Enhancing Structured Light Range Imaging by Adaptation of Color, Exposure and Focus

Alain Boyer, Pierre Payeur
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, ON, Canada
[aboye074, ppayeur]@uOttawa.ca

***Abstract*—This paper presents a series of enhancements to a color-coded structured light range sensor that increases the adaptability to complex and unconstrained scenes. First, the projected pattern is made more visible on colored objects by replacing the unique colored pattern with time-multiplexed pseudo-color channels. Second, an exposure fusion algorithm is used when acquiring images to allow the detection of regions with low and high reflectance characteristics. Finally, the focus planes of the scene are automatically detected and imaged separately, enlarging the sensor's depth of field. Each improvement is detailed and integrated into a custom acquisition procedure. Experimental results demonstrate the improved robustness of the structured light range sensor and validate the proposed design.**

***Keywords*—range imaging; structured lighting; intelligent measurement systems; object modeling; virtual reality.**

I. Introduction

Active vision has been widely researched and is still the preferred method for dense 3D range sensing. Passive vision has a high dependence on the presence of features in the scene and, at best, can only generate sparse 3D range information. To achieve precise measurements with high resolution, an active vision method is required. Two such technologies have gathered much interest. Laser range sensors are accepted as the state-of-the-art standard for 3D measurement since they can achieve very high accuracy with very low computation time. However, they require specialized hardware that is usually expensive and not readily available. Structured light range sensors, on the other hand, remain an affordable solution for 3D range sensing since they can be assembled using common off-the-shelf digital cameras and LCD projectors. Although the precision and resolution of structured light sensors may not be as high as that of their laser-based counterparts, they can produce accurate and dense range scans that are not achievable with passive vision, and exceed what recent mass-market active depth sensors can offer.

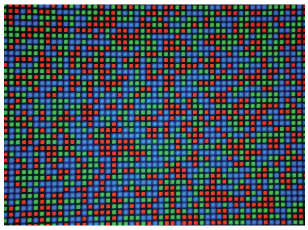
This paper builds on a previous version of a structured light range sensor [1] and aims to improve the acquisition module. The motivation is to make the system more robust while improving its ability to adapt to the scene before it. The goal is also to move away from conventional object-centered to more general scene measurement. As a result, the number of parameters that must be adjusted is reduced and the range sensing system is easier to operate.

Three enhancements that greatly improve the adaptability of the structured light range sensor are presented. First, colored objects and scenes with significant color variation are considered. Second, object brightness and scene reflectance characteristics are compensated for. Third, multiple focus planes of one or more objects are taken into account. These changes are motivated by the need to acquire as much information as possible about the scene during a single capture. Concrete examples are presented that validate the above enhancements and show that the system is robust and capable of dealing with arbitrary scenes in an unconstrained fashion.

II. Previous Work

Before discussing the proposed improvements, the previously developed structured light range sensor [1] is briefly presented. The system is composed of two cameras mounted on a rigid bracket as a stereo pair above an LCD projector as shown in Fig. 1a. The only calibration required is of the intrinsic [2] and extrinsic [3] camera parameters, leaving the projector uncalibrated with the rest of the system. The sole purpose of the projector is to project a pseudo-random (PR) pattern of colored squares [4], shown in Fig. 1b, onto the scene, to generate artificial features. The pattern is defined such that each 3x3 neighbourhood of squares is a unique code and that all codes are separated by a minimum Hamming distance.

(a)



(b)

Fig. 1. (a) Stereoscopic structured light range sensor, and (b) bi-dimensional pseudo-random pattern.

This spatial-neighbouring pattern is combined with a time-multiplexed [5] approach which iteratively marches the entire pattern horizontally and vertically to increase the density of artificial features. During the marching phase, images are captured and stored for analysis. The image processing consists of adaptively segmenting the colored squares, labelling and grouping them to yield 3x3 codes. After a confidence level analysis, the accepted codes are used to solve the correspondence problem and the optimal triangulation [3] is applied to reconstruct a 3D point cloud. Since color images are captured in the process, a color is also assigned to each point, enhancing the richness of data.

Although good results are achieved on single objects, the range sensor's limitations become evident as more complex scenes are introduced. Multicolored objects tend to distort the color of the projected pattern and cause the colored square segmentation algorithm to fail. This is compounded with the fact that objects with multiple brightness and reflectance characteristics produce images with areas that are under- and over-exposed, contributing to a loss of 3D data in those areas. These two common and related problems due to color and light properties must be addressed with parallel solutions. Moreover, the previous sensor can only acquire data from one focus plane and ignores objects or areas of a scene where the pattern is out of focus and unrecognizable by the segmentation algorithm. In some cases, this can greatly reduce the amount of 3D information collected from a scene with a large depth of field. Finally, the previous implementation requires that two major parameters are prior to the acquisition, namely the exposure time of the cameras and the focus of the projector. Manual calibration of the sensor is not acceptable since the goal is to create an easy-to-operate system capable of working autonomously, with the potential to be mounted atop a mobile platform.

The conventional approach of adaptive structured light is to dynamically adjust the projected pattern in response to the scene [6], [7]. This can be achieved by adapting the pixel color [6] and pixel intensity [7] of the pattern. However, extensive calibration between the cameras and projector is necessary, which is incompatible and undesirable with the current structured light system. Another more basic approach is to acquire multiple images while varying the global intensity of the projector and combining the images into a high dynamic range radiance map [8]. Although simple and effective, the problem of selecting a global exposure rate for the image acquisition still remains. This problem is further complicated when highly reflective objects lead to saturated areas in the image, regardless of the projector intensity. The proposed method is inspired by Skocaj and Leonardis [8] but, instead of working at the projector level, it operates at the camera level. The pattern is projected at full intensity and multiple images are acquired while varying the exposure rate. The images are then fused together using the concept of exposure fusion [9], [10], [11], ultimately producing an image with a local exposure rate that compensates for the different colors and reflectance characteristics in the scene.

Most literature on structured light range sensors assumes that the scene is located at a relatively constant distance from the sensor, which assures that the cameras and projector are always in focus. When building a flexible sensor adaptable to any scene, this assumption cannot be made and the focus problem must be considered. This work is inspired by the concept of focus fusion [10], [12], [13] and presents a novel method that acquires data from different focus planes and fuses it to obtain range data from a workspace that exceeds the focus capabilities of most projectors.

The fact that the projector does not require calibration with the stereoscopic camera system is considered an advantage over other similar range sensing systems. This not only increases the flexibility and ease of use of the acquisition system but also allows it to adapt more easily to the scene via zoom and focus. This allows for a more complex and robust acquisition stage which is the objective and main contribution of this work.

III. Adaptive Acquisition Framework

This section presents an acquisition framework that is flexible, customizable and improves the performance and robustness of the sensor. The major design concepts for the three areas of improvement are presented.

A. Time-Multiplexed Pseudo-Color Code Projection

After much experimentation with colored codes on multi-colored objects, the use of a colored spatial-neighbourhood pattern was abandoned. Although not a novel solution, it proved to work reliably only on uniform and lightly colored objects that reflect light well. It did not work on dark objects that absorb most of the projected light, or on multicolored objects. With focus adaptation in mind, it was undesirable to impose a calibration between cameras and projector to perform conventional adaptive structured light.

The proposed approach is to project the spatial-neighbourhood pattern using only white light at full intensity. This not only ensures that the maximum possible amount of light is reflected from dark and colored areas but that it also reflects from areas of the scene that are farther away from the projector, therefore increasing the range of the sensor. Since the spatial-neighbourhood codes are composed of three bits that were encoded using three different color channels (red, green, blue), they are now encoded using a time-multiplexing approach. Assuming that a static scene is being imaged, the three individual code channels are successively projected as white patterns and three sets of images are respectively acquired, simulating simultaneous projection of the three colors. Fig. 2 shows the three individual pseudo-color channels that compose a single complete pattern.

(a) (b) (c)

Fig. 2. Three time-multiplexed projections of the pseudo-color: (a) red, (b) green and (c) blue channels on two objects located side-by-side.

The use of time-multiplexing does not introduce a significant increase in overhead since the colors do not represent unique codes but rather bits, used to construct unique codes. In this case, only two extra projections and acquisitions are needed to obtain a complete pattern. It should be noted that this is a separate time multiplexing, not to be confused with the time multiplexing at the entire pattern level, denoted as marching patterns, which increases the density of the acquired range data.

B. Variable Exposure Fusion

Since the pattern is projected using a maximum intensity, objects and areas with high reflectance properties lead to saturated regions in the acquired images. The previous implementation did not compensate for objects with multiple reflectance characteristics and naively used a global exposure that also had to be adjusted manually. Since the exposure time is the most important parameter to set when capturing digital images, it must be automatically selected. However, with most computer vision algorithms, a single global exposure is not sufficient to properly image the entire scene.

In order to capture the pattern on objects of multiple colors and reflection properties, the proposed solution is to simulate a local exposure when acquiring images using the exposure fusion technique. The technique consists of acquiring several images of the same scene while varying the exposure time to obtain a set of images that contain properly exposed regions of the entire scene. This set of images is fused to produce a single image that has a dynamic range greater than what is possible to obtain from a single image. The exposure fusion technique proposed by Mertens *et al.* [11] consists of computing quality measures of contrast, saturation and well-exposedness at each pixel of each image in the set. These are then combined to produce weight maps for each image and are then normalized. The images are blended together by applying their respective weight maps in a multi-resolution technique using pyramid decomposition. Fig. 3 shows an example of an acquisition and its weight map along with a final exposure fused image.

(a) (b) (c)

Fig. 3. (a) Example of one acquisition at low exposure, (b) its corresponding weight map, and (c) the final composite locally-exposed image.

The Mertens *et al*. technique is used as it operates at the pixel level as opposed to other techniques that process image blocks. The latter techniques were found to require large block sizes that reduced the resolution level of the sensor since it is difficult to retrieve accurate data from blocks, which contain two or more regions of different reflectance properties.

The exposure fusion algorithm is applied every time the two cameras capture the scene. The fusion is performed using only the well-exposedness quality measure, as the other measures do not contribute significantly more information. The resulting composite image is properly exposed and the structured light pattern is uniformly visible regardless of color and reflectance properties. Moreover, the selection of an exposure time parameter disappears since it is inherently selected via the exposure fusion algorithm.

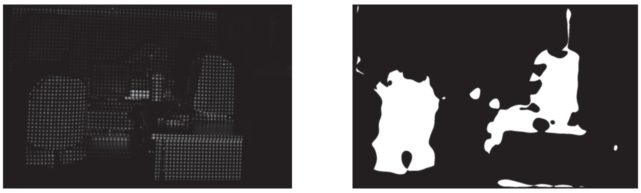
C. Dynamic Focus Fusion

The sensor can now acquire data from multiple surface colors and reflectance properties as long as they all lie within a focus plane and the projector is properly focused to that plane. Such assumptions cannot be made when designing a sensor capable of operating autonomously in unconstrained environments. The workspace of such a system is usually constrained by the focus and intensity of the projector. The minimum distance of the workspace is bound by the focal capabilities of the projector while the maximum distance is bound by the intensity of the projector.

In most realistic applications, cameras and lenses can be configured such that the entire workspace is in focus and the only parameter that must be adapted is the focus of the projector as it varies considerably within the workspace. The proposed solution, which is inspired from the exposure fusion process, is to vary the focus of the projector from the closest to the furthest focus planes, while capturing images at each plane and merging them using a focus fusion algorithm. The focus fusion technique proposed by Hariharan *et al.* [13] consists of computing horizontal and vertical image gradients that are used to determine sharpness masks of the images at each focus plane. The sharpness masks are filtered to reduce noise and increase neighbourhood relevance. Partition masks, which correspond to regions of high focus, are generated for each focus plane by mapping maximum values in the sharpness masks. Finally, the partition masks are used to merge the set of images into a locally-focused composite image.

The Hariharan *et al.* technique is used as it detects focally connected regions as opposed to focused pixels or focused image blocks. This technique is suitable for images where most of the scene is in focus and only the projected pattern squares are in or out of focus, which is the case with the structured light sensor.

Due to the optics of the projector, the projected pattern shifts slightly as the focus is adjusted. Therefore, the focus fusion algorithm as described above is not used in its entirety since this would produce an image that contains a pattern that does not line up across focus regions. Instead, the fusion step is dropped and the partition masks are used to identify focally connected regions during a two-pass acquisition scan. The idea is to perform a first pass, which cycles through each focus plane in one direction and determines focally connected regions at each plane as shown in Fig. 4. A quick analysis is performed and focus planes that have a small overall area in their sharpness mask are dropped. The second pass cycles through the subset of focus planes in the opposite direction, performing the pattern acquisition and processing only in properly focused regions as defined by the sharpness masks. The range data is independently computed for each focus plane and concatenated once all planes are processed.



(a) (b)

Fig. 4. (a) Example of an image collected from one focus plane where closer objects are in focus, and (b) its corresponding sharpness mask.

IV. Detailed Acquisition Procedure

The proposed sensing method operates in two stages. First, the acquisition stage controls the cameras and projector programmatically and saves images to the disk. The process is detailed in pseudo-code below.

Acquisition Stage:

- *Iterate through all focus planes.*
  - *Adjust projector focus to current focus plane.*
  - *Project full pattern using maximum intensity and acquire images using exposure fusion.*
  - *Compute sharpness mask for current focus plane.*
- *Drop focus planes that contain a proportionally small sharpness mask.*
- *Re-compute sharpness masks for subset of focus planes.*
- *Iterate through subset of focus planes.*
  - *Adjust projector focus to current focus plane.*
  - *Perform structured light data acquisition in region specified by current sharpness mask.*
    - *Project time-multiplexed pattern channels at maximum intensity and acquire images using exposure fusion.*
    - *Save images to disk for further processing.*

Processing Stage:

- *Process the saved images and extract 3D range data using previously developed algorithms [1] for each focus plane.*
- *Concatenate 3D range data from all focus planes.*

Essentially, a first pass through the focus planes is performed to determine regions that are in focus. Next, a second pass is made that sequentially collects the projected pattern only from the regions in focus. The three pseudo-color pattern channels are projected separately using a maximum intensity and for each channel, multiple images of increasing exposure are acquired and then fused.

Second, the processing stage analyzes the captured images, identifies the pseudo-random code correspondences and performs a triangulation to extract 3D points. This stage remains mostly the same as in the original design [1]. However, the difference is that it is independently performed on the data at each focus plane and the resulting 3D points are concatenated to produce the final 3D point cloud.

The acquisition procedure is more robust and powerful while remaining flexible. The parameters controlling the number of focus planes, the number of images for exposure fusion and the different exposure levels can all be manually adapted depending on the scene. However, they can also be set to their maximum settings, rendering the structured light sensor completely autonomous in its acquisition regardless of scene colors, reflectance characteristics and depth of field.

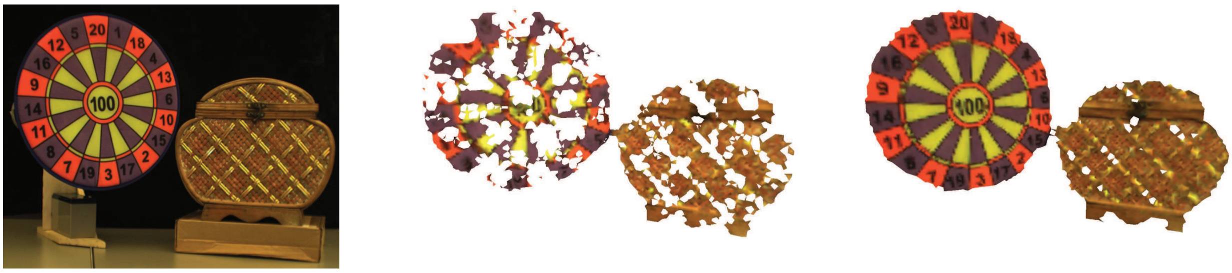
It is also possible to keep marching the patterns horizontally and vertically, during the acquisition stage, to increase the range data density. This extra time-multiplexed acquisition is not shown in the pseudocode above to avoid confusion, but is integrated within the second pass of the acquisition.

V. Experimental Validation

Several tests were performed to demonstrate the increased capabilities of the structured light sensor resulting from the enhancements brought forward with this work. The results of three evaluations that attempt to compare the 3D range data generated with and without the enhancements are presented here. It should be noted that the range data consist of 3D points. These are interpolated to generate a colored surface in the following examples to facilitate the visualization and interpretation of the results.

The first case consists of modeling a brightly colored children's dartboard and a textured wicker basket as shown in Fig. 5a. First, the colored pattern was projected and second, the time-multiplexed pseudo-color channels at maximum intensity were projected to compare performance. Both tests were performed using exposure fusion with the default of 10 images ranging from 5ms to 50ms of exposure time. To increase point density, the pattern was marched 3 times horizontally and 3 times vertically.

The results shown in Fig. 5b demonstrate that the original colored pattern performs poorly when highly colored objects are present such as the dartboard. The yellow and red areas are mostly impossible to detect. On the other hand, the basket contains several holes where shiny stripes appear on its surface. However, when using the proposed time-multiplexed pseudo-color channels, a high percentage of the scene can be modeled, as seen in Fig. 5c, except for only a small number of areas on the basket where there is very high reflection. More importantly, the sensor is not affected by the black numbers and lines on the dartboard as it is when the original colored projection is used.



(a) (b) (c)

Fig. 5. (a) Original image of children's dartboard and wicker basket, (b) its reconstruction using the original colored pattern, and (c) its reconstruction using the time-multiplexed pseudo-color channels at maximum intensity.

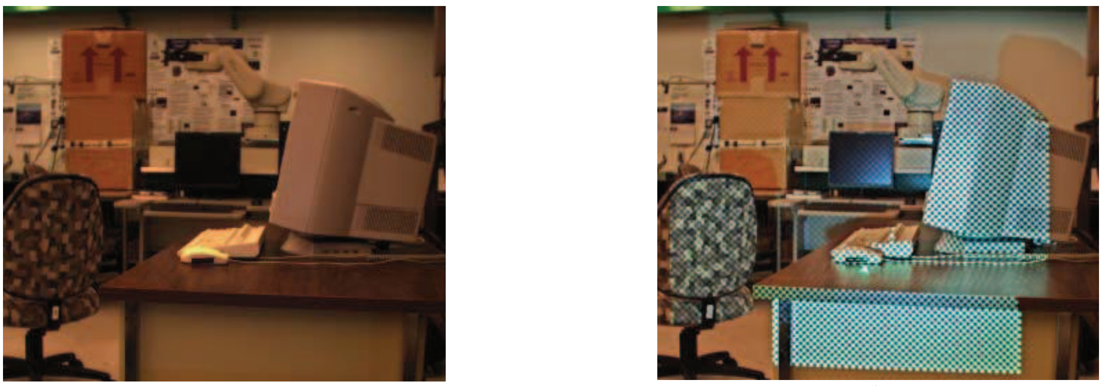
(a) (b) (c)

Fig. 6. (a) Original image of a computer monitor and a robot controller, (b) its reconstruction without using exposure fusion, and (c) its reconstruction using exposure fusion.

The second scenario demonstrates the advantage of using the exposure fusion algorithm when capturing 3D colored models of a computer monitor and a robot controller as shown in Fig. 6a. Two tests are performed using the time-multiplexed pseudo-color channels and a 3x3 marching pattern. In the first attempt, the images are captured using a single exposure of 10ms that was found to be optimal by trial and error. In comparison, the exposure fusion of 10 images ranging from 5ms to 50ms was used during the second image acquisition.

Both objects have bright and dark colored regions as well as surfaces with high reflectance characteristics such as glass and aluminum. When using a single global exposure, such as in Fig. 6b, it is impossible to detect the dark and reflective screen of the monitor as well as the dark grey section of the controller. Also, a specular highlight appears on the right half of the aluminum cross-section of the controller. The sensor has difficulty detecting the pattern in these regions and 3D information is lost. When the exposure fusion algorithm is applied, as in Fig. 6c, the glass screen is detected as well as all regions of the controller, including where the specular highlight appeared. It should be noted that the controller model still has some holes since its small features are below the resolution of the projected pattern's squares.

The third case exemplifies the increased depth of field of the sensor provided by the detection and processing of focus planes. A general scene consisting of an office desk and a robotic workcell, shown in Fig. 7a, is modeled. The depth of the scene from the front desk to the rear wall is roughly 7m. The sensor detected that 3 out of 6 possible focus planes were of interest. First, the front desk was processed, followed by the chair and closest computer monitor and finally the rest of the scene including the farther monitor, the robotic arm and the wall were processed. Again, a 3x3 marching pattern was used and the exposure fusion acquired 15 images from 5ms to 100ms for each retained focus plane.



(a) (b)

Fig. 7. (a) Original image of office desk and robotic workcell, with (b) pattern projected on scene during acquisition.

(a) (b)

Fig. 8. (a) Reconstructed scene from the sensor's viewpoint, and (b) same reconstruction from a side view showing the depth of field.

The results in Fig. 8 demonstrate that the sensor is capable of accurately imaging at several depths of field, with a maximum depth of field that well exceeds that of the popular Kinect sensor, which is limited to a maximum distance of about 3m. Objects that are close, such as the front desk, the chair and the closest computer monitor are modeled with great detail. However, as the distance increases, the projected squares of the pattern get larger, which results in a lower resolution. In addition, important occlusions of the projected pattern are created by the objects themselves, with respect to the projector's location, as can be seen in Fig. 7b. These areas cannot be reconstructed due to the lack of a visible pattern. This explains the missing regions in the models, though these could be recovered by moving the sensor.

VI. CONCLUSION

The proposed range imaging method extends the capabilities of structured light range sensors by introducing several enhancements that allow such technologies to model complex arbitrary scenes, as required in autonomous robotics for exploratory applications. The separate channels of a spatial-neighbourhood colored pattern are projected individually using the maximum intensity of the projector and a time-multiplexing approach to achieve better results on colored objects. Also, an exposure fusion algorithm combines multiple images collected with increasing exposure time for each image acquisition in order to reliably capture regions with low and high reflectance characteristics. Finally, a large depth of field is achieved by automatically detecting several focus planes over the scene, processing the corresponding regions of the field of view independently, and finally concatenating the resulting 3D points.

The enhancements are all strategically integrated into a custom acquisition procedure that remains flexible, easily configurable, and independent from the scene being imaged. When operating autonomously, the sensor is capable of capturing a maximum amount of 3D range data with no a priori knowledge of the scene and no parameter tweaking.

Future work will aim to optimize the displacements of the sensor and register all acquired point clouds into an integrated 3D colored model. Leveraging the sensor's large depth of field and adaptive focal planes, the robot can grasp its environment with minimum displacement. Afterwards, it can concentrate on regions of interest and obtain high-density range data from selected regions, regardless of object colors or reflectance characteristics.

REFERENCES

[1] D. Desjardins and P. Payeur, "Dense Stereo Range Sensing with Marching Pseudo-Random Patterns", in *Proceedings of the 4th Canadian Conference on Computer and Robot Vision, Montreal*, QC, 2007, pp. 216-226.

[2] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334, Nov. 2000.

[3] R. Hartley and A. Zisserman, *Multiple View Geometry*, Cambridge University Press, UK, 2003.

[4] R.A. Morano, C. Ozturk, R. Conn, S. Dubin, S. Zietz, and J. Nissanov, "Structured Light Using Pseudorandom Codes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 322-327, Mar. 1998.

[5] J. Salvi, J. Pagès, and J. Batlle, "Pattern Codification Strategies in Structured Light Systems," *Pattern Recognition*, vol. 37, no. 4, pp. 827-849, Apr. 2004.

[6] D. Caspi, N. Kiryati, and J. Shamir, "Range Imaging with Adaptive Color Structured Light," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 5, pp. 470-480, May 1998.

[7] T.P. Koninckx, P. Peers, P. Dutré, and L. Van Gool, "Scene-Adapted Structured Light," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 611-618.

[8] D. Skocaj and A. Leonardis, "Range Image Acquisition of Objects with Non-Uniform Albedo using Structured Light Range Sensor," in *Proceedings of 15th International Conference on Pattern Recognition*, Barcelona, 2000, pp. 778-781.

[9] A.A. Goshtasby, "Fusion of Multi-Exposure Images," *Image and Video Computing*, vol. 23, no. 6, pp. 611-618, Jun. 2005.

[10] I. Zafar, E.A. Edirisinghe, and H.E. Bez, "Multi-Exposure & Multi-Focus Image Fusion in Transform Domain," in *Proceedings of the IET International Conference on Visual Information Engineering*, 2006, pp. 606-611.

[11] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure Fusion," in *Proceedings of the 15th Pacific Conference on Computer Graphics and Applications*, Maui, HI, 2007, pp. 382-390.

[12] I. De and B. Chanda, "A Simple and Efficient Algorithm for Multifocus Image Fusion using Morphological Wavelets," *Signal Processing*, vol. 86, no. 5, pp. 924-936, May 2006.

[13] H. Hariharan, A. Koschan, and M. Abidi, "Multifocus Image Fusion by Establishing Focal Connectivity," in *Proceedings of the IEEE International Conference on Image Processing*, San Antonio, TX, 2007, pp. III-321-III-324.