

Texture-based Pattern Recognition Algorithms for the RoboCup Challenge

Bo Li and Huosheng Hu

Department of Computer Science, University of Essex
Wivenhoe Park, Colchester CO4 3SQ, United Kingdom
Email: bli@essex.ac.uk, hhu@essex.ac.uk

Abstract. Since texture is a fundamental character of images, it plays an important role in visual perception, image understanding and scene interpretation. This paper presents a texture-based pattern recognition scheme for Sony robots in the RoboCup domain. Spatial frequency domain algorithms are adopted and tested on a PC while simple colour segmentation and blob-based recognition are implemented on real robots. The experimental results show that the algorithms can achieve good recognition results in the RoboCup Pattern Recognition challenge.

1. Introduction:

In many applications, texture-based pattern recognition typically uses discrimination analysis, feature extraction, error estimation, cluster analysis (statistical pattern recognition), grammatical inference and parsing (syntactical pattern recognition). Besides colour, texture is a fundamental character of natural images, and plays an important role in visual perception. However, texture has a non-local image property. The basic methods for texture-based recognition are combinations of both spatial and frequency domains.

Texture-based segmentation algorithms divide the image into regions yielding different statistical properties. Such methods assume that the statistics of each region are stationary and that each region is extended over a significant area. However, most image regions do not present stationary features in the real world. Also, meaningless small-sized regions related to stains, noise or punctual information might appear. Consequently, methods relying on a priori knowledge of the number of textures in a given image [7] [11] often failed, because if any unexpected texture region appears, like the ones related to shadows or boundaries, a wrong fusion of two non-related regions is forced. Unsupervised segmentation [1][3] does not rely on such knowledge, but it is slow because it requires a computationally expensive additional stage to calculate the correct number of regions. Gabor filters and Gabor space are very useful tools in spatial frequency domains. There is some similarity between Gabor process and human vision [8][9][10][6]. Some methods combine statistical approaches and learning algorithms in order to produce good results[8]. Multi-resolution and multiband methods are both very useful in this area.

In the area of texture recognition and segmentation, image modulation and multiband techniques are very useful tools to analyse texture-based patterns. Havlicek has done a lot in AM-FM modulation for texture segmentation [3][4][13]. Images with multi-texture are processed by a bank of Gabor filters. Using the filter responses, Dominant Component Analysis and Channelised Component Analysis are implemented for texture decomposition. There are also other researchers using similar methods.

The rest of the paper is organized as follows. Section 2 describes briefly texture-based recognition, including the pattern recognition challenge in RoboCup and the

texture analysis process. In section 3, the issues related to colour filters and spatial domain processes are investigated. Then, both multi-band techniques and Gabor filters are proposed for texture segmentation in section 4. Section 5 presents image modulation modules based on multi-dimensional energy separation. Experimental results are presented in section 6 to show the feasibility and performance of the proposed algorithms. Finally, a brief conclusion and future work are given in section 7.

2. Texture-based Pattern Recognition

2.1 Pattern Recognition Challenge in RoboCup

In the Sony robot challenge of RoboCup 2002, each Sony robot is expected to recognize the targets with chessboard texture pattern and individual geometrical shapes, which are placed at random positions and orientations. To implement this task, we define a function F to represent the pattern recognition system that can be used in this challenge. More specifically, its inputs are images with dimension of $M \times N$, and its output for each image is the element of a name set or a code set. For example, a name set $S = \{triangle, square, rectangle, T, L\}$ may be defined here for the recognition challenge of Sony robots in the RoboCup competition. Figure 1 shows 4 targets to be identified in the challenge. The size, rotation angle and background are random.

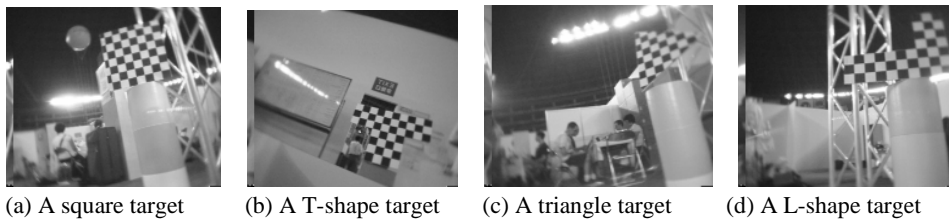


Fig. 1. Target patterns in the RoboCup Sony Robot Challenge

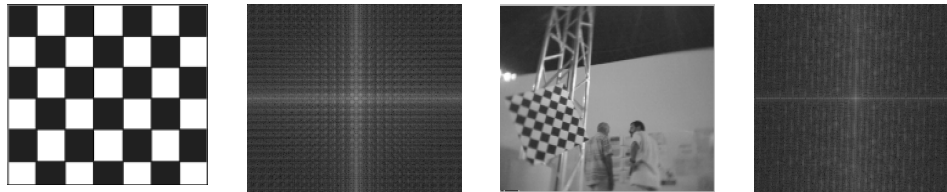
The whole process of recognition can be decomposed to four stages. There are many methods that can be applied at each stage. In general, complex methods can do well, but are expensive. Simple methods are required for real-time systems. More specifically,

- Stage 1 -- *Image processing* stage can enhance image quality and make segmentation or edge detection much easier by removing noises and blurs images. Some images that are not focused well can be sharpened.
- Stage 2 -- *Image conversion* includes filter process, DFT (Discrete Fourier Transform) process, DWT (Discrete Wavelet Transform) process and so on. Its main purpose is to make properties of images easy to be extracted. For instance, skin colour can be used to identify hands of humans. In the RoboCup Sony robot league, the competition field is colour-based. Under natural environment with complex backgrounds, Sony robots may be unable to find objects easily by using colour information only. Further complex methods are required to make it possible [5].
- Stage 3 -- *Information extraction* to obtain useful information for recognition. In most cases, it should output data in a special format for processing at a later stage. For instance, blob-based methods should output useful information for all blobs, including position, size, and so on. In some cases this part is called coding, as it is describing objects with numeric properties.
- Stage 4 -- *Pattern recognition* converts numeric properties to most-like objects and output their name. The process depends on the previous results. If input data is easy to be discriminated, this process can be very simple.

2.2 Texture Analysis Process

Texture is a fundamental character of natural images. It plays an important role in visual perception and provides information for image understanding and scene interpretation. Figure 2 (a) shows the target texture used in the experiments.

Figure 2 (b) shows the magnitude of the DFT result from Figure 2 (a). The result is centered, logarithmically compressed and histogram stretched. From this result, the frequency distribution of the pattern is clear. However, it is difficult to design a filter for recognition. Figure 2 (c) shows an original image captured by the robot.



(a). Chessboard (b) DFT magnitude (c) An original image (d) Along Y-axis

Fig. 2. Texture Analysis Process

Figure 2(d) shows the DFT magnitude for luminance in Figure 4. Because of its complex background, no filtering hint for the pattern can be seen from figure 5. It is impossible to separate the pattern from complex background by using a simple filter in the frequency domain.

3. Colour Filter and Spatial Domain Process

A colour filter can be used to obtain pixels with the concerned colour. Spatial domain processes such as morphology filters cost little, but are very effective for smoothing binary images. Following processes are all based on the image shown in Figure 2(c). The colour filter here are LUT (Look Up Table) based, which was implemented on the robots for the purpose of good performance. Figure 3 shows the result by the LUT method. The concerned colour is white. Figure 4 shows the Horizontal Signature, i.e. the image projection onto the x-axis. The first peak is the dark area and the second part is mainly white part.

It is not easy to separate white and black blobs from the background on the basis of the signatures alone. Figure 5(a) shows the result for black blobs.

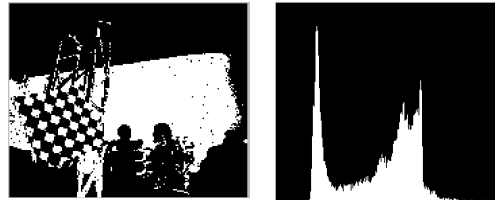


Fig. 3 An image, LUT **Fig. 4.** Horizontal Signature

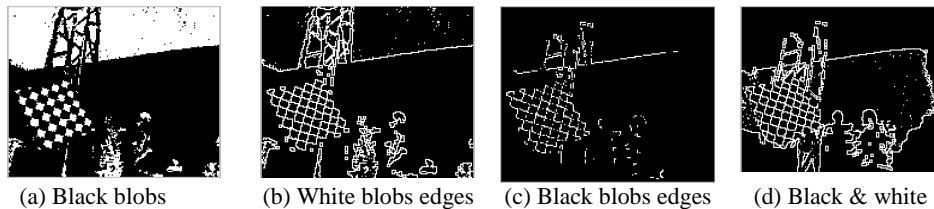


Fig. 5. Spatial domain process

With LUT results, edges can be extracted. Figure 5(b) shows the image with edges of white blobs. The edge detection here is morphology based, i.e. binary calculation. Figure 5(c) shows an image with edges of black blobs. Figure 5(d) shows the overlay of black

and white edges. The main problem with spatial processes is that they are limited by the size of masks. In fact, the target frequencies are fixed with masks. It is therefore difficult to extract useful information for recognition, especially with complex background.

4. Multi-band Techniques & Gabor Filters for Texture Segmentation

In this section, the research focus is placed on multi-band techniques in order to overcome the problem existing in spatial processes described in the previous section. Gabor filters are adopted for texture segmentation in the RoboCup domain. As presented in [6], multi-band techniques are based on the following model:

$$f(m, n) = \sum_{q=1}^Q f_q(m, n) \quad (1)$$

It assumes that an image $f(m, n)$ can be decomposed into Q components and each component is based on a given frequency band. As we know, Gabor filters are wavelets based narrow band filters. With a bank of Gabor filters, it is possible to isolate components from one another. In some cases, they can be used to detect the direction of target's rotation. Gabor wavelets have following properties:

- Gabor functions achieve the theoretical minimum space-frequency bandwidth product; that is, spatial resolution is maximised for a given bandwidth.
- A narrow-band Gabor function closely approximates an analytic function. Signals convolved with an analytic function are also analytic, allowing separate analysis of the magnitude (envelope) and phase characteristics in the spatial domain.
- The magnitude response of a Gabor function in the frequency domain is well behaved, having no side lobes.
- Gabor functions appear to share many properties with the human visual system.

4.1 Two-dimensional Gabor function: Cartesian Form

The Gabor function is extended into two dimensions as follows. In the spatial frequency domain, the Cartesian form is a 2-D Gaussian formed as the product of two 1-D Gaussians:

$$G_C(\omega_x, \omega_y, \omega_{Cx'}, \omega_{Cy'}, \theta, \sigma_x, \sigma_y) = G(\omega_{x'}, \omega_{Cx'}, \sigma_{x'}) G(\omega_{y'}, \omega_{Cy'}, \sigma_{y'}) \quad (2)$$

where θ is the orientation angle of G_C , $x' = x \cos \theta + y \sin \theta$, and $y' = -x \sin \theta + y \cos \theta$.

In the spatial domain, G_C is separable into two orthogonal 1-D Gabor functions that are respectively aligned to the x' and y' axes:

$$g_C(\omega_x, \omega_y, \omega_{Cx'}, \omega_{Cy'}, \theta, \sigma_x, \sigma_y) = g(\omega_{x'}, \omega_{Cx'}, \sigma_{x'}) g(\omega_{y'}, \omega_{Cy'}, \sigma_{y'}) \quad (3)$$

An image could be represented as

$$f(x, y) = \sum_{n_x=-\infty}^{\infty} \sum_{n_y=-\infty}^{\infty} \sum_{m_x=-\infty}^{\infty} \sum_{m_y=-\infty}^{\infty} \beta_{m_x, m_y, n_x, n_y} h_{m_x, m_y, n_x, n_y}(x, y) \quad (4)$$

where $h_{m_x, m_y, n_x, n_y}(x, y) = g(x - n_x X, m_x \Omega_x, \sigma_x) g(y - n_y Y, m_y \Omega_y, \sigma_y)$; $\sigma_x, \sigma_y, X, Y, \Omega_x, \Omega_y$ are constants; and $X \Omega_x = Y \Omega_y = 2\pi$. Assume that the parameters are chosen appropriately, approximation to $\beta_{m_x, m_y, n_x, n_y}$ are obtained by using

$$\hat{\beta}_{m_x, m_y, n_x, n_y} = f(x, y) h_{m_x, m_y, n_x, n_y}(x, y) \approx \beta_{m_x, m_y, n_x, n_y} \quad (5)$$

4.2 Gabor filtering results

In this research, a bank of 72 Gabor filters are implemented on images captured by Sony robots. The filters are of 9 frequencies. For each frequency, there are 8 orientation angles. The parameters of the filters can be set for different applications. Figure 6(a) shows the filter property of its real part and the orientation is 0. Figure 6(b) shows another filter property of its real part. It has higher frequency and the orientation is 90. Figure 6(c) shows the product of Gabor filter.

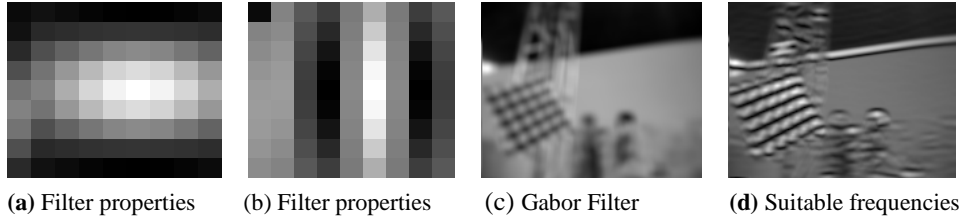


Fig. 6. Gabor filtering results

The responses with only low frequency features remain. By adjusting the frequency and orientation parameters, some results can be used as texture recognition or segmentation. Figure 6(d) shows a suitable frequency with an orientation of 0-degree.

Figure 7 shows different responses by filters with different orientations. Orientation is very important for filtered results. The texture region is stressed greatly in several image results. The next problem is how to find a good result in object recognition without human supervision. It will be discussed in the next section.

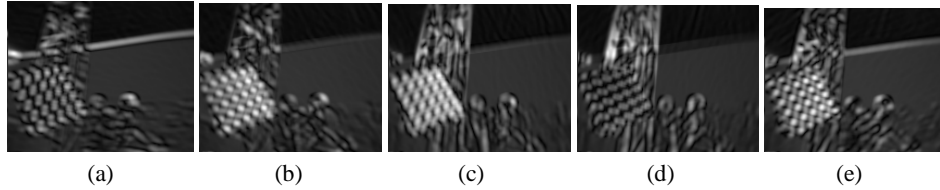


Fig. 7. Different responses of filters at different orientations

5. Image Modulation Models & Multidimensional Energy Separation

The unsupervised image segmentation is very important for computer vision and image understanding, especially for robotic systems. Comparing with colour segmentation, texture based segmentation is much more challenging. Image modulation models are very useful in this area, which may be used to represent a complicated image with spatially varying amplitude and frequency characteristics as a sum of joint amplitude-frequency modulated AM-FM components [4].

5.1 Multidimensional Energy Separation

For any given image $f(n,m)$, there are infinitely many distinct parts of functions $\{\hat{a}(n,m), \nabla \hat{\phi}(n,m)\}$. As an example, we could interpret the variation in $f(n,m)$ exclusively as frequency modulations by setting $a(n,m) = \max|f(n,m)|$ and $\phi(n,m) = \arccos\left(\frac{f(n,m)}{a(n,m)}\right)$.

Teager-Kaiser energy operator (TKEO) could be used to estimate the AM-FM function.

For a 1-D signal $f(n)$ the discrete TKEO is defined by following:

$$\Psi[f(n)] = f^2(n) - f(n+1)f(n-1) \quad (6)$$

When applied to a pure cosine signal $f(n) = A \cos(\omega_0 n + \phi)$, the TKEO yields $\Psi(f(n)) = A^2 \omega_0^2$, a quantity that is proportional to energy required to generate the displacement $f(n)$ in a mass-spring harmonic oscillator. For many locally smooth signals such as chirps, damped sinusoids, and human speech formats, the TKEO delivers

$$\Psi(f(n)) = \hat{a}^2(n) \hat{\phi}_e(n) \quad (7)$$

where $\hat{a}(n)$ and $\hat{\phi}_e(n)$ are good estimates of an intuitively appealing and physically meaningful part of modulations. $\{a(n), \phi(n)\}$ satisfies $f(n) = a(n) \cos[\phi(n)]$. The quantity $a^2(n) \phi^2(n)$ is known as the Teager energy of the signal $f(n)$. A 2-D discrete TEKO is:

$$\Phi[f(n, m)] = 2f^2(n, m) - f(n-1, m)f(n+1, m) - f(n, m-1)f(n, m+1) \quad (8)$$

For a particular pair of modulation functions $\{a(n, m), (n, m) \nabla \phi\}$ satisfying $f(n, m) = a(n, m) \cos(\phi(n, m))$, the operator $\Phi[f(n, m)]$ approximates the multidimensional Teager energy $a^2(n, m) |\nabla \phi(n, m)|^2$. For images that are reasonably locally smooth, the modulating functions selected by the 2-D TKEO are generally consistent with intuitive expectations. With the TKEO, the magnitudes of the individual amplitude and frequency modulations can be estimated using the energy separation analysis (ESA):



$$|\hat{U}(n, m)| = \arcsin \sqrt{\frac{\Phi[f(n+1, m) - f(n-1, m)]}{4\Phi[f(n, m)]}} \quad (9)$$

$$|\hat{V}(n, m)| = \arcsin \sqrt{\frac{\Phi[f(n, m+1) - f(n, m-1)]}{4\Phi[f(n, m)]}} \quad (10)$$

$$\hat{a}(n, m) = |\hat{a}(n, m)| = \sqrt{\frac{\Phi[f(n, m)]}{\sin^2[|\hat{U}(n, m)|^2] \sin^2[|\hat{V}(n, m)|^2]}} \quad (11)$$

Fig. 8. An example of TEKO

Fig. 8 shows the magnitude with the TEKO. The operation is based upon the image in Figure 6(d) and the luminance level was adjusted here for the purpose of presentation.

5.2 Multi-component Demodulation

Multi-component demodulation is based on the following model:

$$f(h, m) = \sum_{q=1}^Q a_q(n, m) \cos[\phi_q(n, m)] = \sum_{q=1}^Q f_q(n, m) \quad (12)$$

One popular approach for estimating the modulating functions of the individual components is to pass the image $f(n, m)$ or its complex extension $z(n, m)$ through a bank of band pass linear Gabor filters mentioned above. Each filter in the filter bank is called a channel. For a given input image, each channel produces a filtered output that is named as the channel response.

Suppose that $h_i(n, m)$ and $H_i(n, m)$ are respectively the unit pulse response and frequency response of a particular filter bank channel. Under mild and realistic assumptions, one may show that, at pixels where the channel response $y_i(n, m)$ is dominated by a particular AM-FM component $f_q(n, m)$, the output of the TKEO is well approximated by

$$\Phi[y_i(n,m)] \approx a_q^2(n,m) |\nabla \varphi_q(n,m)|^2 |H_i[\nabla \varphi_q(n,m)]|^2 = \Phi[f_q(n,m)] |H_i[\nabla \varphi_q(n,m)]|^2 \quad (13)$$

In fact, the results of the filters can be passed into this conversion for the AM magnitude separation. Figure 9 shows the AM modulation result from Figure 15 and Figure 16. No luminance level was adjusted.

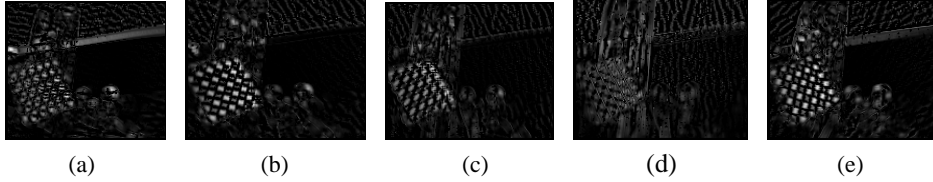


Fig. 9. AM modulation result from Figure 7

For real-time requirement, to select a suitable frequency and orientation of the filter can be based on pre-known data and energy separation. The Sony robot can measure an approximate distance to a target with an infrared range sensor. This means that the frequency and orientation of the target can be estimated by a robot system, which is very helpful for the selection of a suitable filter.

6. Experimental Results

Information extraction from a well-pre-processed image is a simple task. From the image in either Fig. 7 or Fig. 9 both edge-based and blob-based methods can work well. Harris algorithm [2][12] can be used to get the corner and edges. In contrast, a blob-based algorithm can obtain the exact position and the number of blobs in a texture pattern, especially for the image in Fig. 9. Then, the pattern can be recognized based on the position of corners or blobs. The method can be described as follows:

Step 1: Find one edge of the shape & rotate it to make the edge paralleled with Y-axis.

Step 2: Calculate the distribution of blob or corner positions on X-axis and Y-axis.

Step 3: Recognize the pattern by using the rules as follows:

- If position numbers are equal on both X-axis and Y-axis, it is a square target.
- If position numbers are equal on X and Y respectively, it is a rectangle target.
- If position numbers in the middle of X are bigger than both side, and position numbers of one side of Y are bigger than another side, it is a T-shape target. Otherwise, it is a triangle target.
- If position numbers of one side are bigger than another side on both X and Y axes, which is a L-shape target.

Figure 10(a) shows the edges detected from the image in Figure 7(e). Figure 10(b) shows the result of linearisation. Figure 10(c) shows the result of rotation. In fact, from the result of linearisation, it is clear that the pattern is a square. Ambient methods can avoid problems caused by viewpoints. For complex shapes, complex algorithms are necessary.

It is estimated that for each operation of a Gabor filter, with an image dimension $M \times N$, filter dimension $W \times H$, the number of basic calculations can be up to $(M - W + 1)(N - H + 1)W^2H^2$. For the experiment above, $M=176$, $N=144$, $W=9$, $H=9$, it costs about 149905278 basic calculation, 0.5 second on a Pentium III 500 PC. If it runs on the robot with a MIPS 4000 100 CPU, it will take about 1.5 second. The cost can be reduced by 64% if reducing the filter dimension to 7×7 . For the new version of Sony AIBO robot with a supercore CPU, it is possible to implement the filtering processes at a rate of 6Hz, which becomes necessary for complex environments.

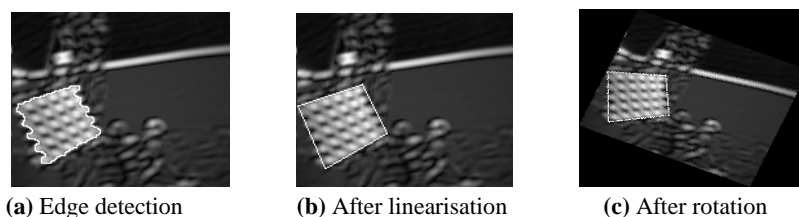


Fig. 10. Experimental results

7. Conclusions and Future Work

In this paper, we have developed spatial frequency domain algorithms that can recognize shapes in chessboard texture in a cluttered environment. Unlike simple algorithms that require clear background and less interfere; spatial frequency domain algorithms can recognize shapes from a cluttered background. These algorithms have been successfully implemented on a PC. In contrast, simple colour segmentation and blob-based recognition are implemented on real robots to satisfy the real-time requirements. When the background is less cluttered, good recognition results can be achieved reliably.

In our future work, the spatial frequency algorithms will be implemented on real robots for real-time performance, especially object recognition under a complex environment. A leaning system will be integrated in order for the Sony robot to recognize colour objects under complex and cluttered environments.

References

- [1] C. Bouman and B. Liu, "Multiple resolution segmentation of textured images", *IEEE Transactions on Pattern Anal. Machine Intel l.*, 13 (2), pages 99-113, 1991
- [2] C.G. Harris and M.J. Stephens. "A combined corner and edge detector", *Proceedings Fourth Alvey Vision Conference, Manchester*, pages 147-151, 1988
- [3] J. P. Havlicek, A. C. Bovik, and D. Chen, "AM-FM image modelling and Gabor analysis", *Visual Information Representation, Communication, and Image Processing*, 343-385
- [4] J.P. Havlicek and A.C. Bovik, "Image Modulation Models", in *Handbook of Image and Video Processing*, A.C. Bovik, ed., Academic Press, pages 305-316, 2000
- [5] B. Li, H. Hu and L. Spacek, An Adaptive Colour Segmentation Algorithm for Sony Legged Robots, *Proc. 21st IASTED Int. Conf. on Applied Informatics*, pp. 126-131, Innsbruck, Austria, 10-13 February 2003
- [6] B. S. Manjunath, G. M. Haley, W. Y. Ma, "Multiband techniques for texture classification and segmentation" in *Handbook of Image and Video Processing*, A.C. Bovik, ed., Communications, Networking, and Multimedia Series by Academic Press, pp. 305-316, 2000
- [7] M. Pietikinen and A. Rosenfeld, "Image segmentation using pyramid node linking", *IEEE Trans. on Systems, Man and Cybernetics*, SMC-11 (12), pages 822-825, 1981
- [8] J. Puzicha, T. Hofmann, and J. Buhmann, "Histogram clustering for unsupervised segmentation and image retrieval", *Pattern Rec. Letters*, Vol. 20, No. 9, pp. 899-909, 1999
- [9] T. Reed and H. du Buf, "A review of recent texture segmentation and feature extraction techniques", *CVGIP: Image Understanding*, Vol. 57, No. 3, pages 359-372, 1993
- [10] Y. Rubner and C. Tomasi, "Texture-based image retrieval without segmentation", *Proceedings of ICCV*, pages 1018-1024, 1999
- [11] O. Schwartz and A. Quinn, "Fast and accurate texture-based image segmentation", *Proceedings of ICIP96*, Vol. 1, pages 121-124, Laussane-Switzerland, 1996
- [12] J. Shi and C. Tomasi. "Good Features to Track", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1994
- [13] T. Tangsuksorn and J.P. Havlicek, "AM-FM image segmentation", *Proceedings of IEEE International Conference on Image Processing*, Vancouver, Canada, pages 104-107, 2000