# Background Modeling Using Phase Space for Day and Night Video Surveillance Systems

Yu-Ming Liang*, Arthur Chun-Chieh Shih*, Hsiao-Rong Tyan+, and Hong-Yuan Mark Liao*

*Institute of Information Science, Academia Sinica, Nankang, Taipei 115, Taiwan

+Institute of Computer Engineering, Chung-Yuan Christian University, Taiwan

**Abstract.** This paper presents a novel background modeling approach for day and night video surveillance. A great number of background models have been proposed to represent the background scene for video surveillance. In this paper, we propose a novel background modeling approach by using the phase space trajectory to represent the change of intensity over time for each pixel. If the intensity of a pixel which originally belongs to the background deviates from the original trajectory in phase space, then it is considered a foreground object pixel. In this manner, we are able to separate the foreground object from the background scene easily. The experimental results show the feasibility of the proposed background model.

## 1. Introduction

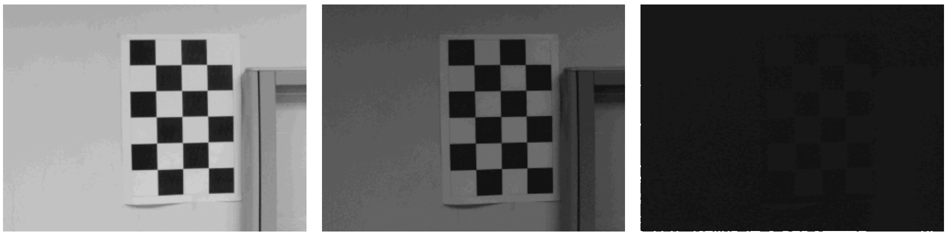
Visual surveillance has become an important research issue in recent years. Since the price of video sensors keeps going down, a great number of researchers have devoted themselves to the development of video surveillance systems [1-5]. A good visual surveillance system should be able to function day and night. However, most of the existing visual surveillance systems cannot work twenty four hours using only one sensor. In order to make a visual surveillance system function well during the night, a night vision sensor is indispensable [6]. In [6], Owens and Matthies proposed to use a night vision sensor (which is actually an infrared sensor) to perform monitoring during the night. However, a visual system that is inexpensive and is able to operate either at the day time or at night is always preferable. A star-light camera, under these circumstances, becomes a good candidate that best fits our requirements.

In the development of a conventional visual surveillance system, foreground object detection is usually the first step that requires to be handled. A correct foreground object extraction process is fairly important because a false detection will make the subsequent processes, such as tracking and recognition, become invalid. There are three types of existing approaches that have been proposed for foreground detection [3-5, 7-14]. They are: temporal differencing [7], background subtraction [3-5, 9-14], and optical flow-based estimation [8]. The background subtraction approach is commonly adopted in most of the existing video surveillance systems

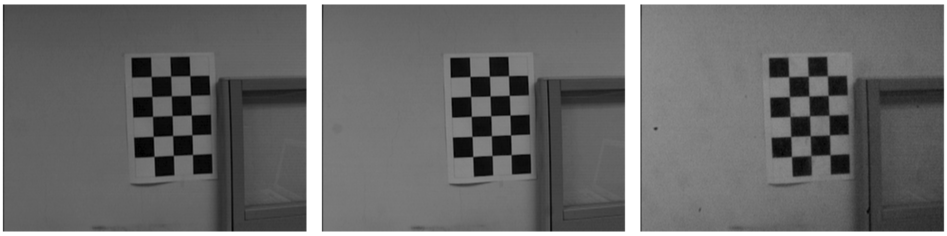
[3-5]. This approach extracts the foreground components by comparing a new frame with the background model, which is represented by a pre-determined background scene. Under the circumstances, an efficient foreground object extraction process depends heavily on a successful background modeling.

There are many existing methods designed for background modeling, and the most famous ones are all statistical-based approaches [3, 4, 9-12]. Among a number of successful models, some used a single Gaussian distribution of intensity (color) for each pixel to represent the background model [9, 10]. If every pixel is resulted from a particular surface in a completely static scene, the pixel intensity can be modeled with a Gaussian distribution due to the nature of noises. However, if the scene is not completely static, multiple surfaces are needed to model a same pixel, such as waving trees. As a consequence, a mixture of Gaussian distributions has to be used to model the above non-static signal [3, 11]. For solving the same problem, Elgammal *et al*. [12] proposed to use a nonparametric kernel density function to substitute the use of the mixture of Gaussian. In [4], Haritaogul *et al*. proposed the so-called $W^4$ system which used a bimodal distribution to model the background scene. The proposed system represents the background scene by three different values for each pixel. These values include: the minimum and the maximum intensity values and the maximum intensity change between two consecutive frames. In [13, 14], the researchers used linear prediction to predict the expected background, in order to separate the foreground from the background.

In this paper, we propose a novel background modeling approach for day and night video surveillance by using a star-light camera. A star-light camera is able to acquire clear images even under very poor lighting conditions. A star-light camera uses Auto Gain Control (AGC) and Auto Electronic Shutter Control (AESC) to maintain the video level on a fixed IRE value, and therefore it can acquire images with similar brightness under various lighting conditions. Fig. 1 shows the acquired images taken by a conventional camera and a star-light camera, respectively, under three different lighting conditions. It is clear that the images taken by a star-light camera can maintain with a satisfactory brightness even under very poor lighting conditions. However, a star-light camera also has its drawback when encountering a light source which changes with respect to time. For example, if the light source is a fluorescent tube, then a star-light camera will activate AGC and AESC mechanism to compensate the changing lighting conditions all the time. As a consequence, the above mentioned compensation will cause the output video signal unstable. Fig. 2 illustrates the phenomenon that the intensity of a pixel over time forms a quasi-periodic signal, and Fig.2 (b) indicates that the intensity contributed by the foreground may replace the signal of the background. Having the above mentioned problem at hand, we are not able to use the existing background modeling methods [3, 4, 9-14] to solve the problem. Since a star-light camera is inexpensive and has been widely used, we shall propose a new background modeling method to increase its accuracy. First, we shall transform the intensity signal from the time domain to the phase space. Then, a phase space trajectory is used to represent the intensity change over time for every pixel in the background scene. By observing the change of this trajectory, one is able to separate the foreground from the background easily.

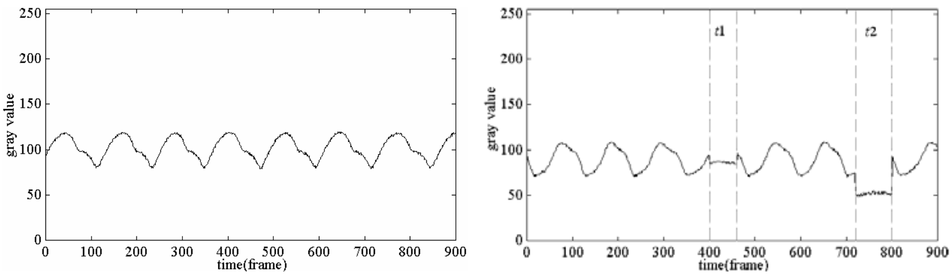


(a)

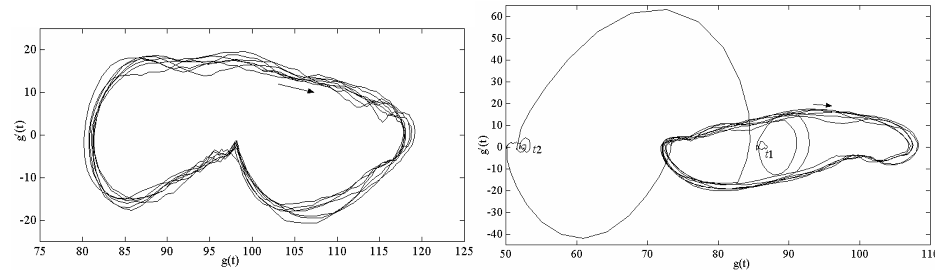


(b)

Fig. 1. Images taken by (a) a conventional camera, and (b) a star-light camera, respectively, under three different lighting conditions.




(a) (b)




(c) (d)

Fig.2. The intensity of a pixel over time, (a) only background signal; (b) the

composite signal that includes background and foreground signals; (c) the phase space diagram corresponding to the signal shown in (a); (d) the phase space diagram corresponding to the signal shown in (b).

The rest of this paper is organized as follows. In Section 2, we shall introduce what the phase space is. Then, the proposed background modeling will be elaborated in Section 3. The experimental results will be reported in Section 4, and the conclusion will be drawn in Section 5.

## 2. Phase Space

In the nature, most dynamic systems are nonlinear, and they are usually difficult to be solved by analytical methods. Under the circumstances, the phase space approach [15] is one of the possible ways that can be applied to analyze the behavior of nonlinear dynamic systems. The phase space approach has been successfully applied to many research fields, e.g. speech [16] and medicine [17]. Fig. 2 shows that the intensity of a pixel over time is actually a nonlinear dynamic problem. Therefore, the phase space method can be applied to analyze the change of intensity over time for each pixel. Suppose the intensity value and the rate of change of intensity are represented by the $X$-axis and $Y$-axis, respectively, in the phase space. Fig. 2(c) shows the phase space diagram that is corresponding to the signal shown in Fig. 2(a). It is clear that the phase space diagram has a particular appearance and trajectory.

It has been made clear that the goal of foreground detection is to separate the foreground from the background. In Fig. 2(b), it is obvious that the background and the foreground are represented by different signals (the durations of $t1$ and $t2$ are foreground and the rest are background). Fig. 2(d) shows the phase space of the signal shown in Fig. 2(b). When some foreground objects appear at the time intervals $t1$ and $t2$, the corresponding trajectories are significantly deviated from the original trajectory. From the above observed phenomenon, it is apparent that the phase space model is indeed a very good tool for modeling the background scene. From the trajectories shown in the phase space, it is very easy to separate a foreground object from the background scene.

## 3. The Proposed Method

In this paper, the phase space trajectory is applied to represent the change of intensity over time for each pixel in the background scene. Therefore, the task of background modeling is to model the phase space trajectory. For the purpose of efficiency, we apply the B-spline curve fitting approach to model the phase space trajectory.

3.1 Background Modeling

Suppose the change of intensity over time for each pixel is an intensity signal function of time, and the phase space is a two-dimensional space that consists of the intensity function (the *X*-axis) and its first order derivative (the *Y*-axis). Since a derivative is quite sensitive to noise, a smoothing process applying to the intensity signal function in advance is necessary. For each pixel *x*, we let the measured intensity signal function be $g_x(t)$ after executing a Gaussian smoothing process. The measured function $g_x(t), 1 \le t \le N$, can be used as a training signal to model the phase space trajectory.

Fig. 2(c) illustrates that a cycle of the phase space corresponds to one period in the intensity signal. Though the periodic durations of the same intensity signal are close to each other, they are not exactly the same. Therefore, a representative period which is the average period of all periods should be decided and then used to represent the period of the signal.

In order to derive the average period, we extract all peaks (or valleys) in a training signal $g_x(t)$, and then use all its constituent peaks to segment $g_x(t)$ into *k*-1 periods:

$$P_i(t) = \begin{cases} g_x(t), & \text{if } \ t_i \le t \le t_{i+1} \\ 0 \quad , & \text{otherwise} \end{cases}, \quad i = 1, \ldots, k-1, \tag{1}$$

where *k* is the number of extracted peaks and $t_i$ represents the instant of the $i^{\text{th}}$ peak in $g_x(t)$. We use $T_i = t_{i+1} - t_i$ to represent the duration of the $i^{\text{th}}$ period, and then we calculate the average period from these *k*-1 periods:

$$T = \frac{1}{k-1} \sum_{i=1}^{k-1} T_i \,, \tag{2}$$

$$P(t) = \frac{1}{k-1} \sum_{i=1}^{k-1} P_i\left(\frac{t - t_k + T}{T} T_i + t_i\right), \quad t_k - T \le t \le t_k \,, \tag{3}$$

where *T* is the duration of the average period. Fig. 3 illustrates an example showing how an average period looks like in two different diagrams.

After an average period is calculated, we apply uniform cubic B-spline curve fitting [18] to model it:

$$G_x(u) = \sum_{j=0}^{n+M-1} N_{j,M}(u) Q_{j \bmod (n+1)}, \quad k_{M-1} \le u \le k_{n+M} \,, \tag{4}$$

where $N_{j,M}(u)$ are the $j^{\text{th}}$ B-spline function of order $M = 4$, $Q_j$ is the $j^{\text{th}}$ control point, and the set of knots are as follows:

$$k_i = \frac{i - M - n}{n+1} T + t_k, i = 0, 1, \ldots, n + 2M - 1 \,. \tag{5}$$

Finally, we use $B_x(u) = (G_x(u), G'_x(u))$ to represent the phase space trajectory of any single pixel *x*.

In addition to background modeling, background updating is also an important task, because the background scene won't stay still forever. In this work, we update the phase space trajectory while a new period emerges. We use the previously derived average period to calculate the new average period and then use it as a new average period.

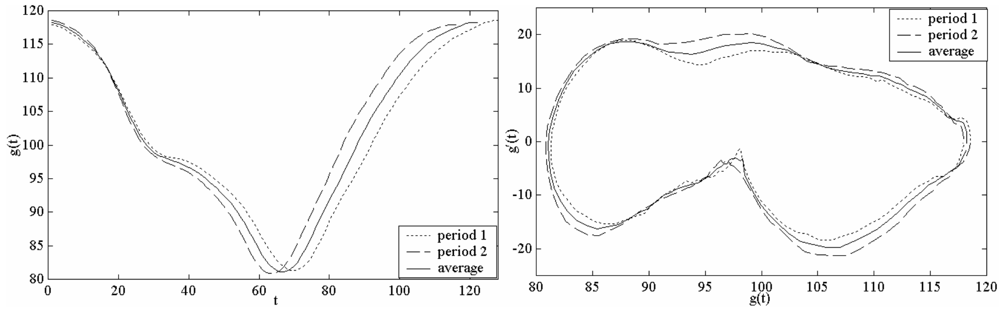



Fig. 3. An example showing how an average period looks like in two different diagrams.

3.2 Foreground/Background Determination

After the training signal is modeled by a phase space trajectory, the trajectory model can be used in the foreground/background determination process for subsequent signals. Fig. 2(d) illustrates clearly that the trajectory deviates from its original (background) path when some foreground objects cover this pixel. In order to make a correct judgement on whether a pixel is occupied by a foreground object or by a background scene, we need to fully utilize the trajectories of the phase space. Suppose the phase space trajectory of the original intensity signal is $v_x(t) = (g_x(t), g'_x(t))$, and the trajectory model is $B_x(u) = (G_x(u), G'_x(u))$. In Fig. 4, it is clear that $v_x(t)$ cannot completely fit $B_x(u)$, but $v_x(t)$ will encircle along the surrounding of $B_x(u)$ following the time order. Thus we can calculate the shortest distance from $v_x(t)$ to $B_x(u)$ under the time order constraint. The distance can be taken as the criterion for determining foreground or background, and the trajectory point $B_x(u)$ with the shortest distance can be represented as the predicted point.

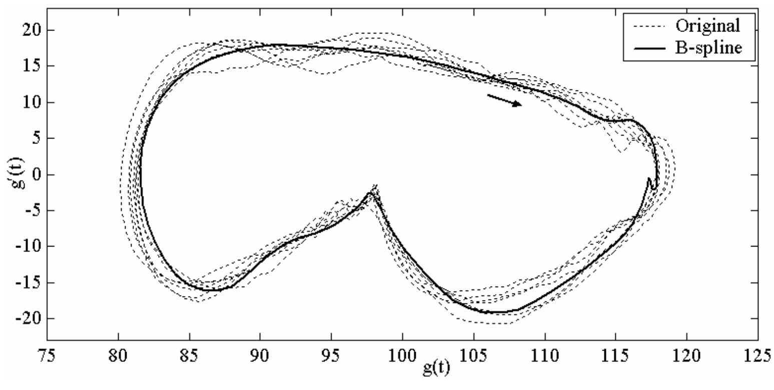


Fig. 4. The phase space trajectory of the original intensity signal and the established trajectory model using B-spline fitting.

Assume that the previously predicted point is $B_x(u')$, and then we calculate $u$ to make the distance from $v_x(t)$ to $B_x(u)$ the shortest under the constraint $u' < u < u' + r$. $r$ here is a given maximum possible change. Therefore, the distance can be calculated as follows:

$$d = \sqrt{(G_x(u) - g_x(t))^2 + (G'_x(u) - g'_x(t))^2}\,, \tag{6}$$

$$d^2 = (G_x(u) - g_x(t))^2 + (G'_x(u) - g'_x(t))^2\,. \tag{7}$$

When the distance is minimal, the first derivative of $d$ is zero.

$$0 = (G_x(u) - g_x(t))\frac{G_x(u)}{du} + (G'_x(u) - g'_x(t))\frac{G'_x(u)}{du}\,. \tag{8}$$

The order of this equation is five, and it is hard to compute. Thus, we use linear interpolation to estimate $u$. Finally, the pixel can be classified as foreground if $d > T$, where $T$ is a given threshold. Otherwise, the pixel is a background pixel. The physical meaning of the above formulation is as follows. When a pixel is part of the background, its corresponding intensity won’t deviate from the original trajectory too far. On the other hand, if there is a foreground object covering this pixel, its corresponding intensity will soon jump away from the original background trajectory.

## 4. Experimental Results

We have conducted a series of experiments to test the effectiveness of the proposed method. The upper half of Fig. 5 shows the intensity signal of a background pixel. The lower half of Fig. 5 shows the estimated distance to the background trajectory. We used the first period of the signal as the training signal to model the background trajectory. After the training process, it is clear that at each instant the estimated intensity distance to the background trajectory was non-zero, but very close to zero. In the second part of the experiment, we used a synthetic signal to conduct the experiment. The upper half of Fig. 6 shows a synthesized intensity signal with $t1$ and $t2$ durations replaced by other signals (this is equivalent to placing an object on the background). Again, we used the first period to train the system. From the estimated distances shown in the lower half of Fig. 6, it is apparent that the estimated distances located in $t1$ and $t2$ durations were much larger than those located in other durations. This means whenever there are any non-periodic signals occurred and some durations of the original periodic background signals were replaced, the corresponding estimated intensity distances of these replaced durations would jump to significantly high. Therefore, we are able to easily separate the foreground from the background by examining the estimated distance change. In the last part of the experiment, we used a real intensity signal to conduct the experiment. From $t$ = 220 to $t$ = 690, the real intensity signal corresponds to the background. At the duration from $t$ = 700 to $t$ = 730, a foreground object emerged and it covered the target pixel. It is clearly seen in the lower half of Fig. 7 that the estimated distances between $t$ = 700 to $t$ = 730 were significantly high in comparison with the estimated distances measured at other instants.

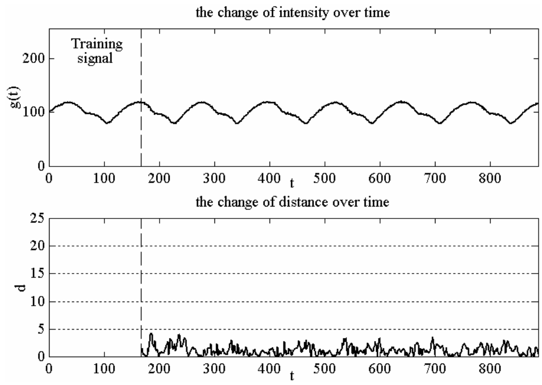



Fig. 5 The distance estimation for a background signal.

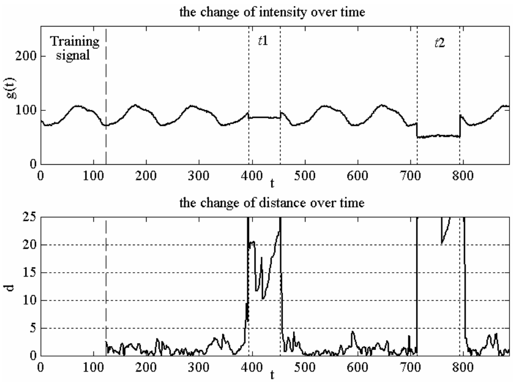



Fig. 6. The distance estimation for a synthetic intensity signal, including background and foreground signals.

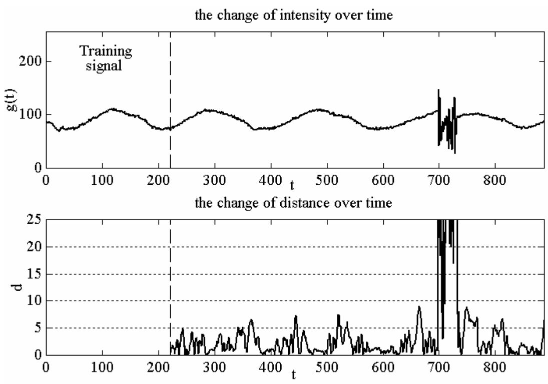



Fig. 7. The distance estimation for a real intensity signal, including background and foreground signals.

## 5. Conclusion

We have proposed a novel background model for day and night video surveillance. We use the phase space trajectory to represent the change of intensity over time for each pixel in the background scene. Furthermore, we detect the foreground by determining whether the current trajectory deviates from the background trajectory or not. The experimental results show the feasibility of the proposed background model. Since there are many issues in the background modeling, such as waving trees and shadows, our future work is to solve these issues based on the proposed background model.

## References

[1] R. T. Collins, A. J. Lipton, and T. Kanade, "Introduction to the Special Section on Video Surveillance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22 pp.745-746, 2000.

[2] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, and O. Hasegawa, "A System for Video Surveillance and Monitoring: VSAM Final Report," Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May, 2000.

[3] W. E. L. Grimson, C. Stauffer, R. Romano, and L. Lee, "Using Adaptive Tracking to Classify and Monitor Activities in a Site," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, pp. 22-29, 1998.

[4] I. Haritaogul, D. Harwood, and L. S. Davis, "$W^4$: Real-Time Surveillance of People and Their Activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 809-830, 2000.

[5] C. J. Pai, H. R. Tyan, Y. M. Liang, H. Y. Mark Liao, and S. W. Chen, "Pedestrian Detection and Tracking at Crossroads," *Pattern Recognition*, Vol.37, pp. 1025-1034, 2004.

[6] K. Owens and L. Matthies, "Passive Night Vision Sensor Comparison for Unmanned Ground Vehicle Stereo Vision Navigation," *Proceedings of IEEE Conference on Robotics and Automation*, San Francisco, CA, 2000.

[7] C. Anderson, P. Burt, and G. van der Wal, "Change Detection and Tracking Using Pyramid Transformation Techniques," *Proceedings of SPIE-Intelligent Robots and Computer Vision*, Vol. 579, pp. 72-78, 1985.

[8] J. Barron, D. Fleet, and S. Beauchemin, "Performance of Optical Flow Techniques," *International Journal of Computer Vision*, Vol. 12, pp. 43-77, 1994.

[9] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Transactions on Pattern Analysis and Machine*

*Intelligence*, Vol. 19, No. 7, pp. 780-785, 1997.
[10] I. Cohen, G. Medioni, and H. Gu, "Inference of 3D Human Body Posture from Multiple Cameras for Vision-Based User Interfaces," *Proceedings of the 5th World Multi-Conference on Systemics, Cybernetics and Informatics*, Orlando, Florida, USA, 2001.
[11] C. Stauffer and W. E. L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 246-252, 1999.
[12] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance," *Proceedings of the IEEE*, Vol. 90, No. 7, July, 2002.
[13] C. Ridder, O. Munkelt, and H. Kirchner, "Adaptive Background Estimation and Foreground Detection using Kalman-Filtering," *Proceedings of International Conference on Recent Advances in Mechatronics (ICRAM)*, pp. 193-199, 1995.
[14] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and Practice of Background Maintenance," *Proceedings of the 7th IEEE International Conference on Computer Vision*, Vol. 1, pp. 255-261, 1999.
[15] W. K. Tang, Y. K. Wong, and A. B. Rad, "Qualitative phase space modeling of non-linear electrical dynamic systems," *Proceedings of IEEE Midnight-Sun Workshop on Soft Computing Methods in Industrial Applications*, Kuusamo, Finland, pp. 140-145, 1999.
[16] M. A. Jackson and I. S. Burnett, "Phase-Space Portraits of Speech Employing Mutual Information and Perceptual masking," *Proceedings of IEEE Workshop on Speech Coding*, pp. 61-63, 1999.
[17] L. Feng, Z. Chongxun, and W. Xiaoyu, "Reconstructing Phase Space for Nonlinear Analysis of Hear Rate," *Proceedings of the 20th Annual International Conference IEEE Engineering in Medicine and Biology Society*, Vol. 20, No. 3, pp. 1576-1578, 1998.
[18] F. Yamaguchi, "Curves and Surfaces in Computer Aided Geometric Design," Springer-Verlad, 1988.