

IMAGE FUSION BASED ON LEVEL SET SEGMENTATION

Filip Šroubek^{1,2}, Gabriel Cristóbal², and Jan Flusser¹

¹ Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4
182 08, Prague 8, Czech Republic
email: sroubekf@utia.cas.cz

² Instituto de Óptica
CSIC
Serrano 121
28006 Madrid, Spain
email: gabriel@optica.csic.es

ABSTRACT

Most of image fusion techniques utilize a key notion called “*decision map*”. This map determines which information to take at what place and therefore governs the fusion process. We illustrate that calculation of decision maps is identical to a segmentation problem. Modifying a state-of-the-art segmentation procedure based on level sets, we obtain more accurate and smooth decision maps. Verification of the proposed method is carried out on wavelet-based multifocus fusion and concluded with an experiment on microscopic multifocal images.

1. INTRODUCTION

The term *fusion* means in general an approach to extract information spontaneously from several sources. An illustration is given by the human system which calls upon its different senses, its memory and its reasoning capabilities to perform deductions from the information it perceives. The goal of image fusion is to integrate complementary multisensor, multitemporal and/or multiview information into one new image containing information, the quality of which cannot be achieved otherwise. The term “quality” depends on the application requirements. The individual images entering the fusion process are called *channels*.

Image fusion has been used in many application areas, e.g., in remote sensing and astronomy, in machine vision and mobile robot navigation, in automatic change detection and monitoring of dynamic processes, and last but not least in optical microscopy (multifocus fusion) and medical imaging (multimodal fusion).

Image fusion usually starts with dividing the channels into subregions (except methods where fusion is used to enhance the spatial resolution of the image), calculating a measure of information level in the regions (in the literature often referred to as a *activity level*) (AL) and then utilizing some fusion rules to combine the channels. The channel comparison can be done at different levels of abstraction. The lowest possible is the pixel level, which refers to the merging of measured physical parameters (intensity values of pixels). One step higher is feature-level fusion, which operates on characteristics such as edge, contrast and texture. Multiscale transforms (MST) are often used for feature extraction and in some sense coefficients of MSTs can be considered as features as well. The most commonly used MSTs are the Laplacian pyramid, contrast pyramid, gradient pyramid and wavelet decomposition. Higher levels of abstraction, e.g. decision-based fusion, are possible but we do not consider them here. The measure of information level in the

subregion is the crucial point in the whole process and several different methods were suggested in the literature. In most of the cases, the AL is proportional to the energy of high frequencies in the channel. It corresponds with an intuitive expectation that high frequencies contain details that are important for our visual perception and understanding of the fused image. Image variance, norm of image gradient, norm of image Laplacian [7], energy of a Fourier spectrum [8], image moments [9], and energy of high-pass bands of a wavelet transform [4, 10, 3] belong to the most popular measures of AL. At each pixel or feature, ALs of all channels are compared and the information (pixel values or MST coefficients) of the channel with the highest activity is preserved (*maximum selection rule*). By this process we create the decision map (DM). Alternatively, the first couple of channels with the highest activity can be preserved and their information is averaged. A consistency verification stage follows to prevent occurrence of outlying decisions. In other words, we want to avoid decision maps that alter too quickly. One can regard this step as smoothing of the DM. Once the DM is ready, we create the multiscale representation of the fused image and perform the inverse MST. A detailed overview of multiscale image fusion is given in [5].

Let us illustrate the role of DM in different fusion applications. Multifocus fusion works with images acquired under differed focus settings. In this case, DM identifies regions in focus and if the focus length of input channels is known, DM also defines a depth map. One can see that DM segments the image according to a distance. Then the distance can be used, for example, for surface reconstruction of the measured object (*2.5D reconstruction*). An accurate DM is not only important for valid reconstruction of the fused image, but it is also critical for the surface reconstruction. Erroneous decisions can produce unrealistic peaks and valleys on the surface. Multimodal fusion deals with images that capture different physical properties of the original scene. In this case, DM identifies regions of interest and it can be used for segmentation and further classification of objects on the scene.

The calculation of the DM very much resembles a segmentation task. Regions of the equal decision define areas that belong to the same segment, i.e., we perform segmentation in the space of ALs. The segmentation approach offers several benefits that emanate from the fact that segmentation techniques are very general, expandable and robust. For example, the segmentation approach can be applied to any AL currently used in image fusion. One can combine different activity measures, which is similar to segmentation of vector-valued images. It is more robust to noise and no consistency

verification stage is necessary since it is inherently included in segmentation. A wide variety of segmentation techniques was proposed in the literature. The most attractive and state-of-the-art are active contour models formulated as level sets [2, 1]. In the sequel, we thus propose to calculate the DM with a segmentation technique based on a multiphase level set model suggested by Samson *et al.* in [6]. In the experimental section, we consider the problem of multifocus fusion, use the framework in [10] and evaluate the performance.

2. CALCULATION OF THE DECISION MAP

The classification procedure in general consists of two steps: defining the classes according to discrimination features and defining a partitioning process. In standard classification problems, we assume that classes have a certain probability distribution of intensity, which is known a priori. We can for example choose in the first step parameters of the Gaussian distribution: mean and variance. However, this assumption is not valid in the case of DMs. ALs in each class do not follow a simple unimodal distribution. A different strategy must be employed here and will be discussed below. For this reason we do not refer to this problem as a classification problem but rather as a segmentation one.

For the partitioning process we choose the active contours model formulated as level sets. A detail discussion can be found in [1, 6]. Let Ω be a support of our DM and $\{\Omega_i\}_{i=1,\dots,K} \subset \Omega$ be a family of sets, where K is the number of input channels. Each channel i has an associated set (region) Ω_i . The goal is to find Ω_i 's (more precisely boundaries of the sets) so that each set Ω_i will define a region where the information level of its associated channel is the highest and simultaneously the regions will be sufficiently "simple". Depending on the type of the fusion task, sets $\{\Omega_i\}_{i=1,\dots,K}$ need not be exactly disjoint and/or cover the whole region Ω . To overcome limitations (e.g. change of topology) related to contour (region boundaries) evolution, we use the level-set formulation, which is based on the the following observation: A curve can be seen as the zero-level of a function in higher dimension. Therefore, let us suppose that for each $i = 1, \dots, K$ there exists a function ϕ_i such that inside the region Ω_i it is positive, at the region boundary it is zero and outside it is negative, i.e., the region Ω_i is entirely described by the function ϕ_i . Then the segmentation process becomes a minimization problem of an energy functional involving $\{\phi_i\}$. This functional contains three terms:

- Partition condition:

$$F^A(\{\phi_i\}) = \frac{\alpha}{2} \int_{\Omega} \left(1 - \sum_i H(\phi_i(x))\right)^2 dx, \quad (1)$$

where H is the Heaviside function ($H(x) = 0$ for $x < 0$, $H(x) = \frac{1}{2}$ for $x = 0$ and $H(x) = 1$ for $x > 0$) and α is a positive constant. The minimization of F^A leads to a solution where the formation of areas with no or more than one set Ω_i associated is penalized.

- Length shortening:

$$F^B(\{\phi_i\}) = \beta \sum_i \int_{\Omega} \delta(\phi_i(x)) |\nabla \phi_i(x)| dx, \quad (2)$$

where δ is the Dirac function and β is a positive constant. This term favors regions with simple (short) boundaries.

- Data term: This is the only term that considers information extracted from the input channels and incorporates any prior knowledge, e.g., in a standard classification model one can assume that in each class the image intensities follow a Gaussian distribution. In the case of segmentation in the AL space, we take as our prior information the decision map (segmentation) obtained by applying the maximum selection rule. Let $A_i(x)$ denote the activity level of channel i at location x , which could be, for example, an energy of wavelet coefficients. The maximum activity among all the channels is $M(x) = \max(A_1(x), \dots, A_K(x))$. We then propose the following data term:

$$F^C(\{\phi_i\}) = \sum_i \gamma_i \int_{\Omega} H(\phi_i(x)) (M(x) - A_i(x))^2 dx. \quad (3)$$

This data term minimizes the mean square error of our decisions from M . Alone the data term would return results equal to the standard fusion procedure, which neglects any spatial relations in DM. Naturally, a more "global" behavior appears after adding the partition condition and length shortening terms.

The complete functional takes the form: $F(\{\phi_i\}) = F^A(\{\phi_i\}) + F^B(\{\phi_i\}) + F^C(\{\phi_i\})$ and we minimize it with respect to $\{\phi_i\}$. It is however necessary to regularize F . This is done by replacing H and δ with their approximations H_{ε} and δ_{ε} , respectively; see [2] for details. We may write formally the associated Euler-Lagrange equations and after embedding them into a dynamical process we obtain a set of PDEs

$$\frac{\partial \phi_i}{\partial t} = \delta_{\varepsilon}(\phi_i) \left[\alpha \left(1 - \sum_i H_{\varepsilon}(\phi_i)\right) + \beta \operatorname{div} \left(\frac{\nabla \phi_i}{|\nabla \phi_i|} \right) - \gamma_i (M(x) - A_i(x))^2 \right]. \quad (4)$$

We discretize the equations by finite difference schemes and solve it iteratively.

Determining parameters α , β and γ_i is a tricky task. At this point we can give only general guidelines. In the case of multifocus fusion, we want to avoid overlapping regions and therefore α should be considerably high in comparison to other parameters. In the case of multimodal fusion, regions with different decisions are plausible and therefore α can decrease. If the complexity of DM is expected to be high, β should decrease. Absolute values of parameters $\{\gamma_i\}$ are defined with respect to α and β . More important are their relative values, which can be useful for taking into account different levels of noise in the channels, e.g., $\gamma_i \propto 1/\sigma_i^2$, where σ_i is the noise variance of channel i .

A possible improvement to avoid over-smoothing of the boundaries between segments is to introduce into F^B a stopping function as suggested in [6]. The stopping function g is inversely proportional to the edge strength, and in our case of multiple channels $\{I_i\}$, it is of the form $g(x) = 1/(1 + \sum_i |\nabla I_i(x)|^2)$. The edge information is taken from the original images while segmentation is performed in the AL space, which is an important difference from the standard segmentation problem. For example in multifocus imaging, we can see an intuitive justification of this procedure: strong edges correspond to object boundaries, the depth of a scene

often changes on object boundaries and therefore decisions in the DM should change as well.

3. EXPERIMENTAL RESULTS

We tested the performance of the proposed method on multifocus images. For the fusion scheme, we chose standard wavelet-based approach proposed, e.g., in [10]. Our implementation utilized stationary wavelet transform (symmlets with 8 vanishing moments and no downsampling) with only one level of decomposition. We defined the AL at a location x as a mean of absolute values of three high-pass band coefficients at the location x , which results in one DM of the size of the original image. We provide comparison with DMs that we obtained by applying the maximum selection rule to the AL spatially averaged. Spatial averaging of the AL is necessary to avoid irregular (noisy) maps and it is a common step in image fusion. The proposed segmentation approach requires the starting point in the iterative process. For this purpose, we used the DM calculated by the maximum selection rule and dilated with a circle of radius 4 (structuring element). In all the following cases, images were gray scaled with intensity values between 0 and 255. We set $\gamma_i = 1$ and $\alpha = 255^2$ to penalize void or overlapping regions and chose $\beta = 0.01 \times 255^2$. A stopping criteria of the iterative process was an insufficient change in the DM.

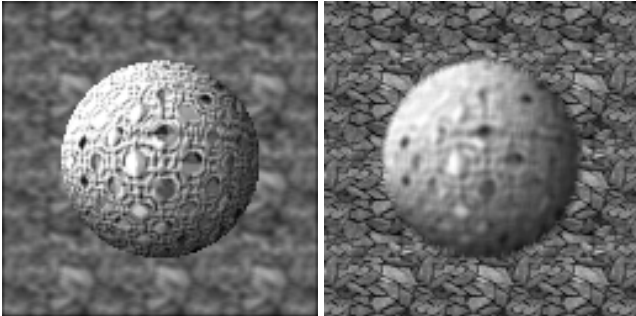


Figure 1: Two synthetically generated images of size 128×128 with different focus settings.

In the first experiment, we generated two synthetic images (see Fig. 1) that depict a spherical object in front of a wall with a focus point set to the sphere and to the wall, respectively. This is a simple two-plane scene. The ideal “ground truth” DM is known in this case and it is a circle that matches the sphere in the image. Results of different DMs are illustrated in Fig. 2. Dark gray (red) denotes areas where the information from image 1 is used and light gray (green) where the information from image 2 is used. To evaluate the results, we define an error measure called *percentage misclassification* (PMC), which is equal to the percentage of pixels in the DM that exhibit an incorrect decision. Fig. 2a shows the DM calculated with the maximum selection rule without spatial averaging of ALs. One can see many outliers compare to the ideal DM in Fig. 2d and PMC is 3.8%. If spatial averaging (5×5 uniform window) of the ALs is performed and then the maximum rule is applied, we obtain a smoother DM in Fig. 2b. The segmentation approach in Fig. 2c outperforms the maximum-rule technique and creates a DM with PMC=1% that closely resembles the ideal DM.

The second experiment outlines the potential pitfall of

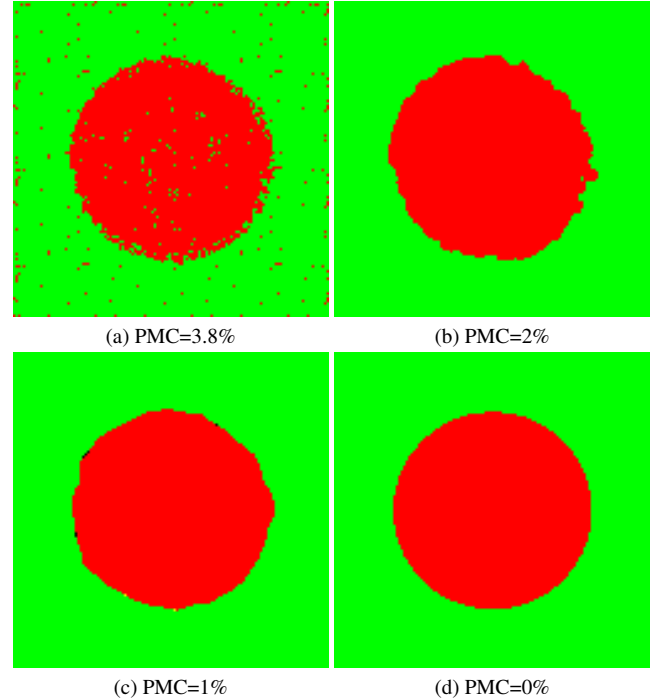


Figure 2: Decision maps: (a) maximum selection rule without averaging; (b) maximum selection rule with a 5×5 averaging window; (c) proposed segmentation approach; (d) ideal decision map (circle).

the segmentation approach with respect to the initial DM. As in the previous experiment, we considered a simple two-plane scene but this time we used real images acquired with a digital camera with different focus settings; see Fig. 3. In this case, the ideal DM corresponds to an image of the Indian segmented from the background. Comparison of different DM estimations, including PMC, is in Fig. 4. First, we used the maximum selection rule and generated to DMs in (a), (c) and (e) with no averaging, 3×3 and 15×15 averaging window, respectively. Clearly, averaging helps to smooth the DM and up to a certain point provides a better estimate. Then we used these results as the initial DMs for our segmentation approach and obtained results in (b), (d) and (f). It is apparent that wrong initial DMs can lead to unsatisfactory results and that one must be careful with choosing the initial DM. Irregular DMs set up too many tiny regions that tend to merge during iterations into larger regions of unpredictable shapes. Over-smoothed initial DMs may prevent small regions to emerge during iterations. Therefore, we conclude that small averaging is necessary for obtaining good results.

We are aware of the fact, that the incorrect decisions calculated by the maximum selection rule occur mainly in flat areas of the images, which is understandable, since wavelet coefficients of high-pass bands are small in these areas and noise alters the decisions. In the segmentation approach, the flat areas contribute to the data term (3) negligibly, the length shortening (2) takes over and we obtain correct decisions.

The last experiment demonstrates the performance of the proposed technique in a real multi-plane scenario. We used 16 multifocus images of a unicellular water organism (“radiolarium”) acquired with an optical microscope. Fig. 5a shows two images from the multifocus stack. The DM es-

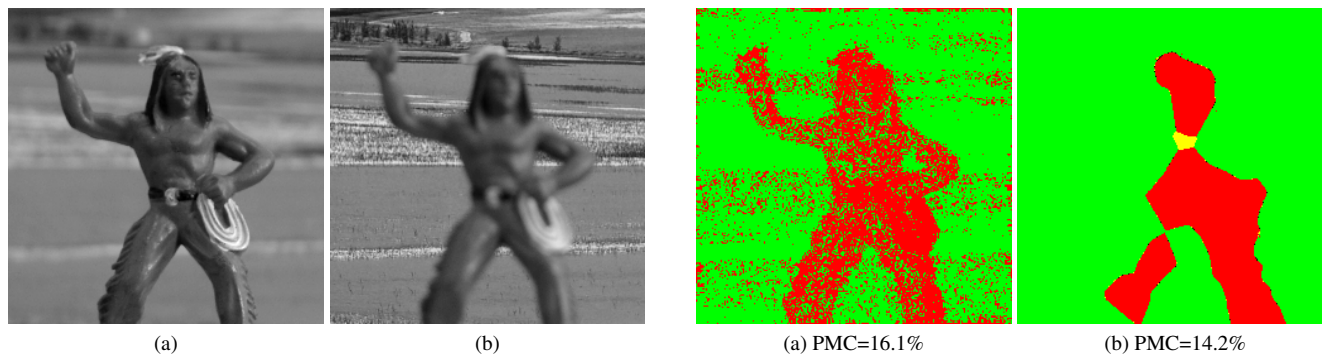


Figure 3: Real two-plane scenario: Two 256×256 real images acquired with different focus points; (a) Indian in focus, (b) background in focus.

timated with the maximum selection rule with 3×3 averaging is in Fig. 6a. This DM served as the initial DM in the segmentation approach and the final result after 40 iterations appears in Fig. 6b. The DM is smooth and individual regions are realistically “simple” in comparison to the DM obtained with the maximum rule. The fused image using the calculated DM is in Fig. 5b. In addition, Fig. 5c shows the 2.5D reconstruction of radiolarium with a surface derived from the reconstructed DM and the fused image as a texture. Since the reconstructed surface is a piece-wise constant approximation, higher-order surface interpolation was necessary to obtain the above results.

4. CONCLUSIONS

We demonstrated that the segmentation techniques, such as level set methods, can be easily modified and applied to the calculation of decision maps. Obtained results are encouraging and show smoother and more accurate decision maps. We may conclude that the segmentation approach outperforms simple selection rules currently used in image fusion. The current drawbacks of the proposed approach are strong dependency on initial conditions and the computational cost which is clearly higher than in the case of simple selection rules. We showed that applications, such as depth reconstruction methods in multifocus imaging, greatly benefit from more accurate decision maps. In this case, any outliers can create unrealistic peaks and valleys on the reconstructed surface and therefore accurate decision maps are crucial for a successful reconstruction. An interesting feature of the level set method, that we did not exploit here but will be a subject of our future investigation, is to allow overlapping regions. We believe that this will help, in particular, with multimodal fusion, where information from different modality blends together at certain areas. Thus the proposed segmentation approach may provide a flexible methodology that solves diverse image fusion problems.

Acknowledgments

Financial support of this research was provided by the Grant Agency of the Czech Republic under the projects No. 102/04/0155, No. 202/05/0242 and by the following Spanish grants: TEC2004-00834; TEC2005-24739-E; TEC2005-24046-E; 20045OE184 from the Spanish Ministry of Education and Science and PI040765 from the Spanish



Figure 4: Two-plane decision maps: (a), (c) and (e) show results of maximum selection rule without averaging, with 3×3 and 15×15 averaging window, respectively; (b), (d) and (f) show results of the proposed segmentation approach with the initial DM from (a), (c) and (e), respectively.

Ministry of Health. The work was also supported in the frame of the bilateral project 2004CZ0009 of CSIC and AS CR. F. Šroubek was also supported by the Spanish States Secretary of Education and Universities fellowship.

REFERENCES

- [1] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing*. Springer Verlag, New York, 2002.
- [2] T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, February 2001.
- [3] J. Kautsky, J. Flusser, B. Zitová, and S. Šimberová. A new wavelet-based measure of image focus. *Pattern Recognition Letters*, 23:1785–1794, 2002.
- [4] H. Li, B.S. Manjunath, and S.K. Mitra. Multisensor

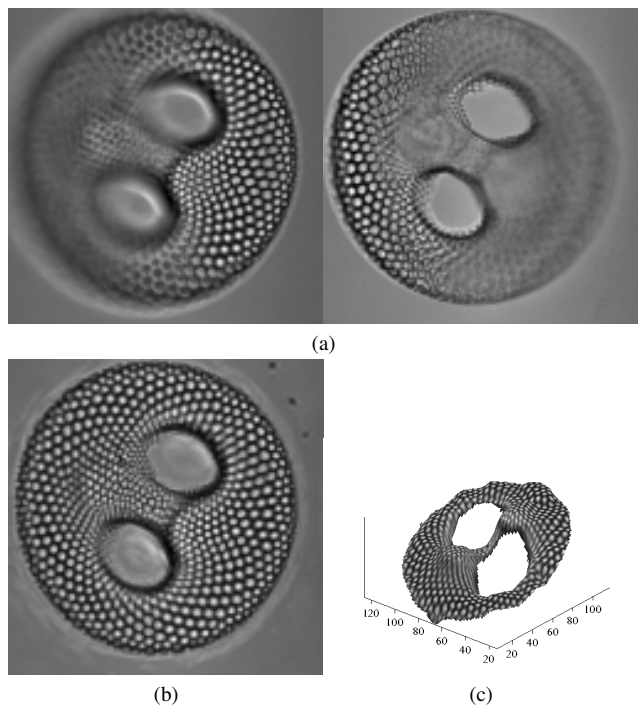
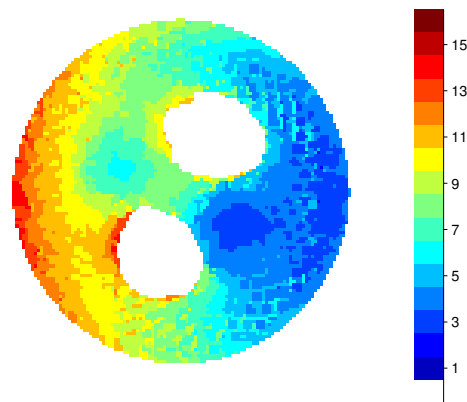


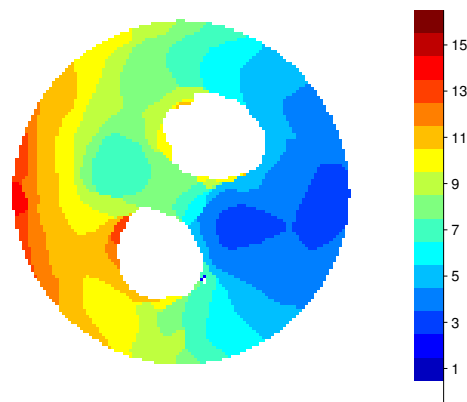
Figure 5: Real multi-plane fusion: (a) shows two of 16 multifocus images; (b) fused images using segmentation approach; (c) 3D reconstruction of the object surface.

image fusion using the wavelet transform. *Graphical Model and Image Processing*, 57(3):235–245, May 1995.

- [5] G. Piella. A general framework for multiresolution image fusion: from pixels to regions. *Information Fusion*, 4:259–280, 2003.
- [6] C. Samson, L.B. Féraud, G. Aubert, and J. Zerubia. A level set model for image classification. *Int. Journal of Computer Vision*, 40(3):187–197, 2000.
- [7] M. Subbarao, T. Choi, and A. Nikzad. Focusing techniques. *Optical Eng.*, 32:2824–2836, 1993.
- [8] M. Subbarao and J. K. Tyan. Selecting the optimal focus measure for autofocusing and depth-from-focus. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20:864–870, 1998.
- [9] Y. Zhang, Y. Zhang, and C. Wen. A new focus measure method using moments. *Image and Vision Computing*, 18:959–965, 2000.
- [10] Z. Zhang and R.S. Blum. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. In *Proceedings of the IEEE*, volume 87, pages 1315–1326, August 1999.



(a)



(b)

Figure 6: Decision maps for multi-plane fusion: (a) maximum selection rule with a 3×3 averaging window; (b) proposed segmentation approach.