

# A GRADIENT-BASED HYBRID IMAGE FUSION SCHEME USING OBJECT EXTRACTION

*Milad Ghantous, Soumik Ghosh and Magdy Bayoumi*

The Center for Advanced Computer Studies

University of Louisiana at Lafayette

{mmg4545, sxx5317, mab}@cacs.louisiana.edu

## ABSTRACT

This paper presents a new hybrid image fusion scheme that combines features of pixel and region based fusion, to be integrated in a surveillance system. In such systems, objects can be extracted from the different set of images due to background availability, and transferred to the new composite image with no additional processing usually imposed by other fusion approaches. The background information is then fused in a multi-resolution pixel-based fashion using gradient-based rules to yield a more reliable feature selection. According to Piella and Petrovic quantitative evaluation metrics, the proposed scheme exhibits a superior performance compared to existing fusion algorithms.

**Index Terms**— Image fusion, Multi-resolution decomposition, object Extraction, surveillance Systems

## 1. INTRODUCTION

The increased importance gained by security and surveillance systems over the recent years has motivated the research community to investigate and explore more effective and flexible solutions to the challenges imposed by such systems. Consequently, image fusion has drawn a lot of attention and became a major part of any surveillance system, due to its important role in combining complementary information from different sources of optical sensors (e.g. Visible and Infrared) into one composite image or video, thus minimizing the amount of data that needs to be stored while preserving all the salient features of the source images, and more importantly, enhancing the informational quality of the surveyed scene. The process of image fusion must ensure that all the salient information present in the source images are transferred to the composite image. Information fusion can be performed at three levels: pixel, object, and decision level. A number of pixel-level fusion techniques, in which the source images are processed and fused on a pixel basis or according to a small window in the neighborhood of that pixel can be found in literature. These range from simple averaging to more complex multiresolution techniques such as pyramids and wavelets [1]. Recent advances include the evolution to region based techniques, in which, the source images are first segmented to yield a set of regions that constitute the image, followed by the fusion of the corresponding regions [2-5]. In spite of the ability of region-based approaches of using more intelligent fusion rules and helping in overcoming the problem of mis-registration, they only achieve performances comparable to their pixel-based counterpart. However, the disadvantage of such schemes lies in the complexity of the multi-resolution segmentation algorithm required prior to the fusion process, which usually outweighs the benefits of using the region-based approach in the first place.

In this paper, a new image fusion algorithm for surveillance applications is presented. It uses the background images collected from different optical sensors, and combines the benefits of pixel and region based approaches into a hybrid scheme, in which, the important objects/regions are first extracted from the source images and classified into two categories: exclusive and mutual. Exclusive objects are transferred to the composite image with no further processing. On the other hand, mutual objects/regions undergo an object/region activity measure to select the suitable fusion rule. And finally, to insure the transferability of all the important visual information present in the source images, including the un-extracted objects, the background information is fused in a pixel-based fashion using gradient activity measures. The rest of the paper is organized as follows. Section 2 discusses the background and the related work, followed by the proposed fusion scheme in section 3. The simulation results are presented in section 4. Section 5 concludes the paper.

## 2. BACKGROUND AND RELATED WORK

### 2.1 Pixel-Based Fusion

The process of combining different types of source images at each pixel location (or according to a small window in the neighborhood of that pixel) is called pixel-level fusion. The most commonly used techniques can be grouped into two major categories: arithmetic and biologically-based methods. The weighted combination (e.g. Average) and principle component analysis (PCA [6]) are widely known in arithmetic pixel-level fusion. Despite the low complexity and the computational efficiency, the fused image suffers from a contrast loss and attenuation of salient features [7]. The limitations of arithmetic methods led to the development of the biologically-based methods which are inspired from the human visual system that is sensitive to local contrast changes such as edges. Multiresolution (MR) decomposition methods were found to be very effective in representing such changes and therefore, they were used for the purpose of image fusion. The process starts by decomposing the source images into different resolutions so that the features of each image are represented at different scales. Intelligent rules (e.g. Select max) are then used to fuse the corresponding MR coefficients and create a decomposed fused image, followed by an inverse decomposition to yield the final fused image. Burt and Adelson proposed one of the earliest MR techniques: the Laplacian Pyramid (LP) originally developed for image compression [8]. Variations of the LP were proposed in the literature with the goal of enhancing the fusion performance such as Ratio of Low Pass Pyramid (RoLP), Contrast Pyramid, Gradient Pyramid, FSD and Morphological Pyramid. One disadvantage of pyramid methods is the over-complete set of transform coefficients. Wavelet decomposition schemes on the other hand, do not suffer from this

shortcoming. Moreover, it became possible, with Mallat's algorithm, to decompose 2-D signals (i.e. image) using 1-D filter banks [9]. The Discrete Wavelet Transform (DWT) is widely used in image fusion since it captures the features of an image not only at different resolutions, but also at different orientations. However, DWT is shift variant due to the sub-sampling at each level of decomposition. The Shift-Invariant DWT (SIDWT) [10] solves this problem at the cost of an over-complete signal representation. Fortunately, the recently introduced Dual-Tree Complex Wavelet Transform (DT-CWT [11]) achieves a reduced over-completeness compared to SIDWT and a better directionality compared to the DWT, by representing the image at six different orientations.

## 2.2 Region-Based Fusion

The motivation for region-based approaches is derived from the fact that a pixel usually belongs to an object or a region in an image. Therefore, it is more reasonable to consider regions instead of individual pixels. This limitation of pixel-based schemes led to the development of region-based schemes, in which, the source images are first segmented to yield a region map comprising all the regions that constitute the image. A set of fusion rules is then applied to the corresponding regions from the different set of images depending on region activity levels and similarity match measures. Region-Based approaches may help to avoid several drawbacks found in their pixel-based counterpart such as sensitivity to noise, blurring effects and misregistration. In [2], the source images are first decomposed using the MR transform  $\psi$ . A segmentation map  $R=\{R^{(1)},R^{(2)},\dots,R^{(K)}\}$ , where  $K$  is the highest level of decomposition, is constructed based on a pyramid linking method [12]. A region activity level is then calculated for each region and a decision map is constructed accordingly. In the same vein, the authors in [3] adopt a texture-based image segmentation algorithm to guide the fusion process. An adapted version of the combined morphological-spectral unsupervised segmentation is used in [4]. Unlike the previous methods, the fusion process is carried in the ICA domain instead of the wavelet domain. In [5], images are initially segmented and several fusion methods are then applied and compared based on the Mumford-Shah energy model. The fusion algorithm with the maximum energy is selected.

## 3. PROPOSED FUSION SCHEME

The aim of this paper is to ensure the transferability of the most relevant information found in source images into the new composite image with the least amount of required processing. A new hybrid scheme that combines the advantages of the pixel-based and the region-based approaches is developed. The basic idea of this work is based on two observations:

- In most applications, few regions/objects contribute to the majority of the important information that needs to be transferred, while the remaining regions belong to the background.
- In surveillance systems, a background image for each sensor type is usually accessible.

### 3.1 Fusion Algorithm

Due to the availability of a background image, we are able to extract the objects of interest (OOI) from the source images by

applying a simple background subtraction that is lower in complexity and more efficient in implementation when compared to the segmentation techniques mentioned in the previous section. Note also that, by applying background subtraction, we are able to separate the OOI from the background and hence apply more intelligent fusion rules to the decomposed images using DT-CWT to ensure that those objects are transferred to the new image. The background information fusion follows a window-based approach to ensure that all the un-extracted objects are conveyed to the composite image as well. The overall fusion process is illustrated in figure 1.

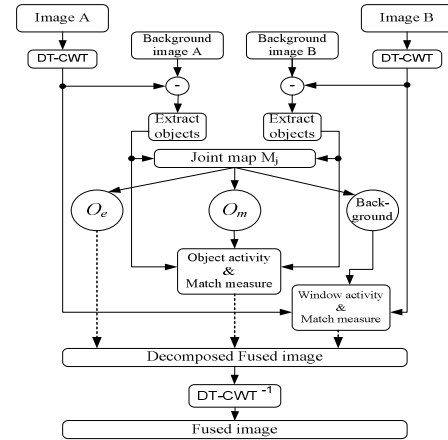


Fig. 1 Proposed image fusion scheme

#### 3.1.1 Object extraction and Categorization

Using DT-CWT, each source image is first decomposed into an approximation image and six different detail images. Let  $R_i$  denote the background approximation image and  $X_i$  the source approximation image for optical sensor  $i$ ,  $i=1,\dots,m$ . Then, the difference image  $D_i$  can be calculated as follows.

$$D_i = |X_i - R_i| \quad (1)$$

A set of binary maps  $M=\{M_i, i=1..m\}$  is then constructed according to a threshold  $\tau$ .  $\tau$  usually depends on the application, however, it can be calibrated online or offline. Equation (2) illustrates this step.

$$M_i(x,y) = \begin{cases} 0 & \text{if } D_i(x,y) < \tau \\ 1 & \text{if } D_i(x,y) \geq \tau \end{cases} \quad (2)$$

In  $M_i$ , '0' denotes a background pixel, while '1' denotes an OOI pixel. It is known that, the objects in a visible image may not appear in an infrared image and vice versa. Since it is vital to transfer the non-background objects to the fused image, it is required that we classify the extracted objects in two categories: mutual and exclusive. An exclusive object ( $O_e$ ) is an object that appears in one image only, while a mutual object ( $O_m$ ) appears in all the images. Hence, we first construct a joint map  $M_{j,k}=U\{M_i, i=1..m\}$  at the coarsest resolution (approximation at level  $K$ ) comprising regions with four different values to separate pixels that belong to the background, an object in the visible image, an object in the infrared image, or a mutual object. Hence, all the regions that belong to the background have the same unique value, as well as the regions that belong to an exclusive object and to a mutual object. Note that, in this paper, we use two source images,  $m = 2$ ; however, the proposed scheme can be easily extended to  $m > 2$ . To obtain the joint maps at higher resolutions ( $M_{j,l} \ l=1..k-1$ ),

$M_{j,k}$  is double-sized by substituting each value in  $M_{j,k}$  by a  $2 \times 2$  matrix in which each element has the same value as  $M_{j,k}$ .

### 3.1.2 Fusion Rules

The overall fusion process can be divided into four sub-processes: the approximation coefficients, the exclusive objects  $O_{e,i}$   $i=A,B$ , the mutual objects  $O_m$ , and the background fusion.

For the fusion of the approximation coefficients, we apply a simple averaging method:

$$C_F^l(x,y) = \frac{C_A^l(x,y) + C_B^l(x,y)}{2} \quad (3)$$

where  $C_F^l$  denotes the fused coefficient,  $C_A^l$  and  $C_B^l$  are the approximation coefficients of the sources images A and B respectively.

For the fusion of  $O_e$ , and since an exclusive object appears in one of the source images and according to the joint map  $M_{j,l}$  at resolution level  $l$ , the detail coefficients of the six different orientation bands of the source image, to which  $O_{e,i}$  belongs, are transferred to the fused image with no further processing as follows:

$$\forall(x,y) \in O_{e,i}, C_F(x,y,l) = \begin{cases} C_A(x,y,l) & \text{if } i = A \\ C_B(x,y,l) & \text{if } i = B \end{cases} \quad (4)$$

Similar to region-based approaches, a mutual object  $O_m$  is considered a separate region and hence, a weighted combination is employed as shown in equation (5).

$$C_F(O_m) = \omega_A C_A(O_m) + \omega_B C_B(O_m) \quad (5)$$

Therefore, a region activity level and match measure should be derived to determine the suitable fusion weights. The region activity is calculated as the local energy (LE) of the region as follows:

$$LE_i(O_m) = \frac{1}{N} \sum_{C_i(x,y,l) \in O_m} C_i(x,y,l)^2 \quad (6)$$

Where  $i=A,B$ ,  $N$  is the area of region  $O_m$ , and  $C(x,y,l)$  is the DT-CWT coefficient at location  $(x,y)$  and level  $l$ . The region match measure is then derived as follows:

$$Match_{AB}(O_m) = \frac{2 \times \left[ \sum_{C_i(x,y,l) \in O_m} C_A(x,y,l) \cdot C_B(x,y,l) \right]}{LE_A(O_m) + LE_B(O_m)} \quad (7)$$

If  $Match_{AB}(O_m)$  is less than a threshold  $\alpha$ , the fusion reduces to a "select max" based on the activity level of the object/region, in which, the object with higher activity level is transferred to the fused image (i.e. the weights reduce to the values 0 or 1)

$$\forall(x,y) \in O_m, C_F(x,y,l) = \begin{cases} C_A(x,y,l) & \text{if } LE_A(O_m) \geq LE_B(O_m) \\ C_B(x,y,l) & \text{if } LE_A(O_m) < LE_B(O_m) \end{cases} \quad (8)$$

On the other hand, if the match measure exceeds  $\alpha$ , the weights  $\omega_A$  and  $\omega_B$  are found by:

$$\omega_A = \begin{cases} \omega_{\min} & \text{if } LE_A(O_m) < LE_B(O_m) \\ \omega_{\max} & \text{if } LE_A(O_m) \geq LE_B(O_m) \end{cases} \quad (9)$$

With  $\omega_{\min} = \frac{1}{2} \left( 1 - \frac{1 - Match_{AB}(O_m)}{1 - \alpha} \right)$ ,  $\omega_{\max} = 1 - \omega_{\min}$ ,  $\omega_B = 1 - \omega_A$

Finally, in order to guarantee that all the remaining information present in the source images including the background and the un-

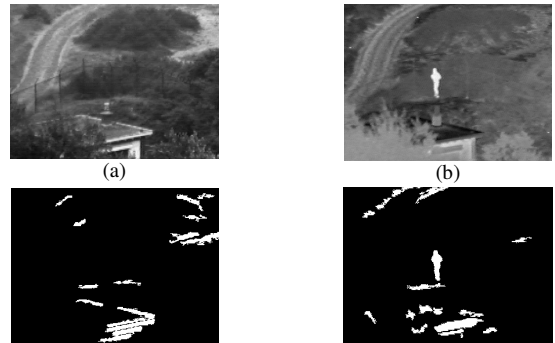
extracted regions, a simple window-based approach is employed. According to  $M_{j,l}$ , for every pixel that belongs to the background, we calculate the activity level of a small neighborhood around that pixel which is usually a size of  $3 \times 3$  or  $5 \times 5$ . However, we propose to use the average gradient of the window as a measure of the activity level of that window. The gradient clearly provides a better understanding of the visual importance of an area in the image. In other words, a larger gradient indicates a possible existence of an edge and hence, results in a more intelligent fusion. The activity level of an  $N \times N$  window  $W$  is calculated as follows:

$$Activity(W) = \frac{\sum_{(x,y) \in W} \sqrt{[C(x,y,l) - C(x+1,y,l)]^2 + [C(x,y,l) - C(x,y+1,l)]^2}}{\sqrt{2(N-1)^2}} \quad (10)$$

A match measure is then calculated for the corresponding windows similar to equation (7). Following the same reasoning for region fusion, the optimal weights  $\omega_A$  and  $\omega_B$  are found. See equations (8) and (9). After all the coefficients of the source images are fused, an inverse DT-CWT is applied to yield the final composite image F.

## 4. EVALUATION AND SIMULATION RESULTS

The proposed fusion scheme was tested on visible (fig. 2(a)) and infrared (fig 2(b)) surveillance images of the same scene. The performance comparison was evaluated through the mutual information (MI) and the newly introduced objective performance metrics:  $Q_p$  proposed by Piella[13], which utilizes local measures to estimate the amount of salient information transferred from the inputs to the fused images using image quality index, and  $Q_x$  by Xydeas and Petrovic [14] which evaluates the fusion performance based on the amount of edge information conveyed from input images to the fused image. The parameters used in the simulations are as follows:  $\tau = 5$ ,  $\alpha = 0.95$ ,  $3 \times 3$  window-based fusions,  $8 \times 8$  sliding window for  $Q_p$  [11]. Our Proposed Algorithm using Local Energy and proposed Gradient Activity levels (dubbed PALE and PAGA respectively), is compared against the average method, window-based Laplacian pyramid (LP), and finally pixel and window-based DT-CWT (DT-CWT software code is provided by Dr. N. Kingsbury). The qualitative evaluation of the proposed fusion scheme is shown in figure 2, while table 1 summarizes the quantitative comparison. Clearly, the proposed fusion algorithm exhibits higher performance (5 to 47% improvement). Moreover, figure 3 illustrates a comparison between PALE and PAGA to evaluate the effect of applying the gradient activity measure as an alternative of the local energy of a window. The simulations show that applying the gradient results in a better fusion according to  $Q_p$  and  $Q_x$ . (Around 10.2% improvement)



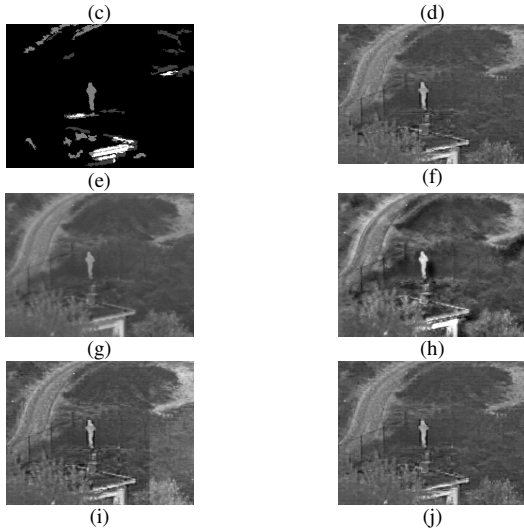


Fig. 2 (a)Visual Image (b)Infrared Image (c) $M_{\text{visual}}$  (d) $M_{\text{infrared}}$  (e) $M_j$  (f) Proposed scheme (g)Average Fused image (h) Window-based Laplacian Pyramid (i)Pixel-Based DT-CWT (j)Window-Based DT-CWT (Images provided by Dr. Lex Toet)

Table 1. Performance comparison

	Qp	Qx	MI
AVERAGE	0.8940	0.2993	2.0040
WINDOW-LP	0.9113	0.3633	2.0160
PIXEL-DT-CWT	0.9334	0.4059	2.0248
WINDOW-DT-CWT	0.9327	0.4126	2.0256
PALE	0.9377	0.4336	2.0280
<b>PAGA</b>	<b>0.9378</b>	<b>0.4411</b>	<b>2.0284</b>

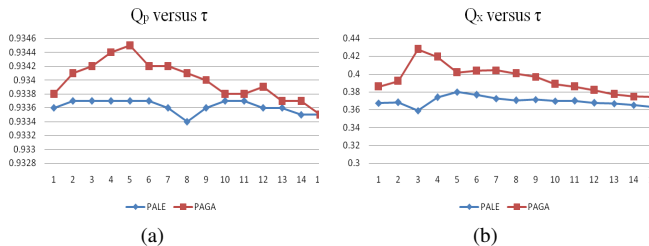


Fig. 3 PALE vs PAGA (a)  $Q_p$  versus  $\tau$  (b)  $Q_x$  versus  $\tau$

## 5. CONCLUSION

In this paper, a new hybrid image fusion scheme that combines features from pixel and region based approaches is presented. The main idea lies in replacing complex multi-resolution segmentation techniques by a simple background subtraction that is applied to only extract the objects of interest found in those images. Furthermore, objects that appear in one of the images need not to be processed or fused; however, they can be directly transferred to the fused image. Objects that appear in more than one source image follow a region-based fusion. Finally, the background is fused using a simple gradient based window approach to ensure the transferability of all the background information and the un-

extracted regions to the fused image. The proposed fusion scheme exhibits higher performance compared to existing fusion algorithms according to the mutual information metric and the objective measures proposed by Piella, Xydeas and Petrovic.

## 6. ACKNOWLEDGMENT

This material is based upon work supported by the U.S. Department of Energy (DoE), the Louisiana Board of Regents contract DOE/LEQSF-ULL, the Governor's Information Technology Initiative and the National Science Foundation under grant No. INF 9-001-001, OISE-0512403.

## 7. REFERENCES

- [1] J. Zeng, A. Sayedelahl, T. Gilmore, M. Chouikha, "Review of Image Fusion Algorithms for Unconstrained Outdoor Scenes", Proc. IEEE Int. Conf. on Signal Processing, Vol. 2, 2006
- [2] G. Piella, "A region-based multiresolution image fusion algorithm", 2002. Proc. of the Fifth Int. Conf. on Information Fusion, vol. 2, pp. 1557-1564, 2002.
- [3] Z. Li, Z. Jing, G. Liu, S. Sun, H. Leung, "A region-based image fusion algorithm using multiresolution segmentation", Proc. IEEE int. Conf. on Intelligent Transportation Systems, Vol. 1, pp. 96-101, 2003
- [4] N. Cvejic, J. Lewis, D. Bull, N. Canagarajah, "Adaptive Region-Based Multimodal Image Fusion Using ICA Bases", Proc. IEEE 9<sup>th</sup> Int. conf. on information fusion, pp.1-6, 2006
- [5] Y. Zhang, L. Ge, "Region-based Image Fusion Using Energy Estimation", Proc. IEEE 8<sup>th</sup> int. conf. on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, Vol. 1, pp. 729-734, 2007
- [6] J. A. Richards, "Thematic Mapping from Multitemporal Image Data Using The Principal component Transformation", Remote Sensing of Environment 16, pp. 36-26, 1986
- [7] M. Smith, J. Heather, "Review of Image Fusion Technology in 2005", Waterfall Solutions
- [8] P.J. Burt, E.H. Adelson, "The Laplacian Pyramid as a Compact Image Code", IEEE Trans. Commun. COM-31, 532-540, 1984
- [9] S.G. Mallat, "A Theory for Multiresolution signal decomposition: The Wavelet Representation", IEEE Trans. Pattern Anal. Machine Intell., 1989
- [10] O. Rockinger, "Image Sequence Fusion Using a Shift Invariant Wavelet Transform", IEEE Trans. Image Processing, vol.3, pp. 288-291, 1997
- [11] N G Kingsbury, "A Dual-Tree Complex Wavelet Transform with improved orthogonality and symmetry properties", Proc. IEEE Conf. on Image Processing, 2000
- [12] P.J. Burt, T.H. Hong, A. Rosenfeld, "Segmentation and estimation of image region properties through cooperative hierarchical computation", IEEE Trans. On Systems, Man, and Cybernetics, vol. 11, pp. 802-809, 1981
- [13] G. Piella, H. Heijmans, "A New Quality Metric For image Fusion", Proc. IEEE Int. Conf. on Image Processing, vol.2, pp. 173-176, 2003
- [14] C.S. Xydeas and V. Petrovic, "Objective Image Fusion Performance Measure", Electronics Letters, Vol. 36, pp. 308-309, 2000