Chapter 1

Introduction to Image Fusion

Human beings possess wonderful sense to appreciate visuals. Eye plays a key role in supporting various human activities. An image capture of a visual scene always conveys much more information than any other description adhered to the scene.

![Human Sensors](image1.png)

Figure 1.1: Human sensors (a) Eyes, (b) Ears, (c) Nose, (d) Tongue, (e) Skin.

Human beings have five sensing capabilities (systems) as shown in Figure 1.1. They are eyes, ears, nose, tongue and skin. These sensors are able to acquire independent information. Eyes can visualize a scene. Ears can sense the data by listening to the sounds. Nose can smell the odour of an object. Tongue can sense the object’s taste. Skin can sense the texture and size of the object. As shown in Figure 1.2, all these five sensing systems act as sensors. Human brain collects data from these individual sensors and fuses or combines it for compact representation or better description about a scenario. This compact data is useful for decision making and task execution.
Data fusion is a process of combining information from several sources for optimal or compact representation of a huge data supporting better description and decision making. Human brain is a best example for a data fusion system.

Even when we take one sensor, for example eye. It can derive many useful details of a scene by looking at the scenario more than once. Brain will integrate the visuals and give the details hidden in a single view. Multiple views will always improve the decisions.

Whenever we take a snap shot of a scene with our digital camera, we will not be satisfied with a single image. We try to take few more images of the same scene, to have more clarity and information. It is not rare to find that none of the images contain all the required qualities. It is common to feel that the positive aspects of these are to be combined to get the desired image. It motivates us to fuse the images, for a desired output. We can use different cameras and fuse the images. Likewise many options are there.

The definition of image fusion is as follows “image fusion is the process of merging or combining or integrating useful or complementary information of several
source images such that the resultant image provides more accurate description about the scene than any one of the individual source images”.

Image fusion finds applications in digital photography (Haghighat et al., 2010), medical imaging (Guihong et al., 2001), remote sensing (Yonghong, 1998), micro imaging (Shreyamsha Kumar, 2013a), concealed weapon detection (Shreyamsha Kumar, 2013b), battle field monitoring (Cui et al., 2015), surveillance (Zhao et al., 2013a) and navigation (Shreyamsha Kumar, 2013a) etc.

1.1 IMAGE FUSION FUNDAMENTALS

Image is a two-dimensional quantity. It can be viewed as the combination of illumination and reflectance. Illumination stands for the amount of light from the source falling on the object and reflectance corresponds to the amount of light that is reflected from the same object.

Sensor is a device which converts incoming energy into electrical signal as shown in Figure 1.3(a). In case of imaging sensors, reflected energy will be converted into corresponding electrical signal. As displayed in Figure 1.3(b), the sensor array gives an array of signals. The sampling in the spatial domain is performed by the sensor array. These signals are quantized to obtain a digital image representation. This entire process is termed as digitization which is shown in Figure 1.3(c). Thus, the visual information present in a scene can be captured as a digital image $f(x, y)$ using a sensor array as shown in Figure 1.3(b). All the elements in the sensor array will be of same modality. Hence image capture using a sensor array is simply referred to as single sensor image capture. We may be interested in the details of a scene using multiple sensor arrays, each operating in a different wavelength range. This is simply termed as multi sensor image capture. In the following discussion, it can be noted that, the term sensor is used simply in place of a sensor array.
Figure 1.3: Single sensor imaging. (a) Single sensor, (b) Single sensor array and its corresponding image represented as a matrix, (c) Digitization process.
Single sensor image capture may not always provide complete information about a targeted scene. Sometimes we need two or more images of the same scene for better visual understanding. These images may be captured by using a single sensor of same modality or by using multiple sensors of different modalities, depending on the application (Ardeshir and Nikolov, 2007). These image captures provide complementary or visually different information. A human observer cannot reliably combine and observe a composite image from these multiple image captures. Useful or complementary information of these images should be integrated into a single image to provide more accurate description about the scene, than any one of the individual source images.

Two different problems where we need to capture multiple images and combine the required information are discussed below.

One is single sensor imaging in which multiple images of the same scene are captured using a single sensor to extract more details of a targeted scene. Another one is multi-sensor imaging which requires multiple images using different sensors for the same propose.

1) Problem 1: Single sensor imaging

In digital photography, objects of a scene at different distances can’t be focused at the same time. If lens of a camera focuses on an object at a certain distance then other objects appear blurred.

Image formation model of a sensor or a camera system is displayed in Figure 1.4. If a point \( P_i \) (green dot) on an object is focused then a dot \( p_i \) corresponding to that particular point will be generated on the sensor plane. Therefore, all the points at the same distance of \( P_i \) from lens will appear sharp. The region of acceptable sharpness of the object is referred to as depth of field (DoF).
Let us consider another point $P_2$ (red dot) on the object behind the point $P_1$. Since this point $P_2$ is out of DoF, it will generate a dot $p_2$ somewhere before the sensor plane. As lens distance from the $P_1$ increases, the object will appear more blurred. As shown in Figure 1, $P_3$ (dark blue dot) is located in front of $P_1$ on the object plane. Since $P_3$ fall out of the DoF, it produces a dot $p_3$ behind the sensor plane, resulting in an unsharp dot on the image plane (sensor plane). For better visual quality images should have all objects in focus. One of the best approaches to do this is: capture images with different focusing conditions and combine them to generate an all-in-one focus image.

Now we discuss another problem where we need to acquire multiple images using different modalities.
2) Problem 2: Multi sensor imaging

Visual information present in a scene can be captured as an image using a charge coupled device (CCD). The wavelength of the visible (VI) light that can be captured by CCD sensor ranges from 400 nm to 700 nm. However, in most of the computer vision applications, CCD image alone is not sufficient to provide all the details of the scene. To extract more details, complementary images of the same scene should be captured by using multiple sensors of different modalities. This can be done by capturing images in wavelengths other than the VI band of the electromagnetic spectrum.

<table>
<thead>
<tr>
<th>Electromagnetic Waves</th>
<th>Wavelength $\lambda$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma Rays</td>
<td>$10^{-15}$ to $10^{-11}$</td>
</tr>
<tr>
<td>X-Rays</td>
<td>$10^{-11}$ to $10^{-9}$</td>
</tr>
<tr>
<td>Ultraviolet</td>
<td>$10^{-9}$ to $4 \times 10^{-7}$</td>
</tr>
<tr>
<td>Visible (VI)</td>
<td>$4 \times 10^{-7}$ to $7 \times 10^{-7}$</td>
</tr>
<tr>
<td>Infrared (IR)</td>
<td>$7 \times 10^{-7}$ to $10^{-3}$</td>
</tr>
<tr>
<td>Microwave</td>
<td>$10^{-3}$ to 0.1</td>
</tr>
<tr>
<td>Radio</td>
<td>0.1 to $10^5$</td>
</tr>
</tbody>
</table>

Figure 1.5: The electromagnetic spectrum

Table 1.1: Electromagnetic wavelength range
The electromagnetic spectrum is illustrated in Figure 1.5. The corresponding wavelengths are presented in Table 1.1. As discussed before, VI light wavelength ranges from $4 \times 10^{-7}$ to $7 \times 10^{-7}$ m. The infrared (IR) spectrum wavelength ranges from $7 \times 10^{-7}$ to $10^{-3}$ m. IR spectrum is further divided into five sub bands as near, short, mid wave, long wave and far IR bands.

Usually objects with more than $0^\circ K$ emits radiations throughout the IR spectrum. The energy emitted by these objects can be sensed by IR sensors and displayed as images for the end users. However, these images alone are not sufficient to provide accurate description about the targeted scene. Hence, required information from VI spectrum also needs to be integrated for better scene understanding.

As shown in Figure 1.6, various scenarios such as digital photography (Zhang and Blum, 1999; Li and Kang, 2012), medical imaging (Guihong et al., 2001), concealed weapon detection (Shreyamsha, 2013a), military applications (Gan et al., 2015) are considered to justify the need of combining information of various source images of the same scene.

In digital photography (Zhang and Blum, 1999; Li and Kang, 2012), more than one objects of a scene can’t be focused at the same time due to inherent system limitations. If we focus on one object, we may lose information of other objects and vice versa. Figure 1.6(a) shows the foreground and background focused images of a bottle dataset. Foreground focused image provides information about the bottle in the foreground whereas background focused image gives the information of the bottle in the background of the same scene. These individual images do not provide complete information about the targeted scene. For better visual understanding, focused regions of these two images have to be combined to result an all-in-one focused image.
Figure 1.6: Problem specifications through various scenarios (a) Multi-focus imaging, (b) Medical imaging, (c) Concealed weapon detection, (d) Battlefield monitoring in military.
In medical imaging, different modalities like positron emission tomography (PET), single photon emission tomography (SPECT), computer tomography (CT) and magnetic resonance imaging (MRI) are used to capture complementary information. These individual image captures do not provide all the required details. Therefore, information from different captures has to be incorporated into a single image. Figure 1.6(b) shows the CT and MR images of a human brain. As shown in Figure 1.6(b), CT can capture bone structure or hard tissue information. Whereas, MRI can capture soft tissue information present in the brain. For a radiologist, fused image obtained from these two images will be helpful in computer assisted surgery and radio surgery for better diagnosis and treatment.

In concealed weapon detection, VI and milli-meter wave (MMW) sensors are used to capture complementary images. In Figure 1.6(c), first one is VI image and second one is MMW image. VI image conveys information of three persons. However, it is not providing any sign of existence of a weapon. MMW image conveys the weapon information alone. From these individual images it is difficult to identify, which person concealed the weapon. To accurately locate and detect the weapon, useful information from these complementary images has to be combined for a single image.

In military and navigation, VI and IR imaging sensors are used to acquire complementary information of the targeted scene. Due to bad weather circumstances such as rain, foggy winter etc., the images captured using VI sensors alone are not sufficient to provide the essential information about a situation. VI image is able to provide background details such as vegetation, texture, area and soil. In contrast, IR sensors provide information about the foreground like weapons, enemy and vehicle movements. For detection and localization of a target as well as to improve the situational awareness, information from both IR and VI images needs to be merged in a single image.

In Figure 1.6(d), first image is VI output and second one is IR image of a battle field. VI image provides the information of a battle field. However, it is incapable of identifying the person near fencing. IR image identifies the person existence but not able to provide sufficient visual information of the battle field. If we integrate useful information of these images in a single image, then we can easily identify and localize enemy or target.
Hence, for better understanding of the scene we need to combine essential visual information in the source images to obtain a meaningful image.

Image fusion (Ardeshir and Nikolov, 2007) is a phenomenon of integrating or combining or merging visually significant or complementary information of several co-registered source images of a scene into a single image. The resulting image provides more accurate description about a scene than any one of the individual source images.

As discussed so far, image fusion is used to obtain fused images in various applications like digital photography (multi-focus imaging), medical imaging, concealed weapon detection and military etc.

In Figure 1.7, all right hand side images are the fused images of various applications. An all-in-one focused image in Figure 1.7(a), obtained from the two out-of-focus images provides visually more information. Fused image in Figure 1.7(b), would assist a radiologist for better diagnosis and treatment than individual CT and MR images. Combined image in Figure 1.7(c) is giving information about the person as well as the concealed weapon. From the fused image in Figure 1.7(c), one can say that the third person from the left concealed the weapon inside his shirt. From the fused image in Figure 1.7(d), one can identify an enemy moment in the battle field near fencing.
Figure 1.7: Fusion results. (a) Multi-focus imaging, (b) Medical imaging, (c) Concealed weapon detection, (d) Battle field monitoring in military.
1.2 TYPES OF IMAGE FUSION SYSTEMS

Image fusion systems are broadly classified as single-sensor image fusion system (SSIF) (Figure 1.8(a)) and multi-sensor image fusion system (MSIF) (Figure 1.8(b)). In SSIF using a single sensor, sequence of images of the same scene are captured and useful information of these several images is integrated into a single image by the process of fusion. In noisy environment and in some illumination conditions, human observers may not able to detect the objects of interest which can be easily found from the fused images of that targeted scene. Digital photography applications such as multi-focus imaging and multi-exposure imaging (Li and Kang, 2012) come under SSIF. However, these fusion systems have their drawbacks. They depend on conditions like illumination and dynamic range of the sensors. For example, VI sensor like digital camera can capture visually good images in high illumination conditions. However, they fail to capture visually good images at low illumination condition such as night, fog and rain.

To overcome the difficulties of SSIF, MSIF systems are introduced to capture images in adverse environment conditions. In MSIF, multiple images of the same scene are captured using various sensors of different modalities to acquire complementary information. For example, VI sensors are good in high lighting conditions. However, IR sensors are able to capture images in low lighting conditions. Required and necessary information of these images are combined into a single image by the process of fusion. Applications such as medical imaging, military, navigation and concealed weapon detection fall under MSIF category.
Figure 1.8: Image fusion systems classification. (a) SSIF system, (b) MSIF system.
Various advantages of MSIF systems are:

1. Reliable and accurate information: These MSIF systems provide reliable and accurate description of the scene compared to source images.

2. Robust performance: Even if one sensor of the MSIF fails, this system generates a composite image by considering the redundant information of all sensors. So it is robust.

3. Compact representation: Fused image of the MSIF is compact and provides all the necessary information of source images in a single image.

4. Extended operating range: The range of operation is extended by capturing images at different operating conditions of the sensors.

5. Reduced uncertainty: Combined information of various sensors reduces the uncertainty present in the individual captures of the scene.

1.3 FUNDAMENTAL STEPS IN IMAGE FUSION

Fundamental steps involved in image fusion process are shown in Figure 1.9. It consists of 5 major steps. 1) Pre-processing, 2) Image registration, 3) Image fusion, 4) Post-processing and 5) Fusion performance evaluation.

1) In pre-processing stage, noise or artifacts introduced in the source images during image acquisition process are removed or reduced.
2) Image registration is the process of aligning or arranging more than one images of a same scene according to a co-ordinate system. In this process, one of the source images will be taken as a reference image. It is also termed as the fixed image. Then geometric transformation will be applied on remaining source images to align them with the reference image.

3) Fusion process can be performed at three levels (Ardeshir and Nikolov, 2007): pixel, feature and decision. Pixel level fusion is done on each input image pixel by pixel. However at feature level, fusion is executed on the extracted features of source images. At decision level, fusion is performed on probabilistic decision information of local decision makers. These decision makers are in turn derived from the extracted features. Pixel level fusion schemes are preferable for fusion compared to other level approaches because of their effectiveness and ease of implementation. In this thesis, our interest is only on pixel level fusion schemes.

4) During the fusion process, some required information of source images may be lost and visually unnecessary information or artifacts may be introduced into the fused image. Hence, fusion algorithms need to be assessed and evaluated for better performance. This performance analysis can be carried out by evaluating them qualitatively by visual inspection and quantitatively using fusion metrics.

5) In post-processing, fused images are further processed depending on the application. This processing may involve segmentation, classification and feature extraction.

In this thesis, we developed new pixel level image fusion algorithms for both multi-focus and multi-modal images.

1.4 ORGANIZATION OF THE THESIS

Remaining chapters of the thesis are outlined as follows:

Chapter 2 : Literature Review

This chapter briefs the classification of image fusion literature. Spatial domain
techniques are briefly reviewed. Various multi-scale fusion methods namely, pyramid, wavelet and edge preserving decomposition methods are discussed. Comments on the merits and demerits of various fusion methods are provided. Motivation for proposing new methods is presented. An overview of petrovic fusion metrics and image datasets that are used in the thesis is also provided.

Chapter 3 : Anisotropic diffusion based fusion of infrared and visible images (ADF)

A new edge preserving image fusion method for IR and VI sensor images is presented. Performance of the proposed algorithm is assessed with the help of petrovic metrics. Computational time is also measured. The results of the proposed algorithm are compared with the traditional and recent image fusion algorithms.

Chapter 4 : Two-scale image fusion of visible and infrared images using visual saliency detection (TIF)

A new image fusion method based on visual saliency detection and two-scale image decomposition is discussed. It is evaluated qualitatively by visual inspection and quantitatively using objective fusion metrics, in addition to the computational time calculation. Outcomes of this method are compared with the state-of-the-art multi-scale fusion techniques.

Chapter 5 : Maximum symmetric surround saliency detection based multi-focus image fusion (MSSSF)

A new multi-focus image fusion method based on two-scale image decomposition and saliency detection using maximum symmetric surround is discussed. This method is tested on several multi-focus image datasets and it is compared with the traditional and recently proposed fusion methods using various fusion metrics.

Chapter 6 : Conclusion

This chapter concludes the thesis and gives directions for the future research in the area of image fusion.