MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11-13, 2002, Nara-ken New Public Hall, Nara, Japan

13—32

Location Extraction and Discrimination of Similar Texture using Gabor Filter for Donut Recognition

Heng Cher Keng[1], Hidenobu IIDA[2], Michio MIWA[3]
Matsushita Communication Industrial Co., Ltd.

Abstract

This paper describes a machine vision system for recognition of bakery donut produce using Gabor filter that is capable of first locating and then analyzing regions of subtle texture differences. In the case when there is similar texture pattern, color analysis can then be used. Compared to conventional methods and applications, what newly presented are:

- Automatic location of small distinctive areas that characterizes the donut produce
- Identification of texture feature based on the spatial and orientation relationship between coarse-to-fine frequency components of the image.

1 Introduction

Texture plays an important role in computer vision and pattern recognition. It is a difficult quantity to be described and numerous researches have been conducted over the past few decades. One particular texture analysis is known as frequency analysis that exploits frequency information from an image patch. Gabor Filter (and Gabor Wavelet) has been a popular tool to extract such frequency components from both color and grayscale images. This is because the Gabor filter is a reasonable model for the operation the simple cell of the human primary vision cortex [1] intuitively. In theoretical sense, Gabor filters are local spatial bandpass filters that are optimal in the sense of minimizing uncertainty in the joint spatial and spatial-frequency domains.

However, the direct application of Gabor filter for recognition of bakery products, like donuts, would not be an easy task. Compared to other texture identification systems [2], texture analysis for donut produce is particularly challenging and difficult. Firstly, the textures of similar donuts are generally random and highly irregular and can consist of different pattern types. Next, different donut types may have the same pattern generally and differ only in slight subtle small regions. Hence, there is a need to extract texture information from area of interest only and discard texture information that is irregular or highly random nature.

The importance of our work lies in the capability of our proposed system that can differentiate donut images with complicated texture structure as described above. The purpose of the system is to identify different donut type by image processing techniques and price the item in an automated checkout or point-of-sale system. Though texture methods have long been employed in machine vision systems, until only much recently, such applications have been relatively restricted to identification of uniform texture or color or shape. As cited in [2], such examples of machine vision applications in food processing would be chocolate chip cookie and orange processing. The "Veggie Vision" system from IBM [3] is yet another example that uses shape, color and texture analyses for produce identification to speed up checkout time and reduce cashier error in supermarkets.

Since most real world objects often consist texture of different types and natural images often consist of textured and non-textured regions, we believe that the ability to identify object of complex composite texture would gain more importance and have great practical value. Hence the future of our work would not be restricted to donut produce alone, but also texture analysis in outdoors scenes or cluttered environments.

2 Brief Description of Methodology

The basic idea is to use Gabor filters as texture descriptors as well as feature detectors. The frequency components are identified and separated based on the extent of correlation of the image signal in the Gabor filter space from a coarse-to-fine scale dimensionality. Global low frequency information is first extracted and is used to guide the succeeding repeating extraction of local high frequency information for texture location and identification. In each stage, frequency features extracted are treated as the elements of a tree-like structure, which is compared against a reference data to identify the unknown donut produce using the decision metric of Euclidean distance. Specific regions of the donut can be extracted and later be identified with color analysis.

2.1 Hierarchical Processing

The RBG image acquired from the camera is converted into a CIELAB[4] image and the three channels (luminance L* and the chrominance a*,b*) of the image are processed separately. We process the image block-wise using a moving square window I of size NxN = 32x32 pixels to locate our area of interest AOI. - see Figure 1.

In the next level of processing, the image window is sub-divided. In our application, we used a total of 3 levels and hence a total of 21 child blocks, I_0 to I_{20}. Our image model assumed that original window I can be decomposed as a summation of its child sub blocks:

$$I(x,y)=\sum_{all\ i} I_i(x,y) \qquad (1)$$

[1] **Address:** Matsushita Communication Industrial Co.,LTD.
600,Saedo-cho, Tsuzuki-ku, Yokohama, JAPAN 〒224-8539
E-mail: hengck@adl.mci.mei.co.jp
[2] **E-mail:** iida.hidenobu@jp.panasonic.com
[3] **E-mail:** miwa.michio@jp.panasonic.com

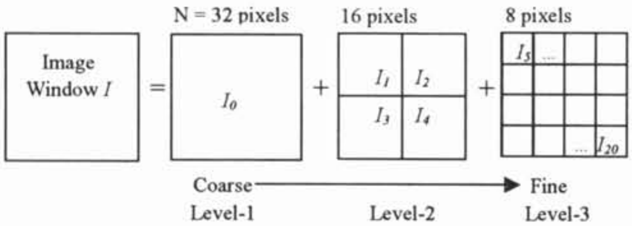


Figure 1.Sub-block model and Hierarchical Processing

Since the smallest sub-block is 8x8 pixels, it is sufficient from sampling theorem[1] to shift window I at intervals of $\Delta x = \Delta y = 4$ pixels while scanning from top to bottom and from left to right.

We shall refer an instance of the image child sub-block of any CIELAB channel as I_i in the following discussion.

2.1 Design of Gabor Filter Banks

For our application, we design three sets Gabor filters $\mathbf{\Psi_m}$ (for m = 1,2,3) for each of the three hierarchies respectively. A single Gabor filter can be formulated as the product of a Gaussian function and a complex sinusoidal modulator as shown in equation (2)

$$\psi_{f,\sigma,\theta,\lambda}(x,y)=\frac{1}{2\pi\lambda\sigma^{2}}e^{-\left[\left(\frac{x}{\sigma}\right)^{2}+\left(\frac{y}{\lambda\sigma}\right)^{2}\right]}e^{2\pi i[fx\cos\theta+fy\sin\theta]} \tag{2}$$

Consider a family of Gabor filters with $\mathbf{\Psi_m} = [\ldots\psi_{mn}\ldots]^t$ having varying orientation θ and aspect ratio λ and the dependent frequency $f = \lambda/(2\sigma)$. Here, the Gabor filter ψ_{mn} is the n-th row of the matrix $\mathbf{\Psi_m}$ in lexicographical order. By varying the aspect ratio, the Gabor filter can be made sensitive to edges when $\lambda \ll 1$ when locating features like holes of the donut and sensitive to repeating spatial sinusoidal texture when $\lambda \approx 1$ – see Figure 2.

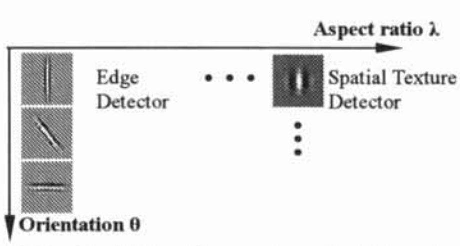


Figure 2. Family of Gabor filters used at a single hierarchy level

We attempt to project the image sub-block I_i onto $\breve{I}_i$ in the Gabor space spanned by $\mathbf{\Psi_m}$. We seek the best representation $\breve{I}_i$ where the reconstruction error $E_i = \| I_i - \breve{I}_i \|^2$ is minimized. The projected $\breve{I}_i$ can be described by:

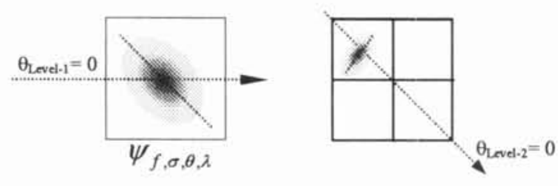
$$\breve{I}_i = \sum_{all\ m,n} w_{i,mn}\psi_{mn} \tag{3}$$

where $w_{i,mn}$ is the weight associated with each Gabor filter ψ_{mn}. Though the family of Gabor filters is not truly orthogonal and it is not possible to obtain the weight values just by direct projection[5][6], we find that it is good enough for our application to approximate the weight values by simply convolving the image sub-block with the filter desired:

$$w_{i,mn} = I_i * \psi_{mn} \tag{4}$$

For all levels, we use $\sigma = N/4$ so that more than 95% of its energy encompassed by the Gaussian function would fit into the block. In Level-1, frequency f is spaced at intervals of $\Delta\theta = 15^\circ$ from 0° to 165°. Aspect ratio λ has $\Delta\lambda = 0.1$ from 0.2 to 1.0.

In the next level, Level-2, the parameters of the Gabor filters are sampled at larger intervals since the sub-lock size is halved – see Figure 3. Orientation of the new family of filters is aligned to the largest notable filter response of the previous level (Level-1), i.e. $\theta_{Level-2} = 0$ in the direction where $w_{1,mn}$ is max. The design for Level-3 filter follows likewise.



ORIENTAION OF GABOR FILTERS ψ WITH STRONGEST RESPONSE $w_{i,mn}$

	Sub-Block Size N	sigma σ	Largest f (λ=1.0)	$\Delta\theta$	$\Delta\lambda$
Level-1	32	8	0.0625	15°	0.10
Level-2	16	4	0.1250	30°	0.25
Level-3	8	2	0.250	45°	0.50

Figure 3. Relation of parameters of Gabor Filter between different hierarchies

2.3 Feature Registering

Figure 4 shows the outline for texture location and discrimination as propose in our algorithm. Location of AOI is achieved by using a search window to scan the image from top to bottom and from left to right. At level-1, convolution is performed with the first and largest sub-block I_0 with the family of Gabor filter $\mathbf{\Psi_1}$ to obtain the weights of the filter $\mathbf{w_0}$. This represents global analysis at the coarsest scale. At the next level, level-2, the sub-block I_0 is sub-divided into its child sub-blocks I_1 ,I_2 ,I_3 and I_4. For each of the child sub-block, convolution is performed with a second family of Gabor filters $\mathbf{\Psi_2}$, which is related to $\mathbf{\Psi_1}$ as described in this section earlier on, to obtain the corresponding weights $\mathbf{w_1}$, $\mathbf{w_2}$, $\mathbf{w_3}$, $\mathbf{w_4}$. This represents local analysis guided by the global analysis. The process is repeated at the last level, level-3, where $\mathbf{\Psi_3}$ is used to obtain $\mathbf{w_5}, \ldots \mathbf{w_{20}}$.

We maintain a tree-like structure between the measured features $\mathbf{w_1}, \ldots \mathbf{w_{20}}$ to preserve their spatial relation. As an example shown in Figure 4-(c), region with distinctive pattern or image features (holes, corners, etc) would yield strongest response $w_{i,mn}$ for Gabor filters at particular orientation θ and aspect ratio λ. The black ellipses mark Gabor filters of strongest response in the case of the donut hole detection.

The tree-like structure can be described using a 3D signature vector. For classification, we used the metric Euclidean distance against previously learned data, defined as:

$$\text{Distance} = \sum_{all\ i,m,n} (w_{i,mn} - w_{i,mn,\ leaned\ data})^2 \tag{5}$$

If the minimum distance is within a threshold, block I_0 would be classified. Otherwise, it would be classified as unknown and regarded as noise.

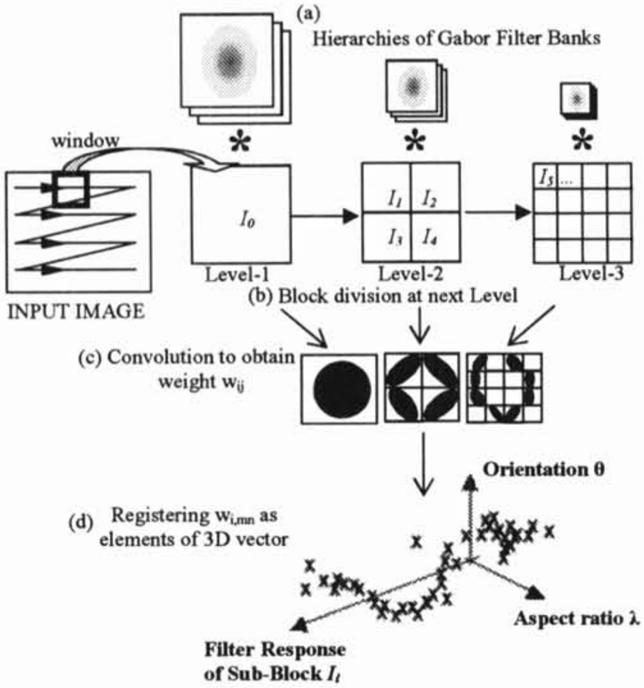


Figure 4. Outline of our proposed algorithm: (a) Design of inter-related hierarchical Gabor filter banks Ψ_m. (b) Image scanning and image block division. (c) Convolution to measure features w_i. (d) Registering w_i in a 3D vector.

2.4 Color Analysis

Color analysis is used after specific AOI's are located. For our application, such AOI's are assumed to consist of few colors (not more that 3). Each pixel of the AOI is mapped into the 3D color space and a K-means clustering algorithm is used to locate the centroids of the color masses. Identification is then simply based on classifying the largest mass closest to previously learned data point.

3 Experiment and Results

3.1 Experimental System

Our system consists of a Panasonic industrial camera (GP-US542 - 1/4" 3CCD High Performance Micro Head Color Camera) mounted on a stand, connected to a Pentium-4 1.2 GHz PC via a standard image video capture card. Images are captured at a resolution of 720 by 480 pixels at a depth of 24 bits per pixel. Each segmented donut image measures approximately 240 by 240 pixels.

For easy segmentation, the donut produces are placed on uniform light blue tray. A binarized image is first obtained from thresholding based on HSV values to remove the background blue pixels. Noises are then removed by morphological operators. A total of 60 images, consisting 10 types of donuts were used in the experiment. Figure 5 shows 6 sample images.

As an example, it can be seen that while certain donuts (types 5 and 6) have distinctive textures, others have similar or random textures (types 1,2 and 3) and can only be identified by locating specific features. Types 1,2 and 3 represent donuts of the same type but of different favor fillings. We attempt to differentiate them by detecting regions where the filling are exposed and analyze such regions by color – i.e. black, yellow or white filling.

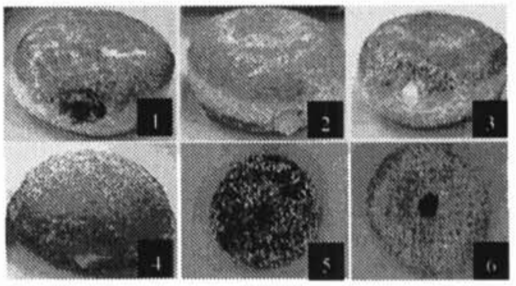


Figure 5. Some Typical donuts used for the experiment

3.2 Experimental Results

Figure 6. shows the decomposing of a sample image into its frequency components at three levels, as described in Sections 2. The images shows the least-error reconstructed image $\check{I}_i$ for the luminance L* channel.

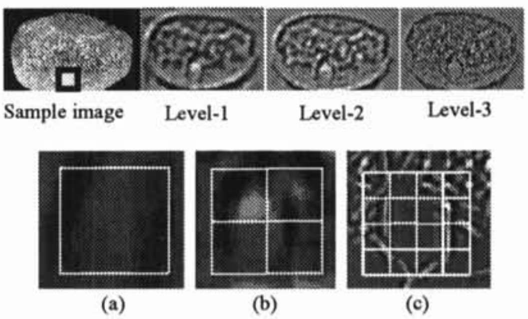
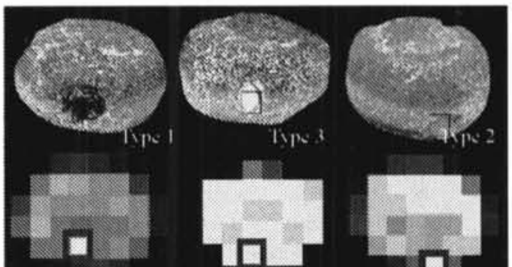


Figure 6. Frequency extraction at different levels. Figures (a) to (c) show the hole Gabor response in the "hole region", as indicated by the marked square in the sample image

Figure 7. shows the successful location of the area of interests. The top images show the target region, marked by a small square. The bottom images show the output of our algorithm, where the input image is divided into sub-block of 32 by 32 pixels and the intensity of each sub-block shows the confidence of locating the target region. In all the three cases, the correct regions are identified, as marked by the square in the bottom images.



	Type 1	Type 2	Type 3
% of correct location of AOI's	100.0	100.0	80.0

Figure 7. Extraction areas of interest AOI's

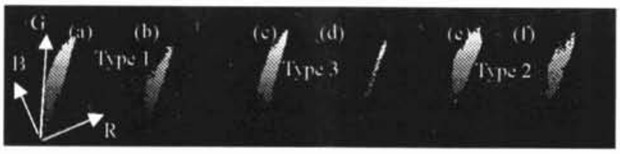


Figure 8. RGB Color distribution in the original donuts (a,c,e). Improved distribution for identification by considering the AOI's (b,d,f) only

Figure 8 shows how the color analysis can be improved by considering the AOI's alone rather than the whole input donut image.

Finally, Figure 9 shows the system overview and results. From the acquired image, segmented objects are first analysed for their global features where donut types A, B, C, D are identified. Different kind of type D donuts would need to be identified which only differs in color of the fillings, exposed at the edges of the donut. Local feature analysis is then performed on type D donuts, by changing the parameters of the Gabor Filter bank. Subtle texture differences due to different donut fillings are captured (marked by the little yellow boxes in the output) which enables color analysis to differentiate between them into D1, D2 and D3.

4 Conclusion

We have showed the use of our proposed algorithm for location and subtle texture discrimination in the donut recognition system using a hierarchy based Gabor filtering process. Most real world objects often consist texture of different types and great variations. In cases when the global texture is the same and only small regions are different, similar to that of the donut, the ability to identify object of complex composite texture would be of great practical importance and uses. Other possible applications would include those that due with natural texture like the skin and rock texture.

References

[1] Bovik, A.C., M. Clark, and W. Geisler. "Multichannel texture analysis using localized spatial filters." IEEE Transaction on Pattern Analysis and Machine Intelligence 12: 55-73,1990.

[2] Davar Pishava, Atsuo Kawai and Tsutomu Shiio, "Shape Bases Segmentation and Color Distribution Analysis with Application to Bread Recognition", Proceedings of MVA2000,IAPR Workshop on Machine Vision Application, pg. 193-196,2000.

[3] R. M. Bolle, J. Connell, N. Haas, R. Mohan, G. Taubin "VeggieVision: A Produce Recognition System" Proc. of the Third IEEE Workshop on Applications of Computer Vision (WACV-96), Sarasota FL, pp. 224-251, December 1996.

[4] Fred W. Billmeyer, Jr and Max Saltzman, "Principal of Color Technology (2nd Edition)", John Wiley 1981.

[5] J.Buhmann, J.Lange and C.v.d Malsburg, "Distortion Invariant object recognition by matching hierarchically labelled graphs ", Proceedings Int Conf. On Neural Networks, Vol 1, pg. 155-159, 1989.

[6] V. Krueger and G. Sommer. Gabor Wavelet Networks for Object Representation . pdf 22. DAGM Symposium, Kiel, Germany, September, 13-15, 2000.

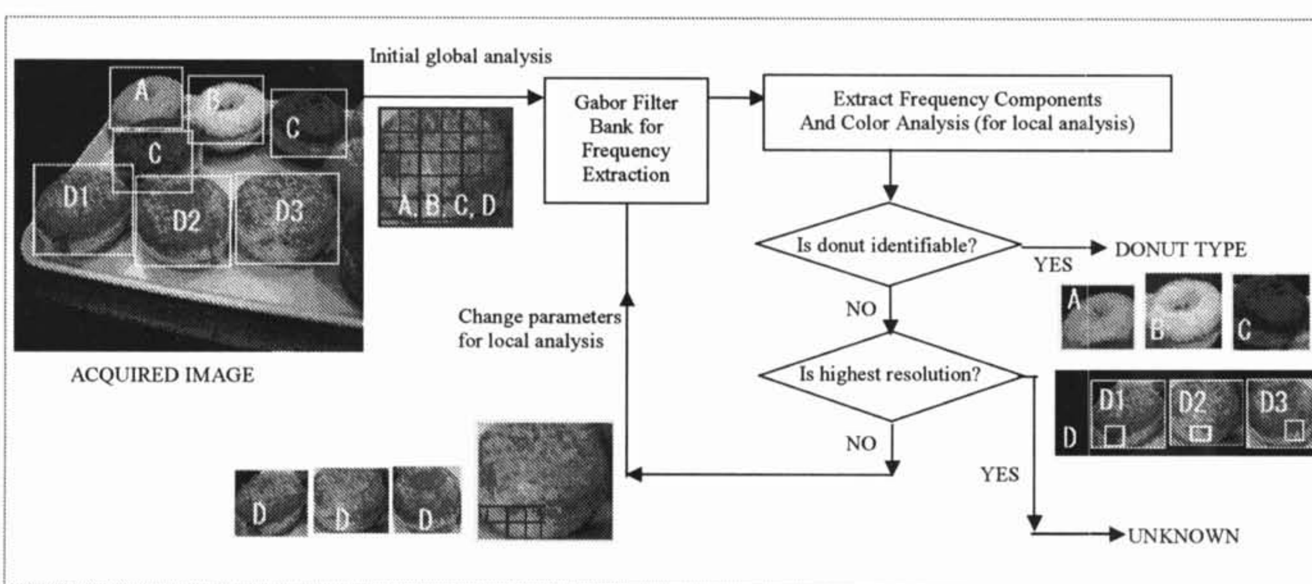


Figure 9. Overall view of implementation