Proceedings of the 2006 IEEE International Conference on Robotics and Automation
Orlando, Florida - May 2006

Extending the Dynamic Range of Robotic Vision

Stephen Nuske*†, Jonathan Roberts†, Gordon Wyeth*
*School of Information Technology and Electrical Engineering
University of Queensland
Brisbane, Australia, 4067
†Autonomous Systems Lab
CSIRO ICT Centre
P.O. Box 883, Kenmore, Australia, 4069
Email: stephen.nuske@csiro.au

***Abstract*— Conventional cameras have limited dynamic range, and as a result vision-based robots cannot effectively view an environment made up of both sunny outdoor areas and darker indoor areas. This paper presents an approach to extend the effective dynamic range of a camera, achieved by changing the exposure level of the camera in real-time to form a sequence of images which collectively cover a wide range of radiance. Individual control algorithms for each image have been developed to maximize the viewable area across the sequence. Spatial discrepancies between images, caused by the moving robot, are improved by a real-time image registration process. The sequence is then combined by merging color and contour information. By integrating these techniques it becomes possible to operate a vision-based robot in wide radiance range scenes.**

I. Introduction

Radiance levels in real environments can span up to ten orders of magnitude, the human eye can capture approximately five orders of magnitude of dynamic range simultaneously and can adapt over time to a radiance range of approximately eight orders of magnitude [1]. Conventional imaging devices, such as cameras (both digital and analog) and monitors, by comparison, have simultaneous dynamic ranges of approximately three orders of magnitude. As a result much information is lost when trying to measure and reproduce intensities outside the range of the imaging device see Fig. 1. The high radiance range problem is well documented, with plenty of literature on attempts to increase the simultaneous dynamic range of conventional low dynamic range sensors. These techniques all have pitfalls they either have reduced spatial resolution [2], require complex optics/hardware [3], [4] or are unable to cope with motion [5]. There is work on true HDR (high dynamic range) sensors [6], though at present are not widely available.

This paper presents a method for an extended dynamic range vision system, enabling a robot to move from a dark indoor to a sunny outdoor environment. The method is unique from other such sequential exposure changing techniques [5], because it allows operation in a non-static, wide radiance range scene in real-time by a conventional digital camera. Authors [7] have presented a method of performing real-time sequential exposure changing, using a camera with modified firmware. This paper shows a similar technique but performed without camera modifications and with the combination of the sequence in real-time.

Fig. 1. Example showing the limitations of conventional LDR cameras. Top: Image taken with a regular exposure, areas outside are completely overexposed. Bottom: Image of the same scene taken with a short exposure, areas inside are completely underexposed.

Because the sequence of images is taken on a moving robot, spatial discontinuity in the image sequence needs to be accounted for before they can be directly manipulated. Kang et al. [7] shows an *offline* technique of pixel boosting images of short exposures, to match image intensities of the reference image, before registration. Conversely this paper, instead of tackling the difficult problem of registering images of differing exposures, presents a technique of registering translational error between exposure-sequences first and then interpolating the error to the other images of differing exposure. By conducting part of the registration process in low-resolution and then performing high-resolution registration only at points of interest, real-time image registration is possible.

Once the sequence of aligned images which together span a wide range of radiance, is obtained the issue is then to preserve all necessary information about the scene in a suitable representation. In [5] HDR radiance map construction is performed on sets of images. However high-bit-depth images contain too much infomation to be used effectively, thus HDR compression is required. Again there has been much literature published on HDR image compression, which is generally aimed at display purposes, either for photography or for conventional monitors, with the principal criteria of making the image *appear* real. Ledda et al. [8] provides an evaluation of existing techniques. These techniques are designed to work offline with the process generally taking several seconds to compute, some methods even need user parameter tweaking to achieve acceptable results. These methods are not designed for real-time robotics this paper aims to show that a sequential exposure technique can be applied to robotics. This paper aims to show that a sequential exposure technique can be applied to robotics. Each image in the sequence is converted into a representation not designed for human viewing, but in a form suitable for robotic vision processing, with contour and chrominance information being retained, as opposed to absolute RGB intensity values.

The aim of this technique is to increase the visual information available for a robot operating in a high radiance range environment, making it possible to perform tasks that involve moving between a dark indoor environment and a sunny outdoor environment. The key challenge is to develop algorithms that achieve this functionality on a conventional, unmodified digital video camera in real-time.

II. Real-Time Exposure Changing with Automatic Control

To form a sequence of differently exposed images the exposure level of the camera is changed between frames. This section explains how this is performed without hardware modification. To achieve real-time operation the number of exposures in the sequence is limited. This restricts the range of radiance covered by the sequence and brings the need for individual exposure control of each image to maximize the information across the sequence.

A. Real-Time Sequential Exposure Changing

Kang et al. [7] shows a technique of modifying a cameras firmware to perform real-time exposure changing on moving scenes. This paper shows that the exposure changing can be performed without camera modification, with any IIDC [9] compliant IEEE1394 digital camera. IEEE1394 digital cameras have the advantage of enabling both video data transmission and the control of the camera over a single connection; exposure settings can be controlled from a PC, see Fig. 3. However if placed in the 'continuous shot' mode from the IIDC camera specification, it is unknown exactly when in the capture-transmit cycle the new exposure setting command will be received by the camera. This is because in this mode the camera's capture-transmit cycle is decoupled from the PC control software. To avoid this problem the alternative 'one shot' capture mode can be used. The PC can send an exposure setting (gain level and exposure time) to the camera before sending a one shot command, which ensures the next image received is captured with that exposure setting. The PC will wait for the image to be transmitted, before sending the next exposure setting followed by another one shot command. The disadvantage of the one shot mode, is that the maximum frame rate of the system is reduced because the camera cannot capture a frame while the previous frame's data is being transmitted.

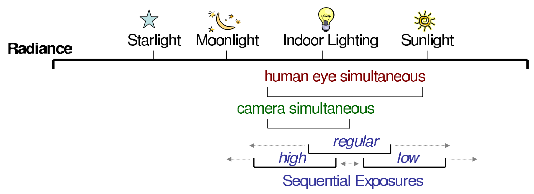

Fig. 2. Example illustrating natural radiance levels compared with the simultaneous radiance range of the human eye and that of a camera. Also shows the extended range of the sequential exposure technique. Rough estimates of range taken from [10], [1]

B. Automatic Multiple Exposure Control

Using the sequential exposure changing process, any number of different exposures can be employed in the sequence, however the more exposures that are used, the lower the cycle rate of the system. Three different exposures (*low*, *regular* and *high* exposures) were deemed to be appropriate to span a wide enough range of radiance, whilst enabling a sufficient cycle rate. In this section the control algorithms for the exposure levels are described.

Generic automatic exposure control (AEC) algorithms work by a grey-world assumption that the surface reflectances of the objects in the field of view will always average out to a constant value. The algorithm will adjust the exposure to return the average image intensity to the desired value, which hopefully negates the effects of changing scene luminance, a downside being that only uniform lighting changes can be accounted. This generic algorithm fails in non-uniform lighting conditions, an example being when the field of view is a combination of a bright outdoor environment and a darker indoor environment.

To maximize the correctly exposed areas of a non-uniformly lit scene we control the exposure level of each image in the sequence. For the *regular* exposure a generic AEC algorithm is employed attempting to keep the mean intensity of the whole image at 50%. It would then be possible to fix the *low* and *high* exposures to a predefined offset either side of the *regular* exposure level, which would give a fixed simultaneous dynamic range. However to cover the full range of radiance in a real world scene more than three exposure levels are needed, see Fig. 2. So to maximize the viewable area across the three exposures and to account for non-uniform lighting changes, individual (de-coupled) control strategies for the *low* and *high* exposures are employed, aimed at viewing the under-exposed and over-exposed areas of the *regular* image.

Our exposure control equation is as follows:

$$L_t = L_{t-1} + (\frac{L_{t-1}}{100} \times P \times E) \quad (1)$$

Where L_t is the exposure level at time t, P is our proportional gain, and E is the error we are trying to control. To make the technique portable between different cameras, the exposure level L is stored as a fraction of the overall exposure range of the particular camera. This range can be calculated at run-time by accessing the minimum and maximim values in the inquiry registers of the camera, listed in [9]. Note in Eq. 1 that the exposure level adjustments need to be scaled by the current exposure level.

The control equation Eq. 1 is essentially the same for the three exposures, however the respective error measurement is different for each exposure. As stated earlier for the *regular* exposure we attempt to keep the mean intensity of the *whole* image at 50%, so the error for our *regular* exposure control is:

$$E_R = \frac{\sum_{k=0}^{n} R_k}{n} - 128 \quad (2)$$

Where n is the number of pixels in the image, R_x is the intensity of pixel x in the regular exposure and also note that the value 128 is 50% of an 8-bit intensity image.

For the *high* exposure we try to control the mean intensity of the *darkest* third of the image at 50% of the intensity range. The error is as follows:

$$E_H = \frac{\sum_{k=0}^{D} HIST_H(k) \times k}{^n/_3} - 128 \quad (3)$$

Where D is the intensity value at which a third of the pixels in the *high* exposure are darker than and where $HIST_H(i)$ is the number of pixels in the *high* exposure that have intensity value i.

For the *low* exposure we try to control the mean intensity of the *brightest* third of the image at 50%. The error is as follows:

$$E_L = \frac{\sum_{k=B}^{255} HIST_L(k) \times k}{^n/_3} - 128 \quad (4)$$

Where B is the intensity value at which a third of the pixels in the *low* exposure are brighter than and where $HIST_L(i)$ returns the number of pixels in the *low* exposure that have intensity value i.

C. Iris Control

It is possible to adjust to an even greater range of lighting conditions by altering the aperture of the iris. This was achieved by controlling a motorized-iris lens through the PC's parallel port, see Fig. 3. The iris is opened and closed with the goal to keep the *regular* exposure level in the middle of the exposure level range. Note that iris control does not directly increase the simultaneous dynamic range of the camera, as iris adjustments occur over longer periods of time. However it does *allow* a greater simultaneous dynamic range by enabling the exposure levels of the sequence to be spread out.

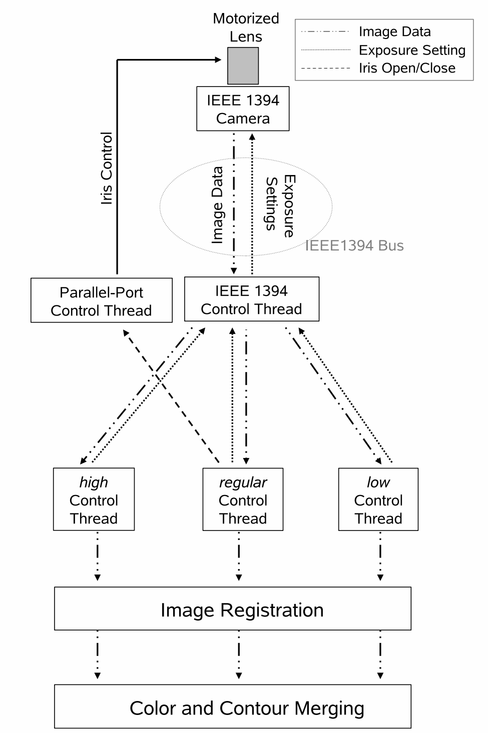

Fig. 3. Diagram of the overall system. Showing automatic control of the three sequential exposures and the lens aperture followed by image registration and merging.

III. Fusion of Gradient and Chrominance Information

It is not possible to directly merge luminance information from differently exposed images because the intensity values represent different ranges of radiance. It is, however, possible to create an absolute radiance map with a large bit-depth, as shown in [5]. HDR image construction is an offline process, often taking seconds to compute and the images contain too much information to be efficiently and effectively processed by robotic vision systems, hence the need for a compressed representation of the scene. Ledda et al. [8] provide an evaluation of existing reduction techniques. These techniques are designed to work offline with the process generally taking several seconds to compute, some methods even need user parameter tweaking to achieve acceptable results. Whon-Ho and Ki-Sang [11] show that HDR radiance map construction can be avoided by fusing low-bit-depth images together by taking intensities from correctly exposed regions and manipulating boundary conditions of the regions to avoid halos. However it is not possible to accurately retain all intensity information in an 8-bit image. Furthermore it is accepted that the human visual system is not sensitive to absolute light levels reaching the retina. The human eye is far more sensitive to local intensity changes [12]; primarily contour, texture and color information is used. This section explains the techniques used to merge the required information from the set of images in real-time, using a conventional color camera with RGB color filters.

Fig. 4. Gradient maps of two exposure sequences, with YCrCb chrominance overlaid. Darker regions correspond to larger gradients. Top-Bottom: *low* exposure, *regular* exposure, *high* exposure, combined gradient and color information.

Raw RGB values are transformed into the YCrCb color space, which separates color and luminance information. Contour and color information of the sequence of images is then merged separately.

A. Multi-Scale Contour Representation

A gradient operator is employed to preserve visual information across the sequence of exposures. The operator works on multiple scales (similar to the method in [13]) to handle image noise and to emphasize important information, first by creating a pyramid of sub-resolution images by starting at the bottom with the original image and creating lower-resolution images by taking the mean intensity of the corresponding pixels from the higher-resolution image. Then the pyramid of images is converted into a pyramid of contour maps using a 3×3 gradient-magnitude operator which takes the mean difference in intensity between the centre pixel and the 8 neighbouring pixels. This operation is performed by starting at the top of the pyramid and working down by taking half the gradient magnitude at the current level and half the corresponding gradient from the level above. The resulting gradient map at the bottom of the pyramid has softer gradients for fine texture/detail and strong gradients on boundaries that occur at larger scales.

Now that a sequence of gradient maps is available, each with different information, the missing details in the *regular* image due to over-exposure or under-exposure, can be recovered in the *low* and *high* images, see Fig. 4. As the gradient maps are simply local intensity differences and are independent of radiance ranges, corresponding gradient values between images can be directly combined without absolute radiance map construction. To merge the information, at each point the maximum gradient magnitude from the sequence is taken, with the concept that the larger the gradient infers more information, and more information infers reliable information.

B. YCrCb Chrominance

The concept that more information infers better information is also applied to color information where more information is a higher saturation of color. In the YCrCb color space, the chrominance channels are Cr and Cb, and when plotted against each other the white point (no color saturation) is located in the centre, the further away from the white point the more color saturation. The Cr and Cb values with the maximum Euclidian distance from the white point are taken when merging the sequence of images.

C. Image Registration

There will be spatial discontinuity between images when there is motion between the camera and the scene, which results in artifacts when merging. This system has been tested with a forward facing camera on a ground robot, and artifacts do appear due to the robot turning. The image motion is predominantly through horizontal translation. A real-time image registration process is presented which finds the translation discrepancies between images.

Kang et al. [7] shows an *offline* technique of boosting images of short exposures to match the intensities of the reference image, before pair-wise registering. The technique requires offline processing and a larger set of images with relatively similar exposure. Pixel boosting will not work with our three images, as they have significantly different exposure levels. Thus a new technique has been developed to avoid this problem, by registering two consecutive sequences of exposures and then interpolating the translational error to images within the sequence.

The method of Kang et al. [7] takes a total of 32 seconds to compute. This paper is concerned with real-time operation and a more efficient technique is required. To enable real-time performance, a coarse to fine registration refinement process is used [14]. In our method registration is performed only at the coarsest and the finest resolutions. Registration is performed using the gradient maps described in section III-A.

To find large translations, *regular* gradient maps from consecutive sequences are registered at the coarsest resolution. As stated in [14] this is benificial not only for computational efficiency but avoids fine scale aliasing problems. The registration is an energy minimization process of finding the translation which produces the most correlated gradients. Before fine-scale registration, points of interest are found at the coarse resolution, to further accelerate the fine registration process. These are the points with the largest gradient magnitude. Registration at the highest-resolution is performed only at the points of interest, using the coarse translation as a starting point. Again using the same energy minimization process to find the resulting fine-scale translation with the most gradient correlations.

The motion between the other images in the sequence, is calculated by interpolating the inter-sequence translation. By assuming constant image velocity, the interpolation is linear based upon the timestamps of each image. See Fig. 8.

IV. Results and Discussion

To demonstrate the system, a forward facing camera was mounted on an experimental four-wheeled robot, and driven from an indoor environment lit by artificial light, to a bright sunny outdoor environment.

A. Multiple Exposure Control Results

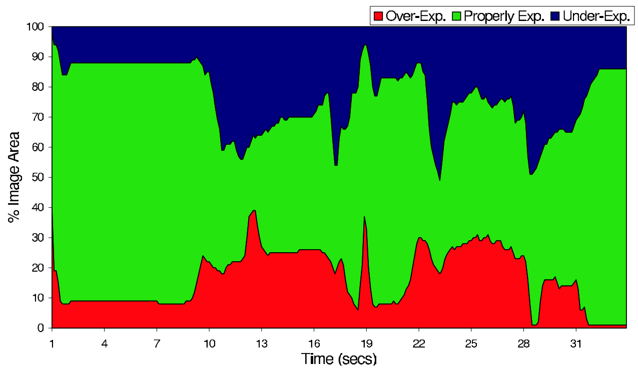

Fig. 5. Graph showing the limitations of a single exposure with a generic exposure control algorithm. At times there is only 20% correctly exposed area in the image. Graph of: over, under and properly exposed areas of our *regular* exposure.

Color and contour images from the experiment can be seen in Fig. 4 which clearly show the information gained from areas both indoors and outdoors. Fig. 5 illustrates the problem of using a single exposure with a generic exposure control algorithm; at times correctly exposed area is limited to around 20% (A properly exposed pixel is deemed to have an intensity above decimal 50 and below 255). Fig. 6 on the other hand shows the increased correctly exposed area our system provides. Fig. 6 also shows that all of the information about a wide radiance range scene cannot be captured in a sequence of three exposures. There is missing information outside the range of the three exposures and intra-range information missing because the radiance ranges do not always overlap (due to our decoupled control algorithm see section II-B).

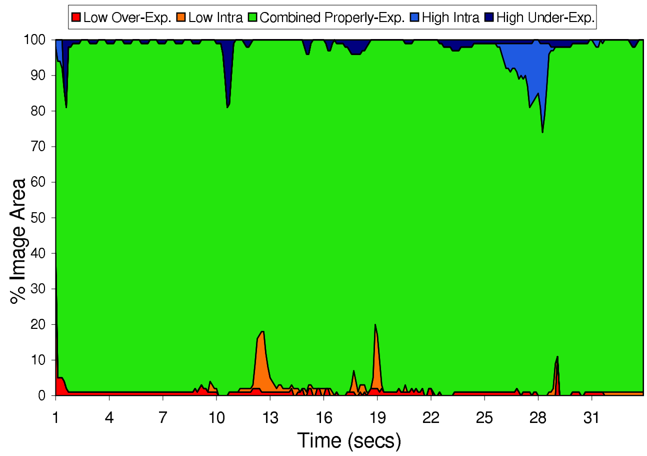

Fig. 6. Graph showing the information gained from our *low* and *high* exposures. Also shows that not all information can be captured with only three exposures. Graph of: the combined correctly exposed areas, over-exposed area in the *low* image, under-exposed area in the *high* image, and intra-range information missed because intensity ranges of the sequence do not always overlap.

B. Image Registration Results

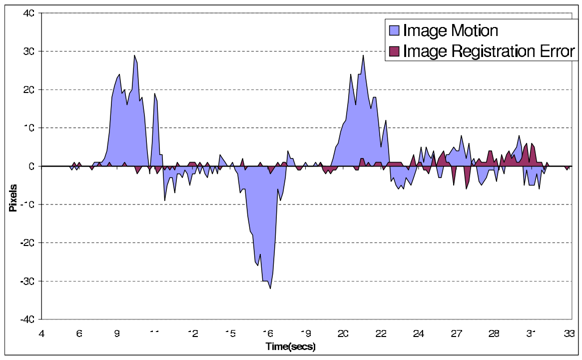

Fig. 7. Horizontal image motion of neighbouring exposure sequences. The motion is calculated by manually aligning the images. Also shown is the corresponding error of the automatic registration process.

The spatial error between images is caused by both motion between camera and scene, and the frame-rate of the camera. To achieve suitable results the camera was configured to capture 500 × 400pixels images enabling frame-rates in excess of 30 Hz. The exposure time also affects the frame-rate; upper limits on exposure time were calculated so that in the worst case scenario, frame-rates would drop to 25 Hz. If dark areas need to be viewed, the limits can be extended but at the expense of the frame-rate. Similarly the resolution could also be increased.

The robot turned corners during the manoevre, which created significant motion between camera and scene, enabling an evaluation of the image registration. Ground-truth image motion was recorded by manually aligning images that were saved during the experiment. The image registration output was compared with this ground-truth motion to give an error measurement. From Fig. 7 and Fig. 8 it is clear that the image registration technique copes with translational image motion with a reasonable degree of accuracy. Fig. 7 also shows that the registration errors are worse towards the end of the experiment. It is during this period that the robot is travelling through the doorway, creating a looming effect, which translational correction cannot resolve. Camera motions such as rotation, affine and projective transformations, are also not accounted. However the effects of these motions are not as significant as translational motions caused by the robot turning, and it is difficult (if not impossible at present) to correct for these in real-time.

Linear image motion is assumed for the interpolation of translational error. If there are significant accelerations in image motion, the interpolation will be incorrect. However, as the cycle rate is kept high, image motion is low.

When operating the system on a 3.6 GHz Pentium 4 machine, image acquisition, sequential exposure changing, exposure control and gradient map construction, are successfully operated at 30 Hz. Image registration and interpolation with contour and color merging of the three image sequence is operated at 10 Hz.

V. Conclusion

In this paper we have presented techniques that make it possible for a robot to operate in a wide radiance range scene. Due to the limited dynamic range of conventional cameras, vision based robots cannot move from an indoor environment to a sunny outdoor environment simply because both environments cannot be viewed at the same time. This paper presents an approach to extend the effective simultaneous dynamic range of a camera, by changing the exposure level of the camera, to form a sequence of images which together cover a wide range of radiance. The individual control algorithms developed for each image in the sequence provide maximal viewable area across the sequence. The sequence is combined by merging color and contour information, which avoids the time expensive operations of constructing absolute radiance maps and image reduction, whilst retaining information useful for a robotic vision system. Spatial discrepancies between images, were improved by developing a real-time image registration method. All these techniques were integrated into one system showing that it is possible for a vision-based robot to be operated in wide radiance range scenes.

Fig. 8. Top: Merged sequence of exposures without registration. Notice the artifacts at the fire hose (top-left), traffic cones (centre) and drawers (right). Bottom: Merged sequence of exposures with automatic real-time registration and interpolation.

References

[1] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 760–768, 2004.

[2] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *Proceedings of the Ninth IEEE International Conference on Computer Vision*, 2000.

[3] S. K. Nayar and V. Branzoi, "Adaptive dynamic range imaging: Optical control of pixel exposure over space and time," in *Proceedings of the Ninth IEEE International Conference on Computer Vision*, 2003.

[4] M. Aggarwal and N. Ahuja, "Split aperture imaging for high dynamic range," in *International Conference on Computer Vision*, 2001.

[5] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH*, 1997.

[6] M. Mase, S. Kawahito, M. Sasaki, Y. Wakamori, and M. Furuta, "A wide dynamic range cmos image sensor with multiple exposure-time signal outputs and 12-bit column-parallel cyclic a/d converters," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 12, December 2005.

[7] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 319–325, 2003.

[8] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen, "Evaluation of tone mapping operators using a high dynamic range display," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 640–648, 2005.

[9] *IIDC 1394-based Digital Camera Specification*, 1st ed., 1394 Trade Association, Regency Plaza Suite 350, 2350 Mission College Blvd., Santa Clara, CA 95054, USA, July 2000.

[10] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg, "A model of visual adaptation for realistic image synthesis," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM Press, 1996, pp. 249–258.

[11] C. Won-ho and H. Ki-Sang, "Extending dynamic range of two color images under different exposures," in *Proc. 17th International Conference on Pattern Recognition*, 2004.

[12] J. DiCarlo and B. Wandell, "Rendering high dynamic range images," in *Proceedings of the SPIE: Image sensors*, 2001.

[13] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Transactions on Graphics*, 2002.

[14] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierachical model-based motion estimation," in *European Conference of Computer Vision*, 1992, pp. 237–252.