Low Level Fusion of Imagery Based on Dempster-Shafer Theory

Xiaohui Yuan and **Jian Zhang** and **Xiaojing Yuan** and **Bill P. Buckles**
Department of Electrical Engineering and Computer Science
Tulane University
New Orleans, LA 70001
{yuanx, zhangj, yuan, buckles}@eecs.tulane.edu

Abstract

An approach to fuse multiple images based on Dempster-Shafer evidential reasoning is proposed in this article. Dempster-Shafer theory provides a complete framework for combining weak evidences from multiple sources. Such situations typically arise in the image fusion problems, where a 'real scene' image has to be estimated from incomplete and unreliable observations. By converting images from their spatial domain into the evidential representations, decisions are made to aggregate evidences such that a fused image is generated. The proposed fusion approach is evaluated on a broad set of images and promising results are given.

Introduction

Multi-sensor systems provide a more comprehensive perception of the monitored event. With many emerging sensor technologies, the capability of integrating evidences from various signals, which are based on different mechanisms, becomes highly demanding. The study of data fusion has experienced an explosion in recent years. (Li, Manjunath, & Mitra 1995; Koren, Laine, & Taylor 1995; Rockinger 1997; Nikolov *et al.* 1998; Petrovic & Xydeas 1999)

Image fusion can be performed roughly at three stages: pixel level, feature level, and decision level. Varieties of approaches have been proposed including weighted average, PCA based fusion, Laplacian pyramid based fusion, contrast pyramid based fusion, etc. Detailed literature reviews can be found at (Zhang & Blum 1999; Yuan, Yuan, & Buckles 2002).

From the evidence point of view, fusion reduces the imprecision and uncertainty by making use of redundancy and complementary information of the source images. That is, weak evidences from inputs are utilized to provide the best estimation.

Evidence theory was first proposed by Shafer in 1970s, which is based on Dempsters research. The advantages are that it distinguishes between the lack of belief and disbelief and allows the probability to be assigned to the union of the propositions in the frame of discernment. Such lack of belief typically arises in image fusion problems where a 'real scene' image is to be estimated from incomplete and unreliable observations.

Evidence theory has been very successful in many areas including pattern classification (Denoeux 1997), object recognition (Borotschnig *et al.* 1998), database integration (Lim, Srivastava, & Shekhar 1994), sensor fusion (Murphy 1999), etc. However, when dealing with pixel level image fusion, critical problems arise. First, a meaningful evidential representation founded on human cognitive grounds is the key to pixel level fusion. Second, the environment-caused and sensor mechanism-caused disparity is inevitable and cause confusion during information aggregation. Generally, a preprocessing step is employed to adjust source images to a comparable status. Techniques include computing first order derivatives of pixel values, applying histogram adjustment, constraints optimization (Socolinsky 2000) etc.

In this article, we propose the evidence functions of images derived from wavelet feature and develop a decision fusion scheme following the Dempster-Shafer combination method. Successful experiments are performed on different kinds of images and results are compared with collateral works.

The rest of this article is organized as follows. Section 2 introduces the background information on Dempster-Shafer theory. Section 3 presents the image representation in the wavelet domain and demonstrates fusion using evidence function as well as developing the aggregation schemes. Examples are illustrated in section 4, followed by the discussion. Section 5 summaries and concludes this article.

Dempster-Shafer Evidence Theory

Let θ represent a finite set of hypotheses for a problem domain, called the *frame of discernment*. A piece of evidence that influences our belief concerning these hypotheses induces a *mass function* m, defined as a function from 2^θ to [0, 1] that satisfies the following constraints:

$$m(\phi) = 0 \quad and \quad \sum_{A \subseteq \theta} m(A) = 1 \tag{1}$$

$m(A)$ can be interpreted as the belief that one is willing to commit to hypothesis A (and to none of its subsets) given the available evidence. The subsets A of θ such that $m(A) < 0$ are called the *focal elements* of m. the belief $Bel(A)$ and

the plausibility $Pl(A)$ of $A \subseteq \theta$ can be computed from m as follows:

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{2}$$

$$Pl(A) = \sum_{A \cap B \neq \phi} m(B) = 1 - Bel(\overline{A}) \tag{3}$$

$Bel(A)$ measures the extent to which one believes in A, given the evidence pointing to that hypothesis either directly through $m(A)$ or indirectly through $m(B)$ for all $B \subseteq A$. The extent to which one doubts hypothesis A is based on the mass of belief committed to $\overline{A}$ or its subsets, and is therefore represented by $1 - Pl(A)$.

An important aspect of D-S theory concerns the aggregation of evidence provided by different sources. If two mass functions m_1 and m_2 induced by distinct items of evidence are such that $m_1(B) > 0$ and $m_2(C) > 0$ for some non-disjoint subsets B and C of θ, then they are combinable by means of Dempster's rule. The orthogonal sum of m_1 and m_2, denoted $m = m_1 \oplus m_2$, is defined as

$$m(\phi) = 0 \tag{4}$$

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B \cap C = \phi} m_1(B) m_2(C)} \tag{5}$$

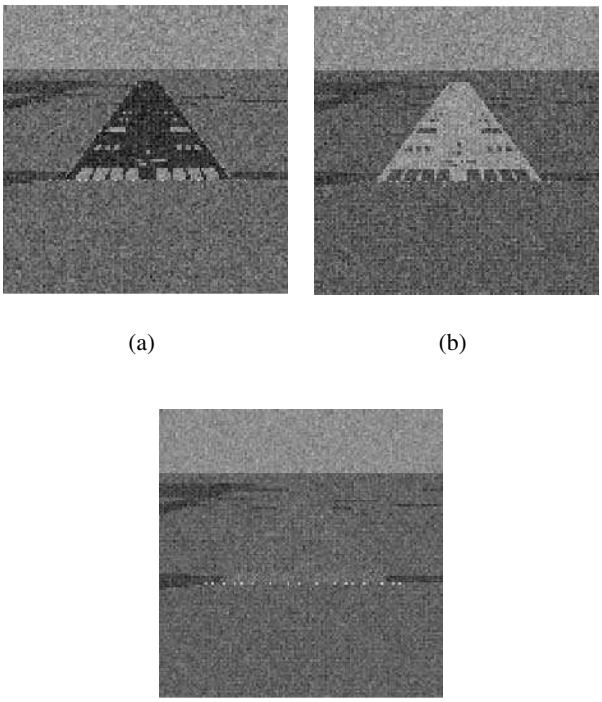
Consequently, the mass function resulting from the combination of several items of evidence does not depend on the order in which the available information is taken into consideration and combined.

Decision Making and Fusion

The direct evidence lies in the brightness of pixels. Typically, images observed through different channels suffer disparity due to the sensor mechanism, resolution, quantization error, error introduced by numerical reconstruction algorithm, etc. In most cases, features in the source images are likely to appear differently from one sensor image to another, but are usually closely related. It is important to note that the relationships between the image features are local in nature. However, the design of effective fusion algorithms is difficult because the optimality of combining different images depends on the local relationship between sensor imagery. For example, fusion based on averaging works well for images that are approximately the same except for the additive noise. On the other hand, averaging can cause a quality reduction in case of reversal of polarity, as shown in Figure 1. Therefore, it is desirable that a fusion algorithm combine different local regions according to local feature relationships.

In contrast, wavelets decomposition transfers an image into spatial-frequency domain, where evidence is implied by the static and dynamic coefficient components. The integration evidence of the dynamic components and static components circumvents the disparity problem as well as provides a human cognitive favored image evidential representation.

In the rest of this section, we first give a brief review of the wavelet description of images, which provides a base for the definition of evidence function. An integration and decision scheme is described in detail at the end.



(a) (b)

(c)

Figure 1: Fusion of reverse polarity images with average combination approach (Sharma 1999). Figure (a) and (b) are source images. Figure (c) illustrates the fusion result with average combination method.

Image Description

The discrete wavelet transform (DWT) of a discretely sampled image is computed using an orthogonal filter base. One of the most frequently used wavelet bases is the Haar wavelet, which contains a high-pass filter and a low-pass filter of length two.

The 2-D wavelet transformation includes a decomposition stage and a reconstruction stage. The decomposition is an iteration process that includes filtering and down-sampling operations. Let j denote the decomposition level. Given the conjugate quadrature filters (CQF), each row of the image undergoes decomposition with high-pass and low-pass filters, resulting in two images whose horizontal resolution is reduced by a factor of two, that is, down-sampled by two. Next, the high-pass and low-pass subband images obtained by two decompositions are each separately filtered column-wise to obtain four subband images: low-low band c^{j+1}, low-high band $d^{H,j+1}$, high-low band $d^{V,j+1}$, and high-high band $d^{D,j+1}$.

After one iteration, the c^{j+1} band is again subjected to the row-column filtering and sampling to obtain a further

decomposition. This process is iterated either until the low-pass image has no more interesting features or a preset number of times is reached.

The reconstruction stage is an inverse process except that instead of down-sampling an up-sampling operation is performed before the filtering. Rows and columns of the high-pass subband images $d^{D,j+1}$, $d^{V,j+1}$, $d^{H,j+1}$ and the low-pass subband image c^{j+1} are filtered with CQF filters to obtain the low-pass subband image c^j at level j. This process is repeated until the original resolution level is reached.

Evidence Representation of Images

One of the essential problems of pixel level fusion with Dempster-Shafter reasoning is construction of the evidential representation of images. Fortunately, DWT transforms a spatial image into a spatial-frequency representation that consists of coefficients having clear meaning with respect to the occurrence of changes. Such changes are generally recognized as edges in the spatial domain. Therefore, the decisions on combining wavelet coefficients essentially preserve edges and aggregate background information. (Mallat & Zhong 1992; Yuan, Yuan, & Buckles 2002; Nikolov *et al.* 1998)

Wavelet decomposition consists of static and dynamic bands. The static band is the band that is processed only with low-pass filters. It contains the background information and most of the bias of the observation, e.g. lighting conditions.

The dynamic bands show the changes over a neighborhood. The evidence of these bands is independent from the static evidence. Therefore, different evidence functions are derived for static and dynamic bands.

Let I_q denote source images, where $q \in \{1, \ldots, Q\}$. Let D_q denote the wavelet decompositions of J levels of image I_q. Given the decomposition level $j \in \{1, \ldots, J\}$ and band number $k \in \{H, V, D\}$, the coefficients of a dynamic band at level j band k is denoted as $d_{k,j}(p)$. Therefore, the evidence function for dynamic bands is given as follows:

$$N_q^{k,j} = \frac{|d_q^{k,j}|}{max(d_q^{k,j})}, \quad q \in \{1, \ldots, Q\} \tag{6}$$

$$m_i^d(p) = \frac{N_i^{k,j}(p)}{max[\sum_{q=1}^{Q}(N_q^{k,j})(p)]}, \quad i \in \{1, 2\} \tag{7}$$

The evidence value of empty set $m(f)$ is defined to be zero. Notice that the evidence function is produced by normalization over multiple decompositions, i.e. cross normalization, such that the unity property of the summation is preserved.

Let $c^L(p)$ denote site p at the static band. The corresponding evidence function is given as follows:

$$m_i^L(p) = \frac{c_i^L(p)}{max[(\sum_{q=1}^{Q} c_q^L)(p)]} \tag{8}$$

Evidence Aggregation

Assume the observations are independent, and uncertainty exists in every image. After images are converted into wavelet based evidence functions, such uncertainty is essentially the insufficiency of the evidence on the occurrence of local changes. Thereby, the belief matrix $Bel_q^{k,j}$ is constructed for each wavelet band as well as its uncertainty matrix $U_q^{k,j}$.

The belief matrix is the aggregation of evidence values at every site p, while the uncertainty matrix is the complementary to the belief, i.e. $1 - Bel(p)$. The discernment of site p, namely $q(p)$, is the union of the belief $Bel_q^{k,j}(p)$ and the uncertainty $U_q^{k,j}(p)$ across the same bands in all decompositions. The decision is made based on the cumulative believes:

$$s^{k,j}(p) = \{d_q^{k,j}(p) | Conf_q^{k,j}(p) = max[Conf_i^{k,j}(p)], \; i = 1, \ldots, Q\} \tag{9}$$

and

$$\begin{aligned} Conf_q^{k,j}(p) &= \frac{1}{1-K}[Bel_q^{k,j}(p) \cap \theta_{q'}^{k,j}(p)] \\ &= \frac{1}{1-K}\prod[Bel_q^{k,j}(p) \times \theta_{q'}^{k,j}(p)], \\ &\quad q \neq q' \end{aligned} \tag{10}$$

where $K = \sum_{q \neq q'} Bel_q^{k,j}(p) \times Bel_{q'}^{k,j}(p)$.

$Conf_q^{k,j}(p)$ is the confidence supported by its belief $Bel_q^{k,j}(p)$. It is the intersection of the belief carried by decomposition q and the discernments of all the other decompositions. Factor K adjusts the probability of the empty set over the confidence that the constraint in the equation 1 is satisfied.

The fusion is the combination of the coefficients that have the highest confidence with respect to the status of pixel brightness, i.e. the pixel gray level remains constant or varies.

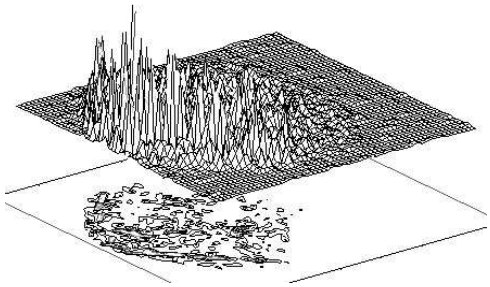
The fusion of the static bands c_q^L is straightforward, since the belief matrices represent the existence of the objects in extremely low resolution. While under such low resolution, the belief at site p is actually the joint beliefs of a region in the original image. Therefore, a different scheme is applied.

$$s^L(p) = \{c_q^L(p) | c_q^L(p) = max[c_i^L(p)], i = \{1, \ldots, Q\} \tag{11}$$

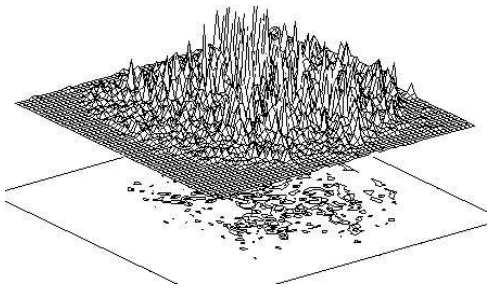
Experiments and Discussion

In this section, we first illustrate fusion examples with medical diagnosis images. Then comparisons with other widely used approaches are discussed.

Our first experiment demonstrates fusion of retinal images. In the computer-aided examination of possible leaky or broken blood vessels, it is prerequisite to fuse multiple

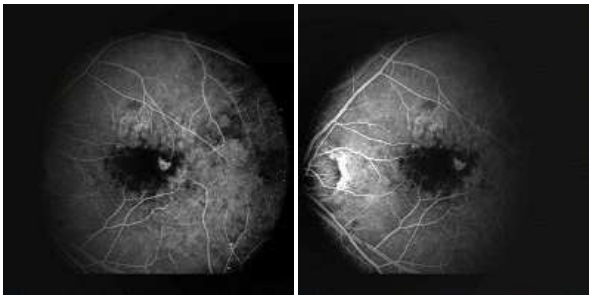


(a)

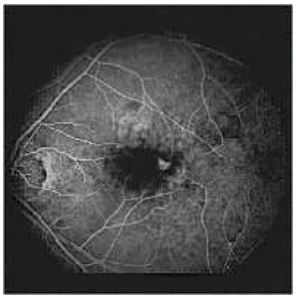


(b)

Figure 2: Evidence functions of the retinal images.



(a) (b)



(c)

Figure 3: Fusion of retinal images.

retinal images into a meaningful combination. Figure 2 illustrates evidence functions of one dynamic band. The corresponding source images are shown in Figure 3. The evidence values are cross normalized on each wavelets decomposition band. The demonstrations contain a 3-D plot of the evidence functions with a 2-D contour image underlying it. From the 2-D contour image, it is clear that the evidence functions provide high belief on different but overlapping regions. The 3-D plot illustrates how much each evidence function supports the belief.

Using the decision scheme developed in section 3, evidential functions of the source images are fused and the final result can be reconstructed using the inverse wavelets transform. Figure 3(a) and 3(b) are source angiographical retinal images. Figure 3(c) illustrates the fused image that successfully aggregates the details from each input.

Another example is illustrated in figure 4. Figure 4(a) is the CT scan that exposes the bony structure. Figure 4(b) is a MRI image of the same patient giving details on soft tissues. The fusion process effectively combines source images into a complete description of the patients inner cranial view as shown in Figure 4(c).

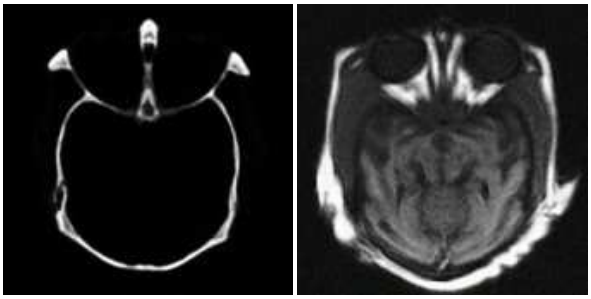
To evaluate the performance, we adopt the mutual information of the reference image and fused image, the root mean square error as well as the normalized cumulative edge distance (Yuan, Yuan, & Buckles 2002) as the measure of fusion performance. Due to the requirement of having a real scene image (as opposed to an artificial 'truth' image) as reference, it is infeasible to use either mutual information or root mean square error for comparison on cases such as the previous medical images.

Table 1 lists the comparison of the performance of five fusion methods, including fusing with average (AVE), PCA based fusion (PCA), Laplacian pyramid based fusion (LAP), contrast pyramid based fusion (CP) and evidential reasoning based fusion (DS). Eight image scenes are used including the two just discussed. With mutual information, larger value represents higher similarity between the reference image and the fusion result, while smaller values in RMSE and NCED stand for less error and therefore represents better fusion performance. It is clear that fusion based on evidence combination outperforms the other methods in most cases.

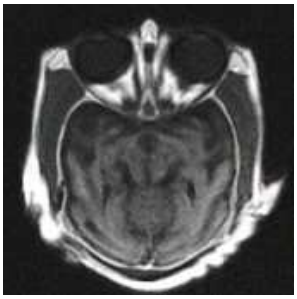
Conclusion

In this article, we present an evidential reasoning driven fusion approach based on Dempster-Shafer theory. The proposed fusion approach is founded on human perception and provides a better interpretation of combining wavelets based evidence than the traditional multi-resolution fusion method.

By converting an image into its evidential representation, a fusion scheme derived from Dempster-Shafer theory is developed and is used to aggregate evidence using belief and plausibility, where the evidence function for combination is generated from the cross normalization over the wavelets



(a) (b)



(c)

Figure 4: Fusion of CT scan and MRI scan.

bands. Compared to other widely used fusion methods, the D-S based evidence combination scheme makes better decision on combining feature evidence. Promising results are given and compared with results generated by several other fusion methods.

Table 1: Fusion performance comparison.

Normalized Cumulative Edge Distance

Images	AVE	PCA	LAP	CP	DS
retinal	2125	2284	1951	2067	1939
med. scan	1131	1281	1123	2311	1090
airplane	1064	1062	795	792	736
APC truck	1370	1385	1226	1210	986
boat	1738	1708	1359	1372	1288
surface	1754	1735	1474	1448	1345
truck	1300	1308	1259	1246	1233
lena	1264	1290	955	964	990

Mutual Information

Images	AVE	PCA	LAP	CP	DS
airplane	2.634	2.707	3.306	3.241	3.477
APC truck	2.233	2.286	2.887	2.890	3.178
boat	2.186	2.228	3.291	3.294	3.600
surface	2.046	2.067	2.801	2.802	3.164
truck	2.120	2.211	2.461	2.460	2.502
lena	2.796	2.804	3.173	3.175	4.453

Root Mean Square Error

Images	AVE	PCA	LAP	CP	DS
airplane	6.339	6.303	2.730	2.700	3.000
APC truck	4.863	4.831	4.223	4.062	3.353
boat	7.332	7.155	2.465	2.479	2.307
surface	4.307	4.306	2.326	2.334	1.871
truck	3.863	3.793	3.739	3.751	3.536
lena	4.453	4.479	2.492	2.455	2.454

References

Borotschnig, H.; Paletta, L.; Prantl, M.; and Pinz, A. 1998. Active object recognition in parametric eigenspace. In *Proceeding of Brit. Mach. Vision. Conference*.

Denoeux, T. 1997. Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition* 30(7).

Koren, I.; Laine, A.; and Taylor, F. 1995. Image fusion using steerable dyadic wavelet transform. In *Proc. IEEE Intl Conference on Image Processing*, 232–235.

Li, H.; Manjunath, B. S.; and Mitra, S. K. 1995. multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing* 57(3).

Lim, E.-P.; Srivastava, J.; and Shekhar, S. 1994. Resolving attribute incompatibility in database integration: An evidential reasoning approach. In *Proc. of 10th IEEE International Conference on Data Engineering*.

Mallat, S., and Zhong, S. 1992. Characterization of signals from multiscale edges. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 14(7).

Murphy, R. R. 1999. Dempster-shafer theory for sensor fusion in autonomous mobile robots. *IEEE Transactions on Robotics and Automation* 14(2).

Nikolov, S. G.; Bull, D. R.; Canagarajah, C. N.; Halliwell, M.; and Wells, P. N. T. 1998. Fusion of 2-d image using their multiscale edges. In *Fusion of Earth Data*.

Petrovic, V., and Xydeas, C. 1999. Cross band pixel selection in multi-resolution image fusion. In *Proc. of SPIE Conference*, volume 3719.

Rockinger, O. 1997. image sequence fusion using a shift-invariant wavelet transform. In *Proceedings of the 1997 International Conference on Image Processing*.

Sharma, R. K. 1999. *Probabilistic Model-based Multisensor Image Fusion*. Ph.D. Dissertation, Oregon Graduate Institute of Science and Technology.

Socolinsky, D. A. 2000. Dynamic range constraints in image fusion and visualization. In *Proceedings of Signal and Image Processing 2000*.

Yuan, X.; Yuan, Z.; and Buckles, B. P. 2002. A computational vision approach to improve fusion performance. In *Proceedings of ICIG'2002*.

Zhang, Z., and Blum, R. S. 1999. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. In *Proceedings of the IEEE*, volume 87.