A Total Variation Based Algorithm for Pixel Level Image Fusion

Mrityunjay Kumar *, *Student Member, IEEE,* Pradeep Ramuhalli, *Member, IEEE,* and Sarat Dass, *Member, IEEE*

Abstract

In this paper a total variation (TV) based approach is proposed for pixel level fusion to fuse images acquired using multiple sensors. In this approach, fusion is posed as an inverse problem and a locally affine model is used as the forward model. A total variation norm based approach in conjunction with principal component analysis is used iteratively to estimate the fused image. The feasibility of the proposed algorithm is demonstrated on images from computed tomography (CT) and magnetic resonance imaging (MRI) as well as visible-band and infrared sensors. The results clearly indicate the feasibility of the proposed approach.

Index Terms

Image fusion, pixel-level fusion, total variation, eigenvector, inverse problem, forward model.

I. Introduction

Data fusion integrates redundant as well as complementary information present in input signals in such a manner that the fused signal describes the true source better than any of the individual signals. The exploitation of redundant information improves accuracy and the reliability whereas integration of complementary information improves the interpretability of the signal [1], [2]. Data fusion plays

Manuscript received March 13, 2008.

Mrityunjay Kumar and Pradeep Ramuhalli are with the Department of Electrical and Computer Engineering, Michigan State University. Address: 2214B Engineering Building, East Lansing, MI 48824 USA. Email: {kumarmri, rpradeep}@egr.msu.edu. Phone: 517-432-4615. Fax: 517-353-1980.

Sarat Dass is with the Department of Statistics and Probability, Michigan State University. Address: A-430 Wells Hall, East Lansing, MI 48824 USA. Email: sdass@msu.edu. Phone: 517-432-5412. Fax: 517-432-1405.

a crucial role in image-based applications such as computer vision, robotics, remote sensing etc. The goal of image fusion is to extract information from input images such that the fused image provides better information for human or machine perception as compared to any of the input images [3]–[6]. Image fusion has been used extensively in various areas of image processing such as remote sensing, biomedical imaging, nondestructive evaluation etc. [7]–[9]. For example, in optical remote sensing, due to physical and technical constraints, some sensors provide excellent spectral information but inadequate spatial information about the scene. On the other hand, there are sensors that are good at capturing spatial information but which fail to capture spectral information reliably. Fusing these two types of data provides an image that has both the spatial and the spectral information. Therefore, only the fused image needs to be stored for subsequent analysis of the scene [10], [11]. As another example, recent developments in medical imaging have resulted in many imaging techniques to capture various aspects of the patient's anatomy and metabolism [12], [13]. However these techniques are complementary in nature. For example, magnetic resonance imaging (MRI) is very useful for defining anatomical structure whereas metabolic activity can be captured very reliably using positron emission tomography (PET). Hence, by using fusion, it is possible to obtain a single image that describes anatomical as well as metabolic activity of the patient effectively [14].

From the perspective of fusion, features of the observed images that are to be fused can be broadly categorized in the following three classes.

1) Common features: These are features that are present in all the observed images.
2) Complementary features: Features that are present only in one of the observed images are called complementary features.
3) Noise: Features that are random in nature and do not contain any relevant information are termed as noise.

Note that the above categorization of the features is local in nature. A fusion algorithm should be able to select the feature type automatically and then fuse the information appropriately. For example, if the features are similar then the algorithm should perform an operation similar to averaging but in the case of complementary features, should select the feature that contains relevant information.

Due to the large number of applications as well as the diversity of fusion techniques, considerable efforts have been put in developing standards for data fusion. Several models for data fusion have been proposed in the recent past [15], [16]. One of the models commonly used in signal processing applications is the three-level fusion model that is based on the levels at which information is represented [17]. This

model classifies data fusion into three levels:

- Data or low level fusion
- Feature or intermediate level fusion
- Decision or high-level fusion

At the data level, raw images obtained from different sensors are fused to generate a new image, which is expected to be more informative than the inputs [18]. Pixel level fusion is an example of low-level fusion. Feature level fusion deals with the fusion of features such as edges or texture while decision level fusion corresponds to combining decisions from several experts [19], [20].

In pixel level fusion the fused pixel is derived from a set of pixels in the various inputs [17], [21]. The main advantage of pixel level fusion is that the original measured quantities are directly involved in the fusion process. Furthermore, algorithms are computationally efficient and easy to implement. However, pixel level fusion algorithms require the input images to be co-registered. An intuitive approach for pixel level fusion is to average the input images. Averaging reduces sensor noise but it also reduces the contrast of the complementary features. More robust algorithms for pixel level fusion such as weighted average, transform based approach, etc. have been proposed in the literature [17], [22]–[26]. In the weighted average approach, the fused pixel is estimated as the weighted average of the corresponding input pixels. However, the weight estimation usually requires a user-specific threshold [17], [22], [23]. Transform based approaches provide a fused image with full contrast, but these approaches are sensitive to sensor noise [24]–[26].

In this paper, a total variation norm [27] based approach has been adopted to fuse the pixels of the noisy input images. The total variation norm has been used in several image processing applications [27]–[30]. In this paper, we propose to use total variation norm for image fusion. The estimation of fused pixels is posed as an inverse problem, with a local affine function proposed by Sharma et al. [31] used as the forward model. Sharma et al. used a Bayesian approach to solve the forward model; however, the *a priori* probability density function of the fused pixel was assumed to be Gaussian which is a limiting assumption. Yang et al. [32] proposed an expectation-maximization (EM) based algorithm to estimate the fused image from the local affine model, with the noise modeled using a K-term mixture of Gaussian distributions. In the proposed approach, the total variation norm is used to solve the forward model and estimate the pixels of the fused image. The total variation norm is based on a nonlinear partial differential equation (PDE) and is robust to noise [33], [34]. Furthermore, it is a data driven approach and does not require knowledge of the probability density function of the fused pixels.

This paper is organized as follows. In Section II, we describe the forward model for fusion. Section III

explains the proposed fusion algorithm, while Section IV summarizes the overall algorithm. Section V presents the results and is followed by concluding remarks.

II. The Image Acquisition Model

Let $f_o(x,y)$ be the true image, which is inspected by n different sensors and $f_1(x,y)$, $f_2(x,y)$, $\cdots$, $f_n(x,y)$ are the corresponding n measurements for $x,\ y \in \Omega$. The local affine transform that relates the input pixel and the corresponding pixel in the measured images is given by [31]

$$f_i(x,y) = \beta_i(x,y) f_o(x,y) + \eta_i(x,y);\ \ 1 \leq i \leq n. \tag{1}$$

Here, $\beta_i(x,y)$ and $\eta_i(x,y)$ are the gain and sensor noise, respectively, of the i^{th} sensor at location (x,y). The goal of fusion is to estimate $f_o(x,y)$ from $f_i(x,y)$, $1 \leq i \leq n$.

In many applications such as radar imaging and visual and IR imaging, the complementary as well as redundant information are available at the local level in the measured images [31], [32]. The main advantage of the local affine transform model is that it can relate this local information content in a mathematically convenient manner. For example, as an extreme case, two sensors i and j $(i \neq j;\ 1 \leq i,j \leq n)$ have complementary information at location (x,y) if $\beta_i(x,y) \neq \beta_j(x,y)$ and $\beta_i(x,y), \beta_j(x,y) \in \{0,1\}$. Similarly, these two sensors have redundant information if $\beta_i(x,y) = \beta_j(x,y)$.

III. Image Fusion

A. Total Variation Norm for Image Fusion

In order to estimate $f_o(x,y)$ from eq. (1), we assume that $f_o(x,y),\ f_i(x,y) \geq 0$ $(1 \leq i \leq n)$. This assumption is valid for many imaging devices such as digital cameras, IR cameras, etc. and does not limit the proposed algorithm in any way since data not satisfying this requirement (*i.e.*, with negative pixel values) can always be transformed using a simple linear transformation to make the pixel values positive. Furthermore, we also assume that sensor noise $\eta_1(x,y)$, $\eta_2(x,y)$, $\cdots$, $\eta_n(x,y)$ are zero mean random variables and are independent of each other. The standard deviation of $\eta_i(x,y)$ is denoted as σ_i, and σ_i is assumed to be known *a priori* and independent of spatial location (x,y).

From eq. (1),

$$\begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} f_o + \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_n \end{pmatrix} \tag{2}$$

$$\Rightarrow \mathbf{f} = \boldsymbol{\beta} f_o + \boldsymbol{\eta} \tag{3}$$

where $\mathbf{f} = [f_1, \cdots, f_n]^T$, $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_n]^T$, $\boldsymbol{\eta} = [\eta_1, \cdots, \eta_n]^T$, and the reference to the pixel location (x, y) has been dropped for notational simplicity. Eq. (3) can be re-arranged as

$$\boldsymbol{\beta}^N \mathbf{f} \quad = \quad f_o + \boldsymbol{\beta}^N \boldsymbol{\eta} \tag{4}$$

where $\boldsymbol{\beta}^N = (\boldsymbol{\beta}^T \boldsymbol{\beta})^{-1} \boldsymbol{\beta}^T$.

The goal of fusion is to estimate f_o $\forall (x, y) \in \Omega$ from eq. (4). Note that $\boldsymbol{\beta}^N$ is unknown and must be estimated from the measurement images $f_i(x, y)$ $(1 \leq i \leq n)$. The method for estimating $\boldsymbol{\beta}^N$ has been discussed in Section III-B. If $\boldsymbol{\beta}^N$ is known, then a common approach to estimating f_o from the measurements $\mathbf{f}$ minimizes the cost function $\|\boldsymbol{\beta}^N \mathbf{f} - f_o\|$. The solution to this least square estimate is $f_o = \boldsymbol{\beta}^N \mathbf{f}$. However, this approach is not robust to noise. As discussed earlier, the TV norm is more robust to noise as compared to L_2 norm based estimators such as least square. A comparison of least square and TV norm based fusion has been shown in Fig. 1. Fig. 1a and 1b show the input noisy images that are to be fused. Fig. 1c and 1d are the fused images obtained using least square and TV norm based approach, respectively. A detailed comparison of the fusion results are presented in Section V. However, a visual comparison of Fig. 1c and 1d clearly shows that the fused image obtained using a TV norm based approach is less noisy and provides pronounced edges as compared to the fused image obtained using a least squares approach.

Therefore, we propose to estimate $f_o(x, y)$ by minimizing its TV norm under suitable constraints as given below:

$$\text{minimize} \int_\Omega \Big| \nabla f_o(x, y) \Big| dx\, dy; \;\; \Big| \nabla f_o(x, y) \Big| = (f_{o_x}^2 + f_{o_y}^2)^{\frac{1}{2}} \tag{5}$$

subject to the following constraints

$$\int_\Omega f_o \, dx \, dy = \int_\Omega \boldsymbol{\beta}^N \mathbf{f} \, dx \, dy \tag{6}$$

$$\int_\Omega (\boldsymbol{\beta}^N \mathbf{f} - f_o)^2 \, dx \, dy = \int_\Omega \sigma^2 \, dx \, dy \tag{7}$$

where f_{o_x} and f_{o_y} are the derivatives of f_o in x and y direction, respectively, and σ^2 is the variance of the transformed noise $\boldsymbol{\beta}^N \boldsymbol{\eta}$. Eq. (5) represents the TV norm for f_o. The constraints given by eq. (6) and (7) involve mean and variance of the noise $\boldsymbol{\beta}^N \boldsymbol{\eta}$. Eq. (6) ensures that the noise $\boldsymbol{\beta}^N \boldsymbol{\eta}$ is of zero mean whereas eq. (7) is based on *a priori* information about the variance of the noise. Note that the sensor gain $\boldsymbol{\beta}^N$ depends on spatial location, which in turn also implies that σ^2 is a function of (x, y). Assuming the sensor gain vector $\boldsymbol{\beta}$ to be independent of f_o and its derivatives, the approach proposed by Rudin et

al. in [27] may be used to derive the following iterative solution to solve eq. (5):

$$f_o^{k+1} = f_o^k - \tau_k \left(\nabla \bullet \left(\frac{\nabla f_o^k}{|\nabla f_o^k|} \right) + \lambda_k \left(f_o^k - \boldsymbol{\beta}^N \mathbf{f} \right) \right) \tag{8}$$

where k is the iteration number, τ_k : is the time step, " $\bullet$ " represents the dot product and

$$\lambda_k = -\frac{1}{\int_\Omega \sigma^2 \, dx \, dy} \int_\Omega \left(f_o^k - \boldsymbol{\beta}^N \mathbf{f} \right) \nabla \bullet \left(\frac{\nabla f_o^k}{|\nabla f_o^k|} \right) dx \, dy.$$

The boundary condition $\frac{\partial f_o^{k+1}}{\partial \boldsymbol{\xi}}\Big|_{\partial\Omega} = 0$ must be satisfied for each iteration of eq. (8), where $\partial\Omega$ represents the boundary of Ω and $\boldsymbol{\xi}$ is the outward normal along $\partial\Omega$. However, the solution to eq. (8) requires knowledge of $\boldsymbol{\beta}$ and σ^2. Since σ^2 is the variance of $\boldsymbol{\beta}^N \boldsymbol{\eta}$ and the variance of the noise vector $\boldsymbol{\eta}$ is assumed to be known, σ^2 can easily be estimated if $\boldsymbol{\beta}$ is known. In general, $\boldsymbol{\beta}$ is unknown and must be estimated prior to solving eq. (8). This is addressed next.

B. Estimation of $\boldsymbol{\beta}$

To estimate $\beta_i (1 \leq i \leq n)$, we assume that the gain of the sensor varies slowly from one spatial location to another. Hence, β_i can be assumed to be constant over a small region of the sensor image. Thus, the input images are split into small regions of size $p \times q$ and sensor gains may be computed for each region. Furthermore, we also assume that $\beta_i \in [0, 1]$ for all $1 \leq i \leq n$.

From eq. (3), it is clear that in the absence of sensor noise, the variance of the observed images $f_1, f_2, \cdots, f_n$ are related to each other through sensor gains $\beta_1, \beta_2, \cdots, \beta_n$. Therefore, the variances of the observed images may be used to estimate $\boldsymbol{\beta}$. Also, it is well known that for a given dataset, the principal eigenvector of the correlation matrix points in the direction of maximum variance [35]. Motivated by this fact, we propose to use a principal eigenvector based approach to estimate sensor gains. The pixels of the observed images are used to construct the correlation matrix and its principal eigenvector is used to estimate the sensor gains. This approach is explained in detail below.

Let us assume that the input images are divided into blocks of size $p \times q$. For one such block denoted by R, assume that $\left(f_1^1, f_1^2, \cdots, f_1^{p \times q} \right)$, $\left(f_2^1, f_2^2, \cdots, f_2^{p \times q} \right)$, $\cdots$, $\left(f_n^1, f_n^2, \cdots, f_n^{p \times q} \right)$ are the lexicographical representation of pixel values of n sensor images. These pixels can be viewed as n-variate random variables $\left(f_1^1, f_2^1, \cdots, f_n^1 \right)^T$, $\left(f_1^2, f_2^2, \cdots, f_n^2 \right)^T$, $\cdots$, $\left(f_1^{p \times q}, f_2^{p \times q}, \cdots, f_n^{p \times q} \right)^T$. Let $\boldsymbol{\mu} = [\mu_1, \cdots, \mu_n]^T$ be the principal eigenvector of the correlation matrix $\boldsymbol{\Sigma}_\nu$, where

$$\boldsymbol{\Sigma}_\nu = \frac{1}{(p \times q) - 1} \sum_{k=1}^{p \times q} \nu_k \nu_k^T. \tag{9}$$

and $\nu_k = \left(f_1^k, f_2^k, \cdots, f_n^k\right)^T$. Assuming noise in eq.(3) to be zero, and due to the assumption that sensor gains are constant over block R, it can be shown that

$$\mathbf{\Sigma}_\nu = \frac{1}{(p \times q) - 1} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} (\beta_1, \cdots, \beta_n) \sum_{k=1}^{p \times q} f_o^2.$$

$$\Rightarrow \mathbf{\Sigma}_\nu = C \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} (\beta_1, \cdots, \beta_n). \tag{10}$$

where $C = \frac{\sum_{k=1}^{p \times q} f_o^2}{(p \times q) - 1}$.

Let ρ and $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)^T$ be the principal eigenvalue and principal eigenvector respectively, of $\mathbf{\Sigma}_\nu$. From eq.(10) it is clear that the rank of the matrix $\mathbf{\Sigma}_\nu$ is one. Therefore, ρ and $\boldsymbol{\mu}$ can be related to the sensor gains as shown below:

$$\rho = C(\beta_1^2 + \cdots + \beta_n^2). \tag{11}$$

$$\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)^T \propto (\beta_1, \cdots, \beta_n)^T \tag{12}$$

where "$\propto$" indicates proportionality. Therefore, we propose the following rule to estimate $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_n]^T$ for region R.

1) Compute $\mathbf{\Sigma}_\nu$ for block R, and estimate $\boldsymbol{\mu}$ such that $\boldsymbol{\mu}^T \boldsymbol{\mu} = 1$.
2) Set $\beta_1 = \beta_2 = \cdots = \beta_n = 1$ if $\mu_1 = \mu_2 = \cdots = \mu_n$, otherwise set $\boldsymbol{\beta} = \boldsymbol{\mu}$.

This process may be repeated for each block. Note that the sensor gains computed this way are assumed to be constant over the region, *i.e.*, $\beta_i(x, y) = \beta_i(r, s)$, where $x \neq r$, $y \neq s$, $x, y, r, s, \in R$ and $1 \leq i \leq n$. Next, we show that this proposed rule for estimating sensor gains can discriminate between common and complementary features. For the sake of simplicity, assume n = 2, *i.e.*, there are only two sensors. Then,

$$\mathbf{\Sigma}_\nu = C \begin{bmatrix} \beta_1^2 & \beta_1 \beta_2 \\ \beta_1 \beta_2 & \beta_2^2 \end{bmatrix} \tag{13}$$

$$\rho = C(\beta_1^2 + \beta_2^2). \tag{14}$$

and $\boldsymbol{\mu}$ can be estimated by solving the following equation:

$$\mathbf{\Sigma}_\nu \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \rho \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}; \ \mu_1^2 + \mu_2^2 = 1. \tag{15}$$

If the two input images have completely redundant information, *i.e.*, $\beta_1 = \beta_2$, then, $\rho = 2C\beta_1^2 = 2C\beta_2^2$ and $\mu_1 = \mu_2 = \frac{1}{\sqrt{2}}$. In this case, we select $\beta_1 = \beta_2 = 1$, *i.e.*, $\boldsymbol{\beta} = [1, 1]^T$. Similarly, if $\beta_1 > \beta_2$ (corresponding to some redundant information between the two measurements), then, from eq. (15):

$$\frac{\mu_1^2 - \mu_2^2}{\mu_1 \mu_2} = \frac{\beta_1^2 - \beta_2^2}{\beta_1 \beta_2}. \tag{16}$$

It is clear that in this case, $\mu_1 > \mu_2$. Therefore, according to the proposed rule, $\beta_1 = \mu_1$ and $\beta_2 = \mu_2$. In the extreme case when $\beta_1 \neq 0$ and $\beta_2 = 0$, it can be shown using a similar procedure that $\boldsymbol{\beta} = \boldsymbol{\mu} = [1, 0]^T$. Similarly, it can be shown that $\boldsymbol{\beta} = \boldsymbol{\mu} = [0, 1]^T$ if $\beta_1 = 0$ and $\beta_2 \neq 0$. Therefore, it is clear that in case of complementary information, the proposed approach selects the region that has relevant information. Due to the assumption that pixel values of the measurement images are always positive numbers, the sensor gains estimated using the proposed approach will always be positive. Note that in the above analysis, we assumed sensor noise to be zero. If noise is present, then the principal eigenvector will provide a noisy estimate of the sensor gains. However, these gains are further used in the TV based framework of eq. (8), which is robust to noise and compensates for the noisy estimate of the sensor gains.

IV. Overall Approach

The flowchart of the proposed pixel level fusion method is shown in Fig. 2. We first split the input images into several blocks, and arrange the pixels inside each block lexicographically. The lexicographically arranged pixels of any given block from all the input images are used to compute the correlation matrix Σ_ν and the principal eigenvector of this correlation matrix is used to determine the gain $\boldsymbol{\beta}$ of that block. This process is repeated for all the blocks. After the sensor gains are estimated, eq. (8) is solved iteratively to estimate the fused pixels, $f_o(x, y)$ for all locations $(x, y) \in \Omega$

V. Results

The results of the proposed fusion algorithm are presented in this section. The proposed fusion algorithm was applied to two different datasets: (i) medical imaging and (ii) aircraft navigation. For each dataset, only two input images were considered for the fusion process and these two inputs were co-registered. The sensor noise was simulated by adding zero mean white Gaussian noise to the input images. For ease of quantitative analysis of the fusion performance, the variance of the noise for each input image was selected appropriately to get the same level of signal-to-noise (SNR) ratio for all the input images, where the SNR was computed using the following expression:

$$\text{SNR} = 10\, log_{10} \frac{\text{Signal Variance}}{\text{Noise Variance}} \text{ dB}. \tag{17}$$

Fig. 3 shows the medical images that were used to validate the proposed algorithm. Fig. 3a and 3b are the images of a human brain obtained using computed tomography (CT) and magnetic resonance imaging (MRI), respectively. These images are available online at [36]. CT and MRI sensors provide complementary information, with CT proving a good visual description of bone tissue, whereas soft tissues are better visualized by MRI. To simulate the sensor noise, zero mean white Gaussian noise were added to these input images. The noisy input images were fused using the least square as well as proposed approach. For both the approaches for fusion, CT and MRI images were divided into non-overlapping regions of size 8×8 and for each region, the sensor gain was computed using the principal eigenvector approach explained in Section III-B. For least square based fusion, the estimated sensor gain was used in eq. (3) to estimate the fused images. To fuse these images using the proposed approach, the estimated sensor gains were used in the total variation algorithm described in Section III-A. Several different levels of SNR were used to validate the robustness of the algorithm. Three sets of fusion results for 23 dB, 12 dB and 0 dB signal to noise ratio (SNR) have been presented in Fig. 4-6. It is clear that for high level of SNR (Fig. 4), the performance of least square and the proposed approach is similar. However, as SNR decreases, proposed approach performs better than the least square approach as shown in Fig. 5 and 6. For very high levels of noise, *i.e.*, low SNR, the proposed algorithm oversmoothes the edges of the fused image as shown in Fig. 6.

The proposed algorithm was also applied to the aircraft navigation images [36] shown in Fig. 7. Fig. 7a was captured using a low-light-television (LLTV) sensor and Fig. 7b was obtained using a forward-looking-infrared (FLIR) sensor. Note that LLTV sensor provides the surface information of the ground as well building and vegetation details around it. However, FLIR image describes the road network details accurately. The goal of fusion is to have all the salient features of both the sensors in the fused image. The sensor noise simulation and the fusion process was exactly the same as that of the previous dataset, and Fig. 8-10 show the fusion results for 23 dB, 8 dB and 0 dB signal to noise ratio (SNR). It is clear from the fused images that the proposed approach is more robust to noise as compared to the least square approach. However, for low SNR, the proposed fusion algorithm tends to oversmooth the edges of the fused images.

To estimate the quality of the fused image quantitatively, the similarity measure based quality index proposed in [37] was used. In this index, a similarity metric is computed between the input images and the fused image. This index lies between -1 and $+1$, with index values close to ± 1 indicating the robustness of the fusion algorithm. The goal of quality assessment is to compare the information content in the fused image and the corresponding input images. Therefore, the similarity index was computed by comparing

the fused images with the noiseless versions of the corresponding input images. The similarity index for the medical and navigation datasets are summarized in Table I, and plots of the similarity quality index for LSE and the proposed approach as a function of SNR are presented in Fig. 11. It is evident from both the Table I as well Fig. 11 that for very high SNR, performance of both the LSE and the proposed TV norm based approaches is comparable. However, as the SNR is reduced, the proposed approach performs better than the LSE. It is apparent that for both approaches, similarity index value approaches zero as SNR drops.

VI. Conclusions

A total variation framework for pixel level image fusion is proposed in this paper. The proposed fusion approach was applied to several different types of datasets. The results on these data indicate the feasibility of the proposed approach. Future work will focus on analysis of the algorithm performance with additional datasets.

References

[1] M. Kokar and K. Kim, "Review of multisensor data fusion architectures," in *Proceeding of the IEEE International Symposium on Intelligent Control*, August 1993, pp. 261–266.

[2] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE*, vol. 85, no. 1, pp. 6–23, 1997.

[3] E. Waltz, "The principles and practice of image and spatial data fusion," in *Proc. 8th Natl. Data Fusion Conf.*, Dallas, TX, March 1995.

[4] R. Blum, Z. Xue, and Z. Zhang, *An overview of image fusion, in: R. Blum, Z. Liu (Eds.), Chapter in the book Multi-Sensor Image Fusion and Its Applications*. Signal and Image Processing Series of Marcel Dekker/CRC Press.

[5] Z. Zhang and R. Blum, "A categorization and study of multiscale-decomposition-based image fusion schemes," *Proceedings of the IEEE*, vol. 87, no. 8, pp. 1315–1326, August 1999.

[6] R. S. Blum, "Robust image fusion using a statistical signal processing approach," *Information Fusion*, vol. 6, no. 2, pp. 119–128, June 2005.

[7] Y. Chibani, "Multisource image fusion by using the redundant wavelet decomposition," *IEEE Geoscience and Remote Sensing Symposium*, vol. 2, pp. 1383–1385, July 2003.

[8] M. Hurn, K. Mardia, T. Hainsworth, J. Kirkbride, and E. Berry, "Bayesian fused classification of medical images," *IEEE Trans. Med. Imag.*, vol. 15, no. 6, pp. 850–858, December 1996.

[9] X. Gros, Z. Liu, K. Tsukada, and K. Hanasaki, "Experimenting with pixel-level ndt data fusion techniques," *IEEE Trans. Instrum. Meas.*, vol. 49, no. 5, pp. 1083–1090, October 2000.

[10] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, "A comparative analysis of image fusion methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1391– 1402, June 2005.

[11] J. Zhou and D. L. Civco, "A wavelet transform method to merge landsat tm and spot panchromatic data," *International Journal of Remote Sensing*, vol. 19, no. 4, pp. 743–757, March 1998.

[12] F. Nebeker, "Golden accomplishments in biomedical engineering," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 3, pp. 17–47, May/June 2002.

[13] D. Barnes, G. Egan, G. OḰeefe, and D. Abbott, "Characterization of dynamic 3-d pet imaging for functional brain mapping," *IEEE Trans. Med. Imag.*, vol. 16, no. 3, pp. 261–269, June 1997.

[14] S. Wong, R. Knowlton, R. Hawkins, and K. Laxer, "Multimodal image fusion for noninvasive epilepsy surgery planning," *International Journal of Remote Sensing*, vol. 16, no. 1, pp. 30–38, January 1996.

[15] D. Hal, *Mathematical Techniques in Multisensor Data Fusion*, 2nd ed. Boston, USA: Artech House, March 2004.

[16] J. Llinas and D. Hall, "A challenge for the data fusion community ii: Research imperatives for improved processing," in *Proc. 7th Natl. Symp. On Sensor Fusion*, Albuquerque, New Mexico, March 1994.

[17] C. Pohl and J. L. van Genderen, "Multisensor image fusion in remote sensing: concepts, methods and applications," *International Journal of Remote Sensing*, vol. 19, no. 5, pp. 823–854, 1998.

[18] M. Daniel and A. Willsky, "A multiresolution methodology for signal-level fusion and data assimilation with applications to remote sensing," *Proc. IEEE*, vol. 85, no. 1, pp. 164–180, January 1997.

[19] B. Dasarathy, "Fuzzy evidential reasoning approach to target identity and state fusion in multisensor environments," *Optical Engineering*, vol. 36, no. 3, pp. 669–683, March 1997.

[20] J. Byeungwoo and D. Landgrebe, "Decision fusion approach for multitemporal classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1227–1233, May 1999.

[21] E. Lallier and M. Farooq, "A real time pixel-level based image fusion via adaptive weight averaging," in *Proceedings of the Third International Conference on Information Fusion*, vol. 2, no. 10, Paris, France, July 2000, pp. WEC3/3 – WEC313.

[22] J. Liu, Q. Wang, and Y. Shen, "Comparisons of several pixel-level image fusion schemes for infrared and visible light images," in *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, vol. 3, Ottawa, Canada, May 2005, pp. 2024– 2027.

[23] G. Pajares and J. D. L. Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognition*, vol. 37, no. 9, pp. 1855–1872, September 2004.

[24] P. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *Fourth Int. Conf. on Computer Vision*, 1993, pp. 173–182.

[25] P. Burt, "A gradient pyramid basis for pattern-selective image fusion," in *Proceedings of the Society for Information Display*, 1992, pp. 467–470.

[26] A. Toet, "Hierarchical image fusion," *Machine Vision and Applications*, vol. 3, pp. 1–11, 1990.

[27] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[28] P. L. Combettes and J.-C. Pesquet, "Image restoration subject to a total variation constraint," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1213–1222, 2004.

[29] T. Chen, Y. Wotao, S. Z. Xiang, S. D. Comaniciu, and T. S. Huang, "Total variation models for variable lighting face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1519–1524, 2006.

[30] G. A. Hewer, C. Kenney, and B. S. Manjunath, "Variational image segmentation using boundary functions," *IEEE Trans. Image Process.*, vol. 7, no. 9, pp. 1269–1282, 1998.

[31] R. K. Sharma, T. K. Leen, and M. Pavel, "Probabilistic image sensor fusion," *Advances in Neural Information Processing Systems 11*, 1999, the MIT Press.

[32] J. Yang and R. S. Blum, "A statistical signal processing approach to image fusion for concealed weapon detection," vol. 1, pp. 513–516, 2002.

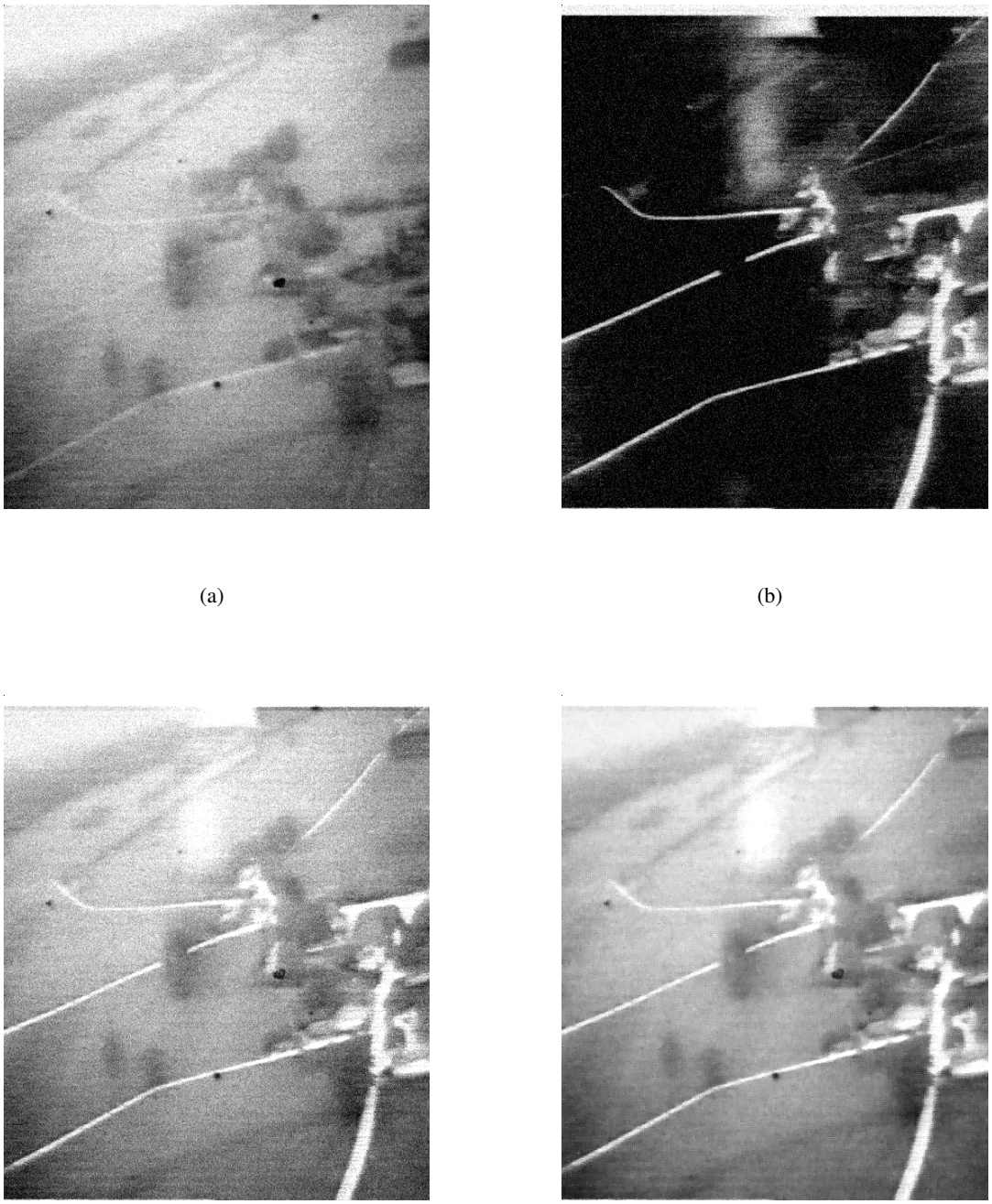
[33] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel, "Axioms and fundamental equations of image processing," *Archive for Rational Mechanics and Analysis*, vol. 123, pp. 199–257, 1993.

[34] L. Alvarez and J.-M. Morel, "Formalization and computational aspects of image analysis," *Archive for Rational Mechanics and Analysis*, pp. 1–59, 1994.

[35] A. C. Rencher, *Methods of Multivariate Analysis*, 2nd ed. Wiley-Interscience, March 2002.

[36] [Online]. Available: http://www.imagefusion.org/

[37] N. Cvejic, A. Loza, D. Bull, and N. Canagarajah, "A similarity metric for assessment of image fusion algorithms," *International Journal of Signal Processing*, vol. 2, no. 3, pp. 178–182, 2005.



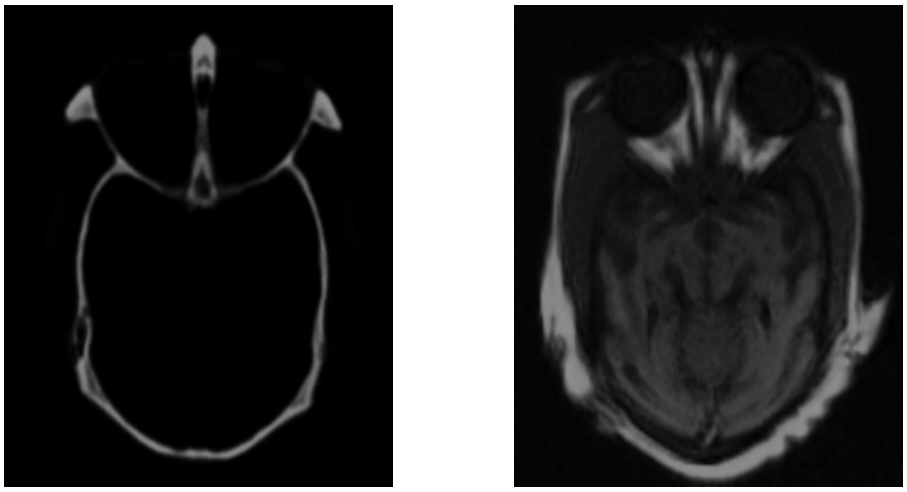
(a) (b)

(c) (d)

Fig. 1: Comparison of least square and TV norm based fusion(a) LLTV Image (b) FLIR Image (c) Fused Image: Least Square (d) Fused Image: Proposed Approach (TV Norm)

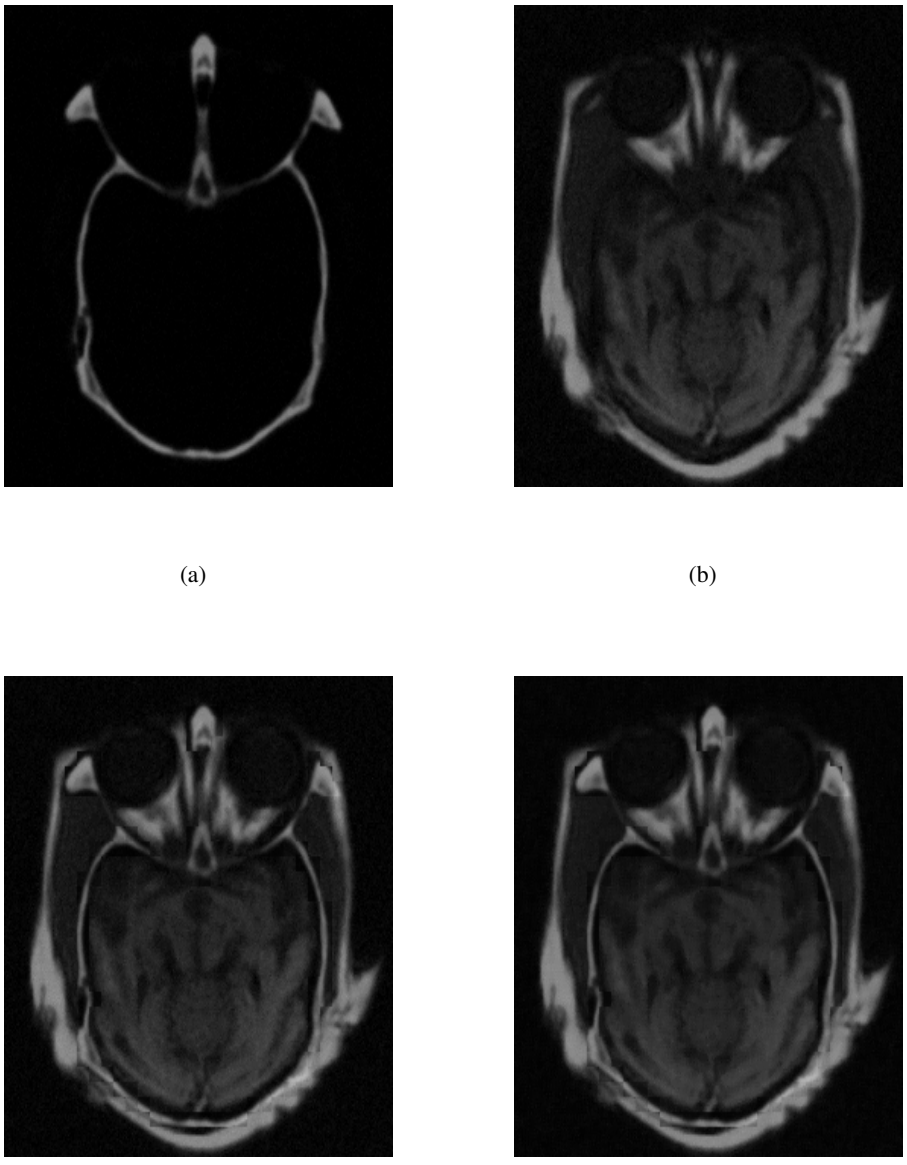
Fig. 2: Flowchart of the Proposed Image Fusion

- **Input:** n co-registered sensor images
- **Output:** Fused image
- **Sensor Gain Estimation:**
  - –) Split each input image into B blocks each of size $p \times q$
  - –) For each block $(i) : i = 1, \cdots, B$
    - Estimate sensor gain for each block using algorithm presented in Section III-B
- **Fusion Process:**
  - –) Solve eq. (8) iteratively for each pixel.



(a) (b)

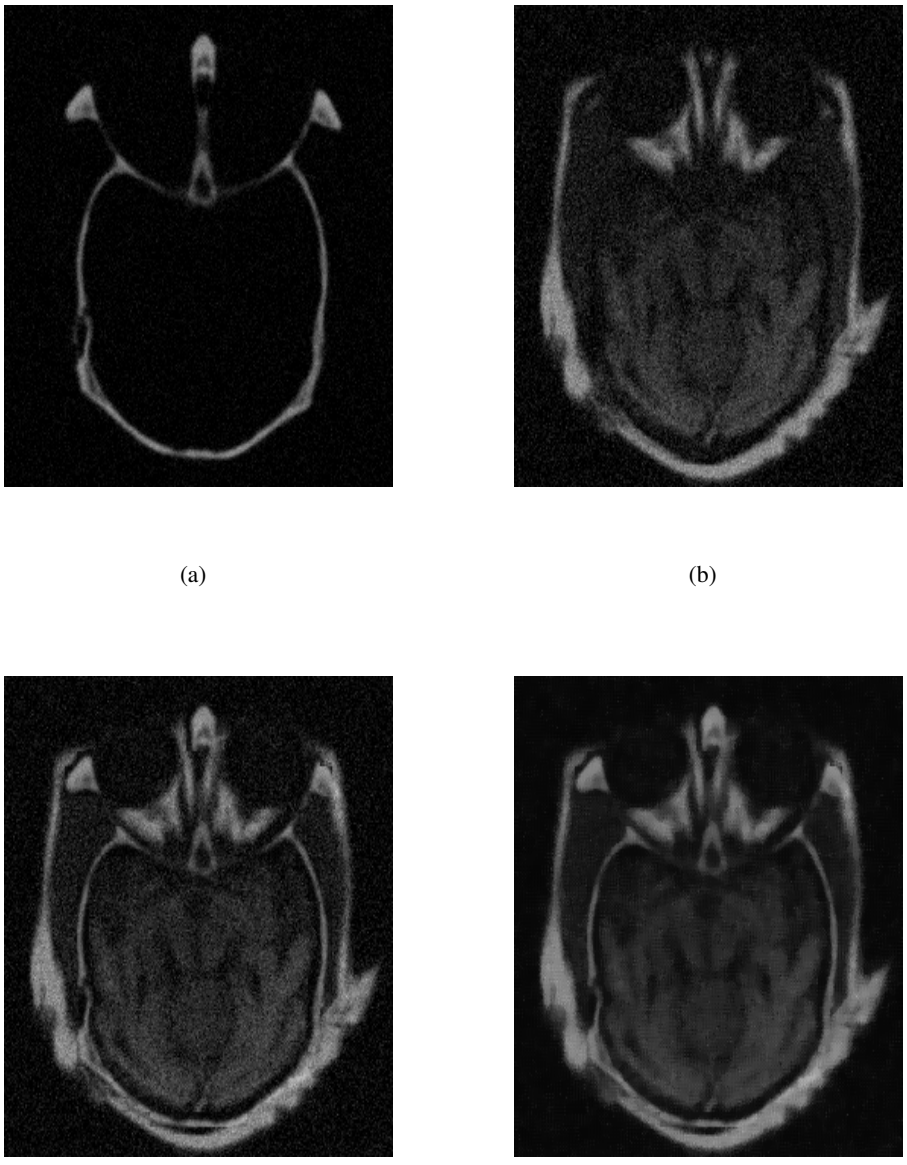
Fig. 3: Medical Images: (a) CT Image (b) MRI Image



(a) (b)

(c) (d)

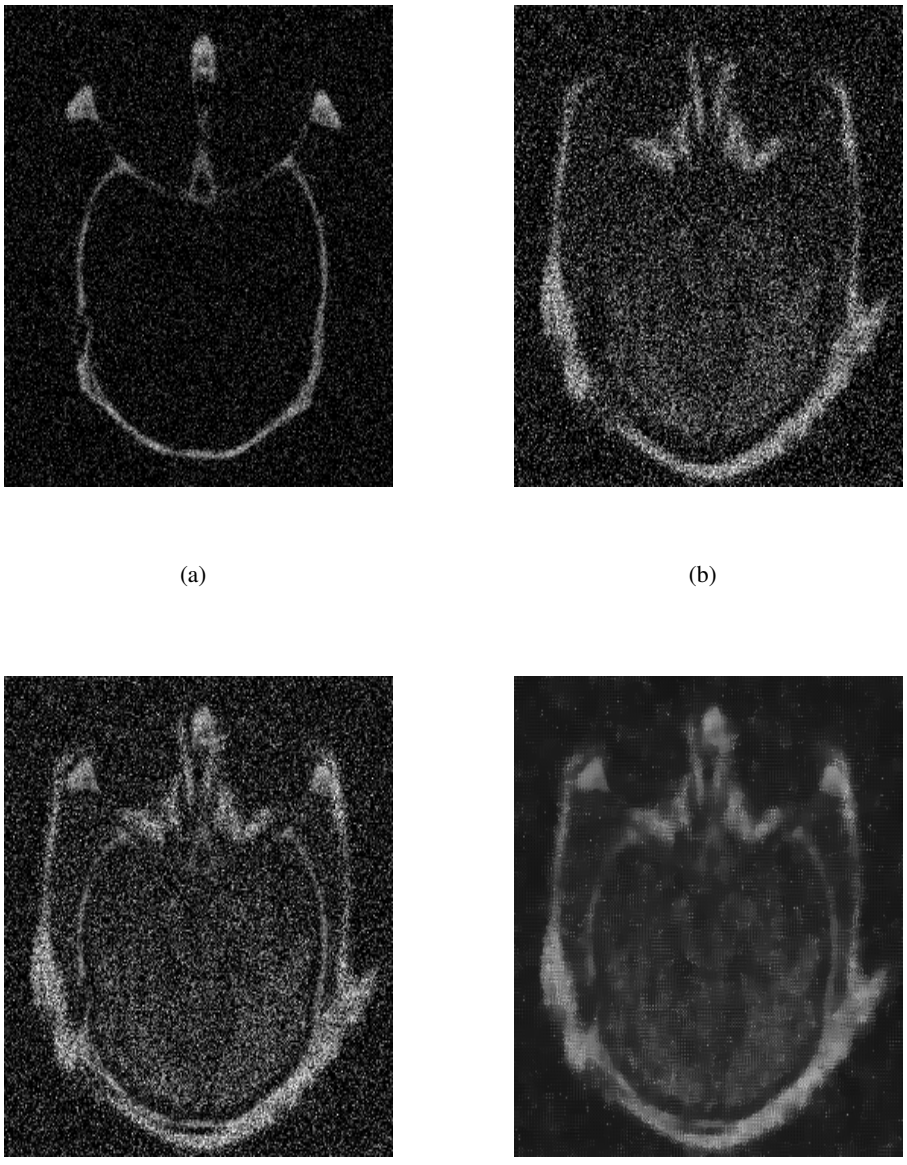
Fig. 4: Medical Images, SNR = 23 dB: (a) CT Image (b) MRI Image (c) Fused Image: Least Square (d) Fused Image: Proposed Approach



(a) (b)

(c) (d)

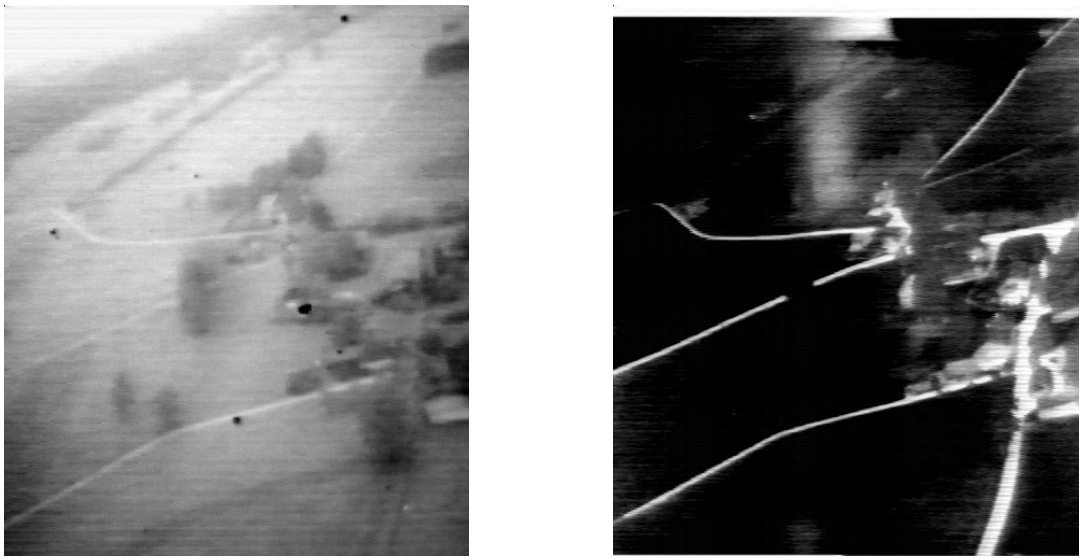
Fig. 5: Medical Images, SNR = 12 dB: (a) CT Image (b) MRI Image (c) Fused Image: Least Square (d) Fused Image: Proposed Approach



(a) (b)

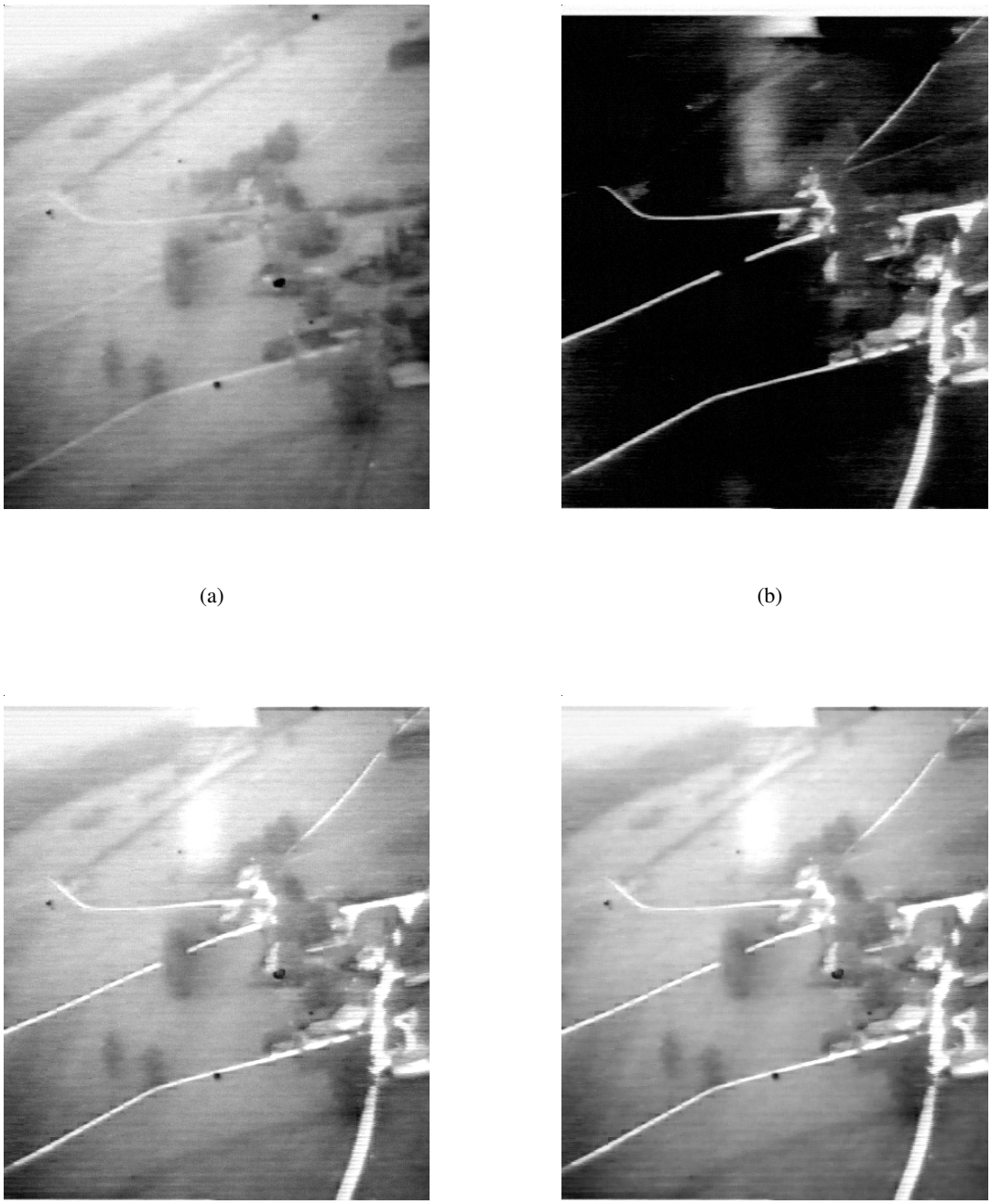
(c) (d)

Fig. 6: Medical Images, SNR = 0 dB: (a) CT Image (b) MRI Image (c) Fused Image: Least Square (d) Fused Image: Proposed Approach



(a) (b)

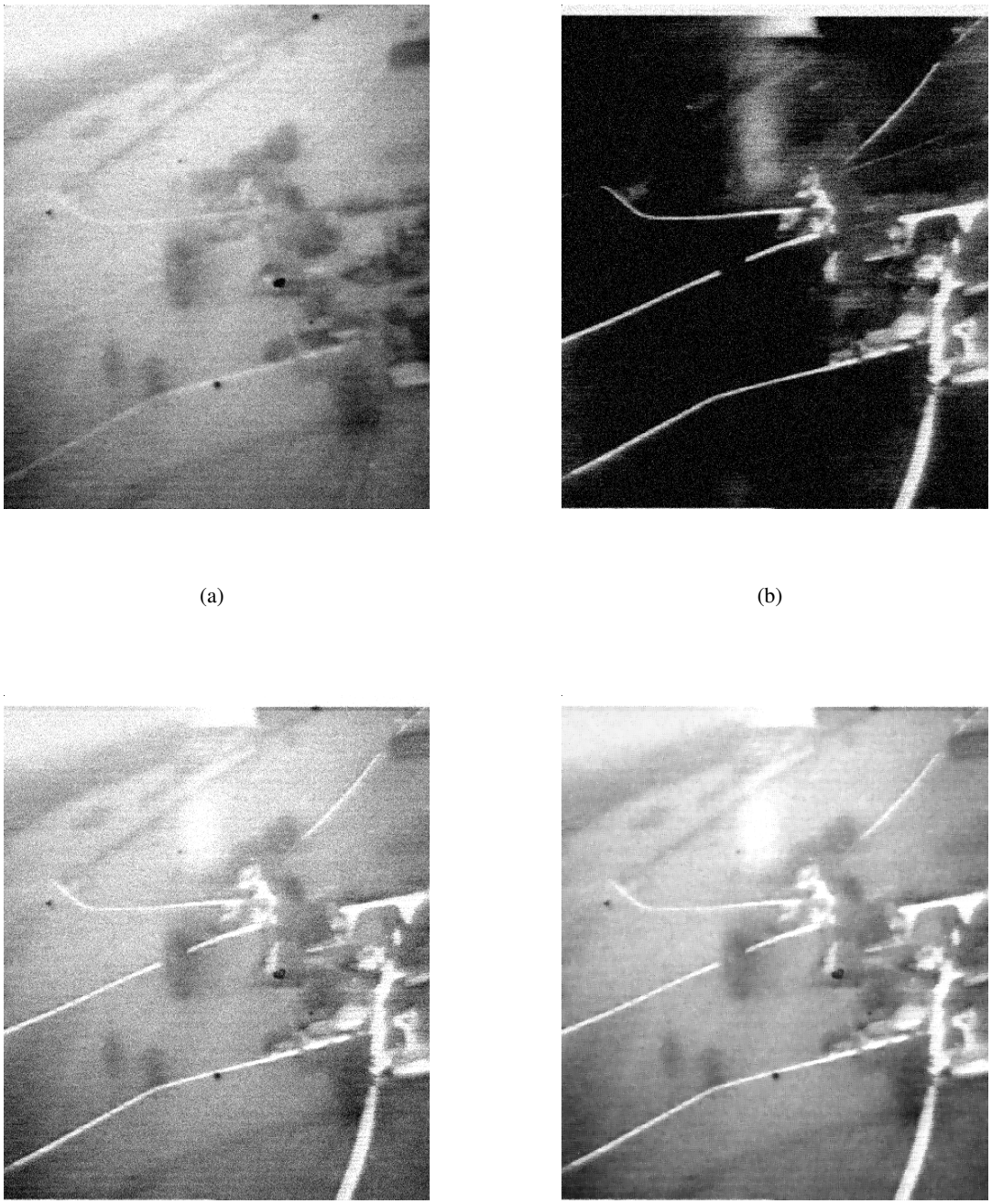
Fig. 7: Aircraft Navigation Images: (a) LLTV Image (b) FLIR Image



(a) (b)

(c) (d)

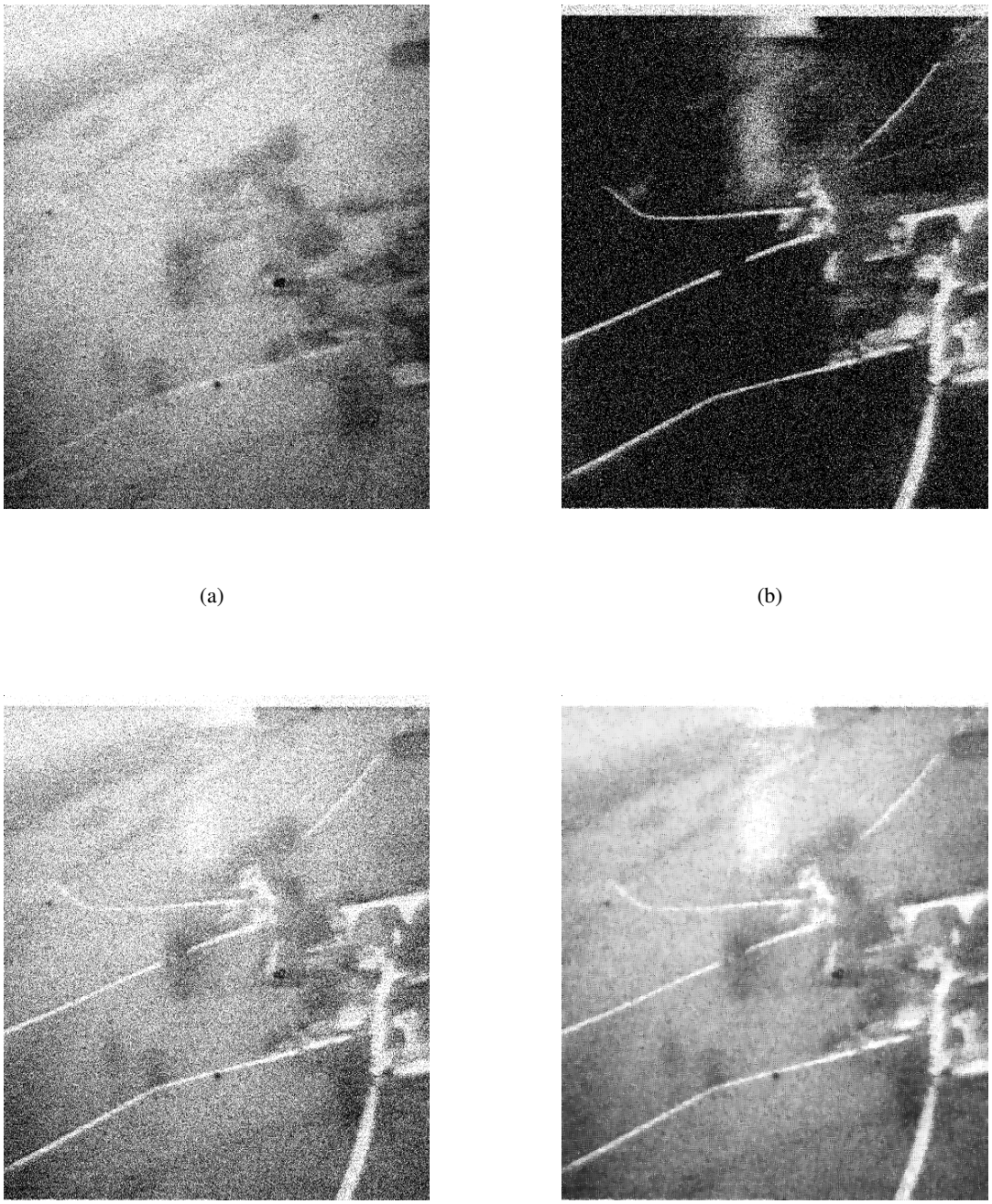
Fig. 8: Aircraft Navigation Images, SNR = 23 dB: (a) LLTV Image (b) FLIR Image (c) Fused Image: Least Square (d) Fused Image: Proposed Approach



(a) (b)

(c) (d)

Fig. 9: Aircraft Navigation Images, SNR = 8 dB: (a) LLTV Image (b) FLIR Image (c) Fused Image: Least Square (d) Fused Image: Proposed Approach



(a) (b)

(c) (d)

Fig. 10: Aircraft Navigation Images, SNR = 0 dB: (a) LLTV Image (b) FLIR Image (c) Fused Image: Least Square (d) Fused Image: Proposed Approach

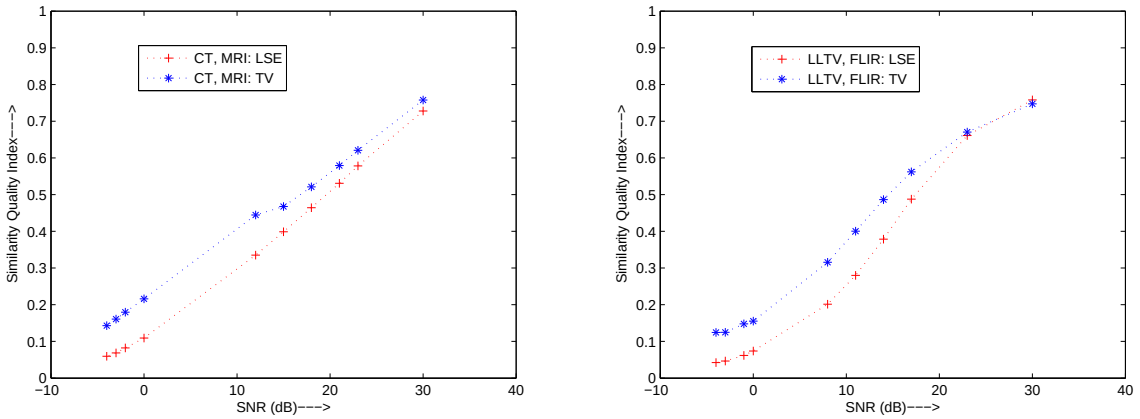


Fig. 11: Plot of SNR vs. Similarity Quality Index (a) CT, MRI dataset (b) LLTV, FLIR dataset

TABLE I: Performance summary of the proposed image fusion algorithm

Image DataSet	**SNR (dB): Sensor 1**	**SNR (dB): Sensor 2**	**Similarity Index: LSE**	**Similarity Index: Proposed Approach**
Medical Image	23	23	0.58	0.62
	12	12	0.34	0.44
	0	0	0.11	0.22
Aircraft Navigation	23	23	0.66	0.67
	8	8	0.20	0.32
	0	0	0.07	0.16