Localized High-Dynamic-Range Method to Enhance Images with Low Reflectance Subjects

Bonnie Tseng
Stanford University
tsengb@stanford.edu

Abstract

Textures and details of low-reflectance objects are often difficult to photograph because the object does not reflect enough light to the image sensor. As a result, low reflectance objects can look underexposed in images even when photographed in a well-lit scene. A common technique to improve such images is high-dynamic-range (HDR) imaging. In this technique, multiple low-dynamic-range (LDR) images captured at different exposures are merged to create a single high-quality image with a greater dynamic range. HDR images typically require additional processing before being displayed due to limitations of display contrast on standard monitors. In this paper, I present a localized high-dynamic-range method which aims to apply HDR processing specifically to the detected low-reflectance subject in the scene. The resulting images show improvement in the amount of detail visible in the subject while preserving the colors of the surrounding scene without additional processing required.

1. Introduction

By merging multiple low-exposure images to create an image that is well-exposed in all areas of the captured scene, high dynamic range imaging can significantly improve the look of scenes with a large range of intensity. Many captivating images can be found online of buildings and landscapes that show incredible detail in the image that would not be visible without HDR processing to correct under- and overexposed areas. It has been demonstrated in the homework problems, however, that producing a good HDR image is a difficult task. The main problem lies in the fact that HDR techniques produce an image that cannot be displayed in displays which have a lower dynamic range capability. Tone mapping methods can approximate the appearance of HDR images into a lower dynamic range for display, but the resulting images often appear very unnatural.

The images that are addressed in this paper are specific to photographs of subjects with low-reflectance. Common subjects include dark-colored animals, hair, clothing and skin tone. These images can be considered a specific type of HDR image. In these images, most of the image is well-exposed, but the subject will appear underexposed because of their low-reflectance nature. Commercial cameras do this to accommodate the dynamic range of most of the scene, not accounting enough for the darker subjects. As a result, the subject ends up being underexposed, and details are lost because the dynamic range of the real scene exceeds the dynamic range of the photo. In such a case, the same scene can be captured at the regular exposure along with one or two higher exposures. This algorithm can use the higher exposures to produce a modestly edited image, where the subject is slightly more detailed and the well-exposed areas of the scene are unchanged.

This report first discusses the related works and datasets consulted, along with how they were selected. The image processing pipeline is then described, consisting of image segmentation, HDR tone mapping, and how these two are merged to form a composite image. There is no ground truth to quantitatively compare the results against, so results are analyzed qualitatively against the original dataset images. Finally, the report suggests future work to improve the pipeline.

2. Related Works

Beran et al. [1] describes a color-based object detection method. In this approach, an image is converted from RGB to HSV color model to separate intrinsic color value from hue and saturation before segmentation. Chen et al. [2] proposes an image segmentation method which utilizes both color and texture information. This method can be used to detect improvement in texture over different exposures which can help with subject identification. Guthier et al. [3] presents an optimized HDR approach in which only badly exposed regions of an image are re-exposed thus reducing the total capture time

3. Datasets

The first image set used is Google's HDR+ Library [4], comprised largely of portraits and everyday scenes. HDR+ process differs from the typical bracketed exposure HDR process such that, rather than capture a burst of images with different exposures, they capture a burst of images with the same exposure. An alignment algorithm is then used to merge the images into a single low dynamic range image with much lower noise. To mimic multiple exposures, gain is applied to the image to simulate longer exposure times

The second image set consists of images I captured myself on an Apple iPhone X. Each scene is photographed at 2-3 different exposures. The images have some added noise due to imperfect alignment when taking the multiple exposures. Both methods of capturing HDR images should lead to resulting images with improvement over the reference images.

In total, 4 images from Google HDR+ dataset and 2 personally captured images are used in this study.

4. Technical Approach

The localized HDR problem can be divided into two independent parts. The first part is image segmentation, using the HSV color model, a mask is created to indicate where the dark subject is located in the reference image. The second part is HDR image fusion and tone mapping. In this part, the burst of images captured are fused to form an HDR image and then tone mapped so the image can be displayed. The output of these two parts are then merged to create a composite HDR image that is localized on the subject. The pipeline was implemented in MATLAB which following two subsections provide more details of.

4.1. Image Segmentation

The image segmentation pipeline implemented is loosely based on the method proposed by Bora et al [1] with experimentation to find the optimal pipeline to isolate for dark subjects. First, the RGB image is converted to HSV color space. In this color space, each pixel is defined by its hue, saturation and value rather its red, green and blue components.

To best isolate a dark colored object from the surrounding scene, the difference between the value and hue channel was used for segmentation. The hue channel is useful for labeling pixels by color because each color is represented by a different hue. In the value channel, dark colors have values closer to zero and brighter colors will have larger values. Subtracting the value channel from hue helps to further isolate the dark subject because brighter are will not be diminished while the dark areas will remain unaffected since its value in the value channel was close to zero. The resulting image is bright in the areas where the subject is predicted to be and dark in all other areas. A challenge with isolating for a dark color in the HSV color space is that the hue channel is very noisy for dark colors making it difficult to define a single boundary of the object. To reduce the noise, a median filter is applied to the image. The benefit of using a median filter is that it preserves edges while denoising the non-edge parts of the image. After filtering, the image is then partitioned by K-means clustering. This is a classification algorithm that sorts into an image into a predefined set of clutters. By specifying a clutter number of 2, the pixels are separated into two groups that distinguish which pixels are part of the image and which are part of the surround scene. This creates the rough outline of the mask. To further improve the mask, small holes in the image are filled into to create a more solid mask to avoid unnecessary pockets that can create artifacts in the final image. A major issue with K-means clustering is that it looks solely that the

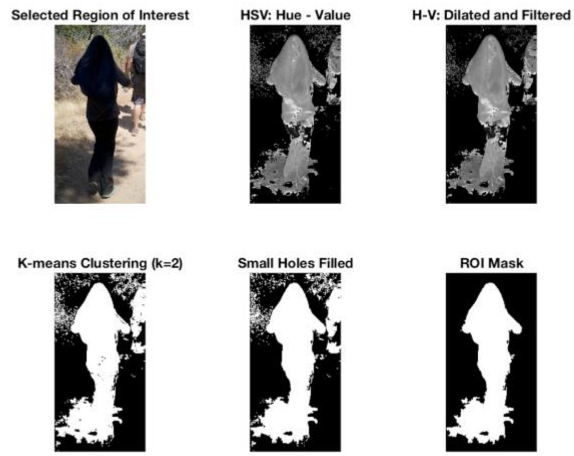

Figure 1: Output of intermediate steps in the image segmentation pipeline to create a mask of the subject.

value of each pixel. This means that there are many parts of the image that are classified as part of the subject even if it may only be a small group of pixels far from the center. To correct for this, only the classified section with the largest area is considered part of the subject. Intermediate outputs of this pipelines are shown in Figure 1 below.

To ensure the correct region of the image is isolated and to reduce the processing time to create a mask, user input is request to select a rectangular region of interest around the subject. It is important to note there the pipeline assumes that the subject will take up at least 10% of the region of interest (ROI). The reason for this is to correct for images where the segmentation fails because the hue channel was too noisy. If at the end of the pipeline, the area of the subject detected is less than 10%, the image will be processed through the pipeline again, but this time the pixels are dilated to blur the image so the k-means clustering algorithm has a higher likelihood of classifying previously rejected parts of the subject more accurately.

4.2. HDR Image Fusion and Tone Mapping

The pipeline for HDR image fusion and tone mapping is identical to the process that was taught in class and that we implemented in the homeworks.

4.2.1 HDR Image Fusion

For each image in a burst of images captured at multiple exposures, a weight map is calculated to determine which parts of each image are best exposed. The weight of each pixel in an image is calculated as follows:

$$w_{ij} = \exp\left(-\frac{4\left(I_{linij} - 0.5\right)^2}{0.5^2}\right)$$

Where I_{linij} is the value of the of the pixel.

The images are then merged into a single image image based on the weights using the following minimization function:

$$minimize_x O = \sum w_i(\log(I_{lini}) - \log(t_i X))^2$$

where w_i is the calculated weight map for each image, I_{lini} is the LDR image, t_i is the exposure of the LDR image and X is the resulting HDR image.

4.2.2 Global Tone Mapping

Create the HDR fused image is only the first part of HDR image processing. Most commercial displays have a lower dynamic range than the HDR image so the image now has to be mapped to a lower dynamic range. In global tone mapping, gamma correction is applied to all channels image. This is a non-linear operation that maps the colors of the image to match how the human visual system perceives color.

4.2.3 Determine exposure times for HDR+ Images

As mentioned in Section 3, a burst of multiple exposures for images taken from the Google HDR+ library is simulated by applying different levels of gain to the aligned and merged reference mage. This means that there is no intrinsic exposure data to extract from the image file. For these images, the exposure is approximated by multiplying the gain by the exposure time of the reference image since intensity of the image is directly proportional to the amount of time the shutter is open. Therefore, a gain of 2 is equivalent to have double the exposure time.

4.3. Merging

Using the mask created by the image segmentation process described in Section 4.1 and the HDR image created in Section 4.2, the two outputs are merged using alpha blending to created the final localized HDR image. The flow chart in figure 2 produces a summary of the localized HDR pipelines proposed.

The output images are analyzed and compared in the following section.

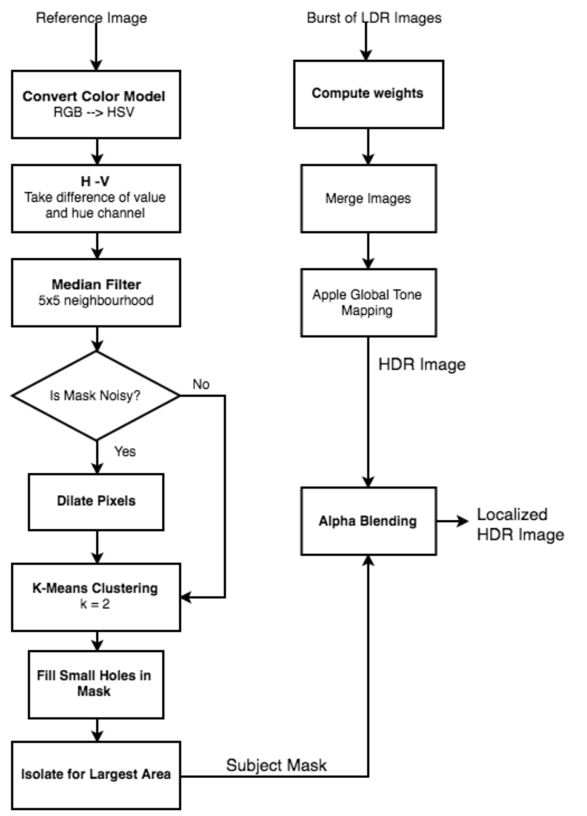

Figure 2: Output of intermediate steps in the image segmentation pipeline to create a mask of the subject.

5. Observations

To evaluate how well the proposed local HDR method works for images with low-reflectance objects, it is compared against global tone mapping and intensity channel gamma correction methods. Intensity channel gamma correction is a method in which gamma correction is applied to the intensity channel only. This is implemented as

$$I_{intensity} = 20 * red + 40 * green + blue/61$$

As with most HDR studies, there is no ground truth that dictates how the result should look making it hard to produce quantitative data for analysis. Instead 4 images are shown in Figure 3.

Localized HDR created a natural result in all cases. In image 1 and image 3, the detail in the subject's clothing is significantly improved with the proposed localized HDR method compared to the original. In image 1, the proposed method shows slightly different color in the pants compared to the jacket. In image 3, a hint of blue in the jacket can be seen which was not visible in the original. More details of the cat's features can be seen in image 4, particularly its right eye and furry tail.

Tone mapping works better on image 2, where the entire scene is brightened and shows more detail in the green pattern on the subject's shirt. With the localized HDR method, the scene stayed dark because we don't alter the areas not identified as the subject. The resulting localized HDR image showed little to no change from the original.

Gamma correction on the intensity channel is used to preserve the colors in the image while adjusting the brightness in the scene. In images 1 through 3, it has created unwanted, visible artefacts. For image 4, the resulting picture shows the cat in warmer colors which can be desirable in some cases. However, this study focuses on enhancing the image only when necessary, and gamma correction does not visibly show higher amount of detail compared to the proposed method.

6. Future Work

Lots of future work can be done this project. First, using convolutional neural networks to segment the image to significantly improve the detection of the subject in the iamge. This is because this method will also take into account the shape of the object and whether it resemble common objects like pets or humans. Second, blending of the two images can be improves. The current method still shows some artifacts around the edges of the merge.Using a gradient blend will reduce this artifact.

7. Acknowledgements

I would like to thank Julie Chang to being a wonderful advisor and providing lots of guidance in defining the scope of the project. I would also like to thank Professor Gordon Wetzstein for teaching me the fundamentals of cameras, imaging and display in this course.

References

[1] L. Beran, P. Chmelar, and L. Rejfek, "Image Processing Methods Usable for Object Detection on the Chessboard," MATEC Web of Conferences, vol. 75, p. 03004, 2016.

[2] J. Chen, T. Pappas, A. Mojsilovic, and B. Rogowitz, "Adaptive perceptual color-texture image segmentation," IEEE Transactions on Image Processing, vol. 14, no. 10, pp. 1524–1536, 2005.

[3] B. Guthier, S. Kopf, and W. Effelsberg, "Capturing high dynamic range images with partial re-exposures," 2008 IEEE 10th Workshop on Multimedia Signal Processing, 2008.

[4] Hasinoff, S., Sharlet, D., Geiss, R., Adams, A., Barron, J., Kainz, F., Chen, J. and Levoy, M. (2016). *Burst photography for high dynamic range and low-light imaging on mobile cameras*.

[5] Bora, D., Gupta, A. and Khan, F. (2019). *Comparing the Performance of L*A*B* and HSV Color Spaces with Respect to Color Image Segmentation*. [online] arxiv.org. Available at: https://arxiv.org/pdf/1506.01472.pdf [Accessed 13 Mar. 2019].

Figure 3: Comparison of Localized HDR Method Against Global Tone Mapping and Intensity Channel Only Gamma Correction