Verging Axis Stereophotogrammetry

Khurram Jawed and John Morris

Electrical and Computer Engineering, The University of Auckland, New Zealand
mjaw002@aucklanduni.ac.nz, j.morris@auckland.ac.nz

Abstract. Conventional stereophotogrammetry uses a canonical configuration in which the optical axes of both cameras are parallel. However, if we follow lessons from evolution and swivel the cameras so that their axes intersect in a fixation point, then we obtain considerably better depth resolution. We modified our real-time stereo hardware to handle verging axis configurations and show that the predicted depth resolution is practically obtainable. We compare two techniques for rectifying images for verging configurations. Bouguet's technique gives a simpler geometry - the iso-disparity lines are straight and the familiar reciprocal relationship between depth and disparity may still be used. However when the iso-disparity lines are the Veith-Muller circles, slightly better depth resolution may be obtained in the periphery of the field of view - at the expense of a more complex conversion from disparity to depth.

1 Preamble

Although the underlying geometry is well understood and mathematical models for verging axis stereophotogrammetry long published[1], the advantages of these configurations - discovered millions of years ago in the evolutionary process as animals learned to swivel their eyes in their sockets[2] - seem to have been substantially overlooked in favour of the trivially modeled canonical configurations in which the optical axes are parallel. Iso-disparity surfaces are also known as horopters[3]. The intersections between the horopters and the plane containing the optical centres and the fixation point are the Veith-Muller circles. Pollefeys *et al.*[4] analyzed these iso-disparity curves for different camera configurations. Olson *et al.*[5] studied the use of the horopter for active stereo heads. Here, we show that verging axis configurations lead to better depth resolution. Further, we implemented a real-time stereo system handling verging camera configurations in an FPGA. We report several experiments to validate the predicted positions and separations of the iso-disparity lines and demonstrate the enhanced depth resolution compared to a canonical stereo configuration.

Stereophotogrammetry systems usually capture 'raw' images from two cameras and then rectify them so that the images correspond to those taken by ideal pin-hole cameras - in a canonical configuration - with their axes parallel and perpendicular to the baseline joining the optical centres of both cameras[6]. This configuration has a significant advantage: scan lines of the rectified images are the epipolar lines so that the search for corresponding points in the two images may be constrained to the scan lines turning an $\mathcal{O}(n^2)$ search into a $\mathcal{O}(n)$

Y.-S. Ho (Ed.): PSIVT 2011, Part I, LNCS 7087, pp. 202–213, 2011.

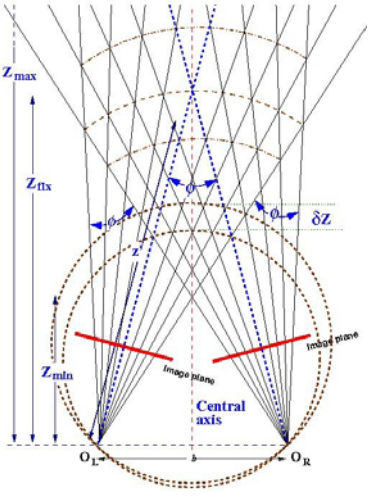


Fig. 1. Verging Axis Geometry showing two Veith-Muller circles. $O_{L|R}$ are the optical centres.

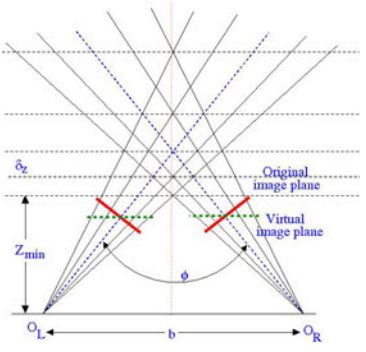


Fig. 2. Verging Axis Geometry using Bouguet's method[7] - the original image plane is transformed to the virtual one parallel to the base line and the principal points moved so that the optical axes intersect at the original fixation point

one. First, we show the theoretical benefits of verged axis systems (principally in enhanced depth resolution) and then show how our real-time stereo hardware was modified to gain these benefits.

If the cameras are deliberately verged, then we can use the same rectification procedures to convert the raw images to those taken by ideal cameras in the canonical configuration, but this loses the enhanced depth resolution of the verging configuration. We compare two techniques for rectifying verged camera images in ways that retain the improved depth resolution. We present some laboratory images of the same object taken with both configurations - empirically demonstrating the benefits and confirming the predicted benefit. Finally, the costs of both configurations were compared.

2 Stereo Geometry

A verging axis configuration is illustrated in Figure 1. For simplicity, we assume that two identical pin-hole cameras are rotated around an axis perpendicular to the baseline joining the optical centers of the two cameras so that the optical axes meet at a *fixation point* in the scene. We use capital letters, (X, Y, Z), for coordinates in a 'world' frame centred on the baseline midway between the optical centres of the two cameras, with its X-axis parallel to the baseline, its Z-axis lying in the same plane as the camera optical axes and its Y-axis perpendicular to the baseline. Lower case, (x, y, z) (with L or R subscripts as needed), is used for camera based coordinates and lengths. Both cameras are rotated about their

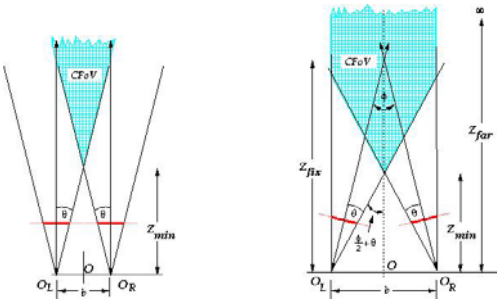


Fig. 3. Image plane use for a canonical configuration (left) *vs* a verging axis one (right): note the large areas outside the common field of view (CFoV) imaged in the canonical configuration

y-axes so that their optical axes intersect at an angle ϕ (the vergence angle) at the fixation point, $(0, 0, Z_{fix})$. Then

$$Z_{fix} = \frac{b}{2} \cot \frac{\phi}{2} \tag{1}$$

where b = baseline length. In the canonical configuration, the optical axes are parallel so $\phi = 0$ and $Z_{fix} = \infty$.

2.1 Depth Resolution

In stereo systems, depth is recovered from a pair of images by measuring the *disparity* or separation between pixels corresponding to the same scene point in the left and right images. In the verging axis configuration, the camera axes intersect at the fixation point in the scene. This point appears at the same position in both image planes and thus has disparity, $d = 0$. The loci of points with the same disparity are the Veith-Muller circles - see Figure 1. Considering only points along the central axis of the system, $X = 0, Y = 0$, the distance to points of disparity, d,

$$Z(d, \phi) = \frac{b}{2} \cot(\frac{\phi}{2} + \tan^{-1}(\frac{d}{2\lambda})) \tag{2}$$

where $\lambda = f/\tau$ (f = focal length and τ = pixel width) is the focal length in pixels.

Most stereo correspondence algorithms measure disparity in integral pixels only, so that the depth resolution at any point on the central axis is

$$\begin{aligned} \delta Z(d, \phi) &= Z(d, \phi) - Z(d-1, \phi) \\ &= \frac{b}{2}(\cot(\frac{\phi}{2} + \tan^{-1}(\frac{d}{2\lambda})) - \cot(\frac{\phi}{2} + \tan^{-1}(\frac{d-1}{2\lambda}))) \end{aligned} \tag{3}$$

Note that, in a verging axis system, unlike the canonical configuration, disparities may be negative: points with $Z > Z_{fix}$ will have $d < 0$.

Some practical constraints govern any stereo configuration. A practical correspondence algorithm will be able to handle disparities in some range, $d_{min} \leq d \leq d_{max}$, thus depth can only be measured in the area between $Z_{max} = Z(d_{max}, \phi)$ and $Z_{min} = Z(d_{min}, \phi)$ along the central axis ($X = 0$) and, in general, between the Veith-Muller circles for d_{max} and d_{min}.

To understand the increase in depth resolution, in Figure 1, observe the intersections of the rays projected through image plane pixels and the central axis. These rays intersect the line $X = 0$ at points of even disparity: thus the distance between any two intersections is roughly twice the depth resolution at that point. As the vergence angle increases, these gaps become smaller and depth resolution improves. We can also observe that the distance over which usable 3D data can be obtained, *i.e.* between Z_{min} and Z_{max}, shrinks as ϕ increases: this distance is divided into $d_{max} - d_{min} + 1$ measurable intervals, so depth resolution increases over the whole usable area. However, note that, in general, Z_{max} is no longer at infinity whereas $Z(d = 0, \phi = 0) = \infty$, so that the increased depth resolution is not without limitations. In practice this is rarely a problem, because the depth resolution, $\delta Z(d = 0, \phi = 0) = \infty$, is of little practical value.

Verging axis configurations also 'waste' less of the image planes of both cameras. Figure 3 shows wide regions of monocular points - for which no depth information can be derived. With a verging axis configuration, the full image planes of both cameras are used effectively.

In a canonical configuration, disparities are constant along straight lines parallel to the baseline, leading to the familiar relationship between depth and disparity:

$$Z = b\lambda/d \tag{4}$$

However in verging axis configurations, disparities are constant along the Veith-Muller circles (*cf.* Figure 1) leading to a more complex transformation, $d \rightarrow Z$[8], For corresponding pixels at $(u_{L|R}, v_{L|R})$ in the left and right images respectively:

$$\begin{aligned} Z &= \frac{b}{\tan\left(\phi_L + \tan^{-1}\frac{u_L}{\lambda}\right) + \tan\left(\phi_R + \tan^{-1}\frac{u_R}{\lambda}\right)} \\ X &= \frac{b\tan\left(\phi_L + \tan^{-1}\frac{u_L}{\lambda}\right)}{\tan\left(\phi_L + \tan^{-1}\frac{u_L}{\lambda}\right) + \tan\left(\phi_R + \tan^{-1}\frac{u_R}{\lambda}\right)} - \frac{b}{2} \\ Y &= \frac{bv_L}{\lambda\left(\tan\left(\phi_L + \tan^{-1}\frac{u_L}{\lambda}\right) + \tan\left(\phi_R + \tan^{-1}\frac{u_R}{\lambda}\right)\right)} \end{aligned} \tag{5}$$

2.2 Rectification

Our first approach to rectification converts the original raw images to ones in which the 'scan' lines are these epipolar lines. Firstly, we remove distortion and align the images so that the optical axes intersect at $(0, 0, Z_{fix})$ (*i.e.* the cameras

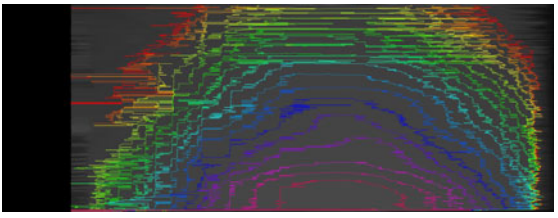


Fig. 4. Disparity map contours - flat panel roughly perpendicular to the system axis. Note that the regions of equal disparity are curved because the flat surface of the object intersects several Veith-Muller circles *cf.* Figure 1.

have been rotated around their y axes only[1]). The epipolar lines are now straight lines crossing the 'raw' images at an angle to the original scan lines (except for the scan line passing through the principal point).

Computing Epipolar Lines. The fundamental matrix, $\mathbf{F}$, was found using the eight point algorithm[9]. We identified corresponding pairs of epipolar lines for each image using $\mathbf{F}$[9]: for any point, p, in the left image, the corresponding epipolar line in the right image is $l' = \mathbf{F}p$. Similarly for a point p' in right image, the corresponding epipolar line in the left image is $l = \mathbf{F}^T p'$.

We now generate images in which the 'rows' are these epipolar lines: they can be fed directly to a correspondence algorithm used for a canonical configuration: it expects epipolar lines - in the canonical configuration, these are the same as scan lines. We simply changed the rectification lookup table so that it generated epipolar lines rather than scan lines. With this method, the depth resolution is the distance between adjacent Veith-Muller circles in Figure 1: it is given by Equation 3 along the central axis $(X = 0, Y = 0)$ of the system.

Figure 4 shows a disparity map obtained by this method. The viewed object is flat but equal disparity regions are curved as expected *cf.* Figure 1.

2.3 Bouguet's Method

An alternative method due to Bouguet[7] also preserves the enhanced depth resolution. It rectifies the two images into a canonical configuration and then re-projects them so that the optical axis meet at the fixation point. It computes a rectification matrix, $\mathbf{R}^{rect}$, that takes the epipole in the left camera to infinity. The rotation matrix, $\mathbf{R}$, computed from calibration, is split into two matrices, $\mathbf{R}_L$ and $\mathbf{R}_R$, rotating each camera by the same amount. From the original camera matrices, $\mathbf{M}_L$ and $\mathbf{M}_R$, rectified camera matrices are then computed $\mathbf{M}_L^{rect} = \mathbf{M}_L\mathbf{R}^{rect}\mathbf{R}_L$ and $\mathbf{M}_R^{rect} = \mathbf{M}_R\mathbf{R}^{rect}\mathbf{R}_R$. $\mathbf{M}_L^{rect}$ and $\mathbf{M}_R^{rect}$ are multiplied by

[1] This implies that small unintended rotations about camera x and z axes have been corrected.

projection matrices, with μ_{x_L} and μ_{y_L} set so that the two optical axis intersect at the fixation point - see Figure 2.

This method gives a slightly better depth resolution along the central axis ($X = 0, Y = 0$) than the verging configuration in Figure 1, but slightly worse depth resolution for points on the periphery of the field of view, where the Veith-Muller circles get closer - see Figure 6. This changes the curved boundaries between regions of equal disparity to straight lines. Transforming disparity to depth uses the re-projection matrix:

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 & -\mu_{x_L} \\ 0 & 1 & 0 & -\mu_{y_L} \\ 0 & 0 & 0 & \lambda \\ 0 & 0 & -\frac{1}{b} & \frac{\mu_{x_L} - \mu_{x_R}}{b} \end{bmatrix} \tag{6}$$

where $(\mu_{x_{L|R}}, \mu_{y_{L|R}})$ is the optical center of the (left|right) camera.

$$\begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = \mathbf{Q} \begin{bmatrix} \frac{u}{2} \\ v \\ d \\ 1 \end{bmatrix} \tag{7}$$

The 3D coordinates are $(X/W, Y/W, Z/W)$. Depth resolution is now:

$$\begin{aligned} \delta Z(d)_{Bouguet} &= Z(d) - Z(d-1) \\ &= b\lambda\left(\frac{1}{d - (\mu_{x_L} - \mu_{x_R})} - \frac{1}{(d-1) - (\mu_{x_L} - \mu_{x_R})}\right) \end{aligned} \tag{8}$$

3 Implementation

3.1 Stereo Hardware

The real time stereo matching hardware uses Gimbelfarb's Symmetric Dynamic Programming Stereo algorithm[10]. It has a pair of Cameralink cameras attached directly to an Altera Stratix III FPGA connected to the host PC via an 8-lane PCIExpress bus. The FPGA removes lens distortion, rectifies the images and produces disparity and occlusion maps. It can compute dense disparity maps with 128 disparity levels at 30fps[11].

A checkerboard pattern was used for calibration[12].

For both rectification methods, all the corrections are combined into a single lookup table containing displacements for every pixel in the left and right rectified image. These lookup tables are reduced[11] and fed to the real time stereo matching hardware that produces the left and right rectified images and the disparity and occlusion maps.

4 Experiments

4.1 Experiment 1 - Stepped Target

For ground truth, we constructed a simple stepped target from Lego blocks, see Figure 5. Lego blocks are, of necessity, produced with the high dimensional accuracy needed to allow thousands of blocks to be used to build complex models. We measured a sample of blocks and confirmed that each block's dimensions were the same to within 0.1mm. Our test structure (Figure 5) has six steps each of two blocks and a height of 15.6 ± 0.1 mm.

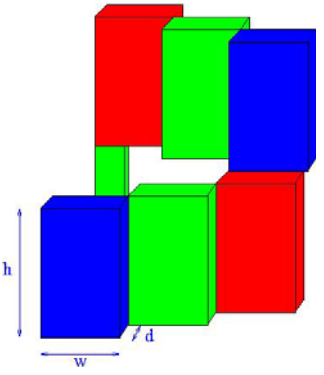


Fig. 5. Lego block structure $h = 63$mm, $w = 40$mm and $d = 15.6$mm

Disparity maps were acquired with both canonical and verging configurations with the target at various depths. The configurations were - canonical: $b = 80$mm, $f = 9$mm and $\phi = 0$ giving a predicted depth resolution from 9.7 mm to 993mm for disparity values from 126 to 12; verging: $b = 427$mm, $f = 9$mm and $\phi = 17.15^o$, fixation point at 1400mm and predicted depth resolution from 1.6mm to 2.3mm for disparity values from 126 to 12. In the verging configuration, a longer baseline was used so that the target fills the field of view at approximately the same distance.

Disparity maps are shown in Figure 7. In every case, the expected disparity was observed and step depths were correct to within the predicted depth resolution for that disparity.

4.2 Experiment 2 - Sphere

In the second experiment, we used a ten-pin bowling ball: bowling balls must be precise spheres[2] so a ground truth can be derived from the geometry of a sphere. The ball was placed 600mm in front of the cameras and disparity maps were captured with both verging and canonical configurations. Configurations

[2] Round to within 0.010" (or 0.25mm in more widely accepted units)[13].

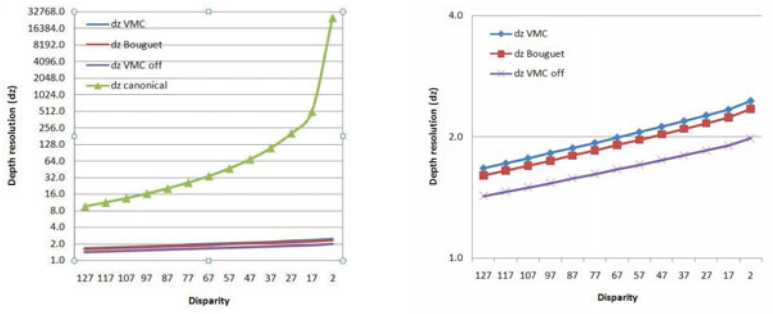
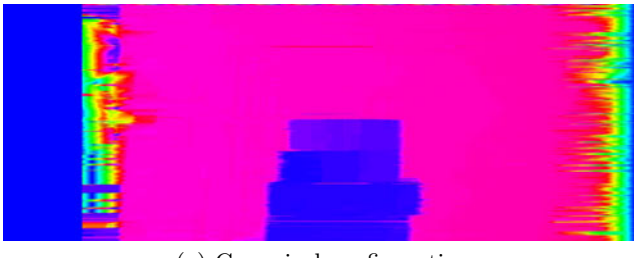
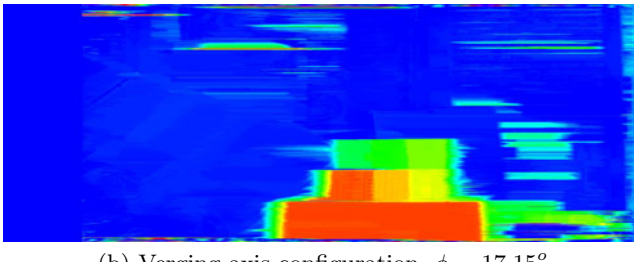


Fig. 6. Depth resolution for configurations of Experiment 1. dz VMC is depth resolution for a verging configuration (Figure 1), dz Bouguet is the depth resolution for Bouguet's method (Figure 2), dz canonical is the depth resolution for a canonical configuration and dz VMC off is the depth resolution for verging configuration but along the periphery of Figure 1. The left figure compares all configurations while the right figure compares verging configurations only at an expanded scale.



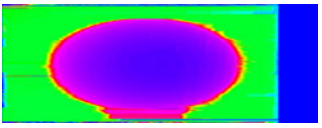
(a) Canonical configuration



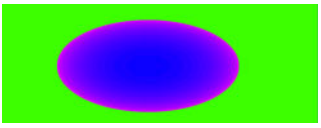
(b) Verging axis configuration, $\phi = 17.15^o$

Fig. 7. Disparity maps for Lego block structure - note the increased number of disparity changes evident for the verging axis configuration

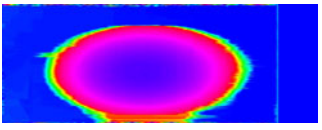
were - canonical $b = 38.7$mm, $f = 9$mm and $\phi = 0$, the depth resolution ranged from 4.9mm to 1.25m for a disparity range of 123 to 2; verging $b = 95.9$mm, $f = 9$mm and $\phi = 5.2^o$, depth resolution of 2.0mm to 5.5mm for a disparity



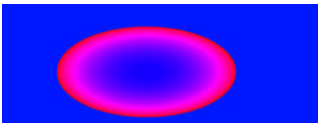
(a) Canonical configuration disparity map



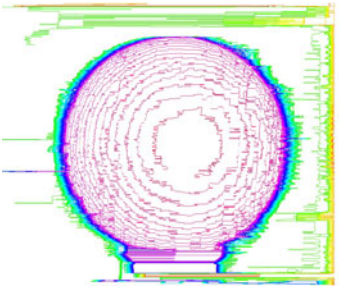
(b) Canonical configuration ground truth



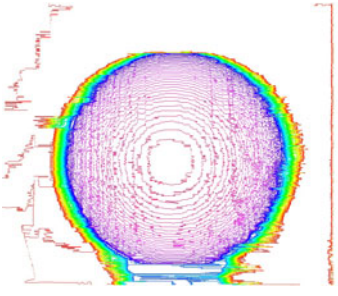
(c) Verging configuration disparity map



(d) Verging configuration ground truth



(e) Canonical configuration contours



(f) Verging configuration contours

Fig. 8. Sphere experiment: contours on disparity maps. Note that SDPS produces a double-width disparity map[10], (a) through (d), leading to the apparently flattened images: when they are rescaled to the same width as the raw images, the contours are circles - see (e) and (f).

Table 1. Bowling ball matching performance

	canonical configuration	verging axis configuration
Threshold	% Bad pixels	% Bad pixels
0.5	70	68
1	44	37
1.5	22	15
2	10	7
RMS error (disparity units)	2.2	2.4

range of 123 to 2. and fixation point at 1055mm. The results of the experiments are summarized in Table 1 and the depth resolution plotted in Figure 9. The disparity maps and ground truth are in Figure 8.

4.3 Experiment 3 - Statue

The third experiment captured disparity maps of a complex object. This experiment demonstrates the increased depth resolution for a 'real' target. The configurations for this experiment were the same as those for experiment 2.

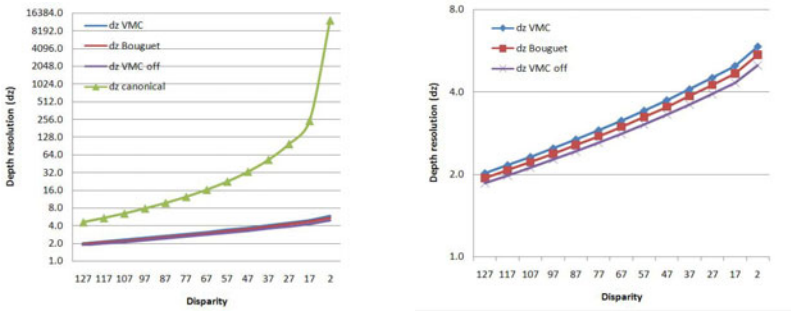


Fig. 9. Depth resolution for configurations used in Experiment 2. Labels are the same as for Figure 6.

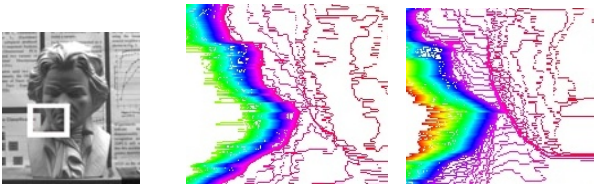


Fig. 10. Contours derived from disparity maps of the statue. From left to right: original raw image, canonical configuration contours and verging axis configuration contours. A small area of the disparity maps has been expanded to show the contour detail.

Fig. 11. System handling negative disparities: the fixation point is on the statue's nose, so that all points on the face and background appear to the left in the right image

Because there is no ground truth, Salmon[14]was used to find contours on disparity maps from both configurations. Contours on the verging axis disparity maps are more closely spaced due to the higher depth resolution. Note that, in the binocularly visible part of the face shown in Figure 10, there are 31 contours in the verging axis configuration disparity map compared to 13 - an increase in depth resolution of ~ 2.5.

4.4 Hardware Costs

In our FPGA hardware, rectification uses a lookup table which maps pixels in the desired configuration to actual image pixels. For either verging configuration,

we simply compute a different lookup table and load it. Negative disparities are handled by trivial changes to the FPGA hardware - adjusting the length of the right pixel delay register - *decreasing* it for negative disparities. Figure 11 shows a pair of images with negative disparities and the contoured disparity map obtained. The total logic utilization was decreased by 4%.

4.5 Computation Costs

Rectification using the epipolar lines requires more complex computation to convert image coordinates to real-world coordinates. In software, conversion of a 1.5×10^6 entry disparity map to real world coordinates using Equation 7 takes 125ms, whereas using Equation 5 requires 325ms (2.0 GHz Pentium Dual Core). Either computation could be moved to the FPGA hardware leading to a negligible ($< 1ms$) increase in latency as pixels of the disparity map are streamed out of the correspondence circuit's back-track module[15].

5 Conclusion and Future Research

We have shown theoretically and verified experimentally that a verging axis configuration gives better depth accuracy. Verging axis configurations also generally produce a more useful common field of view. The depth resolution is essentially similar using either rectification approach. Some applications (*e.g.* collision avoidance, where we may only need a warning that a hazard has encroached an exclusion zone) can work with disparity data. Where conversion from disparities in pixels to world coordinates is required, Bouguet's rectification procedure is faster when the conversion must be performed in software on the host and adds slightly less latency if the hardware is used. The Veith-Muller circles give a slightly better depth resolution at the periphery of the common field of view but take longer to convert from disparity space to world space. In our FPGA system, rectification uses lookup tables, so there is no additional hardware cost or latency for a verging axis configuration: thus, the increased flexibility to design a more useful common field of view combined with superior depth resolution makes verging axis configurations preferable in practical configurations. Allowing the area behind the fixation point to be used (*i.e.* allowing negative disparities) produces a slightly smaller circuit by shortening the right pixel delay register and adds further flexibility in choosing the imaged region.

References

1. Maybank, S.: Theory of Reconstruction from Image Motion. Springer, Heidelberg (1993)
2. Meissner, G.: Beitrge zur Physiologie des Sehorgans. Leipzig, Engelmann (1854)
3. Aguilonii, F.: Opticorum Libri Sex philosophis juxta ac mathematicis utiles. Antwerp (1613)

4. Pollefeys, M., Sinha, S.: Iso-Disparity Surfaces for General Stereo Configurations. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004, Part III. LNCS, vol. 3023, pp. 509–520. Springer, Heidelberg (2004)
5. Olson, T.: Stereopsis for verging systems. In: CVPR 1993, pp. 55–60 (1993)
6. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47, 7–42 (2002)
7. Bouguet, J.Y.: Camera calibration toolbox for Matlab (1999)
8. Woods, A., Docherty, T., Koch, R.: Image distortions in stereoscopic video systems. In: Proceedings of the SPIE: Stereoscopic Displays and Applications IV, vol. 1915 (1993)
9. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press (2004)
10. Gimel'farb, G.L.: Probabilistic regularisation and symmetry in binocular dynamic programming stereo. Pattern Recognition Letters 23, 431–442 (2002)
11. Jawed, K., Morris, J., Khan, T., Gimel'farb, G.: Real time rectification for stereo correspondence. In: Xue, J., Ma, J. (eds.) 7th IEEE/IFIP Intl Conf on Embedded and Ubiquitous Computing (EUC 2009), pp. 277–284. IEEE CS Press (2009)
12. Bradski, G., Kaehler, A.: Learning OpenCV: Computer vision with the OpenCV library. O'Reilly Media, Inc. (2008)
13. United States Bowling Congress: Equipment Specifications and Certification Manual (2009)
14. Khan, T., Morris, J., Javed, K., Gimelfarb, G.: Salmon: Precise 3d contours in real time. In: Proceedings of the 2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, DASC 2009, pp. 424–429. IEEE Computer Society, Washington, DC, USA (2009)
15. Morris, J., Jawed, K., Gimel'farb, G., Khan, T.: Breaking the 'ton': Achieving 1% depth accuracy from stereo in real time. In: Bailey, D. (ed.) Image and Vision Computing. IEEE CS Press, NZ (2009)