Das, S., Zhang, Y-L, Krebs, W.K. (2000). Color night vision for navigation and surveillance, In J. Sutton and S.C. Kak (Eds), Proceedings of the Fifth Joint Conference on Information Sciences, Atlantic City, NJ, February 28th

Color Night Vision for Navigation and Surveillance

Sanjoy Das,

William K. Krebs

ABSTRACT

Image fusion of registered images of night scenery that are obtained from cameras tuned to different bandwidths will be a significant component of future night vision devices. In this article, we describe a new algorithm for such multispectral image fusion. The algorithm performs gray scale image fusion using a method based on principal components. The approach can be easily used for any number of bandwidths. We have provided examples where the algorithm was used to fuse a intensified low light visible image with another image obtained from a single Forward Looking Infrared (FLIR) camera. The algorithm may be implemented readily in hardware for use in night vision devices, as an important aid to surveillance and navigation in total darkness.

Keywords: image fusion, principal components, night vision.

Das, S., Zhang, Y-L, Krebs, W.K. (2000). Color night vision for navigation and surveillance, In J. Sutton and S.C. Kak (Eds), Proceedings of the Fifth Joint Conference on Information Sciences, Atlantic City, NJ, February 28th

INTRODUCTION

The ability to detect targets and obstructions in total darkness is of great importance for vehicle navigation as well as in surveillance and monitoring applications. In military and civilian applications, night vision devices are frequently used to assist operators' night vision. Night vision images are usually either low light images obtained via a CCD camera that are intensified by means of an image intensifier tube, or infrared images obtained by Forward Looking Infrared (FLIR) cameras. Although both sensors are good night vision devices, they do not capture all the available information about the scenery. For instance, warm objects, such as humans and some man-made objects are not easily detected through low-light-visible sensor; however they can be easily detected through an infrared sensor. On the other hand, infrared cameras cannot sense certain finer details about the scenery (such as leaves or grass in natural scenes), and one has to rely on visible imagery to be able to see them. As a consequence, military vessels are often equipped with two or more cameras operating at different bandwidths. Simultaneous viewing of multiple images is a cumbersome task, and it is of little surprise that there has been a major Department of Defense research effort to develop algorithms to merge the images obtained at various bandwidths and obtain single composite images that contain all the information available at the various bandwidths. This operation, known as image fusion has begun to play a very important role in surveillance, monitoring and navigation. It allows the user to view a single composite image that contains all the information gathered from various sources. This obviates the need for separate displays for each sensor.

With the recent advances in digital image processing technology, it is now possible to implement computer algorithms for all sorts of imaging applications. Implementing compact devices for image fusion of these multispectral images is feasible, and image fusion can now be used for a variety of non-military applications, including in the area of transportation. Traffic surveillance centers can easily be equipped with such devices, as can individual vehicles, making navigation in darkness a much easier task.

Mere superposition of all input images does not produce good fused images. This is because, for *n* sensors, a feature that is visible through only a single input camera and not through the others, gets attenuated by a factor of *n* in the output image. Therefore a more elaborate scheme is necessary. A few algorithms for multispectral image fusion have been proposed. A method that is motivated largely by the vertebrate early visual system was developed recently (1,2). In vertebrate vision, color contrast takes place at the inner layers of the retina, where the light sensed by the cones (the photosensitive cells that pick color information, namely red, blue and green) undergo a process called opponent-processing. Opponent processing helps explain many aspects of human color perception, such as why blue stands out in a yellow background and red in a green one. The image fusion algorithm in (1,2) imitates this process in order to obtain maximum color contrast between infrared and low light visible images.

Another method that was proposed recently simply superimposes imagery at three different bandwidths as the red, green and blue components of the fused image (3,4).

Color enhancement is then obtained by performing a linear translation of the color pixels to maximize the color contrast.

This article proposes a new method of image fusion. The proposed algorithm does gray level image fusion by examining the contrast of each input image and performing a weighted combination. The algorithm carries out operations that are highly localized and therefore the proposed method may be realized physically for real-time applications.

OUTLINE OF THE PROPOSED METHOD

Consider n dimensional data in the form of m-vectors $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, ... $\boldsymbol{x}_n$. In order to transform this data into a single m-vector $\boldsymbol{y}$, an n-vector $\boldsymbol{q}$ is selected. The vector $\boldsymbol{y}$ is then obtained as,

$$\boldsymbol{y}_{\boldsymbol{q}} = [\boldsymbol{x}_1\ \boldsymbol{x}_2\ \ldots\ \boldsymbol{x}_n]\cdot\boldsymbol{q}. \tag{1}$$

The subscript $\boldsymbol{q}$ has been added to the $\boldsymbol{y}$ to explicitly show its dependence on it. Let the variance the single vector $\boldsymbol{y}$ be denoted by $\sigma^2(\boldsymbol{y}_{\boldsymbol{q}})$. The (first) principal component is that n-vector $\boldsymbol{p}$ along whose direction, the variance $\sigma^2(\boldsymbol{y}_{\boldsymbol{p}})$ is maximized, i.e. $\forall\boldsymbol{q}$, $\sigma^2(\boldsymbol{y}_{\boldsymbol{p}}) \geq \sigma^2(\boldsymbol{y}_{\boldsymbol{q}})$. The principal component $\boldsymbol{p}$ can easily be computed as the eigenvector of the correlation matrix $E[(\boldsymbol{x}-E(\boldsymbol{x}))\cdot(\boldsymbol{x}-E(\boldsymbol{x})^{\mathrm{T}}]$ having the largest eigenvalue. Although the principal component can be computed in a straightforward manner, a neural network approach based algorithm would be preferred for a hardware implementation.

The monochrome fusion algorithm breaks up the entire images into smaller circular regions or 'partitions' and performs image fusion in a region-wise manner. Fusion is done based on the assumption that the variance of the image within any localized partition is a measure of the information content of the partition. This assumption is valid when the image is noise free and when the partitions are neither too large not too small. We already mentioned that the input images often carry complementary information, e.g., some features in the scenery are visible through one image while some others through another. Therefore, within too large a partition too many features may get included, and as a result, images would show no great difference in information content. In a similar manner, features may not be seen through partitions that are too small, and consequently the algorithm may interpret this as an absence of descriptive features about the scene in the image. In other words, the size of each partition plays an important role in determining the quality of the fused image, and needs to be selected judiciously. Henceforth, we shall treat an image as a two-dimensional array of pixels, and the pixel in the i^{th} row and the j^{th} column shall be denoted by $\mathbf{I}(i, j)$. Using this notation, we define the $(p, q)^{\text{th}}$ partition of an image $\mathbf{I}$ as follows,

$$\mathbf{R}_{\mathbf{I}}(p, q) = \{\mathbf{I}(i, j) | (\alpha p - i)^2 + (\alpha q - j)^2 < \alpha^2\}, \tag{2}$$

where α, an integer, is the region size. Clearly the indices p and q can acquire values in the range $[1, r/\alpha]$ and $[1, c/\alpha]$ respectively where $r\times c$ is the image size. It may be noted from the above equation that the regions are overlapping and that the partitioning process covers the entire image.

In order to determine the localized average pixel values in the images, the infrared (IR) and visible images, $\mathbf{I}_{IR}$ and $\mathbf{I}_{VIS}$ are smoothed by means of convolution with a Gaussian function. The smoothed images are given by,

$$\mathbf{S}_{IR} = \mathbf{I}_{IR} \otimes exp[-(x^2 + y^2)/\sigma^2] \tag{3a}$$

and

$$\mathbf{S}_{VIS} = \mathbf{I}_{VIS} \otimes exp[-(x^2 + y^2)/\sigma^2], \tag{3b}$$

where $\otimes$ is the convolution operation. The width σ of the Gaussian function is closely related to the region size. The deviations from the smoothed images are computed as,

$$\mathbf{D}_{IR} = \mathbf{I}_{IR} - \mathbf{S}_{IR} \tag{4a}$$

and

$$\mathbf{D}_{VIS} = \mathbf{I}_{VIS} - \mathbf{S}_{VIS}. \tag{4b}$$

Next, the principal component of each partition R_{DIR} and R_{DVIS} of the arrays $\mathbf{D}_{IR}$ and $\mathbf{D}_{VIS}$ are computed. The $(p, q)^{th}$ principal component shall be denoted as $\mathbf{p}(p, q)$. Since the data is bivariate, $\mathbf{p}$ shall have two components, $\mathbf{p}_{IR}$ and $\mathbf{p}_{VIS}$. Within each partition, $\mathbf{p}$ shall be appropriately biased towards the image with a higher information content. Therefore, performing a weighted combination of the IR and visible images using $\mathbf{p}$ would lead to a fused image with a high information content. However, before this operation, an $r \times c$ array of weights $\mathbf{W}$ has to be computed from $\mathbf{p}$. This is done in a straightforward manner by mapping the $r/\alpha \times c/\alpha$ of principal components to an $r \times c$ array of weights $\mathbf{w}'$ and then convolving the resulting array with a Gaussian to ensure that the weight variations transition in a smooth manner to obtain the final weight array $\mathbf{w}$,

$$\mathbf{w}'_{IR}(i, j) = \mathbf{p}_{IR}(i/\alpha, j/\alpha), \tag{5a}$$

$$\mathbf{w}'_{VIS}(i, j) = \mathbf{p}_{VIS}(i/\alpha, j/\alpha), \tag{5b}$$

and

$$\mathbf{w}_{IR} = \mathbf{w}'_{IR} \otimes exp[-(x^2 + y^2)/\sigma^2] \tag{6a}$$

$$\mathbf{w}_{VIS} = \mathbf{w}'_{VIS} \otimes exp[-(x^2 + y^2)/\sigma^2]. \tag{6b}$$

Image fusion is carried out as the final step of the monochrome fusion process to obtain the fused monochrome image $\mathbf{M}$. This image is composed by averaging the smoothed input images and adding the weighted, normalized sum of the deviations.

$$\mathbf{M}(i,j)=?[\mathbf{S}_{IR}(i,j)+\mathbf{S}_{VIS}(i,j)] + [\mathbf{D}_{IR}(i,j)\cdot\mathbf{w}_{IR}(i,j)+\mathbf{D}_{VIS}(i,j)\cdot\mathbf{w}_{VIS}(i,j)]/[\mathbf{w}_{IR}(i,j)+\mathbf{w}_{VIS}(i,j)]. \quad (7)$$

In the above equation, the first term in the right hand side is the average of the smoothed IR and visible images. This provides the 'background' image to which the weighted deviations (the information) are added as the second term.

RESULTS AND DISCUSSION

In Figures 1 and 2, we show examples of the performance of the monochrome fusion stage. For the sake of comparison, we also show the image that would be obtained from direct superposition of the input IR and visible images (obtained by averaging the two input images). In order to illustrate more directly the effect of using principal components, we also have provided the image that would be obtained if equal weights were assigned to the IR and visible inputs, that is if $\mathbf{w}_{IR}$ and $\mathbf{w}_{VIS}$ were identical. This image appears more contrast enhanced than the one obtained by superposition because adding the deviations to the smoothed image in Equation (7) has the equivalent effect of high-pass filtering the image to a certain extent.

At the present time, we are exploring ways to reduce the computational overhead of having to compute the principal components separately for each still frame in moving imagery by using motion detection algorithms. The weights obtained from the principal component for a previous frame, then may be translated in the direction of motion. We are also actively looking at using the independent component analysis, a recently developed signal processing technique similar to principal components, for image fusion.

Das, S., Zhang, Y-L, Krebs, W.K. (2000). Color night vision for navigation and surveillance, In J. Sutton and S.C. Kak (Eds), Proceedings of the Fifth Joint Conference on Information Sciences, Atlantic City, NJ, February 28th

REFERENCES

(1) Waxman, A. M., Savoye E. D., Fay D. A., Aguilar M., Gove A. N., Carrick J. E. and Racamato J. P. Electronic Imaging Aids for Night Driving: Low Light CCD, Uncooled Thermal IR and Color Fused Visible/LWIR. *Proceedings of SPIE*, Vol. 2902 No. 2, 1997.

(2) Waxman, A. M., Gove A. N., Fay D. A., Racamato J. P., Carrick J. E., Siebert M. A. and Savoye E. D. Color Night Vision: Opponent Processing in the Fusion of Visible and IR Imagery. *Neural Networks*, Vol. 10, No 1., 1997.

(3) Krebs, W. K., Scribner D. A., Miller G. M., Ogawa J. S. and Schuler J. Beyond Third Generation: A Sensor Fusion Targeting FLIR Pod for the F/A-18. *Proceedings of SPIE*, Vol. 3376, 1998.

(4) Warren P., Scribner D., Schuler J., and Kruer M. Multi-Band Color Fusion. *Meeting of the IRIS Specialty Group on Passive Sensors*, March 1998.

ACKNOWLEDGEMENTS

The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, nor the United States Government.

Das, S., Zhang, Y-L, Krebs, W.K. (2000). Color night vision for navigation and surveillance, In J. Sutton and S.C. Kak (Eds), Proceedings of the Fifth Joint Conference on Information Sciences, Atlantic City, NJ, February 28th

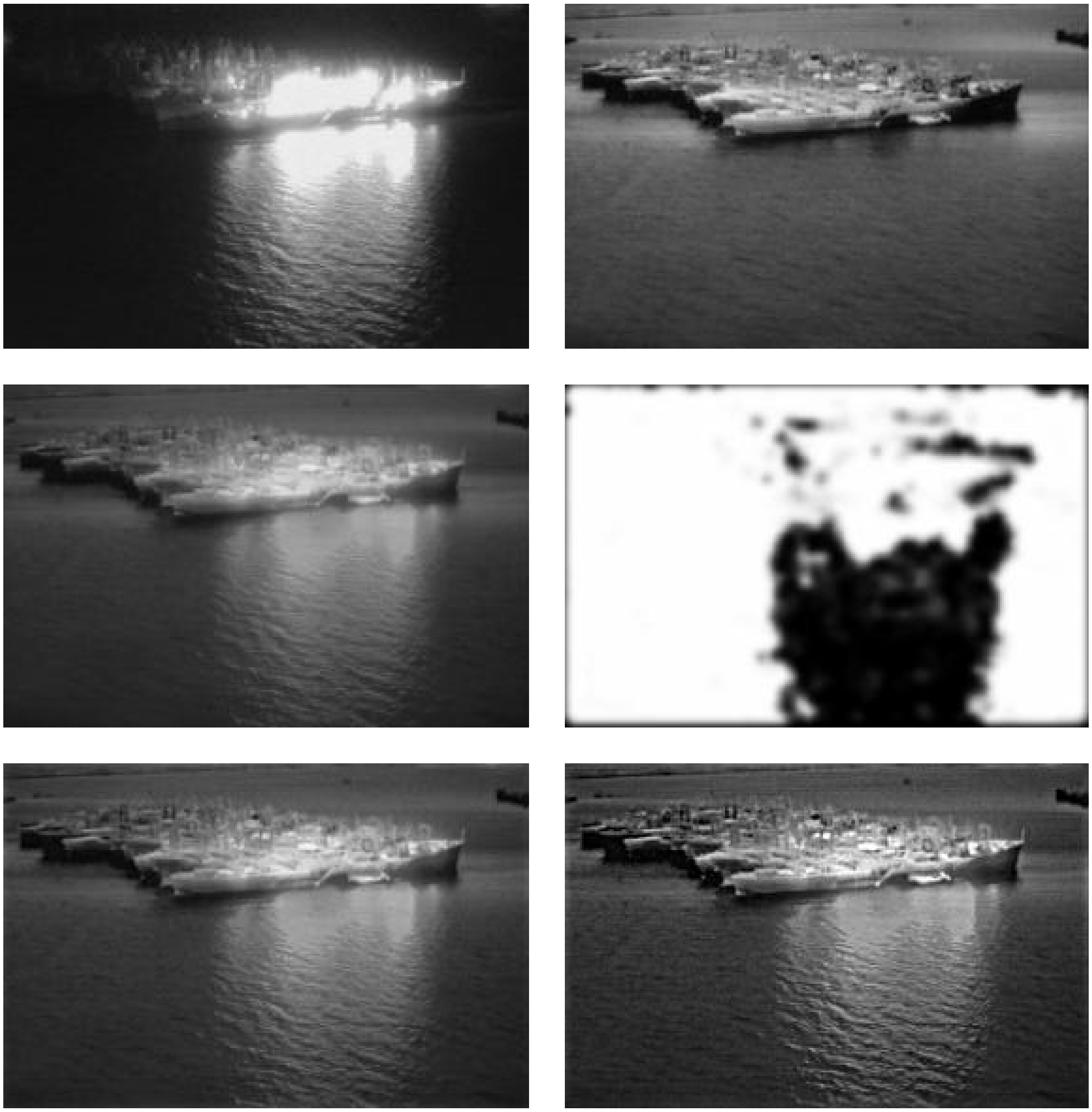

Figure 1. Images obtained from monochrome fusion. ***Top row***: Left - Visible image; Right - Infrared Image; ***Middle row***: Left - Image that would be obtained from superposition of the visible and infrared images; Right - The normalized weight assigned to the input images (**w**). Darker regions are where the visible image is assigned more weight as a result of more information content; ***Bottom row***: Left - Result of image fusion when equal weights are assigned to each input; Right - Result of the monochrome image fusion algorithm (**M**).

Das, S., Zhang, Y-L, Krebs, W.K. (2000). Color night vision for navigation and surveillance, In J. Sutton and S.C. Kak (Eds), Proceedings of the Fifth Joint Conference on Information Sciences, Atlantic City, NJ, February 28th

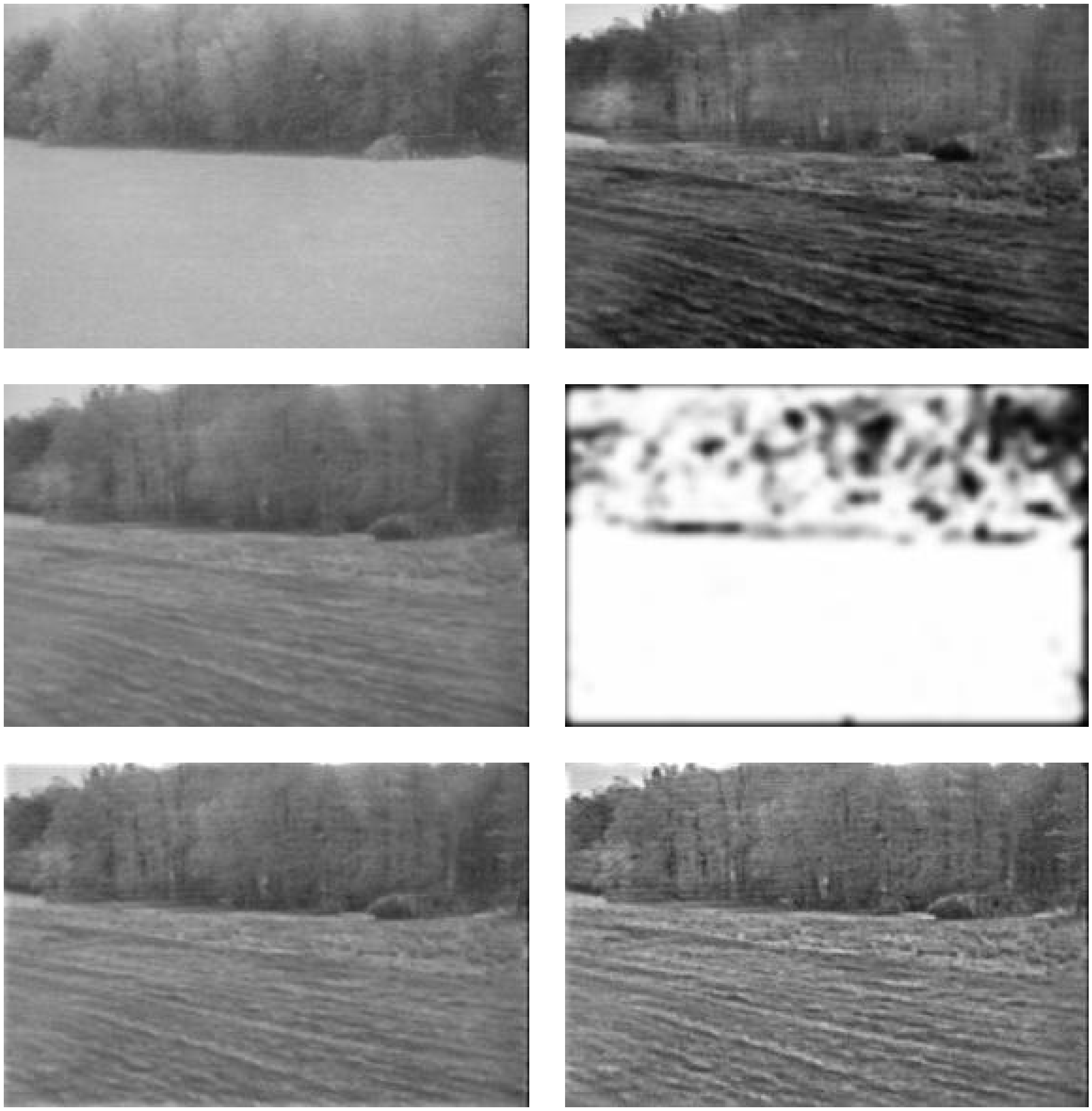

Figure 2. Images obtained from monochrome fusion. ***Top row***: Left - Visible image; Right - Infrared Image; ***Middle row***: Left - Image that would be obtained from superposition of the visible and infrared images; Right - The normalized weight assigned to the input images (**w**). Darker regions are where the visible image is assigned more weight as a result of more information content; ***Bottom row***: Left - Result of image fusion when equal weights are assigned to each input; Right - Result of the monochrome image fusion algorithm (**M**).