# Model-based sensor fusion for aviation

Misha Pavel and Ravi K. Sharma

Department of Electrical Engineering
Oregon Graduate Institute of Science and Technology
P.O. Box 91000, Portland, OR 97291-1000, USA

## ABSTRACT

We describe a sensor fusion algorithm based on a set of simple assumptions about the relationship among the sensors. Under these assumptions we estimate the common signal in each sensor, and the optimal fusion is then approximated by a weighted sum of the common component in each sensor output at each pixel. We then examine a variety of techniques to map the sensor signals onto perceptual dimensions (e.g., color), such that the human operator can benefit from the enhanced fused image, and simultaneously, be able to identify the source of the information. We examine several color mapping schemes.

**Keywords:** sensor fusion, image fusion, color fusion, color vision, color mapping, multisensor fusion, enhanced vision

## 1. INTRODUCTION

The efficiency, robustness, and safety of many visually guided systems, such as flight control, can be improved by providing the operator with the necessary visual information at all times, even in low visibility conditions. One way to achieve this objective, is to use a suite of sensors, each specialized for different environmental conditions, e.g., TV, infrared (IR) camera or a millimeter wave sensor. To generate an appropriate flight control signal, the information from the different sources must be combined (fused). In general, the fusion can be performed by the system or by a human pilot.

In fact, the possible implementations of a fusion system range from a complete automation where a machine controls the flight to the opposite extreme where the information from each sensor is presented to the pilot on a separate display. Unfortunately, at this point in time, neither of these extreme solutions appears to be feasible. Currently available computer vision algorithms are not capable of converting reliably a sensor image output to flight control signals. The other extreme alternative using multiple displays would require the pilot to make complex and difficult cognitive and perceptual decisions during the most demanding phases of a flight. Empirical evidence suggests that humans are not able to perform these tasks well, and so the less than optimal performance together with the increased workload preclude this solution.

Therefore, to effectively deploy multiple sensors we must design a system which can either perform the fusion or can map the sensor information onto perceptual dimensions that are easily fused by the human visual system. In this paper we first describe a fusion algorithm and then examine several strategies for mapping the results onto color dimensions. We, therefore, utilize the capabilities of the human visual system, without increasing the workload.

## 2. BACKGROUND

The design of an effective fusion algorithm for image fusion is difficult because the optimality of a method for combining different images depends on the local relationship among the sensors (for more detail see Section 3. ). For example, fusion based on an additive combination that is optimal in some situations, can frequently result in lower signal-to-noise ratios than a simple local selection of the better sensor. Most of the existing fusion algorithms, therefore, employ some version of signal-dependent nonlinear processes.

Other author information: (Send correspondence to M.P.)
M.P.: Email: pavel@ee.ogi.edu; Telephone: 503-690-1155; Fax: 503-690-1406.
R.K.S.: Email: ravi@ee.ogi.edu; Telephone: 503-690-1121 x7523; Fax: 503-690-1406.

Although the existing methods differ in details, the assumption underlying most algorithms is that the larger gray values (or the larger gradients) represent more salient information. In these algorithms, images are represented by various multiresolution pyramids[2,1,7] and the fused images are constructed by choosing the largest value at each source location. The selection-based approach is optimal under the assumption that at each image location, one of the sensors provides the most useful information. These approaches, therefore, do not discriminate between signal and noise. More recently developed algorithms use sequences of images to estimate noise, and then fuse images optimally using statistical estimates[4,5] of the signals.

The attempt to produce a single gray-level fused image means that, in some situations, information might be lost. Recent attempts to reduce this information loss were based on using color to represent some of the missing information.[10,8] One approach proposed by Waxman[10] is to obtain a pseudo-color image with a natural-like appearance. This use of color may improve the appearance of an image, but does not necessarily deliver the sensor-specific information.

In a recent paper, Toet and Walraven[8] used color to represent the source of information. The shortcoming of their approach applied to visual guidance is that they do not distinguish between reliable signal and noise. We extend their effort using a principled, statistics-based approach to signal analysis and mapping. First we examine the relationships among sensors, the principles that underly the mapping between sensors and display dimensions, and the potential benefits of using such mappings. We then illustrate the general approach using several possible mappings between sensor signals and color. We attempt to use human color vision to fuse signals from different sensors.

## 3. FUSION PROBLEM ANALYSIS

The benefits derived from fusion of images depend on the relationship among sensors and on the manner in which these can be exploited. Since the different sensors are based on different physical phenomena, the relationship among their outputs can be complex. This is especially the case for sensors that use different bands of the electromagnetic spectrum such as Low Light Television (LLTV), Infrared (IR) camera, and millimeter wave radar (MMWR). For the purpose of our analysis we consider two sensors that differ mostly in their spectral sensitivity such as LLTV and IR cameras.

### 3.1. Sensor relationships

Because of the specific phenomenology of individual sensors, different types of sensors generate distinct images of the same scene. Quite often, the images may contain different information or features. For example, the polarity of local contrast changes can be arbitrary between LLTV and IR sensors (local polarity reversals), and a MMWR image may contain significant features that are absent in the LLTV image (inconsistent or complementary features). In the case of local polarity reversals the same information is present in both images, but with reversed contrast. However, in the case of unique or complementary features, there is different information in the images. In addition, the information contained in the sensor images may be corrupted by noise.

These relationships between signals from two sensors can fall into one of the following categories:

1. Common - information that is common to both sensors. A combination of the common components can increase the resulting signal-to-noise ratio.
2. Complementary - information that is unique for a sensor and that can be used in detection, identification and guidance such as temperature of aircraft engines.
3. Irrelevant - complementary (unique) information that is irrelevant and could even interfere with the pilot's tasks. For example, shadows or patches of a runway that was heated by already departed aircraft.
4. Noise - Randomly varying signals, that do not contain any information about the scene being sensed.

To maximize the benefits from the suite of sensors, a fusion algorithm should enhance the common signal, display useful complementary information, and suppress or eliminate noise and irrelevant information. In order to do that we would need to estimate each of these components in each sensor. One approach, presented in the next section, is to construct a model that includes some of the components on our list.

### 3.2. Sensor model

The approach taken in our prior work[5] was based on the assumption that the relationship among sensors can be represented by a local affine transformation at each pixel. The output $a_i$ of the $i$th sensor at pixel location $\vec{x}$ is assumed to be given by:

$$a_i(\vec{x}) = \beta_i(\vec{x})s(\vec{x}) + \alpha_i(\vec{x}) + \epsilon_i(\vec{x}), \tag{1}$$

where $s$ is the common signal, $\alpha$ is the sensor-specific (complementary) signal, $\beta$ is the gain and $\epsilon$ is a random process representing noise. The coefficients of the affine transformation $\beta$ and $\alpha$ are estimated by computing correlations over small neighborhoods of the images. The variance due to the noise at each location, $\sigma_i^2$, is estimated over frames. Note that this sensor model is not linear because the coefficients of the affine transformation depend locally on the sensor images.

### 3.3. Fusion process

Using the description in Equation (1) together with the assumption that $\epsilon_i$ are independent normally distributed random variables, we showed[5] that the most probable estimate $\hat{s}$ of the common component $s$ was a weighted combination of the form,

$$\hat{s} = \frac{\sum_i \frac{\beta_i(a_i - \alpha_i)}{\sigma_i^2} + \frac{s_0}{\sigma_0^2}}{\sum_i \left(\frac{\beta_i}{\sigma_i}\right)^2 + \frac{1}{\sigma_0^2}} = \sum_i w_i(a_i - \alpha_i) + w_0 s_0, \tag{2}$$

where $s_0$ and $\sigma_0^2$ represent the mean and variance of the apriori distribution of $s$, which we assumed to be normal. For the sake of clarity we simplified the notation by dropping the reference to the pixel location.

The relationship between $\hat{s}$ and the desired image to be displayed depends on the interpretation of the common signal $s$ in Equation (1). We can interpret the common component, $s$, in at least two ways, that are closely related to the way that we estimate the coefficients of the affine sensor model. First, if we use a rendered image from a database that represents the *ideal image* to estimate the correlation underlying the coefficients in Equation (1), then $\hat{s}$ represents a direct estimate of the desired image. On the other hand, if we compute the model coefficients by correlating the images of the two sensors, then $\hat{s}$ represents the common component of the two sensors. In that case, $\hat{s}$ might contain little useful signal, and most of the task-relevant information would be in the sensor-specific complementary components.

In either case the difficult problem is to determine (1) the importance of the complementary features and (2) a way to incorporate them in the resulting fused image. A typical approach in existing fusion algorithms is to select one of the complementary feature values $\alpha_1$ or $\alpha_2$ (e.g., the larger one) to construct the fused image. In our original approach we selected the feature with the greatest signal-to-noise ratio as the "useful" complementary component and ignored the other, "irrelevant" one. Unfortunately, signal-to-noise ratio may not always be a reliable discriminant of useful or important features.

## 4. MAPPING

In the absence of a reliable automatic way to distinguish between useful and irrelevant complementary components we would like to find a way to rely on the pilots visual system as much as possible. In particular we have been seeking techniques to map the sensor outputs onto perceptual dimensions such that the pilot could easily perceive the common component and simultaneously assess the usefulness of each complementary component from each source.

### 4.1. Mapping principles

The objective of the fusion process then is to analyze signals into its components as shown in Equation (1) and map them onto perceptual dimensions that would maximally enhance pilots' performance. In our preliminary analysis we arrived at the following set of principles that should underly the choice of mapping.

1. Matching spatial resolution of the sensors with the spatial resolution of the corresponding perceptual dimension. Thus a sensor that has low spatial resolution would be mapped onto a dimension that has low spatial resolution.

2. Matching temporal resolution of the sensors with the temporal properties of the corresponding perceptual dimension. Thus a sensor with slower response or frame rate would be mapped onto a sluggish perceptual dimension.
3. Separability of dimensions — assure that the dimensions for the mapping of the complementary components are perceptually separable and identifiable.
4. Integrability of dimensions — assure that the dimensions of the mapping of the common signal are integrable.
5. Cross-dimensional masking — assure that the inter-dimension masking would not impair the perceptibility of the complementary features.

Implementation of any mapping consistent with a set principles such as the above set requires a model of the pilot's visual system. Although there are several other possibilities we illustrate this principle-based approach using the properties of human color vision.

### 4.2. Color vision

Human color vision is generally accepted to be three-dimensional and, although the details may be subject to controversy, there are several models that seem to be good first order approximations. Because of the three-dimensional nature of human color perception, the color of any light can be represented by a projection of the intensity vs wavelength distribution onto three "primary" color dimensions, for example, $(R, G, B)$. It is important to note that, in general, the color standard $R, G, B$ values do not correspond to the signals driving the guns of an arbitrary color monitor. In fact, the representations discussed below must be modified by calibrating them for the particular monitor used to display the images.

An alternate model of color perception, is the *opponent* color perception model. In this model, the trichromatic dimensions are recoded into one luminance (achromatic) signal and two chromatic signals. Using the notation from vision literature where $S$, $M$, and $L$ correspond to short, medium and long wave cones, the luminance signal $P_Y$, and the two chromatic components are computed as

$$\begin{aligned} P_Y &= L + M, \\ P_{RG} &= L - M, \\ P_{YB} &= L + M - S. \end{aligned} \tag{3}$$

The two chromatic signals represent the "red-green", $P_{RG}$, opponent dimension, and the "yellow-blue", $P_{YB}$, opponent dimension.

In the subsequent discussion, we assume that the cone sensitivities are well described by the functions reported by Smith and Pokorny.[6] Although the transformation between the psychophysically determined color sensitivities of the cones and the standard representation such as $X, Y, Z$ or $R, G, B$ is not perfectly linear, for the sake of simplicity, we will illustrate our approach using a linear approximation. Any of our approaches can refined by introducing transformations for the uniform color spaces.[3]

## 5. EXAMPLES OF COLOR MAPPING

Based on the three-dimensional nature of color representation we consider several possible mappings of the sensor outputs onto various transformations of the color dimensions.

### 5.1. Red-green map

Since the visual system is already fusing the signals from the three types of cones we start with a direct map between a sensor and a cone type. In that case, the additive representation of the luminance component shown in Equation (3) in combination with the red-green opponency suggests that a simple mapping of one sensor onto the $M$ channel and the other onto the $L$ channel would yield the desired fusion effects.

To implement the mapping we first select the gain of each channel in accordance with the optimal fusion Equation (2). Thus at each pixel we let

$$L = w_1 a_1,$$

$$M = w_2 a_2,$$

where the weights $w_i$ are inversely proportional to the variances of the sensor signal to be consistent with those in Equation (2). In order to compute the excitation for each of the three guns of the monitor it is necessary to estimate the calibration matrix for this specific monitor.[9] An example of the fusion process is illustrated using two synthetic images with the resulting fused image shown in Figure 1(c).

This direct mapping approach has at least two drawbacks that are illustrated in Figures 1(a)-1(c). First, polarity reversals can result in incorrect conclusion about the source of the information. The source identification problem is illustrated with two images that contain the same information except for local polarity reversals, as shown in Figures 1(a) and 1(b). The IR image is represented by the "red" channel ($L$) and the TV image by the "green" channel ($M$). The combination of the dark image of the runway from the TV camera with the bright runway from the IR source results in the runway appearing red in the fused image. The red appearance of the runway implies incorrectly that the runway image is from the IR sensor only. This is misleading since the same information is present in both images, albeit polarity reversed. Second, in addition to the attribution error, polarity reversals can result in reduction in detectability of edges or their motion because chromatic edges appear less salient than those defined by luminance. In general, both of these difficulties are due to the fact that the $M$ and $L$ signals form an opponent color direction and are not perceived independently of each other. To rectify the problem we examined alternate mappings of sensors to color directions that would be mutually independent and orthogonal.

### 5.2. Fusion through color mapping in YIQ space

One convenient color space for an alternate mapping is the YIQ space that was developed for commercial color television broadcast. The YIQ color space is a transformation of the RGB color space such that the luminance ($Y$) information is decoupled from the chromatic information ($I$ and $Q$). In particular YIQ color space is derived from the RGB by computing the luminance and subtracting it from the color signals. The chromatic remainder components ($U$,$V$) are then rotated by 33° to maximize the difference in spatial resolution. Consequently, the $I$ component requires higher spatial resolution than does the $Q$ component. In the NTSC color signal the $I$ component gets approximately twice as much bandwidth as does the $Q$ component.

We now describe our implementation of the optimal fusion algorithm in accordance with the principles described in Section 4.for two sensors $a_1$ and $a_2$.

First, we compute zero-mean images $\hat{a}_1$ and $\hat{a}_2$,

$$\begin{aligned} \hat{a}_1 &= a_1 - \overline{a}_1, \\ \hat{a}_2 &= a_2 - \overline{a}_2, \end{aligned} \tag{4}$$

where $\overline{a}_1$ and $\overline{a}_2$ are the averages computed over the entire image of each sensor. We then compute the luminance $Y$ and the chromatic components $I$ and $Q$ by

$$\begin{aligned} Y &= \mid \hat{a}_1 \mid + \mid \hat{a}_2 \mid \\ I &= \hat{a}_1 \\ Q &= \hat{a}_2. \end{aligned} \tag{5}$$

The YIQ representation is converted to the monitor excitations signals by first converting YIQ to RGB space by

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.0000 & 0.9562 & 0.6214 \\ 1.0000 & -0.2727 & -0.6468 \\ 1.0000 & -1.1037 & 1.7006 \end{bmatrix} \begin{bmatrix} Y \\ I \\ Q \end{bmatrix} \tag{6}$$

The final step in the computation is a transformation by the calibration matrix corresponding to the particular monitor to be used for the display of the fused images.

In practice we need to scale the resulting monitor excitation values $R$, $G$ and $B$ to maximize the variance, but to limit the ranges to the integers between 0 and 255. To provide an intuitive insight we demonstrate this mapping using two test images composed of horizontal and vertical gray-level ramps shown in Figure 2. The image in Figure 2(a) is a horizontal graylevel ramp with the pixel values linearly increasing from 0 to 255 from left to right representing the first sensor, $a_1$. The image in Figure 2(b) is a vertical ramp with the values increasing linearly from 0 to 255 from bottom to top representing the second sensor, $a_2$. Figure 2(c) is the corresponding fused image when the YIQ scheme is applied to the ramps.

The source of the information can be identified by associating the different colored regions of the fused image with particular combinations of sensors. For instance, in the region where both $a_1$ and $a_2$ have high intensity, the fused image is rendered in magenta, whereas, when the intensity of $a_1$ is high and that of $a_2$ is low, the resulting color is yellow. Figure 3 shows an example of the YIQ scheme applied to images from medium wave and long wave infrared sensors. Although we have achieved a degree of decoupling between the sensor outputs, the operator, e.g, pilot, must learn the relationship between the color and the combination of sensors. To alleviate this problem, we explored another mapping described in the next section.

### 5.3. Mapping of complementary features

For the mapping in this section we assume that the desired image is closely approximated by one sensor, $a_1$, and, that the relationship between the two sensor images $a_1$ and $a_2$ can be described by a locally (pixel-by-pixel) affine transformation,

$$a_1(\vec{x}) = \beta(\vec{x})a_2(\vec{x}) + \alpha_2(\vec{x}) + \epsilon(\vec{x}), \tag{7}$$

similar to that in Equation (2) in Section 3.. The parameters $\beta$ and $\alpha$ are estimated using standard linear regression techniques and computing correlation over small regions (5 x 5) of the two images. The intercept $\alpha_2$ then represents complementary features from the second sensor.

We let the luminance channel to be equal to the common component $\hat{s}$ computed in accordance with the fusion algorithm in Equation (2),

$$Y(\vec{x}) = \hat{s}(\vec{x})$$

and assign the complementary information to

$$I(\vec{x}) = w(\vec{x})\alpha_2(\vec{x}),$$

where $w$ is inversely proportional to the variance of $\alpha_2(\vec{x})$. To generate a display we convert YIQ to the monitor RGB as described in the previous sections using Equation 6. An example of a fused image using this approach is shown in Figure 4. This image is colored in the regions containing the complementary features.

## 6. CONCLUSION

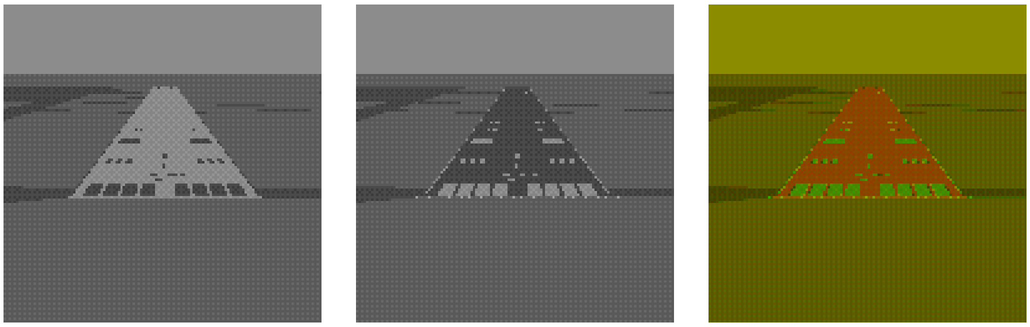
We have demonstrated the feasibility of fusing images from two different sensors using simple assumptions about the relationship among the sensors. Because a fused image hides the source of the information we explored the possibility to provide this hidden information to the observer using color coding as an example of a perceptual code. We were able to use color mappings to generate images in which the common signal was represented by luminance and the sources of information by various color directions. The main difficulty was to find a mapping such that the observer can easily identify the different sensor contributions. We were able to achieve a good separation using the YIQ space, but the observer must learn the mapping. This was also true for another mapping not described here where we used orthogonal color directions and mapped each component onto these orthogonal directions. Finding an optimal mapping would require perceptual experiments on identifying color directions. Although these approaches appear to be promising, empirical studies of pilots' performance using the various mapping schemes is necessary in order to evaluate any of these and similar techniques.

## 7. ACKNOWLEDGMENTS

This work was supported by a grant to the Oregon Graduate Institute from NASA Ames Research Center, NCC2-811.

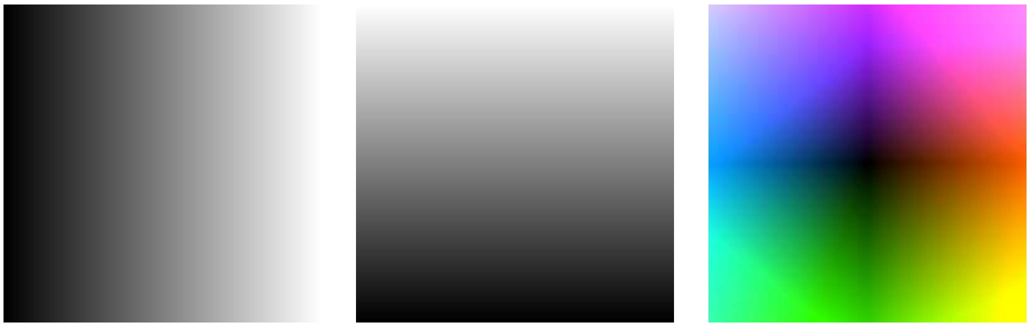
## REFERENCES


1. P. Burt. A gradient pyramid basis for pattern-selective image fusion. In *Proceedings of the Society for Information Display*, pages 467–470. SPIE, 1992.
2. P. J. Burt and R. J. Kolczynski. Enhanced image capture through fusion. In *Fourth Int. Conf. on Computer Vision*, pages 173–182. IEEE Comp. Soc., 1993.
3. *CIE Recommendations on uniform color spaces, color differenve equations, psychometric color terms. Supplement No. 2 to CIE publication No. 15 (E-1.3.1) 1971/(TC-1.3.)*. International Commission on Illumination, 1978.
4. M. Pavel, J. Larimer, and A. Ahumada. Sensor fusion for synthetic vision. In *Proceedings of the Society for Information Display*, pages 475–478. SPIE, 1992.
5. R. Sharma and M. Pavel. Adaptive and statistical image fusion. In *SID International Symposium Digest of Technical Papers*, pages 969–972. Society for Information Display, 1996.
6. V. C. Smith and J. Pokorny. Spectral sensitivity of color-blind observers and the cone photopigments. *Vision Research*, 12:2059–2071, 1972.
7. A. Toet. Hierarchical image fusion. *Machine Vision and Applications*, 3:1–11, 1990.
8. A. Toet and J. Walraven. New false color mapping for image fusion. *Optical Engineering*, 35(3):650–658, 1996.
9. B. A. Wandell. *Foundations of Vision*. Sinauer Associates, 1995.
10. A. M. Waxman, D. A. Fay, A. Gove, M. Seibert, J. P. Racamato, J. E. Carrick, and E. D. Savoye. Color night vision: Fusion of intensified visible and thermal imagery. In *Proceedings of SPIE, Vol. 2463*, pages 58–68. SPIE, 1996.



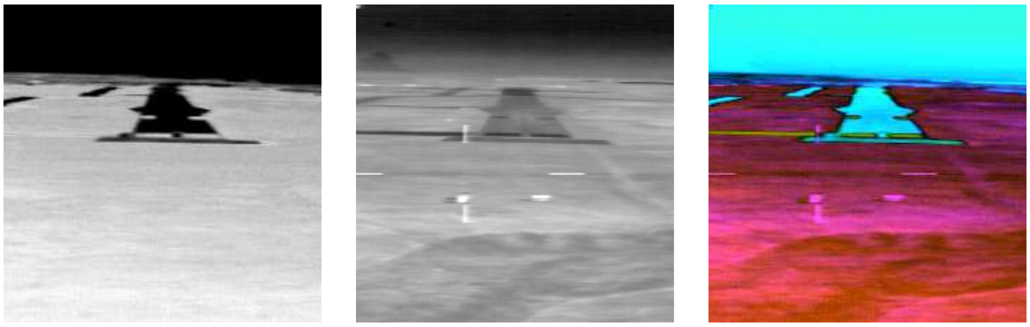
(a) IR image (b) Visual image (c) Fused image

**Figure 1.** Simple color scheme for fusion (simulated images)



(a) Horizontal ramp (b) Vertical ramp (c) Fused image

**Figure 2.** YIQ scheme for fusion



(a) MWIR image (b) LWIR image (c) Fused image

**Figure 3.** YIQ scheme applied to real images (Data from FLIR Systems Inc.)

**Figure 4.** Representing complementary signal using color (Data from FLIR Systems Inc.)