Image Quality Assessment for Performance Evaluation of Image Fusion

Erik Blasch

AFRL/RYAA Evaluation Branch
WPAFB
OH 45433, U.S.A.
erik.blasch@wpafb.af.mil

Xiaokun Li, Genshe Chen, Wenhua Li

DCM Research Resources, LLC
20201 Century Boulevard, Suite 150
MD 20874, U.S.A.
{xli, gchen, wli}@dcmresearchresources.com

Abstract - We present a novel approach on objective non-reference image fusion performance assessment. The Global-Local Image Quality Analysis (GLIQA) approach takes into account local measurements to estimate how well the important information in the source images is represented by the fused image. The metric is an extended version of the Universal Image Quality Index (UIQI) and uses the similarity between blocks of pixels in the input images and the fused image as the weighting factors. When the difference of an image pixel in the input images and its correspondence in the fused image is larger than a threshold and difficult to assess the fusion quality, global measurements will be applied to assist the judgment. The global measurement metric considers a set of properties of human Gestalt visual perception, such as image structure, texture, and spectral signature, for image quality assessment. Preliminary study results confirm that the performance scores of the proposed metrics correlate well with the subjective quality of the fused images.

Keywords: Fusion performance evaluation, image fusion, non-reference quality measures, objective quality measures.

1 Introduction

For the performance evaluation of different image fusion methods, subjective criteria are most commonly used as the human perception of the fused image is of fundamental importance. Objective image fusion performance evaluation is a challenging task due to different application scenarios and the lack of a clearly defined and documented ground truth. Scenarios typically include complex structure over varying operating conditions such as illumination changes in the environment, varying parameters in sensor designs and exploitation algorithms, and a host of target sizes. Each of these variations require

fusion algorithms to adapt in real time. Some fusion algorithms have been evaluated objectively by constructing an "ideal" fused image and using it as a reference to compare with the experimental results. Mean squared error (MSE), peak signal to noise ratio (PSNR), and signal to noise ratio (SNR) based metrics were widely used for these comparisons.

In recent years, several objective performance measures [1-2] for image quality analysis were proposed where the knowledge of ground truth is not assumed. Unfortunately, none of them has shown any obvious advantage over the simple measures such as RMSE and PSNR under strict testing conditions and with different levels of image distortion [3-5]. Despite this, more efforts on objective fusion metrics without knowing display equipment or audience have been made [6-12]. One such evaluation approach is based on the *Universal Image Quality Index* [7] where local image statistics are used to define a similarity between all corresponding 8×8 blocks across input and fused images. Information theoretic measures based on global image statistics such as entropy and mutual information have also been considered within the context of fusion evaluation [8], [10]. But, only considering either local information or global information cannot work well on quality analysis of the fused images in a variety of conditions. Obviously, we need to incorporate them (local and global) together for more reliable and comprehensive performance evaluation. The new Global-Local Image Quality Analysis (GLIQA) performance metric will also provide helpful guidance not only to the performance assessment of image fusion with the same-type imaging sensors, but also to the image fusion with multi-modal sensors worked in 24/7 operational environments [16-17].

In this paper, we propose the global-local image quality analysis (GLIQA) approach to incorporate local image quality evaluation and global quality analysis together for more reliable performance analysis to image fusion

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **JUL 2008**	2. REPORT TYPE	3. DATES COVERED **00-00-2008 to 00-00-2008**
4. TITLE AND SUBTITLE **Image Quality Assessment for Performance Evaluation of Image Fusion**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Research Laboratory,AFRL/RYAA Evaluation Branch,Wright Patterson AFB,OH,45433**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
11th International Conference on Information Fusion, June 30 ? July 3, 2008, Cologne, Germany.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**6**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

without knowing the ground truth. The approach consists of two quality metrics: *local quality metric* and *global quality metric*. In our framework, we use local metric and global metric collaboratively for image quality analysis. The final value of the quality measure will be the values from the two metrics.

The paper highlights an example of image fusion performance evaluation with a focus on reliable and robust metrics. Section 2 details the local quality metric, while Section 3 defines the global quality metric. Section 4 defines the experiment as an exemplar case in image fusion performance analysis. Section 5 draws conclusions.

2 Definition of the Local Image Quality Index

Wang and Bovik [13] proposed a universal objective image quality assessment metric which is easy to calculate and applicable to various image processing applications. In this section, we first give a brief introduction to the method. Then, we present a modified metric which is based on Wang-Bovik's metric.

2.1 A universal image quality index (UIQI)

Instead of using traditional error summation methods, the method proposed by Wang and Bovik was designed to model any image distortion via a combination of three factors: loss of correlation, luminance distortion, and contrast distortion.

More specifically, given two real valued sequences $x = \{x_1,...,x_n\}$ and $y = \{y_1,...,y_n\}$, $\bar{x}$ is the mean of x, σ_x^2 is variance of x, σ_y^2 is variance of y and σ_{xy} is covariance of x, y.

$$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \ , \ \ \sigma_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \quad (1)$$

$$\sigma_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \quad (2)$$

Then, we can compute

$$Q = \frac{4\sigma_{xy}\overline{xy}}{(\bar{x}^2 + \bar{y}^2)(\sigma_x^2 + \sigma_y^2)} \quad (3)$$

The Q can be decomposed into three components as

$$Q = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \cdot \frac{2\overline{xy}}{\bar{x}^2 + \bar{y}^2} \cdot \frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \quad (4)$$

Q = correlation • (luminance) • contrast

The first component is the correlation coefficient between x and y, which measures the degree of linear *correlation*

between and x and y. The second component measures how close the mean *luminance* is between x and y. The third component measures how similar the contrasts of the images are as σ_x and σ_y can be viewed as estimate of the *contrast* of x and y. The value of the three components is in the range of [0, 1]. Therefore, the final value of the quality metric is normalized between [0, 1].

Wang and Bovik apply the metric to local regions using a sliding window for objective image quality analysis. Started from the top-left corner of the image, a sliding window with the size of B × B is moved pixel by pixel horizontally and vertically through all pixels of the image. At the position of (*i,j*), the local quality index Q_{ij} is computed. If the row number and column number of the image are N and M, then the overall normalized quality index is:

$$Q = \frac{1}{N \times M}\sum_{i=1}^{N}\sum_{j=1}^{M} Q_{ij} \quad (5)$$

2.2 Image fusion quality metric

The quality index proposed by Wang-Bovik has been proven very efficient on image fusion performance evaluation as it considers three factors which are crucial in image quality measurement. Besides these three factors, many studies show that in human visual system (HVS) gradient (edge) information plays a very important role when human subject judges the quality of an image. To take the advantage of known characteristics of the human perception, we add the *local gradient information* into the UIQI metric. Before performing quality analysis, we use an edge detector (such as Sobel operator) to quickly process the image and get gradient information for each image pixel, which is denoted as g. Therefore, the new UIQI metric can be presented as:

$$Q = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \cdot \frac{2\overline{xy}}{\bar{x}^2 + \bar{y}^2} \cdot \frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \cdot \frac{2g_x g_y}{g_x^2 + g_y^2} \quad (6)$$

Here, we give the brief introduction on how to apply the new UIQI metric to image fusion performance evaluation. For the fusion of input image A and image B resulting in a fused image F. When the sliding window moves pixel by pixel over the image A and image F at the same time, we can get the following:

$$Q_{A/F} = \frac{1}{N \times M}\sum_{i=1}^{N}\sum_{j=1}^{M} Q_{ij}$$

We do the same procedure to the image B and image F. Then, we can obtain:

$$Q_{B/F} = \frac{1}{N \times M} \sum_{i=1}^{N} \sum_{j=1}^{M} Q_{ij}$$

Then, the fusion quality index can be given as

$$Q_{AB/F} = \frac{1}{N \times M} \sum_{i=1}^{N} \sum_{j}^{M} (\lambda(i,j)Q_{A/F}(i,j) + (1-\lambda(i,j)Q_{B/F}(i,j)) \quad (7)$$

where $\lambda(i, j)$ is a local weight between 0 and 1. To automatically set a suitable value to $\lambda(i, j)$ is also very important. In our research, we calculate $\lambda(i, j)$ according to the *local saliencies* of the sliding window of image A and image B. The value of saliencies is computed from the local contrast and sharpness. If we denote the value of saliencies as *s*, we can have the following equation to compute $\lambda(i, j)$:

$$\lambda(i, j) = \frac{s(A(i, j))}{s(A(i, j)) + s(B(i, j))} \quad (8)$$

3   Definition of the Global Image Quality Index

The fundamental problem of the local measure based approaches is the definition of image quality. In particular, it is not clear that the error/difference visibility based analysis should be equated with loss of quality. For quality measurement via HVS, the evaluation processing heavily depends on both psychophysical and physiological experience of the human subject. Many studies, such as [15], show that the properties of the whole image act very important role in HVS on quality assessment.

Another important reason for considering global properties is when the difference of an image pixel in one input image and its correspondence in the fused image is significantly different from each other, while the image pixel in another input image is very close to its correspondence in the fused image, we cannot say the fusion quality is bad or not. In this case, we need to go back to a larger region or the whole image to decide the fusion quality through analyzing the properties of the entire image. In this section, we propose a global image quality metric for fusion performance measurement. In the metric, the important properties to HVS, such as image structure, texture, and spectral signature, are all considered.

3.1   Image Structure

In computer vision, finding objects or regions of a given type (such as woods, grass, building, sky, etc.) in a photograph, is known as *image segmentation*. After the segmentation, the image structure can be easily obtained

and presented. Finding ways of doing both automatically is of great interest to researchers. Among the current image segmentation methods, *graph cut* is recognized a very efficient and fast algorithm which can optimally segment an image into different patches as illustrated in Fig. 1. Graph cut provides a clean, flexible formulation for image segmentation. It provides a convenient language to encode simple local segmentation cues, and a set of powerful computational mechanisms to extract global segmentation from the simple local (pairwise) pixel similarity. According different parameter settings, the patch number is controllable, which means the user can control the level of image segmentation.
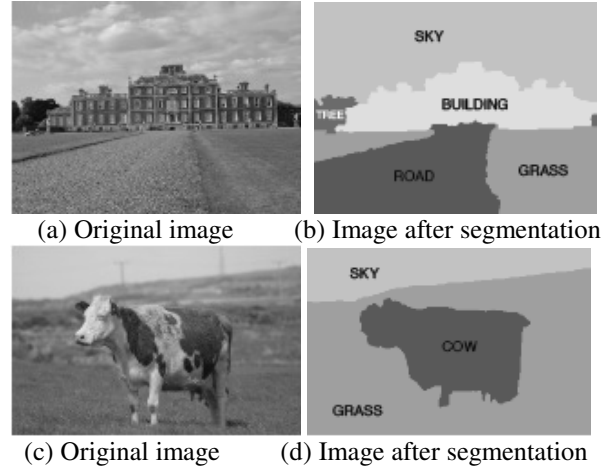

(a) Original image   (b) Image after segmentation

(c) Original image   (d) Image after segmentation

Fig. 1 Image structure (From [18])

3.2   Image Texture

In image processing, simplifying assumptions are made about the uniformity of intensities in local image regions. However, images of real objects often do not exhibit regions of uniform intensities. For example, the image of a wooden surface is not uniform but contains variations of intensities which form certain repeated patterns called *visual texture*. The patterns can be the result of physical surface properties such as roughness or oriented strands which often have a tactile quality, or they could be the result of reflectance differences such as the color on a surface. Fig. 2 shows an example of image texture.

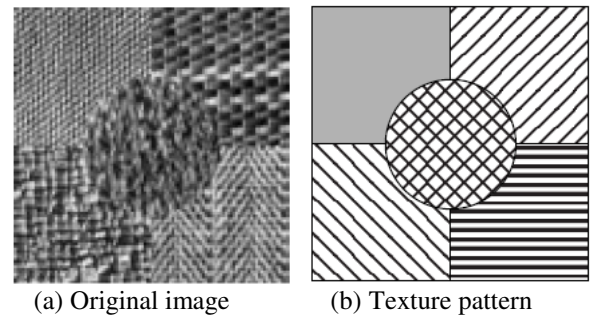(a) Original image   (b) Texture pattern

Fig. 2 Image texture

Image texture, defined as a function of the spatial variation in pixel intensities (gray values), is useful in a variety of applications as well as in image fusion evaluation. Texture is the most important visual cue in identifying types of homogeneous regions. Based on textural properties, we can identify a variety of materials and their pattern in the image as shown in Fig 3..
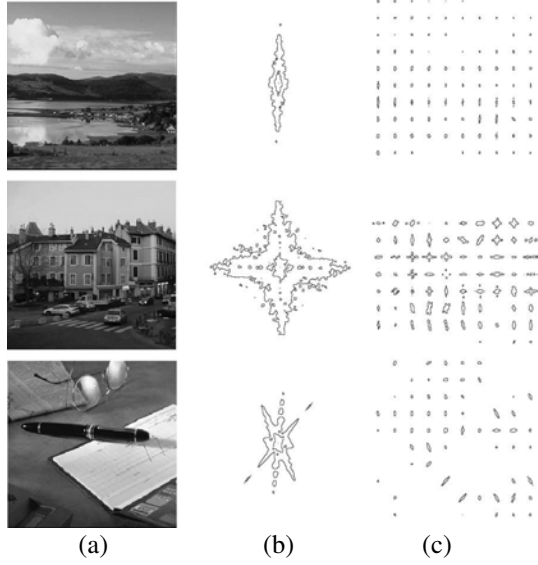

(a)   (b)   (c)

Fig. 3 Three examples for spectral signature (From [19]):(a) Original images; (b) Global magnitude of Fourier Transform; (c) the magnitude of the windowed Fourier Transform

3.3   Spectral Signature

Spectral properties of an image can be described and studied by discrete Fourier transform (DFT) or wavelet transform. Since there are many real-time codes for DFT, we select DFT for our study. The DFT of an image can be defined as:

$$I(f) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} i(x)h(x)e^{-j2\pi<f,x>} \qquad (9)$$

where $i(x)$ is the intensity distribution of the image along the spatial variables $x = (x, y)$; $j = \sqrt{-1}$; and the spatial frequency variables are defined by $f = (f_x, f_y) \in$ [-0.5, 0.5]$\times$ [−0.5, 0.5] (units are in cycles per pixel); and $h(x)$ is a circular window that reduces boundary effects. The amplitude spectrum is defined as the magnitude of the DFT: $A(f) = |I(f)|$. The amplitude spectrum reveals the dominant orientations and textural patterns in the image as shown in Fig. 3.

3.4   Global Image quality index

With the information of image structure, texture, and spectral signature, we can construct a global image quality

index by modifying Equation (6) to the following expression.

$$Q = \frac{2s_x s_y}{s_x^2 + s_y^2} \cdot \frac{2t_x t_y}{t_x^2 + t_y^2} \cdot \frac{2f_x f_y}{f_x^2 + f_y^2} \qquad (10)$$

Q = structure • (texture) • spectral

where s, t, and f are the quality value of image structure, texture, and spectral signature respectively. During image quality assessment, when the difference of an image pixel in the input image and its correspondence in the fused image is significantly large and difficult to estimate the fusion quality, global measurements will be applied to asses the image quality.

4   Experimental results

Our proposed Global-local Image Quality Analysis (GLIQA) method for image fusion performance evaluation was tested by several representative samples of multi-sensor image sequences including both urban and natural scenes. Sequences were fused by three representative fusion schemes: Laplacian-Pyramid (LP) based image fusion method, Discrete-Wavelet-Transform (DWT) based image fusion method, and Dual-Tree-Complex-Wavelet-Transform (DT-CWT) based image fusion method. The evaluation based on our proposed GLIQA metric was compared to two traditional image fusion evaluation metrics: UIQI metric [13] and SSIM (structural similarity) metric [14].

Some experimental results selected from multiple-sensor (CCD camera and IR camera) data are shown in Fig. 4 and Fig. 5. These testing data are provided by OCTEC [20] and TNO Human Factors Research Institute [21].
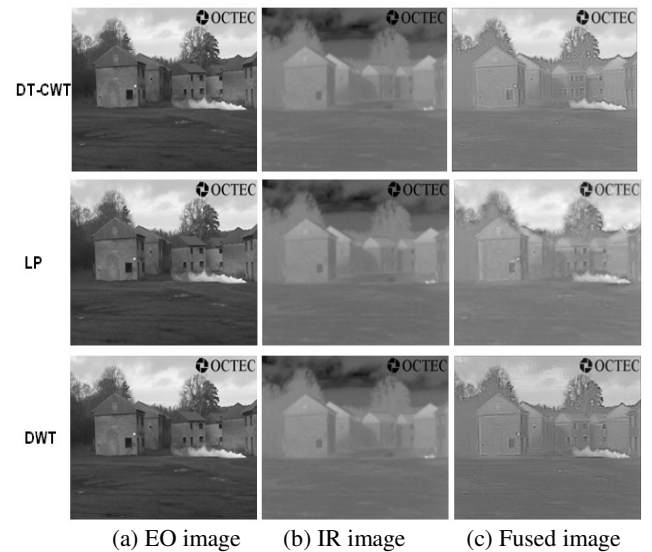

(a) EO image   (b) IR image   (c) Fused image

Fig. 4 Example fused frames from the OCTEC sequence by using three representative image fusion methods

<table>
<tr><td>DT-CWT</td></tr>
<tr><td>LP</td></tr>
<tr><td>DWT</td></tr>
</table>
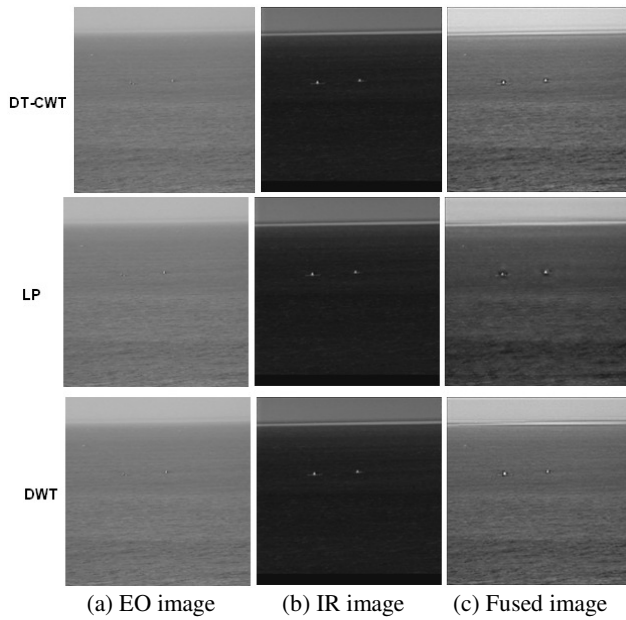
    (a) EO image   (b) IR image   (c) Fused image

Fig. 5 Example fused frames from the maritime sequence by using three representative image fusion methods

The mean performance scores of UIQI, SSIM, and GLIQA are listed in Table 1 to Table 2. The mean subjective scores made by people are in the range of 0.60 to 0.65 for the fused images. From the scores in the tables, we can see that UIQI has the biggest deviation to the scores made by human subjects because it only considers local information of an image. SSIM includes image-structure information in its image fusion quality metric. Therefore, SSIM has much better performance than that of UIQI. Since our proposed GLIQA metric considers both global and local information of an image, its performance scores are very close to the HVS (human visual system) values made by people. From the scores in the tables, the overall performance of GLIQA is the best of the three image quality evaluation methods.

Table 1: OCTEC

Scheme	DT-CWT	LP	DWT
UIQI	0.352	0.321	0.313
SSIM	0.521	0.534	0.543
GLIQA	0.562	0.614	0.601

Table 2: TNO

Scheme	DT-CWT	LP	DWT
UIQI	0.293	0.223	0.208
SSIM	0.585	0.571	0.593
GLIQA	0.610	0.640	0.633

5   Conclusion

In this paper we have discussed some new objective quality measures, which do not require a reference image, for image fusion performance evaluation. The global-local image quality analysis (GLIQA) method correlates well with subjective criteria as with other existing performance measures. Our measures are easy to calculate, and applicable to many sceanrion particular, our measures provide good results on variable quality of input images as it takes into account the local information as well as the global properties of these images.

References

[1]   T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation," in *Handbook of Image and video Processing*, Academic Press, May 2000.

[2]   "Special issue on image and video quality metrics," Signal Processing, (Edited by Murat Kunt), Vol. 70, Nov. 1998.

[3]   J.-B. Martens and L. Meesters, "Image dissimilarity," *Signal Processing*, Vol. 70, pp. 155-176, Nov. 1998.

[4]   VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," http://www.vqeg.org/, Mar. 2000.

[5]   A. M. Eskicioglu and P.S. Fisher, "Image quality measures and their performance," *IEEE Trans. Communications*, vol. 34, pp. 2959-2965, Dec. 1995.

[6]   C. Xydeas, V. Petrovic, "Objective image fusion performance measure," *Electronics Letters*, vol. 36(4), pp. 308-309, IEE, 2000

[7]   G. Piella, H. Heijmans, "A new quality metric for image fusion," *IEEE Conference on Image Processing*, vol. 3, pp. 173-176, 2003

[8]   G. Qu, D. Zhang, P. Yan, "Information measure for performance of image fusion," *Electronics Letters*, vol. 38(7), pp. 313-315, IEE, 2002.

[9]   V. Tsagaris, V. Anastassopoulos, "Global measure for assessing image fusion methods," *Optical Engineering*, Vol. 45, SPIE, 2006.

[10] V. Petrovic, T. Cootes, "Information Representation for image fusion evaluation," *Proceedings of Fusion*, Florence, ISIF, 2006.

[11] N. Cvejic, D. Bull, C. Canegarjah, "A new metric for multimodal image sensor fusion," *Electronics Letters*, Vol. 43(2), pp. 95-96, IEE, 2007

[12] V. Petrovic, C. Xydeas, "Objective evaluation of signal-level image fusion performance," *Optical Engineering*, SPIE, Vol. 44(8), 087003, 2005

[13] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing letters*, vol. 9, no.3, pp. 81-84, March 2002

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Processing*, Vol. 13, No.4, April 2004.

[15] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult," in *Proc. IEEE Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 3313- 3316, FL, 2002.

[16] Y. Chen and R. S. Blum, "Experimental Tests of Image Fusion for Night Vision", *Eighth International Conference on Information Fusion*, Philadelphia, PA, July 2005.

[17] B. Kahler, E. Blasch, D. J. Pikas, and T. Ross, EO/IR ATR Performance Modeling To Support Fusion Experimentation, SPIE Defense and Security Symposium.

[18] C. Galleguillos, A. Rabinovich, S. Belongie, "Object categorization using co-occurrence, location and appearance," *IEEE Conference on CVPR*, 2008.

[19] A. Torralba and A. Oliva, "Depth estimation from image structure," *IEEE Trans. PAMI*, Vol. 24, No.9, Sept. 2002.

[20] The Online Resource for Research in Image Fusion, www.imagefusion.org, 2008.

[21] Toet, A., "Detection of dim point targets in cluttered maritime backgrounds through multi-sensor image fusion," In: W.R. Watkins, D. Clement & W.R. Reynolds (Ed.)*, Targets and Backgrounds: Characterization and Representation VIII, The International Society for Optical Engineering*, pp. 118-129, Bellingham, WA, 2002.