Guest editorial

Image fusion: Advances in the state of the art

What is image fusion?

Image fusion is the process of combining information from two or more images of a scene into a single composite image that is more informative and is more suitable for visual perception or computer processing. The objective in image fusion is to reduce uncertainty and minimize redundancy in the output while maximizing relevant information particular to an application or task. Given the same set of input images, different fused images may be created depending on the specific application and what is considered relevant information. There are several benefits in using image fusion: wider spatial and temporal coverage, decreased uncertainty, improved reliability, and increased robustness of system performance.

Often a single sensor cannot produce a complete representation of a scene. Visible images provide spectral and spatial details, and if a target has the same color and spatial characteristics as its background, it cannot be distinguished from the background. If visible images are fused with thermal images, a target that is warmer or colder than its background can be easily identified, even when its color and spatial details are similar to those of its background. Fused images can provide information that sometimes cannot be observed in the individual input images. Successful image fusion significantly reduces the amount of data to be viewed or processed without significantly reducing the amount of relevant information.

History

Image fusion is a branch of data fusion where data appear in the form of arrays of numbers representing brightness, color, temperature, distance, and other scene properties. Such data can be two-dimensional (still images), three-dimensional (volumetric images or video sequences in the form of spatio-temporal volumes), or of higher dimensions. Early work in image fusion can be traced back to the mid-eighties. Burt [1] was one of the first to report the use of Laplacian pyramid techniques in binocular image fusion. Burt and Adelson [2] later introduced a new approach to image fusion based on hierarchical image decomposition.

At about the same time, Adelson [3] disclosed the use of a Laplacian technique in construction of an image with an extended depth of field from a set of images taken with a fixed camera but with different focal lengths. Later Toet [4] and Toet et al. [5] used different pyramid schemes in image fusion. These techniques were mainly applied to fuse visible and IR images for surveillance purposes [6].


Use of the discrete wavelet transform (DWT) in image fusion was almost simultaneously proposed by Li et al. [11] and Chipman et al. [12]. At about the same time Koren et al. [13] described a steerable dyadic wavelet transform for image fusion. Also around the same time Waxman and colleagues [14,15] developed a computational image fusion methodology based on biological models of color vision and used opponent processing to fuse visible and infrared images. The need to combine visual and range data in robot navigation [16] and to merge images captured at different locations and modalities for target localization and tracking in defense applications [17] prompted further research in image fusion. Many other fusion techniques have been developed during the last decade. Today, image fusion algorithms are used as effective tools in medical, remote sensing, industrial, surveillance, and defense applications that require the use of multiple images of a scene.

For recent surveys of image fusion theory and applications, readers are referred to a paper by Smith and Heather [18] and a collection of papers edited by Blum and Liu [19]. Another excellent source that follows the evolution of image fusion systems and algorithms over the last several years is the special sessions on Image Fusion and Exploitation.
organized by Allen Waxman et al. at the Information Fusion Conferences (2000–2004). The list of papers cited in this introduction is by no means exhaustive but it hopefully provides a flavor of some of the major developments in the field with a focus on recent advances and challenges.

Categorization

Image fusion algorithms can be categorized into low, mid, and high levels. In some literature, this is referred to as pixel, feature, and symbolic levels. Pixel-level algorithms work either in the spatial domain [20,21] or in the transform domain [6,11,22]. Although pixel-level fusion is a local operation, transform domain algorithms create the fused image globally. By changing a single coefficient in the transformed fused image, all (or a whole neighborhood of) image values in the spatial domain will change. As a result, in the process of enhancing properties in some image areas, undesirable artifacts may be created in other image areas. Zheng et al. in this issue describe a method to reduce artifacts by minimizing the ratio of the spatial frequency error. Algorithms that work in the spatial domain have the ability to focus on desired image areas, limiting change in other areas.

Multiresolution analysis is a popular method in pixel-level fusion. Burt [1] and Burt and Kolczynski [23] used filters with increasing spatial extent to generate a sequence of images (pyramid) from each image, separating information observed at different resolutions. Then at each position in the transform image, the value in the pyramid showing the highest saliency was taken. An inverse transform of the composite image was used to create the fused image. Petrović and Xydeas [24] used intensity gradients as a saliency measure. In a similar manner, various wavelet transforms can be used to fuse images. The discrete wavelet transform (DWT) has been used in many applications to fuse images [11]. More recently, the dual-tree complex wavelet transform (DT-CWT), first proposed by Kingsbury [25], was shown by Nikolov et al. [22] and Lewis et al. [26] to outperform most other grey-scale image fusion methods.

While considerable work has been done at pixel-level image fusion, less work has been done at feature-level and symbolic-level image fusion. Feature-based algorithms typically segment the images into regions and fuse the regions using their various properties [26–28]. Feature-based algorithms are usually less sensitive to signal-level noise [29]. Toet [6] first decomposed each input image into a set of perceptually relevant patterns. The patterns were then combined to create a composite image containing all relevant patterns. Nikolov et al. [30] developed a technique that fuses images based on their multi-scale edge representations, using the wavelet transform proposed by Mallat and Zhong [31]. Another mid-level fusion algorithm was developed by Piella [28,29] where the images are first segmented and the obtained regions are then used to guide the multiresolution analysis. High-level fusion algorithms combine image descriptions, for instance, in the form of relational graphs [32,33].

Role of image registration

Low-level fusion algorithms assume correspondence between pixels in the input images. Under certain conditions, it is possible to capture registered images. For instance, if the camera intrinsic and extrinsic parameters are not changed and only environmental parameters change, acquired images will be spatially registered. Some sensors are capable of capturing registered multimodality images [34]. When the images are captured with cameras where extrinsic and/or intrinsic parameters are different, we need to first register them. Although medical scanners that can capture registered multimodality images are becoming available, today, in order to fuse medical images it is required to first register them. Use of image registration in image fusion has been discussed by Zhang and Blum [35] and Goshtasby [36]. Thorough reviews of image registration methods can be found in [21,37–42]. Registration approaches that work in real time on live imagery are discussed by Heather and Smith [43].

Mid-level fusion algorithms assume that correspondence between features in the images is known. Considerable work has been done in feature correspondence, but the methods are very much image and application dependent. Methods for determining correspondence between point features [44–46], line segments [47,48], and regions [49,50] have been developed. Techniques for spatio-temporal alignment of image sequences have also been proposed [51]. High-level fusion algorithms require correspondence between image descriptions for comparison and fusion. Finding this correspondence can be considered a part of the fusion algorithm.

Applications

All imaging applications that require analysis of two or more images of a scene can benefit from image fusion. Reducing redundancy and emphasizing relevant information can not only improve machine processing of images, it can also facilitate visual examination and interpretation of images. Image fusion has been used as an effective tool in medicine. Hill et al. [52], Matsopoulos et al. [53], Wong et al. [54], and Pattichis et al. [55] fused multimodality medical images to improve diagnosis and treatment planning. Image fusion has been used in defense applications for situation awareness [56], surveillance [57], target tracking [58], intelligence gathering [59], and person authentication [60]. Image fusion has also been extensively used in remote sensing in interpretation and classification of aerial and satellite images [21,61–64].

Assessment techniques

The widespread use of multi-sensor and multi-spectral images in surveillance, remote sensing, medical diagnostics, military, and robotics applications has increased the importance of assessing the quality of different fusion techniques.
and relating it to human or computer performance when using the fused images. Better quality assessment tools are needed to compare results obtained by different fusion techniques and to derive the optimal parameters of these techniques.

Often the ideal fused image is not known or is very difficult to construct. This makes it impossible to compare fused images to a gold standard. In applications where the fused images are for human observation, the performance of fusion algorithms can be measured in terms of improvement in user performance in tasks like detection, recognition, tracking, or classification. This approach requires a well-defined task for which quantitative measurements can be made to characterize human performance. However, this usually means time consuming and often expensive experiments with human subjects.

In recent years, a number of computational image fusion quality assessment metrics have been proposed [65,66]. Metrics that accurately relate to human observer performance are of great value but are very difficult to design and, thus, are not yet available at present. In order to objectively compare different image fusion algorithms, what we also need is publicly available multispectral or multi-sensor data sets that can be used to benchmark existing and new algorithms. An increasing number of such large and diverse image collections can be found online at places like ImageFusion.org (www.imagefusion.org) or EquinoxSensors.com (“The Human Identification at a Distance (HID) Database Collection,” www.equinoxsensors.com/products/HID.html).

Special issue contents

This special issue presents some of the most recent advances in image fusion. The first four papers discuss theoretical and algorithmic aspects of image fusion with applications in surveillance and remote sensing, and the next four papers propose new metrics and approaches for evaluation of image fusion algorithms.

Lewis et al. review a number of pixel-based fusion algorithms and compare them with a novel region-based DT-CWT scheme first proposed in [26]. It is found that although region-based algorithms are more complex and usually slower, they offer a number of advantages over pixel-based algorithms, such as the ability to attenuate or enhance regions of interest or adapt to semantic rules used in vision systems for recognition and analysis.

Mitianoudis and Stathaki construct the fused transform image using independent component analysis (ICA) and topographic independent component analysis (TICA) with bases obtained by offline training. The fused transform image is created using pixel-based and region-based rules. They show improved perceptual performance using their algorithm compared to wavelet-based fusion algorithms.

Nencini et al. propose a new image fusion technique for pan-sharpening of multispectral (MS) bands based on nonseparable multiresolution analysis using the curvelet transform (CT). The lower resolution MS bands are resampled to the scale of the higher resolution panchromatic (Pan) image and are then sharpened by introducing high-pass directional details extracted from the higher resolution Pan image by means of the curvelet transform.

Mesev shows classification of multispectral IKONOS urban data by fusing them with labeled point data, leading to higher classification accuracy than that obtained by traditional multispectral analysis. To achieve optimal fusion performance, Petrović and Cootes incorporate objective fusion evaluation into the fusion process. It is shown that this strategy improves the performance of multiresolution algorithms with respect to a number of predefined fusion performance metrics. Images fused in this manner can adapt well to the changing parameters of the input.

Zheng et al. point out problems with transform-domain fusion algorithms and introduce a new metric, the ratio of spatial frequency error, to iteratively minimize over-fusion or artifacts. Chen and Varshney describe a metric inspired by the characteristics of the human visual system for quality evaluation of image fusion algorithms. In a second paper in this special issue, Petrović focuses on the methodology for perceptual image fusion assessment through comparative tests and validation of objective fusion evaluation metrics.

Emerging image fusion technologies and future directions

A number of emerging image fusion technologies are not covered in this special issue but are worth mentioning. These include: (a) the wide spread use of imaging instruments that capture registered multispectral or multimodality images; (b) the increasingly lower cost of sensors and processing power, which allow more complex multi-sensor systems to be developed; (c) the development of specialized image/video fusion boards (e.g. the Acadia Fusion System from Pyramid Vision Technologies, the ADEPT60 image fusion processor from Octe and Waterfall Solutions, the DVP-4000 visible/thermal infrared video fusion board from Equinox Corporation, and the laptop driven fusion and target detection board from BAE Systems); (d) the development of real-time software multi-sensor fusion systems using standard (often off-the-shelf) hardware; (e) recent progress in data mining and pattern recognition using fused imagery; (f) research on 3-D model building from multiple views and view synthesis; and (g) biologically inspired data fusion algorithms that mimic some of the mechanisms the human brain uses to combine information obtained by different senses.

To automate many of the processes that use image fusion, it is required to find better ways to describe images in terms of objects and their relations. Image understanding algorithms developed throughout the years and newly developed image understanding algorithms can greatly benefit machine processing of multiple images of a scene. Image fusion algorithms that take advantage of image
understanding techniques to fuse high-level image descriptions are thus in great demand.

Techniques for the construction of more accurate and richer 3-D scenes or object models are expected in the coming years. Such models will be built from various images captured by different sensors, and possibly augmented by images coming from diverse sources such as specialized image archives and the Internet. We envisage the design and deployment of more adaptive real-time image and video fusion systems in the near future, which can analyze their own performance and feedback this information to adapt their behavior and to improve the quality of the fused images they produce. More theoretical work is also needed to develop quicker and more accurate fully automated image registration techniques, quantify better image properties, and study whether and to what extent image fusion affects further image processing, such as target detection and tracking, and object recognition.

Acknowledgements

Finally, we would like to take this opportunity and thank our anonymous reviewers without whom preparation of this collection would not have been possible. We are grateful for their contributions. We also would like to thank Belur V. Dasarathy for his advice and support.

References

[27] Z. Zhang, R. Blum, Region-based image fusion scheme for concealed weapon detection, in: ISIF Fusion Conference, Annapolis, MD, July 2002.


