Fusion of 2D grayscale images using multiscale morphology

Susanta Mukhopadhyay, Bhabatosh Chanda*

Electronics and Communication Sciences Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700 035, India

Received 29 January 1999; accepted 20 July 2000

Abstract

A scheme for fusion of multi-sensor 2D images based on multiscale morphology is presented in this paper. A point-based registration, using affine transformation is performed prior to fusion. The scale-specific features are extracted from both the images using the morphological towers constructed in course of various types of multiscale filtering. Extracted features are combined to get the fused image. A quantitative measure of the degree of fusion is estimated by cross-correlation coefficient and the error measure obtained by eigenvector fitting between the fused image and each of the constituting images.
reserved.

Keywords: Mathematical morphology; Multiscale morphology; Morphological towers; Registration; Fusion of images; Performance evaluation of fusion algorithm

1. Introduction

In a multisensor data acquisition system the image data of an object consists of information acquired by different sensors from different perspective and possibly at different resolution. The clarities of the object features may be different in different imaging modalities. For example, in the arena of biomedical imaging, two widely used modalities, namely the magnetic resonance imaging (MRI) and the computed tomographic (CT) scan do not reveal identically every detail of brain structure. While CT scan is especially suitable for imaging bone structure and hard tissues, the MR images are much superior in depicting the soft tissues in the brain that play very important roles in detecting diseases affecting the skull base. These two imaging modalities are thus complementary in many ways and no one is totally sufficient in terms of their respective information content. Thus images resulting from a single modality cannot meet the clinical requirement for the purpose of diagnosis and treatment. However, viewing a series of multimodal images of the same object seperately and individually does not provide a better solution because of such inconvenience.

The advantages of multisensor data may be fully exploited by integrating the complementary features seen in the images generated by several sensors. A process of such integration, widely termed as *fusion*, generates an image composed of features that are best detected or represented in the individual modalities. A technically fused image can render itself more successfully for any subsequent processing like object recognition, feature extraction, segmentation, etc. in comparision to images of individual modalities. Image data fusion is a widely demanded work in the field of multisensor data acquisition and interpretation systems.

The first step toward fusion, which may be interpreted as a preprocessing step is the *registration*. Registration is a process of bringing down the constituting images to a common coordinate system subject to some common geometrical references called *ground control points* (GCPs). A feature or object common to all the modalities may have different position and scale in different modalities. Fusion of such images is meaningful only when such common objects (or control points) are made to have identical geometric configuration with respect to size,

location and orientation in all the images. In the next step, the images are combined to form a single fused image through a judicious selection of proportions of different features from different images.

Many researchers have been working in the field of image fusion in connection to biomedical imaging, machine vision, remote sensing applications in defence and atmospheric fields. Li et al. [1] have suggested a multisensor image fusion using wavelet transform in which a cascaded sequence of forward and reverse wavelet transform on multimodal images produces a fused image. Matsopoulos et al. [2] have devised a hierarchical image fusion scheme that integrates features extracted from morphological pyramids of the multimodal images. Generally, data fusion in multimodal images [3–5] is partly redundant as some regions are depicted in all the modalities and is partly complementary as each modality highlights certain features that are absent in images of other modalities. There are various techniques and approaches for fusing multimodal images. Many methods are based on classification of pixels [6]. In this direction Hurn et al. [7] have suggested a Bayesean probabilistic method for biomedical image fusion. While Mukherjee et al. [8] have developed an algorithm for fusing CT and MR images based on entropy.

This paper discusses a new technique of multimodal image fusion based on multiscale morphology. The developed scheme is applied to fuse MR and CT images of brain and the result is compared with fused images resulting from other existing fusion schemes. The proposed method is a feature-based fusion scheme which involves integration of features of different scales coming from a number of morphological towers constructed in course of the process itself. Section 2 discusses a point-based method of image registration adopted as a preprocessing step prior to fusion. In Section 3, we have given a brief introduction to multiscale morphology. The proposed multiscale morphological scheme for fusing multimodal images is described in Section 4. In our experiment we have taken a pair of CT and MR images of human brain as test images. This is followed by Section 5 containing the experimental results and discussions. Concluding remarks are cited in Section 6.

2. Registration

Any kind of image fusion method involves registration as a first and foremost step. In general, different sensors respond to scene characteristics in different and, partially, complementary ways. In a multisensor image acquisition system, the size, orientation and location of an object relative to its own background may not be identical in all the images of different modalities. Integration or fusion of multisensor information is possible only if the images are registered or positioned with respect to a common coordinate system [9]. Image registration (in case of fusing two images) is the process of determining correspondence between all points in two images of the same scene or object. There are various techniques for registering multimodal images [10]. Some of which are area based while some are point based. Point-based methods [11,12], are simpler and more common. These points, i.e. GCPs, are basically spatial features (e.g., corners, junctions, centroid of blobs, etc.) present in the image. The GCPs can be made available by applying standard feature extraction algorithms [13]. Brown [10] made a good survey on image registration.

In a point-based registration scheme a set of GCPs is selected from the images, correspondence is established between them, and from that point correspondence a transformation function is determined. Linear methods use either the affine transformation or the geometric transformation like translation, rotation and scaling so as to minimize the distance measure or to maximize the similarity between these points [14]. Finally, this transformation function is used to map points in one image to points in the other. Note that in this approach more emphasis is given to latter image in reconfiguring the position of the object(s). In this paper we have adopted a point-based method for registering the multimodal images. However, we have given equal emphasis to both the images. Two sets of ground control points from two images are selected. The target image coordinate system is assumed to be intermediate between the given images. Then two affine transformations determined based on these GCPs and following the treatment suggested by Mardia et al. [15,16] are applied on corresponding images. Thus, the adopted method is unbiased to both the images. The parameters of the transformation(s) are determined through exhaustive search for the minimum error.

3. Multiscale morphological operations

Mathematical morphology is a powerful technique in the field of image processing and computer vision. In morphology, the objects in an image are considered as set of points and operations are defined between two sets: the object and the structuring element (SE) [17,18]. The shape and the size of SE are defined according to the purpose of the concerned application. Basic morphological operations are erosion and dilation. Other operations like opening, closing are sequential combination of erosion (dilation) and dilation (erosion). We adopt, here, *function- and set-processing* (FSP) system [19]. FSP dilation of a gray-level image $f(x, y)$ by a two-dimensional set B is defined as

$$(f \oplus B)(x, y) = \max\{ f(x - k, y - l) | (k, l) \in B \}. \quad (1)$$

Similarly, FSP erosion of $f(x, y)$ by B is defined as

$$(f \ominus B)(x, y) = \min\{ f(x + k, y + l) | (k, l) \in B \}. \quad (2)$$

The shape of the structuring element B plays a crucial role in extracting features or objects of given shape from the image. However, for a categorical extraction of features or objects based on shape as well as size, we must incorporate a second attribute to the structuring element which is its *scale*. A morphological operation with a scalable structuring element can extract features based on shape and size simultaneously. Also features of identical shape but of different size are now treated separately. Such a scheme of morphological operations where a structuring element of varying scale is utilized is termed as *multiscale morphology* [20,21]. Multiscale opening and closing are defined, respectively, as

$$(f \circ nB)(x, y) = ((f \ominus nB) \oplus nB)(x, y), \quad (3)$$

$$(f \bullet nB)(x, y) = ((f \oplus nB) \ominus nB)(x, y), \quad (4)$$

where B is a set representing structuring element of a definite shape while n is an integer representing the scale factor of the structuring element. Thus, given $B = \{(x, y)\}$, $nB = \{(nx, ny)\}$ for $n = 1, 2 \ldots$, represents a family of structuring elements. If B is convex, in discrete domain, we obtain nB using Eq. (5):

$$nB = \underbrace{B \oplus B \oplus B \oplus \cdots \oplus B}_{n-1 \text{ times}}. \quad (5)$$

By convention $nB = \{(0, 0)\}$ when $n = 0$.

4. Image fusion using multiscale morphology

We propose a multiscale morphological method of fusing two images of same objects obtained through two different modalities. The *bright* (*dark*) *tophat transformation* originally proposed in Ref. [22] provides an excellent tool for extracting bright (respectively, dark) features smaller than a given size from an uneven background. It relies on the fact that by gray-scale opening, one can remove from an image the brighter areas, i.e. features, that cannot hold the structuring element. Subtracting the opened image from the original one gets another image which comprises of only those features that have been removed by the SE in the opening operation. Similar thing holds good for morphological closing also. Subtracting the original image from the closed image one can extract dark features against a brighter background. Here the structuring element used in both opening and closing is a disk, or, more specifically, a discrete approximation of disk. An ordered sequence of morphological filtering (i.e., opening or closing) of both the image modalities with a disk structuring element at different scales extracts scale-specific features from both the images. These scale-specific features resulting from various image modalities are then compared for the selection of the prominent one for subsequent construction of the fused image. The proposed method is described elaborately in the following subsection.

4.1. One-dimensional case

For a better understanding of the fusion scheme we first elaborate it in the context of one-dimensional functions as shown in Fig. 1. Fig. 1 provides a simplified illustration of the proposed strategy. Here we intend to fuse two functions $f_1(t)$ and $f_2(t)$. Both the functions $f_1(t)$ and $f_2(t)$ have salient features manifested by *crests* and *troughs* of different height (*or depth*) and width located at different position. We use a line segment L of unit length and its higher order dilates kL (where $k = 1, 2, 3 \ldots$) as structuring elements (SE) for extracting the salient features from both the functions as described below:

- Opening by kL removes the crests narrower than length k from both the functions.
- The function $g_k^{O1}(t) = (f_1(t) \circ (k-1)L - f_1(t) \circ kL)$ contains the crests (see Fig. 1) of $f_1(t)$ which are narrower than k but wider than $(k-1)$. Similarly, the function $g_k^{O2}(t) = (f_2(t) \circ (k-1)L - f_2(t) \circ kL)$ contains only the crests of $f_2(t)$ which are narrower than k but wider than $(k-1)$.
- The function $h_k^O(t) = \max(g_k^{O1}(t), g_k^{O2}(t))$ constructed by comparing $g_k^{O1}(t)$ and $g_k^{O2}(t)$ at each sample point t contains the crests from both $f_1(t)$ and $f_2(t)$. Thus, the crests of width lying in between $(k-1)$ and k coming from $f_1(t)$ and $f_2(t)$ are combined in the function $h_k^O(t)$.

The function $S_{op}(t)$ constructed by superposing (*or, simply adding*) the functions $h_k^O(t)$ for $k = 1, 2, 3 \ldots$ contains all the crests of both the functions.

The same set of operations with closing helps us extract the troughs from both the functions and combine them into a single function $S_{cl}(t)$.

Opening the functions with the largest homothetic of the SE L (i.e. kL) removes all the crests from the functions leaving only their locally non-varying parts. While combining we must avoid clipping the highest crests. We therefore consider the minimum of $f_1(t) \circ nL$ and $f_2(t) \circ nL$, where n is the maximum value of k used in the operation. In this example $n = 2$. Let us denote it by $M_N(t)$.

Similarly, in case of closing, we consider the maximum of $f_1(t) \bullet nL$ and $f_2(t) \bullet nL$ for subsequent reconstruction. Let us denote it by $M_X(t)$.

Now, there should be an equal justice to all the crests and troughs of both the functions. Consequently, $M_N(t)$ and $M_X(t)$ should be given equal importance too. With such view we take the average $A(t)$ of $M_N(t)$ and $M_X(t)$ to

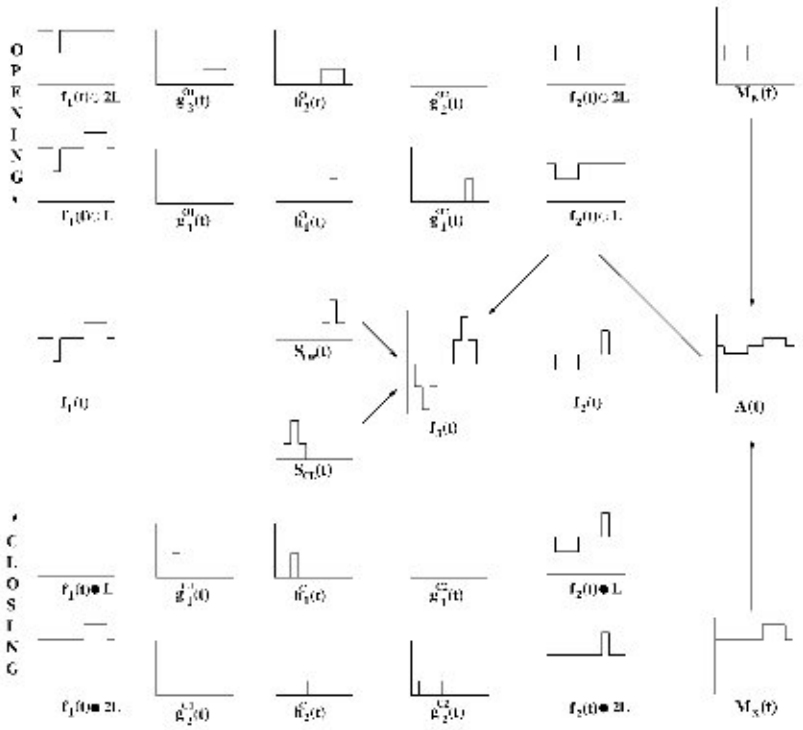


Fig. 1. Fusion of two functions using multiscale morphology.

construct the locally non-varying profile of the fused function. Finally, we form the fused function $f_3(t)$ using Eq. (6). Eq. (6) performs the placement of the crests and troughs at proper locations of the average function:

$$f_3(t) = A(t) + \tfrac{1}{2}S_{op}(t) - \tfrac{1}{2}S_{cl}(t). \tag{6}$$

Thus the fused function $f_3(t)$ is found to contain crests and troughs from both the functions $f_1(t)$ and $f_2(t)$. The scheme explained for this one-dimensional case may as well be extended to two dimension. There we introduce the concepts of different *morphological towers* as discussed in the following sections.

4.2. Construction of morphological towers

The fusion algorithm proposed in this paper is independent of imaging technology. However, in this paper we have described the method for fusing CT and MR images of cross-section of human brain. The CT as well as the MR image are made to undergo a sequence of morphological opening and closing operations with SEs of progressively increasing scale. As an extension of the one-dimensional case to make the SE isotropic in two dimension, here we consider a family of filled circles or *disk*s of increasing radius as structuring elements. For generating the family of filled circles we made use of *Bresenham's* algorithm [23].

We then construct two sets of morphological towers for multiscale opening and closing operations. In each set we have two different towers corresponding to MR and CT images. A *tower* basically comprises of a stack of images produced after morphological opening or closing operations on an image for different values of the scale factor i of the structuring element as shown in Fig. 2. Morphological towers can be used to solve some well-known problems of image processing and computer vision [24,25]. The prime objective is extraction of scale-specific bright and dark features from the images. Thus the ith entry in the multiscale opening tower of the CT/MR image contains the image produced by opening the CT/MR image with a structuring element iB. So, altogether there are four such towers of height n for multiscale opening and closing operations of the CT and the MR image, that are defined as

$$\mathrm{CT} \circ iB = (\mathrm{CT} \ominus iB) \oplus iB, \tag{7}$$

$$\mathrm{MR} \circ iB = (\mathrm{MR} \ominus iB) \oplus iB, \tag{8}$$

$$\mathrm{CT} \bullet iB = (\mathrm{CT} \oplus iB) \ominus iB, \tag{9}$$

$$\mathrm{MR} \bullet iB = (\mathrm{MR} \oplus iB) \ominus iB \tag{10}$$

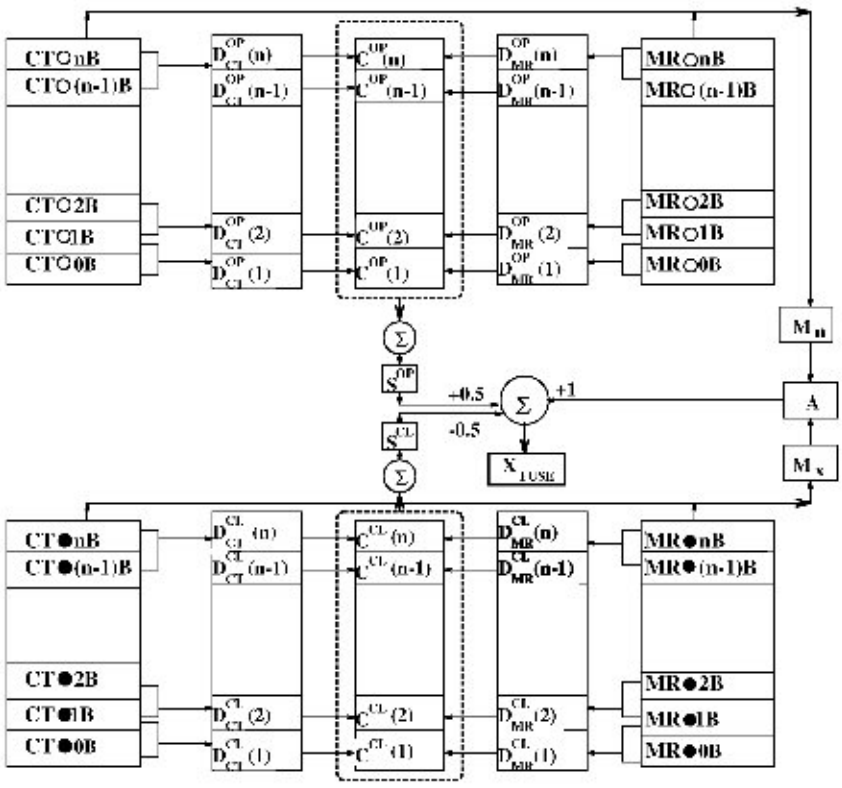


Fig. 2. Morphological towers for image fusion.

for $i = 0,1,\ldots,n$ where $\mathrm{CT} \circ 0B = \mathrm{CT}$, $\mathrm{MR} \circ 0B = \mathrm{MR}$, $\mathrm{CT} \bullet 0B = \mathrm{CT}$ and $\mathrm{MR} \bullet 0B = \mathrm{MR}$. Note that B is a structuring element of unit size.

4.3. Construction of difference towers

As stated earlier the image resulting after morphological opening operation using a structuring element iB contains only those features of the original image (CT or MR) which are at a scale equal to or larger than i. Likewise the image resulting after a morphological opening operation using a structuring element $(i+1)B$ contains all those features of the original image (CT or MR) which are at a scale equal to or larger than $(i+1)$. Thus, a difference of these two resulting images gives rise another image which contains only those features of the original image which are at a scale greater than or equal to i but less than $(i+1)$. This holds good for the multiscale closing operations of the images also.

With such views we construct four difference towers for opening and closing operations of the CT and the MR images by carrying out difference operations between all successive pairs of images resulting after morphological opening (closing) operations using the structuring elements corresponding to two successive scales. Fig. 2 shows four such *difference towers*. Thus the ith entries in the difference tower for morphological operations are defined as

$$D^{op}_{\mathrm{CT}}(i) = (\mathrm{CT} \circ (i-1)B) - (\mathrm{CT} \circ iB), \tag{11}$$

$$D^{op}_{\mathrm{MR}}(i) = (\mathrm{MR} \circ (i-1)B) - (\mathrm{CT} \circ iB), \tag{12}$$

$$D^{cl}_{\mathrm{CT}}(i) = (\mathrm{CT} \bullet iB) - (\mathrm{CT} \bullet (i-1)B), \tag{13}$$

$$D^{cl}_{\mathrm{MR}}(i) = (\mathrm{MR} \bullet iB) - (\mathrm{MR} \bullet (i-1)B) \tag{14}$$

for $i = 1, 2, \ldots, n$.

4.4. Construction of combined tower

The scale-specific features from both the image modalities are now available in the difference towers constructed in the previous step. For constructing the fused image from the CT and the MR images, features from both of them should be combined. At each scale the feature present in at least one image modality should be collected for fusion. However, if the feature is present in both the images we select the prominent one for constructing the fused image. With such views we extract the scale-specific

features from difference towers for opening (closing) operation by taking pixel-wise max of two images as shown in Fig. 2. The set of images obtained through max operation result in two more towers called the *combined towers*. The ith entries of combined tower corresponding to opening and closing operations are basically the images produced, respectively, by

$$C^{op}(i) = \max\{D^{op}_{CT}(i), D^{op}_{MR}(i)\}, \tag{15}$$

$$C^{cl}(i) = \max\{D^{cl}_{CT}(i), D^{cl}_{MR}(i)\} \tag{16}$$

for $i = 1, 2, \ldots, n$. Here max stands for a pixel-wise maximum of two images.

4.5. Reconstruction

For reconstructing the final image we do the following:

- We sum up all the entries in the combined tower corresponding to the opening operation. This results in an image consisting of bright features of all possible scales that are present in at least one modality with relatively better clarity:

$$S^{op} = \sum_{i=1}^{n} C^{op}(i). \tag{17}$$

The summation, here, denotes pixel-wise sum of n images.

- We perform the same operation on the combined tower corresponding to the closing operation. This results in an image consisting of dark features of all possible scales that are present in at least one modality with relatively better clarity:

$$S^{cl} = \sum_{i=1}^{n} C^{cl}(i). \tag{18}$$

- We take the pixel-wise min (max) operation between the CT and the MR images after opening (closing) them with nB. Then we take the average of these two resulting images:

$$M_X = \max\{\mathrm{CT} \bullet nB, \mathrm{MR} \bullet nB\}, \tag{19}$$

$$M_N = \min\{\mathrm{CT} \circ nB, \mathrm{MR} \circ nB\}, \tag{20}$$

$$A = \mathrm{average}(M_X, M_N). \tag{21}$$

Here *average* denotes the pixel-wise average of two images. Finally, the fused image is obtained by combining three images as given by

$$\mathrm{FUSE} = A + \tfrac{1}{2}S^{op} - \tfrac{1}{2}S^{cl}. \tag{22}$$

The '+' and '−' operations are applied between corresponding pixels of three different images.

5. Experimental results and discussions

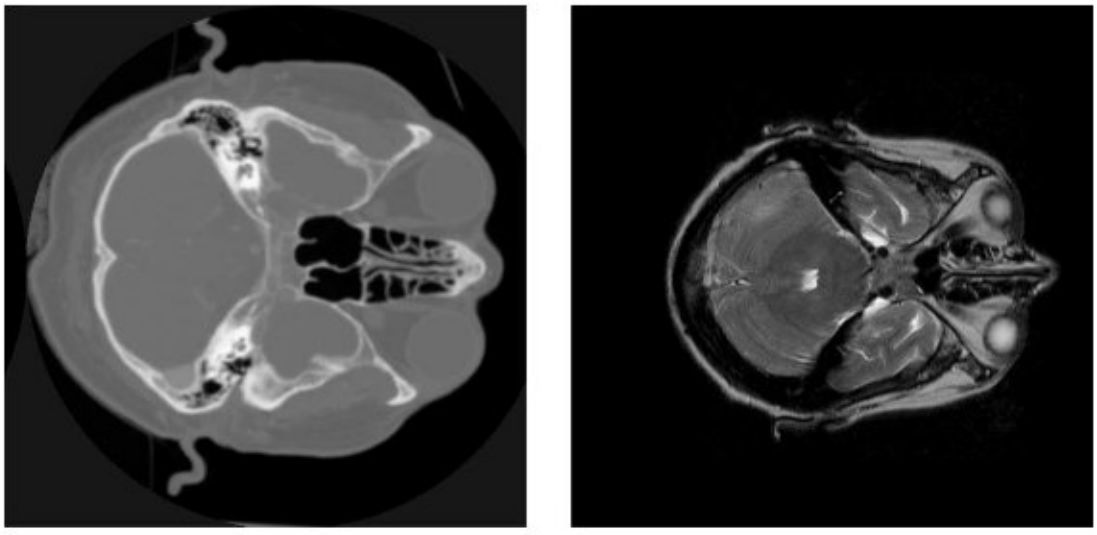
As stated earlier the proposed algorithm is independent of imaging technology. Hence, it can be used for fusing images obtained through any multi-sensor image acquisition system. However, we have demonstrated the performance of the proposed algorithm using CT and MR images of cross-section of human brain. The details regarding the acquisition of the CT and MR images on which we have executed our proposed algorithmic steps are specified below.

The MR image has been obtained on a Genesis Sigma scanner. Window width and the window length of the image are 1120 and 561, respectively. The slice thickness for the image is 5 mm while the magnification ratio is 1.5. The field of vision for the image is 240.00. The MR image in its original form is basically a DICOM image. Finally, a pixel matrix of size 512×512 is obtained. The CT image, which also is originally a DICOM image, is obtained using a Rhapsode scanner with a magnification of 1.25 and field of vision equal to 189.89. The window width and length for the CT image are 2985 and 496, respectively, while the slice thickness is 1.0 mm. Final image is again of size 512×512. The original CT and MR images are shown in Figs. 3(a) and (b), respectively.

Figs. 4(a) and (b) are images obtained by applying registration technique as described in Section 2. The fused image constructed according to the proposed method is shown in Fig. 4(c). In our experiment we have taken B a disk of radius one and $n = 33$. The fused image is found to contain salient/prominent features that are present in either modalities without any distortion. Second, it is found to be clinically more informative as compared to the individual constituting images. The fusing ability of the algorithm is measured quantitatively by means of, say, pixel-gray-level correlation between two images. Let us define correlation between two images f and g by

$$R(f,g) = \frac{\sum_{x,y}(f(x,y) - \bar{f})(g(x,y) - \bar{g})}{\sqrt{\sum_{x,y}(f(x,y) - \bar{f})^2}\sqrt{\sum_{x,y}(g(x,y) - \bar{g})^2}}, \tag{23}$$

where, $\bar{f} = (1/N)\sum_{x,y} f(x,y)$ and $\bar{g} = (1/N)\sum_{x,y} g(x,y)$ and N is the total number of pixels in either of the images. In our experiment the computed values of R(FUSE, CT), R(FUSE, MR) and R(CT, MR) are 0.89, 0.81 and 0.59, respectively. Since R(FUSE,CT) > R(MR,CT) and R(FUSE,MR) > R(MR,CT), the fused image enjoys relatively high correlation with either of the images. This implies features of both the images are transported to the fused image. However, this measure is too simple to yield any strong conclusion. To account for a better measure for the extent of similarity between the fused image and either of its constituting images we carry out the following analysis.



(a) (b)

Fig. 3. Test images used in the experiment. (a) CT image, (b) MR image.

5.1. Similarity measure

Two images, say for example, f and g are said to be similar subject to some *scaling* (by a constant factor c) and a d.c. bias (by a constant d), if the following relation holds for every pixel of f and g. That means f and g are similar if

$$g(x,y) = cf(x,y) + d \tag{24}$$

for all (x,y). Based on this relation, one can estimate one of the images, say, g from the other image f about which one has relatively better knowledge. Thus, the estimated image g_{est} is given by

$$g_{est}(x,y) = cf(x,y) + d. \tag{25}$$

However, the accuracy of this estimation relies on how correctly the parameters c and d are determined. The Euclidean distance or the error between images g and g_{est}, the image estimated from f, is then given by

$$E(g,g_{est}) = \sum_x \sum_y (g(x,y) - cf(x,y) - d)^2. \tag{26}$$

The parameters c and d of this relation can be estimated in such a way that $E(g,g_{est})$ is minimum. However, the relation reflects our bias toward the knowledge about the image f. To get rid of such bias to any of the images under consideration we make use of the following expression for the error:

$$E(g,g_{est}) = \frac{\sum_x \sum_y (g(x,y) - cf(x,y) - d)^2}{1 + c^2}. \tag{27}$$

This is what is known as *eigen-vector fitting* and can be found in Ref. [26]. Parameters c and d are estimated by solving the normal equations generated by taking partial derivatives of $E(g,g_{est})$ with respect to c and d and equating each of them to zero. Finally, we have

$$E(g,g_{est}) = \tfrac{1}{2}(V_f + V_g - \sqrt{(V_f - V_g)^2 + 4V_{fg}^2}). \tag{28}$$

The minimum error $E(g,g_{est})$ gives the measure of the unbiased distance between the two images $f(x,y)$ and $g(x,y)$. Based on this error measure we define similarity as

$$s(f,g) = 1 - \frac{E(g,g_{est})}{V_g + V_{g_{est}}}. \tag{29}$$

When similarity measure, as given in Eq. (29), is not equal to one or, in other words, error between the functions, as given in Eq. (28) is not zero, similarity measure based on only their values is not a sufficient indication. In that case we need to compare spatial variation, too, of those functions. As we know, spatial variation can be analysed by means of derivative operators. Hence, to measure the similarity between two images we should not only compare their pixel values but also compare their spatial derivatives. This is because two very similar images should also have very similar derivative images upto first few orders. Consequently, the error or distance between their derivative images must also be small. To save time and space we have considered here only the first (i.e. *gradient* image) and the second-order (i.e. *laplacian* image) derivatives along with the pair of original images for computing the error and, subsequently, the similarity

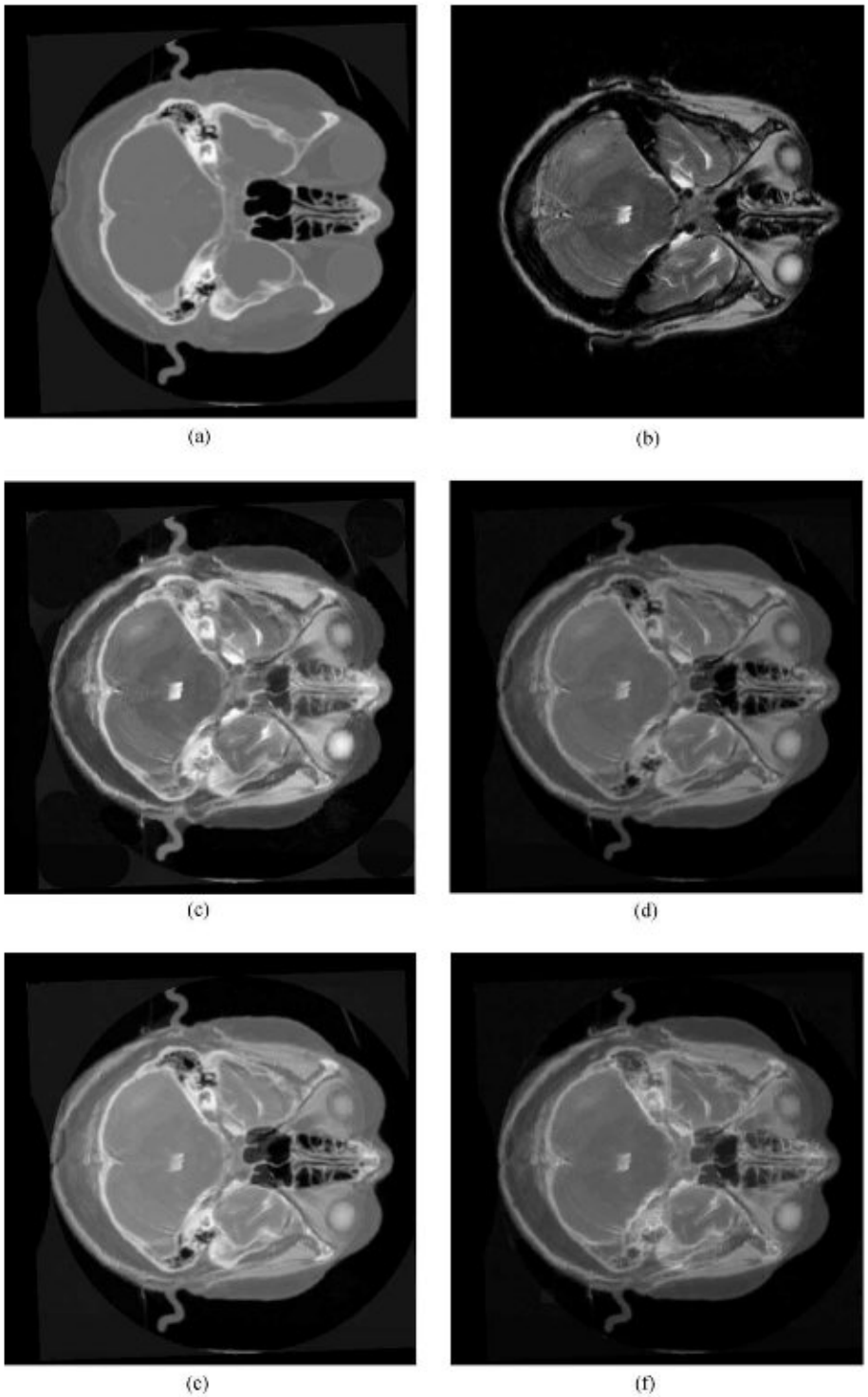


Fig. 4. (a) Registered CT image, (b) registered MR image, (c)–(f) results of fusion, (c) tower (d) simple averaging, (e) KL transform, and (f) pyramid.

between them. As an overall measure of similarity we take the product of the individual similarities. Thus, the *overall similarity* $S(f,g)$ is measured as

$$S(f,g) = s(f,g)s(f',g')s(f'',g''). \quad (30)$$

Finally, the *degree of fusion* is measured as

$$\delta = \min\{S(\text{Fuse, CT}), S(\text{Fuse, MR})\}. \quad (31)$$

At this point we like to present few other widely used image fusion methods with which we have compared the performance of our method.

5.2. A brief description of other schemes

5.2.1. Averaging

Averaging is the simplest technique of fusing two images in which the fused image is formed by a pixel-wise averaging of the constituting images. This method is crude and does not hold good in many cases. The fused image may not contain many salient features of the individual images. As a result there may be a huge loss of information at the time of blending the multi-modal images. Fused image obtained by averaging the images shown in Figs. 4(a) and (b) is shown in Fig. 4(d).

5.2.2. Fusion employing K–L Transform

For constructing the fused image employing K–L transform, we form N vectors $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_N$ where $\mathbf{f}_j = \{f_{1j}, f_{2j}\}$. The first element comes from the first image while the second comes from the second image. We then construct the *mean* vector and the *variance–covariance matrix* from these N vectors. The fused image $F(r,c)$ is constructed from the individual images $F_1(r,c)$ and $F_2(r,c)$ using the following relation:

$$F(r,c) = \phi_1 F_1(r,c) + \phi_2 F_2(r,c), \quad (32)$$

where $\Phi = \{\phi_1, \phi_2\}$ is the normalized eigen vector of the variance–covariance matrix corresponding to its largest eigen value. Thus, this method is nothing but a weighted averaging technique where the weights are determined by the K–L transform. Fused image obtained by using this method on the images shown in Figs. 4(a) and (b) is shown in Fig. 4(e).

5.2.3. Fusion employing morphological pyramid

The method relies on formation of an image pyramid using morphological sampling theorem [18]. Fusion of images using this method can be found in Ref. [2]. Fused image obtained by using this method on the images shown in Figs. 4(a) and (b) is shown in Fig. 4(f).

5.3. Comparison of performance

We have compared the performance of our fusion scheme with that of other well-known schemes such as *averaging*, *morphological pyramid* and K–L *transform*. A visual comparison among Figs. 4(c)–(f) reveals that our method produces the best result giving appropriate emphasis on features of all scales coming from both the images. For quantitative comparison of the results, we have computed the similarity between the *fused* image and the CT (or the MR) image for all the schemes mentioned above and, finally, computed the degree of fusion for each of them. The results are summarized in Table 1. Table 1 indicates that the *degree of fusion* is maximum for the proposed fusion scheme. In that scheme the image pairs along with their first and second-order gradients maintain consistently high similarity values. Morphological pyramidal approach and averaging technique fuse images with comparable degree of fusion. However, experimental results show that former one is much superior to the latter. Performance of

Table 1
Comparative study of different fusion schemes[a]

Image1	Image2	Level of comparison						Overall similarity	Degree of fusion
		Original images		Gradient images		Laplacian images			
		Error	Similarity	Error	Similarity	Error	Similarity		
$FUSE_{mmf}$	CT	295.99609	0.95019	32.25394	0.861920	1115.727539	0.952152	0.779797	0.779797
$FUSE_{mmf}$	MR	324.74359	0.94948	29.61146	0.865361	1223.828857	0.951479	0.781777	
$FUSE_{pyr}$	CT	158.84448	0.95525	19.68276	0.778367	587.132812	0.946611	0.712417	0.691508
$FUSE_{pyr}$	MR	271.66717	0.92718	17.56349	0.801107	1017.848145	0.930977	0.691508	
$FUSE_{avg}$	CT	136.04761	0.95186	16.47837	0.762021	505.472656	0.954643	0.692436	0.692436
$FUSE_{avg}$	MR	200.38049	0.93121	12.53961	0.823891	752.016602	0.934566	0.717018	
$FUSE_{KLT}$	CT	98.84716	0.98376	11.98582	0.920979	367.261719	0.984709	0.892174	0.670788
$FUSE_{KLT}$	MR	380.81756	0.94398	35.01140	0.750388	1418.872314	0.946963	0.670788	

[a] mmf: Proposed approach, pyr: pyramidal approach, avg: averaging, KLT: K–L transform.

averaging technique and K–L transform-based technique are comparable for a large set of images. Moreover, K–L transform-based fusion scheme is slightly better than averaging-based scheme, though Table 1 shows the reverse for this particular example.

6. Conclusion

In this paper we have presented a method for fusing two-dimensional multi-sensor gray-scale images using multiscale morphology. The usefulness of the method is illustrated using CT and MR images of cross-section of human brain. The result of the proposed method is compared with that of some widely used image fusion methods both qualitatively and quantitatively. Experimental results reveal that the proposed method produces better fused image than that by the latters. However, our method is computationally more expensive and needs more space for implementation. It should be noted that the proposed algorithm is domain-independent. That means it uses knowledge of neither the imaging device nor the objects being imaged. Therefore, it can be applied to fusion of other kinds of multi-sensor images, e.g. satellite images, too. Second, as the actual fusion is done during the construction of combined towers by taking the pixel-wise maximum, the algorithm can readily be extended for fusing more than two images. We wish to extend our work for fusing multi-sensor images in three dimension. Also, special attention needs to be paid for reducing space and time requirement for the implementation of the algorithm.

7. Summary

In this paper a scheme for multi-sensor image fusion for 2D images has been proposed by the authors. Quite often an image resulting from a given modality in a multi-sensor image acquisition system is found to be inefficient in detecting some features of the object to be imaged. Also features common to more than one modality may not have identically equal clarity. Consequently, the information content of the individual images suffer from incompleteness. Integration of all such individual images may give rise to a composite image which is enriched with features best detected in the individual images. However, creation of a fused image does not, in general, reduce the importance or the requirement of the individual images.

The proposed method presented in this paper makes use of *multiscale morphology*. Mathematical morphology is a powerful tool in the field of image processing where the filtering technique takes care of the shape of the features present in the image. Multiscale morphology is an extended concept where shape and size (i.e. scale) of the features are taken care of simultaneously.

Prior to fusion the individual images need to be registered so as to make similar objects present in different images appear similar in size, shape, orientation and position. There exists a number of registration schemes. However, in our case we have adopted a landmark point-based registration scheme which is unbiased to all the images to be registered.

As stated earlier some features are best detected in some modalities. The scale specific bright and dark features are extracted from the multi-sensor registered images using multiscale *tophat transformation*. The feature images of same resolution but of different scales are stacked in a number of *morphological towers*. From these towers we pick up the best detected bright and dark features specific to a given scale and form two more feature images corresponding to bright and dark features at that scale. This is repeated for all scales. These bright and dark feature images, so formed, then participate in the formation of the fused image. Thus the fused image is constructed by combining these best features for all scales.

Our proposed algorithm can be used to fuse any type of 2D multi-sensor images. However, in our experiment we have tested our algorithm on a pair of biomedical images. For the purpose of comparison we have implemented few other methods of 2D image fusion and included the respective results after executing them on the same set of brain images. Quantitative comparisons among the methods have been presented through error measure and degree of fusion.

The paper ends with a concluding section where the discussion includes a possible extension of the work for fusion of 3D images, remedial requirements for improving the drawbacks of the implementational aspects of the proposed method.

Acknowledgements

The authors like to thank Prof. C.A. Murthy and Dr. D.P. Mukhopadhyay of Indian Statistical Institute, Calcutta and Dr. P. Ghosh of Jadavpur University, Calcutta for technical discussion.

References

[1] H. Li, B.S. Manjunath, S.K. Mitra, Multisensor image fusion using the wavelet transform, GMIP: Graphical Models Image Process. 57 (3) (1995) 235–245.

[2] G.K. Matsopoulos, S. Marshall, J.N.H. Brunt, Multiresolution morphological fusion of MR and CT images of the human brain, IEE Proc.-Visual Image Signal Process. 141 (1994) 137–142.

[3] M.A. Abidi, R.C. Gonzalez, Data fusion in Robotics and Machine Intelligence, Academic Press Inc., Boston, 1992.

[4] J.J. Clark, A.L. Yuille, Data Fusion for Sensory Information Processing Systems, Kluwer Academic Publishers, Boston, 1990.
[5] R.C. Luo, M. Lin, R.S. Scherp, Multisensor integration and fusion in intelligent systems, IEEE Trans. Systems Man Cybernet. 19 (1989) 901–931.
[6] I. Bloch, Information combination operators for data fusion: a review with classification, IEEE Trans. Systems Man Cybernet. Part A: Systems and Humans 26 (1996) 52–67.
[7] M.A. Hurn, K.V. Mardia, T.J. Hainsworth, J. Kirkbride, E. Berry, Bayesian fused classification of medical images, Technical Report No. STAT/95/20/C, University of Leeds, Department of Statistics, 1995.
[8] D.P. Mukherjee, P. Dutta, D. Dutta Majumdar, Entropy theoretic fusion of multimodal medical images, Technical Report ECSU/2/98, Electronics and Communication Sciences Unit, Indian Statistical Institute, 1998.
[9] C. Shekhar, V. Govinder, R. Chellappa, Multisensor image registration by feature consensus, Pattern Recognition 32 (1999) 39–52.
[10] L.G. Brown, A survey of image registration, ACM Comput. Survey 24 (1992) 325–376.
[11] S. Banerjee, D.P. Mukherjee, D. Dutta Majumdar, Point landmarks for the registration of CT and MR images, Pattern Recognition Lett. 16 (1995) 1033–1042.
[12] A. Colligon, D. Vandermeulen, P. Seutens, G. Marchal, Registration of 3D multimodality medical imaging using surfaces and point landmarks, Pattern Recognition Lett. 15 (1994) 461–467.
[13] D.M. Mount, N.S. Netanyahu, J.L. Moigne, Efficient algorithms for robust feature matching, Pattern Recognition 32 (1999) 17–38.
[14] A.A. Goshtasby, J.L. Moigne, Image registration Guest Editor's introduction, Pattern Recognition 32 (1999) 1–2.
[15] K.V. Mardia, T.J. Hainsworth, Image warping and Bayesian restoration with grey-level templates, in: K.V. Mardia, G.K. Kanji (Eds.), Advances in Applied Statistics, Statistics and Images, Carfax, Abingdon, 1993.
[16] K.V. Mardia, I.L. Dryden, Statistical Shape Analysis, Wiley, Chichester, 1998.
[17] J. Serra, Image Analysis and Mathematical Morphology, Vol. 1, Academic Press, New York, 1982.
[18] R.M. Haralick, L.G. Shapiro, Computer and Robot Vision, Vol. 1, Addison-Wesley, Reading, MA, 1992.
[19] P. Maragos, A Representation theory for morphological image and signal processing, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 586–599.
[20] M. Chen, P. Yan, A multiscaling approach based on morphological filtering, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 694–700.
[21] P. Maragos, Pattern spectrum and multiscale shape representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 701–716.
[22] F. Meyer, in: J.L. Chermant (Ed.), Contrast Feature Extraction, Quantitative Analysis of Microstructure in Material Sciences, Biology and Medicine, Riederer Verlag, Stuttgart, Germany, 1978.
[23] J. Foley, A. Van Daam, S. Feiner, J. Huges, Computer Graphics: Principal and Practice, Addison-Wesley, Reading, MA, 1992.
[24] S. Mukhopadhyay, B. Chanda, A multiscale morphological approach to local contrast enhancement, Signal Process. 80 (1999) 685–696.
[25] S. Mukhopadhyay, B. Chanda, An edge preserving noise smoothing technique using multiscale morphology, Signal Process., accepted for publication.
[26] E. Gose, R. Johnsonbaugh, S. Jost, Pattern Recognition and Image Analysis, Prentice-Hall of India, New Delhi, 1999.

About the Author—BHABATOSH CHANDA was born in 1957. He received B.E. in Electronics and Telecommunication Engineering and Ph.D. in Electrical Engineering from the University of Calcutta in 1979 and 1988, respectively. He received "Young Scientist Medal" of Indian National Science Academy in 1989 and "Computer Engineering Division Medal" of the Institution of Engineers (India) in 1998. He is also a recipient of UN fellowship, UNESCO-INRIA fellowship and fellowship of National Academy of Science, India during his carrier. He worked at Intelligent System lab, University of Washington, Seattle, USA as a visiting faculty from 1995 to 1996. He has published more than 50 technical articles. His research interests include Image Processing, Pattern Recognition, Computer Vision and Mathematical Morphology. Currently, he is working as Professor in Indian Statistical Institute, Calcutta, India.

About the Author—SUSANTA MUKHOPADHYAY received B.Sc.(Hons) in Physics from Presidency College, Calcutta in October 1988. He received B.Tech and M.Tech both in Radiophysics and Electronics from Calcutta University in April, 1992 and December, 1994, respectively. Currently, he is a Ph.D. student in the Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta. His research area and interests include image processing and mathematical morphology.