Split Aperture Imaging for High Dynamic Range

Manoj Aggarwal Narendra Ahuja
University of Illinois at Urbana-Champaign
405 N. Mathews Ave, Urbana, IL 61801, USA
{manoj,ahuja}@vision.ai.uiuc.edu

Abstract

Most imaging sensors have limited dynamic range and hence are sensitive to only a part of the illumination range present in a natural scene. The dynamic range can be improved by acquiring multiple images of the same scene under different exposure settings and then combining them. In this paper, we describe a camera design for simultaneously acquiring multiple images of the same scene under different exposure settings. The cross-section of the incoming beam from a scene point is partitioned into as many parts as the desired degree of split. This is done by splitting the aperture into multiple parts and directing the light exiting from each in a different direction using an assembly of mirrors. A sensor is placed in the path of each beam and exposure of each sensor is controlled either by appropriately setting its exposure parameter, or by splitting the incoming beam unevenly. The resulting multiple exposure images are used to construct a high dynamic range image. We have implemented a video-rate camera based on this design and the results obtained are presented.

1. Introduction

The brightness variation within a real-world scene is usually quite large. A conventional digital camera provides only 8-bits (256 levels) of brightness information which is typically inadequate to faithfully capture the entire range of lummination levels and results in an image with many areas either too dark (under saturated or clipped) or too bright (oversaturated). This range of brightness levels that can be captured by a sensor without clipping or saturation is often referred to as the dynamic range.

The dynamic range of a sensor can be improved using a number of techniques. The basic idea is to acquire multiple images using different exposure settings, thus capturing different sections of the illumination range. These multiple exposure images are then combined (intensity space mosaicing) to cover a larger brightness range which is a union of those covered by the individual images. There are two main types of high dynamic range cameras, depending on whether they have video-rate capability or not. The first type sequentially acquires multiple exposure images, each of which has a resolution identical to that of the sensor. These camera do not have video-rate capability and are suited for only stationary scenes [2, 3, 8, 9, 10]. The second type acquires images at video-rate, which can be achieved using several techniques, each having different tradeoffs. A number of special cameras have been disclosed in patents, which employ a single lens but multiple sensors such that the same scene gets simultaneously imaged on different sensors, preset to different exposure settings. Many of these designs typically require precise alignment of various optical elements and the sensors. In another approach, a special sensor is reported which has multiple sensing elements with different sensitivities for recording light for every pixel on the sensor. This approach uses a single sensor, but requires specialized and expensive hardware design. A special high dynamic range sensor has been proposed in [1], that instead of measuring the amount of charge accumulated in a pixel, measures the time it takes to reach saturation. The recorded times are used to convert them into a high dynamic range image. All the approaches discussed above generate high dynamic range images with resolution that of the sensor. A technique to trade off spatial resolution and quality for dynamic range is presented in [11]. N images with different exposures are captured on a single sensor at $\left(\frac{1}{N}th\right)$ the resolution of the sensor. Interpolation is used to construct a full-resolution image. In this paper, we present a camera design based on multiple sensors capable of acquiring a high dynamic range image at video rate.

The basic idea in multiple sensor based high dynamic range cameras is to split the light refracted from the lens into multiple beams, each of which is then allowed to converge on a sensor suitably placed in the path of the beam. The splitting of the light can be performed by a number of traditional beam-splitting elements such

0-7695-1143-0/01 $10.00 © 2001 IEEE

as semi-transparent mirrors, polka-dot beamsplitters, dichroic cube beamsplitters, pellicle beamsplitters and special prisms [4]. There are two drawbacks of such designs. First, besides splitting the light into multiple beams, the splitters introduce additional lens aberrations because many of them are made of glass with finite thickness and refract light (except pellicle beamsplitters, which is a thin film). Second, most of the splitters are two-way, except for the special prisms which can split the light into three beams. In order to generate more than three beams, the splitters may be used in succession. However, the number of splitters that can be so accomodated physically is severely constrained due to the typically short optical path between the lens and the sensor. We use a design that partitions the cross-section of the incoming beam into as many parts as desired degree of split. This is done by splitting the aperture into multiple parts and directing the light exiting from each in a different direction using an assembly of mirrors. This method avoids both of the above drawbacks encountered using traditional beamsplitters. The use of mirrors for splitting light is not new as they have been used in a patent disclosing a design for a high speed camera [5]. The novelty of this paper lies in the proposed arrangement of mirrors and associated analysis for the distribution of light on the sensors for high dynamic range imaging.

The rest of the paper is organized as follows. Sec. 2 presents how a compact arrangment of mirrors can be used to split the incoming beam into multiple beams. Sec. 3 discusses the properties of the various split beams and their relationship to the arrangement of the mirrors. Sec. 4 describes the suitable mirror arrangment and algorithm for generating high dynamic range images. We provide experimental results and conclude in Sec. 5.

2. Mirror based beamsplitter

Consider a simple camera configuration based on a traditional two-way beam-splitter shown in Fig. 1(a). It shows that every single ray of light from point object P that passes through the lens aperture and then refracts gets split into two rays by the beamsplitter. In this sense, the *entire aperture* contributes to the light that reaches both the sensors. The idea behind the mirror based beam splitter is that rather than splitting every individual ray from an object point into multiple sub-rays, we instead split the set of rays from the object such that disjoint subsets of rays form the different beams. An individual ray is not split but deflected in its entirety. The lens aperture is carved into parts and a sensor is associated with each part. A sensor receives only the portion of the bundle of rays emanating from the object and passing through the corresponding portion of the aperture. The splitting can be achieved for example by a mirror as shown in Fig. 1(b). We note, that a mirror introduces no aberrations. Further as we will show next, multiple mirrors can be very compactly arranged to split the incoming beam/aperture into arbitrarily large number of subdivisions.

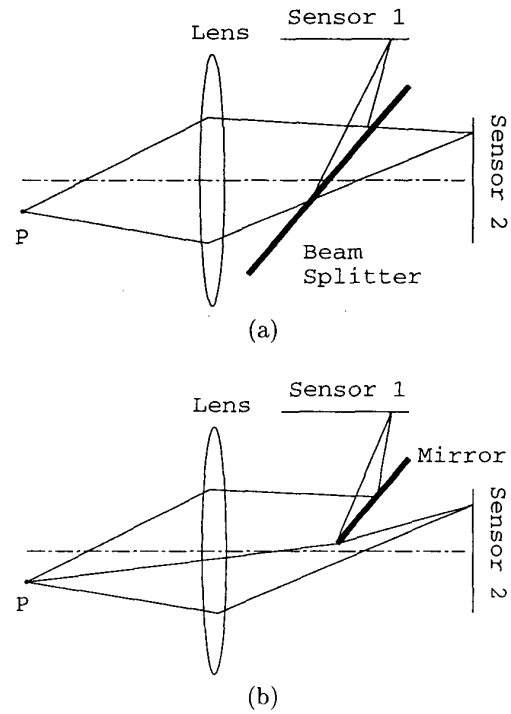


Figure 1. Splitting light into multiple beams. (a) Traditional beam splitter splits every ray of light into multiple rays. (b) Proposed mirror based splitter selectively routes different rays towards different directions.

Consider an n-sided pyramid ($n > 2$) whose top and side view is shown in Fig. 2. The pyramid has n triangular mirrored faces all sharing the pyramid apex or tip. The base of the pyramid is a n sided regular polygon. Let the length of a side of the polygon be denoted by s and height of the pyramid from the tip to the base by h. Consider a arrangement of the lens and the pyramid in which the medial axis of the pyramid is aligned with optical axis, as shown in Fig. 3. Each triangular face of the pyramid is a mirror and deflects only a portion of the total incoming beam. The split beams get deflected into different directions depending on the direction of the normals to the triangular faces, which can then be imaged on sensors appropriately placed in their paths. The cross-section of the incoming beam that gets de-

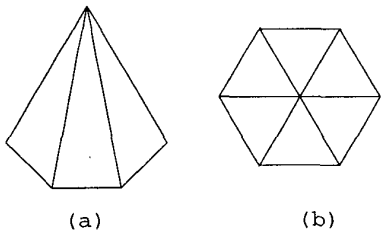


Figure 2. A compact arrangement of mirrors in the shape of an n-sided pyramid for beam splitting (a) Front view of the pyramid, (b) top view.

flected by a particular face depends on the position of the tip of the pyramid and the coordinates of the point object under consideration. We will explore these relationships in the next section, as these will determine the spatial distribution of light on the sensor.

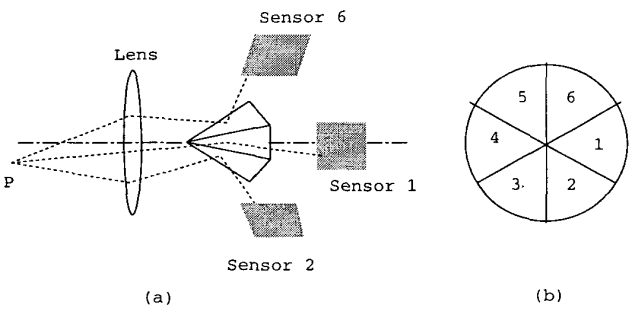


Figure 3. (a) The lens, mirror pyramid and sensors arrangements for capturing multiple images of the same scene simultaneously. (b) The aperture/lens surface is split into parts as shown.

3. Properties of the split beams

The cone of light rays from an object point that enter the lens gets split into smaller cones at the pyramid. The cross-section of rays that get deflected by a face of the pyramid corresponds to a particular region on the aperture which depends on the position of the object point and the tip of the pyramid. If we assume that a point object uniformly illuminates the aperture, the area of the carved region can be used to represent the percentage of light that the face of the pyramid deflects. In this section, we will develop the relationship between the area carved on the aperture, given the location, orientation and dimensions of a face of the pyramid, and the coordinates of a point object.

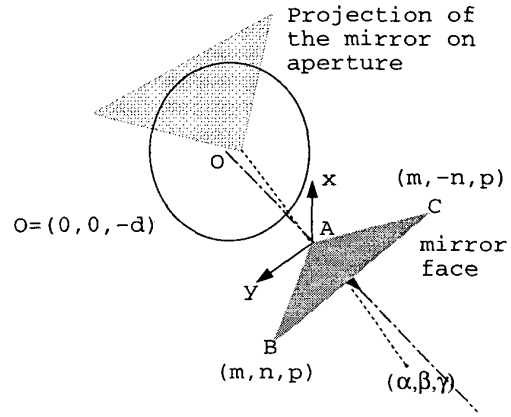


Figure 4. The coordinate system and projection of the pyramid face on the aperture/exit pupil

Since the lens is circularly symmetric and the pyramid is regular, it suffices to perform the analysis for any one of the n faces. Let the tip of the pyramid be at a distance d from the center of the (thin) lens along the optical axis (Fig. 4). For thick-lenses, the center of the lens is not well defined and the distance d must be measured from the effective aperture as seen from the image side of the lens. This effective aperture from the image side is also known as the exit pupil [7]. We define a rectangular coordinate system with origin located at the tip of the pyramid and z axis pointing towards the base of the pyramid. Then the coordinates of a face of the pyramid can be expressed as $(0,0,0), (m,n,p)$ and $(m,-n,p)$, without loss of generality. The coordinates of the center of the aperture/exit pupil are then given by $(0,0,-d)$. Let P denote a point object in the scene, and the location of its focused image in the absence of the pyramid be denoted by (α,β,γ), where γ depends on the z-coordinate of the object point and is determined by the lens law. This implies that in the absence of the pyramid all rays from the point object that refract through the lens converge at (α,β,γ). When the pyramid is introduced some of the rays will be intercepted by the mirrored face under consideration and deflected. This set includes exactly those rays formed by joining any point on the triangular face and point (α,β,γ). The region carved out by these rays on the aperture is thus given by the projection of the triangular mirror on the aperture as viewed from (α,β,γ). The projection of edges $AC = (0,0,0) - (m,n,p)$ and $AB = (0,0,0) - (m,-n,p)$ can be derived using simple

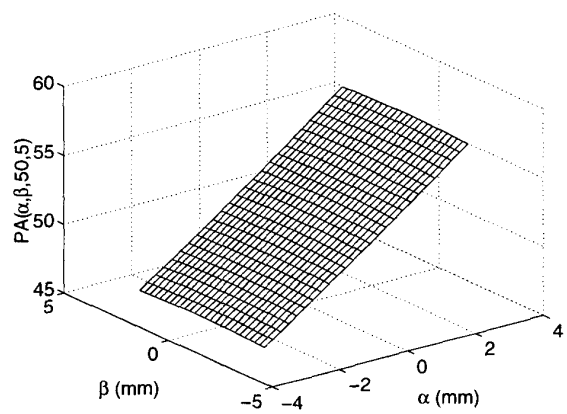
geometry and is given by

$$y = \left(\frac{p\beta - n\gamma}{\alpha p - m\gamma}\right) x + d\left(\frac{m\beta - n\alpha}{p\alpha - m\gamma}\right) \quad \text{and} \quad (1)$$

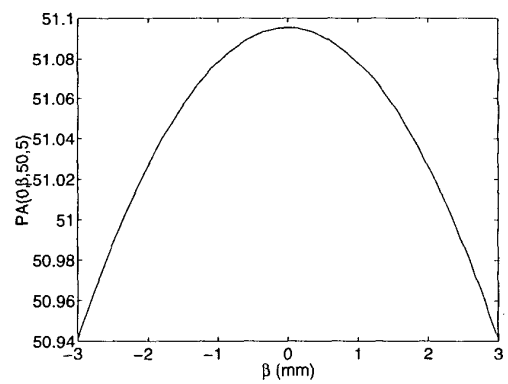
$$y = \left(\frac{p\beta + n\gamma}{\alpha p - m\gamma}\right) x + d\left(\frac{m\beta + n\alpha}{p\alpha - m\gamma}\right), \quad (2)$$

respectively. The point of intersection of the two projected lines is given by $(-\frac{d\alpha}{\gamma}, -\frac{d\beta}{\gamma})$, which clearly is the projection of the tip of the pyramid. If we assume that the triangular mirror is large enough such that third edge CB projects outside the aperture, then the two projected lines and the boundary of the aperture will bound the region carved by the set of rays that are reflected by the triangular face. If the aperture shape is known then the projected area can be determined and analyzed as a function of d and the object point which is represented by its focused image at (α, β, γ). We denote this function as $PA(\alpha, \beta, \gamma, d)$. We evaluated this function for the case of a 3-faced pyramid (Section 3) and a circular aperture of diameter 14mm. The dependence of the function for the pyramid face given by three corners (0,0,0), (m,n,p) and (m,-n,p) is shown in Fig. 4 on some of the parameters is illustrated in Fig. 5 and Fig. 6. Fig. 5(a) shows a plot of $PA(\alpha, \beta, 50mm, 5mm)$, i.e. variation in projected aperture area as a function of (α, β), for a fixed $\gamma = 50$mm and $d = 5$mm. The variation along the β axis is quite small compared to that along α axis and it has been highlighted in Fig. 5(b). Let $diffmax(\gamma, d)$ denote the difference between the maximum and minimum value of $PA(\alpha, \beta, \gamma, d)$ over entire applicable range of (α, β). This value gives a measure of spatial variation in the brightness across the sensor. A plot of the dependence of $diffmax(\gamma, d)$ is shown in Fig. 6(a). Recall that γ depends on the z-coordinate of the object point and thus any dependence of projected area on γ implies that there is a variation in projected area as a function of object depth. To highlight this dependence we plot the maximum difference between the projected area for a scene point when γ is increased by 1mm, as a function of d. This function is given by $diffmax(\gamma, d) - diffmax(\gamma + 1, d)$ and is shown in Fig. 6(b). It shows that when $d = 0$, the dependence of the projected area on the z-coordinate of object position vanishes.

The brightness of a scene point on various sensors could in general be different. This happens because the projection of any two faces of the pyramid on the aperture when viewed from (α, β, γ) could be different. We denote the projected aperture area corresponding to pyramid face i as $PA_i(\alpha, \beta, \gamma, d)$. The distributions $PA_i(\alpha, \beta, 50mm, 5mm)$ for the case of a circular aperture and the 3-faced pyramid are shown in Fig. 7. Each distribution is shown in the form of an image, where the gray level represents the value of the projected area. The range of projected area over the three sensors has been scaled between 0-230 for display. The figure shows that the gray level at corresponding points in the three images are different. However, the mean value of the projected area over any sensor is the same, which was expected because of the rotational symmetry between the three mirror-sensor pairs about the optical axis.



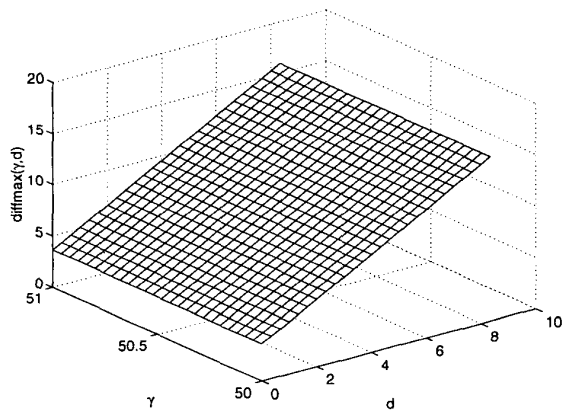
(a)



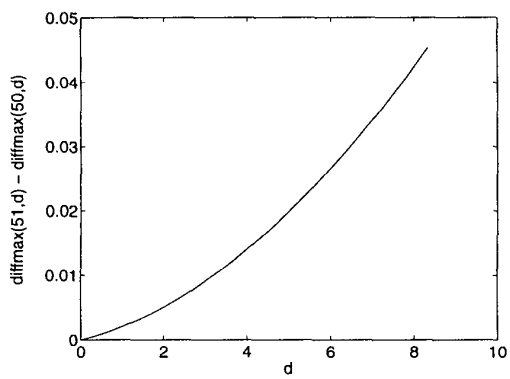
(b)

Figure 5. (a) The projected area PA, as a function of location on the sensor (α, β), for fixed $\gamma = 50$mm and $d = 5$mm; (b) The variation in the projected area highlighted along the β axis, for $\alpha = 0$.

The analysis can be easily extended to the case when the pyramid has a non-zero offset from the optical axis, i.e. medial axis is still parallel to the optical axis, but the tip no longer lies on the axis. Note in such a case, besides the spatial brightness variation across the sensor, there will also be a difference in the amount of light that reaches different sensors. As an illustration



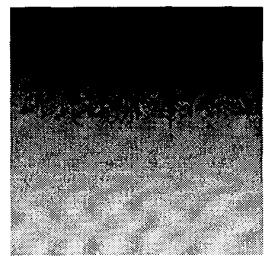
(a)



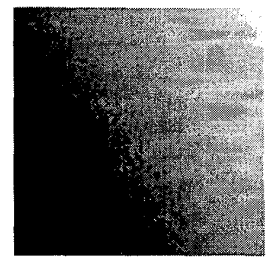
(b)

Figure 6. (a) The dependence of the function $diffmax$ on γ and d. (b) A plot of the maximum difference between the projected area for a scene point when γ is increased from 50mm to 51mm, as a function of d.

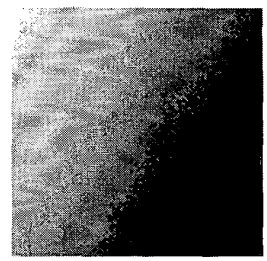
consider a simple two-way mirror split case shown in Fig. 8, if the tip of the mirror is not in the center, unequal areas of aperture will contribute light to different sensors. This can be exploited to obtain different degrees of exposures of different sensors, e.g. for high dynamic range imaging. For example, the distributions of PA_i for the case of a 3-faced pyramid with an exit pupil of diameter 14mm and offsets 2.6mm and 1.2mm along the x and y directions, respectively are shown in Fig. 9. The mean projected areas for the three sensors are 86.20, 43.58 and 23.58, respectively. Any desired proportions of mean projected areas at different sensors and hence equivalent exposure settings can be obtained by selecting suitable x and y offsets.



Range: 53.15 – 49.51
Mean: 51.30
(a)



Range: 53.84 – 48.92
Mean: 51.30
(b)



Range: 53.84 – 48.92
Mean: 51.30
(c)

Figure 7. Images showing $PA_i(\alpha, \beta, 50, 5)$, with the gray level representing the scaled value of PA_i. Scaling is performed by mapping the range of PA, i.e. 48.92 - 53.84 onto the gray level range of 0 - 230. The horizontal and vertical axes are α and β, respectively both with range $(-3, 3)$ mm.

4. High dynamic range imaging

We present two camera designs for performing high dynamic range imaging. In the first, the pyramid is aligned with the optical axis which ensures that each of the split beams has the same brightness. The sensors are placed perpendicular to the split optical axes such that the optical path length to each of the sensors is identical, which ensures that the images obtained from the multiple sensors are geometrically identical. The images could have different brightness variations as discussed in previous sections. In order to get images which do not lose sensitivity to desired parts of the illumination range, either the sensors can be preset to different exposure settings or filters with different transmittances can be introduced just before the sensors. If filters are used, then there is significant amount of light that gets wasted. On the other hand presetting exposure settings would result in different sensors integrating the scene over different durations. This can cause a

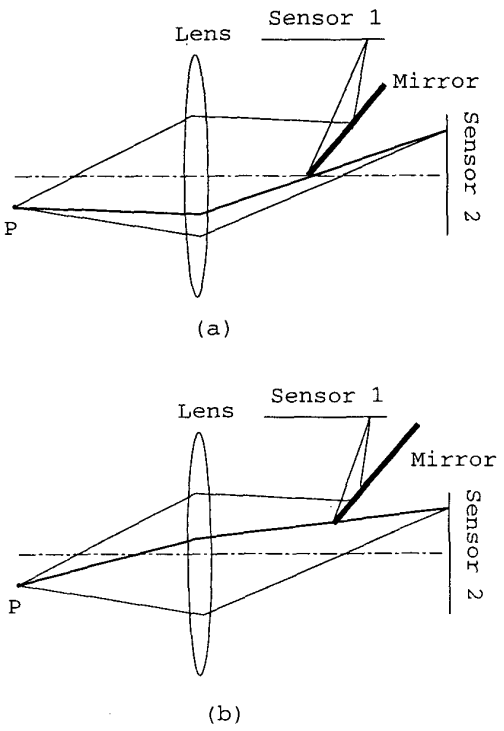


Figure 8. Dependence of the projected area on lateral position of the mirror. (a) Tip of the mirror is on the optical axis, (b) Tip is offset from the optical axis. The amount of aperture that contributes light to any sensor is widely different in the two situations.

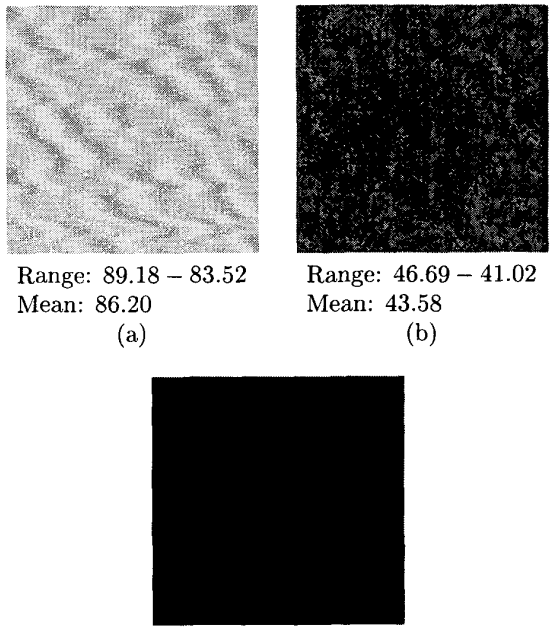


Figure 9. Images of the distribution of $PA_i(\alpha, \beta, 50, 5)$ for the case when the pyramid is offset from the optical axis by 2.6mm and 1.2mm in x and y directions, respectively. The graylevel in the image represents the scaled value of PA, such that the maximum value (i.e. $89.18mm^2$) is represented by gray level 200. The horizontal and vertical axes are α and β, respectively both with range $(-3, 3)$ mm.

problem for dynamic scenes, for e.g. those with moving objects or changing illumination (flourescent lighting), which could result in different scenes being imaged by different sensors. In the second design, the pyramid is placed parallel to the optical axis but with an offset. This design ensures different sensors get different amounts of light and does not require additional filters or a mechanism to suitably set the exposure settings, thereby also overcoming their associated drawbacks. Irrespective of the camera design, the method used to compose a high dynamic range image is the same.

There are three main steps to composing the high dynamic range image. First, we transform the recorded intensities into the actual sensor irradiance values. This mapping can be obtained using radiometric calibration techniques applicable for normal cameras [3, 6, 10]. Second, since the irradiance at corresponding points on different sensors can be different, we need a correction factor to represent a scene point by a unique value independent of the sensor where it gets imaged. This factor is spatially variant and it is different for different sensors. If we select the image from sensor 1 (the brightest image) as reference then the correction factor corresponding to location (α, β) of sensor i is given by $\frac{PA_1(\alpha,\beta,d,F)E_1}{PA_i(\alpha,\beta,d,F)E_i}$, where E_i represents the exposure time or the transmittance of the filter used on sensor i. In the zero-offset camera design E_i varies with i, while in the non-zero offset camera design, E_i is a constant and the effect is absorbed in $PA()$ itself. The third and the last step is fusing the intensity transformed images into a single high dynamic range mosaic. For every location (α, β) we have a set of n intensity values one from each of the images. We discard the values from images in which those location were either saturated or clipped. Since, the undiscarded values may be noisy, we average

them to obtain the final value for location (α, β).

The minimum dynamic range of the resulting image is given by [11]

$$DR = 20\log 2^k + 20\log\left(min_{\alpha,\beta}\left(\frac{PA_1(\alpha,\beta,d,F)}{PA_i(\alpha,\beta,d,F)}\right)\right), \quad (3)$$

where k is the number of bits used to sample the intensity information on the sensor.

5. Experiments

We designed a prototype of the high dynamic range camera using a corner of a cube as a 3-faced pyramid and three sensors. Such glass cube corners are commercially available and marketed as solid retroreflectors. The triangular surfaces can be coated with a metallic coating such as aluminium to obtain the three desired reflective surfaces. Given the height of the pyramid to be h, the radius of the circum-circle of the triangular base is $h\sqrt{(2)}$. Further, the triangular base is equilateral and this can be used to determine coordinates of the three corners and thus the distributions of light on the three sensors. The pyramid was aligned with the optical axis and sensors used were Sony monochrome board cameras CCB-ME37. We used thin-film neutral density filters with transmittances 1, 0.5 and 0.25 in front of the sensors to obtain images capturing different parts of the illumination range. The frame grabber used was Matrox multichannel board capable of synchronizing and capturing three channels simultaneously. A photograph of the final camera is shown in Fig. 10. The lens used for this camera was specially designed with the aperture located at its rear and the pyramid was placed such that its tip is in the center of the aperture, thus ensuring $d = 0$. The positions of the sensors and the pyramid were carefully calibrated to ensure that all the sensors were normal to the split optical axes, equidistant from the tip of the pyramid and images from all sensors overlayed exactly on top of each other.

A sample set of three images of an outdoor parking lot obtained from the camera is shown in Fig. 11(a-c) and the constructed high dynamic range image is shown in Fig. 11(d). The intensity range has been compressed to 256 graylevel using a nonlinear mapping for display purposes. A subsequence of high dynamic images from the video sequence obtained by our camera is shown in Fig. 12.

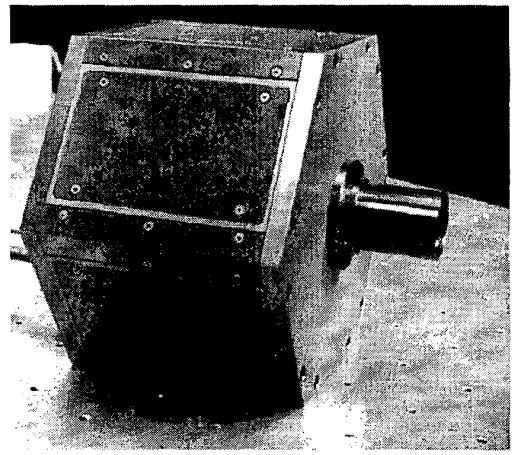


Figure 10. A prototype of the high dynamic range camera based on 3-faced mirror pyramid

References

[1] V. Brajovic and T. Kanade. A sorting image sensor: An example of massively parallel intensity-to-time processing for low latency computational sensors. In *IEEE Conference on Robotics and Automation*, pages 1638–1643, April 1996.

[2] P. J. Burt and R. J. Kolczynski. Enhanced image capture through fusion. In *International Conference on Computer Vision*, pages 173–182, 1993.

[3] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of ACM, SIGGRAPH*, pages 369–378, 1997.

[4] Edmund Industrial Optics. *Optics and Optical Instruments Catalog*, 2000.

[5] R. P. Harvey. Optical beam splitter and electronic high speed camera incorporating such a beam splitter. US 5734507, United States Patent, 1998.

[6] G. E. Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, March 1994.

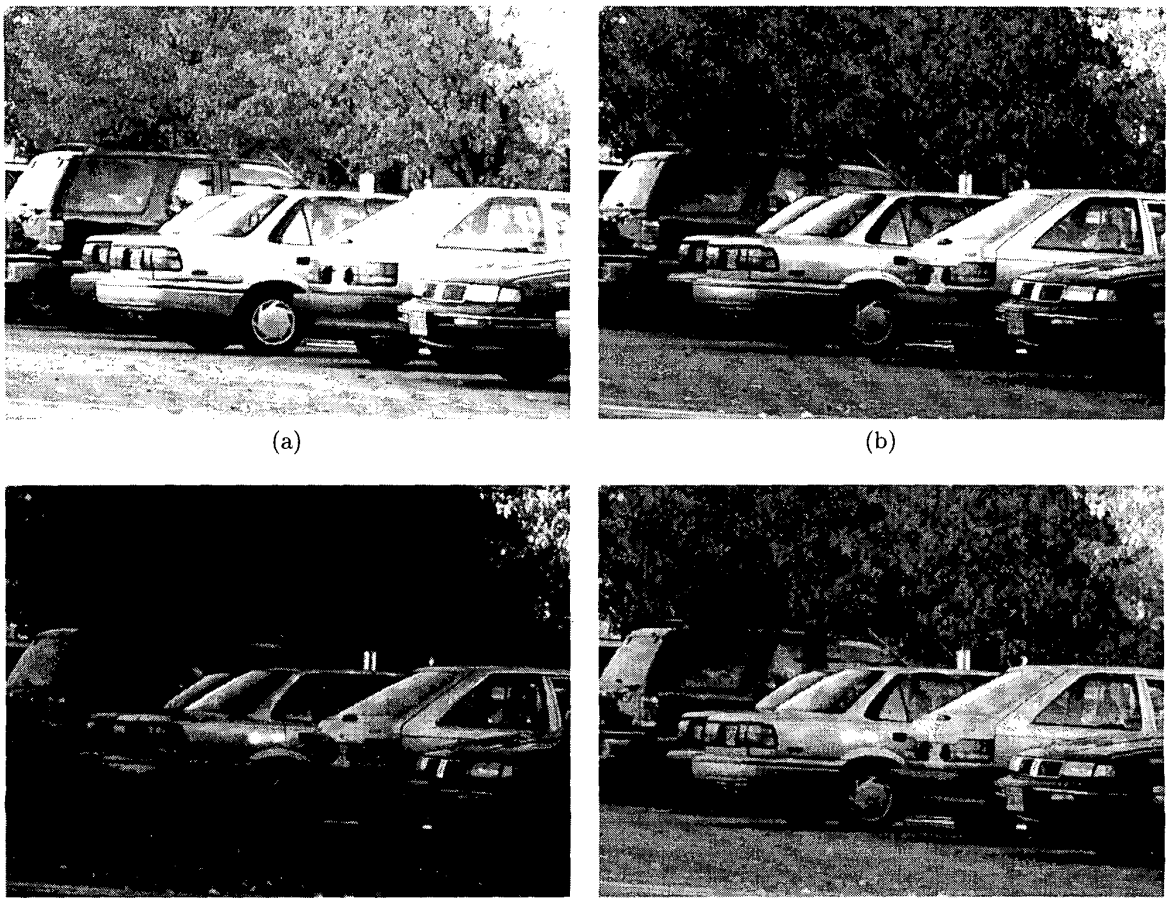
[7] R. Kingslake. *Optical System Design*. Academic Press, 1983.

[8] B. C. Madden. Extended intensity range imaging. Technical Report MS-CIS-93-96, Grasp Lab, University of Pennsylvania, 1996.

[9] S. Mann and R. W. Picard. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of IS&T 46th Annual Conference*, pages 422–428, May 1995.

[10] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In *Conference on Computer Vision and Pattern Recognition*, volume 1, pages 374–380, June 1999.

[11] S. K. Nayar and T. Mitsunaga. High dynamic range imaging: Spatially varying pixel exposures. In *Conference on Computer Vision and Pattern Recognition*, volume 1, pages 472–479, June 2000.



(a) (b) (c) (d)

Figure 11. Experimental results on high dynamic range imaging using mirrors. (a)-(c) The three images of a parking lot scene obtained by our camera employing three 8-bit sensors. The brightness of the three images are in ratios 1:2:4. (d) The high dynamic range image. The intensity range in this image has been compressed to 0-255 using nonlinear mapping for display purposes

(a) (b) (c) (d) (e)

Figure 12. Experimental results on high dynamic range imaging using mirrors. (a)-(e) Subset of the sequence of high dynamic range images of the parking lot with two people moving away from the camera.