Computer Vision and Image Understanding 114 (2010) 168–178

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

A hemispherical imaging camera

Chunyu Gao [a,*], Hong Hua [b], Narendra Ahuja [a]

[a] Beckman Institute, Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[b] College of Optical Sciences, University of Arizona, Tucson, AZ 85721, USA

ARTICLE INFO

Article history:
Received 11 March 2008
Accepted 2 March 2009
Available online 14 March 2009

Keywords:
Panoramic vision
Omni-directional vision
Catadioptric camera
Single viewpoint imaging system
Image mosaicing

ABSTRACT

Most of the existing panoramic cameras have cylindrical field of view (FOV) which has a 360° width in azimuth, but a limited height in elevation. However, a hemispherical (or near spherical) FOV is highly desirable in many applications such as robot navigation and surveillance. This paper presents a planar-mirror-based panoramic camera (HEM) which is capable of acquiring nearly hemispherical FOV panoramic images at high and substantially uniform resolution, from a single viewpoint, at video rates. The camera consists of multiple imaging sensors and a hexagonal prism made of planar mirror faces. The sensors are positioned in such a way that they image different parts of the scene from the same virtual viewpoint, either directly or after reflections off the prism. A panoramic image is constructed by concatenating the images taken by different sensors. The proposed system is designed such that the resolution across entire FOV has the same level of uniformity as delivered by a conventional, non-panoramic camera over a small FOV, and the sensor area utilization is maximized. We also present a novel camera placement technique that helps co-locate the optic points of all seven sensors at a single viewpoint, and a method to compensate for artifacts near mirror boundaries caused by finite size of lens aperture. An implementation of the proposed system as well as resulting hemispherical panoramic images are included.

1. Introduction

A panoramic camera is an imaging device capable of capturing a very large field of view (FOV). It is useful for a wide range of applications including tele-conferencing, surveillance and robot navigation [2,4,15]. Like any other cameras, it is desirable that a panoramic camera acquires the entire FOV from a single viewpoint, in real time, at high and uniform resolution across the FOV, with large dynamic range, and over a large depth of field. Many efforts have been made to achieve various subsets of these properties [1,3,4,6,8,9,13,14,17–21,28]. These methods of capturing panoramic or omni-directional images fall into two categories: dioptric methods, where only refractive elements (lenses) are employed, and catadioptric methods, where a combination of reflective and refractive components is used.

Typical dioptric systems include camera clusters, panning cameras, and fisheye lenses. A camera cluster, which consists of multiple conventional cameras pointing in different directions, is capable of acquiring a spherical panorama with uniform resolution at video rate [2,26]. However, these cameras, being physically separated, do not share a common viewpoint, which makes it impossible to create a seamless panoramic mosaic for scenes with arbitrary depth profiles. A fisheye lens is usually designed for an ordinary camera to capture an extremely large FOV [16,27]. The resulting images, however, suffer from severe spatial compression in certain parts since a very large FOV is mapped onto a single sensor, which results in a corresponding loss of resolution. Furthermore, a fisheye lens is not a single viewpoint imaging system, rather, different parts of the images are associated with different viewpoints that fall along a cusped curve [16]. Panning cameras capture panoramic images from a single viewpoint by rotating a conventional camera around its optical center. They deliver a high-resolution panorama over a wide FOV, as well as omni-focus (i.e., large depth of field) when used in conjunction with non-frontal imaging [13,14]. The time it takes to pan across the width of the FOV determines the panorama acquisition rate, which depends on the panning speed and is usually not real time.

Catadioptric methods include curved mirror systems where a conventional camera captures the scene reflected off a single non-planar mirror (e.g., parabolic or hyperbolic mirror) [3,6,8,11,12,20,21,28], and planar mirror systems such as mirror-pyramid systems where multiple conventional cameras image the scene reflected off the faces of a mirror pyramid [9,12,15,17–19,25]. The cameras that use a parabolic or a hyperbolic mirror to map an omni-directional view onto a single sensor are able to capture a large FOV from a single viewpoint at video rate. However, the FOV shape achieved is ring-like hemispherical minus a central cone due to self-occlusion. The overall resolution of the acquired

* Corresponding author.
E-mail addresses: cgao@uiuc.edu (C. Gao), hhua@optics.arizona.edu (H. Hua), n-ahuja@uiuc.edu (N. Ahuja).

1077-3142/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2009.03.005

images is limited to that of the sensor used, and further it varies with the viewing direction across the ring-like FOV, e.g., from a high just outside the central blind spot to a low in the periphery. The cameras using a spherical or conical mirror have similar properties as those using parabolic or hyperbolic mirrors except that they do not possess a single viewpoint. Srinivasan proposed a class of mirrors that provide uniform resolution in both azimuth and elevation direction. However, the imaging systems based on these mirrors do not possess a single viewpoint and suffer low resolution similar to other curved mirror system [23].

Planar-mirror-based camera systems are capable of capturing high-resolution, wide-FOV panoramas from a single viewpoint at video rate [9,10,15,17–19]. The first planar-mirror-based camera system was proposed by Nalwa [17]. In his design, he uses a four-sided right mirror-pyramid, in which the angle formed between a mirror face and the pyramid base is 45°, to virtually collocate the cameras on the pyramid center axis. Each camera is oriented toward a pyramid face with its optical axis parallel to the pyramid axis, and thus captures the part of the scene reflected off its associated face. Together, all the cameras capture contiguous pictures of a 360° wide panoramic FOV whose height is the same as that of a single camera's FOV. Hua and Ahuja extended Nalwa's design to a dual-mirror-pyramid system, composed of two mirror pyramids with a common base [9]. They use two layers of cameras and all cameras are located such that their virtual optical centers coincide at the center of the common base of the pyramids. These cameras point toward their associated pyramid faces and are oriented such that the FOV of the two layers are contiguous. This modification doubles the vertical FOV while preserving the single viewpoint property. Since the camera viewing directions are not parallel to the pyramid axis, the pyramid face is imaged as a trapezoid on the sensor due to perspective (keystone) distortion. Nalwa also reported an alternative design to achieve a nearly spherical shape panoramic system by placing an additional camera at the location of the virtual optical center looking toward the pyramid base along the pyramid axis [19]. To mosaic the image captured by the additional camera with those by peripheral cameras, this design requires a wide-angle fisheye lens, having more than 130° FOV in both horizontal and vertical directions to cover a disproportionately large FOV around the axis. Therefore, the additional camera will be subject to low and non-uniform resolution as well as the multiple viewpoints problem of fisheye lenses and the center portion of the resulting panoramic image will have much lower resolution than the peripheral parts. These problems result in an overall system that possesses neither a single viewpoint nor uniform resolution.

Majority of the aforementioned systems provide a cylindrical FOV which is 360° in width and has a limited height (Fig. 1a). However, many applications, such as robotic navigation and surveillance, are better served by a seamless hemispherical (or near spherical) FOV (Fig. 1b). Consider security cameras as an example. As illustrated in Fig. 1c, the surveillance camera in a large lobby or airport terminal is often mounted on the ceiling to monitor a large area. It is often desired that one single camera covers the entire space of interest without blind spots. The traditional pan-tilt camera can only monitor a small area at a time such as the entrance of the building. The cylindrical-FOV type camera has a blind spot in the central FOV. The hemispherical FOV type camera is superior for such wide area surveillance. Among the reviewed past approaches, only three—camera cluster, the fisheye lens and Nalwa's design with a central camera—are capable of capturing such FOV. However, these have limitations of low resolution, non-uniform resolution or non-single viewpoint. In this paper, we propose a novel panoramic camera using the combination of a hexagonal prism and a cluster of conventional cameras, which we will also refer to as component imagers. The proposed design is capable of acquiring nearly hemispherical panoramic images at video rate, with high and substantially uniform resolution, from a single viewpoint. Here, by substantially uniform resolution we mean the same level of uniformity as that delivered by a conventional, non-panoramic camera, and for simplicity, will refer to it as "uniform resolution." Our camera prototype is shown in Fig. 2.

2. Design of the hemispherical camera and optimal parameter selection

The main design task for planar-mirror-based panoramic cameras is to properly choose mirror shape and camera parameters for the desired design goal. In [10], Hua et al. proposed an optimal

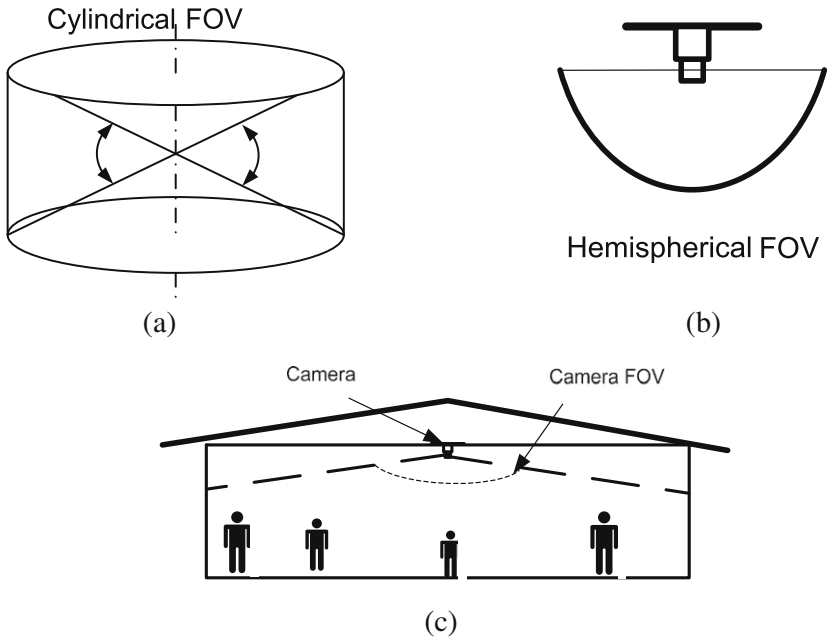


Fig. 1. Two types of FOV shape: (a) cylindrical, (b) hemispherical. (c) The camera with hemispherical FOV for large area surveillance application.

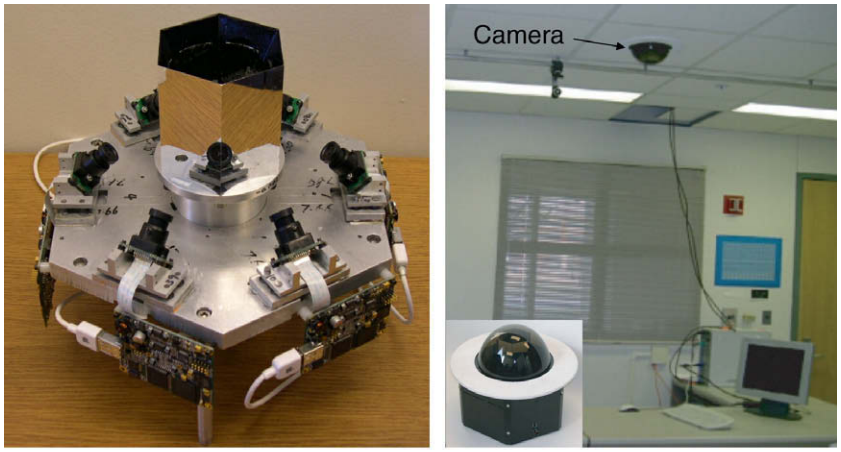


Fig. 2. (a) A prototype of the proposed panoramic camera (HEM). (b) The camera on the ceiling.

approach to designing a dual-mirror-pyramid (DMP) panoramic camera by maximizing the camera resolution uniformity, and sensor utilization, while preserving single view point. In this paper, we extend their approach. Since our goal is to construct a hemispherical FOV, instead of the cylindrical FOV for the DMP system, the actual design constraints and their formulations are quite different. For example, in the DMP system design, the virtual viewpoint offset is considered to be an important variable which allows two layers of cameras sharing the same viewpoint within the pyramid, while the virtual viewpoint offset is irrelevant for us here. We formulate constraints for properties such as the sensor utilization, the mirror shape, and the sensor resolution uniformity consistent with our design goal here.

2.1. Camera co-location and uniform resolution constraint

Constructing a hemispherical FOV with uniform resolution from multiple narrow-FOV conventional imagers is similar to the problem of tessellating the surface of a hemisphere with cells, each corresponding to the FOV of a single imager. Adjacency of the cells would ensure that the FOV's of the imagers are contiguous. Since real imagers are usually rectangular, the problem more specifically amounts to juxtaposing multiple rectangular cones, having a common vertex at the viewpoint, such that together they cover as much of the hemisphere as possible. A straightforward arrangement of the cones is to have a central cell surrounded by one or more rings of additional cells. Since this obviously cannot be achieved by placing multiple cameras physically at the common viewpoint, we virtually collocate them using planar mirrors. In this paper we only consider the case with one ring and a center cone, and will refer to the imager associated with the central cone as the center camera, and those associated with the ring of surrounding cones as side cameras; the common viewpoint is referred to as virtual viewpoint. However, the mirror design principle, the camera alignment technique, the image rendering method and the technique for solving mixing problem discussed in the later sections are applicable to the multiple-ring case with minor or no changes.

To achieve the virtual co-location, the planar mirrors must be placed in such a way that the image of each side camera's viewpoint coincides with that of the center camera. Thus, the center camera is surrounded by a ring of planar mirrors, and the side camera's viewpoints are placed at the images of the center camera's viewpoint into the mirrors. As illustrated in Fig. 3a, we will refer to the angle, ψ, formed by a mirror plane and the base plane (perpendicular to the optical axis of the center camera) as slope angle. The desired location of the side camera varies with the slope angle. To achieve the desired FOV contiguity, the planar mirrors will form a polyhedron which could be a pyramid, a prism or a cone according to whether the mirror planes form an acute, right, or obtuse slope angle with the base plane. We will refer to such a polyhedron containing N mirrors as an N-side polyhedron. A six-side polyhedron (pyramid) is shown in Fig. 3b. In addition to achieving virtual co-location, if the cells in the ring are chosen to be identical, a uni-

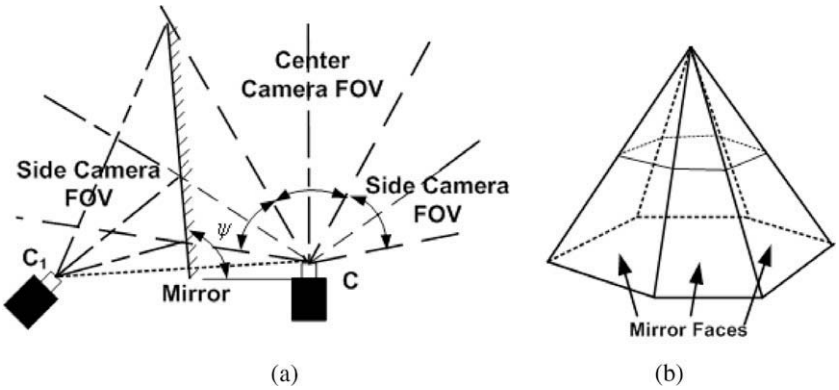


Fig. 3. (a) Co-locating side camera C1 with center camera C using a mirror. The location of C1 varies with the mirror slope angle ψ. (b) A six-side polyhedron.

form design can be obtained by using identical imagers, which means all the side cameras can be placed symmetrically around the center camera. Consequently, the horizontal cross section of the N-side polyhedron formed by these mirrors will be a regular N-polygon. Further, if we choose the vertical FOV of the center camera to be identical to those of the side cameras, the resolution across the entire hemisphere would be uniform. Since the distribution of the side camera is symmetric and the polyhedron is a regular one, in the analysis we only need to consider one camera-mirror pair plus the center camera. Fig. 4 shows one such configuration. In Fig. 4a, the mirror is one of N-side polyhedron faces with a slope angle ψ. The central angle of the mirror face, Φ, which is the angle subtended by the base edge of each face at the center of the base polygon, is $360°/N$. The side camera, with a vertical FOV denoted by φ, is oriented to form a tilt angle θ with the polyhedron base, and it captures the scene reflected off the corresponding mirror to cover one cell of the ring. Fig. 4b shows the vertical cross-sectional view corresponding to Fig. 4a. The vertical FOV of the center camera is identical to that of the side camera. The overall vertical FOV of the system, denoted as $\boldsymbol{\Omega}$, is the sum of the vertical FOV's of the center camera and the side cameras, i.e., $\Omega = 3\ \varphi$. The spatial relation of the side camera tilt angle θ and the vertical FOV of the camera φ is given by:

$$\theta + \varphi = 90^\circ \tag{1}$$

Notice that the FOV arrangement of the side camera vs. the center camera is independent of the mirror slope, which means that the mirror slope has no effect on the resolution uniformity.

2.2. Geometric constraints

As shown in Fig. 4b, since each side camera captures its image through a mirror, it is possible that the side camera sees itself in the mirror if the camera FOV is big or the tilt angle of the mirror is close to 90°, thus causing a blind spot in the side camera's FOV. To avoid such self-occlusion, the mirror face has to be tilted such that the side camera is below the lower limit CB of its vertical FOV. On the other hand, ψ cannot be so large as to make the mirror parallel to the upper limit CA of the side camera's FOV which results a FOV gap between the side camera and the center camera. Thus, the following inequality should hold:

$$90^\circ - \left(\theta - \frac{\varphi}{2}\right) < \psi < 90^\circ + \frac{\varphi}{2} \tag{2}$$

This inequality imposes a constraint on the values of the mirror slope angle, the vertical FOV and the tilt angle of the side camera. In Fig. 4b, the side camera is simplified as a pinhole. Here we can incorporate the camera size by adding an offset value on the left side of the inequality, which results in a tighter lower bound for the side camera FOV.

2.3. Sensor utilization constraint

The projection of a mirror face on the sensor of its associated side camera is in general a trapezoid due to keystone distortion as illustrated in Fig. 5a. Without loss of generality, we assume that the polyhedron is tall enough to cover the entire vertical FOV of the side camera. Under this circumstance, the height of the trapezoid equals to the height of the sensor, i.e., $D_V = d_V$, and the shape and size of the trapezoid is fully determined by the mirror and camera parameters. Here, we used two parameters, aspect ratio β and base angle ϖ, to represent the trapezoid shape and size. The aspect ratio is a function of the central angle of the mirror face, the vertical FOV and the tilt angle of the side camera:

$$\beta = \frac{\sin\left(\frac{\varphi}{2}\right)}{\tan\left(\frac{\phi}{2}\right)\cdot\cos\left(\frac{\varphi}{2}-\theta\right)} \tag{3}$$

The base angle of the trapezoid is a function of the central angle of the mirror face and the tilt angle of the camera:

$$\tan(90^\circ - \varpi) = \tan\left(\frac{\phi}{2}\right)\cdot\sin(\theta) \tag{4}$$

It is obvious that only those pixels inside the trapezoid capture the desired FOV parts reflected off the mirrors, while the pixels outside of the trapezoid capture direct views of other dis-contiguous scene segment. These dis-contiguous segments are properly captured by other sensors and represent wasted pixels. The ratio of the pixel count inside the trapezoid to the overall pixel count on the sensor is referred to as sensor utilization, denoted as Σ, and can be expressed as:

$$\Sigma = \begin{cases} \frac{area\_inside\_trapezoid}{area\_of\_sensor} = \frac{\alpha}{\beta} - \alpha\tan(90^\circ - \varpi) & D_H \le d_H \\ \frac{area\_in\_black}{area\_of\_sensor} = \frac{\alpha}{\beta} - \alpha\cdot\tan(90^\circ - \varpi) - \frac{\alpha}{4}\cdot\left(\frac{1}{\beta^2} - \frac{1}{\alpha^2}\right)\cdot\frac{1}{\tan(90^\circ - \omega)} & D_H > d_H \end{cases} \tag{5}$$

where α is the aspect ratio of the sensor and $a = \frac{d_v}{d_H}$, which is 0.75 for most of CCD cameras and $a = 0.75$ is assumed for our design; d_H are the width of the sensor; β is the aspect ratio of the trapezoid and $\beta = \frac{D_v}{D_H}$; and D_H is the base width of the trapezoid. The upper part of Eq. (5) holds when the bottom edge of the trapezoid is not bigger than the sensor width, the lower part is for the case when the bot-

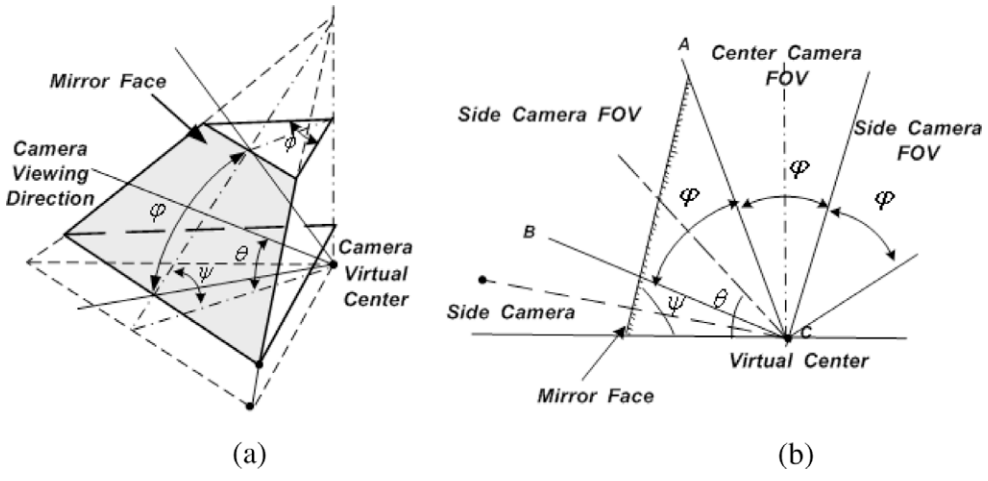


Fig. 4. (a) One mirror face of an N-side polyhedron and the visual field of its associated side camera (the physical imager is not shown). (b) A vertical cross-sectional view of (a).

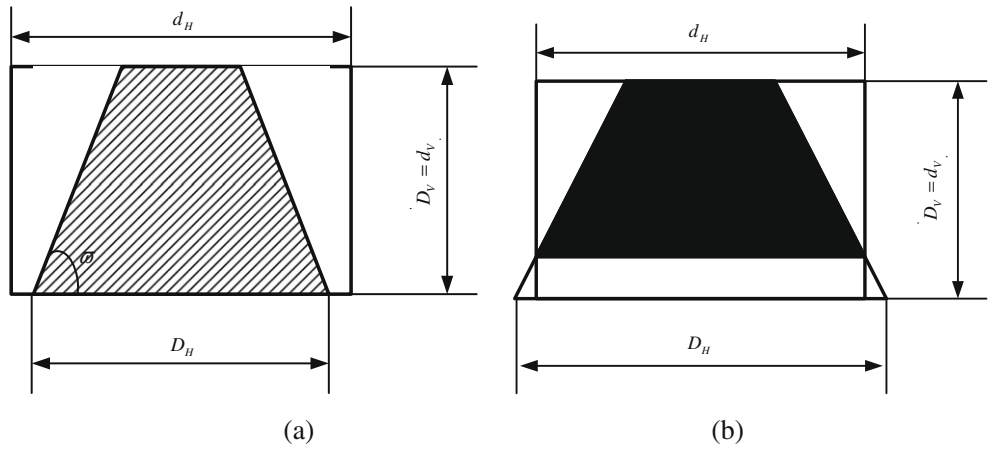


Fig. 5. (a) The image of the mirror face in the side camera sensor is in general a trapezoid shape due to perspective (keystone) distortion. (b) When the trapezoid is bigger than the sensor, only the black area captures useful FOV.

tom edge of the trapezoid is bigger than the sensor width, and only the image area in black represents useful pixels as shown in Fig. 5b.

2.4. Parameter selection

The above analysis makes clear that the design parameters must be chosen to satisfy contradictory constraints on sensor utilization, FOV size and uniformity of resolution. An optimal design is therefore needed. To do so, substituting Eqs. (1), (3), (4) into (5) yields:

$$\sum(\theta) = \begin{cases} 0.75 \cdot \tan\left(\frac{\phi}{2}\right) \cdot \frac{\cos(\theta)}{\tan\left(\frac{90^\circ-\theta}{2}\right)} & D_H \le d_H \\ \frac{3\tan\left(\frac{\phi}{2}\right)\cdot\cos(\theta)}{4\tan\left(\frac{90^\circ-\theta}{2}\right)} - \frac{3\tan\left(\frac{\phi}{2}\right)\cdot\cos^2\left(45^\circ-\frac{3\theta}{2}\right)}{16\sin^2\left(45^\circ-\frac{\theta}{2}\right)\cdot\sin(\theta)} + \frac{1}{3\tan\left(\frac{\phi}{2}\right)\cdot\sin(\theta)} & D_H > d_H \end{cases} \quad (6)$$

For a given N-side polyhedron, the sensor utilization is a function of θ alone. Fig. 6 shows the plot of the function $\Sigma(\theta)$ for $N = 6$. The sensor utilization reaches its maximum value when $D_H = d_H$ which is marked as a star in the plot. The maximum achievable sensor utilization for $N = 6$ is 71.65%. The optimal tilt angle of the side camera for this case is 40.9°, which happens when the bottom edge length of the trapezoid exactly matches with the sensor width. Since the function of sensor utilization $\Sigma(\theta)$ monotonically increases for $D_H < d_H$ and monotonically decreases for $D_H > d_H$, the equation $D_H = d_H$ always gives the optimal solution of the camera tilt angle. Fig. 6b shows the plots of the function $f(\theta) = D_H - d_H$ for $N = 4$–9. The zero crossings give the possible solutions. For example, for the curve with $N = 5$, there are three possible solutions associated with the three zero crossings at $\theta = 25°$, 90°, and 155°. Among these three solutions, $\theta = 25°$ and 155° correspond to the cases for which the side camera is placed on opposite sides of the center camera and thus the two solutions are identical. $\theta = 90°$ yields a trivial solution which corresponds to the case where the side camera FOV overlaps with that of the center camera. Therefore, there is only one valid solution for this case. It is interesting to see that a camera tilt angle $\theta = 90°$ is always a solution for all N (as expected) and the only solution for $N \geqslant 8$ which means that the polyhedron cannot have more than seven sides. However, the above analysis is based on a 0.75 sensor aspect ratio; if the side camera is rotated along its optical axis 90°, the sensor aspect ratio becomes 4/3; and there is still an optimal solution for eight-side polyhedron. In this paper, we will not consider the 4/3 aspect ratio case.

Once the tilt angle of the side camera is selected, its vertical FOV can be computed from Eq. (1). Given the optimal tile angle and vertical FOV, the feasible slope angle of the polyhedron face is constrained by inequality (2) which is plotted in Fig. 7a. The small circles indicate the upper limits of the feasible slope angle and the stars indicate the lower limits. Since the lower limit for N equal to 3 turns out to be larger than their corresponding upper limits,

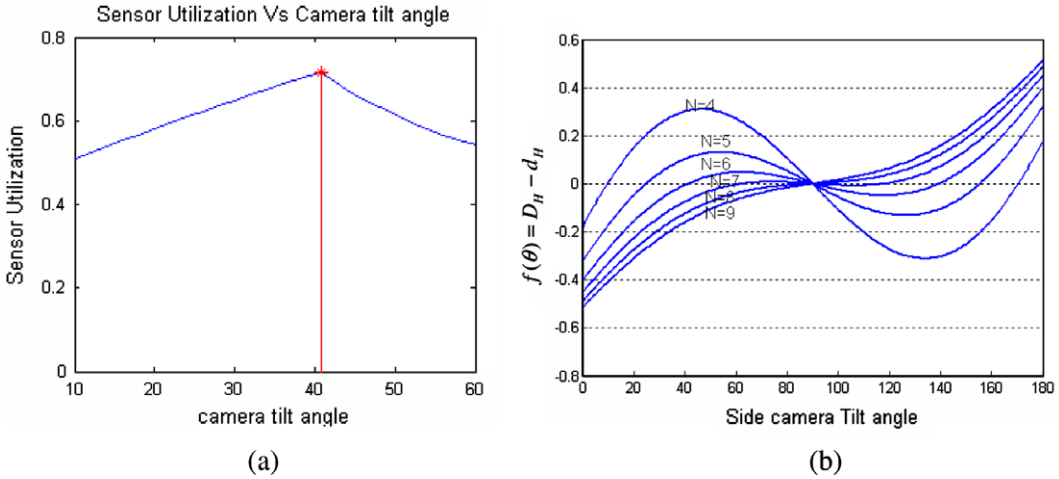


Fig. 6. (a) The plot of sensor utilization vs. the camera tilt angle for $N = 6$. (b) The plot of the function $f(\theta) = D_H - d_H$ for $N = 4$–9.

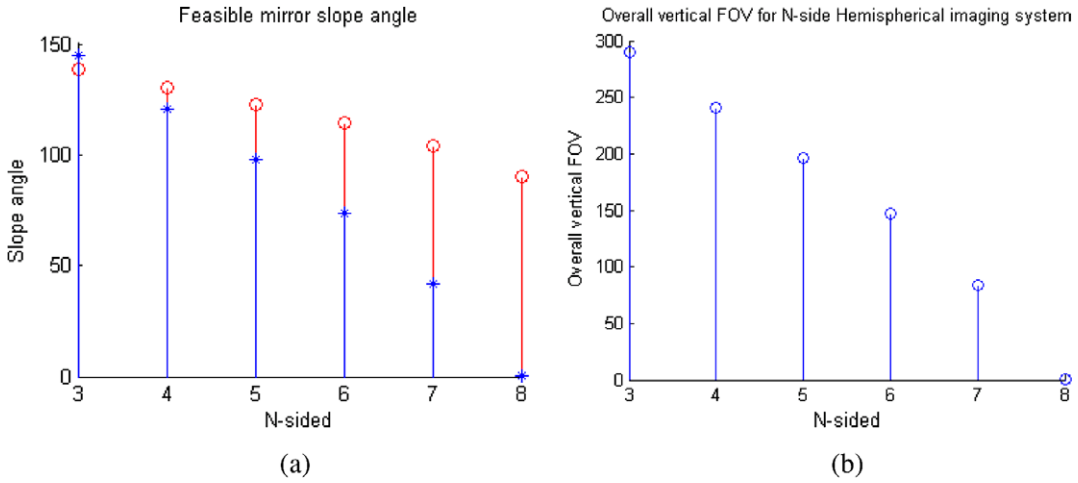


Fig. 7. (a) Feasible mirror slope angle of different *N*-polyhedron. (b) Overall vertical FOV of the system for different *N*-polyhedron.

the problem of camera self-occlusion cannot be avoided and consequently there is no valid solution for $N = 3$ that satisfies all the constraints. Since $N < 8$ from Fig. 6b, we can conclude that N can only be chosen from 4, 5, 6 or 7, corresponding to the camera tilt angle 9.6°, 24.7°, 40.9°, and 62.2°, respectively. For $N = 4$ or 5, the lower bound of the slope angle is more than 90°, which indicates that the polyhedron must take a cone shape. For $N = 6$ or 7, the slope angle could be larger, equal or less than 90°, which indicates that the polyhedron could be a cone, prism, or pyramid shape, respectively. Obviously, prism is the most manufacture-friendly shape of the three. One possible design using a hexagonal prism ($N = 6$) is illustrated in Fig. 8a, where the interior of the prism has a taper at the end so that the prism material thickness does not block the center camera's FOV as shown in Fig. 8b. The overall vertical FOV of the system for different N is plotted in Fig. 7b. For N equal to 5, 6 or 7, the overall vertical FOV's are 196°, 147.3° or 82.5°, respectively.

The analysis so far is based on an ideal camera model and the assumption that the camera FOV could be an arbitrary value ranging between 0° and 180°. In practice, the parameter selection must take into account the facts that only certain values of the parameters may be available, that the cameras may have non-zero center offsets and that lenses have geometric distortion. This may force sub-optimal choices, and as a result, the resolution uniformity and sensor utilization may be compromised. For example, when the tilt angle takes a smaller or a bigger value than the optimal one, the sensor utilization will drop as shown in Fig. 6a, and the overall vertical FOV of the system will increase or decrease. When the vertical FOV of the side camera is chosen to be not equal to the vertical FOV of the center camera, the resolution uniformity is sacrificed, but we gain the flexibility of choosing angle values given by Eq. (2). As a result, the overall achievable vertical FOV and the sensor utilization vary. It is important to note that these modifications do not affect the polyhedron shape significantly, as long as the overall vertical FOV is less than 180°. It is always possible to use the mirror prism to implement all variations. Only when the overall vertical FOV is larger than 180°, a non-prismatic mirror is necessary to eliminate self-occlusion.

3. Camera placement

To achieve a single viewpoint, the side cameras and center camera must be properly positioned with respect to the mirror prism so that the images of the side camera's viewpoints coincide with the center camera's viewpoint. Kawanishi et al. proposed a two-step method which first positions the camera cluster around a mirror pyramid visually, and then the camera orientation and position are finely tuned by analyzing the images of adjacent mirror faces [12]. Majumder et al. have reported an image warping technique to create panoramic images from roughly positioned cameras using pre-calibrated warping parameters [15]. Although such

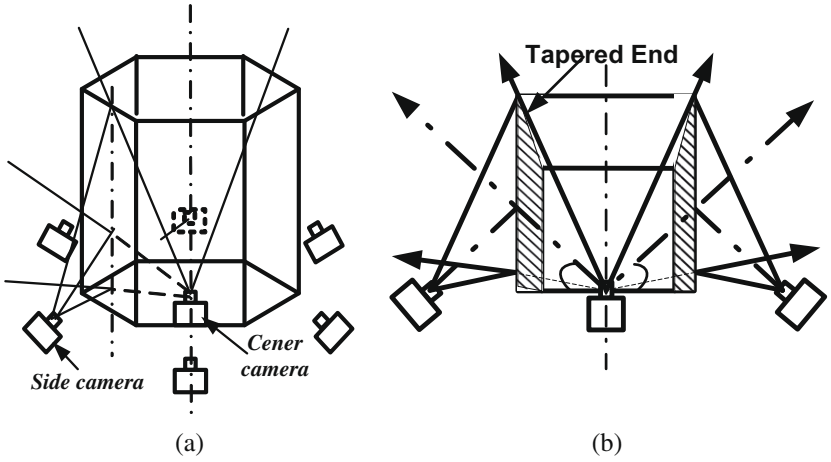


Fig. 8. (a) A concept design with a six-sided mirror prism. (b) Cross-sectional view.

methods may yield acceptable panoramic images for a limited depth range for which the warping parameters are pre-estimated, they will exhibit artifacts for objects far from the calibrated depth, and are not adequate for the general purpose seamless FOV solution aimed at here.

In this section, we describe a simple yet effective calibration method which places each camera independently with respect to its associated mirror face, without requiring direct alignment with other cameras. The method estimates the position and orientation of the side camera relative to its associated mirror face, and the position and orientation of the center camera relative to the top edges of all the prism faces by using correspondences of mirror edges (lines). The estimates directly yield the camera offsets from the desired position, which is compensated for by adjusting the camera location.

A pinhole model is assumed for cameras and the intrinsic parameters of the camera are pre-calibrated [29]. The image is normalized using the calibrated camera intrinsic parameters. Given a 3D point in the mirror frame, $X_M = [x_M \quad y_M \quad z_M \quad 1]^T$, its 2D projection in the image frame, X_I, can be obtained by:

$$X_I = P \cdot T_{C \leftarrow M} \cdot X_M \tag{7}$$

where 3×4 matrix $T_{C \leftarrow M}$ is the transformation from the mirror frame to the camera frame and can be decomposed into the four column vectors $[\vec{r}_1 \quad \vec{r}_2 \quad \vec{r}_3 \quad \vec{S}_{C \leftarrow M}]$ and P is a 3×3 identity matrix (camera projection matrix). Given a line in the image $\vec{A}^T \cdot X_I = 0$ with $\vec{A} = [a \quad b \quad c]^T$, the corresponding 3D line in the mirror frame is:

$$\vec{A}^T \cdot T_{C \leftarrow M} \cdot X_M = 0 \tag{8}$$

Without loss of generality, we choose the mirror face as XY plane of the mirror coordinate system and express the mirror edges in the 2D line form $y_M = kx_M + d$. Comparing $y_M = kx_M + d$ with Eq. (8) yields two equations:

$$\vec{A} \cdot \vec{r}_1 + k \cdot \vec{A} \cdot \vec{r}_2 = 0 \tag{9}$$

$$\vec{A} \cdot \vec{S}_{C \leftarrow M} + d \cdot \vec{A} \cdot \vec{r}_2 = 0 \tag{10}$$

To estimate three rotation angles in $[\vec{r}_1 \quad \vec{r}_2 \quad \vec{r}_3]$ and three translation components in $\vec{S}_{C \leftarrow M}$, it is necessary to have at least three line correspondences. Since each mirror face has three or four well-defined edges, all of these can be utilized to recover the aforementioned rotation and translation. One may notice that Eq. (9) contains no translation components, which indicates that we can use Eq. (9) alone to estimate the rotation angles, and then use Eq. (10) to estimate $\vec{S}_{C \leftarrow M}$. Eq. (9) is nonlinear, in general, requiring a nonlinear solver.

4. Generating panoramic images

After the side and center cameras are properly positioned around or inside the mirror prism, the images captured by these cameras are post-processed and stitched into complete panoramas. Various approaches have been taken to composite multiple camera images into panoramic images [4,5,22,24]. For example, Szeliski and Shum proposed a technique to merge multiple images captured with a pan-tilt camera [24]. Their method registers large set of images using a group of warping matrices estimated from the overlaps between adjacent images. However, adjacent images in our case here are formed after reflection from adjacent mirror faces which have no overlap. As a consequence, it is necessary to calibrate the orientation of each camera with respect to its associated mirror in order to generate the panoramic image. In our implementation, the camera orientations are estimated along with the camera placement using mirror edges as explained in Section 3.

For visualization purpose, the panoramic image is often represented as a texture-mapped surface, where the surface texture is the back projection of the camera images and the surface is centered at the origin which is the viewpoint. Such representation is called environment map [4,7]. Two most widely used environment maps are cylindrical panorama or cubic panorama that uses a cylinder or a cube as the textured surface, respectively. The advantage of these two maps is that the panoramic image can be visualized by Apple's QuickTime VR [7] or other popular visualization tools. However, for the proposed camera, projecting the images on cylindrical or cubic surface will cause significant resolution loss. This is due to the fact that the side cameras are tilted at a large angle with respect to the mirror faces which results in a trapezoidal sensor image (Fig. 9a). When projected on the vertical planar surface of the cube or the curved surface of the cylinder, the top portion of trapezoid must be up-sampled, while the bottom portion down-sampled. The resolution of the resulting rectangular image is, thus not uniform—the resolution of the top portion is much lower than that of the lower portion. This results in loss of visualization quality. To avoid up- and down-sampling, we project camera images on the faces of a truncated virtual pyramid which is obtained by sharing the top and expanding the base of the hexagonal prism by an amount such that the angle between the pyramid face and its base

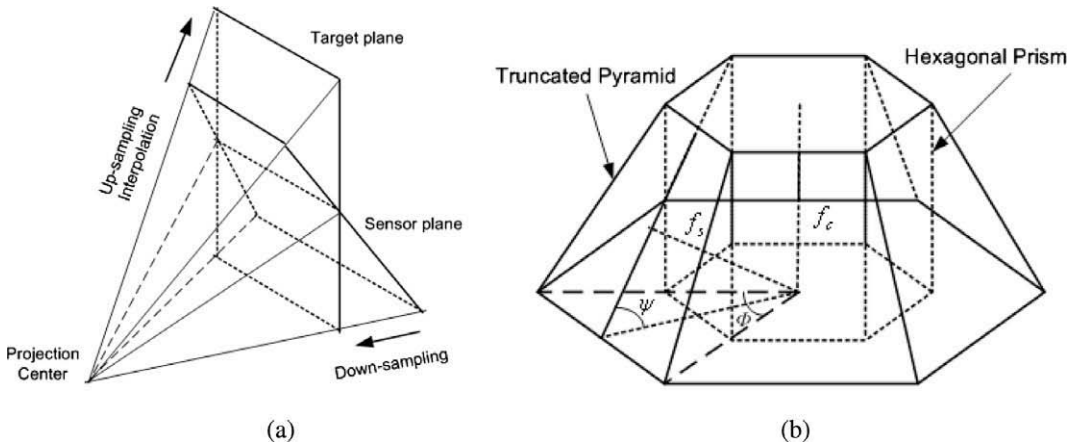


Fig. 9. (a) Projection of a tilted sensor image on a vertical rectangular planar surface requires up-sampling in the top portion of the image, as well as down-sampling in the low portion of the image, resulting in non-uniform resolution. (b) The panoramic image represented by a texture-mapped truncated pyramid with its base as the magnified base of the hexagonal prism. The tilt angle of the pyramid, between the pyramid face and base, is the complementary angle of the chosen camera tilt angle. The size of the polygon faces is determined by the camera focal lengths (f_s is the focal length of the side camera and f_c is the focal length of the center camera).

is the complimentary angle of the side camera tile angle as shown in Fig. 9b. The images from six side cameras are projected on the trapezoid faces of the pyramid and the images from the center camera are projected on the top face of the pyramid. The size of the virtual pyramid is normalized such that the distance from the virtual center to the pyramid side face equals the chosen focal length of the side camera and the distance from the virtual center to the top surface of the pyramid equals the chosen focal length of the center camera, where the focal lengths of all side cameras are standardized to a common value. Since the pyramid faces have the same size and shape as the images of the mirror faces, projection process does not entail any loss of resolution. Further, we consider the images projected on pyramid faces as having been acquired by seven virtual cameras (six virtual side cameras and one virtual center camera) that are located at the virtual viewpoint and whose optical axes are perpendicular to the pyramid faces. The focal lengths of the virtual side and center cameras are set to be the same as the corresponding chosen focal length value. As a result, the projected image of each mirror face on the virtual camera image plane is exactly the same size as the pyramid face. The construction of the virtual pyramid ensures that these textures are aligned geometrically across the pyramid faces after the projection. Fig. 10 shows the flow chart for generating textures using one of the side cameras as an example. The projection of the side camera image on the virtual camera image plane is given by the following equation:

$$U_A = P_v \cdot T'_{v\leftarrow c} \cdot P_c^{-1} \cdot U_c \tag{11}$$

where U_A represents the pixel coordinates in the virtual camera reference; U_c represents the pixel coordinates in the side camera reference after correcting the lens distortion; P_c is the projection matrix of the side camera; P_v is the projection matrix of the virtual camera; and $T'_{v\leftarrow c}$ is the warping matrix. The warping matrix $T'_{v\leftarrow c}$ contains only rotation component which is estimated experimentally during the camera placement process. The translation component in $T'_{v\leftarrow c}$ is set to zero since the side camera is placed with respect to the corresponding mirror.

5. Mixing problem

So far, we have assumed that the cameras are pinholes. In practice, the aperture of a real lens always has a finite size. Instead of having only a single principal ray as in a pinhole model, a bundle of rays from one object point passes through the aperture and forms an image on the sensor. When the object is not at the right distance, the non-pinhole aperture causes defocus as is well known. For mirror-based imaging system, the non-pinhole aperture will introduce not only the defocus blur, but also ray blending phenomena near the mirror edge. The finite aperture allows two different sub-bundles of rays from two different object points to reach the camera and imaged on the same image point. As illustrated in Fig. 11, ray bundle Q_1 reflected from object point S_1 and ray bundle Q_2 directly from object point S_2 pass through lens aperture C_{11} and are concurrently imaged on the same point S. We refer to such blending of the images of two dis-contiguous regions as the mixing problem. This problem affects all mirror-based panoramic systems, causes noticeable artifacts around mirror edges, but has no effect on the rest of the image, away from the mirror edges.

In practice, the width of the mixing area may range from a few pixels to more than 20 pixels, depending on the size of the lens aperture and the distance of the mirror edge points from the aperture. In our implementation, the width of the mixing region is about 30 pixels near the bottom edge of the mirror, while it is only around 6–8 pixels near the top edge.

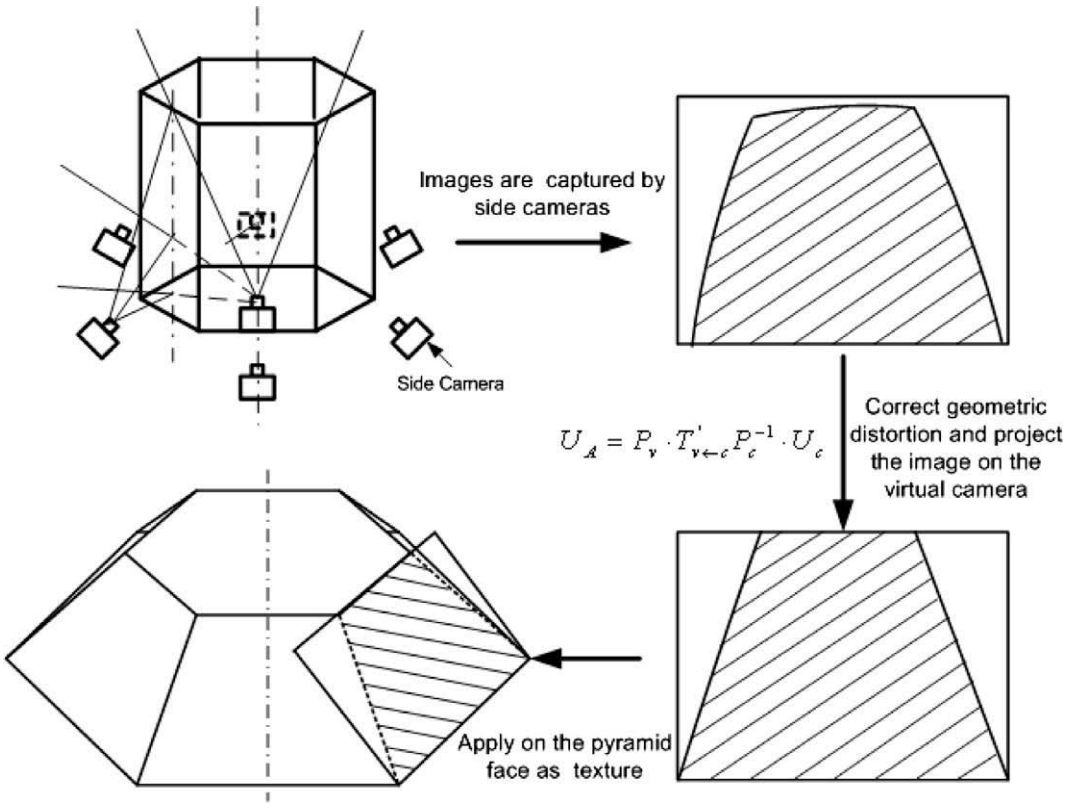


Fig. 10. Generating textures for the truncated pyramid. An image from one of the side cameras is projected on the corresponding face of the pyramid after correcting geometric distortion due to the camera lens.

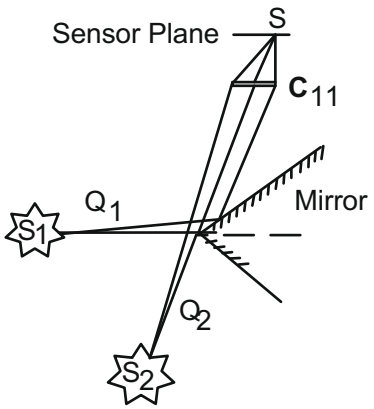


Fig. 11. Illustration of mixing artifacts caused by non-pinhole aperture.

The effect of mixing of two rays can be expressed as a weighted sum of intensities contributed by the two object points:

$$I_a = \alpha \cdot I + (1 - a) \cdot I_n \tag{12}$$

where I_a is the mixing intensity captured by the camera, I_n is the intensity due to the light received directly from an object, I is the intensity due to the light received after reflection off the mirror, and the weight a is the mixing ratio and has a value between 0 and 1. The value of a for a given pixel depends on the pixel location, the shape and size of the lens aperture, as well as the position and orientation of the mirror. The weight can be computed if all these geometry factors are known. Unfortunately, for most of the lens, the shape of the lens aperture could be irregular and the size of the lens aperture is often unknown. In our experiment we took a calibration approach and estimated the weights with a controlled illumination.

Eq. (12) provides a single constraint but contains two unknowns and therefore has infinitely many solutions. To solve this problem, additional constraint, referred to as "local intensity constancy constraint" which states that the neighbors of a given mixing pixel have the same intensity value as the given pixel, is assumed. Using this constraint, the problem can be formulated as a minimization problem. We rewrite (12) in vector form:

$$I_a = \vec{a}^T \cdot \vec{I} \tag{13}$$

where $\vec{I} = [I \quad I_n]^T$ and $\vec{a} = [a \quad 1 - a]^T$

Estimating $\vec{I}$ in a local window is equivalent to minimizing the following quadratic error function:

$$E(\vec{I}) = \sum_w \| \vec{a}^T \cdot \vec{I} - I_a \|^2 \tag{14}$$

where w indicates a small window around the mixing pixel. To obtain an optimal solution, computing the derivative of the error function with respect to the unknown $\vec{I}$, and setting the derivative to zero yields:

$$\vec{I} = G^{-1} \cdot B \tag{15}$$

where $G = \begin{bmatrix} \sum a^2 & \sum a \cdot (1-a) \\ \sum a \cdot (1-a) & \sum (1-a)^2 \end{bmatrix}$ $B = \begin{bmatrix} \sum a \cdot I_a \\ \sum (1-a) \cdot I_a \end{bmatrix}$

For most of mixing pixels, G generally has full rank and is invertible. But when the pixels are far away from the mirror boundaries, a is close to 1, and G may be rank deficient which will add to estimation error. It is necessary to check the rank condition of G when a is close to 1.

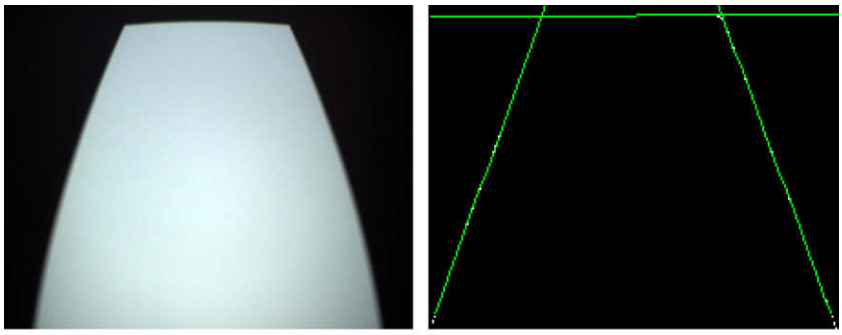
6. Experimental results and discussion

6.1. Implementation

We have implemented the proposed hemispherical camera design using a hexagonal mirror prism. The height of the prism is 74.27 mm and the width of the cross section is 69.54 mm. A total of seven Dragonfly color cameras from Point Grey, each with a resolution of 640 × 480 pixels, are used in the system. Six of them equipped with 4 mm micro-lenses are placed around the mirror prism with a tilt angle 42.5° (slightly different from the optimal value due to the smaller FOV of the available micro-lenses). The seventh camera is equipped with a 3.5 mm lens and placed on the axis of the prism. Each side camera effectively covers 45° vertically, while the center camera covers 50° vertically. Thus, the sides plus center camera have a total vertical FOV of $50 + 2 \times 45 = 140°$, which results in a panoramic system with $360° \times 140°$ overall FOV. As mentioned at the end of Section 2, the value of 140° can be increased up to 180° by reducing the sensor utilization or sacrificing the resolution uniformity. A snapshot of the system implementation is shown in Fig. 2.

6.2. Camera placement

A 3-way translation stage is used to adjust the camera position. The initial estimate given to the nonlinear solver is the ideal orientation (42.5°, 0°, 0°) for side cameras, and (0°, 0°, 0°) for the center camera. Since the actual orientation of the camera is just slightly off from the ideal orientation, using the ideal orientation to initialize the nonlinear solver is very likely to lead to the correct answer. For each side camera, there are three edges (top edge plus two side edges) contained in the image. For the center camera, all six edges are available. Fig. 12a shows a calibration image captured by one of the side cameras. Fig. 12b shows the detected edges using Canny



(a) (b)

Fig. 12. (a) One of mirror image for estimating camera extrinsic parameters. The mirror face is illuminated by a uniform light panel and the camera shutter speed is set to a minimum value. (b) The detected edges using Canny edge detector (lens distortion is compensated).

Table 1
Camera placement results.

Orientation (°)		Position (offset: mm)	
Before	After	Before	After
42.41	42.42	0.397	0.035
−0.165	−0.14	1.073	−0.0028
0.421	0.49	0.0970	0.017

edge detector. Two sets of the estimation results corresponding to before and after adjusting the position of the camera are shown in Table 1. It clearly shows that the camera moves toward the desired location.

6.3. Generating panoramic image

Each of the seven camera images is first rectified and warped to the image plane of the corresponding virtual cameras using its intrinsic parameters and estimated orientation, and then the true intensity values of the pixels in mixing regions are estimated using the method described in Section 5. Finally, the processed images are mapped on the virtual pyramid faces as texture for visualization. Fig. 13 illustrates the image process steps using one of side camera inputs as an example. Fig. 14 shows a frontal panoramic image captured by the virtual camera after texture mapping. The FOV from the top edge of the image to the bottom edge is exactly 140°.

6.4. Mixing artifacts

The true values of the mixing pixels are solved using a 3×3 window. Due to the application of "local intensity constancy constraint", the estimated images in the mixing regions are slightly blurred as expected (Fig. 13d). The bigger a local window, the more the expected blurring. However, it is worth noting that the non-pinhole aperture does not only cause problems; it sometime helps preserve the information. For instance, when the edges of a mirror pyramid have flat transitions, rather than perfect knife edges as desired, perfect pinholes may lead to a dark strip or a gap at the boundary of two adjacent cameras, as noted in [17]. A non-pinhole aperture allows us to see beyond the image boundary and reduce the information loss. In our current implementation, the edges between adjacent mirror faces are sharp, but 0.5 mm flat transitions are present at the top edges. Fig. 14 shows two panoramic images with and without compensating for mixing artifacts. In Fig. 14a, note the boundaries visible between the side camera FOV's, and the much darker strips at the boundary of the center camera caused by the 0.5 mm flat transitions at the top edges. Fig. 14b shows that the compensation process has removed the mixing region artifacts and the images in those areas are well recovered.

7. Conclusion and future work

We have presented a hemispherical camera system. The system consists of multiple imaging sensors and a hexagonal prism made from planar mirror faces with a wedge-shaped interior. The sensors are positioned in such a way that they, collocate at a common viewpoint, image different parts of the scene directly or reflected off the prism. Besides a detail description of the camera design, prototype implementation, and experimental results, we have also presented a novel camera placement technique to achieve single viewpoint panoramic imaging and a method to compensate for mixing artifacts caused by finite lens aperture. While the results

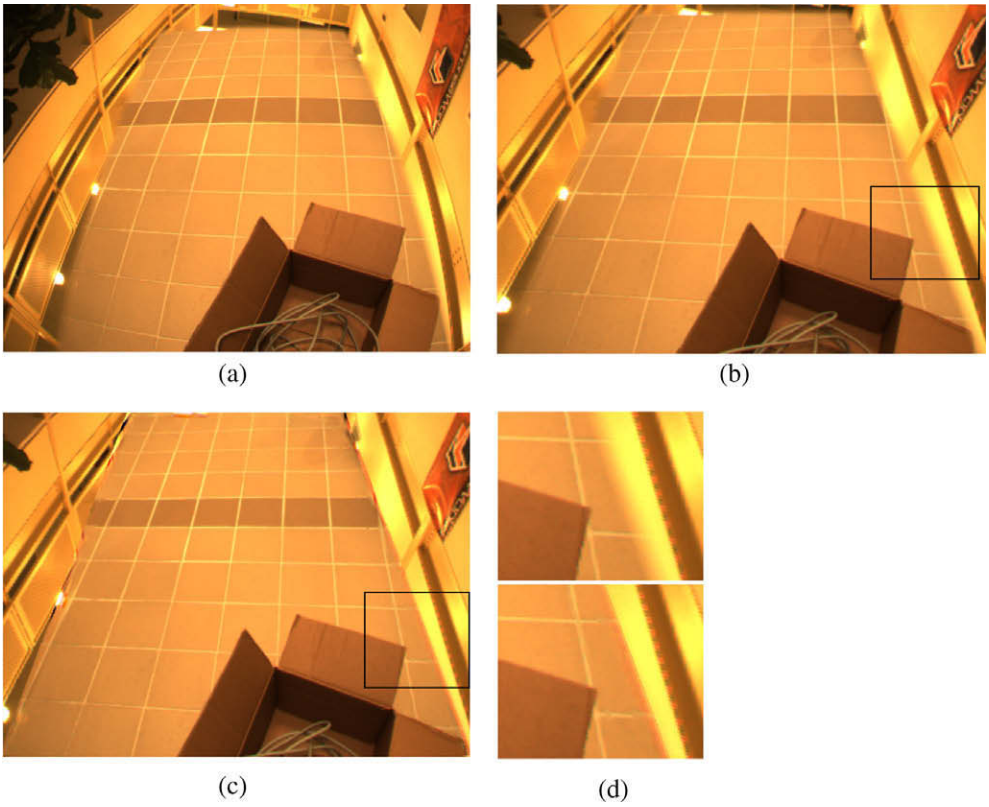


Fig. 13. Steps in composing hemispherical panoramic images from those acquired by the component imagers. (a) The input image taken from one of the side cameras. (b) The image after the warping. (c) The image after compensating for mixing artifacts. (d) Zoomed sub-images from (b) and (c). Upper one (from (b)) shows the mixing effect between yellow wall and orange floor; lower one shows the image after applied proposed mixing algorithm. The scene in mixing area is well recovered while the boundaries of these yellow stripes on the floor are blurred due to finite size window using in the proposed algorithm. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

(a) (b)

Fig. 14. Panoramic images before and after compensating for mixing artifacts. (a) Before compensation; (b) after compensation—the darker strips around the individual image boundaries have been removed by the mixing-compensation process.

show that the proposed method for mixing compensation improves the image quality, it is unable to successfully recover objects boundaries where edges present. In future work, we plan to devise solutions to this problem.

References

[1] M. Aggarawal, N. Ahuja, High dynamic range panoramic imaging, in: Proceedings of International Conference on Computer Vision, Vancouver, Canada, July 2001, pp. 2–9.

[2] P.I. Anderson, From tele-presence to true immersive imaging: into real-life video-now!, Advanced Imaging 10 (7) (1995) 48–50

[3] S. Baker, S.K. Nayar, A theory of single-viewpoint catadioptric image formation, International Journal of Computer Vision 35 (2) (1999) 1–22.

[4] R. Benosman, S.B. Kang, O. Faugeras (Eds.), Panoramic Vision: Sensors, Theory and Applications, Springer-Verlag, 2001.

[5] M. Brown, D.G. Lowe, Recognising panoramas, in: Proceedings of the Ninth International Conference on Computer Vision (ICCV2003), Nice, France, October 2003, pp. 1218–1225.

[6] J.S. Chahl, M.V. Srinivasan, Reflective surfaces for panoramic imaging, Applied Optics 36 (31) (1997) 8275–8285.

[7] S. Chen, QuickTime VR—an image based approach to virtual environment navigation, in: Proceeding of SIGGRAPH, ACM, New York, 1995, pp. 29–38.

[8] R.A. Hicks, R. Bajcsy, Catadioptric sensors that approximate wide-angle perspective projections, in: Workshop on Omnidirectional Vision, 2000, pp. 97–103.

[9] H. Hua, N. Ahuja, A high-resolution panoramic camera, in: Proceedings of CVPR, 2001, pp. 960–967.

[10] Hong Hua, N. Ahuja, C. Gao, Design analysis of a high resolution panoramic camera using conventional imagers and a mirror-pyramid, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (2) (2007) 356–361.

[11] H. Ishiguro, M. Yamamoto, S. Tsuji, Omni-directional stereo, IEEE Transactions On Pattern Analysis and Machine Intelligence 14 (2) (1992) 257–262.

[12] T. Kawanishi, K. Yamazawa et al. Generation of high resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors, in: 14th International Conference on Pattern Recognition, Brisbane, Australia, 1998, pp. 485–489.

[13] A. Krishnan, N. Ahuja, Range estimation from focus using a nonfrontal imaging camera, International Journal of Computer Vision 20 (3) (1996) 169–185.

[14] A. Krishnan, N. Ahuja, Panoramic image acquisition, in: Conference on Computer Vision and Pattern Recognition, 1996, pp. 379–384.

[15] A. Majumder, W. Brent Seals, M. Gopi, H. Fuchs, Immersive teleconferencing: a new algorithm to generate seamless panoramic video imagery, in: Proceedings of ACM Multimedia, 1999, pp. 169–178.

[16] K. Miyamoto, Fish eye lens, Journal of Optical Society of America 64 (1964) 1060–1061.

[17] V. Nalwa, A true omnidirectional viewer, Technical Report, Bell Laboratories, 1996.

[18] V. Nalwa, Stereo panoramic viewing system, US Patent 6141145, 2000.

[19] V. Nalwa, Compact high-resolution panoramic viewing system, US Patent 6,195,204 B1, 2001.

[20] S.K. Nayar, Catadioptric omnidirectional camera, in: Conference on Computer Vision and Pattern Recognition, 1997, pp. 482–488.

[21] S.K. Nayar, V. Peri, Folded catadioptric cameras, in: Conference on Computer Vision and Pattern Recognition, vol. 2, June 1999, pp. 217–223.

[22] H. Shum, R. Szeliski, Construction and refinement of panoramic mosaics with global and local alignment, in: Proceedings of International Conference on Computer Vision (ICCV1998), 1998, pp. 953–958.

[23] M.V. Srinivasan, A new class of mirrors for wide-angle imaging, in: Proceedings of IEEE Workshop on Omnidirectional Vision and Camera Networks, Madison, Wisconsin, USA, June 2003.

[24] R. Szeliski, H. Shum, Creating full view panoramic image mosaics and environment maps, in: SIGGRAPH, 1997, pp. 251–258.

[25] K.H. Tan, H. Hua, N. Ahuja, Multiview panoramic cameras using mirror pyramid, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (7) (2004) 941–946.

[26] Available from: <www.ptgrey.com/products/ladybug2/index.html>.

[27] Y. Xiong, K. Turkowski, Creating image-based VR using a self-calibrating fisheye lens, in: Conference on Computer Vision and Pattern Recognition, 1997, pp. 237–243.

[28] K. Yamazawa, Y. Yagi, M. Yachida, Omnidirectional imaging with hyperboloidal projection, in: International Conference on Intelligent Robots and Systems, July 1993, pp 1029–1034.

[29] Z. Zhang, A flexible new technique for camera calibration, Technical Report MSR-TR-98-71, Microsoft Research, 1998.