

Radiometric Compensation in a Projector-Camera System Based on the Properties of Human Vision System

Dong Wang

Imari Sato

Takahiro Okabe

Yoichi Sato

Institute of Industrial Science
The University of Tokyo
Tokyo, Japan

Abstract

We introduce a novel technique for performing radiometric compensation for a projector-camera system that projects images onto a textured planar surface, which is designed to minimize perceptual artifacts visible to observers according to a model of the human vision system. A projector-camera system has previously been proposed for projecting images onto an arbitrary surface using radiometric compensation, however the dynamic range of a projector is physically limited there are some textures that cannot be compensated correctly. Also, human eyes are sensitive to artifacts introduced in this way. Our technique is designed to provide compensated images with perceptually less noticeable artifacts while preserving enough brightness and contrast in the output. We develop an optimization framework based on a perceptually-based physical error metric to minimize perceptible artifacts in the final compensated images by compressing the contrast of the input images.

1. Introduction

Recently, a variety of novel projected display systems have been developed, such as immersive display systems [20], large seamless display systems [21, 22, 1, 2, 3, 4], and shadow elimination for multi-projector displays [23].

Besides these efforts, several radiometric compensation methods have been proposed to allow projection of images onto arbitrary surfaces, such as surfaces with their own textures [25, 13, 15]. These methods are designed to relax the severe requirement of conventional projection display systems which require high quality screens for projection. These efforts make the projection system significantly more convenient and useful.

Unfortunately, because the dynamic range of the projector is physically limited, the radiometric compensation system will encounter difficulties when the output of the projector saturates. The saturation of output will cause cut-off and result in perceptible artifacts in the final compensated

images. More importantly, observers are sensitive to these artifacts.

What needs to be emphasized here is that, although it is reasonable to assume that the dynamic range problem can be solved by improving the dynamic range of the projector or using multiple projectors, such as [15], it may result in higher financial cost or make the radiometric system overly complex. We want our radiometric compensation system to remain low-cost and simple. We therefore develop a radiometric compensation projection system that can minimize perceptible artifacts caused by the limited dynamic range of the projector while preserving the photometric quality (brightness and contrast) of input images. Our basic principle is to properly compress the contrast of the input image based on the properties of the human vision system.

Relevant to our radiometric compensation problem, several methods have been developed to acquire brightness uniformity in a multi-projector display system. In [1, 3] the brightness uniformity across and within multiple projectors is achieved by matching the brightness of each pixel into a spatially-uniform dynamic range which is the most limited one of all the projectors, which causes significant loss of photometric quality of the input image because of the severe compression of contrast.

On the other hand, a method named PRISM [2] is designed to achieve brightness uniformity while preserving enough brightness and contrast by mapping input images into a properly smoothed spatially-varying dynamic range of the projector. PRISM exploits the property of the human vision system that human eyes are not so sensitive to smooth brightness variation. It should be noted that, although PRISM is a method that incorporates properties of the human vision system, it relies only on the spatially-varying dynamic ranges of multiple projectors and has no regard for the contents of input images.

In contrast to PRISM, we want to develop a method that incorporates more details about the human vision system. When an input image is being projected, we simulate the perception of a human observer based on the properties of

the human vision system. Note that this simulation depends not only on the dynamic range of projector but also on the contents of input images, and it requires better understanding of the human vision system.

The properties of the human vision system have been taken into account in some other research areas. For instance, in order to display high-dynamic-range (HDR) images [17] on a conventional display device such as a monitor, a projector, or a printer etc., a great number of methods [7, 11, 6, 16, 10, 26] have been proposed and these efforts are generally described as tone mapping or tone reproduction methods. Some of these methods, such as [7, 10, 26], are based on visibility matching or some computational model of the human vision system. For instance, in [26] tone mapping is based on a multi-scale model of adaption and spatial vision.

Besides these tone mapping methods, some computational models of the human vision system have been developed such as [24, 12]. These models are usually used to evaluate the perceptual differences between images or used as image quality metrics.

Recently, a perceptually-based physical error metric for accelerating realistic image synthesis has been proposed [14]. Given an input image, a corresponding image called the threshold map is computed based on a computational model of the human vision system. This threshold map can correctly predict the perceptual threshold for detecting artifacts of scene features. After the computation of this threshold map, images can be compared directly in the physical luminance domain, while still accounting for the properties of the human vision system.

In this paper, we develop a radiometric compensation system that can project images onto a textured planar surface. In previously proposed radiometric compensation systems there are some artifacts in the final compensated images caused by the physically limited dynamic range of the projector. This will result in a severe problem because humans are sensitive to these artifacts. Our system is designed to cope with this problem.

2. Our Proposed Method

Our system is designed to provide compensated images with less noticeable artifacts while preserving the photometric quality of the input image. To provide compensated images with less noticeable artifacts, we compress the contrast of the input image to avoid the cut-off caused by the physically limited dynamic range of the projector. To preserve the photometric quality of the input image we need to keep the contrast compression scalar close to 1. We regard this as an optimization problem and our basic idea is to properly compress the contrast of the input image based on the properties of the human vision system. Note that, those

artifacts where human eyes are sensitive will result in significant errors. We develop an optimization framework by incorporating a perceptually-based physical error metric to take account of the properties of the human vision system.

In this section we first state some assumptions. We then introduce two calibration procedures: a simplified radiometric model based on the model developed in [25] to illustrate the idea of radiometric compensation, and a calibration procedure between the physical luminance domain and the pixel value domain which is necessary for computing the perceptually-based physical error metric. After that, we describe the perceptually-based physical error metric, and propose our optimization framework.

2.1. Assumptions

We assume:

1. 8-bit gray-scale images
2. Planar Lambertian projection surfaces
3. No ambient illumination
4. A single global scalar value α to compress the contrast of the input image
5. Linear response properties of the camera

In this paper, we only consider the achromatic artifacts in the compensated images, because the human vision system is more sensitive to luminance variation than to chrominance variation [18, 19, 5]. We will cope with chrominance artifacts in the future. We assume that we have a planar surface for projection because geometric calibration is not the focus of this paper, and we also ignore specular reflectance and ambient illumination. We use a spatially-uniform scalar to compress the contrast of the input image in our optimization framework. Spatially-varying scalars will be considered in the future when accounting for errors caused by local discontinuities. We assume that the camera will be calibrated independently. We have found that the response of our camera is approximately linear so we do not consider any nonlinearity in this paper.

2.2. System Calibration

Our radiometric compensation system, which is similar to the system used in [25] is shown in Figure 1. We use a Sanyo PLC-XP45 projector with a native resolution of 1024×768 pixels. The camera is a Sony DXC-9000 with a resolution of 640×480 pixels. The images through the camera are captured by a Matrox Meteor II frame-grabber.

Because we concentrate on the radiometric compensation part, our geometric calibration part is very simple based on the assumption that we project images onto a planar surface. We project several straight lines and find some point



Figure 1: Our projector-camera system.

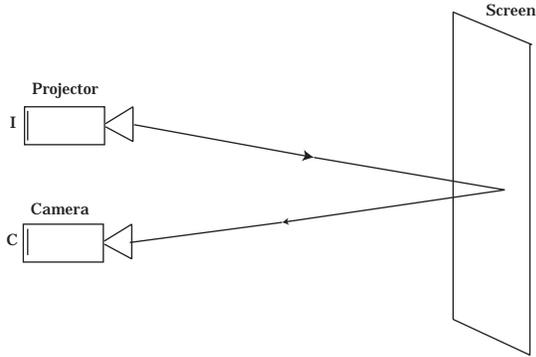


Figure 2: The simplified dataflow pipeline of a projector-camera system

correspondences to compute the homography between the image plane of the projector and that of the camera.

In this radiometric compensation system, we desire the compensated result image to be exactly the same as the input image. Our compensation algorithm is based on the radiometric model developed in [25]. A simplified dataflow pipeline for a projector-camera system is shown in Figure 2.

Because we only consider the special case of projecting gray-scale images, the radiometric model of the whole system can be simplified and represented by using a per pixel non-linear monotonic response function. For a given pixel, we have:

$$C(x, y) = H(I(x, y)) \quad (1)$$

where $I(x, y)$ is the input gray-scale image to be projected, and $C(x, y)$ is the image captured by the camera, H stands for the radiometric correspondence between $I(x, y)$ and $C(x, y)$. If the response function H can be determined, we can achieve any desired image $C(x, y)$ by projecting a com-

penation image instead of $I(x, y)$. That is, to acquire the correct input image $I(x, y)$, we compute a compensation image by:

$$\tilde{I}(x, y) = H^{-1}(I(x, y)) \quad (2)$$

where $\tilde{I}(x, y)$ is the compensation image. If we project this compensation image $\tilde{I}(x, y)$, then ideally the final compensated image captured by the camera will be:

$$I(x, y) = H(\tilde{I}(x, y)) \quad (3)$$

We use a similar calibration procedure as described in [25] to determine this response function H . We use a similar calibration procedure as described in [25] to determine this response function H . We project a set of 256 uniform gray patterns with their gray levels ranging from 0 to 255, and record the corresponding images captured by the camera. This procedure results in a per-pixel radiometric correspondence between $I(x, y)$ and $C(x, y)$.

We then introduce the calibration procedure between the physical luminance domain and the pixel value domain which is necessary for computing the perceptually-based physical error metric. Because the perceptually-based physical error metric (which will be described in Section 2.3) which incorporates the properties of the human vision system has to be computed in the physical luminance domain, given an input image, we have to transform its pixel values to physical luminance values. We determine this correspondence by a simple calibration procedure.

Similar to the radiometric calibration we project several flat patterns then capture them by the camera and simultaneously we use a spectroradiometer to record the physical luminance. Note that, because we focus on the correspondence between the pixel values on the camera plane and their corresponding physical luminance, for simplicity we assume the response of the camera is spatially uniform. We use a high quality screen for projection and this whole calibration procedure is implemented in a dark room.

This calibration procedure finally gives us the correspondence between pixel values on the camera plane and physical luminance values. Then we can compute the threshold map using the threshold model which will be described in Section 2.3. The details about the computation of the threshold map are described in [14] and are beyond the scope of this paper.

2.3. A Perceptually-Based Physical Error Metric

We incorporate a perceptually-based physical error metric proposed in [14] to account for the properties of the human vision system.

Given an input image, a corresponding image called the threshold map (we denote it $T(x, y)$, and describe it in the

next section) is computed based on a threshold model which incorporates the three main properties of the human vision system that human eyes are not very sensitive to those scene features with high background illumination levels, high spatial frequencies, and high contrast levels. These three properties are called threshold sensitivity, contrast sensitivity and visual masking.

Threshold sensitivity is generally specified using a threshold-vs-intensity, or TVI function [10] which describes the threshold sensitivity of the human vision system depending on the background luminance. The TVI threshold for a given uniform background of luminance L is the minimum luminance increment ΔL such that a test spot in the center of luminance $L + \Delta L$ can be detected by an observer.

The contrast sensitivity function, or CSF [26], represents the sensitivity of the human vision system to the range of spatial frequencies found in complex images. The human vision system is most sensitive to scene features with spatial frequencies in the range of 2 to 4 cycles per-degree (cpd) of visual angle and drops off significantly at higher and lower frequencies. This property is generally modelled as the result of spatial processing of the frequency patterns by multiple bandpass mechanisms. Each mechanism processes only a small band of spatial frequencies from the range over which the visual system is sensitive.

Visual masking [9] is the property of the human visual system by which image features with high contrast will dominate over lower-contrast features with similar spatial frequencies and orientations. This compressive non-linearity results in further elevation of threshold with increases in the contrast of the pattern.

These three main properties of the human vision system have been incorporated in the perceptually-based physical error metric we use. Given an input image, a luminance-dependent threshold is computed from the TVI function. Contrast sensitivity and visual masking are used as elevation factors to the luminance-dependent threshold. Implementation details are described in [14].

2.4. Optimization Framework

Based on the radiometric model and the perceptually-based physical error metric, we present our optimization framework. First, we describe our definitions of variables as follows.

1. (x, y) : coordinates on the camera plane
2. $I(x, y)$: input image which is to be projected
3. $R(x, y)$: image captured by the camera when the projector is projecting a uniform white pattern at full power (level 255.)
4. R_{max} : the global maximum of $R(x, y)$.

5. $O(x, y)$: final compensated image measured by the camera. $O(x, y)$ ranges from 0 to $R(x, y)$ based on the assumption that the ambient light can be ignored.
6. $T(x, y)$: the threshold map computed by the threshold model described in Section 2.3.

Let us assume that we have an ideal projection surface with no texture (pure white.) In this case, we can assume that for all x and y we have $R_{max} \equiv R(x, y)$. Then we compute $O(x, y)$ as:

$$O(x, y) = I(x, y) \times \frac{R_{max}}{255} \quad (4)$$

When the projection surface has some texture, we have:

$$O(x, y) = \begin{cases} \alpha I(x, y) R_{max} / 255, & (\alpha I(x, y) R_{max} / 255 \leq R(x, y)) \\ R(x, y), & (\text{otherwise}) \end{cases} \quad (5)$$

where α is a global scalar. Note that, because the dynamic range of projector is physically limited some regions in the output image $O(x, y)$ with pixel values greater than $R(x, y)$ will be cut off. If we use the global scalar α to compress the contrast of input image $I(x, y)$ we need to find a good value of α that minimizes cut-off errors while remaining close to 1 (no contrast compression) as possible.

We assume that when $I(x, y) R_{max} / 255$ changes by less than the threshold value, that is, the change is in the range $[-T(x, y), T(x, y)]$, observers cannot distinguish the difference. When we use a global scalar α to compress the contrast of input image $I(x, y)$, the errors caused by the limited dynamic range of projector will be:

$$e_{ar}(x, y) = \begin{cases} 0, & (\alpha (\frac{I(x, y) R_{max}}{255} - T(x, y)) \leq R(x, y)) \\ (\alpha (\frac{I(x, y) R_{max}}{255} - T(x, y)) - R(x, y))^2, & (\text{otherwise}) \end{cases} \quad (6)$$

where e_{ar} is the error caused by artifacts because of the limited dynamic range of the projector. Note that, because the threshold map must be computed in the luminance domain, to compute $T(x, y)$ we first transform $I(x, y) R_{max} / 255$ to the luminance domain, compute the threshold map using the threshold model, then transform it back to pixel values. We use the correspondence between pixel values on the camera plane and physical luminance values described in Section 2.2 to implement this transformation. Another thing we have to point out is that, based on the Weber's law, when the

input image is compressed by a global scalar α , the corresponding threshold $T(x, y)$ will change to $T(x, y)/\alpha$.

We also have to preserve enough contrast in the input image. The photometric quality degradation caused by the contrast compression can be described as:

$$e_{de} = (1 - \alpha)^2 \quad (7)$$

where e_{de} evaluates the photometric quality degradation of the input image caused by contrast compression.

Then our final error metric becomes:

$$E = \int \int (e_{ar}(x, y) + \lambda e_{de}) dx dy \quad (8)$$

where E is the final error metric, and the integration domain is the whole input image. λ is a constant parameter that can make $\int \int e_{ar}(x, y) dx dy \simeq \int \int \lambda e_{de} dx dy$. With λ chosen in this way we have found that the global scalar turns out to be stable, and subjective evaluations from several observers have indicated that this method compensates effectively for non-uniform surface texture.

We can then calculate the optimal global scalar α by minimizing E . We use this optimal scalar α to compress the contrast of the input image, then compensate this compressed image using our radiometric compensation method. The resulting compensation image is:

$$\tilde{I}(x, y) = H^{-1}(\alpha I(x, y) R_{\max}/255) \quad (9)$$

Because our method permits artifacts in the final compensated images where humans are not very sensitive we can produce compensated images with relatively high brightness and contrast.

3. Results

An example input image that we wish to display is shown in Figure 3. In this image, those regions with high luminance, high spatial frequencies, or high contrast such as the castle and the trees have relatively high threshold values. Namely, human eyes are not very sensitive to these regions. Figure 4 shows the textured screen for projection. Because we only consider projecting gray-scale images, there is no color in our textured screen. Figure 5 shows the uncompensated image. We can see that the input image is modulated by the spatially varying albedo of the screen. The compensated result image without contrast compression is shown in Figure 6. There are some significant artifacts in the final compensated image because in these regions the output of the projector saturates. This is a severe problem because human eyes are very sensitive to these artifacts.

Figure 7 shows the threshold map used in our method. Note that, the threshold values have been adjusted for display. We can see that those regions with high luminance,

high spatial frequencies, and high contrast have relatively high threshold values. For instance, the region of the trees has low luminance but the threshold value is elevated because of its high spatial frequencies.

Figure 8 shows the compensated image which has its contrast compressed by our method. The global scalar α is 0.5678. We can see that the perceptible artifacts are reduced significantly while the photometric quality is preserved.

4. Conclusions and Future Work

We have focused on a severe limitation of radiometric compensation systems, namely that artifacts are produced in the final compensated images by the limited dynamic range of the projector.

We have developed an optimization framework to solve this problem based on a threshold model of the human vision system. Our technique has shown that if input images can be compressed properly based on the properties of the human vision system we can achieve compensated images with perceptually less noticeable artifacts while preserving reasonable brightness and contrast in the input images.

In future work we will implement radiometric compensation of color images and develop an optimization framework that accounts for the chromatic sensitivity of the human vision system. Our method may also be extended to a framework which includes localized scalars and more factors such as offset to account for ambient light, error caused by local discontinuities, spatiotemporal sensitivity and visual attention [8] to generate better compensated images. We also need to accelerate the computation of the threshold map to make our framework easy to deploy.

References

- [1] A. Majumder and R. Stevens, "LAM: Luminance attenuation map for photometric uniformity in projection based displays," *Proceedings of ACM Virtual Reality and Software Technology*, pages 147-154, 2002.
- [2] A. Majumder and M. S. Brown, "Building Large Area Displays," *Eurographics*, 2003.
- [3] A. Majumder, D. Jones, M. McCrory, M. E. Papka, and R. Stevens, "Using a camera to capture and correct spatial photometric variation in multi-projector displays," *IEEE International Workshop on Projector-Camera Systems*, 2003.
- [4] A. Majumder and R. Stevens, "Color nonuniformity in projection-based displays: Analysis and solutions," *IEEE Transactions on Visualization and Computer Graphics*, 10(2):177-188, March/April 2004.
- [5] E. Bruce Goldstein, "Sensation and Perception. Wadsworth Publishing Company," *In Displays*, 2001.

- [6] F. Durand and J. Dorsey, "Fast Bilateral Filtering for the Display of High-Dynamic-Range Images," *In SIGGRAPH*, 2002.
- [7] G.W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Trans. Visual. Comput. Graph.*, vol. 3, pp. 291-306, Oct./Dec. 1997.
- [8] H. Yee, S. Pattanaik, and D. P. Greenberg, "Spatiotemporal sensitivty and visual attention for efficient rendering of dynamic environments," *ACM Transactions on Graphics*, 20(1), January 2001.
- [9] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg, "A Model of Visual Masking for Computer Graphics," *In SIGGRAPH 97 Conference Proceedings*, pages 143-152, Los Angeles, California, August 1997.
- [10] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg, "A model of visual adaptation for realistic image synthesis," *In SIGGRAPH 96 Conf. Proc.*, pp. 249-258, 1996.
- [11] J. Tumblin and G. Turk, "LCIS: A boundary hierarchy for detail-preserving contrast reduction," *In SIGGRAPH 99 Conf. Proc.*, 1999.
- [12] J. Lubin, "A Visual Discrimination Model for Imaging System Design and Evaluation," *Vision Models for TargetDetection and Recognition*, Eli Peli, Editor, World Scientific, New Jersey, pp. 245-283, 1995.
- [13] M. D. Grossberg, H. Peri, S. K. Nayar, and P. N. Belhumeur, "Making One Object Look Like Another: Controlling Appearance Using a Projector-Camera System," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington DC, June 2004.
- [14] M. Ramasubramanian, S. N. Pattanaik, and D. P. Greenberg, "A perceptually based physical error metric for realistic image synthesis," *In Alyn Rockwood, editor, SIGGRAPH 99 Conference Proceedings, Annual Conference Series*, pages 73-82. ACM SIGGRAPH, Addison Wesley, aug 1999.
- [15] O. Bimber, A. Emmerling, and T. Klemmer, "Embedded Entertainment with Smart Projectors," *IEEE Computer*, pp. 56-63, January 2005.
- [16] P. Choudhury, J. Tumblin, "The Trilateral Filter for High Contrast Images and Meshes," *In Proc. of the Eurographics Symposium on Rendering*, Per. H. Christensen and Daniel Cohen eds., pp. 186-196, 2003.
- [17] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," *In SIGGRAPH 97 Conf. Proc.*, pp.369-378. 1997.
- [18] R.A. Chorley and J. Laylock, "Human factor consideration for the interface between electro-optical display and the human visual system," *In Displays*, volume 4, 1981.
- [19] R. L. De Valois and K. K. De Valois, "Spatial Vision," *Oxford University Press*, 1990.
- [20] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, "The office of the future: A unified approach to image-based modeling and spatially immersive displays," *Proc. of SIGGRAPH*, pages 179-188, 1998.
- [21] R. Raskar, M.S. Brown, R. Yang, W. Chen, H. Towles, B. Seales, and H. Fuchs, "Multi projector displays using camera based registration," *Proceedings of IEEE Visualization*, pages 161-168, 1999.
- [22] R. Raskar, "Immersive planar displays using roughly aligned projectors," *In Proceedings of IEEE Virtual Reality 2000*, pages 109-116, 1999.
- [23] R. Sukthankar, T.J. Cham, and G. Sukthankar, "Dynamic shadow elimination for multi-projector displays," *In IEEE Comp. Vis. and Patt. Recog.*, pages II:151-157, 2001.
- [24] S. Daly, "The Visible DiReferences Predictor: An Algorithm for the Assessment of Image Fidelity," *Digital Images and Human Vision*, A. B. Watson, Editor, MIT Press, Cambridge, MA, pp. 179-206, 1993.
- [25] S. Nayar, H. Peri, M. Grossberg, and P. Belhumeur, "A projection system with radiometric compensation for screen imperfections," *In IEEE International Workshop on Projector-Camera Systems*, Oct. 2003.
- [26] S. Pattanaik, J. Ferwerda, M. Fairchild, and D. Greenberg, "A multiscale model of adaptation and spatial vision for realistic image display," *In SIGGRAPH 98 Conf. Proc.*, pp. 287-298, 1998.



Figure 3: The desired input image

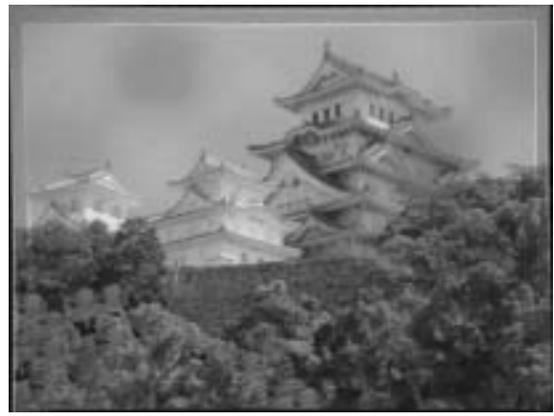


Figure 6: Compensated image without contrast compression. We can see that there are some artifacts that cannot be compensated correctly because of the limited dynamic range of the projector.



Figure 4: The textured screen



Figure 7: The threshold map



Figure 5: The uncompensated image.

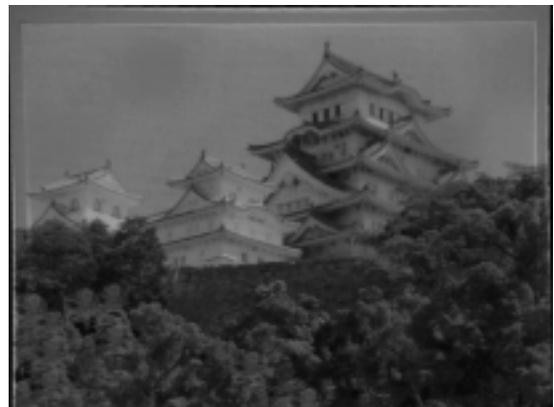


Figure 8: Compensated image with contrast compression. The value of the contrast compression scalar α is 0.5678, as determined by our framework. We can see that the perceptible artifacts in the final compensated image are significantly reduced, while preserving the photometric quality of the input image.