

Single Camera Stereo System Using Prism and Mirrors*

Gowri Somanath, Rohith MV, and Chandra Kambhamettu

Video/Image Modeling and Synthesis (VIMS) Lab, Department of Computer and Information Sciences, University of Delaware, Newark, DE, USA
<http://vims.cis.udel.edu>

Abstract. Stereo and 3D reconstruction are used by many applications such as object modeling, facial expression studies and human motion analysis. But synchronizing multiple high frame rate cameras require special hardware or sophisticated software solutions. In this paper, we propose a single camera stereo system by using a setup made of prism and mirrors. Our setup is cost effective, portable and can be calibrated similar to two camera stereo to obtain high quality 3D reconstruction. We demonstrate the application of the proposed system in dynamic 3D face expression capture, depth super-resolution in stereo video and general depth estimation.

1 Introduction

The knowledge of depth in a scene has been known to simplify many problems in computer vision, for example, face and expression recognition [1,2], tracking [3,4], object discovery [5], 3D video content creation and many other applications. Stereo systems are easy to setup in cases where the scene is stationary. For many applications, high frame rate or video cameras are required. Synchronizing multiple such cameras is a challenging task and requires expensive hardware or sophisticated software solutions. For newly deployed systems, the above solutions can be employed. But in systems where cameras have already been deployed, such as surveillance, changing the existing camera setup is not always practical. Portability and maintenance of multicamera setups is harder than single camera. On the other hand, the calibration of stereo/multi-view systems provides true scale reconstruction which is useful and even essential in many real world applications. A portable, cost effective and single camera system which can be calibrated like stereo to provide metric reconstruction is hence desirable. In this paper, we present a solution using prism and two mirrors, which can be used to turn almost any camera into a stereo camera. The system can be calibrated to obtain high quality metric reconstruction. The proposed solution is cost effective and can be easily used as a detachable accessory for a camera lens. An equilateral prism and two first surface mirrors are used to build the system shown in

* This work was made possible by NSF Office of Polar Program grants, ANT0636726 and ARC0612105.

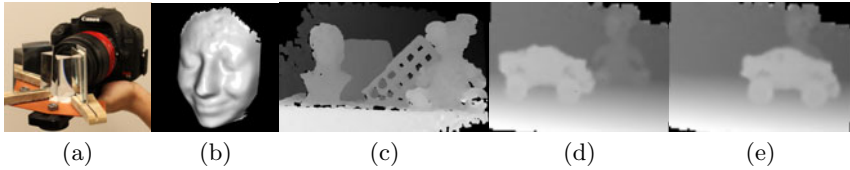


Fig. 1. (a) Our camera setup, (b) 3D reconstruction from a facial expression video sequence, (c) Disparity map of a sample scene, (d),(e) Two frames from a sequence of super-resolved disparity maps of a dynamic scene. The stereo video is 640 X 720 pixels, which is super-resolved to 3168 X 2376 pixels using a low frame rate, motion compensated full resolution still image.

Figure 1(a). The camera used is a Canon T1i DSLR, which can capture HD video (720p) at 30fps and 15MP still images.

The paper is organized as follows. In Section 2 we discuss existing systems. We detail the proposed setup in Section 3. We demonstrate the use of the system for two types of applications which have different needs. In Section 4.1 we show how the system can be used to perform depth super-resolution of dynamic scenes. We present calibrated 3D reconstruction results in Section 4.2. Section 4.3 provides more results and we conclude in Section 5.

2 Previous Work

The idea of converting single camera into stereo using a combination of mirrors, lenses and prisms has been proposed earlier by many researchers. Figure 2 shows a schematic comparison of the different existing setups. One of the oldest known systems, known as the Stereoscopic Transmitter, was built in 1894 by Theodore Brown. A similar system was proposed by Gluckman and Nayar [6,7], which used a pair of hinged mirrors placed in front of the camera with the reflecting surface facing the lens. The scene being imaged was to the back and one side of the system. The placement of prisms and mirrors at considerable distance away from the lens, makes the above system less portable. Also the idea cannot be easily implemented to convert existing single camera systems, like surveillance, to stereo systems. Many catadioptric stereo systems have been proposed using hyperbolic [8] and parabolic mirrors [9]. Nene and Nayar [10] provide some possible configurations using planar, parabolic, elliptic, and hyperbolic mirrors. Mitsumoto et al. [11] and Zhang and Tsui [12] propose a system which images the object and its reflection in a set of planar mirrors. Pyramidal arrangement of planar mirrors was used in [13]. Gluckman and Nayar also proposed systems using single and three mirror configurations [14,15], where the system was constrained to produce rectified images instead of 3D sensitivity (number of levels in disparity map). Gao and Ahuja [16] used a single camera and a rotating planar plate in front of the camera, similar to that proposed by Nishimoto and Shirai [17]. Due to need to capture multiple images with two or more different plate positions, these systems are restricted to static scenes. In [18], Lee et al. used a

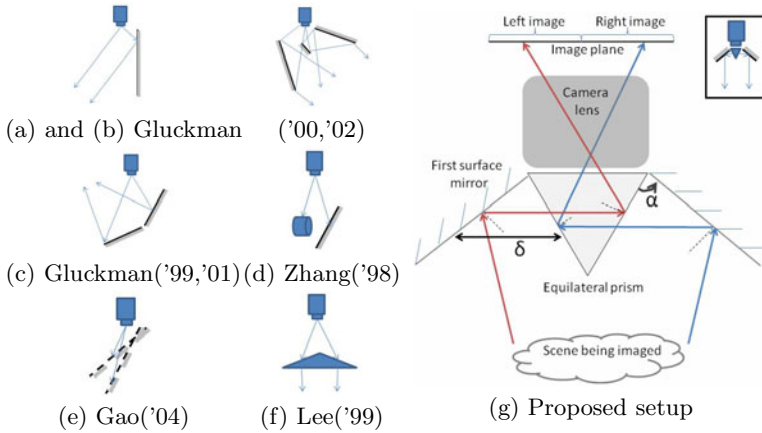


Fig. 2. (a)-(f) Schematic of previous systems. (g) Schematic top view of the proposed mirror and prism arrangement.

biprism placed 15cm in front of the camera to obtain two views of the scene. The idea was generalized in [19] using trinocular prisms. [20,21] proposed a setup using a prism and two mirrors, where the prism edge was used as a reflective surface. This required the use of specialized prisms and also resulted in a setup where the prism had to be kept 12-13cm from the lens of the camera. This made the system less portable and difficult to be used as an add-on attachment for existing systems. Also, due to use of specialized prism, the system is expensive. A commercial product, Tri-Delta stereo adapter, works on similar principle as our system. However the adapter is mounted on the lens and the camera would be placed such that the lens points towards the top (sky) or bottom (ground), to image the scene in front of the observer. Our setup contrasts the adapter in the following manner. The tri-delta is a completely mirror based setup and is compatible with only a few point-and-shoot cameras. Our setup is flexible in terms of baselines and vergence control, which is essential for machine vision studies. Similar to many previous works, the Tri-Delta would also require changes to the way a camera is setup with respect to the viewed scene. Also the cost of our construction is only a fraction of that of the tri-delta. A different approach to obtain stereo video is used in 'NuView Adapter' available for video cameras only. The adapter uses dual LCD shutters, a beamsplitter and a mirror arrangement to provide left and right views as odd and even fields. Our setup is much easier to construct, can be used with both still and video cameras and is much more cost effective.

3 Proposed System

In this paper, we propose a single camera system with the following characteristics:(1) Simple and portable construction: The proposed extension to a camera can be used as a lens attachment. The prism and mirror setup mount in front of

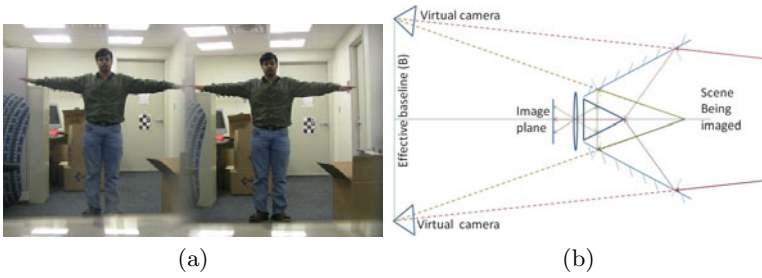


Fig. 3. (a) Raw image captured by our setup. The stereo images are extracted as two halves of the raw image. (b) Ray diagram illustrating the formation of the two images, the virtual cameras and baseline.

the lens and hence the entire system is portable. (2) Easy to deploy on existing single camera systems: Many applications like surveillance and monitoring have single cameras already in place. The proposed setup can be easily attached to such cameras to obtain stereo video/images. Unlike many previously proposed systems, the position of imaged scene with respect to camera does not have to be changed. (3) Can be calibrated to obtain metric reconstruction: Some applications only use depth separation or disparity information, while many others require true scale metric 3D reconstruction (for example, 3D modeling of objects, person tracking etc). The proposed setup can be calibrated like a two camera stereo system. (4) Flexible: Baseline and vergence are some of the flexible parameters in a stereo setup. As will be explained later, the placement of the mirrors can be varied to change the baseline and vergence in our system.

3.1 Setup

The setup is made using an equilateral prism and two first surface mirrors. As shown in Figure 2(g), the prism is placed at the end of the camera lens. The mirrors are placed at an angle α , with respect to the corresponding prism surface. The distance δ , between the mirror and prism can also be varied. Different configurations of α and δ can be used to obtain suitable baseline, vergence and overlap between the left and right images. Intuitively, smaller α and δ are used when objects being imaged are within a few feet of the camera. A detailed analysis of the geometry is discussed in the following section. We have attached the setup to various types of cameras including DSLR, camcorder and network/IP cameras with wide angle lens. Figure 1(a) shows the setup with a DSLR camera. In our experiments, we used a prism of edge length two inches and base of one inch. The mirror is 2 inches by 2 inches.

3.2 Analysis of the Setup

Figure 3(b) shows the ray diagram illustrating the stereo image formation and virtual cameras. We now derive the conditions on the mirror angle and the

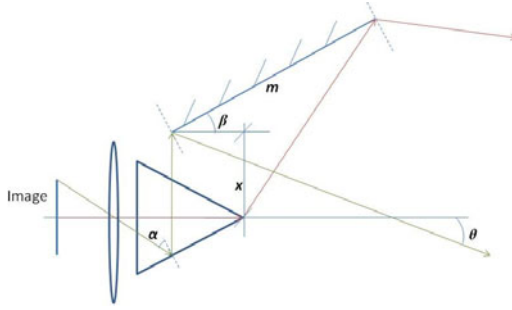


Fig. 4. Ray diagram showing angles and distances used in the analysis (Section 3.2)

effective baseline of our setup. Figures 3(b) and 4 show the ray diagrams with angles and distances used in the following analysis. We define the following angles and distances in the setup: ϕ is the horizontal field of view of camera in degrees, α is the angle of incidence at prism, β is the angle of inclination of mirror, θ is the angle of scene ray with horizontal, x is the distance between mirror and camera axis, m is the mirror length and B is the effective baseline (Fig. 3(b)). To calculate the effective baseline, we trace the rays in reverse. Consider a ray starting from the image sensor, passing through the camera lens and incident on the prism surface at an angle α . This ray will then get reflected from the mirror surface and go towards the scene. The final ray makes an angle of θ with the horizontal as shown in the figure. It can be shown that $\theta = 150 - 2\beta - \alpha$. In deriving the above, it has been assumed that there is no inversion of image from any of the reflections. This assumption is violated at large fields of view. More specifically, it is required that $\phi < 60^\circ$. Since we are not using any lens apart from the camera lens, the field of view in resulting virtual cameras should be exactly half of the real camera. In Figure 3(b), consider two rays from the image sensor, one ray from the central column of image ($\alpha_0 = 60^\circ$, shown in green) and another ray from the extreme column ($\alpha = 60^\circ - \phi/2$, shown in red). The angle between the two scene rays is then $\phi/2$. For stereo, the images from the two mirrors should contain some common part of the scene. Hence, the scene rays must be towards the optical axis of the camera rather than away from it. Also, the scene rays must not re-enter the prism as this would lead to internal reflection and not provide an image of the scene. Applying the two conditions above, we can show that the inclination of the mirror is bound by the following inequality $\phi/4 < \beta < 45^\circ + \phi/4$. We can now estimate the effective baseline based on the angle of scene rays, mirror length and distance of the mirror from the axis as follows.

$$B = 2 \frac{x \tan(2\beta - \phi/2) - m \cos(\beta) - (x + m \cos(\beta)) \tan(2\beta)}{\tan(2\beta - \phi/2) - \tan(2\beta)}.$$

In our setup, the parameters used were focal length of 35 mm corresponding to $\phi = 17^\circ$, $\beta = 49.3^\circ$, $m = 76.2\text{mm}$ and $x = 25.4\text{mm}$. The estimated baseline is

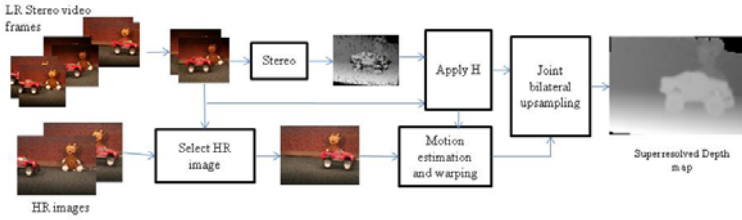


Fig. 5. Depth super resolution pipeline

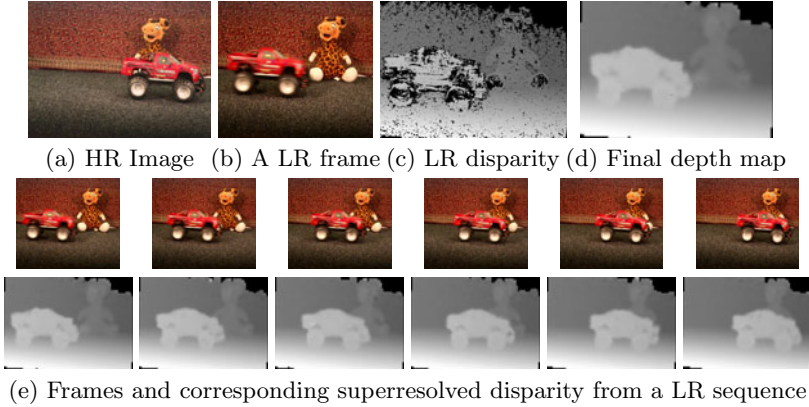


Fig. 6. Super resolution of stereo video

49.62mm which is close to the obtained calibration value of 48.15mm (see Section 4.2). Varying the mirror angle provides control over the effective baseline as well as the vergence of the stereo system.

4 Applications and Results

We demonstrate the system with three applications with different requirements. First, we use the system for depth super resolution of dynamic scenes, which utilizes the high frame rate stereo video capability. Next we apply the calibrated system for the case of metric 3D reconstruction of dynamic and close range scenes. Lastly we show stereo results for different distances ranging from close range (within a feet), medium range (within 5 feet) and far range outdoor scene. We include results for static and dynamic scenes.

4.1 Depth Super Resolution

In this section we discuss the first application of the system for depth super resolution. The resolution of depth map and the time taken to capture is inversely related for most sensors. For example, laser scanners can be used to obtain high

quality and resolution depth maps but are limited to static scenes due to the large scanning time required. For dynamic scenes, methods such as stereo/multi-view or time-of-flight cameras have to be used, which are restricted by the spatial resolution possible for a given frame rate. Many schemes have been thus proposed to enhance the resolution of the depth maps obtained by the above systems, called depth super resolution. In general, the process involves noise filtering, hole filling at missing pixels and enhancing the spatial resolution of the depth map. The depth map is usually upsampled using a higher resolution color image of the scene or using multiple low resolution depth maps taken from slightly different view points. Some of the works using time-of-flight cameras include [22,23,24,25]. Most of these works have been only demonstrated on static scenes. Recently, [24] was extended for dynamic scenes [26]. Many algorithms have been proposed for the case of stereo systems as well [27,28,29]. In [28], the proposed joint bilateral filter was used to upsample both the images and the depth maps. In [29], a hybrid camera setup with low resolution, high frame-rate stereo pair and one high resolution, low frame rate camera was proposed. But the authors demonstrated depth super resolution only for a static case. The DSLR used in our setup is capable of 15MP still image capture at 1fps, during video capture (720 lines at 30fps).

We perform super resolution for a dynamic scene by applying a cross (or joint) bilateral filter with the edge information from the high resolution image. Since that image is only captured at certain instances, we must estimate the motion and warp it to correspond to each stereo frame. The pipeline is shown in Figure 5. Unlike other reported systems, we do not use a separate camera to capture the high resolution (HR) image. The prism setup provides a stereo image even in the still capture. We use one image as the HR image, thus the transform between the low resolution (LR) image and the HR image is a 2D homography. The camera does not use the entire sensor for video capture, since the aspect ratio is different for High-Definition video capture and still image. To obtain the homography H , we use a pre-calibration step with a known pattern in the scene. Using correspondence between the HR image and a LR frame, we estimate the transform H for suitable translation and scaling of the LR frame to HR image. This transformation is fixed given the prism-mirror arrangement and the camera lens zoom, and is thus calculated only once for a given setup. The HR images were captured at random instances for the sequence and are thus non-equally spaced in time. For each LR frame we select the HR image closest in time. We transform the reference stereo frame to the HR image using H and estimate the motion between this pair. Due to the blur caused by the motion of the object and interpolation during the transform, pixel based optic flow schemes do not provide robust results. To maintain high quality edges in the HR image, the warping has to be correct. Since the motion is rigid in our case, we perform motion analysis at the object level. We initially capture an image of the setup without the moving objects. Using this as a reference background image, we obtain the moving objects in the scene. To eliminate errors due to shadows, lighting changes or noise, we use dense SIFT matching instead of simple

Table 1. Calibration results




Errors (mm, mm^2)	Plane pose			
Triangulation error	Mean(Variance)	0.0467 (0.0012)	0.0585 (0.0018)	0.0519 (0.0016)
Plane fit error	Mean(Variance)	0.4076 (0.0965)	0.4324 (0.1891)	0.3429 (0.0866)

image subtraction at this stage. Another feature matching step is used between the objects detected in the LR and HR images to obtain the motion for each individual object. The HR image is then warped to compensate for the object motion. The background image is used to fill the holes in the warped image. The LR stereo pair is rectified using the method proposed in [30] and the initial disparity map is obtained using the Growing Correspondence Seeds method [31]. We then apply joint/cross bilateral filter on the transformed disparity map and the motion compensated HR image. For the disparity map D and HR image I , the joint bilateral filter can be written as

$$D(i) = \frac{1}{W_i} \sum_{j \in N} G_s(\|i - j\|) G_r(\|I(i) - I(j)\|) D'(j)$$

where i, j are i^{th} and j^{th} pixel and N is the predefined neighborhood for a pixel. G_s and G_r are the spatial and range kernel centered at the pixel i and the disparity at i respectively. W_i is the normalization factor defined as, $W_i = \sum_{j \in N} G_s(\|i - j\|) G_r(\|I(i) - I(j)\|)$. The image edge constraint $\|I(i) - I(j)\|$ can be defined in different ways, we use the grayscale value difference at the pixels. Figure 6 shows results for a scene with a toy car moving. Figure 6(a) shows a captured HR image, (b) shows left image of a LR stereo frame, (c) shows the initial disparity map for the image in (b), and (d) the corresponding final super resolved disparity map. Figure 6(e) shows more frames from the sequence, which used HR image in (a) for super resolution (motion compensated as required). Note that we have only showed some intermediate frames from the sequence.

4.2 3D Reconstruction

Here we describe the use of the system for 3D reconstruction which has wide ranging applications in 3D immersive technology [32], face and expression recognition [1,2], action recognition [33], archeology [34]. Many of the above applications demand portability without sacrificing the quality of the reconstruction. Systems with two or more cameras have been proposed for the above, which have been successful in lab setting. In order to employ the algorithms for 3D face analysis, action recognition and others in their real environments such as court rooms or airports, a large change in the camera setup is impractical. Most environments have existing single camera installations in place. Our setup can provide a cost effective and simple mechanism to convert those to stereo. As discussed before, previously proposed single camera stereo systems would require

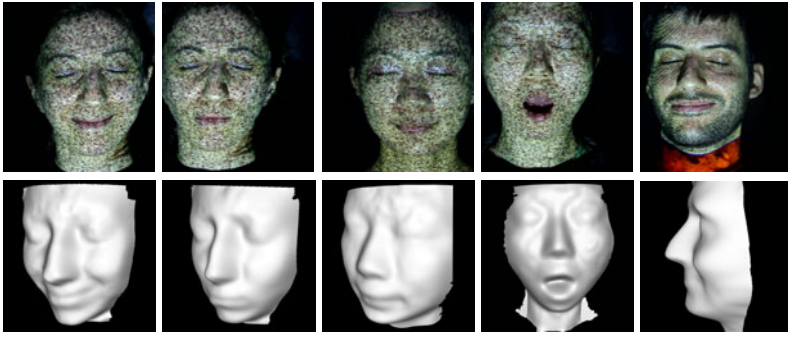


Fig. 7. Snapshots from facial expression video. Top: Reference stereo image, Bottom: Corresponding 3D reconstruction.

change in mounting, cannot be attached easily to the lens of a camera or are have limits on size of object being imaged.

We calibrated our system using a standard stereo calibration technique [35]. Camera parameters were iteratively refined until the reprojection error converged. In our experiments the RMS reprojection error converged to 0.2 pixels. The quality of calibration and stereo is measured using triangulation uncertainty and reconstruction error for a known geometry. The stereo triangulation and plane fitting error for different poses of a plane in the calibrated range is show in Table 1. The triangulation error is the smallest distance between the rays from the two cameras. It must be ideally zero, but due to finite precision in pixel matching, the rays will not intersect. The plane fitting is done using RANSAC such that more than 90% of the points are inliers. All distances are in millimeters(mm). The calibration and plane fitting results clearly show that high accuracy reconstructions can be obtained by calibrating the proposed setup. Some results from face video reconstruction are shown in Figure 7.

4.3 More Results

Disparity maps for other scenes are shown in Figure 8. We obtain approximately 150 levels of disparity for the outdoor tree image and 90 levels for the indoor setup. Figure 9 shows disparity maps for an action sequence, with 100 levels

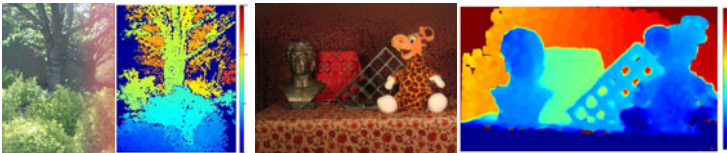


Fig. 8. Disparity maps for general scenes. Left: An outdoor scene and corresponding disparity map, Right: An indoor scene and depth map.

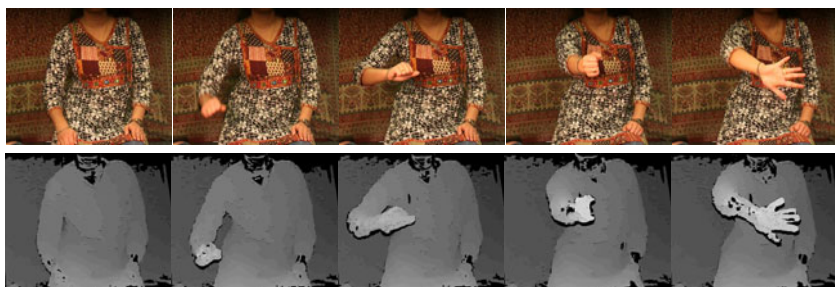


Fig. 9. Disparity maps for an action sequence. Top: Sample captured stereo images, Bottom: Corresponding disparity maps.

obtained between the tip of hand (last image) and background. The quality of results demonstrate that our setup can be used in applications such as autonomous navigation in robots [36], scene understanding [5] and action recognition [33]. For robotic applications, our system can be more effective than using multiple cameras due to limited space and resources on a mobile platform. Our setup also offers flexibility in terms of baseline and vergence, which allows tuning the depth sensitivity to match the application needs. It can also be used in other application such as gesture recognition [37,38] for human-machine interface and 3D immersive technology [32], where the use of a single camera in lieu of multiple cameras greatly simplifies the setup.

5 Conclusion

In this paper, we proposed a setup with prism and mirrors, which allows a single camera to be converted to stereo. The proposed setup was shown to have various key advantages over previous systems. Our setup can be easily attached in front of any camera. Since the prism is placed close to the lens, the portability of the system is not affected. Existing single camera setups can be easily converted to stereo, since the viewing direction of the camera remains unaltered, unlike many of the earlier works. The cost and complexity of construction is much less compared to some commercially available stereo solutions. We demonstrated the use of the system for depth superresolution of stereo video sequences of dynamic scenes. We also showed that high quality 3D reconstructions can be obtained by calibrating the camera. Disparity maps for general indoor and outdoors scenes were provided to show applicability in other areas such as scene understanding, human-machine interface and action recognition.

References

1. Heseltine, T., Pears, N., Austin, J.: Three-dimensional face recognition using combinations of surface feature map subspace components. *Image Vision Comput.* 26, 382–396 (2008)

2. Wang, J., Yin, L., Wei, X., Sun, Y.: 3d facial expression recognition based on primitive surface feature distribution. In: Proc. Conf. Computer Vision and Pattern Recognition, pp. 1399–1406 (2006)
3. Pellegrini, S., Iocchi, L.: Human posture tracking and classification through stereo vision and 3d model matching. *J. Image Video Process.* 2008, 1–12 (2008)
4. Sogo, T., Ishiguro, H., Trivedi, M.M.: Real-time target localization and tracking by n-ocular stereo. In: OMNIVIS 2000: Proceedings of the IEEE Workshop on Omnidirectional Vision, p. 153 (2000)
5. Somanath, G., Rohith, M.V., Metaxas, D., Kambhamettu, C.: D - clutter: Building object model library from unsupervised segmentation of cluttered scenes. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009 (2009)
6. Gluckman, J., Nayar, S.K.: Planar catadioptric stereo: Geometry and calibration. *Computer Vision and Pattern Recognition*. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, p. 1022 (1999)
7. Gluckman, J., Nayar, S.K.: Catadioptric stereo using planar mirrors. *Int. J. Comput. Vision* 44, 65–79 (2001)
8. Chaen, A., Yamazawa, K., Yokoya, N., Takemura, H.: Omnidirectional stereo vision using hyperomni vision. In: Technical Report 96-122, IEICE (1997) (in Japanese)
9. Gluckman, J., Nayar, S.K., Thoresz, K.J.: Real-time omnidirectional and panoramic stereo. In: Proceedings of the 1998 DARPA Image Understanding Workshop, pp. 299–303. Morgan Kaufmann, San Francisco (1998)
10. Nene, S.A., Nayar, S.K.: Stereo with mirrors. In: ICCV 1998: Proceedings of the Sixth International Conference on Computer Vision, Washington, DC, USA, p. 1087. IEEE Computer Society, Los Alamitos (1998)
11. Mitsumoto, H., Tamura, S., Okazaki, K., Kajimi, N., Fukui, Y.: 3-d reconstruction using mirror images based on a plane symmetry recovering method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14, 941–946 (1992)
12. Zhang, Z.Y., Tsui, H.T.: 3d reconstruction from a single view of an object and its image in a plane mirror. In: International Conference on Pattern Recognition, vol. 2, p. 1174 (1998)
13. Kawanishi, T., Yamazawa, K., Iwasa, H., Takemura, H., Yokoya, N.: Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors. *International Conference on Pattern Recognition* 1, 485 (1998)
14. Gluckman, J., Nayar, S.K.: Rectified catadioptric stereo sensors. *Computer Vision and Pattern Recognition*. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, p. 2380 (2000)
15. Gluckman, J., Nayar, S.K.: Rectified catadioptric stereo sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 224–236 (2002)
16. Gao, C., Ahuja, N.: Single camera stereo using planar parallel plate. In: 17th International Conference on Pattern Recognition, ICPR 2004, vol. 4, pp. 108–111 (2004)
17. Nishimoto, Y., Shirai, Y.: A feature-based stereo model using small disparities. In: Proc. Computer Vision and Pattern Recognition, pp. 192–196 (1987)
18. Lee, D.H., Kweon, I.S., Cipolla, R.: A biprism-stereo camera system. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, p. 1082 (1999)
19. Xiao, Y., Lim, K.B.: A prism-based single-lens stereovision system: From trinocular to multi-ocular. *Image Vision Computing* 25, 1725–1736 (2007)
20. Duviolbourg, L., Ambellouis, S., Cabestaing, F.: Single-camera stereovision setup with orientable optical axes. In: International Conference Computer Vision and Graphics (ICCVG), 173–178 (2004)

21. Lu, T., Chao, T.H.: A single-camera system captures high-resolution 3d images in one shot. *SPIE Newsroom* (2006), doi: 10.1117/2.1200611.0303
22. Kuhnert, K.D., Stommel, M.: Fusion of stereo-camera and pmd-camera data for real-time suited precise 3d environment reconstruction. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4780–4785 (2006)
23. Yang, Q., Yang, R., Davis, J., Nister, D.: Spatial-depth super resolution for range images. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2007)
24. Zhu, J., Wang, L., Yang, R., Davis, J.: Fusion of time-of-flight depth and stereo for high accuracy depth maps. In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008*, pp. 1–8 (2008)
25. Schuon, S., Theobalt, C., Davis, J., Thrun, S.: Lidarboost: Depth superresolution for tof 3d shape scanning. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 343–350 (2009)
26. Zhu, J., Wang, L., Gao, J., Yang, R.: Spatial-temporal fusion for high accuracy depth maps using dynamic mrfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 899–909 (2010)
27. Sawhney, H.S., Guo, Y., Hanna, K., Kumar, R., Adkins, S., Zhou, S.: Hybrid stereo camera: an ibr approach for synthesis of very high resolution stereoscopic image sequences. In: *SIGGRAPH 2001: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques* (2001)
28. Kopf, J., Cohen, M.F., Lischinski, D., Uyttendaele, M.: Joint bilateral upsampling. In: *SIGGRAPH 2007: ACM SIGGRAPH 2007 Papers*, p. 96 (2007)
29. Li, F., Yu, J., Chai, J.: A hybrid camera for motion deblurring and depth map super-resolution. In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008*, pp. 1–8 (2008)
30. Fusiello, A., Irsara, L.: Quasi-euclidean uncalibrated epipolar rectification. In: *International Conference on Pattern Recognition, ICPR* (2008)
31. Čech, J., Šára, R.: Efficient sampling of disparity space for fast and accurate matching. In: *BenCOS 2007: CVPR Workshop Towards Benchmarking Automated Calibration, Orientation and Surface Reconstruction from Images*, IEEE, Los Alamitos (2007)
32. Towles, H., Chen, W.C., Yang, R., Kum, S.U., Fuchs, H., Kelshikar, N., Mulligan, J., Daniilidis, K., Holden, L., Seleznik, B., Sadagic, A., Lanier, J.: 3d tele-collaboration over internet2. In: *International Workshop on Immersive Telepresence, ITP* (2002)
33. Roh, M.C., Shin, H.K., Lee, S.W.: View-independent human action recognition with volume motion template on single stereo camera. *Pattern Recognition Letters* 31, 639–647 (2010)
34. Bitelli, G., Girelli, V.A., Remondino, F., Vittuari, L.: The potential of 3d techniques for cultural heritage object documentation. *Videometrics IX* 6491 (2007)
35. Strobl, K.H., Sepp, W., Fuchs, S., Paredes, C., Arbter, K.: DLR CalDe and DLR CalLab
36. Murray, D., Little, J.J.: Using real-time stereo vision for mobile robot navigation. *Autonomous Robots* 8, 161–171 (2000)
37. Agarwal, A., Izadi, S., Chandraker, M., Blake, A.: High precision multi-touch sensing on surfaces using overhead cameras. In: *International Workshop on Horizontal Interactive Human-Computer Systems*, pp. 197–200 (2007)
38. Shimizu, M., Yoshizuka, T., Miyamoto, H.: A gesture recognition system using stereo vision and arm model fitting. *International Congress Series*, vol. 1301, pp. 89–92 (2007)