Towards Autonomous Image Fusion

M. Hossny, S. Nahavandi, D. Creighton, and A. Bhatti
Centre for Intelligent Systems Research
Deakin University
Australia
{mhossny, nahavand, dougc, asimbh}@deakin.edu.au

Abstract

Mobile robots are providing great assistance operating in hazardous environments such as nuclear cores, battlefields, natural disasters, and even at the nano-level of human cells. These robots are usually equipped with a wide variety of sensors in order to collect data and guide their navigation. Whether a single robot operating all sensors or a swarm of cooperating robots operating their special sensors, the captured data can be too large to be transferred across limited resources (e.g. bandwidth, battery, processing, and response time) in hazardous environments. Therefore, local computations have to be carried out on board the swarming robots to assess the worthiness of captured data and the capacity of fused information in a certain spatial dimension as well as selection of proper combination of fusion algorithms and metrics. This paper introduces to the concepts of Type-I and Type-II fusion errors, fusion capacity, and fusion worthiness. These concepts together form the ladder leading to autonomous fusion systems.

Index Terms
Autonomous, Image Fusion, Type I, Type II, Fusion Capacity

I. INTRODUCTION

Image fusion aims to merge two or more images to produce a new image that is better than the original ones. The term 'better' differs from one context to another. In some contexts, it means holding more information. In other contexts, it means getting a more accurate result or reading. In general, an image fusion system takes as an input two or more source images and produces one fused image as an output. The fusion process applies a fusion algorithm, repeatedly, on the source images and/or intermediate output images. As a matter of fact, almost all the present image fusion operators are binary ones. The need for ternary or higher narity fusion operators has not yet been identified as fusing two images at a time with an associative operator does the job.

Researchers have developed several definitions of image fusion. Wald [1] derived a formal definition of image fusion as "a formal framework in which are expressed means and tools for the alliance of data
Fig. 1. Applications of image fusion

originating from different sources. It aims at obtaining information of greater quality; the exact definition of ‘greater quality’ will depend upon the application”. In [2], Pohl and Genderen defined image fusion as “the combination of two or more different images to form a new image by using a certain algorithm”. Li et al. [3] defined image fusion as “fusion refers to the combination of a group of sensors with the objective of producing a single signal of greater quality and reliability”. A good survey about data fusion terminologies and definitions is presented in [4]. In general, as a part of the definition, a fusion algorithm must maintain the closure property, which means its output must be of the same kind as its inputs.

A wide variety of applications can make use of image fusion as illustrated in Fig. 1. A typical scenario is a swarm of cooperative mobile robots operating in a hazardous environment, such as an automated battlefield, disaster area, or even the human body. These environments usually feature limited resources (e.g. bandwidth, processing, battery life, storage, and response time). In automated battlefields where a swarm of robots are gathering information from a sensor network or directly from the field, or human body with a swarm of nano-robots swimming through blood vanes; there must be some scale defining how good
or bad the captured images are before sending them. In general, maintaining autonomy in image fusion systems requires clear identification of the fusion objective function, fusion errors, worthiness of images, and the maximum amount of information to be squeezed into certain spatial dimensions.

During the past decade researchers developed several signal, pixel and decision-level fusion techniques in [5]–[10]. Almost all pixel-level fusion algorithms depend on averaging information or features extracted from source images and reconstruct the fused image at the end. The simplest idea for pixel-level fusion is averaging corresponding pixels in source images. The use of global and local image coefficients in fusion, like outputs from Fourier and Radon transforms, allowed fusion systems to absorb more real and effective pieces of information and present them in the fused image. Multi-resolution transforms such as wavelets and other pyramid transforms [11]–[14] provide more reliable and expressive features to fuse. Analysis $\psi$ and synthesis $\psi^{-1}$ functions define the fusion response [15].

While pixel level fusion produces more visible information from different source images, signal level fusion aims to estimate correct and noise free readings from imaging sensors. The main framework used in signal level fusion is the Kalman filter [16]. It uses previous readings to predict next ones more accurately. A state space with transition functions or matrices to hop from one state to another is used. Predictor and corrector equations are the two basic equations that govern this estimation. Decision level fusion provides image understanding using multiple images to have different opinions or decisions to the same problem.

Image fusion quality metrics evolved from image processing objective quality measures, including mean-squared-error (MSE), entropy and information measures. In fact, developing an objective quality metric is very challenging. Wang has discussed some of the difficulties facing objective quality measures in [17]. These measures have then been realized in image fusion. Finally, researchers concluded that universal quality index (UQI) founded by Wang and Bovik in [18] and improved to be a structural similarity metric (SSIM) in [19], does capture localized structural similarities between images. The simplest idea is to average quality distance of the fused image from each of the source ones. These measures have been used on overlapping portions of the images to maintain localization. In [20], Piella improved UQI and added a saliency factor for each pair of corresponding blocks (a block from each input image) being examined against the corresponding block in the fused image. They proposed the use of simple information measures such as standard deviation, dynamic range and entropy. Many researchers worked on deriving the most suitable and realistic saliency functions to weigh the estimated amount of information being transferred from source images into the fused one. Covariance and quadtree decomposition methods have been investigated in [21] and [22], respectively. In [23], Yang simply applied the weighting function only where source image do not have structural similarities.
Xydeas and Petrovic [25] estimated fusion performance based on edges in the image. Zhang and Blum [26] used a mixture of Rayleigh probability density functions to model image histogram and estimate quality of noisy images. Mutual information measure was examined by Qu [27] and modified by Hossny et. al in [28]. It described the use of a joint histogram between a fused image with each of the source images. Local cross-correlation of feature maps of the source and fused images was studied by Zhao in [29]. Bundilov and Bretschneider [30] applied multi-level thresholding to variance maps in order to identify the spatial blocks holding more information and, probably, should be transferred into the fused image. They concluded that quality measures of image fusion algorithm should be extended to take into considerations segmented regions and weight averaging their contribution in assessment of quality based on their areas and how much
information each region holds. They have derived a thresholding based solution in [31]. An excellent survey about quality measures and its evolution journey are available in [32], [33].

A. Contributions

This paper discusses the concepts that autonomous fusion rely on. The main contributions can be summarized in three points;

1) Formulating the image fusion objective function and isolating the fusion worthiness function from feature extraction.
2) Highlighting what Type-I and Type-II fusion errors are.
3) Introducing to fusion capacity.

The rest of this paper is organized as follows. Section II presents a survey on objective validation of image fusion performance metrics using fusion algorithm/metric duality index [34]. It also derives the argument upon which the need for fusion capacity is justified. Section III introduces to the concept of fusion capacity. Finally, Section IV concludes while Section V discusses future improvements.

II. Fusion Algorithm/Metric Duality

By definition, an image fusion algorithm aims to transfer informative features from source images into the fused image. Obviously, fusing all features from both source images into one fused image is not applicable. This is because it requires squeezing all information from two images into the same spatial space of only one image. This has motivated researchers to focus on transferring important features only where the application domain dictates the fusion worthiness of different features in the source images. Therefore, the objective of image fusion can be redefined into; 1) transferring important features from source images and 2) ignoring the non-important features and minimizing their effect on the fused image. This section presents a survey on objective validation of image fusion metrics. It redefines the objective function of fusion, defines Type-I and Type-II fusion errors, identifies fusion control cases.

A. Image Fusion Objective Function

Throughout the last decade, researchers have developed many image fusion algorithms attempting to only maximize the selection of fusion-worthy features. Interested readers may refer to [35] for a proper survey on image fusion algorithms. On the other hand, in performance evaluation contexts, only two papers have taken the features’ fusion worthiness into consideration. In [25], Xydeas and Petrovic assumed a maximum importance of visible feature. Thus, they used magnitude and direction maps of gradients of source and fused images to estimate fusion worthiness. In [36], Cvejic et al. assumed the perfect fusion result, also
called 'ideal’ fusion, to be the combination of segmented objects in all source images. One can conclude from this discussion that the objective of image fusion algorithms and metrics depends mainly on fusion worthiness and the extracted features dictated by the application domain. Thus, an image fusion objective function can be formulated as follows;

**Definition 1 (Image Fusion Objective Function)**

Let $A$ be an image fusion application domain where a set of images $I$ can be acquired for fusion. Let $\psi_i : I \rightarrow F_i$ be the projection function from images in $I$ onto the $i$th feature dimension $F_i$. Let $\omega_i : F_i \rightarrow \mathbb{R}$ be the fusion worthiness function of the $i$th feature. If $I_f \subseteq I$ is a set of images participating in a multi-source image fusion; then the objective function of fusing the $i$th feature from all images in $I_f$ can be formulated as:

$$\arg \max_{f \in \{ \psi_i(x) | x \in I_f \}} \omega_i(f) = \left\{ f | \forall \hat{f} : \omega_i(\hat{f}) \leq \omega_i(f) \right\}$$

(1)

Multiscale decomposition (MSD) fusion algorithms chose $\{\psi_i\}$ to be the singleton set of a pyramid analysis function [35]. In such algorithms, $\{\omega_i\}$ can be as simple functions as the intensity level of different bands in source images. It can also take into consideration the consistency of selected features with other neighbor features in the fused image as described in consistency verification. In a remote sensing context, researchers chose $\{\psi_i\}$ to be the mapping between Red-Green-Blue (RGB) and Hue-Saturation-Value (HSV) color spaces [37].

Separating feature worthiness function $\omega_i$ from feature extraction function $\psi_i$ adds an extra degree of freedom to ignore a strong feature for the sake of preserving neighborhood features and hence minimizing fusion artifacts [38]. Consistency verification [3] is a direct application of having worthiness function separated from feature extraction.

**B. Type-I and Type-II Fusion Errors**

The fusion worthiness function highlighted in the previous section aims at identifying whether a certain feature in a source image is important enough to be transferred into the fused image. However, the result of this function is subject to two kinds of errors. These errors were defined by Neyman and Pearson as Type-I and Type-II errors [39]. Type-I fusion error, also known as false negative, is an estimation of a number of important features that have not been identified as fusion worthy. This is the type of error that all image fusion performance metrics have been measuring so far. On the other hand Type-II fusion error, also known as false positive, is the error of fusing a feature that is not fusion worthy. This kind of error is also known as fusion artifacts [38]. Figure 2 demonstrates typical image fusion results and illustrates...
examples of Type-I errors (fusion loss) and Type-II errors (fusion artifacts). The test images were obtained from the TNO’s widely popular image fusion test cases [24].

This argument raises a very important question regarding how to measure these two types of errors. Until now, there is no objective non-reference test to measure how far did a fusion algorithm or metric minimize Type-I and Type-II errors. Most researchers used to validate their proposed fusion performance metrics using reference images [20]–[22]. The reference image is used to create two deformed complementary versions, fuse them, and compare the resulting fused image with the original image. However, this comparison is not objective since the ultimate fusion algorithm does not really exist, neither does any performance metric. Therefore, it is analogous to comparing results that are subjective to two sources of errors, namely fusion error and metric error. In [36], Cvejic et al. proposed obtaining the ‘ideal’ fusion result using segmentation maps of source images. However, this ideal fusion result is still subject to segmentation errors. The perfect fusion needs to be carried out on two different source images that we certainly know what the results should be. The perfect fusion requires identifying control cases for both Type-I and Type-II errors.

C. Control Cases

In [34], Hossny and Nahavandi proposed a duality index to measure the suitability of image fusion metrics for different image fusion algorithms. They recommended using fusion with images that are completely
noninformative (0-image) and completely informative (∞-image) as control cases as illustrated on TNO’s infrared images [24] in Fig 3. The “controlled” image fusion test cases can then be formulated as follows;

\[ \forall x \in \mathcal{I} \quad x \oplus 0 = x \]  
\[ \forall x \in \mathcal{I} \quad x \oplus \infty = \infty \]  

Adding image fusion metrics to equations Eq. 2 and Eq. 3 maps the problem from abstract image space into real numbers as follows;

\[ DI_0^\oplus (\oplus, Q_0) = \frac{1}{|\mathcal{I}|} \sum_{x \in \mathcal{I}} Q_0(x \oplus 0, x) \]  
\[ DI_\infty^\oplus (\oplus, Q_0) = \frac{1}{|\mathcal{I}|} \sum_{x \in \mathcal{I}} Q_0(x \oplus \infty, \infty) \]

where \( \oplus \) is the fusion algorithm (operator), \( Q_0 \) is an image dissimilarity metric, \( \mathcal{I} \) is the set of images acquired from a particular application domain, 0 is the completely non-informative zero image, and \( \infty \) is the completely informative infinity image. They concluded that testing duality \( DI_\infty^\oplus \) with an infinity image, if one can identify or approximate it, provides information on the ability of \( \{\omega_i\} \) to capture important features from source images (Type-I Error). On the other hand, the zero-referenced duality index \( DI_0^\oplus \) measures the ability of \( \{\omega_i\} \) to minimize the effect of non-informative features from being added to the fused image (Type-II Error). In [40], Hossny et al. discussed the constraints and equations guiding the selection of zero and infinity images for multi resolution fusion algorithms and metrics. Similar guidelines can be drawn to characterize the performance of other families of image fusion algorithms and metrics.

### III. Fusion Capacity

The discussion of Type-I and Type-II errors raises a question regarding how much information can be stored in limited spatial dimensions. The very common case of image fusion can be summarized in having high frequencies information in a source image at spatial coordinates where the other image holds low frequencies. While this is quite common in fusing infrared and thermal images with visual images (at night), expanding fusion to multi-source multi-modal images highlights the need to study the fusion capacity in order to minimize overlapping of high frequencies causing fusion loss and fusion artifacts. This section studies saturated images, fusion capacity index, and fusion capacity maps.
A. Saturated Images

Considering gray-scaled non-indexed images (8 bits), natural scene statistics suggest using only 5-6 bits of color coding depending on brightness, contrast, exposure, gain, and dynamic range [41], [42]. Theoretically, in order to use all 8 bits the image must maintain uniformly distributed normalized histogram \( U(0, G - 1) \) where \( G \) is number of gray levels \( (2^8) \). However, a uniformly distributed image does not appear in natural scenes it is basically a perfect uniformly distributed noise image or a high entropy gradient images as shown in Fig. 4.

B. Fusion Capacity Index

Therefore, fusion capacity of an image can be measured as how far the image histogram is from a perfect uniformly distributed histogram. Using mutual information measure [27], [28] fusion capacity can be modeled as;

\[
C(x) = 1 - I(X, U(0, G - 1)) \\
I(X, Y) = H(X) + H(Y) - 2H(X, Y)
\]

where capitalized \( X \) is the normalized histogram of a source image \( x \), \( I(X, Y) \) is the mutual information measure [27], \( H(X) \) is the entropy of \( X \), \( H(X, Y) \) is the joint entropy of two histograms \( X;Y \), and \( 0 \leq I(X, Y) \leq H(X, Y) \). In [28], Hossny et al. suggested using normalized version of mutual information (NMI) using Kvalseth’s normalization [43] as follows;

\[
I(X, Y) = 2 \cdot \frac{H(X) + H(Y) - 2H(X, Y)}{H(X) + H(Y)}
\]

One must not confuse a perfect uniformly distributed image with an equalized image because an equalized image does not cover the whole dynamic range of colors.

C. Localized Fusion Capacity Maps

Not only does entropy of the source image matter, but also how the information is distributed across the spatial dimensions. Figure 4 shows three gray-scaled images with maximum entropy (8 bits). Their eligibility for fusion with a natural scene image is illustrated in Fig. 5. In this figure, fusion was carried out using principle component analysis (PCA) fusion. Fusion with gradient pixels image transferred only the very high entropy areas in the image because every block in the uniformly distributed image maintains high entropy. Fusing the source image with block gradient image results more information (e.g. background trees) transferred from the source image because each block maintains a zero entropy in itself. The third
Fig. 4. Three uniformly distributed images with 8-bits entropy and different spatial distribution. From left to right: uniformly distributed pixel gradients, uniformly distributed block gradients, and uniformly distributed random noise.

Fig. 5. Fusion results of a source image [24] (far left) with low capacity saturated images (8 bpp entropy) in Fig. 4. From left to right: source image, fusion with gradient pixels (entropy dropped to 7.89 bpp), gradient blocks (entropy drops to 7.54 bpp), and uniformly distributed random noise (entropy unchanged at 8 bpp).

case is fusing the source image with a uniformly distributed white noise. The fusion algorithm fails to find any informative features compared to the white noise. Hossny et. al called this result fusion with an infinity image in [40].

Therefore, calculating fusion capacity $C$ locally with fixed block size or per regions (e.g. using quadtree decomposition) estimates capacity maps where more information can be fused. Figure 6 demonstrates fusion capacity maps of a source image in Fig. 5 using quadtree decomposition and mutual information of local blocks. Local fusion capacity map can then be defined as follows;

$$C^Q(x) = \sum_{x_i \in Q(x)} I(X_i, \mathcal{U}_i (0, G - 1))$$

where $\mathcal{U}_i$ is a uniform distribution, $x$ is the source image, $x_i$ are quadtree sub-images, capital $X_i$ represent...
normalized histograms of $x_i$, and $Q(x)$ is the quadtree decomposition function that recursively subdivides an image into four quarters if its overall entropy $H(X)$ greater than a cut-off entropy $\epsilon_x$ as follows;

$$Q(x) = \begin{cases} \bigcup_{i=1}^4 Q(x_i) & \text{if } H(X) \geq \epsilon_x \\ \{x\} & \text{otherwise} \end{cases}$$

(10)

Cvajic’s MI variation using Tsallis’s entropy (TMI) suffers from the inconvenience of choosing the tuning parameter $\alpha$ [44] while other gradient-based metrics can be tuned with the kernel window size that control the localization [18]–[23], [25], [26], [30]–[32]. Therefore, MI [27], NMI [28], and TMI [44] comparisons to gradient-based metrics [18]–[23], [25], [26], [30]–[32] are not objective because of the mismatched physical meaning of tuning parameters. Quadtree localized fusion capacity measure also faces the same challenge. The decomposed quadtree topology is very sensitive to the change in cut-off entropy $\epsilon_x$ because of the limited number of the very narrow dynamic range (8-bits per color channel). Thus, choosing a very high cut-off entropy derives very shallow quadtree topologies and looses the structural information. On the other hand, choosing a very low cut-off entropy leads to performing the metric in a pixel by pixel fashion. In general, choosing the correct cut-off entropy depends on many other parameters such as image resolution, spatial distribution of entropy, and the minimum block size (localization) required. Some implementations of quadtree decomposition, such as MATLAB’s, enforce minimum and maximum block sizes as a constraint regardless what the cut-off entropy is. However, such implementation provide a non-differential function that cannot be optimized.

In order to overcome this problem and provide a more intuitive parameter to tune $C^Q$, the cut-off entropy was estimated as the average of local entropies of blocks with a kernel-window of size $w$ applied to the image being decomposed $x$ as follows;

$$\epsilon_x^W = \frac{1}{|W|} \sum_{w \in W} H(X|w)$$

(11)

where $w$ is the localization kernel window, and $|W|$ is the number of all local blocks. The selected kernel size provides a cut-off entropy that maintains a minimum block size.

IV. CONCLUSION

This paper introduced to autonomous image fusion. It highlighted the concepts of Type-I and Type-II fusion errors, fusion capacity, and fusion worthiness. It reformulated the objective function of image fusion separating fusion worthiness functions from feature extraction functions. This separation adds an extra
degree of freedom that allows ignoring a strong fusion worthy feature in order to preserve the consistency with other features in the neighborhood and hence reducing fusion artifacts.

This paper presented an argument about image fusion error characterization. Type-I fusion error is an estimation of the number of important features that have not been identified as fusion worthy. Type-II fusion error is the error of fusing a feature that is not fusion worthy. It also proposed a fusion worthiness test using infinity-referenced $DI^\infty_I$ and zero-referenced $DI^0_I$ duality indices to estimate Type-I and Type-II fusion errors, respectively. The paper also introduced to fusion capacity index. It defined the fusion capacity as the normalized distance, using normalized mutual information, between an image histogram from a uniformly distributed probability distribution.

V. Future Advancements

The arguments presented in this paper open doors for further generalization of image fusion objective function in order to measure intra- and inter-image fusion worthiness and capacity. However, the promising breakthrough improvement of image fusion systems is deriving an automatic selection and tuning of fusion algorithms and metrics in order to equip mobile robots with means for dynamic reconfiguration and tuning while operating in hazardous areas.

A. Fusion Worthy Images

In hazardous environments mobile robots may need to make split second decisions while operating within limited resources (e.g. bandwidth, processing, storage, and battery life). The cost of transferring a non-informative image or fusing more information into a low capacity image is very high and may compromise the success of the mission and/or the safety of other robots and even human lives. Therefore, every captured image has to be tested for fusion worthiness before encountering further processing. Considering the response
time limitation, it is recommended for fusion worthiness functions to employ heuristics to be very quick and precise. Using equation Eq. 2 and Eq. 3, worthiness test can be formulated in maximizing the \( DI_{I}^{0} \) error and minimizing the \( DI_{I}^{1} \) for every tested image given that 0— and \( \infty \)− images have been identified during the calibration and training modes.

B. Optimizing Fusion Operator/Metric Combination

The image fusion operator/metric duality index has a very strong mathematical background obtained from abstract algebra, which Ritter employed to form Image Algebra in [45]. Duality equations can be generalized based on the maturity level of the algebraic structures constructing the algebra. As the algebraic structures mature, more constraints become available reducing the cardinality of the solution set allowing more precise solutions at the cost of longer training procedures and more complicated learning scenarios. For instance, adding a diffusion algorithm (inverse operator) adds one more constraint to the system of equations Eq. 2 and Eq. 3 as follows;

\[
\forall x \in I \quad Q_{0}(x \oplus 0, x) = 0 \\
\forall x \in I \quad Q_{0}(x \oplus \infty, \infty) = 0 \\
\forall x \in I \quad Q_{0}(x \oplus x^{-1}, 0) = 0
\]

where \( Q_{0} \) is a fusion performance metric, \( \oplus \) is the fusion operator (algorithm), boldfaced 0 is the zero (noninformative) image, boldfaced \( \infty \) is the infinity image, \( \cdot^{-1} \) is the diffusion operator, and \( 0 \in \mathbb{R} \) is a real valued zero. This abstractive formulation highlights the need to solve (or possibly train) for the best combination of fusion operators, metrics, and inverse in order to develop solutions on the fly and achieve full autonomy of fusion systems. Using operator algebra one can, theoretically, solve this system given the zero and infinity images in a particular application domain \( I \).

ACKNOWLEDGMENT

This research was fully supported by the Centre for Intelligent Systems Research (CISR) at Deakin University.

REFERENCES
