Chapter 1
Introduction

Abstract. The subject of this book is *image fusion* which we define as the process of combining multiple input images into a single composite image. Our aim is to create from the collection of input images a single output image which contains a better description of the scene than the one provided by any of the individual input images. The output image should therefore be more useful for human visual perception or for machine perception. The basic problem of image fusion is one of determining the best procedure for combining the multiple input images. The view adopted in this book is that combining multiple images with *a priori* information is best handled within a statistical framework. In particular we shall restrict ourselves to classical and robust statistical approaches, Bayesian methods, sub-space and wavelet techniques.

1.1 Synergy

The principal motivation for image fusion is to improve the quality of the information contained in the output image in a process known as *synergy*. A study of existing image fusion techniques and applications shows that image fusion can provide us with an output image with an improved quality. In this case, the benefits of image fusion include:

1. Extended range of operation.
2. Extended spatial and temporal coverage.
3. Reduced uncertainty.
4. Increased reliability.
5. Robust system performance.
6. Compact representation of information.

Traditionally, the input images are captured by the same camera at different times or are captured by different cameras at the same time. However, in the definition of image fusion we shall also include the case when the input images are derived from the same "base" image but which have undergone different processing algorithms.

The following examples illustrate the image fusion synergy process. The first example deals with input images which are captured by the same camera at different times. The next example deals with input images captured by different cameras at the same time, while the third example deals with input images which are derived from the same base image and which are processed differently.

Example 1.1. Multiple Camera Surveillance Systems [9]. The increasing demand for security by society has led to a growing need for surveillance activities in many environments. For example, the surveillance of a wide-area urban site may be provided by periodically scanning the area with a single narrow field-of-view camera. The temporal coverage is, however, limited by the time required for the camera to execute one scan. By using multiple cameras we reduce the mean time between scans and thereby increase the temporal coverage.

Example 1.2. Multispectral Bilateral Video Fusion [2]. A significant problem in night vision imagery is that while an infra-red (IR) image provides a bright and relatively low-noise view of a dark environment, it can be difficult to interpret due to inconsistences with the corresponding visible-spectrum image. In bilateral fusion we enhance a visible video input using information from a spatially and temporally registered IR video input. Our goal is to create a video that appears as if it was imaged only in the visible spectrum and under more ideal exposure conditions than actually existed.

Example 1.3. Color Image Segmentation [11]. A significant problem in computer vision is the reliable segmentation a base image into meaningful labeled segments. In ensemble image segmentation we generate an ensemble of color input images by transforming the base image in different ways. Each input image is separately segmented using a simple segmentation algorithm. By fusing the multiple segmented images we are able to substantially improve both the accuracy and the reliability of the segmentation process.

1.2 Image Fusion Process

Fig. 1.1 shows the principal processes in a generic image fusion processing chain for the case when the output is a single fused image $\widetilde{I}$. The principal processes in the chain are:

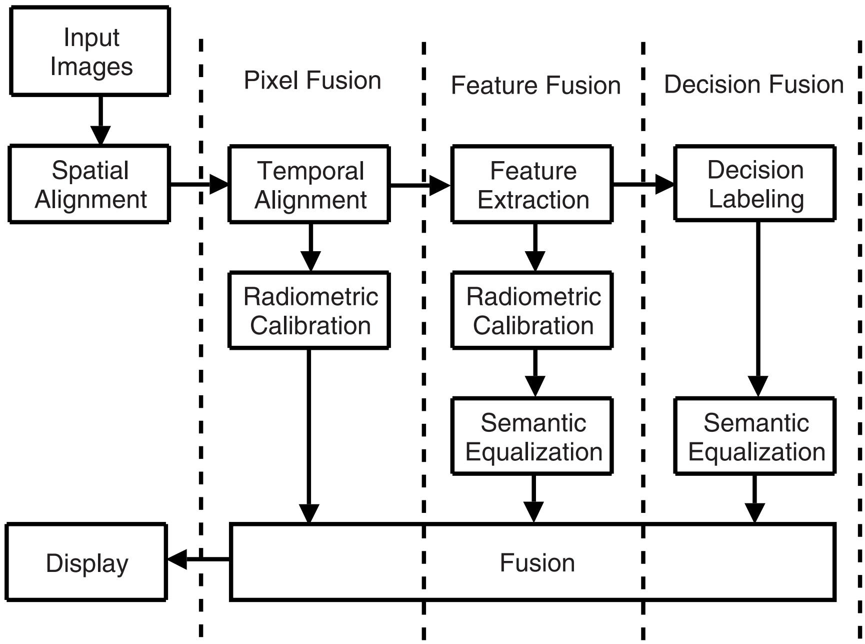

Fig. 1.1 Shows the generic image fusion processing chain. It consists of four principal blocks: (1) Multiple Input Images. Multiple images of the external scene are captured by multiple sensors. (2) Common Representational Format. The input images are transformed into a common representational format. This involves several processes including: spatial and temporal alignment, semantic equivalence, radiometric calibration, feature extraction and decision labeling. (3) Fusion. The multiple images in the common representational format are fused together. The fusion process may be classified into three classes: pixel fusion, feature fusion and decision fusion. (4) Display. The fused output is processed for display.

Multiple Input Images. The external environment is captured by one or more image sensors or cameras. Each camera generates one or more input images.

Common Representational Format. The input images are transformed so they "speak a common language". This involves several processing including: spatial, temporal, semantic and radiometric alignment, feature extraction and decision labeling.

Fusion. After conversion into a common representational format the spatially, temporally, semantically and radiometrically aligned images, feature maps or decision maps are fused together in the fusion block. The output is a fused image $\widetilde{I}$, feature map $\widetilde{F}$ or decision map $\widetilde{D}$.

Display. The fused image, feature map or decision map is processed for display.

1.3 Common Representational Block

The principal function in the common representational format block are:

Spatial Alignment. The input images are spatially aligned into the same geometric base. Without a common geometric base any information derived from a given input image cannot be associated with other spatial information. The accurate spatial alignment of the input images is therefore a necessary condition for image fusion. *Note*. After spatial alignment the input images are re-sampled and if necessary the gray-levels of the input images are interpolated.

Temporal Alignment. The spatially aligned input images are temporally aligned to a common time. This step is only required if the input images are changing or evolving in time. In this case the accurate temporal alignment of the input images is a necessary condition for image fusion.

Feature Extraction. Characteristic features are extracted from the spatially and temporally aligned input images. The output is one or more feature maps for each input image [1].

Decision labeling. Pixels in each spatially and temporally aligned input image or feature map are labeled according to a given criteria. The output is a set of decision maps.

Semantic Equivalence. In order for the input images, feature maps or decision maps to be fused together they must refer to the same object or phenomena. The process of causally linking the different inputs to a common object or phenomena is known as semantic equivalence.

Radiometric Calibration. The spatially, temporally and semantically aligned input images and feature maps are converted to a common measurement scale. This process is known as radiometric calibration.

When the input is a set of K multiple image sequences $I_k(t), t \in [T_1, T_2], k \in \{1, 2, \ldots, K\}$, the output is a fused image sequence $\widetilde{I}(t), t \in [T_1, T_2]$, feature map sequence $\widetilde{F}(t)$ or decision map sequence $\widetilde{D}(t)$. In this case we replace the spatial alignment and temporal alignment blocks in Fig. 1.1 with a single spatial-temporal alignment block which performs both functions simultaneously.

Figs. 1.2–1.4 shows the adaption of the generic processing chain to Ex 1.1–1.3.

1.4 Image Fusion Block

In the image fusion block we fuse together the information contained in the multiple input images after conversion into a common representational format. The common representational format may take the form of an image I, a feature map F or a decision map D. Very often we shall not differentiate between I, F and D and in this case we shall refer to all three as an "image".

We find it convenient to further divide the fusion algorithms into arithmetic, subspace and multi-scale techniques [5].

[1] A feature is any distinguishing property or attribute of an image. Examples of features used in image fusion are: edges, lines, patterns and color.

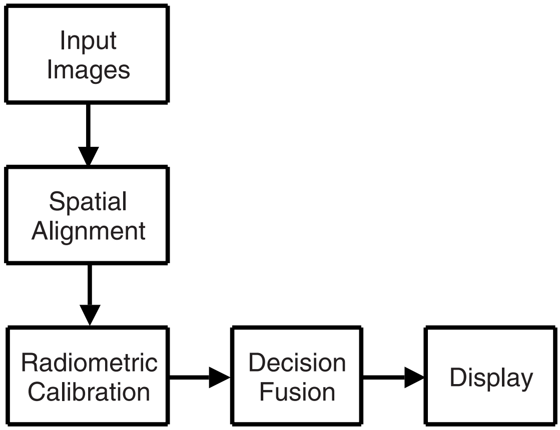

Fig. 1.2 Shows the image processing chain for the multiple camera surveillance system discussed in Ex. 1.1

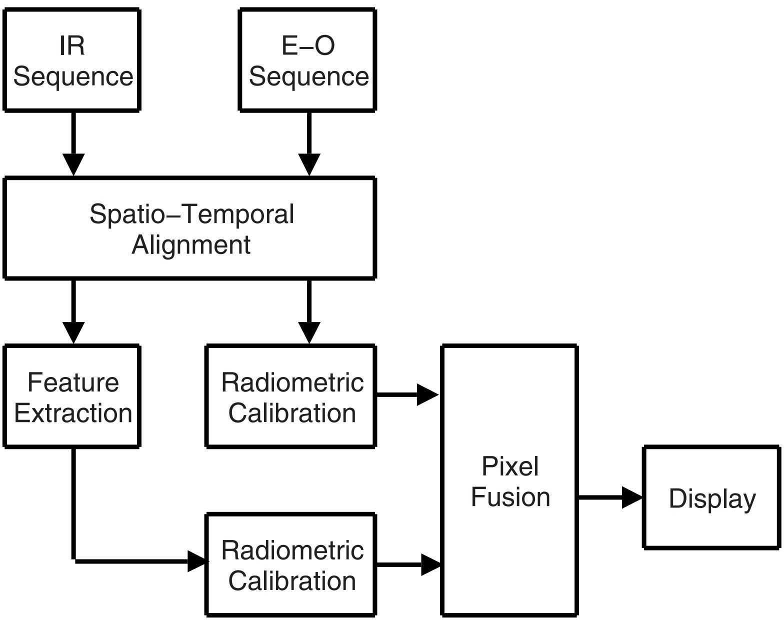

Fig. 1.3 Shows the image processing chain for the multispectral bilateral video fusion system discussed in Ex. 1.2

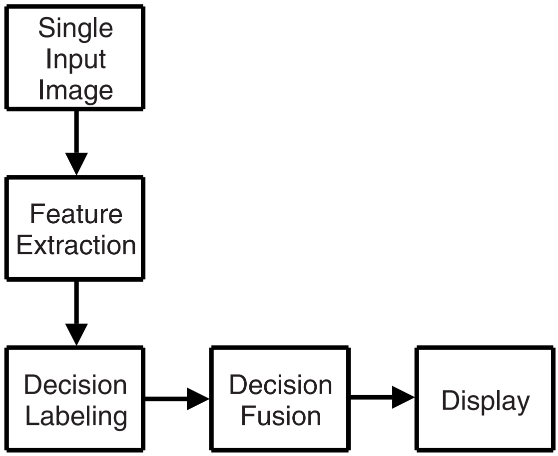

Fig. 1.4 Shows the image processing chain for the color image segmentation system discussed in Ex. 1.3

Pixel Operations. These operations include simple arithmetic operators such as addition, subtraction, division and multiplication as well as minimum, maximum, median and rank. It also includes more complicated operators which are defined by a function or algorithm, such as the expectation-maximization algorithm and Markov random field.

Sub-Space Methods. The sub-space methods are a collection of statistical techniques which remove the correlation which exists between the input images $I_k, k \in \{1, 2, \ldots, K\}$. Important sub-space techniques are: principal component analysis (PCA), independent component analysis (ICA), non-negative matrix factorization (NMF), canonical correlation analysis (CCA) and linear discriminant analysis (LDA).

Multi-Scale Methods. The multi-scale fusion methods are a collection of techniques in which we transform each input image $I^{(k)}$ into a multi-scale representation: $\left(y_0^{(k)}, y_1^{(k)}, \ldots, y_L^{(k)}\right)$.

1.5 Image Fusion Algorithms

For environments which are essentially static and in which the output is a single image $\widetilde{I}$ we often impose the following requirements [15] on the image fusion algorithms:

Pattern conservation. The fusion process should preserve all relevant information on the input imagery in the composite image.

Artifact free. The fusion scheme should not introduce any artifacts or inconsistencies which would distract the human observer or subsequent image processing stages.

Invariance. The fusion scheme should be shift and rotational invariant, i.e. the fusion result should not depend on the location or orientation of an object in the input imagery.

For environments which are evolving in time the input is a set of input sequences $I_k(t), t \in [T_1, T_2], k \in \{1, 2, \ldots, K\}$ and the output is a fused image sequence $\widetilde{I}(t)$. In this case, we often impose the following additional requirements on the image fusion algorithms:

Temporal stability. The fused output should be temporally stable, that is, gray-level changes in $\widetilde{I}(t)$, should be present in at least one of the input sequences $I_k(t)$.

Temporal consistency. Gray level changes which occur in the input sequences $I_k(t)$ must be present in the fused sequence $\widetilde{I}(t)$.

1.6 Organization

Apart from two preliminary chapters, the book is divided into three parts:

Part I: Theories. This consists of Chapts. 3-7 and deals with the conceptual theories and ideas which underlie image fusion. Here we emphasize the concept of a common representational framework and include detailed discussions on image registration, radiometric calibration and semantic equalization.

Part II: Techniques. This consists of Chapts. 8-18 and deals with a wide range of techniques and algorithms which are in common use in image fusion. Among the topics considered are: sub-space transformations, multi-resolution analysis, ensemble learning, bagging, boosting, color spaces, Markov random fields, image similarity measures and the expectation-maximization algorithm. Together Parts I and II provide the reader with an integrated and comprehensive overview of image fusion.

Part III: Applications. This consists of Chapts. 19-22 and deals with applications. In it we examine several real-life image fusion applications. The aim is to illustrate how the theories and techniques of image fusion are used in practical situations.

1.7 Software

The following matlab routines and toolboxes are of general utility and are widely used in image fusion.

IMAGE FUSION TOOLKIT. Image fusion toolbox. Author: Eduardo Fernandez Canga.

MATIFUS. Matlab toolbox for image fusion. Authors: P. M. de Zeeuw, G. Piella and H. J. A. M. Heijmans [14].

MATLAB IMAGE PROCESSING TOOLBOX. Matalb image processing toolbox.

MATLAB WAVELET TOOLBOX. Matlab wavelet toolbox.

1.8 Further Reading

General overviews on multi-sensor image fusion are [1, 3, 4, 5, 6, 7, 8, 10, 12]. For an extended discussion regarding the issues involved in defining multi-sensor image fusion and related terms, see [13, 16].

References

1. Appriou, A., Ayoun, A., Benferhat, S., Bernard, P., Cholvy, L., Cooke, R., Cuppens, F., Dubois, D., Fargier, H., Grabisch, M., Kruse, R., Lang, J., Moral, S., Prade, H., Saffiotti, A., Smets, P., Sossai, C.: Fusion: general concepts and characteristics. Int. J. Intell. Syst. 16, 1107–1134 (2001)
2. Bennett, E.P., Mason, J.L., McMillan, L.: Multispectral bilateral video fusion. IEEE Trans. Image Process. 16, 1185–1194 (2007)
3. Bloch, I. (ed.): Information Fusion in Signal and Image Processing. Wiley, Chichester (2008)
4. Fenoy, G.P.: Adaptive wavelets and their applications to image fusion and compression. PhD thesis University of Amsterdam (2003)
5. Gang, H.: Image fusion, image registration and radiometric normalization for high resolution image processing. PhD thesis, University of New Brunswick, Canada (2007)
6. Hall, D.L., Llinas, J. (eds.): Handbook of Multisensor Data Fusion. CRC Press, Boca Raton (2001)
7. Hall, D.L., McMullen, S.: Mathematical Techniques in Multisensor Data Fusion, 2nd edn. Artech House Inc., Norwood (2004)
8. Hyder, A.K., Shahbazian, E., Waltz, E. (eds.): Multisensor Fusion. Kluwer Academic Publishers, Dordrecht (2002)
9. Jones, G.D., Allsop, R.E., Gilby, J.H.: Bayesian analysis for fusion of data from disparate imaging systems for surveillance. Image Vis. Comp. 21, 843–849 (2003)
10. Luo, R.C., Yih, C.-C., Su, K.L.: Multisensor fusion and integration: approaches, applications and future research directions. IEEE Sensors Journal 2, 107–119 (2002)
11. Mignotte, M.: Segmentation by fusion of histogram-based K -means clusters in different clolor spaces. IEEE Trans. Image Process. 17, 780–787 (2008)
12. Mitchell, H.B.: Multi-Sensor Data Fusion: An Introduction. Springer, Berlin (2007)
13. Oxley, M.E., Thorsen, S.N.: Fusion or integration: what's the difference? In: Proc. 7th Int. Conf. Information Fusion, pp. 429–434 (2004)
14. Piella, G.: A general framework for multiresolution image fusion: from pixels to regions. Inform. Fusion 9, 259–280 (2003)
15. Rockinger, O., Fechner, T.: Pixel-level image fusion. In: Proc. SPIE, vol. 3374, pp. 378–388 (1998)
16. Wald, L.: Some terms of reference in data fusion. IEEE Trans. Geosci. Remote Sensing 37, 1190–1193 (1999)