poor measurement of the intrinsic and extrinsic parameters of the vision system (e.g. baseline, focal length, etc.). The estimation errors for vergence were in the range of the theoretical maximum error values. The same is true for SD in the near and middle ranges.

Inaccuracy in the measurement of the vergence angle, which is mainly caused by measurement noise and low resolution of the vergence motors, is crucial for the vergence method, especially in the far ranges. It is obvious from (1) that even small inaccuracies cause high errors for small vergence angles. This has a high impact on estimation of distant targets. For example, 0.18 degrees (which is the resolution of the vergence motors) of inaccuracy in the vergence angle causes already 90 mm of estimation error at 1000 mm distance. Higher resolution motors may reduce this error. Increasing the baseline may also increase the reliable estimation range. However this may impede the anthropomorphic features of the vision system and may contradict humanoid considerations. Since the active rectification process relies on the vergence angle, the SD method is also affected from inaccurate vergence angles.

The accuracy of FS depends primarily on the object segmentation. Two pixels of error in the segmentation process leads to approximately 32, 83 and 157 mm of mean maximum theoretical errors in the near, middle and far ranges respectively. This may explain the estimation errors in our experiments.

Apart from those, a significant source of error for all methods is due to pixel quantization. Using higher resolution images may reduce this effect.

B. Combinations of Methods

The interpretation of results from the combinations of methods is two-fold. On one hand, the accuracy of individual methods may be improved by combining other methods. For example, combination of vergence with SD or FS gave better results than those two methods deliver alone in the near and far ranges. In the middle range, incorparating SD with FS improved the results of SD. On the other hand, overall comparison of results shows that the combinations of methods did not give the best results. The vergence outperformed any combination of methods in the near and far ranges while FS was the best method in the middle range. The reason for this may be that the integration method requires unbiased and uncorrelated signals to be effective and these requirements have not been met in this application. More advanced cue integration methods may be worth considering.

C. Active Vision vs Static-parallel Stereo

The results showed that all three methods (Vergence, SD and FS) compatible with active vision outperformed the methods using static-parallel stereo setup in the near and middle ranges. In the experiments for the static-parallel stereo case, a systematic underestimation in disparities (i.e. overestimations in depth) in the near and middle ranges for OpenCV block matching algorithm is noticeable (Fig. 8). This is not present in the SVS even though both algorithms use the same block matching method. This is likely due to the fact that the SVS applies post-processing on disparity calculation. Lack of this post-processing in the OpenCV algorithm is likely the cause of this systematic underestimation because for a close object situated at a large disparity, the propensity for search algorithms (performing the search starting from zero) to mismatch within a large disparity search range increases. This biases the disparity search towards disparity underestimation when the true object disparity is large. On the other hand, the estimations with the same OpenCV block matching algorithm using the active vision case did not show any systematic overestimations despite no post-processing was applied. The reason for this is likely that the disparity search is done in a smaller range around zero (i.e. the fixation point) in the active vision case, reducing the bias and evenly spreading it in positive and negative search directions.

The stereo disparity search in the static-parallel stereo camera configuration involves large disparities. Setting the disparity search range, which determines the magnitude of disparities that can be detected, too low impairs depth estimation in close ranges. High errors of the SVS in the near range are due to this phenomenon (Table I, Fig. 8). Removing erroneously labeled disparity values in disparity maps computed via the SVS reduced the mean estimation error to 20.28 mm in the close range. The search range can be increased at the expense of computation time, since increasing the search range means higher number of correlations to compute. This outlines a trade-off between the disparity range and computational resources. Active vision eliminates this trade-off by fixating on the object of interest and constraining the disparity search around zero. A virtual tracking system may solve this trade-off problem for the static-parallel stereo case only if the object is in the field of view of both cameras.

Another benefit of conducting a disparity search around zero for a given search range, as an active vision approach offers, is having better quality of depth information for a specific scene volume than the static-parallel stereo disparity search provides, especially in the near range. This may be especially beneficial in humanoid grasping tasks. Disparities

calculated with the active vision setup are distributed around zero at multiple values more smoothly, implying more continuous and detailed depth transitions across the object (Fig. 10(a)). Disparity values calculated with the static-parallel stereo setup are more discontinuously clustered and show artefacts in the representation of depth across the object (Fig. 10(b)).

VI. CONCLUSION

Depth estimation in the near visual field to support tasks like grasping for humanoids has been a major focus of this work. Three depth estimation methods incorporating active vision on a biomimetic head were tested. The accuracy of the methods were high in the near range and diminishes further away. Compared to static-parallel stereo methods examined in this work, active vision methods better results in the near and middle ranges. These results suggest active vision approach may be better suited for depth estimation within close and middle ranges where grasping and manipulation occur. It has also been shown that estimation of each method may be improved by combining with other methods. The accuracy of estimations may further be improved by using more precise hardware (e.g. higher resolution motors) and software (e.g. advanced object segmentation methods) solutions. The estimations that were based on pure sensory data in this work may be enhanced using common post-processing methods.

VII. ACKNOWLEDGMENTS

The work described was supported by the Honda Research Institute Europe.

REFERENCES

[1] J. Alomoinos, I. Weiss, and A. Bandopadhay. Active vision. *International journal on Computer Vision*, 1988.

[2] A. Bernardino and J. Santos Victor. Correlation based vergence control using log-polar images. In *Intelligent Robotic Systems, 4th Int. Symposium*, 1996.

[3] Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, pages 648–655, 1998.

[4] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, Inc., 1st edition, October 2008.

[5] M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993–1008, 2003.

[6] A. Dankers, N. Barnes, and A. Zelinsky. Active vision - rectification and depth mapping. In *Australasian Conf. on Robotics and Automation*, 2004.

[7] A. Dankers, N. Barnes, and A. Zelinsky. Map zdf segmentation and tracking using active stereo vision: Hand tracking case study. *Comput. Vis. Image Underst.*, 108(1-2):74–86, 2007.

[8] A. Denecke, I. Ayllon-Clemente, H. Wersing, J. Eggert, and J. Steil. Figure-ground segmentation using metrics adaptation in level set methods. In *European Symposium on Artificial Neural Networks (ESANN)*. D-Side, 2010. accepted.

[9] M. O. Ernst and M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, January 2002.

[10] A. Gasteratos, R. Martinotti, G. Metta, and G. Sandini. Precise 3d measurements with a high resolution stereo head. In *Image and Signal Processing and Analysis, First Int. Workshop*, pages 171–176, 2000.

[11] C. Goerick, J. Schmuedderich, B. Bolder, H. Janssen, M. Gienger, A. Bendig, M. Heckmann, T. Rodemann, H. Brandl, X. Domont, and I. Mikhailova. Interactive online multimodal association for internal concept building in humanoids. In *IEEE-RAS International Conference on Humanoids 2009*. IEEE, 2009.

[12] S. Kirstein, H. Wersing, and E. Koerner. A biologically motivated visual memory architecture for online learning of objects. *Neural Networks*, 21(1):65–77, 2008.

[13] K. Konolige. Small vision systems: hardware and implementation. In *Eighth Int. Symposium on Robotics Research*, page 111–116, 1997.

[14] S. E. Palmer. *Vision Science: Photons to Phenomenology*. The MIT Press, 1 edition, May 1999.

[15] D. Scharstein, R. Szeliski, and R. Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, 2001.

[16] C. Zhang, V. Willert, and J. Eggert. Tracking with depth-from-size. In *Neural Information Processing, 15th Int. Conference*, pages 274–283, 2009.

APPENDIX

Theoretical maximum errors are calculated and compared with the results from the experiments. For the vergence method main sources of error are the baseline (b) and the vergence angle (Θ_v). The maximum error for this method is calculated from (1) as

$$\Delta z_{verg} = \left|\frac{\partial z}{\partial b}\right| \cdot \Delta b + \left|\frac{\partial z}{\partial \Theta_v}\right| \cdot \Delta \Theta_v.$$

The individual components are calculated as

$$\frac{\partial z}{\partial b} = \frac{z}{b},$$
$$\frac{\partial z}{\partial \Theta_v} = -\frac{b}{4sin^2(atan(\frac{b}{2z}))}.$$

The SD is affected by inaccuracies in the baseline (b), focal length (f), disparity (d) and the distance between the rotation center of the camera and the image plane (r). The maximum error for the parallel stereo disparity method is derived from (2) as

$$\Delta z_{sd} = \left|\frac{\partial z}{\partial b}\right| \cdot \Delta b + \left|\frac{\partial z}{\partial f}\right| \cdot \Delta f + \left|\frac{\partial z}{\partial d}\right| \cdot \Delta d + \left|\frac{\partial z}{\partial r}\right| \cdot \Delta r.$$

The individual components are calculated as

$$\frac{\partial z}{\partial b} = \frac{z - r - f}{b},$$
$$\frac{\partial z}{\partial f} = \frac{z - r}{f},$$
$$\frac{\partial z}{\partial d} = -\frac{(z - r - f)^2}{b \cdot f},$$

and $\frac{\partial z}{\partial r} = 1$. The FS is affected by inaccuracies in the focal length (f), object size on the image (w) and physical size of the object(W). The maximum error is computed from (3) as

$$\Delta z_{fs} = \left|\frac{\partial z}{\partial f}\right| \cdot \Delta f + \left|\frac{\partial z}{\partial W}\right| \cdot \Delta W + \left|\frac{\partial z}{\partial w}\right| \cdot \Delta w.$$

The individual components are calculated as

$$\frac{\partial z}{\partial f} = \frac{z - r}{f},$$
$$\frac{\partial z}{\partial W} = \frac{w - r - f}{W},$$
$$\frac{\partial z}{\partial w} = -\frac{(z - r - f)^2}{W \cdot f}.$$