Object Detection Using Gabor Filters

Anil K. Jain and Nalini K. Ratha
Department of Computer Science
Michigan State University
East Lansing, MI 48824

Sridhar Lakshmanan
Department of Electrical and Computer Engineering
University of Michigan - Dearborn
Dearborn, MI 48128

Abstract

This paper pertains to the detection of objects located in complex backgrounds. A feature-based segmentation approach to the object detection problem is pursued, where the features are computed over multiple spatial orientations and frequencies. The method proceeds as follows: A given image is passed through a bank of even-symmetric Gabor filters. A selection of these filtered images is made and each (selected) filtered image is subjected to a nonlinear (sigmoidal like) transformation. Then, a measure of texture energy is computed in a window around each transformed image pixel. The texture energy ("Gabor features"), and their spatial locations, are inputted to a squared-error clustering algorithm. This clustering algorithm yields a segmentation of the original image – it assigns to each pixel in the image a cluster label that identifies the amount of mean local energy the pixel possesses across different spatial orientations and frequencies. The method is applied to a number of visual and infrared images, each one of which contains one or more objects. The region corresponding to the object is usually segmented correctly, and a unique signature of "Gabor features" is typically associated with the segment containing the object(s) of interest. Experimental results are provided to illustrate the usefulness of this object detection method in a number of problem domains. These problems arise in IVHS, military reconnaissance, fingerprint analysis, and image database query.

Keywords: texture-based segmentation, object detection, even-symmetric Gabor filters, fingerprint, target recognition

1 Introduction

The detection of objects in a complex background is a well-studied, but yet unresolved problem. A number of different techniques for solving this problem have been proposed [(1)]. Typically, the problem of object detection is posed as an equivalent problem of segmenting a given image based on the discrimination between, and the spatial continuity of, locally computed image feature

vectors. The success or failure of these methods is intricately tied to the types of features used, the reliability with which these features are extracted, and the criteria used for merging (and discriminating between) the extracted features [1].

A useful set of features commonly used for segmenting images pertain to texture. Texture-based segmentation involves identifying regions with uniform texture in a given image. Since image texture possesses spatial continuity at both local and global scales, it is impossible to give a universal definition of texture or assign a universal measure of textural uniformity to a given image region. Over the last two decades, a large number of techniques for analyzing image texture have been proposed [2, 3]. In this paper, we focus our attention on a multi-channel filtering approach that was introduced in [4]. The approach is inspired by a multi-channel filtering theory for processing visual information in the early stages of the human visual system, and it is inherently multi-resolutional. This multi-resolution characteristic allows image texture to be analyzed at both local and global scales simultaneously.

The main issues involved in the multi-channel filtering approach to texture analysis are:

1. The functional characterization of the different channels and the number of channels,
2. The extraction of texture features from the filtered images,
3. The relationship between the channels (dependent vs. independent), and
4. Integration of texture features from the channels to produce the final segmentation.

Various multi-channel techniques proposed in the literature [5–11] differ in their approach to one or more of the above four issues. See Jain and Farrokhnia [4] for a complete discussion.

The multi-channel filtering technique in this paper uses a bank of even-symmetric Gabor filters to functionally characterize the different channels. Those filtered images that possess a significant component of the original image are selected, and the rest are discarded. The select filtered images are then subjected to a non-linear transformation that behaves like a blob detector. An energy measure is defined on the transformed images in order to compute different texture features for the different blobs. The measure calculates a sum of texture energies over a small window around each transformed pixel in the select filtered images. Using the energy measure as features, a vector of Gabor features is defined for each pixel in the original image. The Gabor features, and the spatial co-ordinates of the corresponding pixels, are subsequently inputted to a squared-error based clustering algorithm. This clustering algorithm

yields a segmentation of the original image - it assigns to each pixel in the original image a cluster label. This label identifies the amount of texture energy that a pixel possesses across the different spatial orientations and frequencies (see Fig. 1 for an overall schematic of the proposed technique).

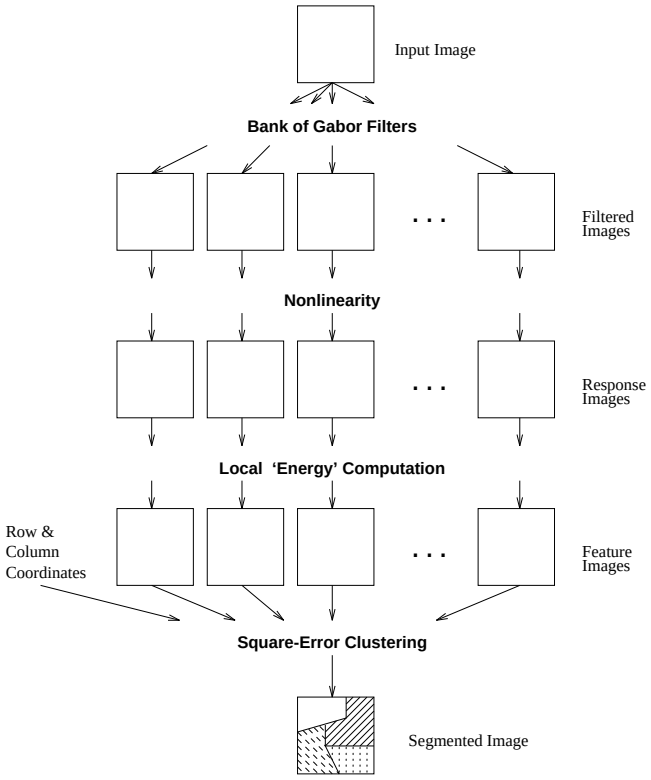


Figure 1: An overview of the texture segmentation algorithm.

We segment a number of visual and infrared images, which contain one or more objects, by using our multi-channel filtering technique. The images have been selected to illustrate the usefulness of the technique in a number of problem domains. These problems arise in automotive applications, military reconnaissance, fingerprint analysis, and image database query - see Figs. 2 - 9. We observe that in all the images considered, the region(s) corresponding to the object(s) of interest is usually segmented correctly. Furthermore, a unique signature of Gabor features is typically associated with the segment containing the object(s) of interest. No special tuning was done to accommodate the variations that are expected in each image - the very same algorithm was applied across all the images. These experimental results suggest that our Gabor filter-based scheme holds much promise in the area of object detection. In particular, the uniqueness of the object's Gabor features enhances the utility of our technique from an applications standpoint.

We note that a number of other researchers have also recently explored the utility of Gabor filters for object detection tasks [12, 13].

The rest of the paper is organized as follows. In section 2 we present a mathematical description of the multi-channel Gabor filtering scheme and the issues relevant to it. In section 3 we present extensive experiments on applying this multi-channel filtering technique for object detection in visual and infrared images. The paper closes with some concluding remarks in section 4.

2 Multi-Channel Filtering and Segmentation

Given an image, we first pass it through a bank of Gabor filters. The impulse response of the Gabor filters is given by,

$$h(x,y) = exp\{-\frac{1}{2}[\frac{x^2}{{\sigma_x}^2} + \frac{y^2}{{\sigma_y}^2}]\}cos(2\pi u_o x + \phi), \tag{1}$$

where u_0 denotes the radial frequency of the filter, and σ_x and σ_y represent the space constants of the filter. The filter $h(x,y)$ shown in Eq. (1) is real-valued, even-symmetric, and oriented along the horizontal x-axis. Filters with arbitrary orientation can be obtained from such a $h(x,y)$ by rigid rotations of its $x-y$ coordinates.[1] The Fourier representation of $h(x,y)$ is given by,

$$H(u,v) = A(exp\{-\frac{1}{2}[\frac{{(u-u_0)}^2}{{\sigma_u}^2} + \frac{v^2}{{\sigma_v}^2}]\} + exp\{-\frac{1}{2}[\frac{{(u+u_0)}^2}{{\sigma_u}^2} + \frac{v^2}{{\sigma_v}^2}]\}), \tag{2}$$

where $\sigma_u = 1/2\pi\sigma_x$, $\sigma_v = 1/2\pi\sigma_y$, and $A = 2\pi\sigma_x\sigma_y$. By passing the original image through such a filter, we obtain all those components in the image that have their energies concentrated near the spatial frequency point[2] $(\pm u_0, 0)$ within a frequency bandwidth of B_r octaves and orientation bandwidth of B_θ degrees. These bandwidth quantities are given by

$$B_r = log_2(\frac{u_0 + {(2ln2)}^{1/2}\sigma_u}{u_0 - {(2ln2)}^{1/2}\sigma_u}), \tag{3}$$

$$B_\theta = 2(tan^{-1}\frac{{(2ln2)}^{1/2}\sigma_v}{u_0}). \tag{4}$$

[1] We denote the angle of such rotations by θ_0.

[2] Or a rotation of $(\pm u_0, 0)$ by an angle θ_0

Selecting a relatively large bank of such Gabor filters results in a nearly uniform coverage of the spatial-frequency domain. The resulting decomposition of the original image is nearly orthogonal as the amount of overlap between the filters in the spatial-frequency domain is small [4].

Once such a multi-channel decomposition of the original image is obtained, we then select a subset of the filtered images and discard the rest. The motivation behind this pruning comes from the fact that some of the filtered images contain little or no information about the original image. Let $s(x, y)$ denote the reconstruction of the original image by combining all of the filtered images, and $\hat{s}(x, y)$ denote the reconstruction achieved by combining only a subset of the filtered images. The fraction of the intensity variations in $s(x, y)$ that is explained in $\hat{s}(x, y)$ can be measured by the quantity:

$$R^2 \;=\; 1 - \frac{\epsilon}{\mathcal{E}}, \tag{5}$$

where $\epsilon = \sum_{x,y}[\hat{s}(x, y) - s(x, y)]^2$ and $\mathcal{E} = \sum_{x,y}[s(x, y)]^2$. So, we use this quantity R^2 in our filter selection process as follows:

1. Select the filtered image that results in the highest value of R^2;
2. Select the next filtered image that together with the previously selected one(s) yields the highest value of R^2;
3. Repeat step (2) until $R^2 \geq 0.95$;

Although this filter selection process is stated in the spatial domain, it is more easily implemented in the frequency domain without the need for any expensive inner product (inverse-Fourier) computations [4].

Having pruned the multi-channel filtered images, we next subject the select filtered images to a non-linear point-wise transformation. This transformation is given by,

$$\psi(t) \;=\; \tanh(\alpha t) \;=\; \frac{1 - \exp(-2\alpha t)}{1 + \exp(-2\alpha t)}, \tag{6}$$

where α is a constant. This non-linear transformation changes the sinusoidal variations in the filtered image into square variations, and so it can be viewed as a blob detector.

We then define a texture energy measure over a small window around each transformed pixel

in the select filtered images:

$$e_k(x,y) \;=\; \frac{1}{M} \sum_{(a,b)\in W_{xy}} |\psi(r_k(a,b))|, \tag{7}$$

where $r_k(a,b)$ denotes the k^{th} filtered image, and $e_k(x,y)$ is the corresponding texture energy.

Using $e_k(x,y)$, we can associate with every pixel (x,y) in the original image a feature vector $[e_1(a,b),\ e_2(a,b),\ \ldots e_d(a,b)]$, where d is the number of selected filters or features. We input these feature vectors and their (x,y) positions into a squared-error clustering algorithm, called CLUSTER [14]. The CLUSTER algorithm iterates through two phases. Phase 1 (the K-mean pass) creates a sequence of clusterings containing $3, \ldots,\ k_{max}$ clusters, where k_{max} is specified by the user. Phase 2 (the forcing pass) then creates another set of clusterings by merging existing clusters two at a time to see if better clusterings can be obtained. After each pass through phases 1 and 2, the squared-errors of the clusterings are compared with the squared-error of the clusterings that existed before that pass. (Each new clustering is compared with an old clustering with the same number of clusters.) If any of the squared-errors are smaller than before, another pass through phases 1 and 2 is initiated. This continues until the squared-error cannot be decreased.

3 Experimental Results

We segment a number of visual and infrared images, which contain one or more objects, by using our multi-channel filtering technique. The images have been selected to illustrate the usefulness of the technique in a number of problem domains. A brief description of these images is given below.

- The visual image in Fig. 2, and the infrared image in Figs. 3 are associated with survivability analysis in the military reconnaissance context. The objects of interest are tanks and other armored vehicles.

- The visual images in Figs. 4 and 5 are associated with a computer vision solution to traffic management problems. The objects of interest are automobiles.

- The images in Figs. 6 and 7 are associated with content-based query of image databases. The objects of interest include various animate and inanimate entities such as the fish,

and the saxophone.

- The images in Figs. 8 and 9 are associated with fingerprint analysis. The object of interest are individual fingerprints.

The segmentation results are shown in Figs. 2-9. Part (a) in each figure represents the original grey-tone image containing one or more objects of interest. Part (b) in each figure represents the segmentation of the original grey-tone image, where regions having different Gabor features are designated by different grey-values. In all the images the region corresponding to the object(s) is (are) usually segmented correctly, and a unique set of Gabor features is typically associated with the segment containing the object(s) of interest. This fact is illustrated in part (c) - the segmentation is overlaid on top of the original image, and only those portions of the original image that have the same Gabor features as the object of interest are shown. The size of the segment containing the object of interest is typically invariant to the number of chosen clusters, as shown in Fig. 10. We note, however, that when the number of chosen clusters is large and the object(s) of interest occupies a substantial portion of the image, our technique fails and the object of interest is segmented into several smaller regions. This point is illustrated in Fig. 11 where part (b) shows a correct segmentation of the image in part (a), where the object of interest is segmented as a whole into one region (this segmentation is obtained when the number of clusters chosen is 4); Part (c) shows an incorrect segmentation of the image in part (a), where the same object of interest is now segmented into several smaller regions (obtained by choosing the number of clusters to be 8).

We provide below a description of several of the experimental issues:

- We start with 20 Gabor filters in each case, and they correspond to radial frequencies (u_0) of $4\sqrt{2}, 8\sqrt{2}, 16\sqrt{2}, 32\sqrt{2}, 64\sqrt{2}$ cycles per image-width and spatial orientations (θ_0) of $0°, 45°, 90°, 135°$.

- The scale factors (σ_x and σ_y) in each of these filters are chosen in such a manner as to make the orientation and frequency bandwidths (B_r and B_θ) 45° and 1 octave, respectively.

- We use our filter selection scheme from section 2 to determine a subset of filtered images that achieves an R^2 value of at least 0.95. Thenumber of filtered images (features) actually used varies from 8 to 20.

- In all of the experiments, the nonlinear transform applied on the (selected) filtered image corresponds to the parameter value of $\alpha = 0.25$.

- In computing the texture energy, described in section 3, we used Gaussian weighted windows with a space constant of $\sigma = 0.5N_c/u_0$, where N_c denotes the number of columns in the image.

- Rather than clustering clustering all the pixels in the image, we cluster only 4000 to 8000 randomly selected patterns to save computation time. The remaining pixels are assigned to the closest cluster center (in the feature space).

The mean values of the twenty Gabor-filter features for each of the three clusters in the segmentation of Figure 8 are shown in Figure 12. The cluster 2 corresponds to the fingerprint regions. Note that the filter responses to the fingerprint regions are quite different from the other background regions in the image. This explains the good discrimination which we have obtained between the regions of interest and the background.

4 Concluding Remarks

In this paper the problem of detecting objects in complex background was studied on several visual and infrared images. Our segmentation approach to the object detection problem relies on features computed over multiple spatial orientations and frequencies. A given image was passed through a bank of even-symmetric Gabor filters. A selection of these filtered images was made and each (selected) filtered image was subjected to a nonlinear (sigmoidal like) transformation. Then, a measure of texture energy was computed in a window around each transformed image pixel. The texture energies (Gabor features), and spatial locations of pixels, were inputted to a squared-error clustering algorithm. This clustering algorithm yielded a segmentation of the original image. The proposed method was applied to a variety of images, each one of which contained one or more objects. The region corresponding to the object was usually segmented correctly, and this region also typically possessed a unique set of Gabor features specified over different spatial orientations and frequencies. The images were chosen so as to verify the usefulness of our object detection algorithm in a number of problem areas such as IVHS, military reconnaissance, fingerprint analysis, and image database query.

References

[1] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognition,* Vol. 29, pp. 1277–1294, 1993.

[2] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE,* Vol. 67, pp. 786–804, 1979.

[3] L. Van Gool, P. Dewaele and Oosterlinck, "Texture analysis anno 1983," *Comput. Vision Graphics Image Process.,* Vol. 29, pp. 336–357, 1985.

[4] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition,* Vol. 24, pp. 1167–1186, 1991.

[5] A. C. Bovik, M. Clark and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Mach. Intell.,* Vol. 12, pp. 71–82, 1990.

[6] J. M. Coggins and A. K. Jain, *A spatial filtering approach to texture analysis,* Pattern Recognition Letters, Vol. 3, pp. 195-203,1985.

[7] O. D. Faugeras, "Texture analysis and classification using a human visual model," *Proc. IEEE Int. Conf. Pattern Recognition, Kyoto,* pp. 549–552, 1978.

[8] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms," *J. Op. Soc. Am. A,* Vol. 7, pp. 923–932, 1990.

[9] T. R. Reed and H. Wechsler, "Segmentation of textured images and Gestalt organization using spatial/spatial-frequency representations," *IEEE Trans. Pattern Anal. Mach. Intell.,* Vol. 12, pp. 1-12, 1990.

[10] H. Voorhees and T. Poggio, "Computing texture boundaries from images," *Nature,* Vol. 333, pp. 364–367, 1988.

[11] M. Porat and Y. Y. Zeevi, "The generalized Gabor scheme of image representation in biological and machine vision," *IEEE Trans. Pattern Anal. Mach. Intell.,* Vol. 10, pp. 452–468, 1988.

[12] R.N. Braithwaite and B. Bhanu, "Hierarchical Gabor filters for object detection in infrared images", In Proc. of Computer Vision and Pattern Recognition, Seattle, June 1994, pp. 628–631.

[13] D. Dunn, W.E. Higgins, and J. Wakeley, "Texture segmentation using 2-D Gabor elementary functions", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, 1994, 130-149.

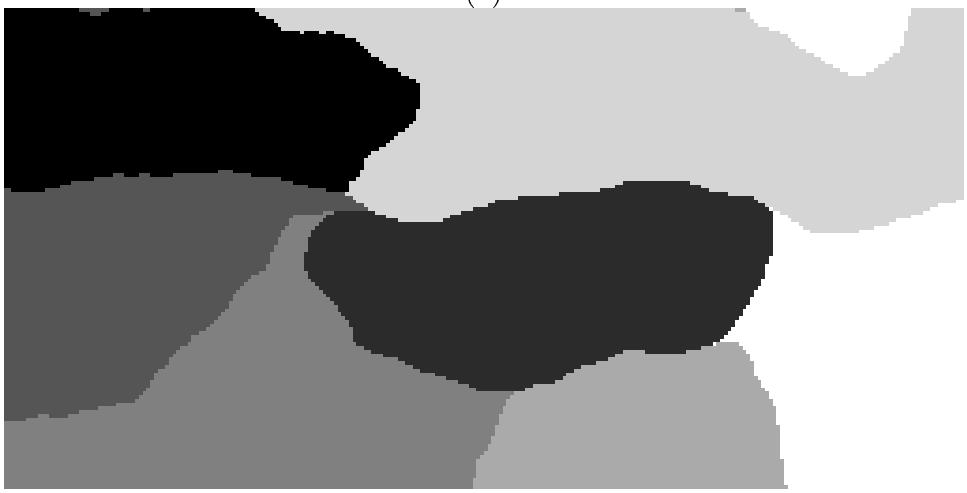
[14] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data,* Prenctice-Hall, 1988.

[15] A. K. Jain and S. Bhattacharjee, "Text Segmentation using Gabor filters for automatic document processing ", Machine Vision Applications, vol. 5, 1992, pp. 169–184.

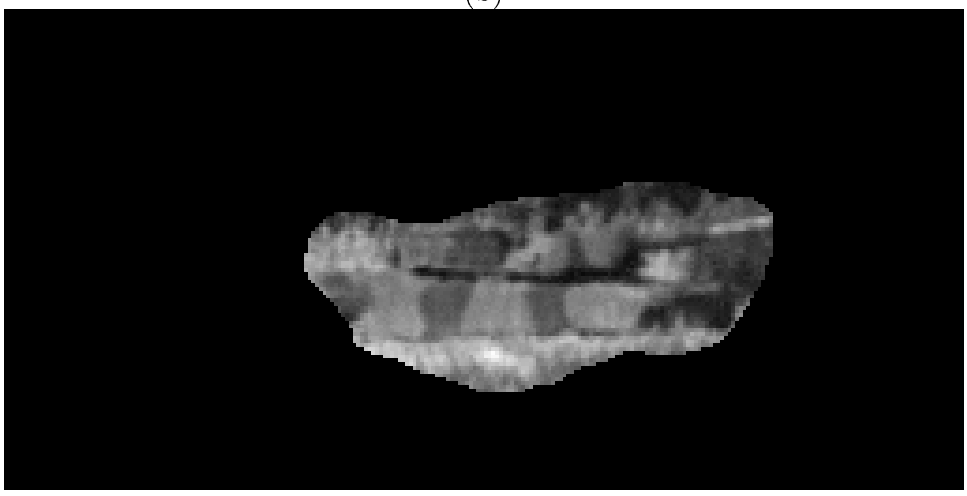
[16] M. Tuceryan and A. K. Jain, Texture Analysis, in C. H. Chen L. F. Pau and P. S. P. Wang (Eds.), Handbook of Pattern Recognition and Computer Vision (World Scientific Publishing Company, 1992).

[17] R. Chellappa and A. K. Jain (Eds.), *Markov Random Fields: Theory and Applications,* Academic Press, 1993.

(a)

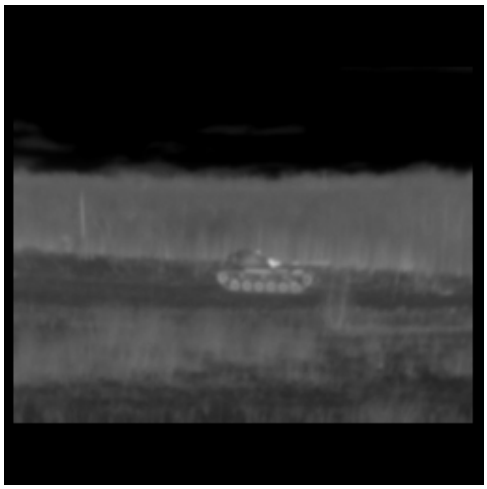


(b)

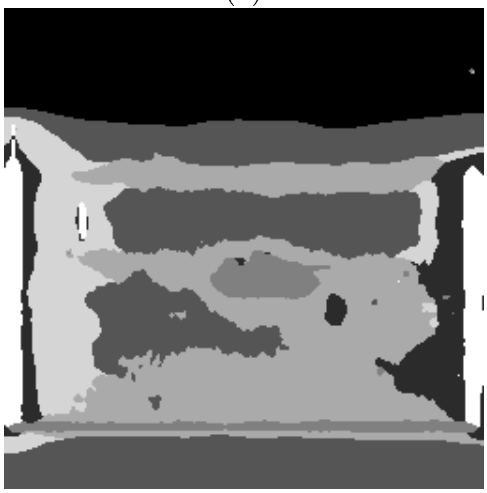


(c)

Figure 2: Detecting a tank in a visual image: (a) original image (128 × 256) with a tank, (b) 7-class segmentation result, and (c) the cluster (image region) associated with the tank.



(a)



(b)

(c)

Figure 3: Detecting a tank in an infrared image: (a) Original image (256×256) with a tank, (b) 7-class segmentation result, (c) the cluster associated with the tank.

(a)

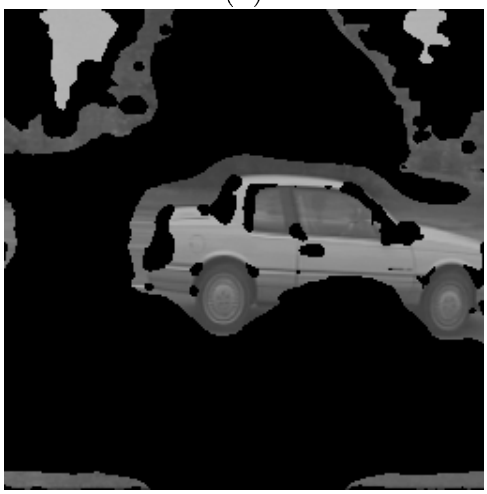
(b)

(c)

Figure 4: Detecting a car in a visual image: (a) Original image (256 × 256) with a car, (b) 5-class segmentation result, (c) the cluster associated with the car.

(a)

(b)

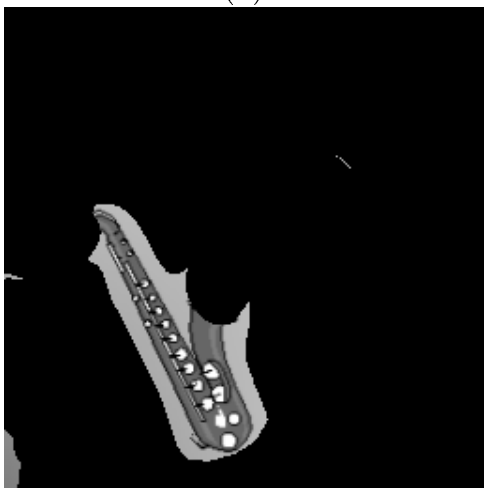


(c)

Figure 5: Detecting a car in a visual image: (a) Original image (256 × 256) with a car, (b) 3-class segmentation result, (c) the cluster associated with the car.

(a)

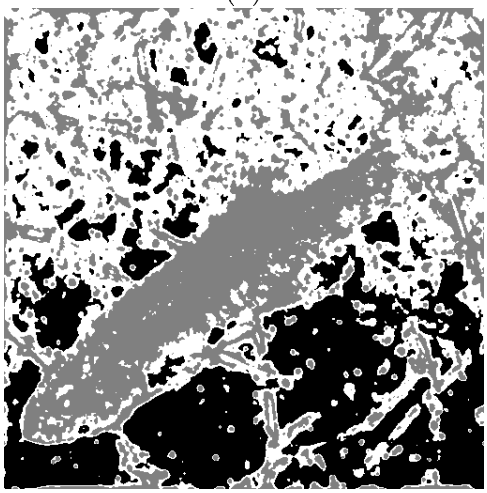
(b)



(c)

Figure 6: Detecting a saxophone in a visual image: (a) Original image (256 × 256) with a saxophone, (b) 7-class segmentation result, (c) the cluster associated with the saxophone.

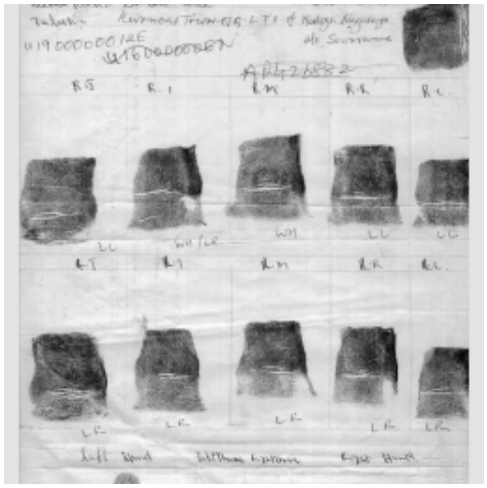
(a)



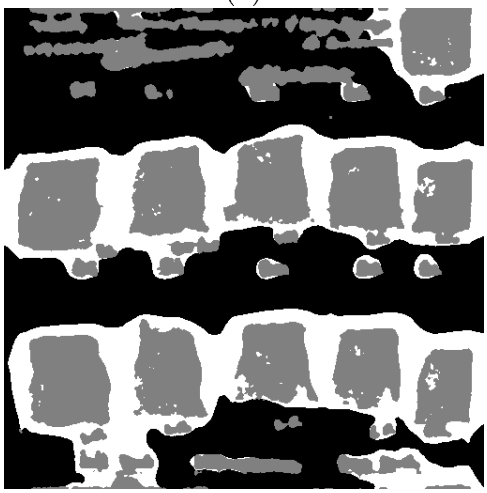
(b)

(c)

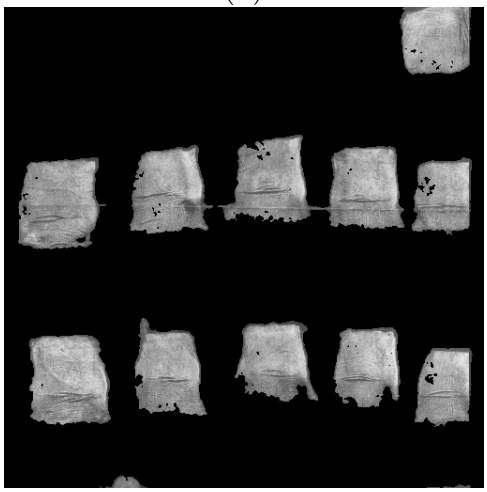
Figure 7: Detecting a fish in a visual image: (a) Original image (512 × 512) with a fish, (b) 3-class segmentation result, (c) the cluster associated with the fish.



(a)



(b)



(c)

Figure 8: Detecting fingerprints: (a) A typical fingerprint card image (512 × 512), (b) 3-class segmentation result, (c) the cluster associated with fingerprints.

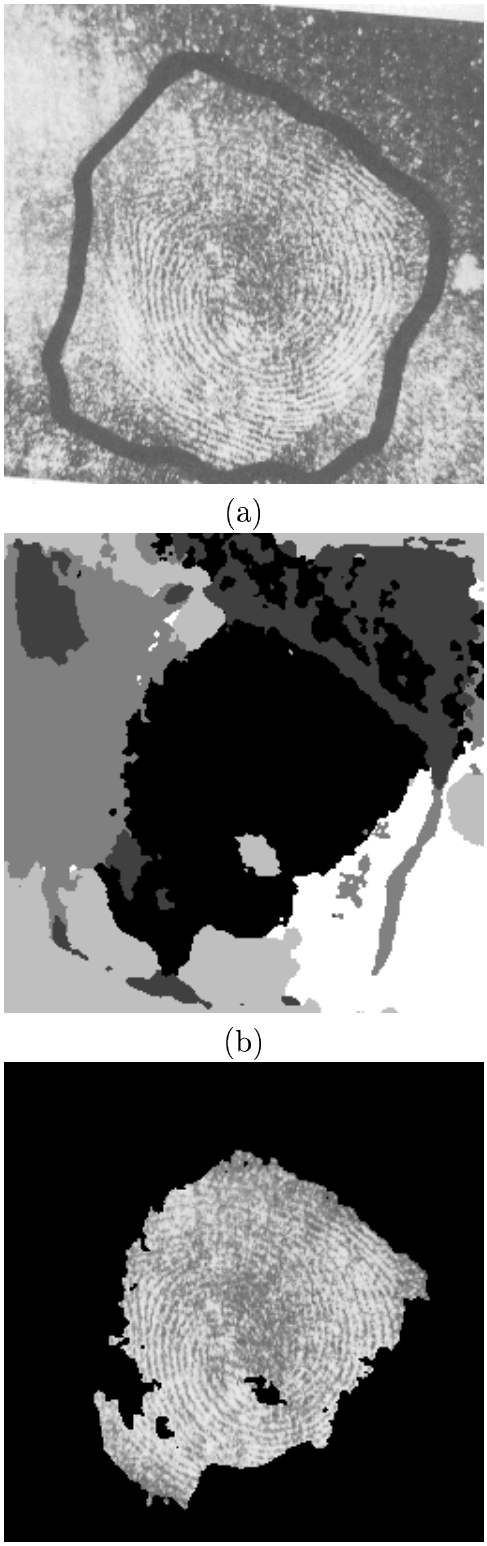


Figure 9: Detecting fingerprints in a latent image: (a) Original latent image (256 × 256) containing a fingerprint, (b) 5-class segmentation result, (c) the cluster associated with the fingerprint.

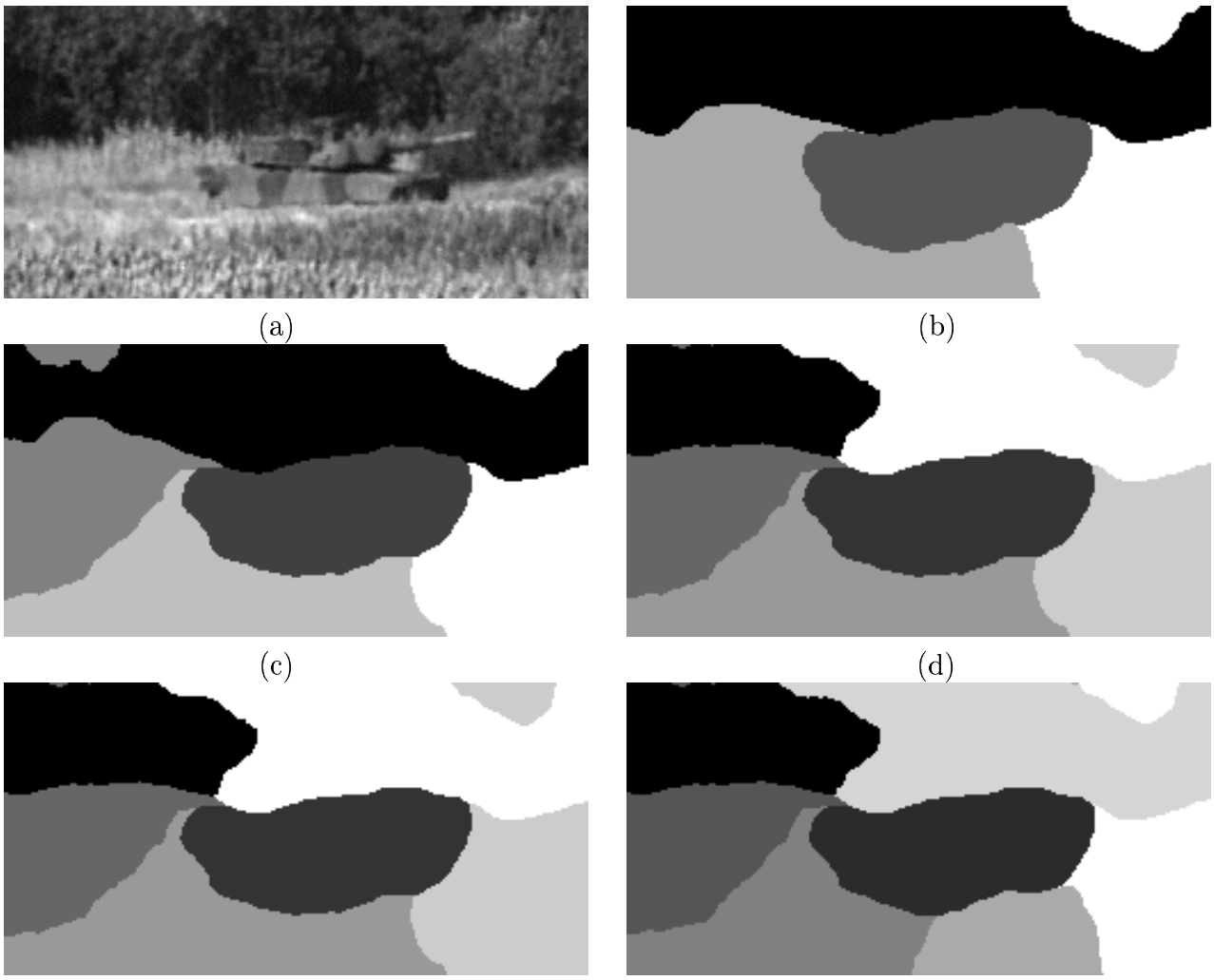
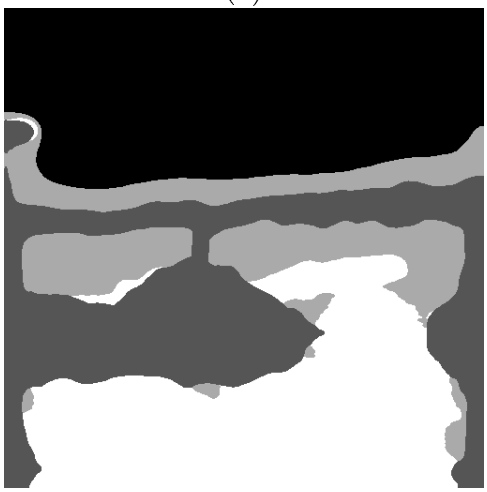


Figure 10: Effect of increasing the number of clusters: (a) original image (128×256) with a tank, (b)-(f) sequence of segmentations with increasing number of clusters (4-8).

(a)



(b)

(c)

Figure 11: Correct and incorrect segmentations: (a) original infrared image (512 × 512) with a tank, (b) a reasonable segmentation (number of clusters equal 4), and (c) incorrect segmentation (number of clusters equal 8).

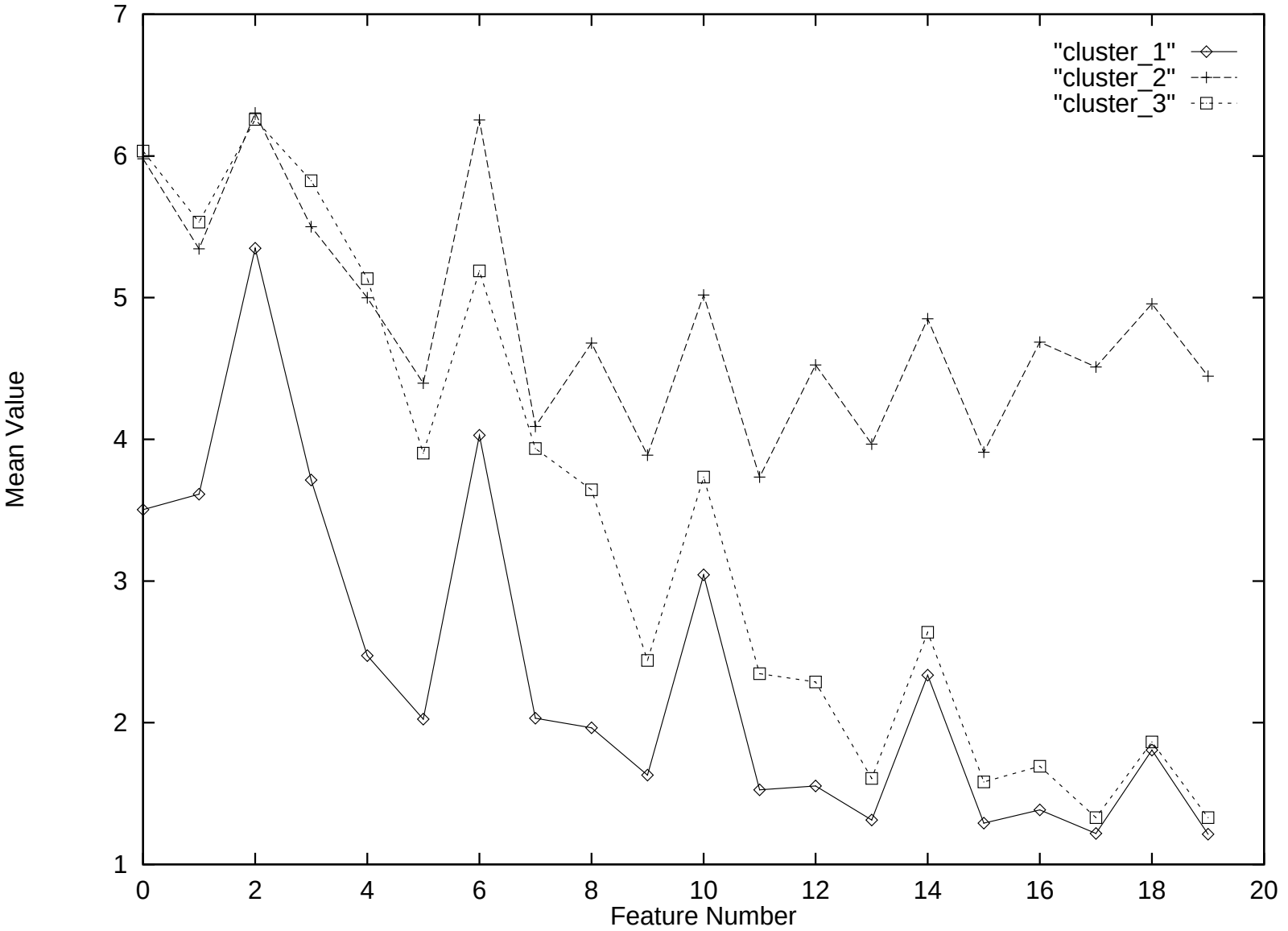


Figure 12: Mean feature values of the twenty Gabor filters for the fingerprint image (Figure 8). Cluster number 2 corresponds to the fingerprint regions in the image.