



Research Journal of  
**Information  
Technology**

ISSN 1815-7432



Academic  
Journals Inc.

[www.academicjournals.com](http://www.academicjournals.com)

## Moving Target Detection in Complex Background

Yu-Long Qiao and Xian-Rui Song

College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China

*Corresponding Author: Xian-Rui Song, College of Information and Communication Engineering, Harbin Engineering University, 145 Nantong Street, Nangang District, Harbin, Heilongjiang 150001, PR China  
Tel/Fax: +86-13766903335*

### ABSTRACT

Moving target detection is a key problem in many image based applications. For these scenes whose background is complex caused by stochastic motion, it becomes more difficult to well detect the moving targets. This study proposes a center-surround discriminant saliency based detection method using the Stationary Wavelet Transform (SWT) and the generalized Gamma distribution (GFD). It makes use of the fact that the properties of the moving target and the complex background is different in the stationary wavelet domain and can be described by the GFD model. The center-surround discriminant saliency which is to measure the difference between the current location and its surround, is determined with the Symmetrized Kullback-Leibler Distance (SKLD) of the GFD models of the center and surround in the wavelet transform. Then the moving target is detected based on this discriminant saliency. The experimental results demonstrate the effectiveness of the proposed method.

**Key words:** Moving target detection, discriminant saliency, stationary wavelet transform, generalized gamma distribution

### INTRODUCTION

For a human visual system, we not only focus on static image features significantly including luminance, color, contrast to the surrounding area, but also pays attention to the movement of the interesting objects. Therefore, moving target detection from the background became an important research topic in the image sequence or video processing. There are lots of algorithms for moving target detection. However, for these scenes whose background is complex caused by stochastic motion, it becomes more difficult to well detect the moving targets (Kaawaase *et al.*, 2011; Nagarajan and Balasubramanie, 2008; Sulaiman *et al.*, 2008).

In order to detect the moving target in image sequences, there are some typical algorithms such as mixture Gaussian model (Hui *et al.*, 2011). The mixture Gaussian model based method can realize the background estimation and adaptive background update and uses background subtraction between frames to get the moving target. However, for complex dynamic background, e.g., crowd, rain, swaying trees and other moving objects such like birds and smoke-filled environment, this traditional method may detect the dynamic scene as moving targets. Therefore, Gao *et al.* (2008) proposed the center-surround discriminant saliency algorithm. They also demonstrated the performance of the method with the discriminant saliency detection in static imagery and discriminant saliency for dynamic scenes. The wavelet domain Generalized Gaussian

Distribution (GGD) is used to describe the interesting region during the saliency detection. However, Van de Wouwer *et al.* (1999) verified that some wavelet detail histograms may not be fitted well by the GGD. Because the observed histograms do not have a sharp peak at zero but shift toward left or right (Do and Vetterli, 2002). In such cases, the GFD model which is a three-parameter (scale, shape and index shape parameters) distribution model, is a good choice to replace the GGD model. Furthermore, Song (2008) presented that the GFD model often outperforms the GGD for modeling Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) coefficients.

In this study, we proposed a moving target detection algorithm based on the center-surround discriminant saliency by using the Stationary Wavelet Transform (SWT) and the GFD model. The SWT is performed on the image sequences (video) to obtain the multiscale decomposition. Then the GFD is used to model the wavelet coefficients of the center and surround windows in wavelet transform domain. The center-surround saliency map is formed by combing the center and surround discrepancy that is measured with the Symmetrized Kullback-Leibler Distance (SKLD). Finally, the moving target is determined from the saliency map.

### THEORY BACKGROUND

**Stationary wavelet transform:** The standard DWT is a non-redundant and compact representation of signal in transform domain. Although the DWT possesses many usefully characteristics and achieves satisfying result (Chen *et al.*, 2003; Venkateswaran and Rao, 2007; Loum *et al.*, 2007; Logashanmugam and Ramachandran, 2008), it is translation-variant. That is to say, when a signal is translated, its numerical descriptors will be not translated but modified. The reason is that there is a decimation step after filtering operation. The Stationary Wavelet Transform (SWT) has a similar tree structured implementation without any decimation (subsampling) step. Therefore, the SWT is a shift-invariant transform that is suitable for the signal analysis.

The implementation of the stationary wavelet transform is similar to that of the discrete wavelet transform, except that there is no downsampling operation. For a one-dimensional signal  $x$ , the block diagram of SWT is shown in Fig. 1. It can also be expressed as follows:

$$\begin{aligned} s_{i+1}(k) &= h_{\uparrow 2^i} * s_i(k) \\ d_{i+1}(k) &= g_{\uparrow 2^i} * s_i(k) \end{aligned} \tag{1}$$

with  $S_0 = x$ .  $*$  is the convolution operator. Equation 1 involves two basic filters  $h$  and  $g$ . At each step, a branch output of the last step,  $S_i$ , convolutes with the filters  $h_{\uparrow 2^i}$  or  $g_{\uparrow 2^i}$ , where  $[\cdot]_{\uparrow 2^i}$  denotes the upsampling by a factor of  $2^i$  which also means inserting  $2^i-1$  zeros between two taps of the basic filters.

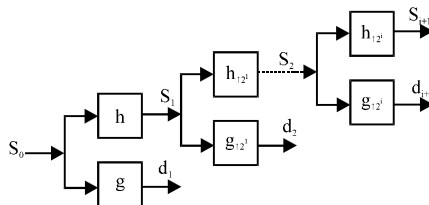


Fig. 1: 1D stationary wavelet transform (1D SWT)

For the 2D separable wavelet transform of an image can be easily derived from 1D wavelet decomposition. One-level decomposition of an image can be obtained by first applying 1D SWT to each row of the image and then each column of the resulting image. Successively perform one-level decomposition on the resulting low frequency image and then obtain the discrete wavelet representation of this image. The 2D SWT produces three detail subbands at each level, each of which is strongly oriented at orientations 0, 45 and 90°.

**Generalized gamma distribution (GFD):** The GFD is presented in 1962 (Stacy, 1962) and applied in different fields, such as speech spectra characterization, texture retrieval and texture classification (Maliani *et al.*, 2010). Given the complex structure and enormous variation of the image, Choy and Tong (2010) demonstrated it is better than GGD for the histogram model of image variations. This study uses GFDs to model the coefficient distributions of the interesting regions. The Probability Density Function (PDF) of GFD is:

$$f(x; \alpha, \beta, \lambda) = \frac{\beta}{2\alpha^{\beta}\Gamma(\lambda)} x^{\beta\lambda-1} e^{-\frac{x}{\alpha}{}^{\beta}} \quad x \geq 0, \alpha, \beta, \lambda > 0 \quad (2)$$

where,  $\alpha$  and  $\beta$  are scale and shape parameters,  $\lambda$  is the index shape parameter and  $\Gamma(\cdot)$  is the Gamma function. The GFD model covers many distributions such as the generalized Gaussian, Gamma and Weibull distribution.

Given the wavelet coefficients  $x_i$  of the interesting region, the GFD parameters can be estimated either by using the Method of Moments (MM) or by using Maximum Likelihood Estimation (MLE). It is well known that MLE is better than MM for the parameters estimation. Thus, we use MLE approach (Choy and Tong, 2010) to estimate parameters  $\alpha$  and  $\lambda$  as follows:

$$\hat{\alpha} = \left[ \frac{1}{N\lambda} \sum_{i=1}^N x_i^{\beta} \right]^{\frac{1}{\beta}} \quad (3)$$

$$\hat{\lambda} = \frac{1}{\beta} \left[ \frac{\sum_{i=1}^N x_i^{\beta} \log x_i}{\sum_{i=1}^N x_i^{\beta}} - \frac{\sum_{i=1}^N \log x_i}{N} \right]^{-1} \quad (4)$$

where, N is the number of the wavelet coefficient, The shape estimator  $\hat{\beta}$  can be computed by the sample Scale-in-dependent Shape Estimation (SISE) equation, defined by:

$$S(\beta) = \log \left( \frac{1}{N} \sum_{i=1}^N |x_i|^{2\beta} \right) - 2 \log \left( \frac{1}{N} \sum_{i=1}^N |x_i|^{\beta} \right) - \log \left[ 1 + \beta \left( \frac{\sum_{i=1}^N |x_i|^{\beta} \log |x_i|}{\sum_{i=1}^N |x_i|^{\beta}} - \frac{\sum_{i=1}^N \log |x_i|}{N} \right) \right] \quad (5)$$

For measure the similarity between two GFD models of the interesting regions, we adopt the Symmetrized Kullback-Leibler Distance (SKLD):

$$\begin{aligned}
 \text{SKLD}(f_1 \| f_2) &= \text{KLD}(f_1 \| f_2) + \text{KLD}(f_2 \| f_1) = (\beta_1 \lambda_1 - \beta_2 \lambda_2) \left( \log \left( \frac{\alpha_1}{\alpha_2} \right) + \frac{\psi(\lambda_1)}{\beta_1} - \frac{\psi(\lambda_2)}{\beta_2} \right) \\
 &+ \left( \frac{\alpha_1}{\alpha_2} \right)^{\beta_2} \frac{\Gamma(\lambda_1 + \beta_2 / \beta_1)}{\Gamma(\lambda_1)} + \left( \frac{\alpha_2}{\alpha_1} \right)^{\beta_1} \frac{\Gamma(\lambda_2 + \beta_1 / \beta_2)}{\Gamma(\lambda_2)} - \lambda_1 - \lambda_2
 \end{aligned} \tag{6}$$

where,  $\alpha_i$ ,  $\beta_i$  and  $\lambda_i$  ( $i=1,2$ ) are the parameters of the GFD.

### PROPOSED ALGORITHM

The discriminant center-surround hypothesis, introduced by Gao *et al.* (2008), is that the most salient locations of the visual field are those that enable the discrimination between feature responses in center and surround with smallest expected probability of error. It has been shown that the discriminant saliency based algorithms achieve state-of-the-art performance in some image processing problems, such as interest point detection, object recognition and background subtraction.

The discriminant center-surround saliency can be formed as a binary classification problem (Gao *et al.*, 2008; Zhang *et al.*, 2008). At each position  $l$  of the video, there are two classes: a class of stimuli of interest and a surround. The stimuli of interest are the observations within a neighborhood  $w_1^l$  of the location  $l$  and referred to center (with label  $c = 1$ ). The surround (with label  $c = 0$ ) is composed of all the stimuli that are not salient which is the observation within a surrounding window  $W_1^0$  of the neighborhood  $W_1^1$ . Based on the feature extracted from the center and surround regions, the saliency of the location  $l$  can be determined by the discriminant power of the features between the two regions. The interest object in the center window means that there has large disparity of features responses between center window and surround window. Therefore, we use this center-surround formulation to detect the moving target.

Specifically, for a video to detect the moving target, we perform  $k$ -level SWT on each frame of this video which produces  $3k$  detail subbands with the same size as the original frame. At the position  $l$  of the current frame of the video, the center window  $w_1^0$  and surround window  $W_1^1$  are shows as Fig. 2. These windows have the similar regions  $W_{1i}^0$  and  $W_{1i}^1$  in each detail subbands. The wavelet coefficients in the windows  $W_{1i}^0$  and  $W_{1i}^1$  are bandpass features which will be modeled with the GFD. Mathematically, the bandpass features in the center windows  $W_{1i}^0$  are assumed to be drawn from the condition probability density GFD  $p_{Y|c}(y_i|1)$  and those in surround windows  $W_{1i}^1$  from  $p_{Y|c}(y_i|0)$ .

The saliency  $S(l)$  is defined as the power that the features  $Y$  can discriminate between the two classes. Based on the minimum probability of the Bayes classification error,  $S(l)$  is quantified by the mutual information between the feature response  $Y$  and the class label  $C$ :

$$\begin{aligned}
 S(l) = I_1(Y, C) &= \sum_c \int p_{Y,c}(y, c) \log \frac{p_{Y,c}(y, c)}{p_Y(y)p_c(c)} dy \\
 &= \sum_c p_c(c) \int p_{Y|c}(y|c) \log \frac{p_{Y|c}(y|c)}{p_Y(y)} dy \\
 &= \sum_c p_c(c) \text{KL}(p_{Y|c}(y|c) \| p_Y(y))
 \end{aligned} \tag{7}$$

where,  $\text{KL}(\cdot \| \cdot)$  is the Kullback-Leibler divergence between two probability distributions. A larger  $S(l)$  implies there has large disparity between center and surround which indicates a (part) moving target in the center.

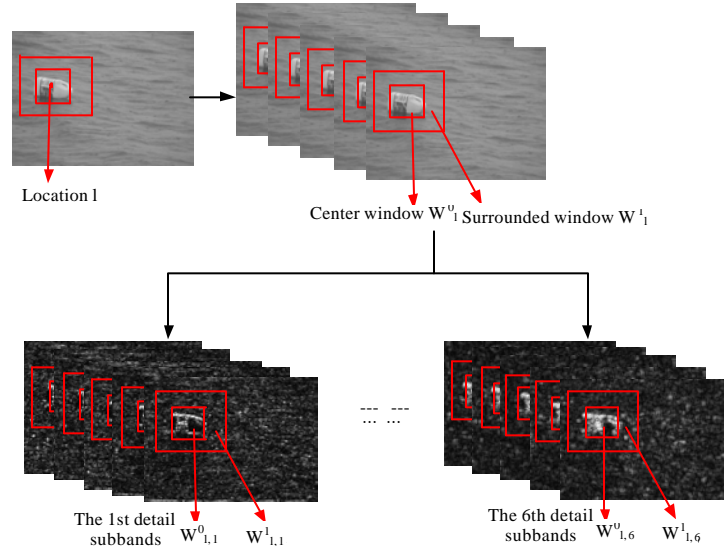


Fig. 2: Center and surround windows

Because performing the SWT on the image sequences, the feature of a location should be multi-dimension. According to the result, the saliency  $S(l)$  can be approximated as follows (Mahadevan and Vasconcelos, 2010):

$$\begin{aligned}
 S(l) &= I_l(Y, C) = \sum_i I_l(Y_i, C) \\
 &= \sum_i \sum_c p_c(c) \text{KL}(p_{y_i|c}(y_i|c) \| p_{y_i}(y_i))
 \end{aligned} \tag{8}$$

The GFD model is quite sensitive to its parameters. That is to say, a small change in model parameter may lead to subtle differences in its distribution. Therefore, we replace  $\text{KL}(\cdot \| \cdot)$  in saliency  $S(l)$  with Symmetrized Kullback-Leibler Distance (SKLD):

$$\text{SKLD}(p \| q) = \text{KL}(p \| q) + \text{KL}(q \| p) \tag{9}$$

Then  $S(l)$  is rewritten as:

$$S(l) = \sum_i \sum_c p_c(c) \text{SKL}(p_{y_i|c}(y_i|c) \| p_{y_i}(y_i)) \tag{10}$$

In general, when the size of a center window is identified as  $n \times n \times t$ , the size of the surround window is set to  $6n \times 6n \times t$ . In order to generate  $S(l)$ , we should estimate the GFD parameters of each window and compute  $p_c(l) \text{SKL}(p_{y_i|c}(y_i|l) \| p_{y_i}(y_i))$ . However, due to the unbalance of the sample size (the number of the wavelet coefficients in the window), the estimated parameters of the center window are different from those of the whole window  $W^0_{l,i} \cup W^1_{l,i}$ , even if they belong to the same category. Therefore, this study uses the parameters estimated from  $y_i \in W^0_{l,i} \cup W^1_{l,i} \cup W^l_{l,i}$ , obtains the corresponding PDF  $p_{y_i}(y_i)$  and replaces  $p_c(l) \text{SKL}(p_{y_i|c}(y_i|l) \| p_{y_i}(y_i))$  with  $p_c(l) \text{SKL}(p_{y_i}(y_i) \| p_{y_i}(y_i))$ . Then we have:

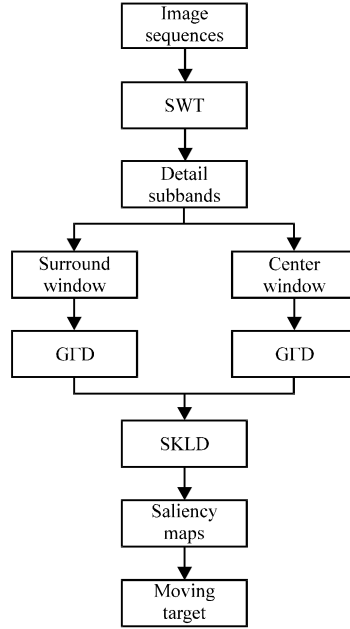


Fig. 3: Moving target detection algorithm

$$S(l) = \sum_i \left( p_c(1) \text{SKL}(p_{y_i}(y_i') \| p_{y_i}(y_i)) + p_c(0) \text{SKL}(p_{y_i c}(y_i | 0) \| p_{y_i}(y_i)) \right) \quad (11)$$

Though Eq. 11 is different from Eq. 8, it could be used to describe the discriminant power of the features extracted from the location  $l$  in differentiating the two classes.

In summary, the discriminant center-surround saliency maps are generated by using the SWT and GFD models. When a moving target moves to the center window, the saliency of this location will become larger. Based on this fact, the target could be detection from the saliency map. The entire detection algorithm is shown in Fig. 3 and summarized as follows:

- **Step 1:** Perform  $k$ -level SWT on each frame of the image sequence and there will generate  $3k$  detail subbands
- **Step 2:** For each position  $l$ 
  - Identify two corresponding windows in each detail subbands: the center window  $W_{li}^0$  and the surround window  $W_{li}^1$
  - Estimate the GFD model parameters of the surround window  $W_{li}^1$ , the whole window  $W_{li}^0 \cup W_{li}^1$  and the window  $W_{li}^0 \cup W_{li}^1 \cup W_{li}^1$  with MLE method
  - Compute the saliency  $S(l)$  by using Eq. 11
- **Step 3:** Determine a threshold value, the region whose saliency values are larger than the threshold  $T$  is the moving target

## EXPERIMENTS AND RESULTS

To evaluate the performance of the proposed algorithm, we chose two image sequences with the dynamic backgrounds which are downloaded from the Internet (Mahadevan and Vasconcelos,

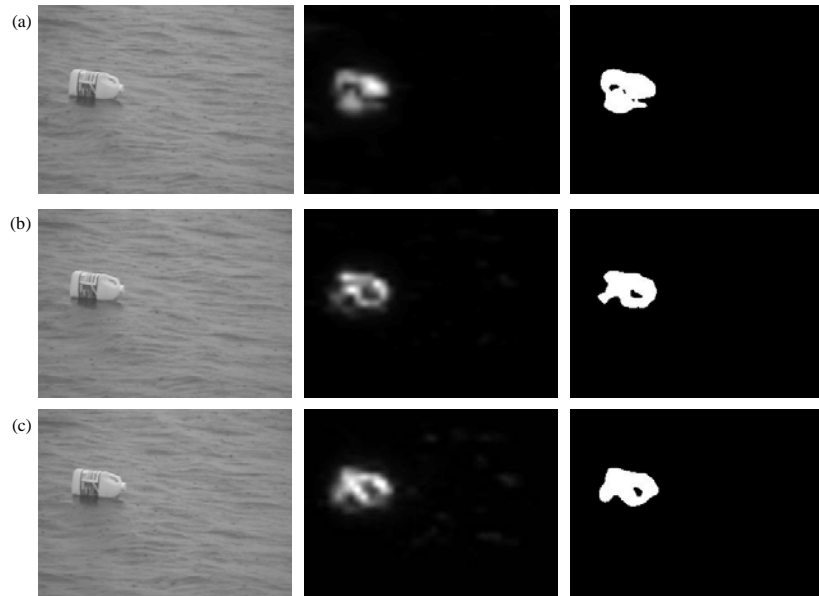


Fig. 4(a-c): Image sequences “bottle” and the detection results

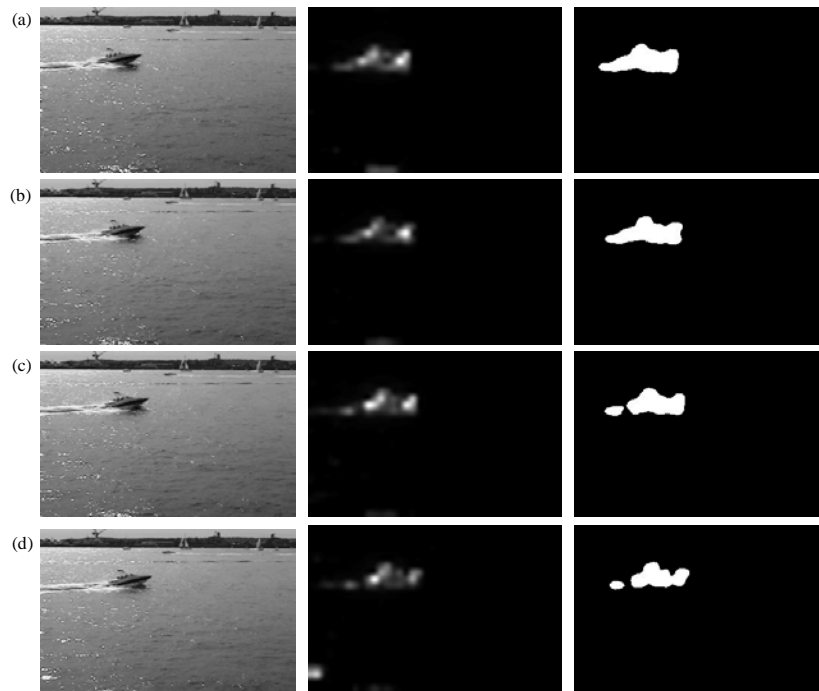


Fig. 5(a-d): Image sequences “boat” and the detection results

2008). Each image sequence is converted to grayscale images. Figure 4 shows a bottle floating in the waving water. Figure 5 shows a driving boat which has the sweeping wave. Firstly, each frame of the image sequence is decomposed with the 2-level SWT. At each detail subband, we use the



center window of size  $16 \times 16 \times 5$  pixels and surround window of size  $96 \times 96 \times 5$  pixels and adopt the MLE to estimate the GFD model parameters. The experimental results are shown in Fig. 4 and 5. In both cases, the background is dynamic and the saliency maps completely show the foreground objects and does adapt well to the dynamic scene. The result shows that the proposed algorithm can almost completely ignoring the dynamic background and clearly detect the moving target. In other words, this method can obtain better detection results.

## CONCLUSIONS

In this study, we propose a moving target detection method based on the discriminant center-surround saliency that is generated by using the Stationary Wavelet Transform (SWT) and generalized Gamma distribution (GFD). The center-surround saliency has various advantages, such as, complete unsupervised, robust to the complex dynamic background and invariant to the camera motion. Meanwhile, the distribution GFD provides a good model for the stationary wavelet coefficients. Therefore this method can effectively detect the moving target based on the saliency map. The experimental results demonstrate the performance of the proposed method.

## ACKNOWLEDGMENTS

This study is partially supported by National Natural Science Foundation of China 60902064 and 61172159, Postdoctoral Science-research Developmental Foundation of Heilongjiang Province LBH-Q09128. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers which have improved the presentation.

## REFERENCES

- Chen, T.H., G. Horng and S.H. Wang, 2003. A robust wavelet-based watermarking scheme using quantization and human visual system model. *Inform. Technol. J.*, 2: 213-230.
- Choy, S.K. and C.S. Tong, 2010. Statistical wavelet subband characterization based on generalized  $\Gamma$  density and its application in texture retrieval. *IEEE Trans. Image Process*, 19: 281-289.
- Do, M.N. and M. Vetterli, 2002. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Trans. Image Process*, 11: 146-158.
- Gao, D., V. Mahadevan and N. Vasconcelos, 2008. On the plausibility of the discriminant center-surround hypothesis for visual saliency. *J. Vision*, Vol. 8. 10.1167/8.7.13.
- Hui, S., S. Kai and H. Qing-Yu, 2011. Fast moving small target tracking based on local background Gaussian mixture model. *Inform. Technol. J.*, 10: 1381-1387.
- Kaawaase, K.S., F. Chi, J. Shuhong and Q.B. Ji, 2011. A review on selected target tracking algorithms. *Inform. Technol. J.*, 10: 691-702.
- Logashanmugam, E. and R. Ramachandran, 2008. An improved algorithm for image compression using wavelets and contrast based quantization technique. *Inform. Technol. J.*, 7: 180-184.
- Loum, G., C.T. Haba, J. Lemoine and P. Provent, 2007. Texture characterisation and classification using full wavelet decomposition. *J. Applied Sci.*, 7: 1566-1573.
- Mahadevan, V. and N. Vasconcelos, 2008. Background subtraction in highly dynamic scenes. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, June 23-24, 2008, Anchorage, AK USA., pp: 1-6.
- Mahadevan, V. and N. Vasconcelos, 2010. Spatiotemporal Saliency in Dynamic Scenes. *Pattern analysis and machine intelligence. IEEE Trans. Pattern Anal. Mach. Intell.*, 32: 171-177.

- Maliani, A.D.E., M.E. Hassouni, N. Lasmar and Y. Berthoumieu, 2010. Texture classification based on the Generalized  $\Gamma$  distribution and the dual tree complex wavelet transform. Proceedings of the 5th International Communications and Mobile Network, September 30-October 2, 2010, Rabat, Morocco, pp: 1-4.
- Nagarajan, B. and P. Balasubramanie, 2008. Neural classifier for object classification with cluttered background using spectral texture based features. *J. Artif. Intell.*, 1: 61-69.
- Song, K.S., 2008. Globally convergent algorithms for estimating generalized  $\Gamma$  distributions in fast signal and image processing. *IEEE Tran. Image Process.*, 17: 1233-1250.
- Stacy, E.W., 1962. A generalization of the  $\Gamma$  distribution. *Ann. Math. Statist.*, 33: 1187-1192.
- Sulaiman, S., A. Hussain, N. Tahir, S.A. Samad and M.M. Mustafa, 2008. Human silhouette extraction using background modeling and subtraction techniques. *Inform. Technol. J.*, 7: 155-159.
- Van de Wouwer, G., P. Scheunders and D.P. Van Dyck, 1999.. Statistical texture characterization from discrete wavelet representations. *IEEE Trans. Image Process.*, 8: 592-598.
- Venkateswaran, N. and Y.V.R. Rao, 2007. K-means clustering based image compression in wavelet domain. *Inform. Technol. J.*, 6: 148-153.
- Zhang, M., Z. Lu and J. Shen, 2008. Salient region: Presentations of image main contents and its exaction algorithms. *Inform. Technol. J.*, 7: 992-1000.