

# Statistical Multisensor Image Segmentation in Complex Wavelet Domains

Tao Wan<sup>1</sup> and Zengchang Qin<sup>1,2</sup>

<sup>1</sup> Robotics Institute, Carnegie Mellon University, Pittsburgh, USA,

<sup>2</sup> Intelligent Computing and Machine Learning Lab,  
School of ASEE, Beihang University, Beijing, 100191, China  
{taowan,zcqin}@andrew.cmu.edu

**Abstract.** We propose an automated image segmentation algorithm for segmenting multisensor images, in which the texture features are extracted based on the wavelet transform and modeled by generalized Gaussian distribution (GGD). First, the image is roughly segmented into textured and non-textured regions in the dual-tree complex wavelet transform (DT-CWT) domain. A multiscale segmentation is then applied to the resulting regions according to the local texture characteristics. Finally, a novel statistical region merging algorithm is introduced by measuring a Kullback-Leibler distance (KLD) between estimated GGD models for the neighboring segments. Experiments demonstrate that our algorithm achieves superior segmentation results.

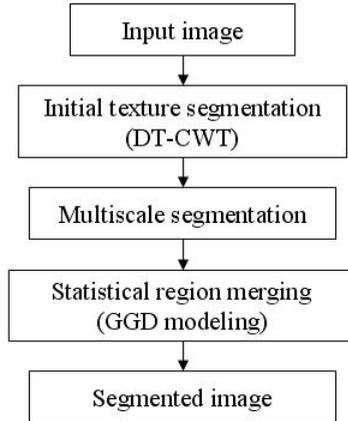
**Keywords:** multisensor image segmentation, statistical modeling, complex wavelets.

## 1 Introduction

The purpose of image segmentation is to produce a partition of the underline scene so that there are similar intensity or texture characteristics within each partitioned region. These features would favour a multisensor image fusion algorithm performed on each region in the fusion process. For this specified application, the segmentation should ideally have the following properties:

- All the salient objects, especially small-sized, in the input images are detected by the segmentation algorithm;
- All required features should be segmented as single separate regions;
- The segmented images should be able to provide precise region boundaries.

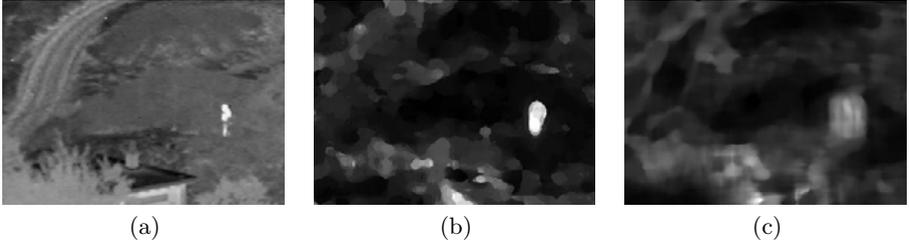
Although the above criteria can also be applied to a general task for segmenting natural images, they are the key factors of obtaining effective segmentations from the available sensors to perform a successful region-based image fusion (Lewis et al. 2007, Wan et al. 2009). In this paper, we present an automated image segmentation algorithm to fulfill such goals.



**Fig. 1.** The proposed image segmentation algorithm

In the previous work (Wan et al. Apr. 2007), we have developed a multi-scale image segmentation algorithm based on the dominant color and homogeneous texture features. A statistical approach (Wan et al. Sep. 2007) using a non-Gaussian model was also proposed for segmenting natural color images. This paper introduces alternative approaches to the texture feature extraction and the region merging components of the algorithm in (Wan et al. Apr. 2007, Wan et al. Sep. 2007) to process multisensor images. The texture feature extraction is improved by performing in the wavelet domain, while the region merging is based on the Kullback-Leibler distance (KLD) (Do and Vetterli 2002) as a similarity metric between two neighboring segments.

As shown in Fig.1, the segmentation algorithm comprises three components. First, the image is roughly segmented into textured and non-textured regions using the dual-tree complex wavelet transform (DT-CWT) (Kingsbury 2001). The initial texture map is generated to discover the small salient objects that may be neglected by the following segmentation procedure. A multiscale segmentation is then applied to the resulting regions by considering different local texture features. A boundary refinement step is utilized to improve the segmentation performance. Finally, a statistical region merging approach is developed to measure the KLD between two generalized Gaussian models corresponding to the adjacent segments, ensuring the important features are grouped into individual regions. This three-component method is specifically designed to meet the criteria that could be useful in the segmentation-driven image fusion application. The key to the success of the proposed algorithm is that the statistical modeling techniques and wavelet textures are integrated into an unified framework to achieve precise and robust segmentation for multisensor images.



**Fig. 2.** The resulting images by using the truncated median filter. (a) “UNCamp” IR image. (b) The filtered image using DT-CWT. (c) The filtered image using Gabor decomposition.

## 2 Initial Texture Segmentation

Wavelets have emerged as an effective tool to analyze texture information as they provide a natural partition of the image spectrum into multiscale and oriented subbands. In this work, we use a three-scale DT-CWT (Kingsbury 2001) with six orientations, which is able to provide approximate shift invariance and directional selectivity while preserving the usual properties of perfect reconstruction and computational efficiency. The feature value  $T(x, y)$  at the pixel location  $(x, y)$  is defined as:

$$T(x, y) = \{w_{l,\theta}(x, y)\} \quad l = 1, 2, 3 \quad \theta = 1, 2, \dots, 6 \quad (1)$$

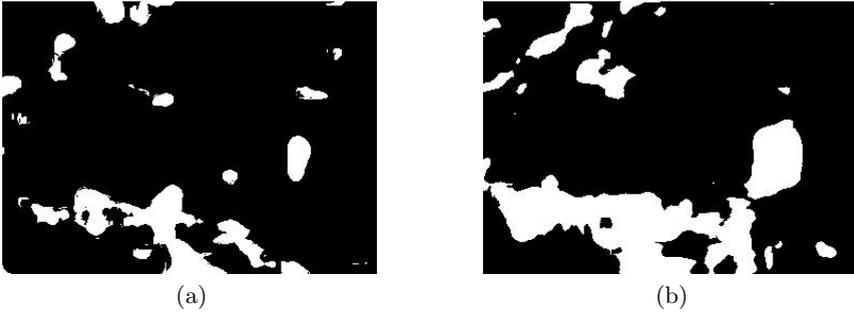
where  $w_{l,\theta}$  are the DT-CWT coefficients in the  $l^{th}$  level and  $\theta^{th}$  orientation band.

A truncated median filter (Nixon and Aguado 2008) is applied to  $T(x, y)$  of each subband to filter out the texture associated with transitions between regions. Fig.2 shows the filtered images using the DT-CWT and Gabor decomposition on an infrared (IR) image. By examining the figure, we can see that the DT-CWT produces more desirable results by highlighting the salient objects as well as maintaining the clear and smooth boundaries, which leads to a better quality image texture map.

A two-level K-means algorithm is used to define textured and non-textured regions. A pixel is classified as textured if the proportion of the number of the subbands belonging to the textured region is above a threshold  $P$ . Our experiments show that an appropriate value for thresholding the multisensor images is assigned to  $P = 0.7$ . The final texture maps are illustrated in Fig.3, in which the DT-CWT demonstrates a strong capability to detect the small-sized salient objects.

## 3 Multiscale Image Segmentation

The textured and non-textured regions are further segmented into relatively small and homogeneous regions while retaining the boundaries between the two



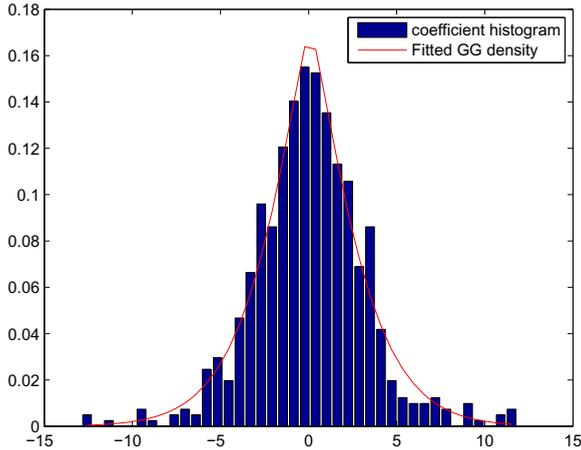
**Fig. 3.** The texture maps of “UNCamp” IR image. (a) The texture map using DT-CWT. (b) The texture map using Gabor decomposition.

regions. The dominant grayscale values are first extracted based on the peer group filtering and generalized Lloyd algorithm (Deng and Manjunath 2001). Then, the *JSEG* algorithm proposed in (Deng and Manjunath 2001) is used to minimize the cost associated with partitioning an image at different scales. A bigger window size is used for high scales, which are useful for detecting texture boundaries, while lower scales are employed in order to localize the intensity of grayscale edges. It is reasonable to apply the lower scales to the non-textured region, which has a more or less homogeneous texture, while higher scales are adopted for the textured region to find the texture boundaries. In contrast with the *JSEG*, which does not account for the local texture difference between the image regions, the strength of this approach is that we are able to apply the multiscale segmentation simultaneously to the same image according to the local texture characteristics.

Nevertheless, the resulting boundary locations between textured and non-textured regions are not the actual boundaries due to the fact that K-means clustering can only segment the image into rough regions. Moreover, multiscale segmentation provides accurate results only within the textured and non-textured regions. Consequently, a boundary refinement step is employed to adjust the boundaries between the two regions. A pixel is assigned to the neighbor class that has the minimum  $D$  value using the following function:

$$D = \text{Dist}(G^0, G^j) + a(S_4^j - D_4^j) + b(S_8^j - D_8^j) \quad (2)$$

where  $\text{Dist}$  refers to the Euclidean distance measure,  $G^0$  and  $G^j$  are the grayscale intensity of the current pixel and its  $j^{\text{th}}$  neighboring segment,  $S_4^j$  and  $S_8^j$  are the numbers of 4 and 8-neighbor pixels belonging to the  $j^{\text{th}}$  segment, while  $D_4^j$  and  $D_8^j$  are the numbers of 4 and 8-neighbor pixels belonging to the different classes of the  $j^{\text{th}}$  segment.  $a$  and  $b$  represent the strength of the spatial constraint. Specifically, as  $a$  and  $b$  increase, a pixel is more likely to belong to the class to which many of its neighbors belong. Thus region boundary smoothness is achieved. In all the experiments,  $a$  and  $b$  are assigned values of 0.8 and 0.6.

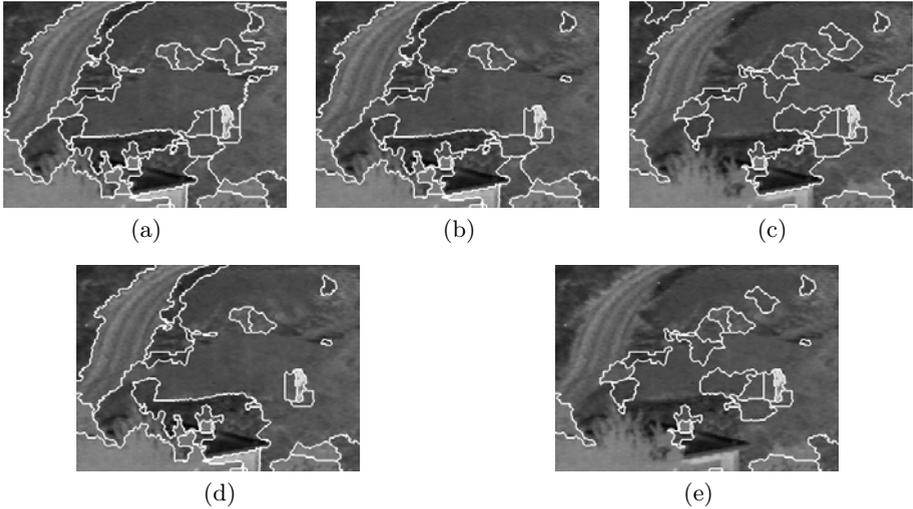


**Fig. 4.** An example of a region of the  $1^{th}$  level and  $2^{th}$  orientation DT-CWT subband coefficients histogram fitted with a generalized Gaussian density on a log scale. The estimated parameters are:  $\alpha = 3.1370$ ,  $\beta = 1.2782$ .

## 4 Statistical Region Merging

After applying a multiscale segmentation on textured and non-textured regions, there are usually some very small regions or more than one neighboring regions exhibiting similar attributes. A merging step is therefore used to eliminate such instances by generating bigger regions. Do and Vetterli recently introduced a statistical framework for texture retrieval in content-based image retrieval applications (Do and Vetterli 2002), where the subband marginal density of the coefficients was approximated by the generalized Gaussian distribution (GGD), and KLD was computed as a similarity measurement. In this study, a statistical method is implemented by varying two model parameters  $\alpha$  and  $\beta$  of the GGD to appropriately model the wavelet coefficients within the segmented regions. The segments with more than 80% of their pixels belonging to the non-textured area are categorized as non-textured segments, and the remaining segments are classified as textured segments. Therefore, segmented regions are considered individually rather than globally.

A corresponding merging criterion is provided for each category. The difference lies in the way the features are extracted within the regions. Non-textured segments are merged based on their grayscale intensity similarity. To achieve this, the Euclidean distance of the gray-level histograms extracted from the neighboring non-textured segments is calculated. For textured segments, region similarity is measured using the statistical model parameters followed by computing the Kullback-Leibler distance. We model the wavelet coefficients of each textured



**Fig. 5.** The region merging results for “UNCamp” IR image. (a) The input segmented image. (b) The segmented image using KLD with 31 segments. (c) The segmented image using Euclidean distance with 31 segments. (d) The segmented image using KLD with 26 segments. (e) The segmented image using Euclidean distance with 26 segments.

region independently by a generalized Gaussian distribution. The KLD between two adjacent textured segments is defined as:

$$\begin{aligned}
 KLD(s_1, s_2) = & \frac{1}{18} \sum_{l=1}^3 \sum_{\theta=1}^6 \log \left( \frac{\beta_1^{l,\theta} \alpha_2^{l,\theta} \Gamma(1/\beta_2^{l,\theta})}{\beta_2^{l,\theta} \alpha_1^{l,\theta} \Gamma(1/\beta_1^{l,\theta})} \right) \\
 & + \left( \frac{\alpha_1^{l,\theta}}{\alpha_2^{l,\theta}} \right)^{\beta_2^{l,\theta}} \frac{\Gamma((\beta_2^{l,\theta} + 1)/\beta_1^{l,\theta})}{\Gamma(1/\beta_1^{l,\theta})} - \frac{1}{\beta_1^{l,\theta}}
 \end{aligned} \quad (3)$$

where  $\Gamma(\cdot)$  is the Gamma function,  $s_1$  and  $s_2$  are the adjoining texture segments, and  $l$  and  $\theta$  denote the index of the decomposition level and orientation. The characteristics of the regions can be completely defined via two model parameters  $\alpha$  and  $\beta$  of the GGD.  $\alpha$  and  $\beta$  can be estimated using the second ( $m_2$ ) and fourth ( $m_4$ ) order moments of image coefficients (Simoncelli and Anderson 1996).

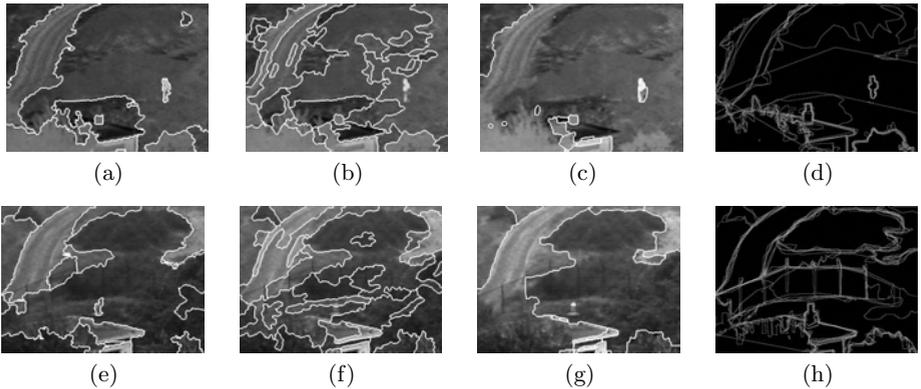
$$m_2 = \frac{\alpha^2 \Gamma(\frac{3}{\beta})}{\Gamma(\frac{1}{\beta})} \quad m_4 = \frac{\Gamma(\frac{1}{\beta}) \Gamma(\frac{5}{\beta})}{\Gamma^2(\frac{3}{\beta})} \quad (4)$$

Fig.4 demonstrates a typical example of a histogram of detail coefficients in a particular subband within a region together with a plot of the fitted estimated GGD model. The fits are generally good. As a result, with only two parameters of the GGD, we can accurately capture the characteristics of marginal densities of wavelet subband coefficients in each region.

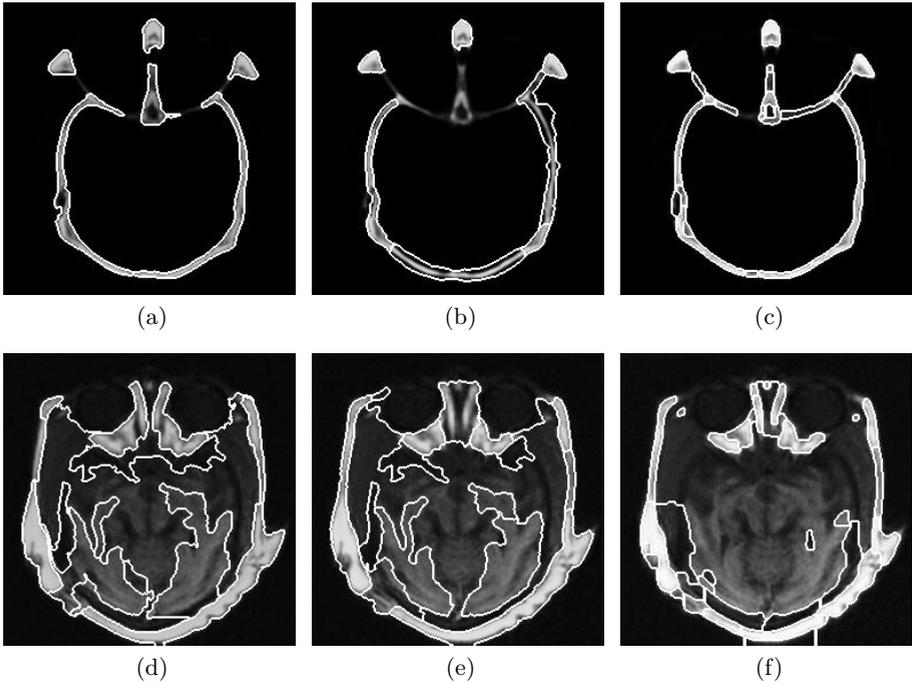
Fig.5 shows the mediate merging results for “UNCamp” IR image after few iterations of the region merging algorithm using the KLD and Euclidean distance. Each pair of images for the comparison contain the same number of regions. The figure displays the actual merging process as similar neighboring regions are integrated into one segmented region. It is clear that the KLD provides better results than the Euclidean distance in terms of human perception.

## 5 Experimental Results and Discussions

The segmentation algorithm has been subjectively evaluated on various images of the same scene captured by different sensors. In Fig.6 and Fig.7, we demonstrate two examples of segmented results in comparison with the *JSEG* (Deng and Manjunath 2001) and the *Watershed* (O’Callaghan and Bull 2005). The *JSEG* examples shown are optimized using the best option values through visual inspection. The *Watershed* applies the region-depth threshold to the gradient surface which is set as 0.15 times the median gradient. Two pairs of images are used in the experiments, including the IR and visible images, a magnetic resonance image (MRI) and a computer tomography (CT) image. For the first pair, the hand-labeled ground truth segmentation are available from the multisensor image segmentation dataset (Lewis et al. 2006). In Fig.6(a), the important regions, such as road, house, and human figure shown in the ground truth, are well segmented by the proposed method, but over-segmented in the *JSEG*, and half merged with the scene background in the *Watershed*. This also happens to the visible image. For the CT image, there are no distinct differences between our method and the *Watershed*. However, in the *JSEG* segmentation shown in Fig.7(b), the central part of the structure has not been segmented and merged into the background. In Fig.7(d), both inner soft tissues and outer structure in



**Fig. 6.** The segmentation results for “UNCamp” IR and visible images. From left to right: the proposed segmentation algorithm, *JSEG*, *Watershed*, hand-labeled ground truth segmentation.



**Fig. 7.** The segmentation results for CT and MRI images. From left to right: the proposed segmentation algorithm, *JSEG*, *Watershed*.

the MRI are better segmented than the *JSEG* and *Watershed*. The experiments demonstrate that the proposed method provides an effective way to accurately segmenting the meaningful objects, even in small size, into separate regions.

## 6 Conclusions and Future Work

We have demonstrated that robust and meaningful image segmentation can be achieved by integrating the statistical modeling with the wavelet feature extraction and multiscale segmentation. The proposed algorithm is evaluated on a variety of multisensor images. The results show the effectiveness of the segmentation for multisensor images. In future, we are interested in applying the segmentation algorithm to a region-based image fusion scheme.

**Acknowledgment.** This work is partially funded by the NCET Program of MOE, China, the SRF for ROCS and the China Scholar Council for visiting fellowship (No. 2010307502).

## References

- Deng, Y., Manjunath, B.: Unsupervised Segmentation of Color-Texture Regions in Image and Video. *IEEE Tran. Pattern Anal. Machine Intell.* 23, 800–810 (2001)
- Do, M., Vetterli, M.: Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance. *IEEE Tran. Image Process.* 11, 146–158 (2002)
- Kingsbury, N.: Complex Wavelets for Shift Invariant Analysis and Filtering of Signals. *Appl. Compt. Harmon. Anal.* 10, 234–253 (2001)
- Lewis, J., Nikolov, S., Canagarajah, N., Bull, D., Toet, A.: Uni-Modal Versus Joint Segmentation for Region-Based Image Fusion. In: *Proc. of the Int. Conf. Information Fusion*, pp. 1–8 (2006)
- Lewis, J., O’Callaghan, R., Nikolov, S., Bull, D., Canagarajah, N.: Pixel- and Region-Based Image Fusion with Complex Wavelets. *Information Fusion* 8, 119–130 (2007)
- Nixon, M., Aguado, A.: *Feature Extraction and Image Processing*. Academic Press, Oxford (2008)
- O’Callaghan, R., Bull, D.: Combined Morphological-Spectral Unsupervised Image Segmentation. *IEEE Tran. Image process.* 14, 49–62 (2005)
- Simoncelli, E., Anderson, E.: Noise Removal via Bayesian Wavelet Coring. In: *Proc. of the IEEE Int. Conf. Image Process.*, pp. 378–382 (1996)
- Wan, T., Canagarajah, N., Achim, A.: Multiscale Color-Texture Image Segmentation with Adaptive Region Merging. In: *Proc. of the IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, pp. 1213–1216 (2007)
- Wan, T., Canagarajah, N., Achim, A.: Statistical Multiscale Image Segmentation via Alpha-Stable Modeling. In: *Proc. of the IEEE Int. Conf. Image Process.*, pp. 357–360 (2007)
- Wan, T., Canagarajah, N., Achim, A.: Segmentation-Driven Image Fusion Based on Alpha-Stable Modeling of Wavelet Coefficients. *IEEE Tran. Multimedia* 11, 624–633 (2009)