7-5-1995

Pixel-registered image fusion

Rufus H. Cofer

Samuel Peter Kozaitis

Pixel-registered image fusion

Rufus H. Cofer
Samuel Peter Kozaitis

PIXEL-REGISTERED IMAGE FUSION

R. H Cofer and S. P. Kozaitis
Dept. of Electrical/Computer Engineering
Florida Institute of Technology
150 W. University Blvd.
Melbourne, FL 32901-6988
(407) 768-8000
rhcofer@ee.fit.edu
spk@ee.fit.edu

ABSTRACT

One of the highest potential uses of image fusion is that of recognition of critical targets. The continuing image fusion question then is how to make optimal use of the often disparate forms of encountered image detail during fusion. Toward this end, many techniques have been advanced for fusion to a single viewable image. Fewer techniques have been suggested toward fusion with the goal of directly improving target detection or recognition.

Based upon emerging trends in pixel accurate registration of images, we show the theoretical foundations required to optimally fuse target imagery for recognition. Results obtained can be applied to both the cases of automatic target recognition and image analysis.

Keywords: image fusion, pixel-registration, recognition, ATR

1.0 INTRODUCTION

Most target recognition systems, whether machine or vision based, suffer from a paucity of available information with which to make the recognition determination. One way to overcome such limitation is through fusion of imagery from multiple sensors and views. Every sensor is ideally suited to the detection of some specific form of target phenomenology which is difficult to impossible to pick up with an alternate sensor. For instance IR sensors are most sensitive to thermal characteristics, LADAR sensors respond to surface geometry, electo-optical sensors can provide color and SAR sensors can be exquisitely sensitive to metallic objects. Various sensor views can also provide additional insights into the target's phenomenology including a better appreciation of the 3-D target shape and articulations. Patently it is impossible to adequately fuse such a wealth of possibilities into a single viewable image or even to overcome the inherent geometric projective conflicts involved. A new paradigm must be sought.

Actually, the ultimate goal of much of image fusion is clear: improve the detectability and recognition of target objects and conditions. If one focuses on this goal, then the need to directly fuse multiple images into a single viewable image lessens. It is replaced by the need to output better detection and recognition decisions through proper fusion of the underlying evidence gathering processes leading up to the decision. This new

 0-8194-1837-4/95/$6.00

view of image fusion is greatly aided by the recent advances in image registration of multiple images to sub-pixel accuracies across differing image sensor platforms. Here we assume that such registration[1] has been accomplished. As a result, we are free to consider the primary problem of image fusion of pixel-registered imagery for target detection and recognition.

Since the underlying goal of sensor fusion is now seen to be more accurate target detection and recognition, there is need for fusion techniques that will optimize the probability of detection and recognition while minimizing the probability of false alarm. As a result, Section 2 briefly reviews the required decision theory.[1] There it is shown that optimal target decision processing depend upon the determination of certain probabilities. In the case of sensor fusion, these probabilities are the joint probability of the output of each sensor. However, these joint probability can be quite difficult to calculate given usual techniques.[2] What is needed and what we supply are simplifying techniques that essentially maintain decision optimality during the sensor fusion calculations. To a large degree, we show how each sensor's output can be individually processed and how these individual results can then be probabilistically fused. We then briefly discuss how the results can be used in an example scenario.

2.0 FUNDAMENTAL THEORY

If we are to make a decision on the basis of "fused imagery," then it should be optimal in terms of some weighting of the costs of proper and erroneous decisions. For instance the cost of mis-recognizing a target should be more than the cost of correct recognition. In this realm, the optimal[2] decision process has long been known [4] as

$$\text{If} \quad \Lambda > \eta \quad \begin{array}{l}\text{then recognize } T \\ \text{else recognize } B\end{array} \tag{1}$$

where Λ, the likelihood function, is

$$\Lambda = \frac{p(X|T)}{p(X|B)} \tag{2}$$

X is the set of all imagery data to be fused.

η, the decision threshold, is

$$\eta = \frac{(c_{10} - c_{00})P(T)}{(c_{01} - c_{11})P(B)} \tag{3}$$

1. The concept of pixel registration refers to registration of the ground datum to common pixel locations within all of the images to be fused. In practice, this usually implies a detailed knowledge of the sensor to scene sensing geometry. Most modern automatic registration processes either know or develop knowledge of this geometry to permit improved registration accuracies
2. No other decision process has lower cost on the average.

where the c's are the relative costs of correct and erroneous recognitions of T and B.

Let T and B be the cases where there is or is not a target to be found in the fusion data set, X. Then the fused decision process of Eq. (1) can be applied to the problem of detecting the presence or absence of the "fused" target. This is accomplished by properly evaluating Λ.

To evaluate Λ, one needs the probability density functions of $p(X|T)$ and $p(X|B)$. Again, X itself is composed of imagery X_1, X_2,...,X_n from the individual sensors. Thus more specifically, one needs $p(X_1, X_2, \ldots, X_n|T)$ and $p(X_1, X_2, \ldots, X_n|B)$. These joint probabilities can be broken down to a meaningful form through an ordered introduction of auxiliary information including that afforded by pixel level registration.

3.0 TARGET INDEXING

The question now arises as how to determine $p(X_1, X_2, \ldots, X_n|T)$ and $p(X_1, X_2, \ldots, X_n|B)$ so as to provide optimal sensor fusion. For example, assume that is possible to find the single sensor probabilities $p(X_1|T)$, and $p(X_2|T)$. On the basis of simplicity, one would hope to be able to probabilistically fuse these probabilities into the desired joint probability, $p(X_1, X_2|T)$ as

$$p(X_1, X_2|T) = p(X_1|T)\,p(X_2|T) \tag{4}$$

but this is not generally possible as the individual sensor outputs are seldom independent.

Fortunately, it is still usually possible to combine individual sensor probabilities into a proper fused joint probability density function upon first introducing indexing variables[3]. An indexing variable is a target attribute such as position, orientation, articulation, size, etc. where every value of the index is mutually exclusive. An overall target index, I_i, can be considered an element of the Cartesian product of all of the target indexing variables.

As a result of target indexing, the joint probability of the fusion imagery data set can now be expressed as

$$p(X|T) = \sum_i p(X|I_i, T)\,p(I_i|T) \tag{5}$$

This in turn infers that the final fused likelihood, Λ, is given by

$$\Lambda = \sum_i \Lambda_i p(I_i|T) \tag{6}$$

where Λ_i is the fused likelihood of looking for a specific target indexing, I_i, within the fusion data set. Eq. (6) shows that the final stage of developing a fused detection or recognition can be the probabilistic summation of likelihoods, Λ_i, of the fused image data sets conditioned upon the expectation of specific individual target indexings.

4.0 PIXEL REGISTRATION

Pixel registration of the various images within the fused image data set now provide an important source of information.

Since we assume pixel level registration of the imagery sensors, the image may be partitioned into various regions, x_j, on the ground with approximately independent behaviors. This allows Eq. (5) to be re-expressed as

$$p(X|T) = \sum_i \prod_j p(x_j|I_i, T)\, p(I_i|T) \tag{7}$$

when a target, T, is assumed present.

On the other hand, if no target is present, B, the fused image data set, X, portrays only scene background with probability $p(X|B)$. Again partitioning the registered image fusion data set into the regions x_j gives

$$p(X|B) = \prod_j p(x_j|B) \tag{8}$$

Substituting Eq.'s (7) and (8) into Eq. (6) gives the fused indexed conditioned likelihood, Λ_i, as

$$\begin{aligned} \Lambda_i &= \prod_j \frac{p(x_j|I_i, T)}{p(x_j|B)} \\ &= \prod_j \Lambda_{ij} \end{aligned} \tag{9}$$

where Λ_{ij} is the likelihood ratio of the region, x_j, within the fusion image data set as conditioned on target index, I_i. Thus Eq. (9) shows that the next to last stage of an optimally fused detection or recognition is the multiplication of the likelihoods of the regions of the image as conditioned on the various target indexings. This computation can be improved greatly by noting that for any region, x_l, not affected by target presence, we will have

$$p(x_l|I_i, T) = p(x_l|B) \tag{10}$$

and so

$$\Lambda_{ij} = 1 \tag{11}$$

Thus all regions unaffected by target presence need not enter into the $\prod$ calculation of Eq. (9) and thus allows greatly improved computational efficiency and retention of precision during the fused likelihood calculation.

5.0 OPTIMAL SENSOR FUSION

Based upon the results of the prior two sections, we now complete the breakdown of the overall optimal sensor fusion problem to one based upon the optimal detection or recognition problem on each separately considered image of the fusion image data set.

In sensor fusion, x_j is actually composed of data from each of the n sensors and so Λ_{ij} can be rewritten as

$$\Lambda_{ij} = \frac{p(x_{j1}, x_{j2}, \ldots, x_{jn} | I_i, T)}{p(x_{j1}, x_{j2}, \ldots, x_{jn} | B)} \tag{12}$$

Since $x_{j1}, x_{j2}, \ldots, x_{jn}$ of $p(x_{j1}, x_{j2}, \ldots, x_{jn} | I_i, T)$ is conditioned on I_i and T, this is equivalent to saying that the only influence the target has on $x_{j1}, x_{j2}, \ldots, x_{jn}$ is through its j'th surface patch being interrogated. Once this influence has been determined by target modeling, the $x_{j1}, x_{j2}, \ldots, x_{jn}$ become mutually independent[1] resulting in

$$\begin{aligned} p(x_{j1}, x_{j2}, \ldots, x_{jn} | I_i, T) = p(x_{j1} | I_i, T) \\ \bullet\, p(x_{j2} | I_i, T) \\ \bullet \quad \bullet \quad \bullet \\ \bullet\, p(x_{jn} | I_i, T) \end{aligned} \tag{13}$$

Once a suitable model for the background has been found, $p(x_{j1}, x_{j2}, \ldots, x_{jn} | B)$ can be similarly decomposed. This then gives the likelihood function, Λ_{ij}, as

$$\Lambda_{ij} = \prod \Lambda_{ijk} \tag{14}$$

that is, the average of the product of the likelihood functions, Λ_{ijk}, where Λ_{ijk} is simply the likelihood, Λ_{ij}, taken for the k'th sensor. Thus the first step of developing the fused target detection or recognition solution is simply that of multiplying the single sensor likelihoods, Λ_{ijk}.

Eq.'s (6), (9) and (14) are obviously the simplest general forms of optimal sensor fusion possible being based on the product of the likelihood outputs of each individual sensor given target indexing certainty. They are also particularly appealing sensor fusion solutions as they require minimal additional stages of processing over those required for individual sensor optimal detection and recognition.

---

1. The temperature distribution of a target in an IR image is independent of its range in a LADAR image given conditional knowledge of the target's indexing.

6.0 AN EXAMPLE

Although the approach shown above is optimal, this does not imply that the resulting fused target detection will be without error - but rather that no other image fusion process can give better results. The question then arises as to the effectiveness of the optimal fused image detection or recognition process relative to the best possible given just a single image of the fusion image data set. General results can not be given here; however, representative results can.

Assume that N images are to be fused, there are n pixels on target in each image, the average target minus background graylevel is μ, the image variability on each pixel of the target region Gaussian with standard deviation σ, each pixel is independent of the others, and the target indexing is known. Then Figure 1. holds where $k = \frac{\sqrt{n}\mu}{\sigma}$. As is verifiable by common sense, the figure shows that fusing poor imagery does little to

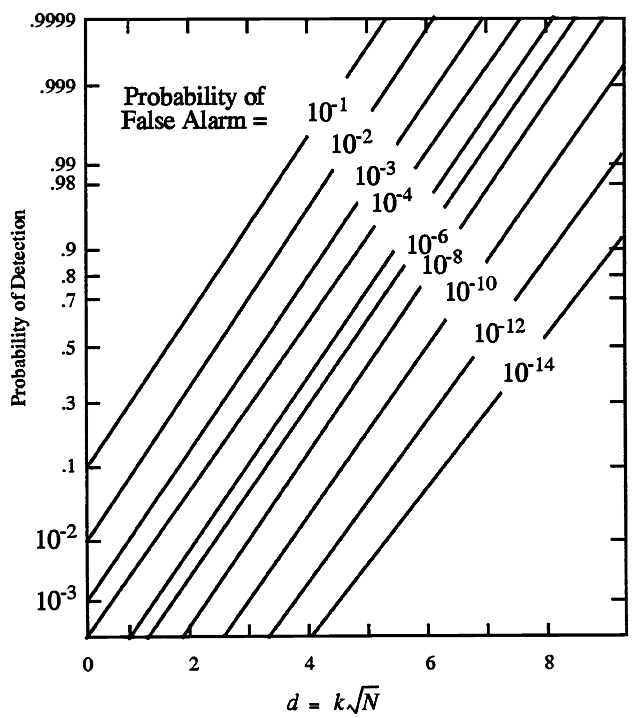

Figure 1. Detection Probability versus Normalized Number of Images to Fuse

improve overall detection accuracy. For instance assume $k = 1$, $N = 1$ and a desired probability of false alarm of .1, the probability of correct detection will be .3. Under the same conditions, raising the number ,N, of images to be fused to 4 will only increase the probability of detection to .55 - a relatively poor gain.

On the other hand fusing good images will indeed greatly improve the detection accuracy. Now assuming $k = 4$, $N = 1$ and a desired probability of false alarm of 10^{-14}. Then the probability of correct detection will be .003. Under these new conditions; however, raising the number, N, of images to be fused to 4 will now increase the probability of detection to .43. This is an improvement in the probability of correct detection of more than 1000 - solely due to use of the optimal image fusion technique presented here.

7.0 CONCLUSIONS AND FURTHER WORK

The approach to sensor fusion discussed here results in the ability to dedicate early sensor specific processing to individual processors while conducting the optimal fusion of their results via a conceptually simple final process. The approach is capable of incorporating target indexing variabilities and is highly optimal when appropriate target and background modeling is possible.

Although we have show the optimality of the fusion process advanced here, it is more difficult to develop the relative gain in performance obtained. Nevertheless, we provide a resulting optimal sensor fusion performance curve for one rather general class of image fusion data sets. With due caution, these results may be used as a rule of thumb in predicting the performance detection and recognition gains in other cases.

Our future work will center on two areas. The first is the use of region likelihoods as an overlay for more conventional image fusion products in aiding the image analyst. The second is the continued development of higher-speed approximations to the optimal fusion technique set forth here.

8.0 References

1 P. S. James, *Bayesian Statistics*, John Wiley, 1989

2 R. H. Cofer, "Probability Event Spaces For ATR," *Internal Report, USAF-UES Summer Faculty Research Program*, 1989

3 R. H. Cofer, "Probabilistic IR Evidence Accumulation," *Final Report, USAF-UES Summer Faculty Research Program*, 1990

4 H. L.Van Trees, *Detection, Estimation, and Modulation Theory*, Vol. 1., New York: John Wiley, 1968

5 J. Wozencraft and I. Jacobs, *Principles of Communication Engineering*, John Wiley, 1967