Entropy-based image merging

A. German, M. R. Jenkin, Y. Lespérance
Department of Computer Science and Engineering and Centre for Vision Research
York University, Toronto, Ontario, Canada.
{german,jenkin,lesperan}@cs.yorku.ca

Abstract

Spacecraft docking using vision is a challenging task. Not least among the problems encountered is the need to visually localize the docking target. Here we consider the task of adapting the local illumination to assist in this docking. An online approach is developed that combines images obtained under different exposure and lighting conditions into a single image upon which docking decisions can be made. This method is designed to be used within an intelligent controller that automatically adjusts lighting and image acquisition in order to obtain the "best" possible composite view of the target for further image processing.
Keywords: Image Entropy, High Dynamic Range.

1 Introduction

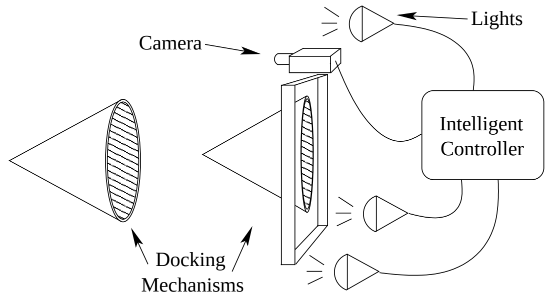
Perhaps the most interesting vision tasks involve guiding semi-autonomous vehicles such as unmanned underwater vehicles, mining machines and spacecraft. Given the widely varying and often poor lighting conditions encountered in such tasks, the remote video camera is often associated with one or more (typically fixed) but controllable light sources. The camera itself often has a variety of controllable parameters such as shutter speed and aperture. Given the controllable intrinsic camera parameters, and the controllable light sources, the remote operator manipulates the various camera parameters and lighting options in order to be able to carry out the required task. This task may be performed directly by a human operator or it may be performed by a software agent with or without human intervention. In either case, the operator manipulates the camera parameters and the available lighting in order to ensure that those portions of the image that are critical to the task at hand are illuminated appropriately (see Figure 1a).

Choosing an appropriate illumination for a human operator is an extremely complex problem. Maximizing one illuminant may place portions of the scene in high relief, while at the same time casting shadows over other portions of the image. Interactions between the illuminants and gain control within the camera itself complicates the task even further. Perhaps the most common version of this problem is the lighting problem portrait photographers encounter: How should the various illuminates be lit and the camera controlled in order for the camera to best capture the subject? Note that what "best" *is* depends significantly on the specific task on hand.

In the machine vision domain, the task becomes even more complex. Cameras typically have a limited dynamic range so they often cannot be used to effectively image the whole scene in one acquisition. Unlike natural settings, one simplifying assumption that is often made is that the only active agent in a teleoperated setting is the teleoperated agent. Assuming that the scene is static, then it is possible to illuminate different parts of the image under different illuminates and camera capture parameters, and then to combine different parts of the image captured under different conditions into a single composite image.

To consider this illumination problem in its simplest form, consider a spacecraft equipped with a camera-light arrangement like that given in Figure 1b. If one assumes that the underlying camera capture and scene geometry is static, i.e., the spacecraft are not moving relative to each other and the position of the camera, the lights and object being viewed remain unchanged, then the camera's intrinsic parameters and the level of illumination provided by each light can be manipulated. Furthermore, if the aperture, focus and focal length of the camera remain unchanged, then over a set of images taken under different lighting and camera parameters, a given pixel (u, v) in the camera will always image the same scene point and image blur will remain constant. Under these conditions, the process of combining multiple images into a single image can be expressed at the pixel level – how should a specific pixel values at (u, v), taken under different illumination and camera

(a) A computer graphics rendering of the space shuttle docking procedure.



(b) An intelligent controller can manipulate lighting intensities and camera intrinsics in order to derive an accurate model of the relationships between the spacecrafts involved in a docking procedure.

Figure 1: Illumination issues in teleoperation. How can the scene be best illuminated and captured in order to dock the two vehicles?

parameters, be combined to obtain a composite pixel value at (u, v)?

1.1 Formal Statement of the Problem

Given a set of images $\{I_1, \ldots, I_N\}$, a function ϕ is desired that combines the set into a single image $\tilde{I}_{1\ldots N}$. Notationally, we seek a function $\phi()$ that operates at the pixel level and that has the following properties:

$$\begin{aligned} \tilde{I}_1 &= \phi(I_1) \\ \tilde{I}_{1\ldots N} &= \phi(I_1, I_2, \ldots, I_N) \end{aligned}$$

In order for the image merging to operate in an efficient, online manner ϕ should have the property that

$$\tilde{I}_{1\ldots N+1} = \phi(\tilde{I}_{1\ldots N}, I_{N+1})$$

That is, it should be efficient to compute the $N+1^{th}$ image of $\tilde{I}$ given the computation for the N^{th} image.

2 Related Work

The problem of combining multiple images taken under varying sensor/lighting conditions has received considerable attention in the literature, although not in the limited scope of the algorithm being considered here. High dynamic range images have many properties in common with the task being considered here (see [1] for an introduction to the problem of high dynamic range images). A commonly considered problem in high dynamic range images is the task of rendering the wide range of data available at a given pixel (u, v) given the limited display range of the intended displays: That is, given the set I how to compute an image $\tilde{I}$ that best represents the input images. In the high dynamic range image case, the various images I are typically captured before the image processing takes place and an offline version of the algorithm is appropriate – the display is not updated as new bands of image information are obtained. Several approaches to this 'rendering of high dynamic range images' problem have been devised and implemented in both hardware and software. Cameras like the QinetiQ High Dynamic Range Logarithmic CMOS Cameras[1] compress the dynamic range of the image using on-board logarithmic intensity compression. The system described in [1] uses several images under different exposures to recover the camera's response function and from that is able to fuse the images into a single, high dynamic range radiance map. In [2] a contrast compression algorithm using a coarse-to-fine hierarchy is described. In [3] a system is developed that performs gradient attenuation to reduce the dynamic range in the image. The algorithm described in this work is based in part on the approach of Goshtasby [4]. The basic approach of [4] is to combine images in a manner that maximizes the entropy of the resulting combined image, while using a smoothing function to ensure that the resulting image does not exhibit intensity discontinuities that were not present in the input images.

3 Basic Approach

The basic approach developed for the online combination of images builds upon Goshtasby's entropy based high dynamic range reduction algorithm [4],

[1] QinetiQ LogCMOS Camera - http://www.qinetiq.com

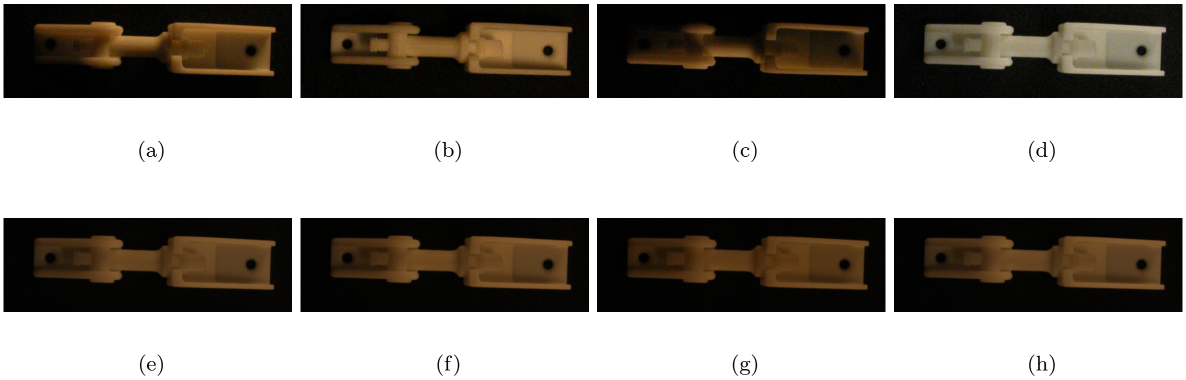


Figure 2: Top row: (a) Illumination by illuminant 1 only at 100% (b) Illumination by illuminant 2 only at 100% (c) Illumination by illuminant 3 only at 100% (d) Illumination by illuminant 1,2 and 3 all at 100%. Bottom row: (e)-(g) The composite image at various stages during the addition of the 512 source images. (h) The final composite image after all 512 images have been added.

but differs in how the images are combined. In this system, the images are merged on a pixel per pixel basis by weighting the local pixel values by their local entropy estimate.

Entropy was chosen as a measure of the detail provided by each picture. The *entropy* (see [5]) is defined as the average number of binary symbols necessary to code a given input given the probability of that input appearing an a stream. High entropy is associated with a high variance in the pixel values, while low entropy indicates that the pixel values are fairly uniform, and hence little detail can be derived from them. Therefore, when applied to groups of pixels within the source images, entropy provides a way to compare regions from the different source images and decide which provides the most detail.

The method developed for this task, though simple, is both flexible and powerful. Every pixel in the final image is computed as the weighted average of the corresponding pixels in the source images where each value is weighted by the entropy of the surrounding region. For each pixel $p = (u, v)$ in the final image there are corresponding pixels $p_1, p_2, \ldots, p_N$, one for each source image. For each pixel p_i in each image, the local entropy (measured within a fixed window) v_i is computed, and the weighted average p is computed as

$$p = \frac{\sum_{i=1}^{N} p_i v_i}{\sum_{i=1}^{N} v_i}$$

The entropy for the pixel window is computed as

$$v_i = \sum_k -q_k \times \log_2(q_k)$$

where q_i is the probability that a random pixel chosen from the window centered on p_k will have intensity i.

3.1 Online Computation

When docking a spacecraft feedback is necessary in order to adjust the myriad of parameters required for the task. The system must function before all images are available. It may even be desirable to use a partial computation of I to aid in the choice of future camera capture parameters.

For every pixel p in the final image, two running sums are maintained for each colour component: The first is G_p^r the sum of the intensity of channel r at location p in the source images, multiplied by the entropy of the surrounding pixel window in each source image. The second sum I_p^r is the sum of the entropy in each source image channel.

$$\begin{aligned} G_p^r &= \textstyle\sum_{i=1}^{N} p_i^r \times v_i^r \\ I_p^r &= \textstyle\sum_{i=1}^{N} v_i^r \end{aligned}$$

$$\begin{aligned} G_p^g &= \textstyle\sum_{i=1}^{N} p_i^g \times v_i^g \\ I_p^g &= \textstyle\sum_{i=1}^{N} v_i^g \end{aligned}$$

$$\begin{aligned} G_p^b &= \textstyle\sum_{i=1}^{N} p_i^b \times v_i^b \\ I_p^b &= \textstyle\sum_{i=1}^{N} v_i^b \end{aligned}$$

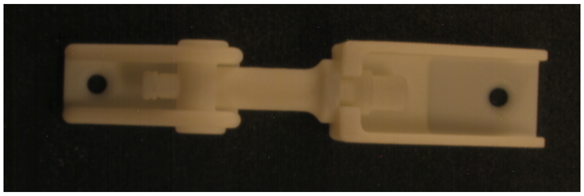


Figure 3: The combined result of all 512 images after gamma correction

The final image pixel p can be computed from G_p^r, G_p^g, G_p^b and I_p^r, I_p^g, I_p^b as

$$
\begin{aligned}
p^r &= \frac{G_p^r}{I_p^r} \\
p^g &= \frac{G_p^g}{I_p^g} \\
p^b &= \frac{G_p^b}{I_p^b}
\end{aligned}
$$

Storing the final image as a set of sums provides the system with the capability to add and remove images from the collective quickly and easily, while at the same time ensuring that the final image can also be generated quickly.

The agent developed to control the addition and removal of images as well as the intensities of the related light sources is described in [6].

4 Results

The images in Figure 2 are take from a set of 512 images of the same orbital object (a latch from the Hubble Space Telescope) taken under different lighting conditions. The 512 images were captured under 512 different illumination conditions (three light sources each with 8 possible light values). The second row of the image shows four different merged images. These were obtained by merging different subsets of the images together. Image (h) is the result of the merger of all 512 images. Figure 3 is an expanded version of Figure 2 (h) after gamma correction. The fine detail of the latch is quite visible.

The process of merging the images together on a pixel-by-pixel basis permits the complex set of 512 images to be rendered as a single image that can be viewed by either a human operator or used as the input to later computational stages. The merging process weights the images by the local entropy. Figure 4 shows the relative entropy of each of the 512 source images as well as the entropy of the combined image made from images 0..N (where N is in the range 0 to 511). The wide variability in the Entropy of individual images is evident in Figure 4 as is the stability of the merged image after a small number of images have been combined.

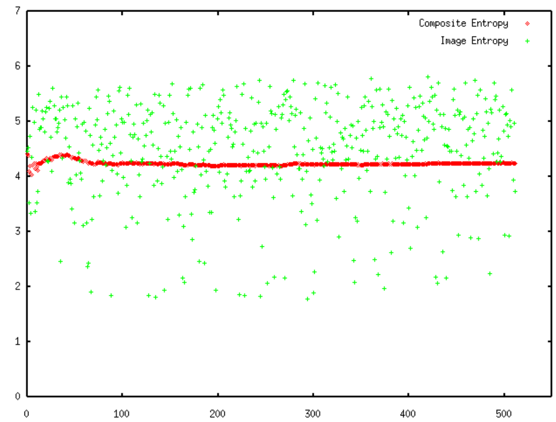


Figure 4: A graph of the evolving composite entropy and the entropy of the 512 source images

A second example is given in Figure 5. The source images are of a satellite docking module and are taken at increasing luminosities. The composite images in Figure 5 are computed with window sizes 5, 11, 21, and 41 respectively. Clearly, since entropy can be sampled over a larger area, larger windows yield a smoother image.

Also of interest is the effect of both random noise and blank images on the composite. Figure 6 provides such a comparison. While blank frames (of any colour) have no effect on the composite due to their low entropy, as expected images with random noise have a detrimental effect on the composite. However, we are assuming all images in the input set will be highly correlated and should therefore not exhibit such random noise.

5 Discussion

In this paper an on-line method for combining multiple images taken under different lighting conditions was presented. The method involves weighting each pixel by the surrounding entropy such that each element in the final image is in high relief but lacks abrupt contrast changes due to the different light sources that might introduce spurious lines and other artifacts. The method also allows for fast addition and removal of images from the collective.

The resulting merged image can be used by either a human operator or by a software agent. Of particular interest is the development of a software system that automatically adjusts the camera and illumination parameters in order to obtain the 'best' combined image automatically using a small number of different illumination settings. The development of such a system is the subject of ongoing research (see [6]).

Acknowledgments: We would like to acknowledge the support of Mark Obsniuk, Andrew Hogue and Olena Borzenko. The financial support of CITO, MDRobotics and NSERC is greatly appreciated.

References

[1] Debevec, P. E. and Malik, J., "Recovering High Dynamic Range Radience Maps from Photographs," *Proceedings of SIGGRAPH 1997*, ACM Press / ACM SIGGRAPH, 369-378, 1997.

[2] Tumblin, J. and Turk, G., "LCIS: A Boundary Hierarchy for Detail-Preserving Contrast Reduction," *Proceedings of SIGGRAPH*, ACM Press / ACM SIGGRAPH, 83-90, 1999.

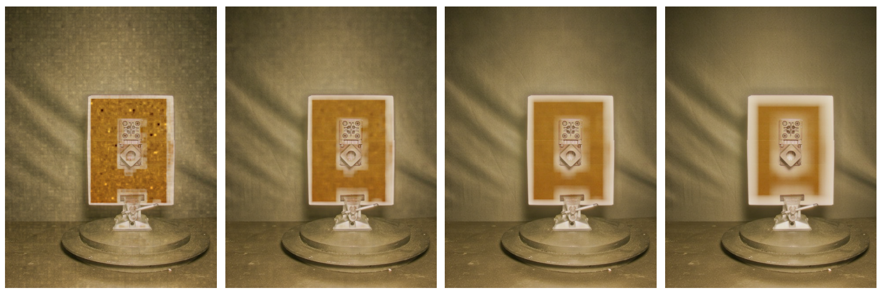
[3] Fattal, R., Lischinski, D., and Werman, M., "Gradient Domain High Dynamic Range Compression," *Proceedings of SIGGRAPH*, ACM Press / ACM SIGRAPH, 249-256, 2002.

[4] Goshtasby, A. A., "High Dynamic Range Reduction Via Maximization of Image Information.", http://www.cs.wright.edu/ agoshtas/hdr.html

[5] Shannon, C. E., "A Mathematical Theory of Communication," *Bell System Technical Journal*, Vol. 27, 379-423, 623-656, 1948.

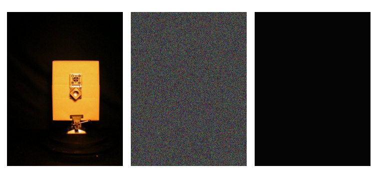
[6] Borzenko, O., Lesperance, Y., and Jenkin, M. R., "Controlling Camera and Lights for Intelligent Image Acquisition and Merging.", IEEE Second Canadian Conference on Computer and Robot Vision (CRV 2005), 2005.

(a) (b) (c) (d) (e) (f) (g)

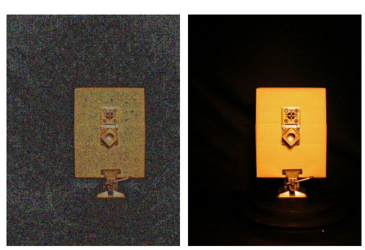


(h) (i) (j) (k)

Figure 5: Top row: (a)-(g) Images taken as luminosity increases. Bottom row: (h)-(k) Composites of (a)-(g) images with window sizes 5, 11, 21, 41 respectively



(a) (b) (c)



(d) (e)

Figure 6: Shows the effect of random noise as well as blank images on the composite. Top row: The source images (a) an image taken of the object, (b) an image of random noise, (c) a blank image. Bottom Row: (d) A composite comprised of (a) and (b), (e) a composite comprised of (a) and (c)