IMAGE FUSION

M. Pavel[1]
Western Aerospace Laboratories
and
NASA Ames Research Center

1 Motivation

In order for a pilot to fly an airplane, she or he must combine information from a large number of different sources. Useful information for this purpose may be available as readouts from avionics instruments, symbology on a HUD, or from the image of an airport scene seen through a window. The workload of the pilot is frequently increased as the number of sources of information and the complexity of the data increases. Because humans do not necessarily combine information optimally, effective automatic combination of the data may lower the load and thereby free the pilot to be ready if necessary to make critical decisions. The combined data are frequently more useful because the combination may reduce variability, or use complementary information from the different sources.

It is interesting to note that fusion of information is a common process in both natural and machine vision. Consider these examples of fusion:

1. Combining images obtained from different locations, e.g., binocular stereopsis.

2. Combining images obtained from different sources — flight instruments and an image of a scene.

3. Combining information from one source over time, i.e., temporal filtering.

4. Combining information from one source over space, i.e., spatial filtering.
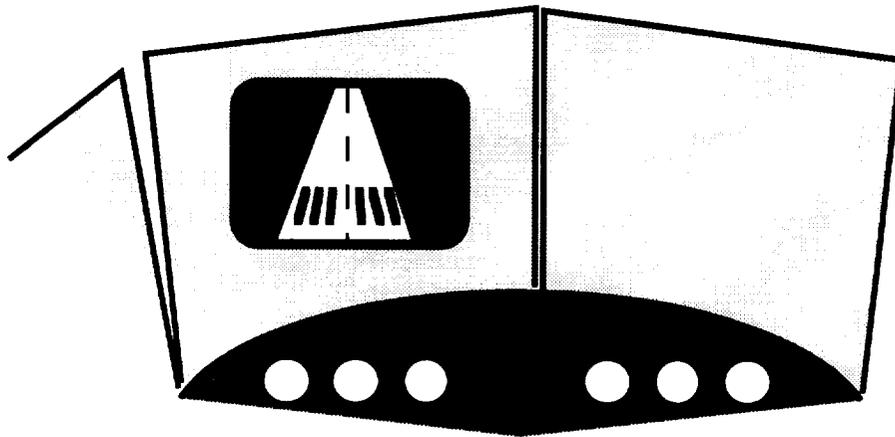

Figure 1: Schematic representation of the HUD arrangement.

These considerations are among those motivating the development of systems that augment the traditional display system. One approach, schematically depicted in Figure 1, illustrates one possible implementation of the AVID system.

2 System Overview

Figure 2 illustrates the basic components of a system designed to improve the ability of a pilot to fly through low-visibility conditions such as fog.

The underlying principle is based on the fact that atmospheric attenuation is greatly reduced for millimeter waves (MMW) relative to the radiation in the visible spectrum. In the proposed system the information (images) from sensors operating in the MMW regime are combined with other information such as a global positioning system (GPS) and a stored database. The fusion process is necessary because the spatial and temporal resolution of the MMW sensors is greatly limited.
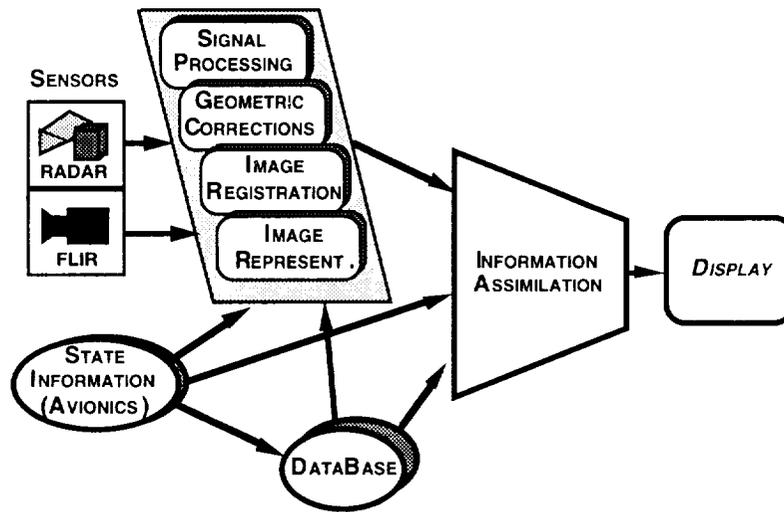
Figure 2: Diagram of ESAS.

2.1 Role of Visual Sciences

A successful design of a system such as the one illustrated in Figure 2 requires a combination of expertise ranging from radar engineering to human factors and psychology.

Life sciences are critical for the development and design of such a system in at least three ways. First, knowledge of the visual system must be used to optimize the design of displays used by the pilot in all phases of flight operations. Second, understanding the human visual information processing can guide the development of solutions to many system design problems. For example, biological fusion may be used in the process of reverse engineering to guide the design of fusion algorithms. Finally, psychology of measurement, combined with the models of the visual system, can be used to develop methodology for evaluation of the complete system.

It is also important to note that the solution of the particular problems associated with AVID gives rise to questions whose answers will enhance our basic understanding of the human visual system. For example, displaying information *on* a HUD without impairing significantly the information viewed *through* the HUD requires a good understanding of perception of transparent images. Although recent results[2] provide useful information for the

designer, additional basic research is required to develop a model of transparency perception.

2.2 Fusion Issues

The first prerequisite for a successful design and evaluation of fusion algorithms is a definition of a goal specified in terms of desired images and an objective function. The ultimate desired image is one that contains all necessary information for flight control. To achieve (or to approximate) this goal requires a convenient representation of data, optimal fusion algorithms, and a effective display of the resulting images. System evaluation can be performed by comparing the obtain image to the desired one with respect to the objective function.

Unfortunately, our knowledge to date is not sufficiently complete to specify a unique desired image and an objective function. Rather, we define a gray-level image s(x,t) to be an image that would be obtained under uniform illumination with unlimited visibility. Using simulator test results, one can easily demonstrate that this image is sufficient, but not necessary, for a pilot to land an airplane.

3 Sources of Information

There are many sources of information that could be used to support the functions of the enhanced situational awareness. For the purpose of this project, we consider the following sources of information:

- High resolution sensors of visible spectrum (Video)

- High resolution sensors of infrared spectrum (IR)

- Low resolution millimeter wave sensors (Radar, PMMW)

- Terrain database

- Inertial navigation system (INS)

- Global positioning system (GPS)

3.1 Sensor Characterization

Effective fusion of information from different sources requires the comprehensive characterization of the sources. The following is a list of sensor characteristics that are important in the design of image processing and fusion algorithms.

3.1.1 Signal Characteristics

These characteristics describe the properties of the signals generated by the sensor:

- Spatial and temporal transfer functions

- Sensitivity

- Relationship between visual and sensor images

- Noise, drift, changes in gain

- Atmospheric attenuation

- Temporal sampling / dynamics

- Inhomogeneity of sensor image

3.1.2 Geometric Properties

Knowledge of the imaging geometry of the sensor is critical in order to generate conformal images from different sources. In addition to the imaging geometry of each sensor, its location and orientation is also critical. These effects are illustrated in Figure 3. Geometric corrections to compensate for the variety of geometric distortions can be implemented, for most sensors, by simple transformations. One notable exception is an active radar which requires special considerations.
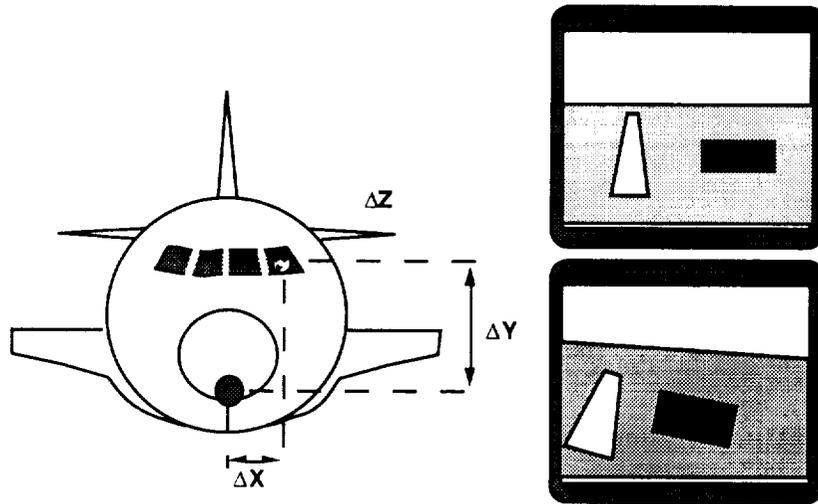
197

Figure 3: Diagram of geometric distortions due to sensor viewpoint placement

3.1.3 Imaging Radar Distortions

Radar is an active device that illuminates a scene, detects reflections, estimates delays associated witn the reflections, and thereby estimates the distances of the reflecting objects. Since a radar measures ranges (b-scope representation), a geometric transformation is necessary to convert the range image to a perspective projection of the scene (c-scope image). As shown in Figure 4, this transformation is, unfortunately, underconstrained because measured distances do not specify position uniquely.

A typical solution, used to regularize this problem, is to assume that all reflections are from objects located on the surface of flat earth. Of course the flat-earth assumption results in errors whenever the actual reflections are generated by objects at some vertical distance from the earth surface (Figure 4).

Recently we have been able to demonstrate a theoretical approach to reduce the problem by eliminating the flat earth assumption. The computational method is based on integrating information from multiple frames of b-scope images. We are currently examining the practical implications of these theoretical efforts.
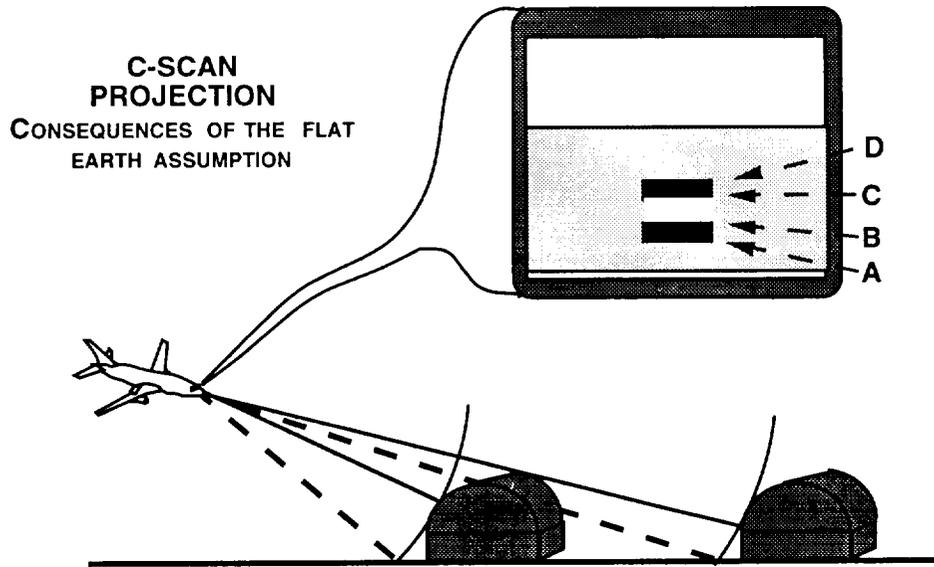
198

Figure 4: An illustration of the effects of flat-earth assumption in the rectification of returns from two elevated structures.

3.2 Simplified Sensor Model

Under the assumption that it is possible to correct all geometric distortions in images obtained from a sensor, the output of the sensor can be approximated by

$$m\left(\vec{x}\right) = h * \left\{a\left[r\left(\vec{x}\right)\right]\ b\left(\vec{x}\right)\ s\left(\vec{x}\right) + n_m\left(\vec{x}\right)\right\} \tag{1}$$

where m is the sensor image
$\vec{x}$ image coordinates
h spatial impulse response
a atmospheric attenuation
r range (distance) from sensor to an object
b sensor-to-visual factor
s objective image
n_m noise

199

3.3 Database

The database (DB) consists of the best available information (model) of the landing terrain. The database includes the airport, the runway, and some surrounding stationary objects. The models of the objects are represented in terms of polygons. The geometric model of the terrain includes color information and it is rendered by the geometry engine of a graphics workstation, such as the Silicon Graphics Inc. (SGI) machine.

When the rendered scene is converted to a gray-level representation of the landing scenario, the resulting image can be approximated by:

$$d(\vec{x}) = [1 - c(\vec{x})] s(\vec{x}) + c(\vec{x}) g(\vec{x}) + n_d(\vec{x}) \tag{2}$$

where
d computer generated image obtained from the DB
c obstacle indicator function
s objective image
g obstacle image
n_d] noise, quantification of DB inaccuracy.

In this simple model, the difference between a real image of the scene and the DB rendering is expressed by the noise term in equation (2).

4 Image Processing

Prior to fusion, information from each sensor is processed by algorithms specialized for that sensor. These algorithms are designed for:

1. Noise reduction: Linear and non-linear filtering

2. Image enhancement: Histogram equalization, edge enhancement.

3. Uncertainty (Noise) Estimation: Estimation of variability and consitency within and across sources.

4. Prediction: Recursive estimation of expected and observed image.

5 Image Fusion

There are many ways to combine information from different sources. The optimal technique to be selected depends on prior knowledge of the signal characteristics, the objective, and the required robustness. The following is a list of examples of candidate techniques:

1. Additive, linear combination

2. Selection (1/0)

3. Additive, nonlinear combination

4. Bayesian update of information

I will first discuss briefly the first two techniques which have been considered by several investigators [1, 3].

5.1 Linear Additive Combination

Linear additive rule is a pixel by pixel combination of two sources that can be expressed by

$$\langle s\left(\vec{x}\right)\rangle = \alpha\, d\left(\vec{x}\right) + \beta\, m\left(\vec{x}\right).$$

There are several reasons why a linear additive combination is particularly important. First, additive combination is an optimal rule when the individual sources can be characterized by normal distributions. Second, additive combination is easily implemented in real-time hardware. Finally, additive combination occurs naturally when an image is displayed on a HUD.

5.2 Disadvantages of Additive Fusion

There are several shortcomings of the simple linear additive approach:

Obstacle Detection: Whenever information is present in one, but not in the other image, the fused signal-to-noise ratio is lower than that in the original image with the signal.
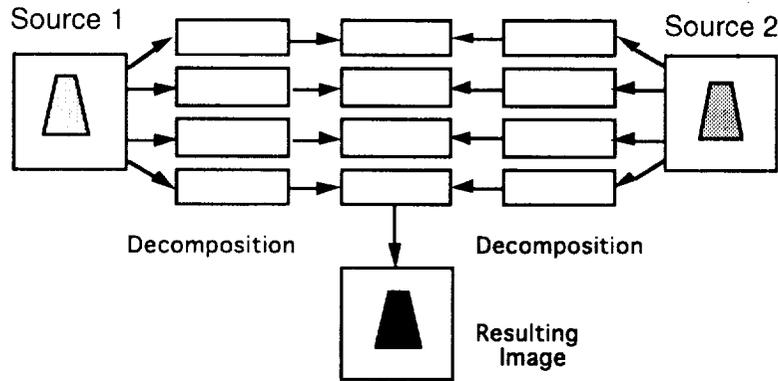
201

Figure 5: A diagram of fusion by components.

Polarity Changes: The relationship between the polarity of two images may vary for different locations and may depend on environmental conditions.

Spatial Frequency: Signal-to-noise ratio may vary for different spatial frequency bands and different spatial locations.

Because of these shortcomings of the linear additive rule, we consider more complex, nonlinear rules.

5.3   Fusion by Components

One approach that can be used to remedy the disadvantages of the linear additive rule is to decompose each image into components and then perform the combination by combination rules specific to the components. This general approach is shown in Figure 5.

Depending on the specific application, there are numerous ways of decomposing images into components. Multiresolution representation of images is one way of decomposing images into its components.

5.4   Multiresolution Representation

A typical multiresolution representation can be thought of as a decomposition of an image into a set of spatial frequency bands as illustrated in Figure 6.
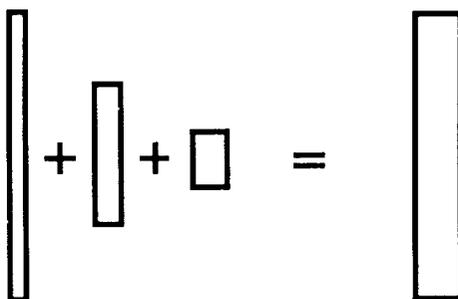
Figure 6: Illustration of a pyramidal representation.

The size of the blocks in the diagram in Figure 6 indicates that the lower spatial resolution bands require fewer samples.

One way to construct such representation consists of recursive applications of the following steps:

1. low-pass filter,

2. subsample,

3. interpolate,

4. compute difference between two adjacent levels, until the representation reduces to a single sample.

In this particular multiresolution representation, each resolution level is insensitive to local orientation of features. There are other schemas for the decomposition such that the information at each resolution level is further decomposed to several subimages, one for each of a set of diretions [1, 4].

Given the multiresolution representation, there are many alternative ways to fuse the images.

5.5 Sample Selection

One way to fuse two images consists of examining each pixel in both images at each level, and selecting the pixel with a particular property. For example, one can select the pixel with the greater gray level value [1]. Alternatively, it is possible to compute contrast at each level and select the pixel with

203

greater contrast value [3]. Although these methods have been shown to be successful they do not eliminate all the problems listed in Section 5.2. We are, therefore, considering a more general, statistical approach to fusion.

5.6 Optimal Fusion Approach

The goal of the optimal fusion approach is to use the best models of the sources together with the desired image and determine the combination that minimizes the difference between the fused and the desired images. Although there are questions concerning the particular metric to be used for the measurement of the difference, our initial development is based on maximizing aposteriory probability.

This approach requires either prior knowledge or on-line estimation of the variability of the sensor images. Limited spatial resolution and the physical phenomena underlying some sensors, e.g., MMW radar, results in spatial correlation that can be utilized in fusion.

Our current approach consists of the following steps:

1. Compute multiresolution pyramid for each image.

2. Predict image from the database.

3. Predict image from prior frames.

4. Estimate the variances at each pixel $\vec{x}$ at each level l.

5. Estimate correlation with the expected image from the database.

6. Combine pixels using optimal weights for each pixel and each level.

To the extent that the underlying assumptions are valid, this approach determines statistically optimal fused images. In addition, this statistically-based approach can be used directly to identify specific features of interest, for example, unexpected obstacles or runway incursions.

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] J. B. Mulligan. Nonlinear combination rules and the perception of motion transparency. *Vision Research*, 33:2021–2030, 1993.

[3] A. Toet. Hierarchical image fusion. *Machine Vision and Applications*, 3:1–11, 1990.

[4] A. B. Watson. Efficiency of a human image code. *Journal of the Optical Society of America A*, 4:2401–2417, 1987.