2014 Fifth International Conference on Intelligent Systems, Modelling and Simulation

Image Fusion: An Overview

Zaid Omar
Faculty of Electrical Engineering,
Universiti Teknologi Malaysia,
Johor, Malaysia
zaid@fke.utm.my

Tania Stathaki
Communications and Signal Processing Group,
Imperial College London,
United Kingdom
t.stathaki@imperial.ac.uk

Abstract - **An extensive overview of the field of image fusion is presented in this paper. The study firstly delves into the problem of multiple modalities that form the motivation for fusion, and discusses the main advantages of image fusion. Further, it discusses in detail the history of fusion algorithms that comprise various transform-domain and data driven methods. A section on image fusion applications, ranging from geo-spatial, medical to security fields, is also presented. Overall the paper aims to bring to light the advances and state-of-the-art within the image fusion research area so as to benefit other fields.**
Keywords: Image fusion, data and information fusion, multi-modal, multi-sensor.

I. FUSION FOR MULTI-MODALITY SENSORS

In real world applications where various optical sensors are used for image acquisition, it is often difficult to obtain a good quality image from a single sensor alone. Decisions pertaining to system conditions are very rarely made upon the measurement of a single parameter. This condition remains true across many branches of modern technology be it medicine, geography or the military.

The term 'good quality' itself tends to encompass various elements of the image scenery: illumination, sharpness, noise and contrast among others. There exists a multitude of sensor tools which include the optical camera, millimeter wave (MMW) camera, infrared (IR) and near-infrared (NIR), X-ray, radar, magnetic resonance imaging (MRI) among others, each of which tends to emphasise a different aspect of a captured image. In addition to sensor modalities, the pluralistic nature of input images is also necessary due to many other factors - occlusion of objects of interest due to smoke, fog and other unwanted objects, changing illumination in a scenery for photography applications e.g. daylight exposure at different times of day, and adjustable parameters within the sensors themselves, such as focal length.

Notwithstanding the above, large amounts of data tend to contribute to problems common in signal processing including storage requirements, computation time, limited bandwidth and inconsistency with decisions pertaining sensor systems [1], as well as the lack of a standard assessment criteria to measure sensors belonging to different modalities. It therefore makes sense to reduce these multi-dimensional data into just a compact single image that preserves relevant information and whose quality exceeds any of its inputs. The successful merging of contrasting, but complementary, features from multiple sensors should therefore be the goal of the image acquisition system.

A simple example would be of the UN Camp sequence set in Figure 1. A landscape is captured during night time using two image acquisition techniques. An NIR camera that detects strong thermal presence (within the 1.5-15μm spectrum) such as humans comprise one input. In general though, NIR sensors suffer from lower image resolution, prevalent image noise and the lack of availability of data sets which render them unsuitable for solitary use [2]. The second input is a standard image of the same scene, which is taken by a visual camera that captures strong textural background details (in the 0.45-0.7μm spectrum) but is severely limited in sparse illumination conditions. In this case the purpose is to enhance the lighting conditions in the scene and improve image qualities, so as to facilitate the detection of various moving objects and isolate pre-specified objects of interest i.e. tracking.

Figure 1. Example of Visual-NIR image fusion

Depending on user requirements, a good image must possess the ability to detect the human figure against the detailed backdrop of terrain and forestry. The formation of a good quality image is crucial as it enables us to have a proper understanding of the scenery context, which may prove

[1] Image provided by TNO Human Factors Research Institute

2166-0662/14 $31.00 © 2014 IEEE
DOI 10.1109/ISMS.2014.58

decisive in real world surveillance and target recognition systems.

Therein lays the concept of fusion. As part of the grand challenge in image processing, image fusion aims to merge the salient aspects of two or more source images from these sensors to produce a singular output image, that contains all pertinent image features [3] and has a higher visual and numerical quality than any input, which may be essential in critical applications such as military surveillance. In this case, a simple correlation of pixels will not suffice due to the extremely diverse modalities employed for image acquisition.

The purpose of fusion transcends the output image quality alone. It essentially enables users to visualise different sets of data under one scene. An important factor why fusion has been so successful is that engineers, developers and users are able to save costs by utilising signal processing techniques in lieu of designing an expensive system for image acquisition. Fusion reduces data dimensionality while preserving salient information content, thus reducing storage costs [4].

Sensory systems [2] focus mainly on how information can be extracted from sensory data. Fusion is therefore desired so as to improve visual accuracy and imply specific inferences that could not be achieved by a single sensor. The fusion framework resembles the way humans locate their surroundings by using various cues from multiple modatilies. For instance, humans utilise binocular vision whereby they combine visual content from both the left and right eyes for visual processing.

Fusion also improves system reliability by reducing uncertainty in variables, thereby increasing accuracy [4] and aiding situational awareness [2]. It enables better decision making, localisation and discrimination of objects of interest - due to the potentiality of more complete information [2]. Most importantly, it enhances and enables us to distinguish complementary information for the purpose of detection and segmentation in many security applications.

II. HISTORY OF IMAGE FUSION

The advent of multisensory applications in the 1980's particularly in the field of remote sensing, coinciding with extensive research discoveries in pyramid-based transform methods, introduced image fusion as a research area for the acquisition of higher quality images for human visualization [5].

As a research topic, image fusion can be uniquely perceived as in two ways: firstly *abstraction-wise*, as a sub-branch of data fusion, which also includes fusion of other data types such as audio, video, multi-dimensional and complex numerical data [6]. Several examples have been put in practice, namely an intrusion detection system in cyberspace which fuses network data. In advanced vehicles, location estimates from a global positioning system (GPS) chip are merged with the on-board diagnostics (OBD) system of the car for automated navigation [6]. In medicine, electroencephalography (EEG) signals are combined with electrooculography (EOG) and respiratory signals for fatigue modelling of patients [7].

Algorithm-wise, fusion is an extension of image analysis methods that also comprise other image processing tools such as compression and coding, feature extraction, registration, recognition and segmentation. As it is, many of the image fusion approaches discussed here can generally be applied towards other image and signal processing applications. Examples include ICA for blind source separation of EEG and ECG signals in medicine [8] and for facial recognition [9] and the wavelet transform for dimensionality reduction for use in image coding.

A. Early Fusion Systems

The primary purpose of fusion was initially restricted to human observation and decision making. The earliest and most basic form of fusion is pixel averaging, whereby each pixel of all input images is individually summed up and their average pixel value is incorporated into the fused image. However this method is extremely crude and its results proved unsatisfactory. The averaging method introduces artifacts especially when features present in only one input image is 'superimposed' on the fused output, as can occur in photographic multiple exposure. It also causes pattern cancellation and contrast reduction in the case where two inputs have features of equal salience but opposite contrast.

Most approaches are thus categorised under intermediate level fusion (ILF), also called fusion of features as the process involves extracting relevant features from the image using techniques such as multi-resolution analysis (MRA) and signal decomposition [4]. In [10], Nikolov et al. has classified image fusion algorithms into spatial and transform domain methods. Almost all fusion algorithms have since been based on a specific type of the transform domain, whereby a transform is performed on each input image and the transform coefficients undergo a fusion step. The resulting composite image is obtained by applying the inverse transform of the coefficients. The central idea of transform-based fusion methods is to modify the magnitude of the source image coefficients, so that edges and gradients are maximised.

B. Pyramid-based Methods

A more tangible approach to image fusion is by pyramid decomposition. An image pyramid, an early form of multi-resolution analysis (MRA), comprises a set of filtered and scaled representations of the image. Fusion is performed through selection of coefficients at every scale from the source image pyramids, followed by the inverse transform of the resulting pyramid [11].

The pyramid method was first proposed by Burt (1984) [12], who introduced the low-pass Laplacian pyramid for binocular fusion. A Laplacian pyramid is the bandpass equivalent of the Gaussian pyramid, and is obtained by the subtraction between two successive lowpass Gaussian pyramid levels. In ratio of low-pass (RoLP) pyramid the level is scaled to a ratio of two from its preceding level, whilst contrast pyramids are similar to RoLP but measure the ratio of luminence of a certain region within an image to the local background luminence. Eventually a host of improved pyramid-based schemes, including filter-subtract-decimate (FSD), morphological and gradient pyramids [13] have been proposed and used in fusion literature.

1988 saw the first application of fusion on visible, thermal and infrared images through works by Lillquist, Nandhakumar and Aggarwal and Rogers et al. [14], whilst Ajjimarangsee and Huntsberger have suggested utilising neural networks for fusion of these modalities. A weakness of the neural network method is the large overhead entailed from processing whole images. MRA techniques overcome this by decomposing images into details and average channels, where fusion can be performed in the wavelet coefficient space.

To that end, pyramid decomposition approaches have been in widespread use within the image fusion community, though it is not without drawbacks. It was often found that fused images tend to contain blocking artifacts particularly in regions where the multi-sensory input data are significantly different in modality. Another problem of pyramids is their lack of flexibility, i.e. lack of anisotropy and directional information.

C. Wavelet-based Methods

In 1993 Hunstberger and Jawerth [15] introduced us to a wavelet-based image fusion approach. Thereafter in 1995 a study by Li et al. [16] also used wavelets as alternative basis functions for multisensor image fusion, which is able to overcome the limitations of pyramid-based schemes by virtue of its directionality. This means an image is decomposed into its low frequency approximation and horizontal, vertical and diagonal edges. Essentially, wavelets enable spatial information to be incorporated into the transform process. These efforts and another work by Chipman et al. [17] kick-started a trend of using wavelet transforms for image fusion, on which a majority of current fusion algorithms in existence is based.

As opposed to the non-local and non-finite sinusoidal representations used by Fourier analysis, wavelets can be suitably used in finite domains and are a good fit for approximating data with sharp discontinuities or edges.

The wavelet transform works on the same premise as the pyramid, where the sub-band coefficients of the corresponding frequency content are merged and the subsequent inverse transform generates a synthesised fused image. Wavelet methods that are based on critically-sampling image signals, namely the discrete wavelet transform (DWT), suffer from shift variance whereby a discontinuity by a source signal could adversely affect its transform equivalent. This is a direct consequence of the down-sampling operation of wavelet transforms. An alternative to this is to use oversampled schemes, though that would increase redundancy and add to the processing time and cost. A more feasible solution was proposed by Rockinger [18] in 1997 for a shift-invariant DWT (SI-DWT) method which discards the subsampling step, therefore rendering it overcomplete. Further, Chibani [19] introduced a redundant wavelet transform (RWT), using an undecimated form of the dyadic filter tree which is best implemented via the à trous algorithm. Elsewhere the Haar wavelet solves the shift variant DWT by circumventing the down-sampling move in the decomposition process and utilising a set of new filters throughout the process of each decomposition.

The development of complex-based wavelet transforms, namely the dual-tree complex wavelet transform (DT-CWT) was able to overcome poor directional and frequency selectivity issues surrounding previous wavelet models, in addition to reducing over-completeness and easily achieving perfect reconstruction. Its complex property means the phase information derived from the transformation can be utilised for further analysis if necessary. Offshoots derived from the wavelet transform include contourlet, ridgelet and curvelet transforms that incorporate anisotropic behaviour and directional sensitivity to better facilitate the analysis of essential image features like edges.

D. Data Driven Methods

A perceived weakness of the wavelet transform, and similar transforms such as Fourier and Gabor, is the constant dependence of the basis functions on a fixed mathematical property that bears no relation statistically to the input data at hand, which often are non-linear and non-stationary. In this regard, independent component analysis (ICA) [cvejic07ICA], empirical mode decomposition (EMD) and other non-parametric and data-driven methods are considered superior as its features are directly derived from the training of data. For instance, instead of a standard bases system using wavelets, a set of bases that are suitable for particular types of images may be trained for ICA.

In contrast, EMD is an entirely adaptive fusion approach that makes no assumptions of the data. EMD works in the spatial domain where it recursively deconstructs an image into intrinsic mode functions (IMF) at different frequencies. The decomposition method utilises envelopes associated with the local maxima and minima respectively. Fusion takes place through a weighted combination of IMF's of input images. Compared to wavelet coefficients, the IMF's are not fixed and can be suited to fit the data at hand.

ICA and EMD are both examples of non-parametric regression that requires a larger sample size of data, which in turn supplies the model structure and model estimates. The works on data modelling have since prompted a study into using biologically-inspired models, namely colour vision for fusion in 1995 by Waxman and colleagues [20], in which opponent processing was applied on fusing visible and infrared images. Since then, new directions in image fusion have led to the development of a number of biological models, which are based on the way the human brain processes and combines information obtained by many different senses. ICA, for example, assumes the solution to the 'cocktail party' problem in which auditory signal sources are distinguished by the human cognitive system.

E. State-of-the-art in Image Fusion

Another emerging trend in image fusion is the application of region-based methods. This approach is borne from the understanding that regions, or wholesome objects within an image, tend to carry information of interest. It therefore makes sense to focus on regions as opposed to just individual pixels, since pixels can be processed more efficiently if they are treated as a collective group within a region rather than separate entities. Region-based fusion may therefore help to overcome some drawbacks of pixel-based fusion, like blurring, susceptibility to noise and misregistration.

Other notable approaches for image fusion include those based on statistical and estimation theory, as first proposed by Sharma et al. [21] using Bayesian fusion. Particle models and Bayesian-based fusion achieve superior performance with high requirement applications, though this often comes at a cost of higher computational complexity.

Today a plethora of image fusion methods are in existence, each with their advantages and suitability designed for certain fusion applications. Significant improvements have taken place to close the gap between computer vision and the human perception of image quality. However, the search for the gold standard in fusion remains.

III. IMAGE FUSION APPLICATIONS

It is generally acknowledged that all imaging applications that comprise analysis of multiple image inputs may benefit from image fusion. Indeed, image fusion has been found to be very useful in a variety of critical applications such as remote sensing, medical imaging, industrial defect detection and military surveillance.

Remote sensing (RS) applications are concerned with the acquisition of geo-spatial images using aerial photography by satellites and airborne sensors, such as SPOT, QuickBird, IKONOS and IRS. RS aims to deliver high quality geographic images in terms of both spatial and spectral resolutions for purposes such as urban planning, agriculture and geology. Developing a high performance sensor camera to perform such tasks is unfeasible due to factors such as the radiation energy absorbed by the sensor and the limited data transfer rate from satellite platform to ground. Rather, signal processing methods are utilised to achieve similarly high quality results.

In [22] a detailed review of fusion techniques for RS was presented. In pan-sharpening, acquired data of a given scene comprise two modalities: a panchromatic (PAN) image depicting the scene in a high spatial resolution but in a single frequency, and a multispectral/hyperspectral (MS/HS) image that captures the landscape in a multitude of spectral resolutions across the wavelength spectrum though at 1:4 the spatial resolutions of PAN. Fusion offers a practical and cost effective method to aid in distinguishing objects with salient information from RS imagery; by means of injecting the detailed spatial resolutions of PAN into a resampled version of multispectral images using methods such as the wavelet transform.

Classical fusion techniques in RS applications also include the intensity-hue-saturation (IHS) method in which the red-green-blue (RGB) coloured domain of the original MS imagery is transformed into IHS to obtain a better separation of colour for fusion with PAN images, though it often produces spectral degradation.

Others include the principal component analysis (PCA), in which the MS image is decorrelated into several components. Fusion occurs by replacing the first/principal MS component with the PAN image, coupled with the Brovey transform that multiplies each MS band by the PAN image, and finally by the division of each product by the sum of the MS bands. However these methods tend to ignore the need for high quality outputs of spectral information, which has proven essential in applications such as lithology and soil and vegetation analysis. High pass filtering (HPF) or modulation (HPM) of PAN inputs added to multispectral images are able to overcome this drawback. More recently, given the conciliatory nature of RS fusion between spatial resolution of PAN and spectral resolution of MS images, wavelet-based fusion techniques were found to be better equipped to handle this trade-off.

In medicine, image fusion and other technological advances are increasingly being relied upon for diagnostics and treatment of patients. An overview of medical image fusion was given by Pattichis et al. [23]. Fusion aids medical imaging by providing a complementary composite of various image formats stemming from multiple modalities, such as ultrasound, magnetic resonance image (MRI), computed tomography (CT), positron emission tomography (PET), and single photon emission computed tomography (SPECT), which in turn helps to delineate and distinguish targeted objects of interest such as tumours and blood vessels. In radiation oncology, a treatment plan for radiotherapy involves CT data primarily for patients' dose calculation, while the outlines of tumour are better represented in MRI. For medical diagnosis, CT best illustrates denser tissues with low distortion, while MRI offers more comprehensive information on soft tissues with higher distortion and PET provides better information on blood flow with a generally low spatial resolution [7]. Using image fusion helps to distinguish important anatomical objects of interest of both sources.

IV. CONCLUSION

An introduction and background study of the field of image fusion has been presented in this study. It aims to brief the general reader as to the basics of fusion and its motivation. The history of fusion, the advent of which coincides with the development of pyramd-based decomposition methods in the 1980's, has also been discussed. The paper concludes with a discourse on the contribution of fusion in diverse fields of technology, ranging from remote sensing to medicine.

REFERENCES

[1] A. Loza, "Statistical model-based fusion of noisy images", Computer Science Departmental Seminar, University of Bristol, 22 June 2007.

[2] N. Cvejic, D. Bull and N. Canagarajah, "Improving fusion of surveillance images in sensor networks using independent component analysis", IEEE Transactions on Consumer Electronics, Vol. 53, No. 3, 2007, 1029-1035.

[3] Z. Omar, N. Mitianoudis and T. Stathaki, "Two-dimensional Chebyshev polynomials for image fusion", 28th Picture Coding Symposium, Japan, 2010, pp. 426 -429.

[4] R.S. Blum, Z. Xue and Z. Zhang, "An Overview of Image Fusion", in R.S. Blum (Ed.), Multi-Sensor Image Fusion and Its Applications, Marcel Dekker, 2005.

[5] Y. Zhang, "Understanding image fusion", Photgrammetric Engineering and Remote Sensing, 2004, pp. 657-661.

[6] D.P. Mandic, D. Obradovic, A. Kuh, T. Adali, U. Trutschel, M. Golz, P. De Wilde, J. Barria, A. Constantinides and J. Chambers, "Data fusion for modern engineering applications: An overview", International Conference on Artificial Neural Networks 2005, LNCS 3697, 2005, pp. 715-721.

[7] G. Qu, D. Zhang and P. Yan, "Medical image fusion by wavelet transform modulus maxima", Optics Express, Vol. 9, No. 4, 2001, pp. 184-190.

[8] B. Van Dun, J. Wouters and M. Moonen, "Improving auditory steady-state response detection using independent component analysis on multichannel EEG data", IEEE Trans. On Biomedical Engineering , Vol. 54, Issue 7, 2007, pp. 1220-1230.

[9] K.C. Kwak and W. Pedrycz, "Face recognition using enhanced independent component analysis approach", IEEE Trans. on Neural Networks, Vol. 18, No. 4, 2007, pp. 530-541.

[10] S. Nikolov, P.R. Hill, D.R. Bull, C.N. Canagarajah, "Wavelets for image fusion". In A. Petrosian and F. Meyer, editors, Wavelets in Signal and Image Analysis, from Theory to Practice. Kluwer Academic Publishers, 2001.

[11] F. Sadjadi, "Comparative image fusion analysis", IEEE International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum, 2005.

[12] P.J. Burt and E.H. Adelson, "The Laplacian pyramid as a compact image code", IEEE Transaction on Communications, 31(4), 1983, pp. 532-540.

[13] J. Zeng, A. Sayedelahl, T. Gilmore and M. Chouikha, "Review of image fusion algorithms for unconstrained outdoor scenes", Proc. Intl. Conf. on Signal Processing, Vol. 2, 2006.

[14] S.K. Rogers, C.W. Tong, M. Kabrisky and J.P. Mills, "Multisensor fusion of ladar and passive infrared imagery for target segmentation", Optical Engineering, 28(8), 1989, pp. 881-886.

[15] T. Huntsberger and B. Jawerth, "Wavelet-based sensor fusion", in Proc. SPIE, Vol. 2059, 1993, pp. 488-498.

[16] H. Li, B.S. Manjunath and S.K. Mitra, "Multi-sensor image fusion using the wavelet transform", Graphical Models and Image Processing 57 (3), 1995, pp. 235-245.

[17] L.J. Chipman, Y.M. Orr and L.N. Graham, "Wavelets and image fusion", Proc. Intl. Conf. Image Processing, USA, 1995, pp. 248-251.

[18] O. Rockinger, "Image sequence fusion using a shift invariant wavelet transform", Proc. Intl. Conf. on Image Processing, 1997, pp. 288-291.

[19] Y. Chibani and A. Houacine, "On the use of the redundant wavelet transform for multisensory image fusion", Proc. IEEE Int. Conf. Electronics, Circuit and Systems, 2000, pp. 442-445.

[20] A.M. Waxman, D.A. Fay, A.N. Gove, M. Siebert, J.P. Racamoto, J.E. Carrick and E.D. Savoye, "Color night vision: fusion of intensified visible and thermal IR imagery", Proc. SPIE Conference on Synthetic Vision for Vehicle Guidance and Control, 2463, 1995, pp. 58-68.

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

[22] C. Pohl and J.L. van Genderen, "Multisensor image fusion in remote sensing: Concepts, methods and applications", Int. J. Remote Sensing, Vol. 19, No. 5, 1998, pp. 823854.

[23] C.S. Pattichis, M.S. Pattichis and E. Micheli-Tzanakou, "Medical imaging fusion application: an overview", Asilomar Conference on Signals, Systems and Computers, 2001, pp.1263-1267.