Visible and Long-Wave Infrared Image Fusion Schemes for Situational Awareness

Multi-Dimensional Digital Signal Processing

Literature Survey

Nathaniel Walker

The University of Texas at Austin

nathaniel.walker@baesystems.com

April 20, 2008

Abstract

Image fusion schemes are still a popular area of academic research, but the growing field has few agreed upon standards. Most schemes rely upon multi-scale decompositions. These algorithms contain many steps and complex decision models that may be difficult to implement in real-time video systems. They are also susceptible to artifacts and noise enhancement because they treat source images as equally likely contributors to the fused result. This paper will propose the study of a new brand of image fusion scheme for situational awareness applications where a visible light camera and an infrared camera are used. In the proposed scheme, the image with the most content will be defined as the primary image and used as the base for the fused image. The other source image is defined as secondary and used to overlay high contrast object information onto the primary image. During daylight hours the algorithm would use the visible light camera as primary and add objects of interest from the infrared sensor. This model maintains consistency with the intention of situational awareness systems to complement the human visual system with infrared content. During nighttime operation, the infrared sensor would be primary and the algorithm would add objects of interest from the visible light camera.

I. Introduction

Situational awareness is the ability to see, understand, and interpret your surroundings in such a way as to be able to make informed decisions. In environments where the human senses are obscured, distracted or overwhelmed, it is desirable to deploy electronic systems that perform the functions of the human senses and communicate the received information to a user. Such systems also provide an opportunity to enhance sensing capabilities beyond that of a typical human. It is no easy task to build electronics that recreate the complexity of the human senses to take in information, or the power of the human brain to combine this information and interpret it to make rational decisions. We have made remarkable progress in sensing technology that provides visual awareness. High-resolution optical cameras have long captured the world as we see it. Infrared detectors now allow us to provide imagery in spectral bands that the human eye cannot see. This allows us to see thermal content in a scene, as well as provide visibility in low-light and adverse weather conditions.

However, much work remains to determine the best way to combine these two sources of information for interpretation by the human visual system. Situational awareness systems are often limited not by their ability to sense the environment, but by their ability to effectively combine and communicate information to the user [1]. Often, image fusion techniques are evaluated subjectively by peer reviewers [2]. It would be helpful to develop quantitative criteria for objectively evaluating fusion techniques. Some work has been done to define a set of quantitative evaluation criteria [3] but the ideal fused image quality metric is not yet agreed upon.

Since humans do not normally see emissions in the infrared spectrum, we are not used to interpreting that data. Fusion algorithms must strike a balance between taking advantage of the new information that is available from infrared sensors, while presenting the information in a way that is familiar, not distracting, and easy to interpret. Specific applications must be considered, because the optimal fusion approach may depend on the scene content or application objectives.

Effectiveness is just one important characteristic of image fusion systems, although it has received most of the academic community's attention on this topic. Speed, efficiency and complexity of algorithms are also a concern in situational awareness applications. Most implementations in situational awareness operate on streaming video from multiple sensors for near real-time display. Low-latency and high frame rates (30 fps) are essential, driving a trade-off between image quality and algorithm complexity and speed. Also, many situational awareness applications are for mobile platforms (vehicles, UAVs, humans) where size, weight and power must be drivers for design. Sacrifices in image quality are often accepted if an algorithm can save system resources.

The focus of this paper will be to propose an image fusion algorithm that maintains image quality while improving simplicity and speed over those used in existing situational awareness systems. I am interested in applying the strengths of leading high-quality image fusion algorithms, while developing a simpler implementation for use in mobile situational awareness systems.

II. Background

A number of image fusion techniques have been proposed and studied. There have also been multiple frameworks proposed for classifying and categorizing techniques [4] [5]. Proposals for image quality metrics and fusion image evaluation have also varied [2] [6] [7]. The fragmentation in the research is reflective of the topic's early stage of development. The area of image fusion technology is still immature and growing with few standards in place and a limited history of existing implementations outside of academia.

The simplest form of fusion algorithm is additive. The fused result is a linear combination of the source images. This approach is not commonly used. In fact, there is little analysis available in the research because it is rarely used.

Most popular image fusion algorithms are based on multi-scale decomposition (MSD) techniques. An excellent survey of existing methods can be found in [4]. The MSD representation decomposes an image into contributions from different spatial frequencies. An MSD transform is performed on each source image, and then the coefficients are combined in some intelligent manner as determined by the fusion algorithm of choice. Finally, an inverse MSD transform produces the fused image.

MSD fusion schemes generally combine source image data by identifying edges and local areas of high contrast in each source, then transferring those edges to the fused image. One trouble is that many algorithms inadvertently add noise from both sources, in addition to actual scene content. The fused image then has more scene details than either source image alone, but it also has more noise than either source image alone. In an attempt to make the algorithms more robust to noise, they are layered with decision and verification mechanisms which increase execution time and latency.

The difficulty is that most algorithms attempt to transfer all fine detail from both source images to the fused image. Noise, depending on the fusion algorithm used, can look very much like fine detail and the algorithms pass it on to the fused image. This paper will propose research into algorithms that employ more stringent decision thresholds for image fusion. The aim is to continue to transfer prominent features to the fused image while reducing noise transfer. This will be at the cost of losing fine detail from the source image. However, this issue will also be addressed in the future work proposed. First we must review some existing methods.

III. Additive Technique, α-Blending

The additive approach to image fusion is the simplest of methods. The fused image can be as easy as a pixel by pixel average of the source images. A slightly more advanced variation will use weighting factors to weight contributions from the source images differently. The weighting factors can be applied to the image as a whole, or on a pixelwise basis. This is sometimes called an α-blend since the weighting factor is represented by α as shown in equation 1.

$$y(i,j) = \alpha x_1(i,j) + (1-\alpha)x_2(i,j) \qquad (1)$$

where: i,j are pixel indices in the images
α is the weighting factor
y is the output (fused) image
x_1, x_2 are the source images

The biggest downside to this approach is reduced contrast. Given two images, if one has a flat intensity or low contrast in an area of the image then it will flatten out the contrast in that area of the fused image. The weighting factors can help, but when applied globally you have to choose between limiting the low contrast effects and sacrificing the high contrast contributions. We'd rather only use the parts of each source image that have high contrast and edge information. This is the goal of multi-scale decomposition.

IV. Multi-Scale Decomposition Techniques

With multi-scale techniques, the source images are initially decomposed by spatial frequency content using a multi-scale transform. An overview of the Laplacian Pyramid decomposition can be found in [8]. The basic decomposition will filter the source into a low-pass and high-pass component of the image. The high-pass portion can be easily compressed because it is largely decorrelated. The low-pass portion can be down-sampled since it contains less information. After down-sampling, the process can be repeated on the low-pass image. The number of times the filtering and down-sampling is performed is called the level of decomposition.

The output of the transform is a series of image frames. Each frame contains coefficients that describe the source image content at a different frequency. But the multi-scale transform choice is just one step in the multi-scale fusion technique. There are many different ways to combine the source image coefficients into a fused image. An excellent review can be found in [4]. This reference presents a framework for differentiating multi-scale fusion methods, shown in Figure 1.

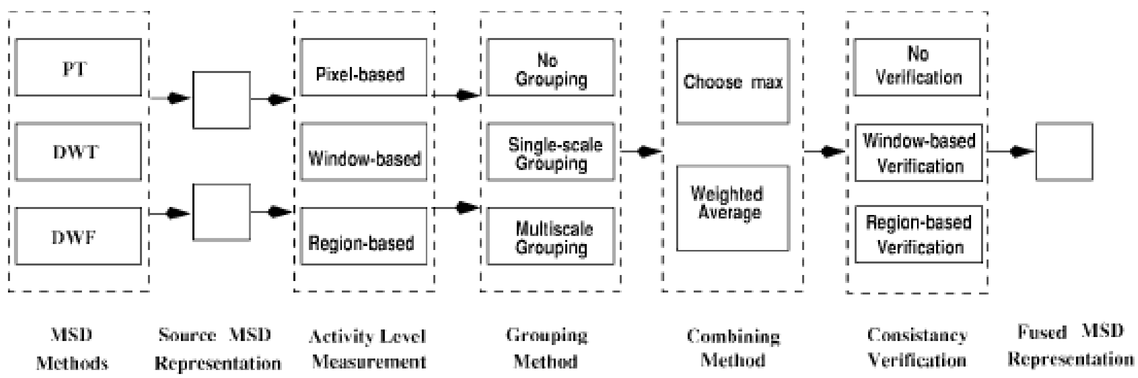

Figure 1: Multi-scale image fusion framework presented by Zhang and Blum [4]

The multi-scale decomposition is only the first step of the algorithms. It converts the source data into a format that allows the design of intelligent fusion algorithms based on spectral content. The goal of these algorithms is to identify areas in each source image that contain lots of information and add these to the fused image.

After decomposition, the activity level measurement of each coefficient determines how much 'information' is present at that coefficient in the source image. These measurements can be made at each coefficient individually, in a small window around each coefficient, or in the region that the coefficient belongs. Region-based measurements require a separate image segmentation filter that divides the image into a discrete number of regions. The activity level measurements are used to determine how each source will be combined to form the fused image. In some algorithms, coefficients with the highest activity levels are added to the fused image, but in others the fused coefficient is a weighted average of the source coefficients. Optionally, verification can be performed to ensure that coefficients in a window or region of the fused image are taken from the same source image. The final image is achieved with an inverse multi-scale transform.

V. Comparison of Existing Scheme Performance

Chen and Blum [2] put many algorithms to the test with a series of subjective and objective evaluations. The algorithms they used are listed in Figure 2. The source images and some samples of the fused results are given in Figure 3.

Fusion Scheme	Reference	MSD level	Activity Measure	Grouping	Combining	Weights(IITV/IR)	Verification
ADD	[3]	None	None	None	None	70%/30%	None
DWT1	[13, 7]	4	Coefficient based	None	Weighted Average	70%/30%	None
LAP1	[8, 7]	4	Coefficient based	None	Weighted Average	70%/30%	None
FSD1	[9, 7]	4	Coefficient based	None	Weighted Average	70%/30%	None
GRAD1	[10, 7]	4	Coefficient based	None	Weighted Average	70%/30%	None
MORPH1	[11, 7]	4	Coefficient based	None	Weighted Average	70%/30%	None
SiDWT1	[14, 7]	4	Coefficient based	None	Weighted Average	70%/30%	None
DWT2	[13, 7]	4	Coefficient based	None	Choosing Maximum	None	None
LAP2	[8, 7]	4	Coefficient based	None	Choosing Maximum	None	None
FSD2	[9, 7]	4	Coefficient based	None	Choosing Maximum	None	None
GRAD2	[10, 7]	4	Coefficient based	None	Choosing Maximum	None	None
MORPH2	[11, 7]	4	Coefficient based	None	Choosing Maximum	None	None
SiDWT2	[14, 7]	4	Coefficient based	None	Choosing Maximum	None	None

Figure 2: Fusion schemes evaluated by Chen and Blum [2]

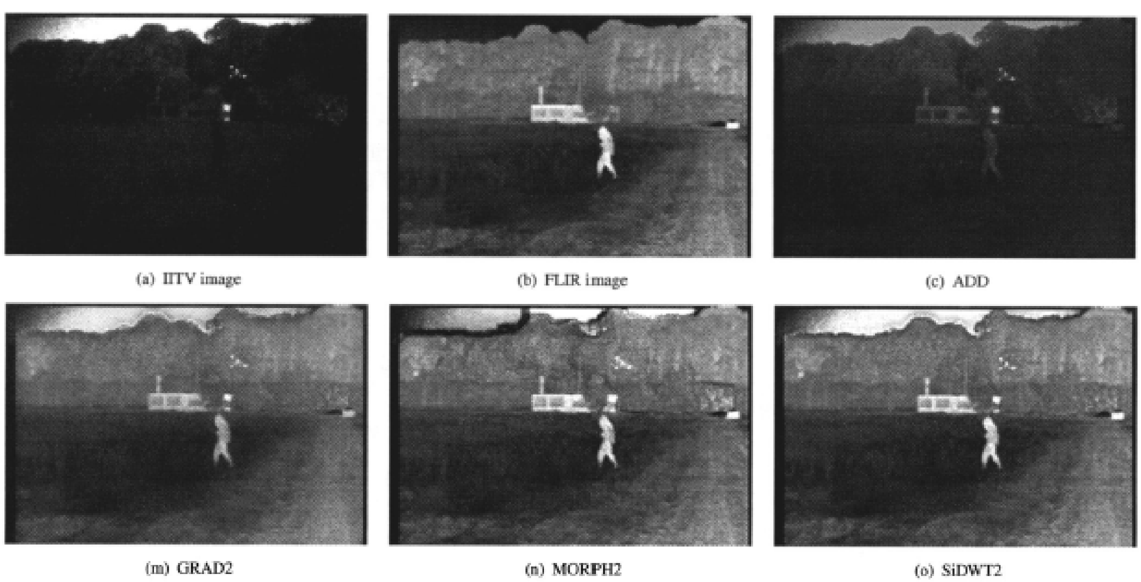

Figure 3: Source images (a) and (b) and results of four fusion algorithms from Chen and Blum [2]

The additive method is shown in image (c). It does contain elements of both source images. A good fusion algorithm would maintain the contrast and information from at least the best source image, but the additive result shows much lower contrast and detail than the FLIR source image alone.

The three multi-scale algorithms are all improvements over the additive method in contrast. However, we can also see strange edge effects in all of them, particularly where the sky meets the trees. This is because the sky is black in one source and white in the other. These opposite representations of the same scene compete for inclusion in the fused image resulting in undesired artifacts. In these images, both sources are noisy so it's difficult to prove that the fused image noise is a combination of the source noises. However, in [9], the writers observe in their experiments that their fused images combine noise and artifacts from both source images.

VI. Future Work

Existing algorithms assume that both source images will contribute to the fused image equally, or that it isn't known which of the source images will provide most of the fused image content. In most scenarios this assumption for situational awareness is invalid. Consequently, conflicting information from both sources competes in the fused image creating artifacts. Also, image noise and reduced contrast from the poorer source image effects the fused result.

During the daytime, the resolution, sensitivity and dynamic range of the visible light camera will usually be superior to the image of the infrared sensor. Conversely, at night the infrared provides a clear image while the visible sensor contains very little useful information. In designing a fusion algorithm for situational awareness, we can choose one source image as the primary image and use it as a basis for the fused result. The other source can be considered the secondary image and used to add complementary information to the fused image that is not included in the primary source image.

I propose a framework that defines the two source images as primary and secondary images. This determination can be made using a histogram analysis or an entropy measure. The secondary image will then be region-segmented using an existing segmentation algorithm. Regions in the secondary image exceeding a threshold of activity or contrast level will be identified and fused into the primary image. The threshold can be adjusted to add more or less secondary image content.

This approach should minimize noise contribution from the secondary image since only identified regions of interest from that image will be added. Also, this approach is consistent with the intention of a situational awareness system to act as an enhancement to human visual capabilities. During daytime conditions the fused image would be similar to the view seen by the human eye with contributions from the infrared detector only where feature detections unique to the infrared domain are present. This scheme also eliminates some steps from the multi-scale schemes, including the transforms, so implementation should be comparatively less time-consuming and resource intensive.

References

[1] M. Gilger. 'Addressing Information Display Weaknesses for Situational Awareness,' *Conference on Machine Learning and Applications*, pp. 109-116, Dec 2006.

[2] Y. Chen, R. Blum. 'Experimental Tests of Image Fusion for Night Vision Applications,' *International Conference on Information Fusion*, Volume 1, pp. 25-28, July 2005.

[3] Q. Wang, Y. Shen, Y. Zhang, and J. Qiu Zhang. 'A Quantitative Method for Evaluation the Performances of Hyperspectral Image Fusion,' *IEEE Transactions on Instrumentation and Measurement*, Volume 52, Issue 4, pp. 1041-1047, Aug 2003.

[4] Z. Zhang, R.Blum. 'A Categorization of Multiscale-Decomposition-Based Image Fusion Schemes with a Performance Study for a Digital Camera Application,' *Proceedings of the IEEE*, Volume 87, Issue 8, pp. 1315-1326, Aug 1999.

[5] Z. Wang, D. Ziou, C. Armenakis, D. Li, Q. Li. 'A Comparitive Analysis of Image Fusion Methods,' *IEEE Transactions on Geoscience and Remote Sensing*, Volume 43, Issue 6, pp. 1391-1402, June 2005.

[6] C.S. Xydeas, V. Petrovic. 'Objective Image Fusion Performance Measure,' *Electronics Letters*, Volume 36, Issue 4, pp. 308-309, 17 Feb 2000.

[7] G. Piella, H. Heijmans. 'A New Quality Metric for Image Fusion,' *International Conference on Image Processing*, Volume 3, pp. 173-176, Sept. 2003.

[8] P. Burt, E. Adelson. 'The Laplacian Pyramid as a Compact Image Code,' *IEEE Transactions on Communications*, Volume 31, Issue 4, pp. 532-540, April 1983.

[9] L. St-Laurent, X. Maldague, D. Prèvost. 'Combination of Colour and Thermal Sensors for Enhanced Object Detection,' *International Conference on Information Fusion*, pp. 1-8, 9-12 July 2007