Extending Dynamic Range of Monochrome and Color Images through Fusion

Luca Bogoni
Vision Technology Laboratory
Sarnoff Corporation
201 Washington Rd., Princeton, NJ 08540, USA
lbogoni@sarnoff.com

Abstract

This paper presents an approach to enhance the dynamic range of monochromatic and color images. The goal of this method is that of overcoming the inherent physical limitations of the acquisition process. The approach uses pattern-selective image fusion as a means to integrate information from images acquired under different shutter speeds or apertures. The result of this process is a composite in which none of the portions of the image is over or under saturated and the underlying information has been preserved. The results show the performance of this enhancement process from imagery taken from consumer-grade video camera equipment.

1. Introduction

Known methods for adjusting a sensor's integration time and aperture, including methods such as automatic gain control and automatic iris selection, are commonly used to adjust for the overall illumination in a scene, but cannot compensate for large local variations in scene brightness. As a result, very bright scene regions may become saturated, and dark regions can lose most or all of their detail.

The proposed method extends the dynamic range of a sensor through the fusion of multiple images of the same scene. The consideration underlying is that each of the participating image will have, at least some portion with high-dynamic resolution. Hence, the goal of this process is that of generating a composite in which each portion of the image will have an enhanced dynamic range.

To accomplish this task, a pattern-selective image fusion process has been employed. Prior work on the image fusion process has focused on operating on multiple intensity images, either as sets of monochrome images or sets of intensity images from different sensors such as visible and infrared sensors [5, 7] and using other basis functions [6].

Recently, [9], images obtained at different shutter speeds were combined into an image in which the full dynamic range contributed by each image is preserved. In this process, the underlying radiometric response function is recovered and images are fused into a single image spanning the full extent of the dynamic range. The method presented is very effective and through a special viewing window it is possible to locally view the composite. Global histogram equalization can be employed to bring the image into a normal eight bit viewer but the rescaling leads to loss of information since the full range is equally weighted. Local histogram equalization is very sensitive to the size of the window employed and may lead to either emphasize too many details, when the windows are too small, or introduce ghosts.

Our approach extends and generalizes the standard fusion approach so that both monochrome and color images may be fused. In the context of the color domain, the application of fusion has been thought as a direct extension of the original technique whereby the operation applied to one image frequency band could be directly replicated to the other ones [1]. However, it is necessary to treat each pixel in a color image as a vector in the respective color spaces, since the values in the separate planes are correlated, to be able to generate images that preserve the original information while taking advantage of the multi-resolution framework.

In the application here described, the pattern-selective fusion process allows to extract from the participating images those areas that satisfy desired properties. The desired composite image is one that incorporates portions from the various images which are neither over- or under- saturated and best reveal the details. The key result is that the composite image has brought into the displaying range aspects from the images which would otherwise be beyond the capability of any single image.

The remainder of this paper is organized as follows. The methodology of pattern selective fusion is reviewed in Section 2. Section 3 describes the fusion process as applied to monochrome images and Section 4 discusses the effec-

0-7695-0750-6/00 $10.00 © 2000 IEEE

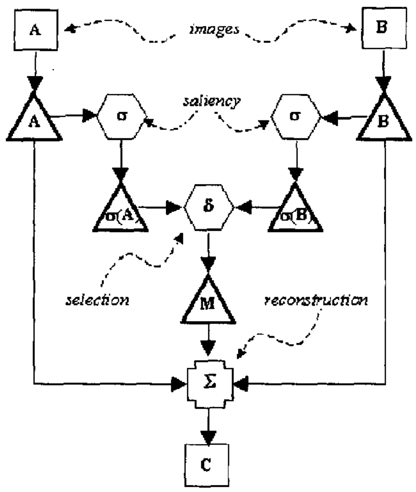

Figure 1. Block diagram of the pattern selective image fusion process.

tiveness of fusion as applied to the dynamic range domain. Section 5 describes the operation as extended to color images. The paper concludes, Section 6, with a review of the presented results.

2. Review of Pattern-Selective Image Fusion

The foundation for combining multiple images into a single, enhanced result is the pattern-selective fusion process itself, see [3] for more details. This section provides an overview of the generalized fusion process as it can be applied to color imagery. To simplify this discussion, we assume the fusion process is to generate a composite image C from a pair of source images denoted with A and B. The process used to fuse the source images is shown schematically in Figure 1, and is summarized below.

- *Pyramid Construction* Pyramids are formed for the input images A and B,see [4]. The highest level of resolution (level 0) refers to the image at the original resolution, while the lowest level of resolution (level N) refers to an image whose resolution has been reduced by the pyramid formation process by a factor of 2^N. The fusion methods described within this paper use a Laplacian pyramid representation.
- *Feature saliency* The feature saliency computation process, labeled σ, expresses a family of functions that operate on the pyramids of both images yielding saliency pyramids. The saliency function captures the importance of what is to be fused.
- *Selection Process* The selection process, labeled δ, expresses a family of functions that operate on the saliency pyramids, $\sigma(A)$ and $\sigma(B)$, obtained from the saliency computation process. The result of this process is a pyramid of coefficients, M, which defines contributions between the pixels in the two images to yield the result.
- *Reconstruction* The fused image result C is reconstructed from the pyramids of the original images A and B subject to the coefficients generated by the selection process. The reconstruction process, labeled Σ, operates on each level of the pyramid of the original images in conjunction with the pyramid-coefficient mask to generate the composite image, C. The highest level in the pyramid, N, may be blended using a function β. One possible function for β could be the average of the two Gaussians.

2.1. Registering Images with different Dynamic Range

It should be noted that this process is very sensitive to errors in the registration of the source images. Errors in registration will result in the expected "doubling" of features that are out of alignment. For the results shown in this paper, the source images are registered to a selected reference source image prior to the fusion process.

The image registration process can be described as a two step process and can be summarized as follows:

- *Global Parametric Registration* First, each of the source images are registered into a common coordinate system using affine global motion estimation as described in [2, 8]. This global registration process eliminates small global displacements in the source images that can occur from instabilities in the sensor that can occur when a sensor is hand-held or settings are manually changed.
- *Local Unconstrained Motion Estimation* Second, local unconstrained motion estimation is performed to register local scene structure in the source images that may be in motion. This is accomplished through estimating an optical flow field between each source image and a common reference image, and using that flow field to warp the source images such that all scene features are in accurate registration. Once the source images are registered, the pattern-selective color image fusion process is applied to the set of registered source images to determine the best representation for each scene point, and the final output image is generated.

Figure 2. Subsequence of the images used in generating the composite. They were acquired using a b&w Sony Camera with fixed 25 deg FOV lens. The exposure time were 10 msec, top left, 2 msec, bottom right, 0.2 msec, bottom left, and 0.125 msec.

3. Extending Dynamic Range for Multiple Monochrome Images

This section presents the fusion process as applied to multiple monochrome images. The images, see Figure 2, show a subsequence of the images participating in the fusion process.

The process for fusing multiple images to extend dynamic range is described as follows:

- The Laplacian pyramid is constructed for the images.
- Select the contributing Laplacian image amongst K images for each level λ, $\lambda < N$, based on maximum strength of the edge $Max()$. Let Ind denote the index function, then the selection functions for each position i, j of the K Laplacians and level λ is specified by

  $$M_\lambda(i,j) = Ind(Max(L_k(i,j))) \;\; k = 1 \ldots K$$

- The goal of the basic function, β, that of balancing the areas of saturation of the images. At any position at the highest level of the pyramid, N the value at each position in the image for resulting Gaussians is determined by the product between the Gaussian value at i, j weighted by the magnitude of the gradient of the Gaussian images. Let

  $$W_I(i,j) = \frac{\mid \nabla G_I(i,j) \mid}{\sum_{k=1}^{K} \mid \nabla G_k(i,j) \mid}$$

  denote the weight contribution based on the magnitude of the gradient of the Gaussian image at position i, j across the different images in the sequence for image I, then

  $$G_N(i,j) = \sum_{k=1}^{K} W_k(i,j) * G_k(i,j)$$

  Defines the resulting value for the composite of the Gaussian.

- The reconstruction process begins with the Gaussian composite, then at each position in each level, the value determined by the mask determines the contribution of the images.

Figure 3. Composite image resulting from a sequence of 9 images. Figure 2 shows a selection of the images.

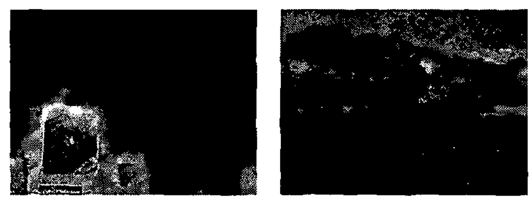

Figure 4. Selection images showing the contribution for two of the images in the sequence.

The selection of the maximum value for Laplacian for any given points effectively generates a mask in which the strongest edges have been identified. The edge is strongest where the dynamic range is neither under- or over-saturated.

The choice of the β function was determined by having all the images contributing in proportion to the amount of information contained in the Gaussian image. An ordered sequence of images provides most contribution whenever it neither under or over-saturated but present most change in value with respect to the other ordered members in the sequence. When only two images are present then an average of the values for the Gaussian are performed.

Figure 5. Test image generated to test the fusion process.

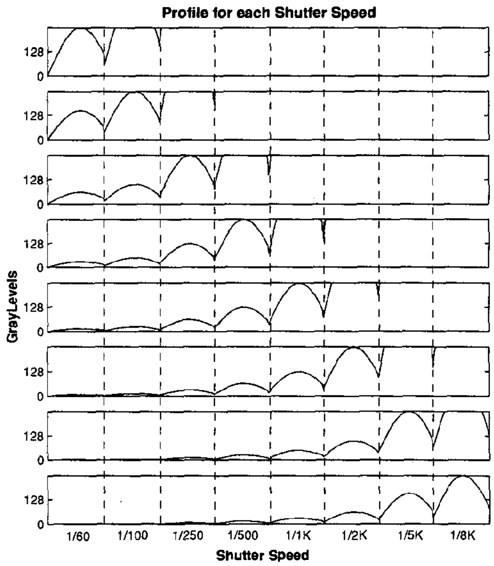

Figure 6. Row profiles for the various images for each of the shutter speeds.

Figure 3 shows the result of the fusion process. We can see that both the content of the box as well clouds present in the sky are visible.

Figure 4 shows the contribution for two images in the sequence is providing to the final composite. These were obtained by subtracting the fused image obtained from the complete sequence and the fused image with one image in the sequence missing. Clearly we can see that the image on the right, for instance, having shutter speed of 0.2 msec provides details information for the portion of the sky which is otherwise not available.

4. Effectiveness of Fusion

In order to evaluate the effectiveness of the fusion process, several test images have been generated. In this section we show one such images and demonstrate how the fusion process is applied this domain.

The test image is assumed to represent the intensity of a scene, see Figure 5. The image can be thought to represent a set of cylinders painted in decreasing values of grayness, left to right. The values were generated so that for any of the shutter-speeds selected will obtain the most dynamic range extent when integrated over the individual period. The test image is horizontally divided into shutter speed bands (8

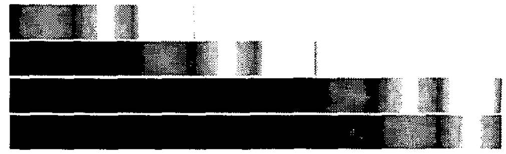

Figure 7. Selection of the images obtained for fusion from the original for shutter speeds of 10 msec (2nd cylinder), 2 msec (4rth cyl.), 0.2 msec (7th cyl.), and 0.125 msec. (8th cyl.)

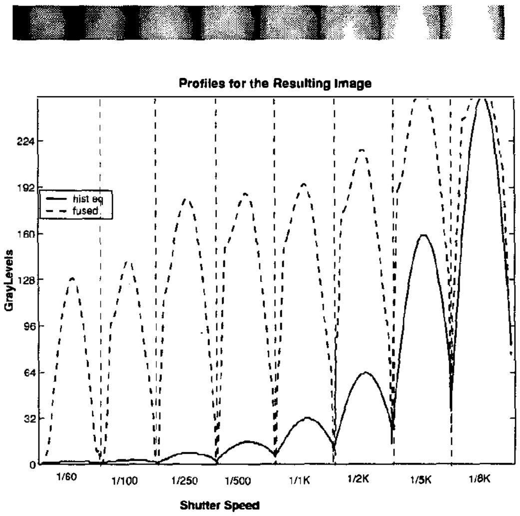

Figure 8. Top: fused image; bottom: profiles for the fused image (dash-lines) and the histogram equalized image.

in all). The values of the shutter speed were selected to reflect the same setting used in the previous example, Figure 2. Each band will integrated to the full dynamic range when the shutter speed is set for the particular value. For instance, a shutter speed of 1/250 of a second will generate an image which will exhibit the full dynamic range for the chosen band. If it were possible to have a perfect reconstruction of the scene, the intensity values would range [0.0-33,999.39]. When, this image is histogram equalized to fit in the displaying range [0-255], see Figure 5, the resulting image exhibits the over-all structure of the dynamic range of the scene, yet the details of the individual shutter-speed bands have been lost. While performing local histogram equalization may provide a better profile for this particular image, in practice, when multiple frequecies are present, noticeable side-effects, such as ghosting and noise augmentation occur.

As discussed in the introduction and in the previous section, the effect of the fusion process, is that of emphasizing the individual bands. This tends to equalize across the vari-

Figure 9. Selection images showing the contribution from each of the images in the sequence.

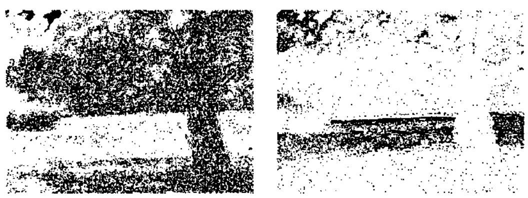

Figure 10. Selection images showing the contribution from two images in the sequence. These correspond to the top-left and bottom-right in Figure 9.

ous integration times rather than equalize across the over-all time-span. Figure 6 shows the cross section of the images when the shutter speed is optimized for the particular band. Figure 7 shows the actual intensity images generated. Notice that portions of the image which are to the left of the selected shutter speed band are under-exposed while those to the right are over-saturated.

Figure 8 shows the fused image and the profile for the fused image and the histogram equalized image. We can notice that the histogram equalized image clearly preserves the overall structure for the dynamic range of the scene; however, due to the quantization process, the details for slower shutter speeds have been mostly lost to benefit those in the highest dynamic range.

5. Extending Dynamic Range for Multiple Color Images

Extending dynamic range for color images requires a choice of an appropriate color space and the definition of the functions used in the fusion process. We have chosen to

Figure 11. Resulting fused image from sequence shown in Figure 9.

develop this algorithm using HSV space. The V component captures the luminance, S describes the color saturation and H the hue. The sequences of images were obtained with short time separation between them, during this time it is assumed that the position of the luminant, its intensity and scene geometry do not change during the acquisition process. Hence by varying the shutter speed of the camera, the acquired images will differ primarily in saturation.

The functions used in the fusion process are defined as in the case of the monochromatic images. The difference being that instead of luminance, the V component is used and that the selection for S and H components are performed as vectors in three-space. See [3] for a detailed description on fusion of color images.

Once the image has been reconstructed, the S-component is rescaled locally to account for the change in color-saturation. The reconstruction process tends to attenuate the color-saturation by selecting components from images which might not have the highest color saturation and having medium intensity saturation. Thus, in order to compensate for this attenuation the color-saturation is rescaled to acquire the maximum of color saturation already present in the various images in the sequence.

Figure 9 shows four of the images used in sequence. Notice that in the top left image details of the trunk of the tree are visible, while in the bottom right image details of the branches are visible. Figure 10 emphasizes the selection of the respective components from Figure 9. The selection masks actual refer to the zeroth in the reconstruction pyramid. Figure 11 shows the resulting composite.

Figure 12 shows another example, image sequence courtesy of [9]. It is clearly visible that the composite has brought into the visible range portions from different images.

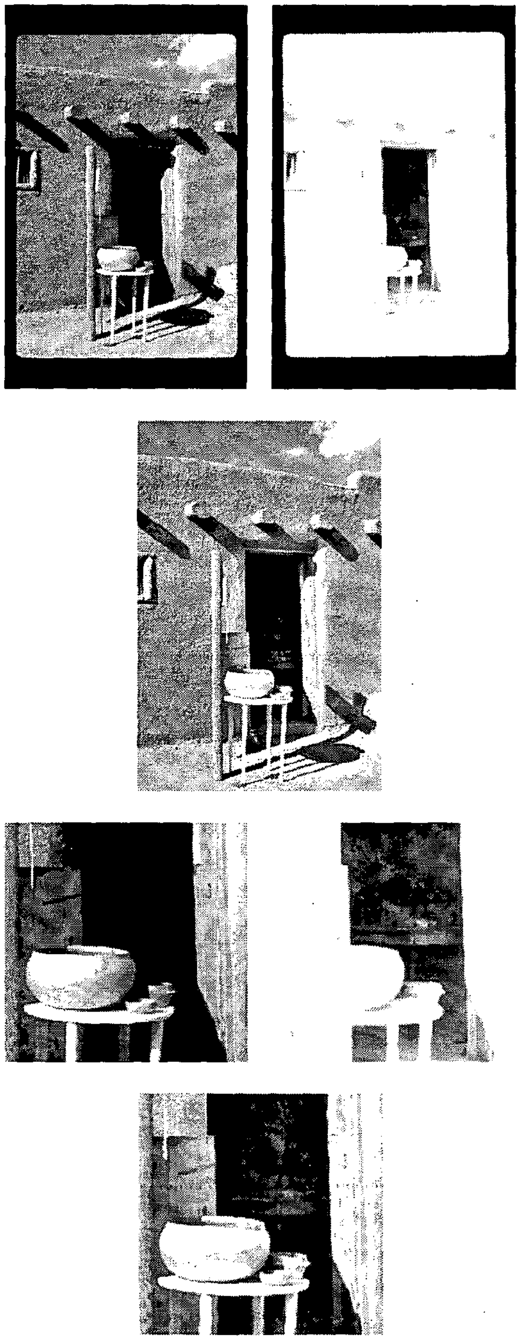

Figure 12. Top: two images from a sequence of 5 images spanning the dynamic range, used with permission [9]. The left image clearly shows the adobe building and sky while the right image shows the inside; next the resulting composite. The following images show two insets from the top images and inset of the composite image.

6. Conclusions

This paper has presented the extension of dynamic range applied both to monochrome as well as to color image sequences. The registration process for the aligning images of the same scene having different dynamic range was outlined. The fusion process was shown to be an effective method to incorporate into the composite information from portions of images having large variation of the dynamic range. We have seen that while the fusion process alters the dynamic range profile of the image it enhances the over-all image to generate a composite in which the features appearing across the different visual bands are given equal importance. This type of processing is most suited for surveillance and/or visual inspection applications in which the dynamic range of the scene is not known a priori and may vary dramatically.

The images that have been acquired using a consumer-grade video-camera were subject to manual control of the settings in order to change the shutter speeds. Future applications of this algorithm will automatically control the camera settings and adapt to maximize the quality as well as the run-time execution of the fusion process.

References

[1] E. Adelson. *Depth-of-focus imaging process method.* U.S.P No. 4,661,986 , 1987.

[2] J.R. Bergen, P.J. Burt, R. Hingorani and S. Peleg. Computing two motions from three frames. In *Proceedings of the Third ICCV*, pages 27–32, 1990.

[3] L. Bogoni, M. Hansen and P. Burt. Image enhancement using pattern-selective color image fusion. In *Proceedings of the 10th International Conference Image Analysis and Processing*, pages 44–49, 1999.

[4] P. Burt. *The pyramid as structure for efficient computation, Multiresolution Image Processing and Analysis.* Springer Verlag, pages 6–35 1984.

[5] P. Burt. Pattern selective fusion of ir and visible images using pyramid transforms. In *National Symposium on Sensor Fusion*, pages 0–0, 1992.

[6] P. Burt and E. Adelson. The laplacian pyramid as a compact image code. *IEEE Transaction on Communications*, 31(4):583–540, 1983.

[7] P. Burt and R. Kolczynski. Enhanced image capture through fusion. In *International Conference on Computer Vision*, pages 173–182, 1993.

[8] M. Hansen, G. van Der Wal, K. Dana and P. Burt. Real-time scene stabilization and mosaic construction. In *Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision*, pages 54–62, 1994.

[9] T. Mitsunaga and S. Nayar. Radiometric self calibration. In *Proceedings of Computer Vision and Pattern Recognition*, pages 374–380, 1999.