Camera Calibration and Shape Recovery from videos of Two Mirrors

Quanxin Chen and Hui Zhang

Dept. of Computer Science, United International College, 28, Jinfeng Road, Tangjiawan, Zhuhai, Guangdong, China.

Abstract. This paper addresses the problem of motion and shape recovery from a two-mirror system which is able to generate five views of an object. Different from existing methods, this paper uses a short video instead of static snapshots so that it can help with action recognition once the 3D visual hull model is reconstructed. In order to solve the problem, this paper shows the geometry relationship between the two-mirror system and circular motion, so that the two-mirror system can be solved as circular motions. Different from the approach of Zhang et al.[22], we avoid using the vanishing point of X-axis which would cause accumulate error when calculating the epipoles of two views. Results of comparative experiments and the 3D visual hull of model show the feasibility and the accuracy of the proposed approach.

Keywords: video, motion and shape recovery, camera calibration, two-mirror system, circular motion

1 Introduction

In recent years, videos have been studied to recover the shape and motion of models in different aspects. Guillemaut et al.[5] proposed an approach based on view-dependent segmentation and reconstruction which is used in challenging outdoor environments with moving cameras. Plänkers et al.[14] developed an interesting approach for motion and shape recovery of articulated deformable objects from silhouettes by using trinocular video sequence of complex 3D motions. Similarly, we also use articulated deformable objects but using normal video camera and only analyze single frame alone. Moreover, self-calibration is another important aspect in video study. Camera Calibration in video can be used as a pre-processing step for many later video analysis and event recognition applications [10][17][18].Lv et al.[12] calibrated the video camera from three vanishing points in the image. This paper also calibrates the camera but in a more flexible way. Sinha et al.[19] calibrated the video camera from dynamic silhouettes by camera network. However, this approach should keep synchronizing for every video camera.

A system consisting of refracting (lens) and reflecting (mirror) elements are called catadioptric systems [8]. This system is used to be studied because it provides symmetric relationships for object and its reflections which enable us to use

only a single camera. Therefore, we do not need to consider camera synchronization any more. Martin et al.[13] studied camera calibration and 3D reconstruction by one mirror. However, one-mirror system provide only two views for object which is not enough for 3D reconstruction. Gluckman et al.[4] exploited the geometry of the two-mirror system and recover the focal length via corresponding points of two views. Different from [4], Forbes et. al.[3] introduced an approach based on silhouettes alone which avoid exacting corresponding points of views in image. However, both this two papers should figure out the reflection matrix of mirror which is in a traditional way but difficult for every situation. Zhang et al.[22] and Ying et al.[20] recovered the geometry relationship between the two-mirror system and the circular motion which makes the calibration easier.

Inspired by [3][20][22], this paper aims to recover the motion and shape of object from video other than static snapshots with two-mirror system because the reconstructed 3D model can help to boost the accuracy of action recognition and avoid processing large dataset. A novel approach which is cross-view action recognition without body part tracking was proposed in [6]. However, it still required a large collection of mocap data. Lu et al.[11] proposed binary range-sample feature in depth for action recognition which is robust to occlusion and data corruption. In this paper, two equal but unknown angles could be figured out from the two-mirror system so that they can be used to recover the image of circular points [9]. The rotation axis of circular motion is initialized from the intersections of epipolar bitangent lines of two views. Once the camera is calibrated, the projection matrices of views can be obtained. Therefore, the 3D visual hull model is recovered.

This paper is organized as follows. §2 gives the notation and geometry of the two-mirror system and the relationship between it and circular motion. This section also presents several image invariants which play important roles in our approach. §3 and §4 present the details of the camera calibration and motion recovery. §5 and §6 show the experimental results and the conclusion of this approach.

2 Notation and Geometry

2.1 Geometry of Two Mirror Setup

Considering a two-mirror setup for picturing five views of an object in a single image (see Fig. 1(a)). Note there is a real object O and its four mirror reflections O_1, O_2, O_{12} and O_{21}. The virtual object O_1 is the reflection of O in Mirror M_1; O_2 is the reflection of O in Mirror M_2; O_{12} is the reflection of O_2 in Mirror M_1; and O_{21} is the reflection of O_1 in Mirror M_2. Note there are two virtual mirrors M_{12} and M_{21} which reflect O_1 to O_{12} and O_2 to O_{21}, respectively. There are also several virtual cameras C_1, C_2, C_{21}, C_{12} can be similarly formed as the way of the virtual objects.

According to the lines joining corresponding points between O and O_1 are perpendicular to the mirror M_1 and those between O and O_2 are perpendicular

to the mirror M_2, we can easily find that the three objects O, O_1 and O_2 lie on a plane Π which is perpendicular to both the mirror plane M_1 and M_2, and the intersection line of the mirror M1 and M_2. Similarly, O_{21} and O_{12} also lie on the plane Π. Therefore, the two virtual mirrors M_{21} and M_{12} are both perpendicular to the plane Π. Similarly, the five camera centers should all lie on a plane parallel to the plane Π. More importantly, the plane Π is generally not the desk plane where the real object O lying on.

Let's denote the vanishing line of Π as l_h (see Fig. 1(b)), then we will show how to get l_h according to vanishing points of the directions perpendicular to

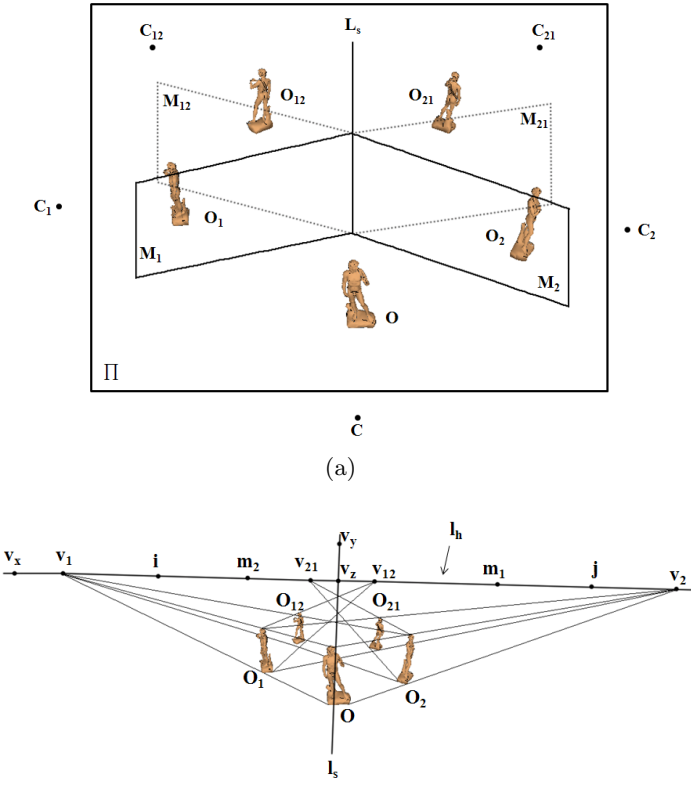

Fig. 1. Geometry of the two mirrors. (a) Two mirror setup. (b) The image invariants of the two mirror setup.

all four mirror planes. For the mirror M_1, four outer bitangents (two between the object O and O_1, two between the object O_2 and O_{12}) intersect at the

vanishing point v_1, which is their vanishing point on the l_h. Similarly, for the mirror M_2, we can get vanishing point v_2 which the four outer bitangents (two between the object O and O_2, two between the object O_1 and O_{21}) intersect at; for the mirror M_{12}, the two outer bitangents between the object O_1 and O_{12} intersect with the l_h at the vanishing point v_3; for the mirror M_{21}, the two outer bitangents between the object O_2 and O_{21} intersect with the l_h at the vanishing point v_4. Hence the horizon l_h passes through vanishing points v_1, v_2, v_{12}, v_{21}.

Moreover, the plane Π intersects mirror M_1 and M_2 at two lines which intersect with vanishing line l_h at point m_1 and m_2 respectively(see fig. 1(b)).

2.2 Relationship between the Two Mirror System and the Turntable Motion

Considering the top view of the real and virtual objects on the plane Π and let the real camera C lying on the positive Z-axis of the world coordinate system (see Fig. 2). Let the mirrors be M_1, M_2, M_{12} and M_{21}, and Y-axis intersects Π at point O. Then the real camera C has its reflections C_1, C_2, C_{12} and C_{21} with respect to the M_1, M_2, M_{12} and M_{21} respectively. Let's denote the angle between C and the mirror M_1 as α, denote the angle between C and the mirror M_2 as β. Hence, the angle between M_1 and M_2 is $\theta = \alpha + \beta$. Since all the views are generated by mirrors, it can be easily found that the length $|OC_1| = |OC_2| = |OC_{12}| = |OC_{21}| = |OC|$ and the angle $\angle COC_1 = \angle C_2OC_{21} = 2\alpha$, $\angle COC_2 = \angle C_1OC_{12} = 2\beta$. According to the symmetric relationship of mirrors, we can easily divide the five views into two turntable motions based on the image direction. By thinking of this, we can find that C_{21}, C and C_{12} are in similar direction, and C_2 and C_1 are in another similar direction. Hence, we can get two turntable motions: one is C_{21} rotates to C about the Y-axis with angle of 2θ and then rotates to C_{12} with the same angle of 2θ. Another one is the C_1 rotates to C_2 with the angle of 2θ. Note it can be observed from the Figure 2 that the angles:

$$\angle CC_1C_{12} = \angle C_1CC_2 = \angle CC_2C_{21} = \pi - \theta \qquad (1)$$

Turntable motion is always regarded as a scene and a camera rotating about a fixed axis. The invariants related to the geometry of turntable motion have been well established in [1] (see Fig. 1(b)). l_h is the vanishing line of the turntable plane. l_s is both the image of rotation axis and the vanishing line of the Y-axis. v_z is the vanishing point of the Z-axis. v_x is the vanishing point of the X-axis and v_y is the vanishing point of the Y-axis. v_x lies on l_h, v_y lies on l_s, and v_z is the intersection point of l_h and l_s. Note v_x and is form a pole-polar relationship with respect to the image of the absolute conic [2]. The pair of imaged circular points i and j for the turntable plane Π also lie on l_h. i.e., $l_h = i \times j$. All the aforementioned image entities will be fixed if the camera intrinsics are invariants.

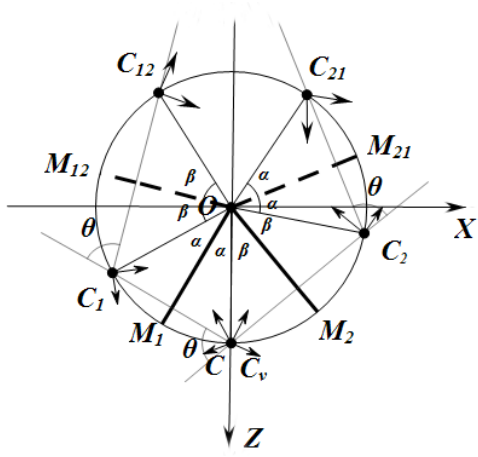

Fig. 2. Top view of the two-mirror setup.

3 Camera Calibration

3.1 Recovery of the Imaged Circular Points and the Angle between Two Mirrors

In order to rectify the five-view image to affine, we should construct a "pure projective" transformation matrix P [9] by use the vanishing line $l_h = [l_1\ l_2\ l_3]^T$.

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{bmatrix} \tag{2}$$

Denotes the circular points on the affine space as $l_h = [\alpha \mp \beta i\ 1\ 0]^T$. The five-view image can be further rectified from affine to metric by another transformation matrix A [9].

$$A = \begin{bmatrix} \frac{1}{\beta} & -\frac{\alpha}{\beta} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

From section 3.1, we can know there are equal unknown angles one the world plane and according to [9]. We can find α and β both lie on a circle with center(c_α, c_β).

$$(c_\alpha, c_\beta) = (\frac{a_1b_2 - a_2b_1}{a_1 - b_1 - a_2 + b_2}, 0) \tag{4}$$

and squared radius

$$r^2 = (\frac{a_1b_2 - a_2b_1}{a_1 - b_1 - a_2 + b_2})^2 + \frac{(a_1b_1)(a_1b_1 - a_2b_2)}{a_1 - b_1 - a_2 + b_2} - a_1b_1 \tag{5}$$

where the angle between two lines imaged with directions a_1, b_1 is the same as that between lines imaged with directions a_2, b_2. Hence the circular points are imaged on the vanishing line at $i, j = [\,(\alpha \mp \beta i)l_3 \; l_3 \; -\alpha l_1 - l_2 \pm \beta l_1 i\,]^T$.

The angle θ can be directly obtained from the rectified metric image by Laguerre Formula[16]:

$$\theta = \frac{1}{2j} log\{l_1, l_2; i, j\} \tag{6}$$

3.2 Recovery of the imaged rotation axis and vanishing point v_x

In order to initialize l_s accurately, [20] propose an effective method to guess l_s. Inspired by it, we induce a formula to compute epipoles for two views under circular motion. Let's denote e_{v_1} and e_{v_2} as the epipoles for $\{C, C_{12}\}$ and $\{C, C_{21}\}$. From section 3.1, we know that the angle between C and C_{12} is 2θ, so is C and C_{21}. It is easy to get that the angle between OC and CC_{21} is $\frac{\pi}{2} - \theta$, and the angle between OC and CC_{12} is $-\frac{\pi}{2} + \theta$(see Fig. 3(a)). From (7), we can get

$$\frac{\pi}{2} - \theta = \frac{1}{2j} log\{v_z, e_{v_2}; i, j\}, \tag{7}$$

and

$$-\frac{\pi}{2} + \theta = \frac{1}{2j} log\{v_z, e_{v_1}; i, j\} \tag{8}$$

Similarly, denote $e_{v_{12}}$ and $e_{v_{21}}$ as the epipoles for $\{C_{21}, C_{12}\}$(see Fig. 3(b)). We can get

$$\frac{\pi}{2} - 2\theta = \frac{1}{2j} log\{v_z, e_{v_{21}}; i, j\} \tag{9}$$

and

$$-\frac{\pi}{2} + 2\theta = \frac{1}{2j} log\{v_z, e_{v_{12}}; i, j\} \tag{10}$$

Since v_1 and v_2 are the vanishing points of tangent lines which are perpendicular to mirrors. Obviously, we can get

$$\frac{\pi}{2} = \frac{1}{2j} log\{v_1, m_1; i, j\} \tag{11}$$

and

$$\frac{\pi}{2} = \frac{1}{2j} log\{v_2, m_2; i, j\} \tag{12}$$

From (11) (12), we can easily get m_1 and m_2, so that we can search for v_z by using Golden Section search[15] between m_1 and m_2.

Therefore, the epipoles e_{v_1} and e_{v_2} for $\{C, C_{12}\}$, $\{C, C_{21}\}$ and $\{C_1, C_2\}$; and the epipoles $e_{v_{12}}$ and $e_{v_{21}}$ for $\{C_{21}, C_{12}\}$ can be recovered by (9) (10) (11) (12). Draw bitangent lines on corresponding silhouettes from this epipoles, so that we can get 8 intersection points. Adding v_z, a hypothesis $\overline{l_s}$ is fitted by 9 points.

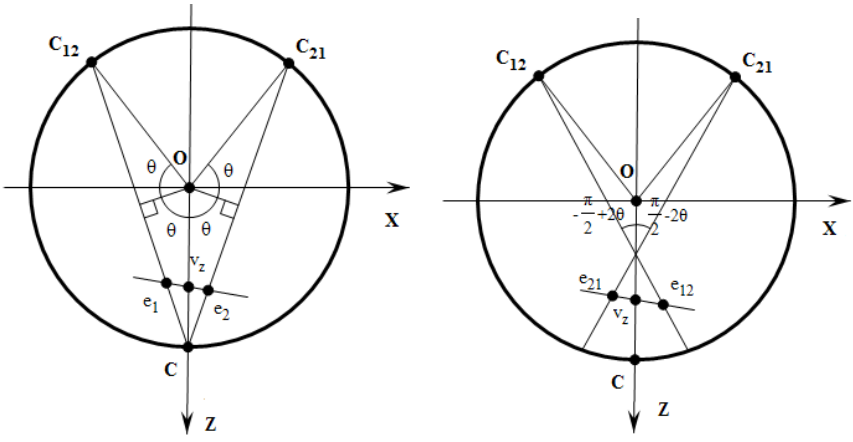

Fig. 3. Epipoles of two views in circular motion. (a) Rotation with 2θ. (b) Rotation with 4θ.

Using the sum of the distances between the 9 points and $\overline{l_s}$ to be the cost for Golden Section method can obtain a better rotation axis l_s.

After the initialized imaged rotation axis l_s is recovered, then the vanishing point v_z can be obtained as the intersection between l_s and l_h. The vanishing point v_x can also be easily recovered from the cross ratio $\{i, j; v_x, v_z\} = -1$.

And the optimization of the imaged rotation axis l_s may be carried out as two dimensional optimization problem by minimizing the distance from the transformed line $l_i' = W^{-T} l_i$ to the profile in the second image(see fig. 4). Where W is harmonic homology:

$$W = I - 2\frac{v_x l_s^T}{v_x^T l_s} \qquad (13)$$

l_s and v_x can be further optimized by minimizing the transformation error of the silhouette brought about by the harmonic homology W.

3.3 Recovery of the intrinsics

From section 3.2, the obtained imaged circular points can be used to find the camera intrinsics since they lie on the image of the absolute conic (IAC)ω . Besides, the imaged rotation axis l_s and the vanishing point v_x define a pole-polar relationship respect to ω [7]. It can then be estimated from the following constraints:

$$\begin{cases} i^T \omega i = 0 \quad \text{and} \quad j^T \omega j = 0 \\ l_s = \omega v_x \end{cases} \qquad (14)$$

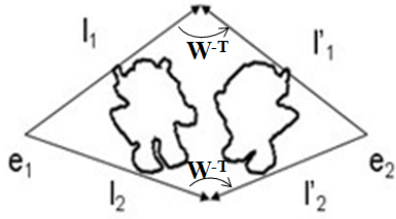

Fig. 4. The overlapping of two silhouettes and their epipolar tangents. l_1, l_1', l_2 and l_2' are the outer epipolar tangent lines.

At last, the intrinsics K can be obtained by Chloesky decomposition of IAC ω ,

$$\omega^{-1} \sim KK^T \tag{15}$$

4 Shape Recovery

In mathematical expression, the camera projection matrix P can be written as:

$$P = KR[R_y(\theta)| - T], \tag{16}$$

where K is camera intrinsic parameter, R is the camera pose, $R_y(\theta)$ is the rotation matrix about rotation axis with angle θ and $T = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$.
In order to unify the two different directions of turntable motions, let's image that there is a virtual mirror which passes through O and C. Let C_v be the reflection of C_1 according to the virtual mirror(see fig. 2). The camera projection matrices for C and C_v can be written as:

$$P_C = KR[R_y(\theta)\Sigma| - T], \tag{17}$$

$$P_{C_v} = KR[R_y(\theta)| - T] \tag{18}$$

where $\Sigma = diag(\begin{bmatrix} -1 & 1 & 1 \end{bmatrix})$. The projection matrices for C_{21} and C_{12} can be derive easily in a similar way. So far, we have obtained all the parameters except camera pose. Zhang et al. [21] introduced a traditional method to obtain camera pose(see fig. 5).
According to this method, we firstly rotate the camera respect to Y-axis of so that the principal point x_0 can lie on rotation axis l_s. Secondly, we rotate the camera respect to Z-axis to enable rotation axis l_s to be perpendicular. At last, we rotate the camera respect to X-axis to enable the principal point x_0 to be the intersection of vanishing line l_h and rotation axis l_s.
Finally, we can use the projection matrix P to get a visual hull of a model.

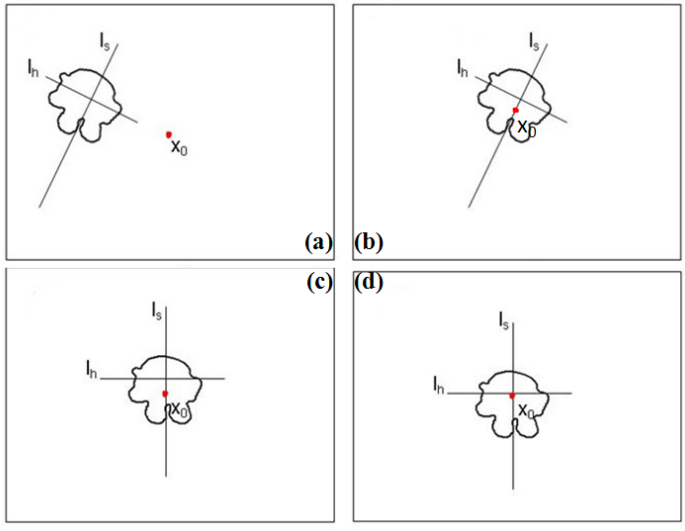

Fig. 5. Three steps of rectification. (a)Before rectification. (b)After rotated the camera respect to Y-axis. (c)After rotated the camera respect to Z-axis. (d)After rotated the camera respect to X-axis.

5 Experimental Result

In this part, we show three experimental results of applying the approach to reconstruct 3D object models from three frames of video which contain three different motions of robot. The video had a resolution of 640 × 425. The natural camera was assumed to be zero-skew and unit aspect ratio when calibrated. The angle between the mirrors was computed to be 72.2 degrees. Table 1 shows the differences of results between our method and Zhang's checkerboard calibration method[23]. It can be told that focal length f and the u_0 coordinate of the principal point were both estimated with around 5% error compared to Zhang's method while v_0 was larger. It results from the accumulative error in the estimated v_x. Figure 6 shows the frames of video we used and the visual hull model of robot.

Table 1. Comparative results of the intrinsics between Zhang's method and our method

	f	u_0	v_0
Zhang's method	507.85	319.5	212
Proposed method	488.21	301.40	230.92
Error to Zhang's method	3.87%	5.67%	8.92%

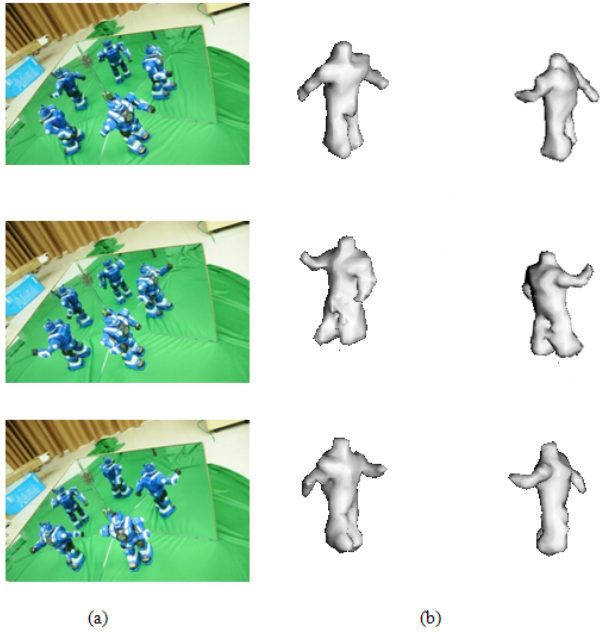

Fig. 6. (a) Three frames from video. (b) Visual hull models for corresponding frame.

6 Conclusion

In this paper, we have introduced a novel approach which transforms five views generated by two mirrors to turntable motion. Difference from previous silhouette-based methods, the approach requires a short video instead of a static snapshot. This can allow us to build a moving 3D model based on the motions of object. We used a simple formulation of the imaged circular points to obtain the image of absolute conic. A closed-form solution for the image invariants, the mirror angle and the camera intrinsics is obtained. After calibration, a visual hull of the objet can be obtained from the silhouettes. Experimental results demonstrated the performance of our approach.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (Project no. 61005038) and an internal funding from United International College (Project no. R201312).

References

1. A. W. Fitzgibbon, G.C., A.Zisserman: Automatic 3d model construction for turntable sequences. In: Koch, R., Gool, V.L. (eds.) 3D Structure from Multiple Im-

ages of Large-Scale Environments, pp. 155–170. LNCS 1506, Springer Verlag (June 1998)
2. Faugeras, O., Luong, Q., , Maybank, S.: Camera self-calibration: Theory and experiments. In: Proc. 2nd European Conf. on Computer Vision. pp. 321–334 (1992)
3. Forbes, K., Nicolls, F., G.D., J., Voigt, A.: Shape-from-silhouette with two mirrors and an uncalibrated camera. In: Proc. 9th European Conf. on Computer Vision. pp. 165–178 (2006)
4. Gluckman, J., Nayar, S.K.: Planar catadioptric stereo: Geometry and calibration. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 1 (1999)
5. Guillemaut, J.Y., Kilner, J., Hilton, A.: Robust graph-cut scene segmentation and reconstruction for free-viewpoint video of complex dynamic scenes. In: 12th International Conference on Computer Vision. pp. 809–816 (2009)
6. Gupta, A., Martinez, J., Little, J.J., Woodham, R.J.: 3d pose from motion for cross-view action recognition via non-linear circulant temporal encoding. In: Proc. Conf. Computer Vision and Pattern Recognition (2014)
7. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, UK (2000)
8. Hecht, E., Zajac, A.: Optics. Addison-Wesley (1974)
9. Liebowitz, D., Zisserman, A.: Metric rectification from perspective images of planes. In: Proc. Conf. Computer Vision and Pattern Recognition. pp. 482–488 (1998)
10. Liu, L., Shao, L., Rockett, P.: Boosted key-frame selection and correlated pyramidal motion-feature representation for human action recognition. Pattern Recognition 46(7), 1810–1818 (2013)
11. Lu, C., Jia, J., Tang, C.K.: Range-sample depth feature for action recognition. In: Proc. Conf. Computer Vision and Pattern Recognition (2014)
12. Lv, F., Zhao, T.: Self-calibration of a camera from video of a walking human. In: 16th International Conference on Pattern Recognition. vol. 1, pp. 562–567 (2002)
13. Martins, N., Dias, J.: Camera calibration using reflections in planar mirrors and object reconstruction using volume carving method. The Imaging Science Journal 52, 117–130 (2004)
14. Plänkers, R., Fua, P.: Articulated soft objects for multi-view shape and motion capture. In: Pattern Analysis and Machine Intelligence, IEEE Transactions on 25(9). pp. 1182–1187 (2003)
15. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes 3rd Edition: The Art of Scientific Computing. New York: Cambridge University Press (2007)
16. semple, J., Kneebone, G.: Algebraic Projective Geometry. Oxford University Press, USA (1952)
17. Shao, L., Jones, S., Li, X.: Efficient search and localization of human actions in video databases. IEEE Transactions on Circuits and Systems for Video Technology 24(3), 504–512 (2014)
18. Shao, L., Zhen, X., Tao, D., Li, X.: Spatio-temporal laplacian pyramid coding for action recognition. IEEE Transactions on Cybernetics 44(6), 817–827 (2014)
19. Sinha, S.N., Pollefeys, M., McMillan, L.: Camera network calibration from dynamic silhouettes. In: CVPR (1). pp. 195–202 (2004)
20. Ying, X., Peng, K., Y.B. Hou, S.G., J. Kong, H.Z.: Self-Calibration of Catadioptric Camera with Two Planar Mirrors from Silhouettes. IEEE Trans. on Pattern Analysis and Machine Intelligence (2012)

21. Zhang, H., Mendonça, P.R.S., Wong, K.Y.K.: Reconstruction of surface of revolution from multiple uncalibrated views: A bundle-adjustment approach. In: Proc. Asian Conference on Computer Vision (ACCV2004). pp. 378–383. Jeju Island, Korea (January 2004)
22. Zhang, H., Shao, L., Wong, K.Y.K.: Self-calibration and motion recovery from silhouettes with two mirrors. In: ACCV (4)'12. pp. 1–12 (2012)
23. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 22(11), 1330–1334 (2000)