

Multi-image Processing Technologies for Better Images in Challenging Conditions

Mei Chen*, Suk Hwan Lim, Qian Lin, Hewlett-Packard Laboratories, Palo Alto, CA, USA

Abstract

Despite rapid development in digital photography, what is captured still often does not quite match what is perceived by the eye. This is particularly noticeable if the dynamic range of the scene is higher than what the imaging device is capable of capturing, in which case either detail in the shadows and/or highlight regions are lost. Another common scenario is under low light conditions where flash is not permitted, such as in a museum or at a concert, and one often ends up with a blurry or under-exposed photo. We present a multi-image processing system that registers a set of image sequence, extracts appropriate information from them, and forms a composite image with enhanced image quality that significantly exceeds the quality of each individual image. With our multi-image processing system, we address the problems of capturing images in scenes with low light and extended dynamic range. We also apply our solution to remove unwanted cluster from multiple image capture. We illustrate the effectiveness of our methods with experimental results from the three cases.

1. Introduction

Despite rapid development in digital photography, what is captured still often does not quite match what is perceived by the eyes. This is particularly noticeable if the dynamic range of the scene is higher than what the imaging device is capable of capturing, in which case, details in the shadows and/or highlight regions are lost. Another common scenario is under low light conditions where flash is not permitted, such as in a museum or at a concert, and one often ends up with a blurry or under-exposed photo.

Dynamic range is defined as the ratio between the maximum and minimum light intensity that can be reliably captured by the camera, from the darkest shadows to the brightest highlights. Modern cameras, even the most expensive ones, can only capture a limited dynamic range as compared to the human eye. With just a single shot, current cameras, film or digital, still do not allow different exposure settings on different areas of the scene. For example, when you take a photograph of a temple beneath a clear blue sky, the resulting picture, more often than not, will have a good exposure of either the temple's interior or the sky but not both. The exposure settings used for capturing the correct color and details of the interior are very likely to produce a washed-out (i.e. over-exposed) sky while the exposure settings used for capturing a blue sky will make the temple's interior under-exposed. Many image enhancement methods such as histogram equalization

and global/local contrast enhancement algorithms increase the contrast of under-exposed areas such as the shadow regions. However, they do not truly extend the dynamic range of the image. Note that in many cases when the contrast of the image is extended to the extreme, artifacts such as contouring and noise amplification can occur. Also, the image may look unnatural after applying contrast enhancement in the extreme cases.

While the dynamic range or the exposure of a single capture is limited by the intrinsic parameters of the imaging sensor, we can overcome this limitation by using multiple captures. For example, to capture a scene with high dynamic range, one can take two shots. One contains details in the highlight areas; the other contains details in the shadows. Most related work on dynamic range enhancement is typically limited to multiple captures on a tripod, or multiple captures with minimal image motion. Paul Debevec at SIGGRAPH 97 showed how to take multiple photographs at different exposures and merge them to create a single high-dynamic range image. This technique is now incorporated in products such as Photogenics. At the time of this presentation, the technique only works if the camera is mounted on a tripod. Adobe Photoshop recently also added a dynamic range enhancement feature by combining multiple frames into a photo with details both in the highlight areas and in the shadows. However, its ability to handle image motion is very limited.

Low light, no-flash conditions are also challenging for current image sensors. The automatic exposure mode will likely set the exposure time longer than what a user can hold the camera steadily, which results in a blurry photo. However, artificially shortening the exposure time would result in an under-exposed, low contrast image, which is not an option either. Unfortunately, some of the best opportunities for taking interesting photos occur indoors and in situations where a flash is not allowed. (e.g., live music concert, a children's play, at a museum, or at a wedding, where no flash during the ceremony is common.) Similar to the scenarios in dynamic range enhancement, a tripod would be required to obtain a sharp and clean image by extending the exposure time or by accumulating multiple frames.

To address these problems, we developed a multi-image processing system that automatically registers and synthesizes multiple frames to obtain a better image without the requirement of a tripod. In this paper, we describe a system that registers a sequence of images, extracts appropriate information from them, and forms a composite image with enhanced image quality. With this approach, we address the problems of capturing images in scenes with low light and extended dynamic range. Furthermore, we expanded our system to other applications such as unwanted

* now with Intel Research, Pittsburgh, PA, USA

object removal using multiple images. We illustrate the effectiveness of our methods with experimental results from these scenarios.

2. Our Approach

Multiple images of the scene can be acquired using various capture modes, such as manual, burst mode, video, etc. Whether the goal is to enhance the dynamic range of an image or to improve low light/no-flash exposure, in order to analyze and synthesize multiple images captured without a tripod, the first and critical step is *image motion estimation/registration* to account for camera motion (and optionally, potential object motion in the scene). Once the images are *registered* and *resampled* to a common reference frame, application-specific analysis can be done to extract the desired information from each image. Masks are generated for each image to determine the contribution of each pixel in the combination process. Finally, appropriate information from each image is synthesized for the particular application to generate an output composite image. Figure 1 shows the block diagram of the overall system.

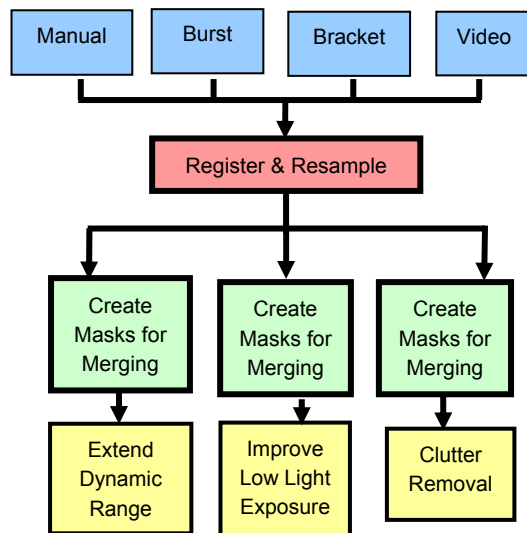


Figure 1: Block diagram of overall system

2.1. Image Registration

We adopt the affine model for camera motion, and use a multi-scale image alignment technique. Note that this is more challenging than the case for super-resolution, largely because it is not possible to compute motion for saturated pixels whose brightness information is lost. To solve this problem, we perform motion estimation only using non-saturated areas, and assume that image motion in the saturated region is coherent with its neighboring regions. Laplacian pyramids are constructed for the input images, and motion parameters are computed in a coarse-to-fine manner. At each resolution, the sum of squared differences integrated over regions of interest serves as the objective function

to be minimized. Additionally, to make images of different exposures more *comparable*, multi-level intensity equalization is carried out along with the coarse-to-fine motion estimation process.

2.2. Dynamic Range Enhancement

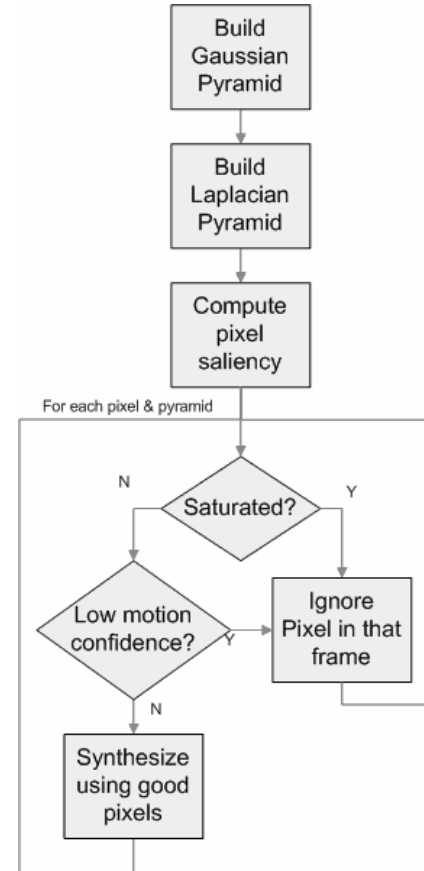


Figure 2: Block diagram of multi-frame dynamic range enhancement

Our technology extends image dynamic range by composing an optimized image using several captures of the same scene taken with different exposures. After image motion estimation and compensation, information from multiple exposures is synthesized to render an image with extended dynamic range. Specifically, corresponding unsaturated pixels from multiple exposures are weighed by the confidence of their motion estimates, their feature saliency measure, and combined proportionally. Figure 2 is the flow diagram of the process. The reference image here is chosen as the image with the most number of unsaturated pixels.

The feature saliency function can be the absolute value (amplitude), square value (energy), or a weighted neighborhood average (e.g. a 3-tap Gaussian filtering) of the Laplacian filtering response. From our experience, the amplitude is a good compromise between efficiency and performance. Image synthesis is done in a multi-resolution, coarse-to-fine style, so as to avoid introducing inconsistencies/artifacts when integrating information from multiple images (as illustrated in the lower diagram of Figure 2). There are two modes for generating the composite image, one is *pick 'n choose*, in which case the pixel with the highest saliency

is used in the output image; the other mode is *blending*, in which case a weighted combination of all corresponding un-saturated pixels is computed. In general, *pick 'n choose* yields more dramatic improvement in areas where a majority of the captures have low texture/contrast, whereas *blending* is more reliable when there exists considerable high frequency noise in input images.

2.3. Low Light Exposure Enhancement

For the low light no-flash situation, our approach is to capture several under-exposed yet sharp photos, and use multi-image processing to synthesize them into a sharp image with boosted exposure level. Simply accumulating multiple images without accounting for camera or object motion may result in a blurry photograph. We register the subsequent frames to the reference frame and compensate for both camera and object motion. Note that since most of the images taken under low light conditions have low signal-to-noise-ratio, it is challenging to obtain accurate motion estimates between frames.

2.4. Clutter Removal

It is sometimes difficult to capture a clear image without things getting in the way. Rather than having to wait for an unobstructed view, we apply our multi-image processing system to clear up the scene by distinguishing the obstructor from the obstructed using multiple captures. An earlier paper on this subject showed that it is possible to remove clutter using a minimum of two photos [15].

The key to successfully performing this task is to make image registration step robust even in the presence of potentially large occlusion (e.g. cars, people). Once the images are registered, we use majority voting to identify which pixels belong to the stationary background scene and use them to reconstruct the final image.

3. Experimental Results

3.1. Dynamic Range Enhancement

This technology has demonstrated robust performance on images captured under a wide variety of lighting conditions using different cameras. Figure 3 shows the result on the challenging example mentioned earlier. It overcomes image sensor limitations in dynamic range that photographers frequently encounter when shooting scenes that contain dramatic differences between highlights and shadows. It produces a *well-exposed* image from pictures of the same scene taken with different exposure values.

3.2 Low Light Exposure Enhancement

In Figure 4, the photograph in the middle was captured with a handheld camera inside a museum where no flash was permitted. Because of the dim available lighting, the exposure time was automatically set to 1/6 second, which was longer than what the camera could be held steadily for, and resulted in a blurry image. As an alternative, we took six under-exposed photos (four at 1/60 second, one at 1/30 second, and one at 1/15 second), one of which is displayed at the top of Figure 4. Using our multi-image processing system, we could align these photos, and combine the corresponding pixel information from the motion-compensated images to render a sharp, well exposed image, as shown in the bottom of Figure 4. Note that during the image synthesis process,

the combined pixel values in certain areas may exceed the range of the display device (typically 255 per channel). Instead of simply *clipping* them at the maximum value, we conduct local contrast adjustment to preserve details.

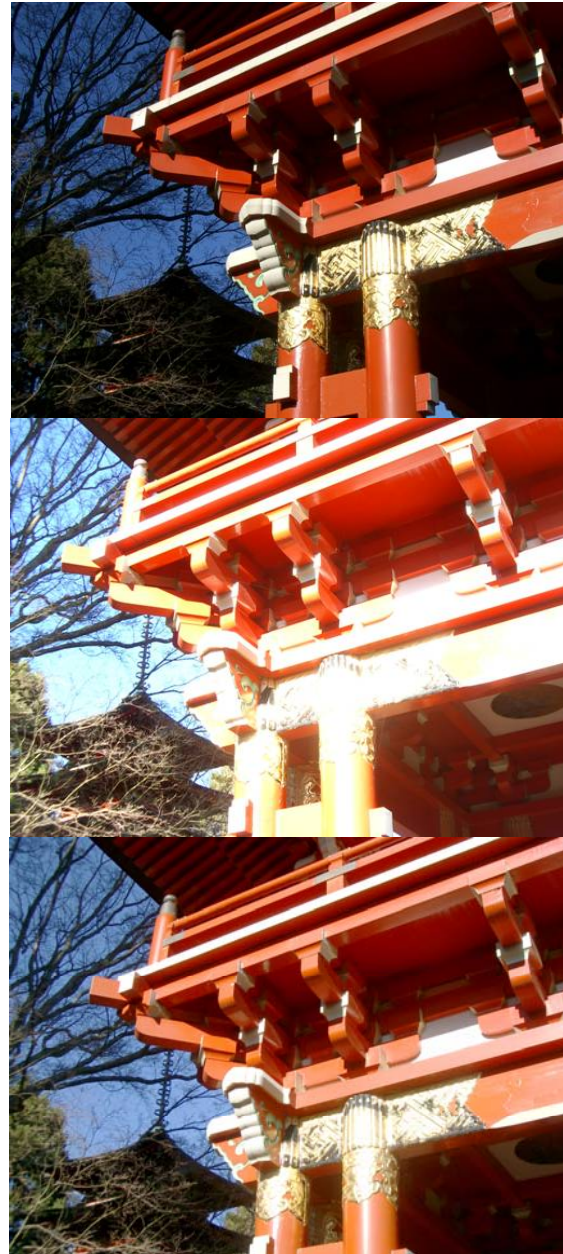


Figure 3. From top to bottom: Short exposure image, Long exposure image and resulting composite image



Figure 4. From top to bottom: Underexposed image, blurry image from long exposure time and composite image after accumulation along motion trajectory

3.3. Clutter Removal

We show the result of clutter removal in Figure 5. The top four images are shots of the same store with people walking by, taken by a handheld camera. The bottom image is generated by registering the four photos and removing the passers-by from the images, producing a clear image of the store.



Figure 5: Results of using multiple images to remove clutter: the top 4 images are captured sequentially, where people are walking in the front of the store. The bottom image is processed from the top 4 images to remove clutter.

4. Summary

We present a multi-image processing system that can automatically register a sequence of images, extract appropriate information from them, and synthesize a composite image with quality significantly exceeding that of each individual image. We applied our system to applications such as extending the dynamic range, faithfully capturing low light scenes without a flash, as well as removing moving objects in a scene. Experimental results demonstrate significant improvement in image quality.

References

- [1] Yadid-Pecht, O. Widening the dynamic range of pictures using a minimum spanning tree algorithm, MVA(12), No. 1, 2000, pp. 1-6.
- [2] Aggarwal, M.[Manoj], Ahuja, N.[Narendra], Split Aperture Imaging for High Dynamic Range, IJCV(58), No. 1, June 2004, pp. 7-17.
- [3] Aggarwal, M.[Manoj], Ahuja, N.[Narendra], High Dynamic Range Panoramic Imaging, ICCV01(I: 2-9).
- [4] Hua, H., Ahuja, N., A High-Resolution Panoramic Camera, CVPR01(I:960-967).
- [5] Won-ho, C.[Cho], Ki-Sang, H.[Hong], Extending dynamic range of two color images under different exposures, ICPR04(IV: 853-856).
- [6] Brajovic, V., Brightness Perception, Dynamic Range and Noise: A Unifying Model for Adaptive Image Sensors, CVPR04(II: 189-196).
- [7] Pal, C., Szeliski, R., Uyttendaele, M., Jovic, N., Probability models for high dynamic range imaging, CVPR04(II: 173-180).
- [8] Nayar, S.K., Branzoi, V., Adaptive Dynamic Range Imaging: Optical Control of Pixel Exposures over Space and Time, ICCV03(1168-1175).
- [9] Nayar, S.K.[Shree K.], Mitsunaga, T.[Tomoo], High Dynamic Range Imaging: Spatially Varying Pixel Exposures, CVPR00(I: 472-479).
- [10] Kang, S.B., Uyttendaele, M., Winder, S., and Szeliski, R., ACM Transactions on Graphics, vol.22(3), pp.319-325. 2003.
- [11] Robertson, M.A.[Mark A.], Borman, S.[Sean], Stevenson, R.L.[Robert L.], Dynamic range improvement through multiple exposures, ICIP99(III:159-163).
- [12] Debevec, P.E., Malik, J., Recovering High Dynamic Range Radiance Maps from Photographs, SIGGraph-97(369-378). 1997
- [13] Madden, B.C., Extended Intensity Range Image, UPennMS-CIS-93-96, 1996.
- [14] Chen, M., "Dynamic content adaptive super-resolution," Int. Conf. Image Analysis & Recognition, Sep. 2000.
- [15] Herley, C., "Automatic Occlusion Removal from Minimum Number of Images", Image Processing, 2005. ICIP 2005. IEEE International Conference on Image Processing, Volume 2, 11-14 Sept. 2005
Page(s):1046 – 1049

Author Biography

Mei Chen is a senior scientist at Intel Research and an adjunct faculty at the School of Computer Science at Carnegie Mellon. She is the principal investigator on the medical decision support project and leads research on biomedical imaging and machine learning. Prior to joining Intel, she was a research scientist at Hewlett-Packard Laboratories. Focusing on computational photography, she developed technologies that have been transferred to HP printers and all-in-one units. Before her career at HP Labs, Dr. Chen was a Member of Technical Staff at Sarnoff Corporation, where she designed innovative video rendering techniques. She holds a Ph.D. and a M.S. from the School of Computer Science in Carnegie Mellon, and a M.S. and B.S. from Tsinghua University in China.

Suk Hwan Lim (S'98-M'03) received the B.S. degree in electrical engineering from Seoul National University, Seoul, Korea, in 1996, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1998 and 2003, respectively. He is currently a Senior Member of Technical Staff at Hewlett-Packard Laboratories, Palo Alto, CA. At Stanford University, he worked on the Programmable Digital Camera (PDC) project and his research focused on the high-speed CMOS image sensor and its video processing applications. He co-designed a 10,000 frames/s CMOS image sensor and developed video processing applications that exploit the high-speed imaging capability. At HP Labs, his research topics include imaging pipeline algorithms for cameras and camera-phones, computational photography for low light imaging, image/video analysis, motion estimation and video enhancement for motion pictures. His research interests include image sensors, computational photography, optical flow estimation, and image/video analysis/enhancement.IE.

Qian Lin received her B.S. degree in Control and Automation from Xi'an Jiaotong University, her M.S.E.E. degree from Purdue University, and her Ph.D. degree from Stanford University. Her PhD thesis is in the area of image processing for remote sensing. She has been with Hewlett-Packard Company since 1992, doing research in the areas of image and video processing and rendering. She is currently manager of the Imaging Technology Department at HP Labs, leading a team of researchers in developing advanced imaging technologies from capture to management and display.