INFRARED-VISIBLE IMAGE FUSION USING THE UNDECIMATED WAVELET TRANSFORM WITH SPECTRAL FACTORIZATION AND TARGET EXTRACTION

Andreas Ellmauthaler[1], *Eduardo A. B. da Silva*[1], *Carla L. Pagliari*[2] *and Sergio R. Neves*[3]

[1] Universidade Federal do Rio de Janeiro, PEE/COPPE/DEL, Rio de Janeiro, Brazil
[2] Instituto Militar de Engenharia, Department of Electrical Engineering, Rio de Janeiro, Brazil
[3] Instituto de Pesquisas da Marinha, Rio de Janeiro, Brazil

ABSTRACT

In this work we propose a fusion framework based on undecimated wavelet transforms with spectral factorization which includes information about the presence of targets within the infrared (IR) image to the fusion process. For this purpose a novel IR segmentation algorithm that extracts targets from low-contrast environments and suppresses the introduction of spurious segmentation results is introduced. Thereby, we ensure that the most relevant information from the IR image is included in the fused image, leading to a more accurate representation of the captured scene. Moreover, we propose the use of a novel, hybrid fusion scheme which combines both pixel- and region-level information to guide the fusion process. This turns the fusion process more robust against possible segmentation errors which represents a common source of problems in region-level image fusion. The combination of these techniques leads to a novel fusion framework which is able to improve the fusion results of its pure pixel-level counterpart without target extraction. Additionally, traditional pixel-level fusion approaches, based on state-of-the-art transforms such as the Nonsubsampled Contourlet Transform and the Dual-Tree Complex Wavelet Transform, are significantly outperformed by the use of the proposed set of methods.

Index Terms— Image Fusion, Undecimated Wavelet Transform, Image Segmentation, Spectral Factorization

1. INTRODUCTION

Recent advances in the image sensor field led to the availability of a vast amount of information about a captured scene. However, in many scenarios no single sensor is able to capture all the information depicted in a single scene. For example, images taken in the visible spectrum usually exhibit good textural information but tend not to contain camouflaged targets or objects-of-interest, concealed in low light conditions. On the other hand, IR imagery does not suffer from these shortcomings but generally lacks textural information. Thus, the combination of visible and IR sensors may lead to a composite representation where both textural information (represented by the visible image) and target information (represented by the IR image) is depicted in a single image [1].

The majority of image fusion strategies can be classified into pixel- and region-level approaches [2]. As for pixel-level image fusion, each pixel in the fused image is determined by a set of pixels in the source images. Region-level fusion approaches typically segment the images into regions and perform the fusion based on the extracted regions. This usually yields advantages compared to pixel-based techniques since some drawbacks, such as blurring effects, high sensitivity to noise and misregistration can be avoided [3].

In the literature plenty of pixel-level fusion methods can be found, reaching from simple averaging techniques to more complex methods employing e.g. multiscale (MS) transforms [3]. As for the latter group, the actual fusion takes place after transforming each source image into the transform domain (e.g. wavelet domain). The final fused image is obtained by taking the inverse transform of the composite representation. Often, the fusion performance of MS approaches deteriorates when dealing with input images exhibiting overlapping or adjacent, contrast complementary information, commonly found in multimodal image pairs. In a previous work [4] we showed that for these images, fusion results can be considerably improved using an Undecimated Wavelet Transform-based fusion approach with spectral factorization (UWT-SF) of the analysis filter pair. In spectral factorization, the fusion is performed after applying a first filter pair (represented by the first spectral factor), exhibiting a significantly shorter support size than the original filter. Thus, the coefficient spreading problem that usually complicates the feature selection process, is successfully reduced.

In this paper we propose an extension of the fusion framework of [4] by including information about the presence of targets within the IR image into the fusion process. For this purpose we introduce a novel IR segmentation method which is able to detect targets in low-contrast environments without introducing spurious results. Steered by the segmentation we ensure that the targets are properly incorporated in the fused image, constructing a more accurate representation of the captured scene. Since the target extraction is performed on the decomposed images obtained after application of the first spectral factor, it can be embedded directly within the existing fusion framework. Additionally, a new hybrid fusion scheme is proposed in this work which utilizes both pixel-level and region-level information to fuse the source images. This mix turns the fusion process more robust against possible segmentation errors which may corrupt the final composite image.

The structure of this paper is as follows: Section 2 introduces the target extraction algorithm whereas its inclusion into the existing fusion framework is presented in Section 3. Section 4 discusses the obtained results and compares them with other state-of-the-art fusion schemes. Finally, our conclusions are given in Section 5.

2. TARGET EXTRACTION ALGORITHM

A number of segmentation techniques have been proposed in the context of image fusion, e.g. [1], [2], [3] and [5]. Most of these methods first employ a MS transform to the source images and extract the regions from the transform coefficients. This segmentation can be performed either separately or jointly. For separate segmentation, an independent segmentation map is generated for each input image. Joint segmentation results in a single segmentation map, incorporating regions from all source images.

In general, the final fusion performance of region-based image fusion methods highly depends on the quality of the segmentation process. For example, objects-of-interest which are concealed within other regions may not be incorporated in the fused image. On the other hand, features which are split into more than one region may cause unwanted side effects such as ringing effects in the fused image [5]. Unfortunately, in case of IR-visible image fusion, a proper segmentation map for all input images is difficult to achieve due to the different nature of the imaging sensors.

The approach taken in this work differs substantially from the

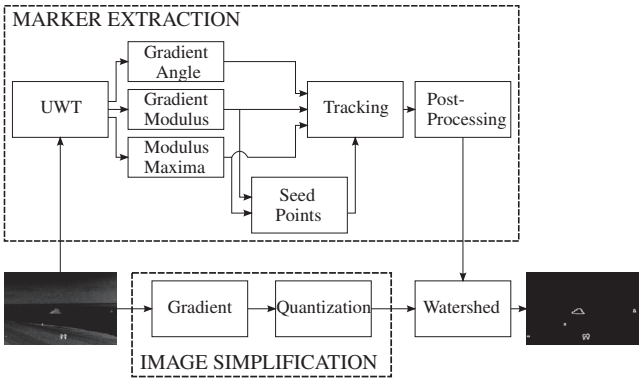

Fig. 1: *Block diagram of the proposed target extraction approach*

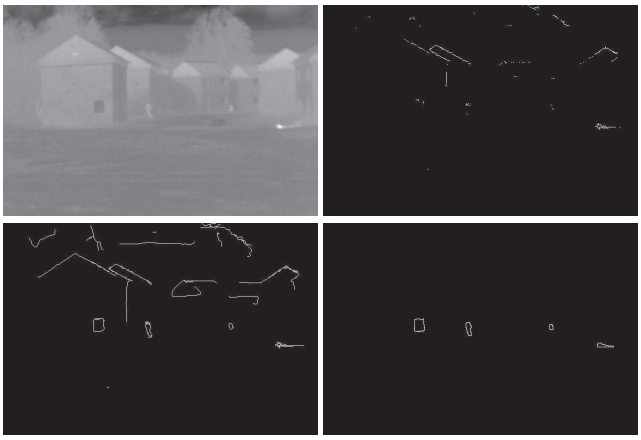

Fig. 2: *Results of the target extraction. (Top-left) Original image. (Top-right) Seed points. (Bottom-left) Result of the tracking operation. (Bottom-right) Final result after application of the watershed transformation.*

aforementioned methods. The main difference is that we will not segment both IR and visible images with the help of a single segmentation algorithm. Instead we will use *a priori* knowledge of the properties of IR images to successfully extract all objects-of-interest. An IR image is the result of the acquisition of thermal radiation of a scene, producing a two-dimensional map representing the temperature, emissivity and reflexivity variation of the respective scene [6]. Thus, we can define an object-of-interest (or target) as an enclosed region with either a larger or smaller temperature than the environment which is situated beyond transient regions such as edges.

In this work we propose the use of a marker-controlled watershed transformation to extract targets from the IR image. The marked image will be computed using the gradient modulus maxima of the Undecimated Wavelet Transform (UWT) in combination with a novel edge tracking approach. The block diagram of our proposed target extraction method is given in Fig. 1. It can be considered as consisting of three main parts: marker extraction, image simplification and watershed transformation. In the remainder of this section, these steps will be explained in detail.

2.1. Marker Extraction

The direct application of the watershed transformation usually leads to a considerable over-segmentation of the input image. One way to improve results is the use of the watershed transformation in combination with a marker image, limiting the segmentation process to some "marked" areas. Since targets in IR images are usually bounded by transient regions such as edges, we first apply an UWT-based edge detector to the input image, as described in [7]. As a result three images are obtained at each scale: the modulus of the gradient vector $\overrightarrow{\nabla}(x_j * h_{2^j})$ (x_j and h_{2^j} represent the approximation image and the upsampled low-pass filter at scale j, respectively), the angle of the steepest ascent of the gradient vector, and a binary image containing the positions of the local modulus maxima of the gradient vector. Next, the binary modulus maxima images are multiplied with the corresponding gradient modulus images and a first threshold is applied. This results in a binary image containing only those modulus maxima above the chosen threshold. After combining the thresholded images of the 1^{st} and 2^{nd} decomposition level, we obtain a first course segmentation (top-right corner of Fig. 2). From this segmentation the seed/starting points for the edge tracking operation are computed.

The proposed tracking algorithm starts by taking a seed point from the seed point list and follows the target boarder in the direction perpendicular to the gradient angle, marking each encountered pixel on its way as belonging to a target edge. Thus, in order to track a target it is sufficient that a single seed point is located on the target edge. This allows for the selection of a high first threshold, minimizing the introduction of false targets in the segmentation process. At each new point, the tracking algorithm multiplies the 8-connected neighborhood of the tracked pixel with a directional mask, discarding those pixels which do not agree with the mask's angle. Note that the used angle is computed by averaging the gradient angles of all decomposition levels. The directional masks with their corresponding directions are given in Fig. 3.

From all candidate pixels, the tracking algorithm chooses the one with the highest gradient modulus and marks it as tracked. Additionally, the 4-connected neighbors as well as all remaining candidate pixels arising from the previously tracked pixel are marked as "discarded", avoiding the use of these pixels as candidate pixels again. Note that this step is of prime importance since it circumvents the ambiguity problem of the gradient angle (there always exist two tracking directions). The tracking stops if: 1) the new point is 8-connected to a previously tracked point or 2) the averaged gradient modulus of the new point is below a certain threshold. The bottom-left image of Fig. 2 shows the result of the tracking operation on the IR image given in its top-left corner.

After the tracking operation all "true" targets from the input image form a bounded region. Next, a post-processing step cleans the tracked image from all wrongly tracked portions by removing all edge-segments which do not form a closed region. Furthermore, in order to make the result more robust against spurious targets, we remove all small objects (≤ 40 pixels) from the tracked image. Finally, the marker image is generated from the cleaned, tracked image by a morphological dilation combined with a skeletonization followed by an area opening, as described in [6].

2.2. Image Simplification and Watershed Transformation

Before performing the marker-controlled watershed transformation it is advantageous to simplify the original IR image [8]. For this purpose we compute the morphological gradient of the source image and quantize it to 100 gray levels. Once the simplified image is obtained, it is combined with the marker image and the watershed transformation is computed. The final result of the target extraction stage can be seen in the bottom-right corner of Fig. 2.

3. FUSION FRAMEWORK

In most region-level image fusion methods, the actual fusion process is solely concerned with the proper combination of the segmented regions. This is usually done by weighted averaging of associated regions within the source images. Even though this technique has been shown to be effective, its performance highly depends on the quality of the computed region map. In other words, segmentation errors such as under- or over-segmentation may lead to the absence or degradation of certain features in the fused image, respectively.

In this work, the use of a hybrid fusion scheme is proposed. Here, all extracted target regions are fused using a region-level fu-

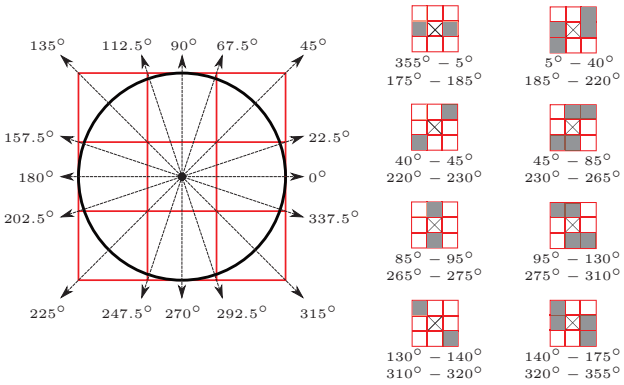

Fig. 3: *Directional masks of the tracking operation*

sion rule whereas the remaining image portions are fused using the pixel-level fusion rules, given in eqs. (2) and (3). This turns the fusion process more robust against the introduction of segmentation-induced fusion errors since we can still rely on the pixel-level algorithm to correctly incorporate an object-of-interest in the fused image, in case it was "missed" by the segmentation process. Fig. 4 shows the implementation of the proposed, overall fusion framework, combining the UWT-SF with the target extraction algorithm of Section 2, for the 1st decomposition level. In this work we will solely be concerned with the fusion of a single, registered IR-visible image pair. However, the presented fusion scheme can easily be extended to the case of multiple input images.

We start our discussion by noting that the low- and high-pass analysis filters $H(z)$ and $G(z)$ can always be expressed in the form

$$\begin{aligned} H(z) &= (1+z^{-1})P(z) \\ G(z) &= (1-z^{-1})Q(z) \end{aligned} \tag{1}$$

within any undecimated, perfect reconstruction filter bank. Hence, the image decomposition process can be split into two successive filtering operations. In this setup the first spectral factors $(1 \pm z^{-1})$ exhibit a significantly smaller support size than the original filter pair. Thus, if fusion is performed right after filtering with just the first spectral factors [4], the spreading of coefficient values around image singularities is successfully minimized. In case of complementary features located close to each other in the input images this may lead to an unwanted coefficient overlap in the transform domain. Since such coefficient overlaps generally can not be resolved by the fusion algorithm, the introduction of artifacts is greatly reduced.

After decomposing the input images using the first spectral factor, we apply the target extraction algorithm to the horizontal and vertical detail images of the IR image. Subsequently, the extracted target information is used to guide the fusion process. In this context we differentiate between two fusion scenarios. The first scenario is concerned with the fusion of transform coefficients not belonging to any extracted target. In this case the following fusion rules are used: The detail coefficients $y_I^j[m,n|p]$ and $y_V^j[m,n|p]$ of the IR and visible image, respectively, are fused using the pixel-level "choose max" fusion rule with intra-scale grouping, given by

$$y_F^j[m,n|p] = \begin{cases} y_I^j[m,n|p] & \text{if } \sum_{i=1}^{3} y_I^j[m,n|i] > \sum_{i=1}^{3} y_V^j[m,n|i] \\ y_V^j[m,n|p] & \text{otherwise} \end{cases}. \tag{2}$$

Here, m,n represents the spatial location in a given orientation band p at decomposition level j. The approximation coefficients $x_I^J[m,n]$ and $x_V^J[m,n]$ at the coarsest decomposition level J are combined

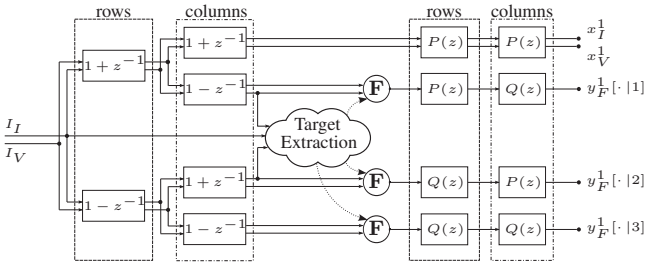

Fig. 4: *Implementation of the 1st stage of the proposed fusion framework*

using a simple averaging operation,

$$x_F^J[m,n] = \frac{x_I^J[m,n] + x_V^J[m,n]}{2}. \tag{3}$$

A different approach is taken for all transform coefficients belonging to an extracted target region. First, a measure of the matching degree between the transform coefficients (belonging to a single target region) of the IR and visible image is calculated. Consequently, each extracted target is classified as being present only in the IR image or in both source images. Based on this classification the following semantic region-level fusion rule is derived: If the extracted target is not evident in the visible image, all detail and approximation coefficients of the corresponding region are directly transferred from the IR to the fused decomposition. Otherwise, the fusion will be handled by the pixel-based fusion scheme, given by eqs. (2) and (3). Please note that we expand the extracted target region in each decomposition step by $(2^{j-1}–1)$ pixels in all directions.

In order to calculate the match metric between the same target regions within the IR and visible image, two metrics are considered. The first one measures the normalized correlation between the transform coefficients averaged over the target region $\mathcal{R}_k$ for each decomposition level and direction as described in [3]

$$M_1^j(\mathcal{R}_k|p) = \frac{2 \sum_{(m,n)\in\mathcal{R}_k} y_I^j[m,n|p] y_V^j[m,n|p]}{\sum_{(m,n)\in\mathcal{R}_k} \left|y_I^j[m,n|p]\right|^2 + \left|y_V^j[m,n|p]\right|^2}. \tag{4}$$

The final match measure is obtained after taking the absolute value of the averaged metrics, thus, bounding the final result to the interval $[0,1]$ with a value close to one suggesting a high similarity between the compared regions.

Alternatively, a second match metric is implemented which first models the wavelet coefficients of each target region as symmetric alpha-stable (SαS) random processes. Next, the similarity between the two corresponding target regions at a given decomposition level and direction is calculated by means of the Kullback-Leibler distance (KLD) between the extracted model parameters. Note that in this case a value close to zero indicates a high resemblance between the two target regions. Closed form expressions for the calculation of the model parameters and the KLD can be found in [5]. The final classification is obtained after applying a threshold where in case of the first (second) metric all targets with a match measure below (above) it, are transferred directly to the fused decomposition.

After the fusion step is complete, the filter pair represented by the 2nd spectral factor ($P(z)$ and $Q(z)$ in eq. (1)) is applied to the approximation and fused detail images. After the desired number of decompositions is reached, the approximation images are fused and the fused image is computed by applying the inverse UWT, using the corresponding synthesis filter bank without spectral factorization.

4. RESULTS

The performance of the proposed image fusion scheme was compared to the pixel-level fusion results obtained by applying the Non-

Fusion metric	DTCWT	NSCT	UWT-SF	Proposed
$Q_{AB/F}$	0.5705	0.5757	0.6008	**0.6021**
Q_P	0.7841	0.7899	0.7981	**0.7995**

Table 1: *Performance comparison of the achieved fusion metrics*

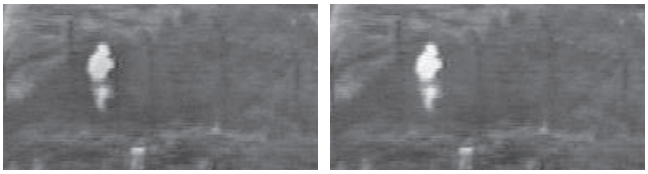

Fig. 5: *Magnified fusion results of a sample image from the "UN Camp" sequence. (Left) UWT-SF. (Right) UWT-SF with target extraction*

subsampled Contourlet Transform (NSCT), the Dual-Tree Complex Wavelet Transform (DTCWT) and the UWT with spectral factorization (UWT-SF). As for the NSCT and DTCWT, we followed the recommendations published in [9] regarding the filter choices and (in case of the NSCT) number of directions. In case of the UWT-based fusion schemes, we chose a non-orthogonal filter bank with the following coefficients [4]:

$$\begin{aligned} h[n] &= [1,1]/2 & g[n] &= [-1,2,-1]/4 \\ \tilde{h}[n] &= [1,3,3,1]/8 & \tilde{g}[n] &= [1,6,1]/4 \end{aligned} . \quad (5)$$

Please note that in this approach both synthesis filters $\tilde{h}$ and $\tilde{g}$ are positive and do not oscillate. This lack of oscillations provides a reconstruction less vulnerable to ringing artifacts. One should note that this filter bank does not obey the anti-aliasing condition, imposed by the DWT, and can therefore only be used in the undecimated case. Four decomposition levels were chosen for all transforms.

We performed simulations for 5 IR-visible image pairs using the fusion rules given in eqs. (2) and (3) for the NSCT, the DTCWT and the UWT-SF. As for the proposed fusion scheme these fusion rules also got extended with the region-level fusion rule of Section 3. The most frequently used fusion metrics, namely the $Q_{AB/F}$ [10] and Q_P [3], were employed to provide an objective evaluation of the results. Table 1 lists the average results of the simulations. It can be noticed that the UWT-SF as well as the proposed extension of the UWT-SF significantly outperform the fusion results obtained by state-of-the-art transforms such as the DTCWT and NSCT for both fusion metrics. Furthermore it can be seen that, by including target information into the fusion process, the fusion results of the UWT-SF can be further improved. This is most evident when looking at the magnified fusion results of Fig. 5. It can be seen that the proposed extension produces fused images which show improved contrast around target regions. This is particularly visible when observing the person depicted in the center of the image, which was correctly identified as an unambiguous target (not present in the visible image) by our target extraction algorithm.

Additionally, the proposed fusion method can be used to artifically "enhance" the extracted targets within the fused image. This can be accomplished by multiplying all high-pass coefficients of the UWT belonging to a target region by a constant larger than one. The corresponding effect is shown in Fig. 6, where a multiplicative factor of 2 is used. Please note that this approach may lead to the introduction of additional artifacts in the fused image. However, due to the non-oscillating nature of the synthesis filters of eq. (5), these artifacts are successfully reduced.

Both tested match metrics were able to successfully distinguish between targets solely visible in the IR image and targets contained in both source images. However, for very small targets (e.g. second target from right in Fig. 2) the SαS model-based match metric exhibits unreasonable high differences between the target regions.

Fig. 6: *Magnified fusion results of a sample image from the "Octec" sequence. (Left) UWT-SF with target extraction. (Right) UWT-ST with target extraction and additional target enhancement.*

This is due the fact that their small number of pixels makes it difficult to extract meaningful model parameters from these regions, subsequently leading to unstable KLD values.

The MATLAB implementation of the UWT-SF fusion algorithm with target extraction, together with the fusion results for all tested images, is available for download at http://www.lps.ufrj.br/profs/eduardo/fusion/ICIP_2012.

5. CONCLUSION

In this paper an extension of the UWT-based pixel-level image fusion framework with spectral factorization (UWT-SF) of [4] is introduced. The proposed algorithm extracts all targets which are not contained in the visible image from the IR image and transfers them directly to the fused image. We showed that our solution is able to improve the objective fusion results of the UWT-SF and perceptively leads to a fused image with increased contrast around target regions. Furthermore, our proposed extension can be used to artifically enhance the visibility of the extracted targets within the fused image, improving target detection, localization and identification.

6. ACKNOWLEDGMENTS

The authors would like to thank the Brazilian Funding Agencies CAPES, CNPq and FINEP, and the Ministry of Science Technology and Innovation-MCTI for their financial support, as well as the Army Technological Center-CTEx for cooperating with this research.

7. REFERENCES

[1] J.J. Lewis, R.J. O'Callaghan, S.G. Nikolov, D.R. Bull, and N. Canagarajah, "Pixel- and region-based image fusion with complex wavelets," *Information Fusion*, vol. 8, no. 2, pp. 119–130, April 2007.

[2] Z. Zhang and R.S. Blum, "Region-based image fusion scheme for concealed weapon detection," in *Proceedings of the 31st Annual Conference on Information Sciences and Systems*, March 1997, pp. 168–173.

[3] G. Piella, *Adaptive Wavelets and their Applications to Image Fusion and Compression*, Ph.D. thesis, University of Amsterdam, 2003.

[4] A. Ellmauthaler, E.A.B. da Silva, and C.L. Pagliari, "Multiscale Image Fusion Using the Undecimated Wavelet Transform With Spectral Factorization and Non-Orthogonal Filter Banks," *submitted for publication*.

[5] T. Wan, N. Canagarajah, and A. Achim, "Segmentation-driven image fusion based on alpha-stable modeling of wavelet coefficients," *IEEE Transactions on Multimedia*, vol. 11, no. 4, pp. 624–633, June 2009.

[6] S.R. Neves, E.A.B. da Silva, and G.V. Mendonca, "Wavelet-watershed automatic infrared image segmentation method," *Electronics Letters*, vol. 39, no. 12, pp. 903–904, June 2003.

[7] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710–732, July 1992.

[8] F. Meyer and S. Beucher, "Morphological segmentation," *Journal of Visual Communication and Image Representation*, vol. 1, no. 1, pp. 21–46, 1990.

[9] S. Li, B. Yang, and J. Hu, "Performance comparison of different multi-resolution transforms for image fusion," *Information Fusion*, vol. 12, no. 2, pp. 74–84, 2011.

[10] C.S. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, no. 4, pp. 308–309, February 2000.