Extended Intensity Range Imaging

Brian C. Madden

MS-CIS-93-96
GRASP LAB 366

Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA USA

Extended Intensity Range Imaging

Brian C. Madden

GRASP Laboratory
University of Pennsylvania
3401 Walnut Street, Room 301C
Philadelphia, PA 19104, USA

December 17, 1993

Abstract

A single composite image with an extended intensive range is generated by combining disjoint regions from different images of the same scene. The set of images is obtained with a charge-coupled device (CCD) set for different flux integration times. By limiting differences in the integration times so that the ranges of output pixel values overlap considerably, individual pixels are assigned the value measured at each spatial location that is in the most sensitive range where the values are both below saturation and are most precisely specified. Integration times are lengthened geometrically from a minimum where all pixel values are below saturation until all dark regions emerge from the lowest quantization level. The method is applied to an example scene and the effect the composite images have on traditional low-level imaging methods also is examined.

1 Introduction

While it is true that reflectance values range over less than two orders of magnitude (often much less), variations in patterns of illumination can distribute the information in natural scenes over luminance values having a much greater diversity. Interposition of opaque objects and concavities within them, as well as interreflections among opposing surfaces, all contribute to the variety of local light levels. The difficulty is in capturing this diversity with sufficient resolution to represent the information in an image that is acquired with a linear sensor of limited quantization resolution.

Conventional CCD cameras provide a voltage that is proportional to the irradiance on each photosensitive region (see [HealIP]). The signal to noise ratio is usually guaranteed to be smaller than the smallest quantization level over the specified operating temperature range. In critical low-light applications such as astronomy, the noise introduced by the camera electronics can be reduced by cooling; however, temperature control is not practical in many weight- and size-sensitive environments. Most commonly, digitization of a solid-state camera image results in an 8-bit (256 level) pixel representation. Higher resolution converters are available, although often at less than frame rate speeds. The effective resolution of any subsequent digitization is limited by the settling time of the conversion and the signal to noise ratio of the analog signal.

Such 8-bit/pixel images are appropriate for capturing *single surface* scenes, scenes of uniform illumination and Lambertian reflectance with no specularities or shadowed obscurations. Historically, scenes with greater intensive range were captured by altering the lens aperture [Born81] or by actively controlling the source of illumination [YiSe90]. The down side of these methods was

that they altered the optical wavefront presented to the sensor array or they altered components of the scene such as the position of shadows and specularities.

Advances in camera technology offer electronic alternatives that may close the gap between sensor sensitivity and the distribution of luminance values in the scene. Sensors have been suggested that allow the sensitivity of individual pixels to be altered [Huan91] or allow flux differentials to be directly encoded at the sensor [Ando91]. These adaptive sensors, however, do not offer a way to label the extent to which the local gain is adjusted, thus information is lost. In addition, adaptive systems can create nonmonotonicities where none exist. Intense, localized features can depress the values of the representation in the region, making a gradient appear.

The availability of cameras that allow the temporal interval over which flux is integrated in the CCD sensor to be varied now offers another alternative. With variation in temporal integration time, sensitivity of the image may be controlled without either altering the composition of the scene or distorting the optical waveform within the camera. A method is proposed here that uses this *electronic shutter* to generate a series of images of varying sensitivity and then to create a single composite image of extended intensive resolution. If the series of images varies geometrically in the gain applied in their acquisition, the resulting composite will have a compressive intensity-to-pixel value relation with the width of the quantization level approximately proportional to intensity. The existence and utility of such a compression in biological vision is described in a companion paper [Madd94] and its application to computer vision will be addressed here in a later section.

2 Extended Intensity Range Imaging

If the goal is to represent the spatial position and magnitude of all flux gradients in a scene above a criterion contrast without loss (i.e., fusion of information within overly large quantization levels or fusion of information due to saturation), then a decision needs to be made as to the allocation of computational and storage resources for the task. The method proposed here allocates representational resources so that local contrast is equally preserved in the scene independent of the absolute value of the local mean luminance. Local contrast is a measure of the difference in magnitude between two sets of pixel values weighted by a third set that is often (but not necessarily) the union of the difference sets. Weighting of individual pixels prior to pooling can be used to fine tune the differencing and scaling operations. It is important in this process, however, not to discard the absolute luminance information, nor allow its representation to be saturated so that chromatic information will be distorted.

It is proposed that the increased cost in expanding pixels from the usual 8-bit integer representation to larger integer or floating point formats (commonly a factor of 4 increase) is offset by the benefits associated with the increased retention of information. The additional information comes from images that have a different sensitivity. With all quantization scales starting at zero (no DC offset), changes in the sensitivity of a fixed number of levels trade range for resolution (see Figure 1). By weighting the pixel values in each image by their associated sensitivity, pixels acquired in different images can be combined in a manner that preserves local luminance relations even though they extend beyond the scale of a single image. When a more sensitive scale saturates, the composite response shifts to a less sensitive scale, mapping additional quantization levels to larger intensity values at proportionately coarser resolution (see Figure 2). This method is compressive in the number of quantization levels but allows the retention of the absolute level of intensity. This compressive nonlinearity should not be confused with gamma correction of displays with nonlinear voltage to luminance functions.

The distribution of intensity resolution within an extended intensity range image itself depends

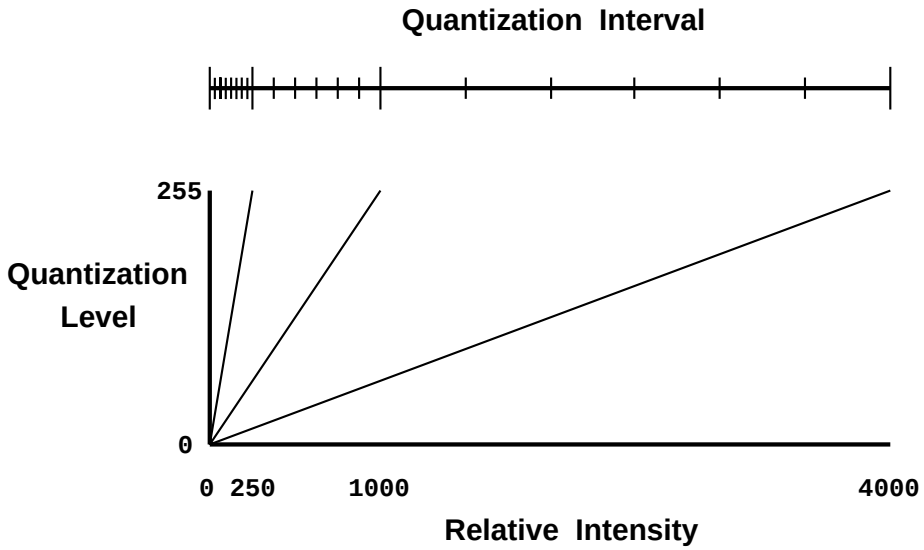

Figure 1: Distribution of Quantization Levels with Changes in Sensitivity. When sensitivity is increased, the limited quantization range saturates proportionately earlier while, at the same time, affording greater resolution to flux changes in the darker regions of the image. A decrease in sensitivity extends the range of intensity values represented without saturation at the cost of lower discrimination. The scale at the top reflects the best relative quantization available at each intensity level. The bottom graph shows the intensity/pixel map of three images obtained with gain differentials of 4.

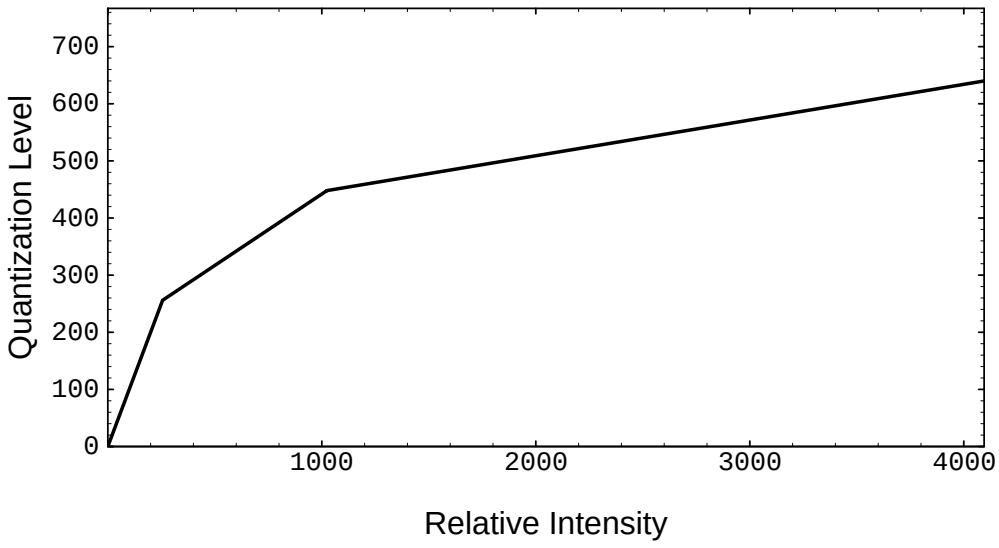

Figure 2: Combination of Linear Response Ranges. Three linear response function are combined to make a single compressive response function. The resulting compressive function allows high quantization resolution at low intensity levels without the cost of maintaining that resolution over the entire range.

on four factors: the absolute sensitivity of the camera, set by optical and electronic factors that determine the amount of flux integrated per pixel; the number of quantization levels per image (q), limited by electronic noise and digitization time; the gain differential between the images (m), and the number of images (n). These factors can be adjusted to match the requirements of the task and the scene (see [Kamg89]).

Given the criteria for the detection of the desired information in the scene (the minimum contrast), the number of component images and their relative sensitivity remain to be determined. In order to maximize the capture of the intensive gradient information, the sensitivity of the images is adjusted to match the values present in the scene. To accommodate the largest intensity values, the sensitivity of the initial image is reduced until all pixel values are less than saturation. Conversely, for lower intensities, sensitivity is raised successively until all pixel values are above zero and the quantization level is smaller than the difference required for the criterion contrast level. When the local intensity mean is at the high end of the linear scale, pixel differences correspond to smaller contrast changes than the same difference does at the low end of the scale and it is easier to determine if there is gradient information present in the region. This variation of contrast resolution is a consequence of the piecewise linear compression of quantization levels. The decisions as to what constitutes a meaningful contrast and how many pixels are enough to be considered a region with their own lighting characteristic are task dependent and may not always be determined from the image statistics alone. With a fixed number of quantization levels there is a limit to how small a gradient may be reliably extracted from an image. In the piecewise linear composite image the contrast ($\Delta I/I$) resolution represented by adjacent quantization levels goes from $1/q$ to m/q each time the sensitivity is decreased. In addition, contrast resolution degrades below this in the lowest $1/m$ portion of the most sensitive image since this range is not overlapped by a further reduction of the quantization interval.

2.1 Reduction in Quantization

The benefit of a sliding scale of sensitivity that is made possible by the use of local contrast as the measure of information in the image is considerable. The advantage of scaled compression versus full representation of the entire range at the most sensitive quantization is:

$$relative\ savings = \frac{m^n}{m + (m-1)(n-1)} \tag{1}$$

where m is the differential gain between images and n is the number of images in the sensitivity series.

Note that the reduction in the quantization requirement is independent of both the number of available quantization levels in the sensor as well as the absolute sensitivity. A combination of just three sensitivity levels with a gain differential of 4, results in a more than 6 to 1 savings. In another example, if 8 images were acquired, each a factor of 4 in sensitivity, the resulting savings ratio would be in the thousands (see Figure 3).

2.2 Merging of Images

Starting with the least sensitive image as a base, nonsaturated portions of successively more sensitive images are masked off and overlaid until each pixel in the resulting mosaic has the best available resolution. As sensitivity increases, more and more pixels emerge from the lowest quantization level and move up the available quantization scale becoming more and more precisely defined.

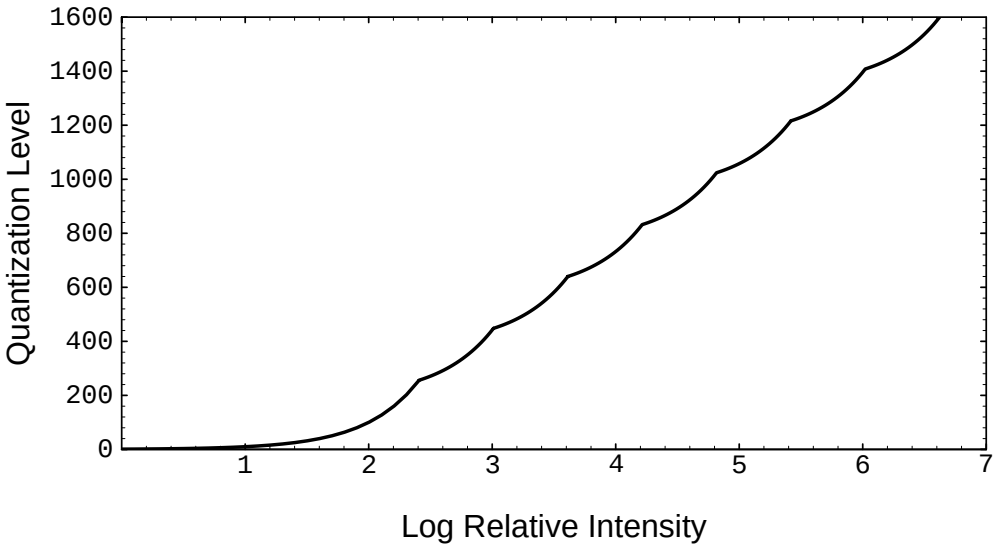

Figure 3: Combined Response Function. The combined response function made up of portions of 8 different sensitivities ($\Delta gain = 4$) has a slightly scalloped appearance due to its piecewise linearity and a semilogarithmic presentation. There is a savings of over 2600:1 in this representation (1600 quantization steps versus over 4 million at full resolution).

Ultimately, the intensity of a given pixel will pass beyond the range of conversion and the response will be saturated. If the images are acquired in order of increasing sensitivity, there is no need to store them since the masking and replacement operations may proceed sequentially.

2.3 Gain Measurement

The proper scaling and combination of pixel values from different images depends critically on an accurate measure of the differential gain. In the discussion to here, all intensity to pixel conversions were assumed to be linear and of the correct magnitude. This is seldom the case (see Section 4). Fortunately it is possible to obtain a measure of both the gain and the conversion linearity from the intensity distribution of two images alone (see Figure 4). Discounting noise and other conversion anomalies, a pixel value at a given spatial location in a less sensitive image can be paired with one of m values in the next more sensitive image. When this pairing operation is done for all corresponding pixels in two images, the resulting two-dimensional histogram of pixel values will define a straight line with a slope (as well as a vertical thickness) that corresponds to the relative gain between the two images. Care should be taken in obtaining this empirical estimate since small differences in the less sensitive images are amplified by the cascaded gains in the formation of the composite image creating contours where none should exist. Small errors in the assumed gain or linearity of the transformation could prevent the seamless combination of pixels.

3 Example

An indoor scene was created with as much diversity as possible to test the extended intensity range algorithm. The collage was arranged to contain concavities, obscurations, interreflections,

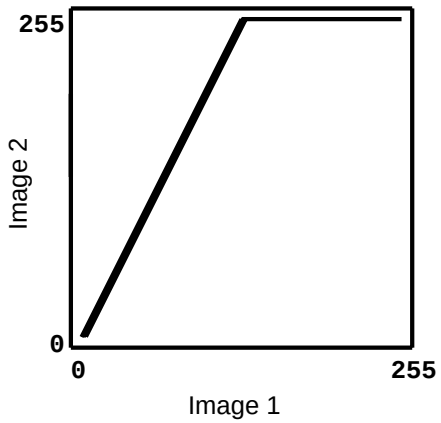

Figure 4: Determination of Effective Gain Differences. A 2-dimensional histogram of corresponding pixel values in two images of differing sensitivity will fall in a straight line with a slope that corresponds to the gain differential. Pixels in the less sensitive image (plotted on the abscissa) that have values beyond the range of the more sensitive image will be matched with saturated responses and will fall in a horizontal line at 255. The slope of 2 corresponds to a doubling of the sensitivity.

and a range of surfaces, matte and glossy, opaque and transparent - all at various orientations with respect to the principal illumination (1000 watts of light positioned to the right of the scene).

3.1 Methods

The images were obtained with a CCD camera (SONY XC-77RR) and a Cosmicar 25 mm lens with the manual aperture fully open (1:1.8). The camera's sensor array provided 93% coverage and 0.8 lux sensitivity as well as a 768 by 493 usable pixel matrix that was read out at a 14.3 MHz scanning rate [SONY90]. The temporal integration time of the sensor was changed manually on the camera adaptor box. While it is possible to electronically control the integration time, such an interface was not assembled for this test. The integration times went from 4 microseconds to 32 milliseconds (mostly in factors of 4). The camera output was digitized by a Data Translation framegrabber (DT1451) which has a 10 MHz resampling rate. The mismatch between the scanning and resampling rates resulted in each output pixel being a combination of the voltages from 2 or 3 sensor cells along a raster line. The underlying cell substrate for a given output pixel varied from image to image (and line to line) due to a $+/-50$ nanosecond range of jitter offset.

Much of the image manipulation was done with the HIPS Image Processing Package (SharpImage Software) and the XV Interactive Image Display System.

3.2 Results

A series of 8 images with varying sensitivity were acquired of the same scene (see Figure 5). At the lowest sensitivity, only the interior of the lamp and the outline of the cardboard box can be seen. As sensitivity increases, more detail emerges but at the cost of losing some information to saturation. Beginning as soon as the third image, information on the shape of the light bulb is lost. By the fourth image, the writing and all detail of the interior of the lamp are gone and the specularity on the rim of the cup is saturating. In the fifth image, much of the wall and part of the box are saturating but the writing on the cap, slide and box is discernible. In the sixth image, the corner of the cardboard box is lost and is fused with the wall but the stereo slideholder begins to emerge from the shadow inside the box. In the final two images, more features are eroded by

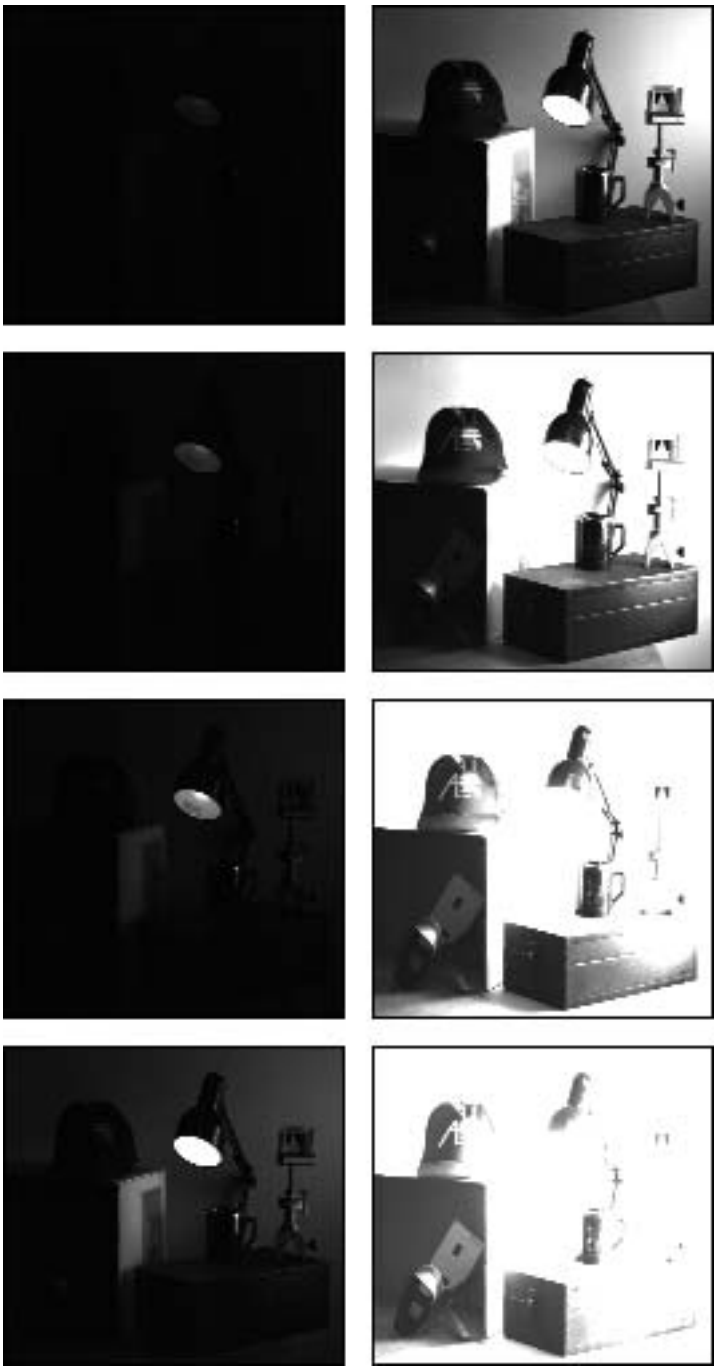

Figure 5: Variation of Integration Time. This series of images reflects changes in the temporal integration time of the CCD array by a factor of over 8,000. The shortest time was 4 microseconds; the longest, 32 milliseconds. This variation can only be approximated by the figure which has a range of less than 25 to 1.

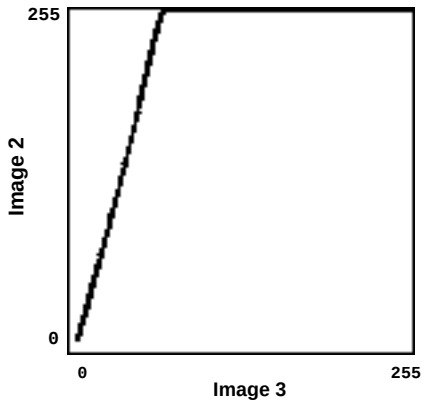

Figure 6: 2D Histogram from Levels 2 and 3 of Example. The best linear fit to the histogram was 3.98 (with an intercept of -10.89). The graph represents all occurrences of 10 or more correspondences. Note the thickening of the slanted portion of the graph. This vertical spread is a consequence of the mapping of several pixel values in the image with the greater gain to a single pixel value in the image with the lower gain. The thickness of the plot increases in proportion to the differential gain.

saturation while the objects in deep shadow become more defined.

According to the settings provided by SONY, the differential gains were all 4, except between the two most sensitive images when it was 2. The ratio of the luminance represented by the highest pixel of the least sensitive image to the lowest pixel of the most sensitive image is more than 10^6. In order to test these settings, 2D histograms were obtained for each adjacent pair of images in the sensitivity series. The obtained linear fits (starting from the most sensitive image pair) for the example images were: 0:1 (2.17, -8.03), 1:2 (3.80, -9.00), 2:3 (3.98, -10.89), 3:4 (4.17, -12.87), 4:5 (3.67/3.75, -12.35/-12.75), 5:6 (2.78/3.26, -5.34/-6.78), 6:7 (2.13/2.71, -3.39/-5.13) (see Figure 6). Similar results were obtained for a second camera. The observed gain differentials for the less sensitive images were less than that expected from the camera settings. At the shorter flux integration times, the fixed 5.04 microsecond difference in the collection time between the even and odd fields becomes appreciable. This electronic irritation can be eliminated by fitting the odd and even lines separately in the gain calculations, hence the double estimates for the slope and intercept values of the lower sensitivity images (see Section 4).

Using the obtained best linear fits to match the intensities of different sensitivities, a series of overlays produced a single floating point representation of the full range of intensities in the scene. Since the selection of sensitivity is done on a pixel by pixel basis, little spatial information is distorted and since each quantization level is weighted by the associated sensitivity, little intensive information is lost. The percentage of pixels in the final composite image from least to most sensitive image was: 0.0, 0.05, 0.9, 8.5, 32.2, 22.1, 9.6 and 26.6 (total pixels: 245,760) (see Figure 7). Note that the two most sensitive images are only a factor of 2 away from one another, thus the reduced percentage of contributed pixels in the second most sensitive image.

The pixel values in this floating point representation of the indoor scene range over 2500:1, up to a a value of 112,123 (see Figure 8). This difference is appreciably more than that available from a single 8-bit image (and it is likely an underestimate since the veiling luminance is an additive distortion and disproportionately increases the lowest pixel values, see Section 4). The distribution of pixel values is very skewed (90% of the pixel values reside in the lower 5% of the range). This is certainly due to the specular component of the image. Even larger ranges of pixel intensities can be expected from daylight scenes where more candlepower is available for distribution across

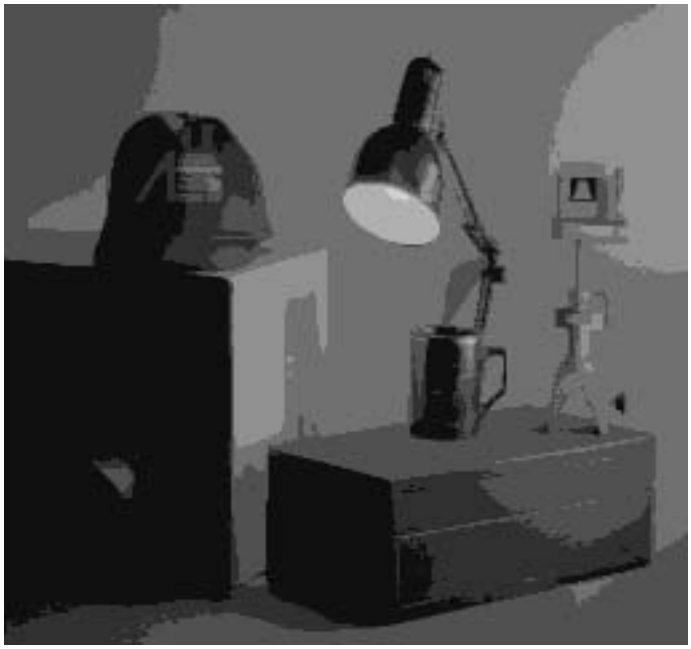

Figure 7: Contribution of Different Sensitivities to Composite Image. Reflecting the large range of intensities, the composite image is a mosaic of regions of varying quantization resolution. The grey scale in this image is proportional to the size of the quantization interval in the respective source images. In those portions of the image with higher luminance, sensitivity is decreased in order to extend the range of representation. In the darker portions of the image, the decreased range requirements allow greater resolution.

the scene (resulting in perhaps a less skewed distribution of pixel values). If the specularities were not of interest in the example and their information could be discarded (distorted), adequate representation of the intrinsic surface contrast would still require a series of several images. Once the scene is represented in a single image having this expanded resolution, arbitrary extraction of slices of the image can be taken, either multiplicative or subtractive, or both. The inclusion of all scene information on a single scale and in a single image makes locally adaptive transformations especially convenient.

3.3 Application of Low-Level Operators

With extended intensity range images, low-level operators need not be stymied by edges vanishing into shadows or specularities. For the most part, such disappearances have been due to quantization or range limitations of the digitization process. To demonstrate the utility of the extended representation, a simple bandpass edge detector was convolved with the composite image of the example scene (see Figure 9). A 7 by 7 difference of Gaussians (DOG) operator was used to obtain an estimate of the local edge gradients. The magnitude of the filter response was normalized by the mean intensity of the pixels under the operator. The result is a dipole at each step function within the bandpass of the filter that has a magnitude that is proportional to the local contrast. This extremely simple adaptive operator is able to extract the patterns on the stereo slide in the shadow as well as on the interior of the lamp. Other, more spatially diffuse changes are ignored (e.g., the intensities on the back wall vary over a 7:1 range). It is quite likely that even this performance would be improved on by more sophisticated filters or a better measure of local contrast (especially in regions of large and rapid intensity change).

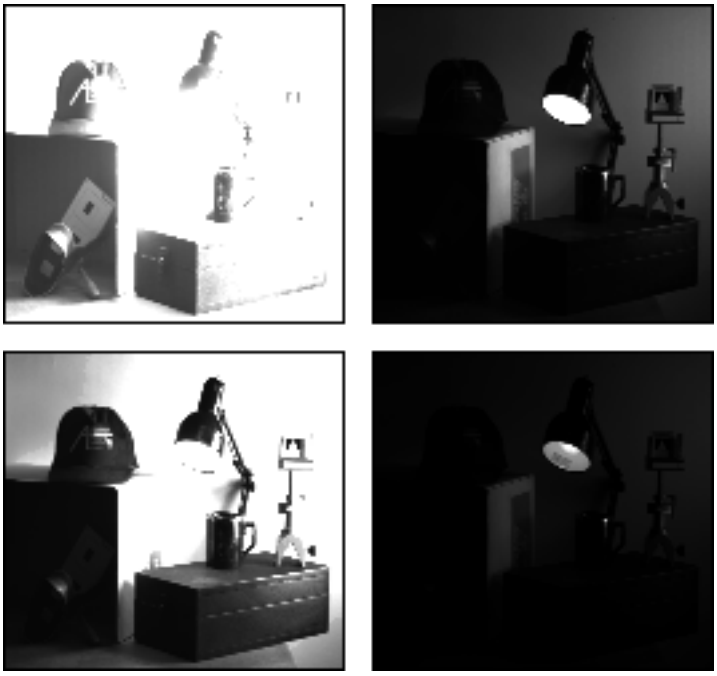

Figure 8: Slices of the Scaled Image. Different views of the scaled image were obtained by dividing the floating point representation by 1, 8, 64 and 256, respectively. All of the nonsaturated grey values in the image at the upper left are compressed into the lowest quantization level of the image at the lower right. Conversely, all of the pixels greater than 0 in the lower right are saturated (255) in the upper left. Even when compressed by 256 (lower right), there are still saturated pixels in this indoor scene (the luminous source in the lamp and the specular component on the rim of the cup). The pixel values in this floating point representation range from 44 to 112,123.

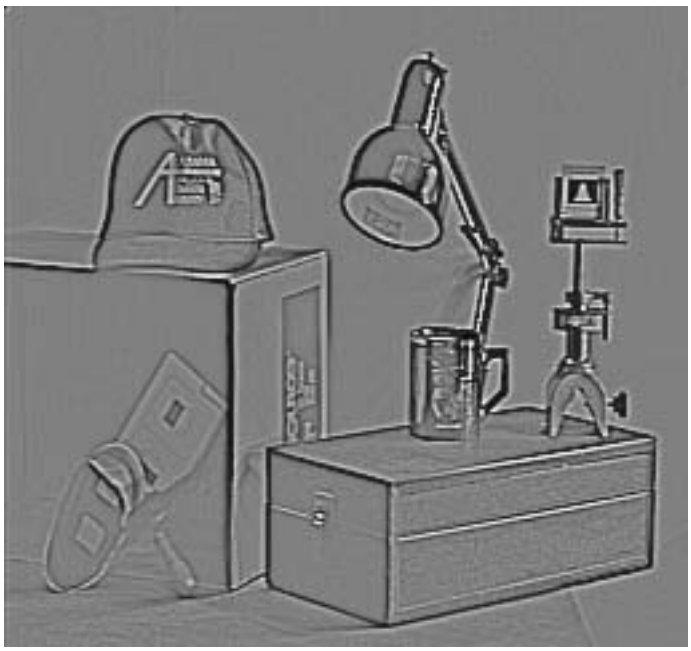

Figure 9: Normalized Edges in the Scaled Image. Intensity gradients in the scaled image were obtained by convolving the image with a difference of Gaussians operator and normalizing the resulting values by the local mean luminance of the pixels under the operator. In this example, the operator was 7 by 7 and the ratio of the two Gaussians was 1.6:1.

The majority of edge junctions in natural scenes involve edges that have similar absolute gradient magnitudes. Even these similar edges, when intersecting, present problems to segmenting algorithms that attempt to form closed contours. In the past, algorithms have only had to deal with differences in edge magnitudes that could be spanned by an 8-bit representation - at most a 0 to 1 ridge running orthogonally into a 0 to 255 ridge. Both these edges potentially are equally *important* in that they have the same local contrast, varying only in mean luminance. While the one pixel step may be a minor variation in the dark side of the larger step (under uniform illumination), it may also be due to a large reflectance difference itself (but in deep shadow). In cases where the local contrast criterion is applied, the extended intensity range representation raises the ante of this task by requiring the simultaneous processing of much greater edge differences.

3.4 Representation of Specularities

Although the presentation heretofore has focused on greyscale images, they can be thought of also as one channel of a color image. Given the dimensionality of of human color matches, three color channels are required to represent the perceived hue corresponding to the broad distribution of wavelengths reflected from a point in the scene. The linear combination of wavelengths within a channel and the linear combination of the three channels is an adequate first-order approximation of human color processing. One fundamental problem in color imaging (and human vision) is the assessment of the intrinsic spectral reflectance of the surfaces in a scene in the presence of an unknown illuminant. The problem is underconstrained. Not only do the potential degrees of freedom of the illuminant and reflectance functions dwarf the degrees of freedom available to represent the chromaticity of each pixel, the intrinsic reflectance is contaminated by an additive specular surface reflection complicating the extraction of object properties [Shaf85, Nov92b]. By examining the apparent wavelength of the specular reflections in the image, it is possible to extract an estimate of the nature of the illuminant [DZmu86, LeeH86, Gers86, Klin87]. In the current context, performance on the extraction task is degraded if any of the color channels is saturated (see [Klin87]). As the intensity is increased for a given camera configuration, saturation will result in all specularities, and thus all illuminants, appearing white. While the range of intensity spanned by an 8-bit representation may be increased in order to reduce the amount of specularity saturation, it is done at the cost of reduced local contrast resolution. Adequate quantization resolution is not only required to discern objects under varied illumination conditions but it is also necessary for the accurate determination of the chromatic distribution of the highlight so that the intensity of the illuminant may be estimated [Nov92a]. With the intensity and chromaticity of the illuminant known, the albedo of the surface may be factored out of the distribution of light reflected from the object. It is only when the specular component can be accurately determined that the intrinsic reflectance of real objects under different illumination conditions can be tested against the various theoretical predictions (e.g., [Beck63, Torr67, Cook81, Naya91, Wolf92]).

Other studies have proposed that specular components can play a role in the assessment of form in machine vision [Thri83, Heal88, Park90, Naya93] and can contribute to human vision as well [Blak90]. The local geometric information available in highlights in a single image affects the perception of surface curvature. In glossy and metallic surfaces, especially in areas of high curvature, the specularity tends to exceed the other reflection components by far. The intensive changes that these algorithms rely upon for local structural cues are an order of magnitude or more larger than the intrinsic properties. Constraints on the relation between the specularity distribution and surface curvature can be extended through binocular viewing or motion of the observer [Blak88, Ziss89, Park90, LeeS92] or of the illuminant [Ikeu81]. Again, performance on these tasks is degraded by saturation of the representation of the highlight. Extending the intensity range so that the contrast of the matte component will retain sufficient resolution and yet allow

the full amplitude of the specularity to be represented will reduce or eliminate many of the errors.

In the example, a localized specularity can be seen on the lip of the cup. At a sensitivity level where the stereo viewer is still fused together in the low quantization levels, the specularity on the cup is already saturating (bottom left image, figure 5). Only at successively lower sensitivities is the distribution of intensities on the lip undistorted by the upper limit of the camera's response range. While this example is due to the polished ceramic surface of a man-made object, equivalent degrees of specularity may be encountered in natural scenes where silicates, bodies of water or moisture on surfaces could approximate this degree of glossiness.

4 Practical Considerations

As with the dancing dog, the surprise is not how well it works, but that it works at all.

It is of interest to examine how vulnerable the merging algorithm is to limitations in both the environment and the equipment. The images in the example were taken in an open terminal room with uncalibrated cameras and framegrabber in as-is-from-factory condition. Multiple banks of fluorescent ceiling lights complimented the incandescent floodlights to illuminate the scene. The poorly balanced three-phase electrical system in the lab building added to the general variability of the West Philadelphia power grid.

There are several consequences of operating a CCD camera at its greatest sensitivity while acquiring images of moderate to high intensity. Most of these effects are not apparent when the cameras are operated in a more conventional manner. The first of these is lens flare. Reflections of the bright sources illuminating the scene bounce off of the internal supports of the optical components and end up providing an additive veiling luminance. The lowest 44 quantization levels of the most sensitive image in the example were unused. The number of unused levels decreased in proportion to the decreases in sensitivity (also taking into account a fixed digitization offset of between 2 and 3 that can be seen in the linear fits of the 2D pixel value histograms). In order of decreasing image sensitivity, the number of unused levels are: 44, 23, 8, 3, 3, 2, 2, 2. Some of this veiling luminance (which need not be uniform across the sensor array) can be eliminated if care is used in positioning the camera with respect to the major sources of illumination. In addition, external baffling can be added to the lens to reduce the off-axis scatter.

Another source of image distortion is due to the sensor electronics. When the 2D histogram of pixel intensities is expanded to include the infrequent correspondences (< 10), there appears a *comet's tail* of matches between the linear ridge caused by the gain differential and the major diagonal (see Figure 10). Once the best linear fit to the ridge of the 2D histogram is known, it is then possible to localize within the images all occurrences that do not fall within the m pixel width of the linear ridge. When this is done, all of the *outliers* were found to fall on a raster line next to a large and rapid bright-to-dark or dark-to-bright transition. Distortion of greyscale values near highly saturated ones reflect the inability of the sensor electronics to recover sufficiently fast during the shifting of large voltages out of the array (see [Witt88]). The presence of this nonlinearity is evinced by the appearance under some circumstances of pixel values *lower* than the veiling luminance should allow.

In addition to the step response nonlinearity, any variability in the sampling of the camera output along a raster line relative to the scanning rate of the sensor would alter the manner in which an output pixel weights the voltage produced by the sensor cells (see [Beye90]). In the current setup, this variability is exacerbated by the mismatch between the scanning and sampling rates where 2 to 3 cells influence a pixel value. The consequence of any phase jitter is amplified by

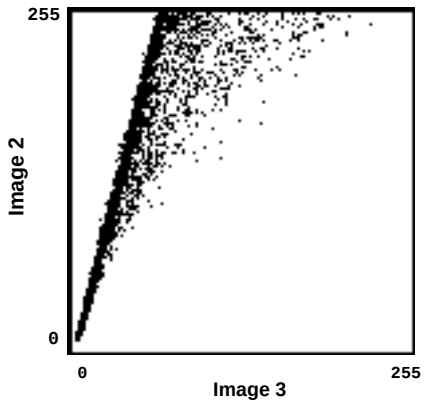

Figure 10: 2D Histogram from Levels 2 and 3 with Nonlinearities. Transitions from saturated pixels to unsaturated pixels result in nonlinearities. These nonlinearities are not reflective of the gain differential between the two images and thus do not fall on the sloping line. When this *comet's tail* is localized on the image they all fall laterally on a raster line next to a bright-to-dark or dark-to-bright transition. As in Figure 6, the pixel values of the less sensitive image are plotted on the abscissa.

the magnitude of the flux gradient; however, the distribution of the *comet's tail* almost entirely to one side of the ridge argues against jitter as the major source of intensity error in the images.

Because the large aperture used in the example resulted in a very small depth of field, relatively few of the contours were in exact focus. It would be extremely difficult, therefore, to discriminate distortions due to limitations of temporal response or phase stability from gradations in the objects or illumination or from optical blurring. Nonetheless, the localization of significant distortion to regions of steep gradients allowed a remedy. In the formation of the composite image, nonsaturated pixels (< 255) in the next more sensitive image were used to replace the spatially corresponding pixels in the unfinished composite. Prior to replacement, the greyscale portion of the image being added was eroded by two pixels to either side of any saturated pixel value. In the example, all but 5 nonsaturated pixels (out of 245,760 total image pixels) eroded by this method had corresponding pixels available in less sensitive images that could be used to take their place. The replacement process is, in large part, a trade between the error introduced by the nonlinear distortion and an error introduced by the increase in the quantization interval of the less sensitive images. In the extreme this process produced, at regions of steep intensity gradients in the example, strings of adjacent pixels that all come from different images - the edge of the cardboard box (5 in a row) and the specularity on the lip of the cup (6 in a column).

As the image sensitivity decreases and image capture enters a more conventional regime, these effects diminish (and others become more apparent). With CCD cameras there is often a smearing phenomenon that creates vertical bands in the image whenever there is a bright object in the field, especially one that reflects or emits long wavelength (infrared) light. These photons generate charges deep within the photosensors and the effect is cascaded along a column because of the interline-transfer of the vertical shift registers. Small, single pixel level effects can be seen even at the shortest integration times. This effect is most perceptually apparent at low pixel levels where a single quantization level constitutes a proportionately much greater proportion of the total luminance of a given pixel. In addition, at low sensitivity (short integration times), there is a constant temporal difference between the flux integration times of the even and odd fields. The 5 microsecond increase in integration is significant when the integration time itself is only tens of microseconds or less. The result is an interline gain differential that creates two linear distributions

in the 2D histograms. It is possible to eliminate most of the effect by fitting the odd and even lines separately.

Additional nonlinearities will introduce minor distortions of the linear distribution in the 2D histograms. For example, there is often a deviation of the ridge in the 2D histogram at the high values. This error may be due to an electronic *feature* such as white clipping. It is difficult to accurately partition the multiple sources of error between the images when distortions are present in both. As a practical matter, when significant distortions are present, it is probably best to fit the 2D distribution with a supralinear function in an attempt to maintain a seamless match between images of different sensitivity. For most applications, the introduction of a contour where none existed is probably much more damaging than the distortion of the intensity assigned to a uniform region.

5 Discussion

Natural scenes contain a diversity of lighting conditions which must be accommodated in the image representation if structural information about the scene is not to be lost because of the chance state of illumination. Often, complicated scenes are processed piecemeal, in regions of uniform illumination. Still more often, extrema are truncated by Procrustean digitization methods. The method proposed here is more accommodating. For the cost of acquiring a series of images with increasing sensitivity and then masking the nonsaturated portions so they can be sequentially superimposed, arbitrarily diverse scenes can be represented in a single image. Creation of the composite images is simplified by the self-calibrating nature of the relation between two versions of the same scene taken over different integration times. The preservation of specularity and edge information are examples of what can be accomplished with a relatively small increase in format.

The combination of new camera technology and the continuing reduction in memory costs make the acquisition of extended intensity range images an increasingly attractive alternative. The time will come when the reduced costs will make 8-bit representation the image equivalent of the black and white television.

5.1 Biological Analogs

How is this problem solved in biological vision? In human retinae, the large range (10^{10}:1) of intensities in the external world to which the visual system differentially responds is accommodated by adaptive nonlinearities. One class of photoreceptors, the cones, mediate most of the sensitivity in photopic (daylight) vision. Cones are thought to exhibit at least three mechanisms of adaptation [Vale83]. At extremely high levels of illumination, sensitivity is reduced by a proportional depletion of available photopigment. The analogy here to CCDs, however, is weak. Even photon densities sufficiently large to cause damage to the sensor do not lack for sites to generate electron-hole pairs. A second adaptation mechanism is the multiplicative shift in the cone response function with changes in mean luminance. In biological systems this shift is a way to respond to ever larger intensities while constrained by a fixed output range. Absolute intensity information is sacrificed while relative information (contrast) is preserved. If adaptation is incomplete, a measure of the mean can be maintained in the encoding by shifting the operating point as luminance increases, altering the balance between increment and decrement response range. The analogous constraints do not apply here either. By combining quantizations of the scene of different sensitivities, it is possible to effectively extend the output range of the initial point-intensity transform.

The third mechanism of adaptation in the cone is instantaneous response compression. Viewed

on linear coordinates, the Naka-Rushton intensity voltage relation is increasingly compressive:

$$R = \frac{I^n}{I^n + \sigma^n} \tag{2}$$

where σ is the luminance level that elicits a relative response of 1/2 and n alters the slope of the function.

While even a 100% contrast sinewave is subjected to only moderate distortion (26% second harmonic content) by this transform, it is the extreme spread of the top third of the response range that is exceptional (see [Madd94]). It is this component of the cone adaptation mechanism that is useful in encoding natural scenes. The shape of the curve in Figure 2 is very similar to Equation (2). Certainly, the visual system does not allow a third of the response range to go to waste. The spreading of response increments over ever larger intensity ranges allows both biological and computer vision systems to generate proportional responses to local contrast as well as to contain the sometimes extreme demands of specularities.

5.2 Ideal Sensor Characteristics

This paper has focused on the assembly of composite images with extended intensity range. Is there a way in which this composition can be captured in a transform within the sensor? A compressive function integrated into the sensor would reduce the need for some of, or perhaps eliminate entirely, the series of images of different sensitivities. While it may be possible to use the gamma correction feature to redistribute quantization resolution in an approximation of the instantaneous compression of photoreceptors, gamma correction does not extend the intensity range of the transform. A more fundamental change in the transduction mechanism is required. Of all the possible sites (flux to voltage transform within the cell; the array output amplifier; the resampling digitization at input to the framegrabber; or, the output look-up table of the framegrabber), if the locus of the initial transformation in the cell could incorporate the same form as the Naka-Rushton equation, it would go a long way toward representing local contrast with equipoise. A compressive function at the sensor would eliminate the scalloped variation in contrast resolution due to the piecewise linear approximation and should present little difficulty for most low-level computer vision methods.

5.3 Future Directions

While it is clear that natural scenes require much more than 8-bits/pixel to represent their structure, it is not as clear what the optimal distribution of quantization levels is over the intensity range nor even whether there is a need to retain the *correct* labeling of absolute luminance. Also of interest is the relation between the content of the scene and the relative gain of the series of images. Is it possible that the gains might be automatically determined, perhaps by the outcome of the previous masking stage in a manner that would minimize the number of required images?

What edge operators best extract contours across a large range of luminances? Adaptive operators will be needed that do not allow their response to evaporate at T-junctions with bright edges. Extended intensity range images provide a greater challenge to the traditional collection of edge operators. At large, bright contours many factors contribute to raise the baseline response of linear filters and force small increments below criterion. Perhaps nonlinear filters offer a better chance to track weak edges as surrounding conditions worsen. Does the fact that steep gradients are composed of pixels from many images improve or hamper the detection of intersecting edges?

Also, data from a variety of scenes (especially daylight) need to be collected. The use of a CCD camera with variable flux integration times will facilitate this; however, the portability of a tethered camera is limited. It is also of interest to determine the distribution of naturally occurring specularities. If no color CCD cameras are available that have an electronic aperture, color filters can be used to obtain reasonable RGB partitioning of the scenes. In addition, if unsaturated measures of specularities are obtained, physical models of color clusters can be tested and the performance of edge operators in the absence of specularities can be observed. Different tasks (e.g., object recognition versus image quality) may benefit by different representations of the color gamut (RGB intensities versus Uniform Chromaticity Space (see [Judd75])). Reductions could be made in the residual error caused by the simplifying planar assumptions of specularity extraction algorithms (see [Klin87]).

Are there better combination methods to compensate for hardware deficiencies? When there are pixels at risk of distortion and erosion might be used, what is worse, the quantization error or the nonlinear error? Many of the errors are a result of sensor weaknesses made apparent by the extreme signal strength. Is there a good model for the *class* of CCD sensors incorporating all nonlinearities and thus enabling seamless composites viewable at any arbitrary scale? Complicating the problem is the fact that there are nonlinearities that are characteristic of this particular camera, this particular model of camera, this particular class of sensor, and electronic cameras in general (see [Ande88]). How robust can the combination of images be made with such variation in the underlying equipment?

With the advent of smart sensors, much of the sequential assembly of composite images could be done by a microprocessor proximal to the sensor. The availability of camcorder subassemblies with a color sensor, control microprocessor, and motorized zoom, focus and aperture lens, all weighing 175 grams, will certainly form a good part of the next generation of robotics image acquisition. It would not be difficult to incorporate the serial merging of successive frames, each acquired at a different sensitivity and adaptively modified, all distal to central control.

6 Conclusions

• Natural scenes contain a range of intensities that is far in excess of that which can be covered with conventional solid-state light sensors while still affording sufficient quantization resolution.

• CCD cameras with electronic shutters can be used to create single composite images of natural scenes with extended intensity range from a series of images that varies in sensitivity.

• These composite images have been shown to have advantages in the extraction of important properties of the scene such as local contrast and specularities.

• It is possible to overcome sensor nonlinearities brought on by the presentation of unbuffered high intensities and form a single representation of the scene that is free of artifacts from the composition.

Acknowledgements

The equipment used in this paper was supported by Navy Grant N00014-92-J-1647; ArmyDAAL 03-89-C-0031PRI; NSF Grants CISECDA 88-22719, IRI 89-06770, ASC 91-08013, MSS-91-57156, and CISECDA 90-2253; NATO Grant 0224/85; A.I. duPont Institute, Barrett Technology Inc., duPont Corporation, and General Motors.

References

[Ande88] Anderson, Helen L. (1988) Grasp Lab camera systems and their effects on algorithms. University of Pennsylvania GRASP Laboratory Report 161, MS-CIS-88-85.

[Ando91] Ando, T., Yamamoto, K. and Sawada, K. (1991) Contrast edge detection using a differential input technique of a CCD. *Elec. Lett.* **28**, 2190-2191.

[Beck63] Beckmann, P. and Spizzochino, A. (1963) *The Scattering of Electromagnetic Waves from Rough Surfaces.* Pergamon, New York.

[Beye90] Beyer, Horst A. (1990) Linejitter and geometric calibration of CCD-cameras. *ISPRS J. of Photogram. and Remote Sensing* **45**, 17-32.

[Blak88] Blake, A. and Brelstaff, G. (1988) Geometry from specularities. *ICCV*, 394-403.

[Blak90] Blake, A. and Bulthoff, H. (1990) Does the brain know the physics of specular reflection? *Nature* **343**, 165-168.

[Born81] Born, M. and Wolf, E. (1981) *Principles of Optics*, Sixth Edition. Pergamon Press, New York.

[Cook81] Cook, Robert L. and Torrance, Kenneth E. (1981) A reflectance model for computer graphics. *Comp. Graph.* **15**, 307-316.

[DZmu86] D'Zmura, M. and Lennie, P. (1986) Mechanisms of colour constancy. *J. opt. Soc. Am.* **A3**, 1662-1672.

[Gers86] Gershon, R., Jepson, A. D. and Tsotsos, J. K. (1986) Ambient illumination and the determination of material changes. *J. opt. Soc. Am.* **A3**, 1700-1707.

[Heal88] Healey, Glenn and Binford, Thomas O. (1988) Local shape from specularity. *CVGIP*, 62-86.

[HealIP] Healey, Glenn and Kondepudy, Raghava (In press) Radiometric CCD camera calibration and noise estimation. *IEEE Trans. PAMI.*

[Huan91] Huang, Zhong-Shou and Ando, Takao (1991) Image sensor operating in a persistence-integration mode. *Appl. Opt.* **30**, 4636-4642.

[Ikeu81] Ikeuchi, Katsushi (1981) Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Trans. PAMI*, **3**, 661-669.

[Judd75] Judd, Deane B. and Wyszecki, Gunter (1975) *Color in Business, Science and Industry.* Wiley, New York.

[Kamg89] Kamgar-Parsi, Behrooz and Kamgar-Parsi, Behzad (1989) Evaluation of quantization error in computer vision. *IEEE Trans. PAMI* **11**, 929-940.

[Klin87] Klinker, G. F., Shafer, S. A. and Kanade, T. (1987) Using a color reflection model to separate highlights from object color. *Proc. ICCV*, pp. 145-150.

[LeeH86] Lee, H-C. (1986) Method for computing scene-illuminant chromaticity from specular highlights. *J. opt. Soc. Am.* **A3**, 1694-1699.

[LeeS92] Lee, Sang Wook and Bajcsy, Ruzena (1992) Detection of specularity using colour and multiple views. *Image and Vision Comput.*, **10**, 643-653.

[Madd94] Madden, Brian C. (1994) Light adaptation and the perception of contrast, luminance and the illuminant. University of Pennsylvania GRASP Lab Technical Report.

[Naya93] Nayar, S. K., Fang, X-S. and Boult, T. (1993) Removal of specularities using color and polarization. *CVPR*, 583-590.

[Naya91] Nayar, S. K., Ikeuchi, K. and Kanade, T. (1991) Surface reflection: Physical and geometrical perspectives. *IEEE Trans. PAMI*, **13**, 611-634.

[Nov92a] Novak, Carol L. and Shafer, Steven A. (1992) Estimating scene properties from color histograms. Carnegie Mellon University Computer Science Department Technical Report, CMU-CS-92-212.

[Nov92b] Novak, Carol L. and Shafer, Steven A. (1992) Anatomy of a color histogram. *CVPR*, 599-605.

[Park90] Park, Jong-Seok and Tou, Julius T. (1990) Highlight separation and surface orientations for 3-D specular objects. *Proc. ICPR*, 331-335.

[Shaf85] Shafer, Steven A. (1985) Using color to separate reflection components. *COLOR res. and applic.* **10**, 210-218.

[SONY90] SONY Corporation (1990) XC-77RR/77RR-CE Service Manual (XCM-77RR).

[Thri83] Thrift, Philip and Lee, Chia-Hoang (1983) Using highlights to constrain object size and location. *IEEE Trans. Sys., Man and Cybern.*, **13**, 426-431.

[Torr67] Torrance, K. E. and Sparrow, E. M. (1967) Theory for off-specular reflection from roughened surfaces. *J. opt. Soc. Am.* **57**, 1105-1114.

[Vale83] Valeton, J. M. and van Norren, D. (1983) Light adaptation of primate cones: An analysis based on extracellular data. *Vision Res.* **23**, 1539-1547.

[Witt88] Wittles, N., McClellan, J. R., Cushing, K., Howard III, W. and Palmer, A. (1988) How to select cameras for machine vision. *Proceedings of Optics, Illumination, and Image Sensing for Machine Vision III, SPIE* **1005**, 44-53.

[Wolf92] Wolff, Lawrence B. (1992) Diffuse reflection. *CVPR*, 472-478.

[YiSe90] Yi, S., Haralick, R. M. and Shapiro, L. G. (1990) Automatic sensor and light source positioning for machine vision *Proceedings of ICPR*, 55-59.

[Ziss89] Zisserman, A. Giblin, P. and Blake, A. (1989) The information available to a moving observer from specularities. *Image and Vision Comput.*, **7**, 38-42.