Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

A framework of region-based dynamic image fusion*

WANG Zhong-hua†, QIN Zheng, LIU Yu
(*Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*)
†E-mail: zhwangtsinghua@126.com
Received July 20, 2006; revision accepted Oct. 7, 2006

Abstract: A new framework of region-based dynamic image fusion is proposed. First, the technique of target detection is applied to dynamic images (image sequences) to segment images into different targets and background regions. Then different fusion rules are employed in different regions so that the target information is preserved as much as possible. In addition, steerable non-separable wavelet frame transform is used in the process of multi-resolution analysis, so the system achieves favorable characters of orientation and invariant shift. Compared with other image fusion methods, experimental results showed that the proposed method has better capabilities of target recognition and preserves clear background information.

Key words: Dynamic image fusion, Region segmentation, Non-separable wavelet frame
doi:10.1631/jzus.2007.A0056 **Document code:** A **CLC number:** TP751

INTRODUCTION

Image fusion is a sub-area of more general topics of Information Fusion and Data Fusion and deals with static images and dynamic images (image sequences or video data) (Maitre and Bloch, 1997). By combining the information from a range of image data obtained by different sensors or the same sensor with different imaging schemes, the composite image always presents the complementary information of different input images at the same time and provides us the integrated description of the scene or object of interest. The quality of the fusion image makes it easier for vision perceiving and computer processing. Up to now, image fusion has attracted much attention all over the world and has been successfully applied to many fields such as military affairs, medical imaging, remote sensing, security and monitoring (Huang and Chen, 2002; Jan *et al.*, 2005). It can be used to improve the performance of space awareness and the accuracy of target detection and identification, reduce the operating load and enhance the system stability (Smith and Heather, 2005).

There are several requirements that can be imposed in the process of image fusion. Firstly, all major information of the input images should be preserved in the composite image. Secondly, any artifacts or inconsistent information cannot be introduced into the fusion process. Furthermore, the fusion process should be shift invariant and the composite image should be temporally stable and consistent with the input images (Rockinger, 1997). The shift invariant, which is especially important for dynamic image fusion, means that the fusion result should not depend on the location of the object. In dynamic image fusion, a shift dependent system leads to unstable and inconsistent fusion result in both spatial domain (intraframe) and time domain (interframe). However, the human visual system is especially sensitive to the interframe moving of the image sequences.

Based on different extracting extents of the information, image fusion can take place at pixel, feature, or decision level (Abidi and Gonzalez, 1992). Pixel-based fusion is the lowest level, which consists of comparing information among the pixels in the same location or pixels in the same region in different images. So far, pixel-based fusion has attracted much

* Project (No. 2004CB719401) supported by the National Basic Research Program (973) of China

attention and many interrelated methods have been proposed such as weighted means and multi-resolution analysis (Toet, 1990; Burt and Kolczynski, 1993; Li *et al.*, 1994; Chipman *et al.*, 1995; Koren *et al.*, 1995; Rockinger, 1997; Hill *et al.*, 2002). Feature-based fusion can be achieved by the region-based fusion framework. Since more intelligent rules are applied depending on different region features, the region-based image fusion outperforms many other traditional pixel-based methods (Zhang and Blum, 1997; Matuszewski *et al.*, 2000; Piella, 2002; 2003; Piella and Heijmans, 2002; Li *et al.*, 2003; Lewis *et al.*, 2004; Cardinali and Nason, 2005).

In this paper, a new framework of region-based dynamic image fusion is proposed, which is applicable for experiment in the real world. Firstly, the technique of target detection is used to separate the object from background in the dynamic image (image sequences). Then, different fusion rules are applied in different regions to build the composite image.

FRAMEWORK OF PIXEL-BASED IMAGE FUSION

Fig.1 shows the schematic framework of the pixel-based image fusion. Sequenced images are initially processed using some multi-resolution analysis method, such as pyramid transform (Toet, 1990; Burt and Kolczynski, 1993), discrete wavelet transform (Li *et al.*, 1994; Chipman *et al.*, 1995; Koren *et al.*, 1995), discrete wavelet frame transform (Rockinger, 1997) and dual-tree complex wavelet transform (Hill *et al.*, 2002). Then, decomposition coefficients obtained by multi-resolution analysis are processed depending on fusion rules. During this stage, pixel-by-pixel maximum selection rule and area-based selection rule are both widely used, while the latter is more robust (Burt and Kolczynski, 1993; Li *et al.*, 1994). Subsequently, a fusion image is constructed by performing an inverse transform on the fused decomposition coefficients. For more information regarding the framework of pixel-based image fusion, please refer to (Toet, 1990; Burt and Kolczynski, 1993; Li *et al.*, 1994; Chipman *et al.*, 1995; Koren *et al.*, 1995; Rockinger, 1997; Hill *et al.*, 2002).

FRAMEWORK OF REGION-BASED DYNAMIC IMAGE FUSION

Since the useful information concerned is not only one pixel, but also features in the whole image, the basic pixel-based methods such as those concerning decomposition coefficients pixel-by-pixel (maximum selection rule) or performing filtering within certain window (area-based selection rule) may not be appropriate. It will be more suitable if the useful features of images in the fusion process (Piella and Heijmans, 2002) are considered. Based on this supposition, several frameworks of region-based image fusion were introduced recently (Zhang and Blum, 1997; Matuszewski *et al.*, 2000; Piella, 2002; 2003; Piella and Heijmans, 2002; Li *et al.*, 2003; Lewis *et al.*, 2004; Cardinali and Nason, 2005), in which the feature is an object or region of interest which we define as a region and the regions are elementary units in the fusion process. However, most of them are used for static image fusion. If they are applied in dynamic image (image sequences) fusion, the important moving information of image sequences will not be utilized sufficiently.

As shown in Fig.2, a new framework of region-based dynamic image fusion is proposed in this paper (the source images are supposed to be registered). The source image sequences are preprocessed firstly. Then, multi-resolution analysis is used for every preprocessed frame, and the source images are decomposed by means of steerable non-separable wavelet transform. At the same time, the technique of target detection is applied to separate the target from

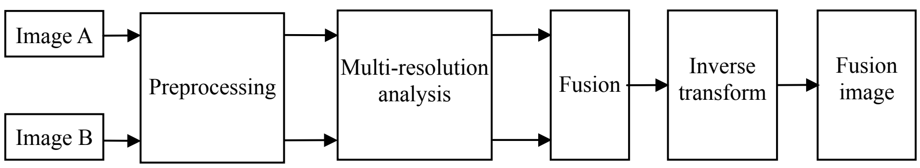


Fig.1 Framework of pixel-based image fusion

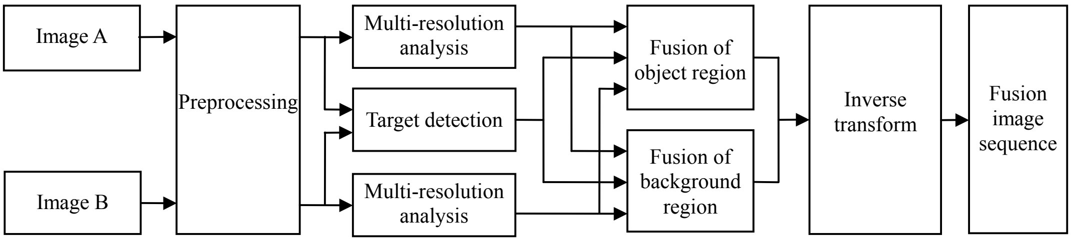


Fig.2 Framework of the region-based dynamic image fusion

background in the source image sequences and obtain a set of candidate regions. Subsequently, the decomposition coefficients are obtained by using different fusion rules in different regions. The fusion image sequences are constructed by performing an inverse transform on the fused decomposition coefficients in the end.

Steerable non-separable wavelet frame transform

In the case of dynamic image fusion, some factors must be considered before using wavelet transform. On the one hand, standard discrete wavelet transform is shift dependent because of the sampling (Simoncelli *et al.*, 1992; Strang, 1989). So the traditional framework for image fusion based on the standard discrete wavelet transform is not shift invariant and cannot be used for dynamic image fusion (Rockinger, 1997). On the other hand, the 2D discrete wavelet transform, which is extended from 1D by separable filter, only does 1D transform on every line and every row. Essentially, it is a kind of 1D transform which decomposes 2D space $L^2(Z^2)$ into the tensor product of two 1D spaces $L(Z)\otimes L(Z)$. Compared with separable wavelet transform, non-separable wavelet transform is a more general method which considers the image as a region, not several separate lines and rows (Kovačević and Vetterli, 1995). The non-separable wavelet transform has more favorable harmonic response characters and can be designed flexibly. Furthermore, the standard discrete wavelet transform is not separable (Laine *et al.*, 1995; Koren, 1996), so in the case of 2D transform the sub-bands are fixed in horizontal, vertical and diagonal directions.

To avoid shift dependence, the algorithm proposed in (Rockinger, 1997) introduced discrete wavelet frame transform into the framework of image fusion. Sampling was removed from the process of discrete wavelet frame transform. Then the shift invariant character was obtained by adding redundancy. Depending on the same principle, a framework of non-separable discrete wavelet transform is proposed, which also removes the process of sampling. The framework is shown in Fig.3b.

In order to get the useful image information from different directions, the high-pass filter $H_1(z_1, z_2)$ in Fig.3a is replaced by a set of directional band-pass filters $\{H_{1k}(z_1,z_2)|k=1, 2, \ldots, 2K+1, K\in\mathbf{Z}_+\}$, in which $H_{1k}(z_1,z_2)=H_1(z_1,z_2)\cdot A_k(z_1,z_2)$.

$A_k(z_1,z_2)$ is a kind of filter with the character of directional selection. In the frequency-domain, it is defined as follows:

$$A_k(\omega_1,\omega_2)=\frac{1}{C_{2K}}\cos^{2K}\left[\arg(\omega_1,\omega_2)-\frac{k-1}{2K+1}\pi\right],$$

where $C_{2K}=\sum_{k=1}^{2K+1}\cos^{2K}\left(\frac{k-1}{2K+1}\pi\right)$. Given any value of (ω_1,ω_2), $\sum_{k=1}^{2K+1}A_k(\omega_1,\omega_2)=1$ is always true.

Moving target detection extraction of cellular protein

The intention of moving target detection (TD) is to find the region of moving object in two sensor image sequences. The object region always contains very important information such as illegal intrusion objects in the field of secure monitoring, panzers or fighter planes in the war field.

To begin with moving TD, the first frame of the source image sequences is initially segmented by region merging (Frank and Richard, 2003). If the object is not in the initial frame, the next one is processed sequentially in the same way. Since high contrast always exists between object and background, it is appropriate to judge the target region from several candidate regions acquired by segmenting.

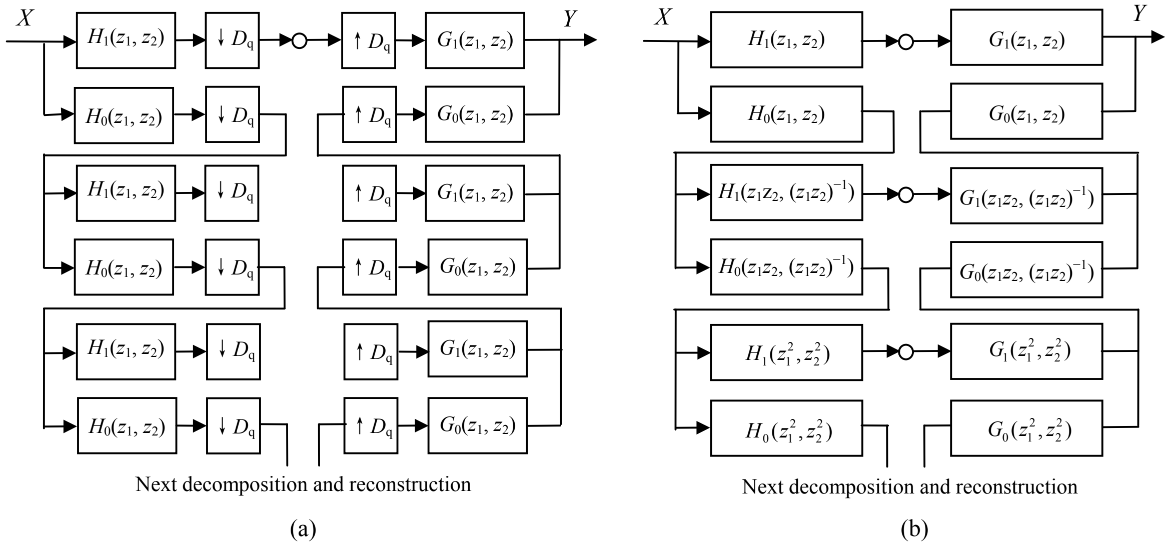


Fig.3 Framework of non-separable discrete wavelet transform (a) and non-separable discrete wavelet frame transform (b)

For each candidate region, a confidence measure is defined as the product of two functions (Yilmaz *et al.*, 2003):

$$C_i = [1 + \mathrm{e}^{-\lambda_1(\mu_\mathrm{f}-\mu_1)}]^{-1} \cdot [1 + \mathrm{e}^{-\lambda_2(\mu_\mathrm{f}-\mu_\mathrm{b}-\mu_2)}]^{-1},$$

where μ_f and μ_b are the mean of the foreground and background of the ith target, respectively. λ_1 and λ_2 are both constants which control the covariance of the Gaussian function while μ_1 and μ_2 are constants which control the offsets of Gaussian function. If the brightness ratio between a candidate region and its neighborhood goes higher, the value of the confidence will go higher at the same time and close to 1, otherwise it will be close to 0. Candidate regions with high confidence are selected as target regions.

However, if the method of image segmentation is applied to each frame, the complexity will increase greatly. In order to avoid doing the image segmentation again and again, a pattern matching method is used to locate the target regions in the next frame depending on the target regions detected in the current frame. The new method is described as follows:

(1) Let i=1.

(2) Use the method of image segmentation to process the ith frame of the source image sequences.

(3) Note the centroid of the target region in the ith frame. Use the intensity-based information of the target regions as target pattern (Make sure that the area of the pattern should be a little larger than that surrounded by the contour of the target, and that the target should be inside the pattern).

(4) If the ith frame is the last frame in the image sequence, stop; else go to Step 5.

(5) Let i=i+1. Do the pattern matching in the ith frame according to the target pattern and the region which the centroid is located in the (i−1)th frame.

(6) If the target region of the ith frame is obtained, go to Step 3; else go to Step 2.

Since the target pattern is a little larger than the real target region, the target regions obtained are always larger than the target. The target region which is approximate to the real target can be obtained by doing threshold segmentation simply.

Region-based fusion rule

A specific fusion rule is used to preserve as much useful information as possible. In this method, two source image sequences A and B are prepared to be fused. The current frames in A and B are denoted as I_A and I_B respectively. The target regions detected in I_A are represented by $T_\mathrm{A} = \{t_\mathrm{A}^1, t_\mathrm{A}^2, \cdots, t_\mathrm{A}^M\}$ and the target regions detected in I_B are represented by $T_\mathrm{B} = \{t_\mathrm{B}^1, t_\mathrm{B}^2, \cdots, t_\mathrm{B}^N\}$, where M and N are the number of the regions in I_A and I_B respectively. The union target region can be calculated by $T_\mathrm{J}=T_\mathrm{A}\cup T_\mathrm{B}$. There are three region aggregates in the current frame: single target region set T_S, superposed target region set T_O and background region set B. They are defined as follows:

$$T_O = T_A \cap T_B, \quad T_S = T_J \cap \overline{T}_O, \quad B = \overline{T}_J.$$

Obviously, T_J can also be calculated by $T_S \cup T_O$.

In the single target region, the fusion rule is described by the following formula:

$$c_F(x,y) = \begin{cases} c_A(x,y), & \text{if } (x,y) \in T_A, \\ c_B(x,y), & \text{if } (x,y) \in T_B, \end{cases}$$

where $c_A(x,y)$ and $c_B(x,y)$ are the decomposition coefficient of multi-resolution analysis for I_A and I_B respectively, $c_F(x,y)$ is the fused coefficient and (x,y) is the coordinate position of the pixel.

For the closed region t in the superposed target region set ($t \in T_O$), $M(t)$ is used to denote the similarity between the region t in I_A and I_B.

$$M(t) = \frac{2 \sum_{(x,y)\in t} [I_A(x,y) \cdot I_B(x,y)]}{\sum_{(x,y)\in t} [I_A(x,y)]^2 + \sum_{(x,y)\in t} [I_B(x,y)]^2}.$$

Then, the coefficient energy index S_i of the region t is calculated

$$S_i(t) = \sum_{(x,y)\in t} [c_i(x,y)]^2, \quad i = \text{A, B}.$$

Let $\alpha \in [0, 1]$ be the close threshold, which is usually set as α=0.85. If the similarity $M(t)$ is smaller than α, the fusion rule in superposed target region $t \in T_O$ is defined as follows:

$$c_F(x,y) = \begin{cases} c_A(x,y), & \text{if } S_A(t) \ge S_B(t), \\ c_B(x,y), & \text{otherwise.} \end{cases}$$

If $M(t) \ge \alpha$, the fusion rule is calculated by the method of weighted means

$$c_F(x,y) = \begin{cases} \varpi_{\max}(t) \cdot c_A(x,y) + \varpi_{\min}(t) \cdot c_B(x,y), \\ \qquad \text{if } S_A(t) \ge S_B(t), \\ \varpi_{\min}(t) \cdot c_A(x,y) + \varpi_{\max}(t) \cdot c_B(x,y), \\ \qquad \text{if } S_A(t) < S_B(t), \end{cases}$$

where $\varpi_{\max}(t)$ and $\varpi_{\min}(t)$ are the maximal and minimal weighted coefficients, respectively. And

$$\varpi_{\min}(t) = \frac{1}{2}\left[1 - \frac{1 - M(t)}{1 - \alpha}\right], \quad \varpi_{\max}(t) = 1 - \varpi_{\min}(t).$$

In the background region, the basic MS (Maximum Selection) rule is applied for fusion.

EXPERIMENTS

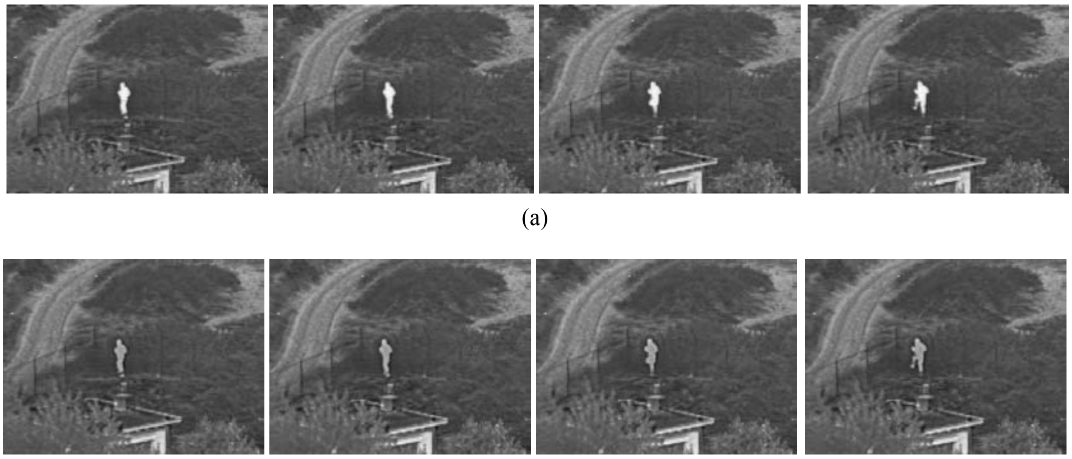
A forward looking infrared (FLIR) image sequence and an optical image sequence were used in our experiment. Fig.4a shows the four continuous frames in the fusion image sequence (the source image sequences have 32 frames). Direction-operated inseparable wavelet frame transform were applied in the process of multi-resolution analysis. Our experiment was set as follows: the number of layers for the wavelet frame transform n_l=6, the number of sub-bands in each layer n_s=3. The filter was obtained by the McClellan transform of the Biorthogonal Wavelet Filter (Kovačević and Vetterli, 1992). Meanwhile, the same input image sequence was fused by the pixel-based method in (Rockinger, 1997), which utilizes a shift invariant extension of the discrete wavelet transform. Fig.4b shows the result of the method proposed in (Rockinger, 1997), in which n_l was set to 3, and the Daubechies 9-7 Biorthogonal Wavelet Filter was used.

As listed in Table 1, the performance of the two methods was evaluated by calculating the value of information entropy, mutual information (Qu *et al.*,

Table 1 Performance of the experiment of multi-sensor dynamic image fusion

Frame sequence	Method	Information entropy	Mutual information	Edge information preservation
1	Method 1	6.3792	1.4640	0.4317
	Method 2	6.4291	1.4727	0.4382
2	Method 1	6.4285	1.4763	0.4259
	Method 2	6.4718	1.4825	0.4326
3	Method 1	6.4183	1.4668	0.4269
	Method 2	6.4641	1.4749	0.4327
4	Method 1	6.4149	1.4707	0.4237
	Method 2	6.4576	1.4759	0.4293

Method 1: the method in (Rockinger, 1997); Method 2: the method proposed in this paper



(a)

(b)

Fig.4 Fusion image sequence based on (a) method proposed in this paper and (b) method in (Rockinger, 1997)

2002), and edge information preservation (Xydeas and Petrovic, 2000). All of them were used to represent how much information is obtained from the input images. So the larger the value is, the better the fusion algorithm performs. The results of the two image fusion methods showed that the proposed algorithm in this paper outperforms the method in (Rockinger, 1997) and preserves more integrated target information and background information.

CONCLUSION

In this paper, we have proposed a new framework of region-based dynamic image fusion. Firstly, the source image sequences are preprocessed. Then the technique of target detection is applied to each image sequence to separate the targets from background and lead the following fusion process. The steerable non-separable wavelet frame transform used in this paper ensures that the decomposition coefficients have favorable harmonic response characters, and make the system shift invariant. A FLIR image sequence and an optical image sequence were used in our experiment. The results showed that the proposed algorithm outperforms traditional dynamic image fusion methods. Furthermore, since the method has better capabilities of target recognition and preserves clear background information, it can be used effectively to enhance the inspectors' target awareness capability.

References

Abidi, M., Gonzalez, R., 1992. Data Fusion in Robotics and Machine Intelligence. Academic Press, USA, p.10-50.

Burt, P.J., Kolczynski, R.J., 1993. Enhancement with Application to Image Fusion. Proceeding of the 4th International Conference on Computer Vision, p.173-182.

Cardinali, A., Nason, G.P., 2005. A Statistical Multiscale Approach to Image Segmentation and Fusion. 8th International Conference on Information Fusion, p.475-482. [doi:10.1109/ICIF.2005.1591893]

Chipman, L.J., Orr, T.M., Lewis, L.N., 1995. Wavelets and image fusion. *IEEE Transactions on Image Processing*, **3**:248-251.

Frank, N., Richard, N., 2003. On Region Merging: The Statistical Soundness of Fast Sorting with Applications. IEEE Conference on Computer Vision and Pattern Recognition, p.19-26.

Hill, P., Canagarajah, N., Bull, D., 2002. Image Fusion Using Complex Wavelets. Proceeding of the 13th British Machine Vision Conference, p.487-496.

Huang, X.S., Chen, Z., 2002. A Wavelet-based Scene Image Fusion Algorithm. Proceedings of IEEE TENCON'02, p.602-605.

Jan, M.L., Chuang, K.S., Chen, G.W., Ni, Y.C., Chen, S., Chang, C.H., Wu, J., Lee, T.W., Fu, Y.K., 2005. A three-dimensional registration method for automated fusion of micro PET-CT-SPECT whole-body images. *IEEE Transactions on Medical Imaging*, **24**(7):886-893. [doi:10.1109/TMI.2005.848617]

Koren, I., Laine, A., Taylor, F., 1995. Image Fusion Using Steerable Dyadic Wavelet Transforms. Proceeding of IEEE International Conference on Image Processing. Washinton DC, p.232-235.

Koren, I., Laine, A., Taylor, F., Lewis, M., 1996. Interactive Wavelet Processing and Techniques Applied to Digital Mammography. Proceeding of IEEE International Con-

ference on Acoustics Speech Signal Process, p.1415-1418.
Kovačević, J., Vetterli, M., 1992. Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for Rn. *IEEE Transactions on Information Theory*, **38**(2):533-555. [doi:10.1109/18.119722]
Kovačević, J., Vetterli, M., 1995. Non-separable two and three-dimensional wavelets. *IEEE Transactions on Signal Processing*, **43**(5):1269-1273. [doi:10.1109/78.382414]
Laine, A., Koren, I., Wang, W.H., Taylor, F., 1995. A Steerable Dyadic Wavelet Transform and Interval Wavelets for Enhancement of Digital Mammography. SPIE, p.736-749.
Lewis, J.J., O'Callaghan, R.J., Nikolov, S.G., Bull, D.R., Canagarajah, C.N., 2004. Region-based Image Fusion Using Complex Wavelets. Proceeding of the 7th International Conference on Information Fusion, p.555-562.
Li, H., Manjunath, B.S., Mitra, S.K., 1994. Mulitsensor Image Fusion Using the Wavelet Transform. Proceeding of IEEE International Conference on Image Processing. Austin, Texas, p.51-55.
Li, Z.H., Jing, Z.L., Liu, G., Sun, S.Y., Leung, H., 2003. A Region-based Image Fusion Algorithm Using Multiresolution Segmentation. IEEE International Conference on Intelligent Transportation Systems, p.96-101.
Maitre, H., Bloch, I., 1997. Image fusion. *Vistas in Astronomy*, **41**(3):329-335. [doi:10.1016/S0083-6656(97)00035-4]
Matuszewski, B., Shark, L.K., Varley, M., 2000. Region-based Wavelet Fusion of Ultrasonic, Radiographic and Shearographyc Non-destructive Testing Images. Proceeding of the 15th World Conference on Non-destructive Testing. Rome, p.358-363.
Piella, G., 2002. A Region-based Multiresolution Image Fusion Algorithm. Proceeding of the 5th International Conference on Information Fusion, p.1557-1564.
Piella, G., Heijmans, H., 2002. Multiresolution Image Fusion Guided by a Multimodal Segmentation. Proceeding of Advanced Concepts of Intelligent Systems, p.175-182.
Piella, G., 2003. A general framework for multiresolution image fusion: from pixels to regions. *Information Fusion*, **4**(4):259-280. [doi:10.1016/S1566-2535(03)00046-0]
Qu, G., Zhang, D., Yan, P., 2002. Information measure for performance of image fusion. *Electronics Letters*, **38**(7): 313-315. [doi:10.1049/el:20020212]
Rockinger, O., 1997. Image sequence fusion using a shift invariant wavelet transform. *IEEE Transactions on Image Processing*, **3**:288-291.
Simoncelli, E.P., Freeman, W.T., Adelson, E.H., Heeger, D.J., 1992. Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, **38**(2):587-607. [doi:10.1109/18.119725]
Smith, M.I., Heather, J.P., 2005. A Review of Image Fusion Technology. Proceeding of SPIE, p.29-45. [doi:10.1117/12.597618]
Strang, G.T., 1989. Wavelets and dilation equations: a brief introduction. *SIAM Review*, **31**(4):614-627. [doi:10.1137/1031128]
Toet, A., 1990. Hierarchical image fusion. *Machine Vision and Applications*, **3**(1):1-11. [doi:10.1007/BF01211447]
Xydeas, C., Petrovic, V., 2000. Objective image fusion performance measure. *Electronics Letters*, **36**(4):308-309. [doi:10.1049/el:20000267]
Yilmaz, A., Shafique, K., Shah, M., 2003. Target tracking in airborne forward looking infrared imagery. *Image and Vision Computing*, **21**(7):623-635. [doi:10.1016/S0262-8856(03)00059-3]
Zhang, Z., Blum, R., 1997. Region-based Image Fusion Scheme for Concealed Weapon Detection. Proceeding of the 31st Annual Conference on Information Sciences and Systems, p.168-173.