

# Multi-Spectral Image Fusion for Visual Display

Tamar Peli<sup>a</sup>, Eli Peli<sup>b</sup>, Ken Ellis<sup>a</sup>, Robert Stahl<sup>a\*</sup>

<sup>a</sup>Atlantic Aerospace Electronics Corporation  
470 Totten Pond Road, Waltham, MA 02154

<sup>b</sup>Schepens Eye Research Institute, Harvard Medical School  
20 Staniford Street, Boston, MA 02114

## ABSTRACT

This paper describes a contrast-based monochromatic fusion process. The fusion process is aimed for on board real time application and it is based on practical and computationally efficient image processing components. The process maximizes the information content in the combined image, while retaining visual clues that are essential for navigation/piloting tasks. The method is a multi scale fusion process that provides a combination of pixel selection from a single image and a weighing of the two/multiple images. The spectral region is divided into spatial sub bands of different scales and orientations, and within each scale a combination rule for the corresponding pixels taken from the two components is applied. Even when the combination rule is a binary selection the combined fused image may have a combination of pixel values taken from the two components at various scales since it is taken at each scale. The visual band input is given preference in low scale, large features fusion. This fusion process provides a fused image better tuned to the natural and intuitive human perception. This is necessary for pilotage and navigation under stressful conditions, while maintaining or enhancing the targeting detection and recognition performance of proven display fusion methodologies. The fusion concept was demonstrated against imagery from image intensifiers and forward looking infrared sensors currently used by the U.S. Navy for navigation and targeting. The approach is easily extendible to more than two bands.

**Key Words:** Image Fusion, contrast, visual model, night pilotage

## 1. INTRODUCTION

Significant improvements were made in night vision devices/sensors over the last two decades to aid military forces in conducting night operations. There are two types of night vision devices currently in use; image intensifiers ( $I^2$ ) and forward looking infrared (FLIR) sensors. Each sensor provides a monochrome image that can be used for either navigation and/or targeting. Image intensifiers collect reflected energy in the 600 to 900 nm range, while FLIRs collect emitted energy from the far infrared spectrum (typically 8 to 12 $\mu$ m). Targets as well as terrain features may exhibit noticeable contrast in either domain dependent on a variety of conditions and environments. The two sensors tend to excel under different conditions and as a result of their complementary nature there may be a strong desire and benefit from having imagery from the two sensors during a night flight mission.

In addition to improved visibility at night, detection performance of targets in concealment camouflage and deception (CC&D) conditions is greatly improved, if multi-spectral imagery is available. This was demonstrated [Peli 1997a] for automatic target detection using multi-spectral imagery collected by the ERIM M7 and Daedalus 1268. For a human operator the multiple sources of imagery need to be fused and displayed in a form that is easy and natural to interpret and that will result in improved targeting and navigation performance.

Our long-term objective is to develop a robust color coding method that utilizes the output of a monochrome multi-spectral fusion process. This paper describes the new contrast-based monochromatic fusion process that maximizes the information content in the combined image while retaining visual clues that are essential for navigation/piloting tasks. The monochrome fusion process selects features from each band based on visually relevant contrast measures, but performs the fusion in the amplitude domain. The resulting algorithm was tested by fusion of imagery from image intensifiers and forward looking

---

\*former employee

infrared sensors currently used by the U.S. Navy for navigation and targeting. The monochrome fusion stage was tested against a wide range of imagery conditions, and was shown to capture the details from both inputs using a single parameter setting (no image-by-image tuning). The fused imagery was also shown to preserve the depth (shading) information of the visible input. We developed multiple methods for color coding which are derivatives of our technique of polarity preserving color coding. These methods will be reported separately.

### 1.1. Current fusion technologies

Current fusion procedures can be divided into multiple categories based on two main variables; color vs. monochrome, and single scale vs. multi-scale. The use of color in image fusion was frequently advocated under the argument that color contrast can provide improved detection performance when added to luminance contrast. This was demonstrated for very low, near threshold contrast [Gur 1993]. This improvement, however, disappeared when contrast is increased. In a targeting function, both detection and discrimination are required. The contrast level required for discrimination may be sufficiently higher to reduce the noticed improvement in detection performance. The investigation of Perconti et al [Perconti] also indicates that the benefits of color representation of fused imagery is not clear for pilotage tasks, but may be of utility for targeting.

A typical color fusion approach presents the multiple (monochrome) bands or a derivative of them using color displays by transforming (projecting) them onto display variables such as the RGB or the Luminance-chroma-saturation. This approach is taking advantage of the observer's color vision to introduce additional dimensionality for interpretation. For up to three bands, the apparent advantage is that one does not have to make a choice and discard information - all of the information from the multiple bands can be presented in a way that the visual system will make the selection/decision itself. Both the Naval Research Lab [Scribner 1993] and Lincoln Laboratory [Waxman 1995] have developed color fusion procedures for human observers.

The monochrome approach generates a single image from the multiple (typically two) bands. The rules for combination vary from selecting pixels from a single image to weighting of the two images [Morgan 1991]. The combination rule/decision may be based on the amplitude to make this selection (i.e. Paval et al), or on a contrast measure [Toet 1992], or on a contrast-like measure [Waxman 1995]. In either of these cases the fusion is performed in the decision domain (e.g. fusion in contrast space if decision is based on contrast).

Once a mode of display (color or monochrome) has been selected, the decision is between "single" scale and multi-scale image decomposition for fusion. Waxman's [Waxman 1995] shunting approach is essentially a bandpass filtering method. Each band is processed with such filtering to yield a single scale, which is combined using the same method with one of the other bands. The specific spatial scale to be selected is not specified in the various reports, but it is determined by a few of the free parameters defined for this method. The application of the Peli and Lim algorithm [Peli 1979] for multi-band fusion developed by [Scrofani 1997] also represents a single scale (possibly wide band) fusion of the two images. These single scale approaches can be extended to a multi-scale fusion variant.

The single and multiple scale fusion methods discussed above include explicitly a spatial preprocessing or "enhancement" stage before fusion [Toet 1992; Waxman 1995; Ryan 1995]. The preprocessing may be needed for two reasons. Since the two sensors may have widely varying dynamic range and result in highly varying image histograms and contrast, some normalization is required before the bands can be fused into a common space. We will refer to this process as "normalization". In addition, the spatial preprocessing itself may be used to improve the visibility of desirable features in the image and thus serves as an image enhancement module. Waxman's processing is based on the model of early processing stages of the visual system. Toet [Toet 1992] achieved normalization by combining "contrast" rather than amplitude, which was shown by Peli [Peli 1990] to provide a form of enhancement. In particular, this tends to increase the visibility of low-contrast features in low-luminance areas of the image (in shadows). At the same time, as shown by Peli [Peli 1990], this approach reduces the visibility of details in high luminance areas. The latter problem can be resolved with the proper set of parameters applied to the Peli and Lim adaptive enhancement algorithm as was demonstrated in [Ryan 1995] at the pre-fusion processing stage. Note however that the Peli and Toet enhancements are parameters free, while the Peli-Lim and Scrofani approaches require specific tuning of the enhancement parameters.

The method of multi scale fusion by [Paval 1991] provides a combination of pixel selection from a single image and a weighting of the two/multiple images. In Paval's approach, the spectral region is divided into spatial sub bands of different

scales, and within each scale a combination rule is applied for the corresponding pixels taken from the two components. Even when the combination rule is a binary selection [Toet 1992], the combined fused image may have a combination of pixel values taken from the two components at various scales since it is taken at each scale. A combination rule that takes a weighted sum of the two pixel values has also been applied [Paval 1991].

Perconti et al [Perconti] have recently reported the result of task evaluation of various methods of two band fusion. They found that object recognition tasks (targeting) are performed best with the FLIR alone and with the Waxman color fusion method that incorporated the FLIR image, however the value of the color aspect was not clear. In evaluating the horizon perception task, no difference was found between the various imaging and presentation modes. In the geometric perspective task, which is important in navigation, the monochrome version of Waxman provided faster response. Scribner's color format that maintains the phase (polarity) for the visible image and separates it from the IR by color was found significantly better than Waxman color algorithm and not different than the Waxman monochrome version. None of the formats were found to be desirable for night helicopter pilotage from the short video segments. They concluded also that the color fusion might have its greatest utility as a targeting aid.

Ryan et al [Ryan 1995] reported on a comparison of FLIR only and Texas Instruments proprietary monochrome fusion algorithm (FLIR & intensified). In extensive field tests they found a substantial preference for the fused configuration for the specific pilotage maneuvers and for overall ranking. It should be noted that the study found no significant differences in any of the tasks and measures between the Lincoln Lab color and monochrome versions of the algorithm.

### 1.2 Multi-Spectral fusion process

Our fusion process for two bands is a two step process; monochrome fusion followed by pseudo chromatic color enhancement. The overall approach is illustrated in Figure 1.1.

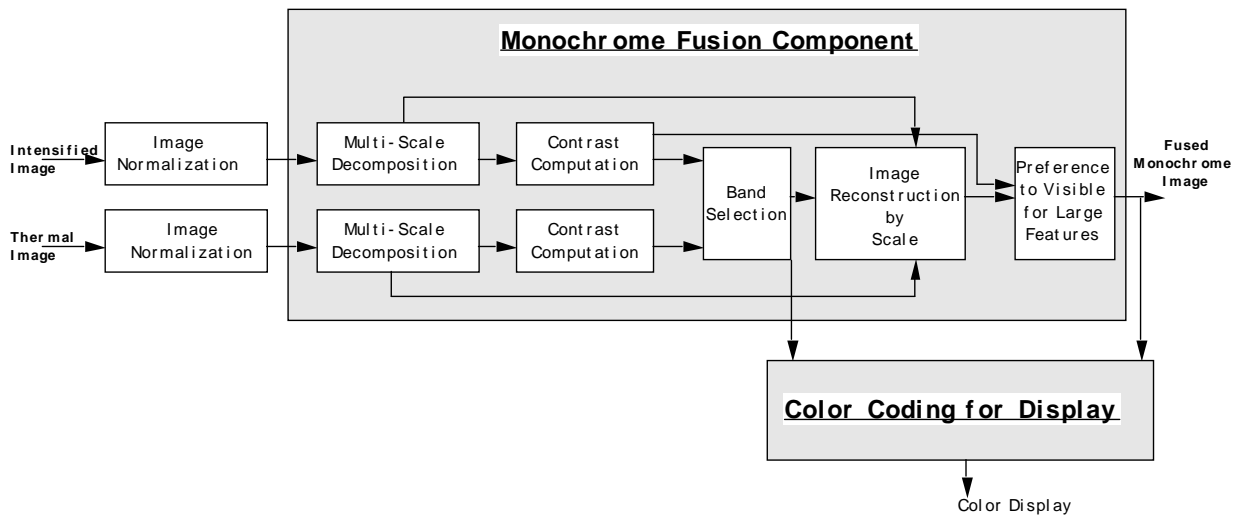


Figure 1.1: Fusion Process for Two Bands

The monochrome fusion has three key attributes

- 1) scale-by-scale fusion using oriented filters
- 2) decision based on contrast in each scale, fusion in amplitude domain
- 3) preference for visible band at least for large scales to preserve shape from shading information.

Points 2 and 3 above distinguish our process from others in the literature. The color-coding illustrated in Fig 1.1 is applied to the monochrome fused image. The color-coding aims to help the user identify the identity of the band that contributed to colored features. The color coding is not reported here.

## 2. METHODS

### 2.1 Imagery for algorithm development and evaluation

Two sources of imagery were used in the study; imagery provided by the Naval Post Graduate School (NPS) and existing collections of static multi-spectral imagery (collected by the Daedalus 1268 MS sensor). The sets of images obtained from NPS included static images and video segments.

### 2.2 Image normalization

The fusion process is applied to normalized imagery. The two bands may result in images with differing characteristics including dynamic range and contrast range. The goal of the normalization stage is to bring the multiple outputs from the different bands into a common framework. Since an enhancement may be simultaneously achieved, depending on the normalization scheme, the normalized/enhanced imagery has been used for fusion evaluation to assure critical evaluation of the fusion results separately from any other processing effect. Thus, the fused image was compared to the individual normalized inputs and not to the raw imagery.

It is possible to apply either global or local image normalization procedures. We applied a global histogram based normalization procedure. The normalization of images was accomplished by computing a global mean ( $m$ ) and standard deviation ( $\sigma$ ) for each image. The statistics ignore a percentage of the two tails of the histogram. Once these statistics are obtained, each pixel ( $i,j$ ) is normalized using the following equation:

$$p_o(i, j) = \frac{D}{2} + \left( \frac{2^n}{2n_o} \right) \left[ \frac{p_i(i, j) - m}{\sigma} \right]$$

where  $p_i$  is the input,  $p_o$  the output,  $D$  the maximum range of the data (for eight bits this would be 256),  $n$  the desired bits of dynamic range and the number of standard deviations to cover the dynamic range.

### 2.3 Image registration

The video images suffered from considerable misregistration between the two sources. There was a mismatch in both aspect and field of view. For the static images, registration mismatch was also observed; however, this was only in the form of translation. As might be expected, the accuracy of registration is key to successful image fusion, misalignment reduces the sharpness and contrast of the fused images. Using our alignment process described below, we were able to register these image pairs and achieved a substantial improvement in fused image quality.

Since the relative geometry between the two sensors was unknown, we warped the images using manual selection of tie-points. Common points were selected in both images and these coordinates were used to warp one image to the other (visible to IR). Selecting tie-points can be made difficult due to the mismatch in the image content and contrast differences, therefore an automatic scheme has been implemented to fine-register the images after the manual warping. To register the motion video imagery, a single set of manual tie-points was chosen and used to warp the visible to the IR.

The fine-registration process was only partly successful. This was quite apparent when viewing the output sequences in a movie format. One observes in the movie sequence a considerable amount of jitter due to the random misalignment of the visible image. These effects are unlikely to appear when the two sensors have the same field of view (as opposed to the large difference that existed in the set of video segment used in this study).

### 3. IMAGE DECOMPOSITION FOR MULTI-SCALE FUSION

We have implemented a multi scale image fusion algorithm. Each of the input images is divided into spatial sub bands of different scales. Within each scale a combination rule for the corresponding pixels taken from the multiple (one per sensor input) components is applied. However, we use a contrast measure to determine the combination rule and apply the rule to the amplitude image (in each scale). These changes offer higher sensitivity to visual image aspects and assure reconstruction of a single channel if the other channel has no signal or very weak signal. In addition, this approach permits contribution for each pixels to vary among the source images depending on scale, thus even when a binary combination rule is applied, each pixel of the combined fused image may have a contributions to its values from all sensors.

We have compared the use of isotropic and oriented multi-scale oriented filters in the multi-scale decomposition. We first calculated a set of isotropic filters to obtain band-pass amplitude scale representations. This kind of decomposition is similar to the image decomposition that takes place in the eye’s retina. We also applied four oriented filters that result in a set of scaled oriented images that are motivated by current cortical visual system models. In addition, we have tested the use of two oriented filters. The filters are designed to permit complete image reconstruction by simple sum of all the filtered oriented versions. Normalization of the amplitude signals by the local luminance completed the decomposition into visually relevant scale representations –local band limited contrast (Peli 1990).

#### 3.1 Filter implementation

Isotropic one- octave band-pass filters have been implemented along with a DC low-pass band and a high-pass band. One-dimensional representation of these filters for a is shown in Figure 3.1 on both a linear and log (base 2) frequency scale. For perfect reconstruction of the input images, filters sum to unity. Thus, a summation of all of the scales from single image decomposition reconstructs the original image. The equation used to generate the filters for the  $r^{th}$  spatial frequency at the  $i^{th}$  scale is:

$$G_i(r) = 0.5 [1 + \cos(\pi \log_2 r - \pi i)]$$

To implement the filters in a two dimensional plane,  $r$  must now represent the distance from the origin of the Fourier domain, or at pixel  $(f_x, f_y)$ . Figure 3.2 depicts filters for a choice of four orientations.

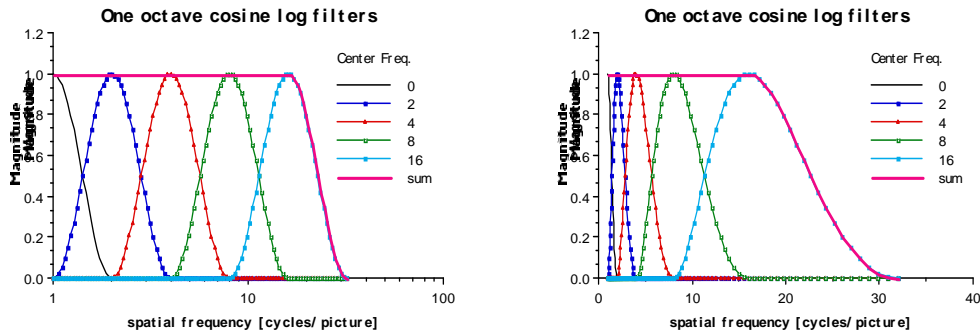


Figure 3.1: Isotropic filters response functions on linear (left) and log (right) scale

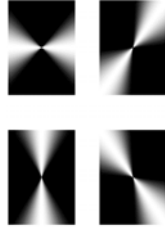


Figure 3.2: Oriented filters for four orientations

### 3.2 Contrast-based image fusion

The contrast based fusion algorithm calculates, for each of the oriented bandpass filtered versions of the images, a corresponding oriented band-pass contrast image similar to the isotropic contrast measure computed by Peli (1990). For each pixel at each scale  $i$ , for each spectral band (electromagnetic)  $j$ , and filter orientation  $l$ , the contrast measure is,

where  $a_{i,j,l}(x,y)$  is the band-pass filtered image in the  $i^{\text{th}}$  spatial frequency octave, the  $j^{\text{th}}$  orientation and the  $l^{\text{th}}$  band and  $l_{i,l}(x,y)$  is the lowpass filtered image which represents the local luminance mean.

The basic fusion process compares the calculated contrast measure for each pixel, at each scale and orientation, and selects the spectral band that should dominate these variables in the fused image. The simplest binary rule will select the component with the higher contrast. Once a contrast component  $c_{i,j,l}(x,y)$  is selected the corresponding amplitude component  $a_{i,j,l}(x,y)$  is added to the fused image. As can be seen if one spectral component has no signal, and thus no contrast or just very low contrast, this process will reconstruct the active spectral band image.

A number of variations of the basic algorithm were evaluated. The use of 4 orientations was found to provide the best performance (least artifacts) versus two orientations (horizontal and vertical) or isotropic filtering alone. It is noted that the gain from two orientations to four is not overwhelming. As a result, in analyzing the requirements for a real-time implementation, it is likely that the saving in processing only two orientations far outweighs any performance gain.

We implemented a strong preference for the intensified visible band at low scales representing large features. This was done under the assumption that large features are likely to be surface landmarks needed for pilotage and navigation. Thus, their correct interpretation of shape from is needed mostly for piloting tasks shading under stressful situations.

## 4. RESULTS - MONOCHROME FUSION

This section includes fusion results using both static and video imagery provided by NPS.

### 4.1 Results of fusion of static images

Figures 4.1 through 4.3 show the images intensified (II), IR and fused images for a sequence denoted by “up3”. Here, we can see that the object, marked in the output image, is transferred from the  $I^2$  image. Also, information from both images is used to provide better detail to the road and trees. The fused image maintains the shape from shading of the visual band for large piloting related objects (hills and valleys). This preference for the visual band does not preclude the use of high frequency small features from the IR.



Figure 4.1: up3  $I^2$



Figure 4.2: up3 IR



Figure 4.3: up3 Fused Image

Figures 4.1 through 4.3 show the images intensified (II), IR and fused images for the up3 image. The bright object, marked in the output image, is transferred from the II image, while the dark clouds next to it are from the IR image. Information from both images is used to provide better detail to the road and trees. The 3-D shape-from-shading information available in the II image and not in the IR image is preserved in the fused

## 4.2 Results of fusion of video images

Two sequences were focused on for the video fusion: named tow and pas. The biggest obstacle in processing these images was the large misalignment of the source images. For each image set (tow and pas), only one set of manual tie-points was used to warp all images (visible to IR), and then each image was fine-registered using our automated procedure. While the quality of image registration was acceptable, there were still misalignment problems. This became obvious when viewing a "movie" of the registered and fused images; quite a bit of image jitter was present in the registered visible image and as a result the jitter was noted also in the fused images.

One of the sequences (pas) presented a large component of scene motion from a low flying craft. The motion of the whole scene effectively masked the registration jitter providing a satisfactory presentation of the fused image sequence. The other sequence (tow) presented a relatively static scene. In this sequence the registration error jitter was very apparent both in the registered IR band displayed alone and in the fused sequence. The relative motion of the two components, however, provides a fortuitous method to illustrate the level of image fusion. With the relative motion between the components it is easy to note the image source of various displayed features. Because of the high correlation between the two images, it is sometimes hard to see the effect of the registration in a static image as provided here. However, when the two are jittered relative to one another the relation becomes immediately apparent. With the aid of the jitter we could see that the fusion of the two images is much more uniform across image features than could be determined from static frames. This effect may be used as a very effective evaluation tool using purposely induced jitter between perfectly aligned sensor images. It may even be used operationally if induced on command to aid the pilot in identification of image sources in the fused image. However, the color coding we are developing will provide these cues in a more convenient manner.



Figure 4.4: Tow000 normalized visible image (warped and fine-registered)



Figure 4.5: Tow000 normalized IR image

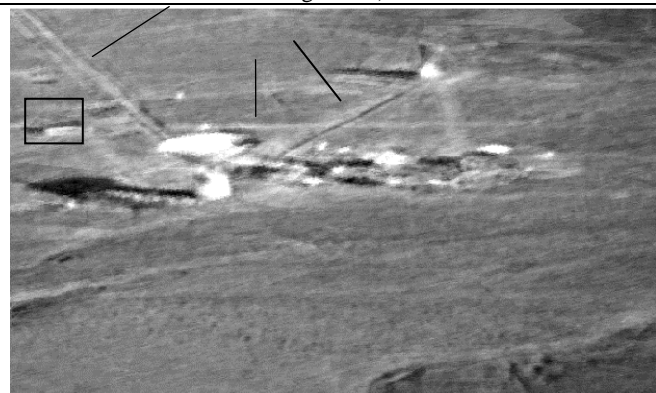


Figure 4.6: Tow000 fused output image

The visible, IR and fused images respectively for tow000. In the fused image, some features have been highlighted. In particular, a building that is not present at all in the visible image comes through from the IR, as well as some road features. In addition, we see a horizontal road, marked by a dashed line that comes through from the visible image. Also worth noting is the detail in the terrain in the lower half of the image. This portion takes the best qualities of both images, presenting the shape from shading of the visible band enhanced by the texture taken from the IR image





Figure 4.7: Pas299 normalized visible image (warped and fine-registered)



Figure 4.8: Pas299 normalized IR image

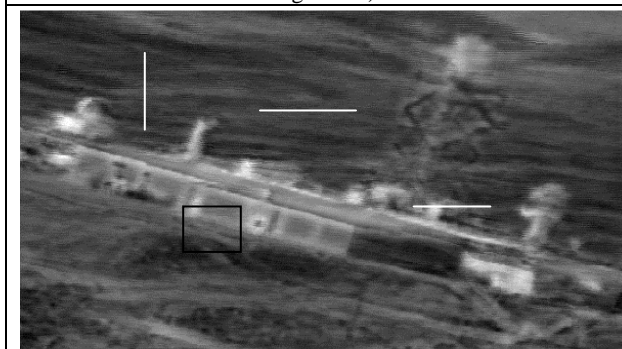


Figure 4.9: Pas299 output image

Images for pas299. In the fused image, we observe that the tower structure is coming through from the IR image, as well as window features in the building. As in the previous image, terrain in the lower half of the image acquires the best features of both images and provides an enhanced view over either.

## 5. SUMMARY

A robust multi-spectral fusion process that consists of two stages was developed. The monochrome fusion stage was shown to capture the details from all of its inputs across a diverse set of inputs using a single algorithm setting. The algorithm maintains the natural visual appearance of large scale terrain features which are necessary for navigation and pilotage and increase the visibility of small scale targets obtained from both bands. A color coding, which followed this fusion serve to highlight target and at the same time indicate to the pilot the source of the target and its polarity in the more sensitive sensor.

## ACKNOWLEDGEMENTS

This work was performed under the sponsorship of the U.S. Navy (ONR) under SBIR Contract No. N00014-97-C-0423. We wish to thank William K. Krebs of the Naval Postgraduate School (Monterey, CA) for providing us with a large set of imagery and his support through out the execution of the program. E. Peli was supported in part by NIH grants EY10285 and EY05957 and NASA contract NCC 2-1039 during the preparation of this manuscript.

## REFERENCES

- [AAEC 1997] "Practical AGC for Quasi-Invariant Target Signature," Fin. Rep. to Army/MICOM.
- [Braje 1997] Braje W.L. "Do shadows influence recognition of natural objects" (ARVO 1997) Investigative Ophthalmol. & Vis. Sci. Suppl. pg s1002.
- [Georgeson 1992] Georgeson M.A. "Human vision combines oriented filters to compute edges" Proc. R. Soc. Lond. B 235-245 1992.
- [Gur 1993] Gur M and Syrkin G Color enhances Mach Band s detection threshold and percieved brightness Vision Res 33; 2313-2319 1993
- [Jenkin 1997] Jenkin and Howard "Interpreting shape from shading under different visual frames and body orientation" (ARVO 1997) Investigative Ophthalmol. & Vis. Sci. Suppl. pg s1002.
- [Katz 1987] Katz et al, "Application of Spectral Filtering to Missile Detection Using Staring Sensors at MWIR Wavelengths", Proceedings of the IRIS Conf. on Targets, Backgrounds, and Discrimination, Feb. 1987
- [Morrone 1989], Morrone M.C. and D.C., "Discrimination of spatial phase in central and peripheral vision", *Vision Research* 29 : 433-445.
- [Morgan 1991] Morgan M.J., Ross J., and Hayes A. "The relative importance of local phase and local amplitude in patchwise image reconstruction" Biol. Cybern. 63:113-119, 1991.
- [Paval 1991] Paval M., Larimer J., and Ahumanda A. " Sensor fusion for synthetic vision" AIAA conference "Computing in Aerospace 8" Baltimore, MD, October 21-24, 1991.
- [Peli 1990] Peli E "Contrast in complex images," J. Opt. Soc. Am. A, 7: 2032-2040, 1990
- [Peli 1992] Peli E, "Display nonlinearity in digital image processing," Opt. Engineering 31; 1992.
- [Peli 1997c] Peli E (1997) In search of a contrast metric: matching the percieved contrast of Gabor patches at different phases and bandwidths. *Vision Research* In press.
- [Peli 1979] Peli, T., and Lim, J. S., "Adaptive Filtering For Image Enhancement," Proc. of ICASSP Conf., 1979.
- [Peli 1997d] Peli, T., Monsen, P., Stahl, R., and Pauli, M., Long Range Detection and Tracking of Missiles, IRIS/IRCM, May 1997
- [Peli 1997a] Peli, T., Young, M., and Ellis, K" Combining Linear and Nonlinear Processes for Multi-Spectral Material Detection/Identification," SPIE Conf. on Algorithms for Multispectral and Hyperspectral Imagery II, 1997 .
- [Peli 1997b] Peli, T., Young, M., and Ellis, K., "Multi-Spectral Change Detection," SPIE Conf. on Algorithms for Multispectral and Hyperspectral Imagery II, Orlando, FL, 1997.
- [Perconti] Perconti P., and Steele P.M. "Part Task Investigation of Multispectral Image Fusion using Gray Scale and Synthetic Color Night Vision Sensor Imagery for Helicopter Pilotage", undated report.
- [Ryan 1995] Ryan D., and Tinkler R. " Night pilotage assessment of image fusion" SPIE vol 2465, 1995.
- [Scribner 1993] Scribner D. A., Satyshur M.F., and Kruer M.R. "Composite Infrared color images and related processing" IRIS Targets, Back. and Discrim. January 1993, San Antonio TX.
- [Scrofani 1997] Scrofani J.W. "An adaptive method for the enhanced fusion of low-light visible and uncooled thermal infrared imagery" MSc thesis 1997
- [Stotts 1990] Stotts, L.B., Winter, E.M. and Reed, I.S., "Clutter Rejection Using Multispectral Processing," SPIE Aerospace Sensing, Conference 1305 (16-20 April) 1990, Orlando, FL.
- [Subramaniam 1997] Subramaniam and Biederman "Effect of Contrast Reversal on Object Recognition (ARVO 1997) Investigative Ophthalmol. & Vis. Sci. Suppl.
- [Toet 1992] Toet A " Multiscale contrast enhancement with application to image fusion" Optical Engineering 1992 vol 31 1026-1031.
- [Waxman 1995] Waxman M.A., Fay D.A., Gove A.N., Seibert M., Racamato J.P., Carrick J.E., and Savoye E.D. "Color Night Vision: Fusion of Intensified Visible and Thermal IR Imagery" SPIE Vol 2463, 58-68 1995.