

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

# INFORMATION TECHNOLOGY JOURNAL

**ANSI***net*

Asian Network for Scientific Information  
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

## Video Forgery Detection Based on Non-Subsampled Contourlet Transform and Gradient Information

Richao Chen, Qiong Dong, Heng Ren and Jiaqi Fu  
School of Information Science and Engineering, Hunan University, No. 252,  
Lushan South Road, Changsha, 410082, China

---

**Abstract:** For digital video, object-based manipulations, such as adding, removing or changing objects, are usually malicious tamper/forgery operations. Compared with the conventional double compression or frame-based operation, it makes more sense to detect these object-based manipulations because they might directly affect video content. This paper concentrates on video object contour and its Adjustable Width Object Boundary (AWOB), digs the trace of forgery in small-scale by analysing detail coefficients of Non-Subsampled Contourlet (NSCT) and gradient information, of which feature vectors are obtained and combined as the input of Support Vector Machine (SVM), thus natural objects and forged ones will be successfully classified. The proposed algorithm turns out to be effective with a high accuracy of correct detection up to 95%.

**Key words:** Video forgery, object detection, digital morphology, non-subsampled contourlet transform, average gradient, support vector machine

---

### INTRODUCTION

In the era of digital multimedia, the development of image/video editing tools makes the tampering or forgery of digital media much easier. Even ordinary users can produce forged digital media and spread them over internet maliciously. This leads to increasing concerns about the trustworthiness of public digital media. It is in urgent need to verify the originality and integrity of digital video. In response to this challenge, digital media forensics (Bing *et al.*, 2011) occurs. The techniques for digital forensics can be divided into two categories: active and passive forensic. Since active forensics depends on the auxiliary data such as digital watermark or signatures (Jin and peng, 2006), passive forensics is becoming a hot research topic in the field of information security (Abdulfetah *et al.*, 2010).

In the domain of digital image forensics, a mass of relative works have been conducted. In general, the tampering of digital video is more sophisticated and time-consuming. With an increasing number of video editing tools, such as Video Edit Magic, it becomes easier for video tampering. Correspondently, the research on digital video forensics is still in its infancy. The most representative works are summarized as follows: (1) forensics by the in-consistent trails during the imaging process of digital video, (2) forensics by the traces of video forgery process, such as ghost shadow (Zhang *et al.*, 2009), block artefacts (Abboud, 2006; Luo *et al.*, 2008), noise residue (Shitong *et al.*, 2005; Kobayashi *et al.*, 2010), similarity between image regions

(Wang *et al.*, 2007) and GOP periodicity (Su *et al.*, 2009) etc. They are effective to detect those traditional forgery operations, including copy-paste, double compression and frame-based manipulation.

Object-based manipulations are usually malicious for digital video. For example, if an object is added, or deleted from digital video, it might have direct influence to the viewer's understanding of video content (Renuga and Sadasivam, 2009). To the best of knowledge, there is still no work reported in literatures about the passive forensics of object-based forgery in digital video (Xiang-Wei *et al.*, 2009).

In this study, a forgery passive forensic method based on video object is proposed, which analyses statistical property of object and variable width of its surrounding area, then digs the trace of forgery in small-scale by extracting non-subsampled contourlet coefficients and gradient information, which will effectively exploits the image edges localized in both location and direction while producing significant coefficients and these features vectors are formed as input of Support Vector Machine (SVM), then natural objects and forged ones will be successfully classified, the procedure of proposed method is listed in Fig. 1.

### VIDEO OBJECT DETECTION

Usually, video forgery based on object consists of four steps: object detection, object manipulation, motion interpolation and background in-painting. For digital

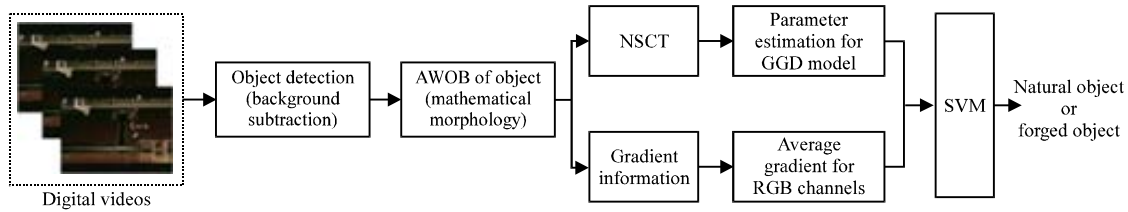


Fig. 1: The procedure of proposed method

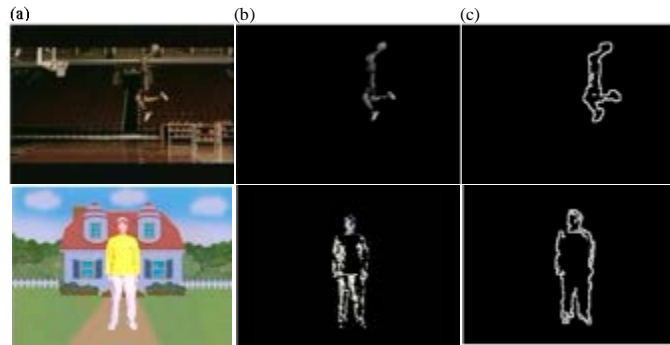


Fig. 2(a-c): Detection of video object and AWOB, (a) Original frame, (b) Background subtraction and (C) Contour of AWOB

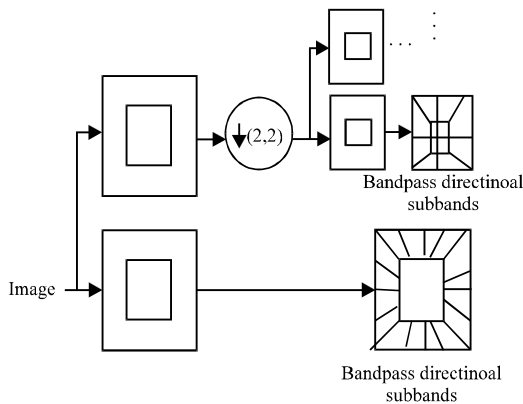


Fig. 3: Filter bank structure of contourlet transform

video forensics, the first step will also be object detection. Then, the object contour and its bounding area can be located and the statistical features are extracted in order to verify the originality and integrity of digital video.

For moving object detection, the three typical methods are: optical flow (Yao *et al.*, 2012), frame difference (Xue *et al.*, 2011) and background subtraction (Sulaiman *et al.*, 2008; Jodoin *et al.*, 2007). Optical flow method can get accurate results when tracking fast-moving object but with complex computation. Frame difference method is computationally efficient but too sensitive to the scene change such as

light. It is easy to lead to ghost effect. Therefore, it is relatively reasonable to utilize background subtraction, especially for the motion object detection from static background.

Since the left traces for object-based video forgery always exist near the object boundary, a new concept of Adjustable Width Object Boundary (AWOB) is introduced by mathematical morphology. Let  $l$  be the extracted binary object,  $\oplus$  be the dilation operation and  $\delta_s$  be the symmetric structure element, AWOB is defined as follows:

$$AWOB = \delta_s \oplus (\dots \delta_s \oplus (l)) \quad (1)$$

From Eq. 1, it is obvious that the object area gets larger with the increase of  $n$  for dilation. In Fig. 2, two examples are given for the above-mentioned steps in which object of original frames is obtained by background subtraction and subsequently the contour of its AWOB, where  $n$  equals 2.

### NON-SUBSAMPLED CONTOURLET TRANSFORM

The contourlet transform (Xin *et al.*, 2011) is a new two-dimensional extension of the wavelet transform using multi-scale and directional filter banks. The contourlet expansion is composed of basis images oriented at various directions in multiple scales, with flexible aspect

ratios. Figure 3 shows a flow graph of the contourlet transform. It consists of two major stages. The first stage is Laplace Pyramid (LP) and the second stage is Directional Filter Banks (DFB).

Figure 4 shows an example of the frequency decomposition achieved by the DFB. It offers a flexible multi-resolution and directional decomposition for images, since it allows for different number of directions at each scale. For the contourlet transform to satisfy the anisotropy scaling law, as in the curvelet transform, the number of directions is doubled at every other finer scale of the pyramid (Do and Vetterli, 2002).

Figure 5 shows an example of the contourlet transform on the Zoneplate image. The image is decomposed into a lowpass subband and several bandpass directional subbands.

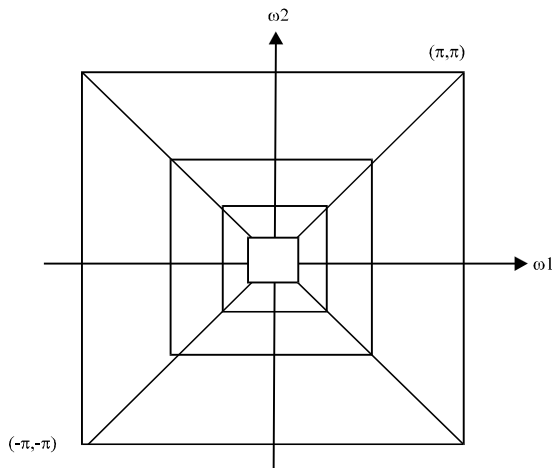


Fig. 4: Frequency partitioning by contourlet transform

It is noticeable that only contourlets that match with both location and direction of image contours produce significant coefficients. Thus, the contourlet transform effectively exploits the fact image edges are localized in both location and direction.

### PARAMETER ESTIMATION OF GENERAL GAUSSIAN DISTRIBUTION

In this study, the distribution of detail subband coefficients is described by General Gaussian Distribution model, whose model parameters are estimated by Maximum Likelihood Estimate (MLE) (Lee, 2010).

For the given subband coefficients of contourlet transform ( $N$  indicates the number of coefficients), the likelihood function having independent component is defined as:

$$L(x; \alpha, \beta) = \log \prod_{i=1}^N p_m(x_i; \alpha, \beta) \quad (2)$$

where,  $\alpha$  and  $\beta$  are parameters to be estimated, in this case the following likelihood equations have a unique root in probability, which is indeed the maximum likelihood estimator:

$$\frac{\partial L(x; \alpha, \beta)}{\partial \alpha} = -\frac{N}{\alpha} + \sum_{i=1}^N \frac{\beta |x_i|^\beta \alpha^{-\beta}}{\alpha} = 0 \quad (3)$$

$$\frac{\partial L(x; \alpha, \beta)}{\partial \beta} = \frac{N}{\beta} + \frac{N \Psi(1/\beta)}{\beta^2} - \sum_{i=1}^N \left[ \frac{|x_i|^\beta}{\alpha} \right] \log \left[ \frac{|x_i|}{\alpha} \right] = 0 \quad (4)$$

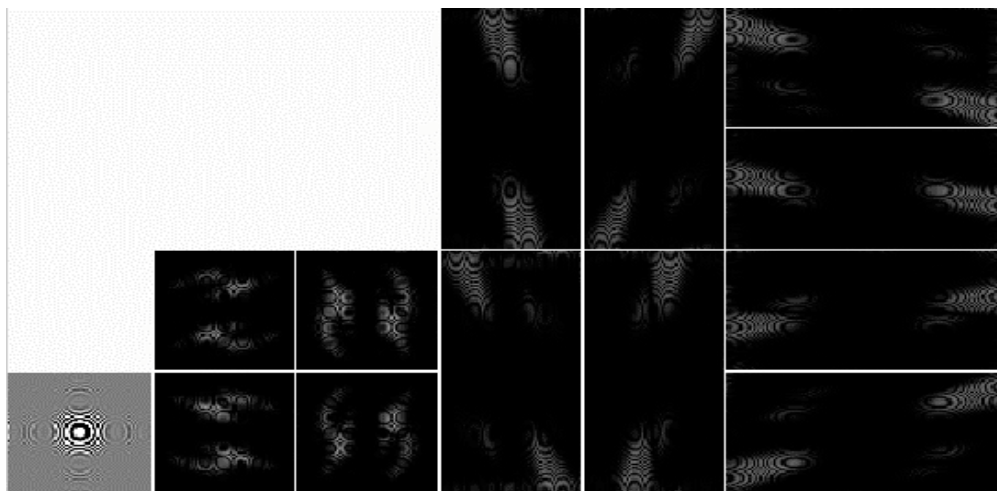


Fig. 5: Contourlet transform of the zoneplate image

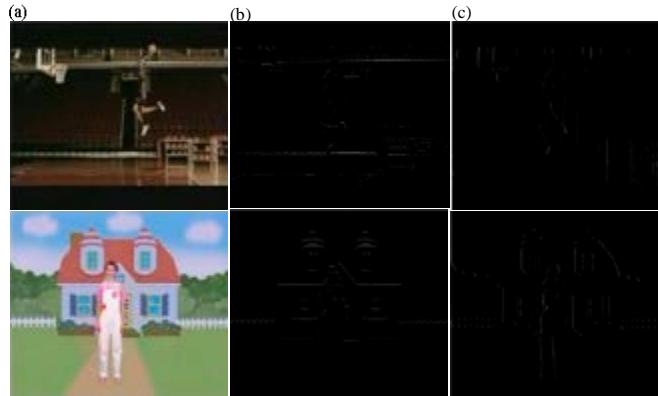


Fig. 6(a-c): Gradients in X-direction and Y-direction, (a) Original frame, (b) Gradients in X-direction and (c) Gradients in Y-direction

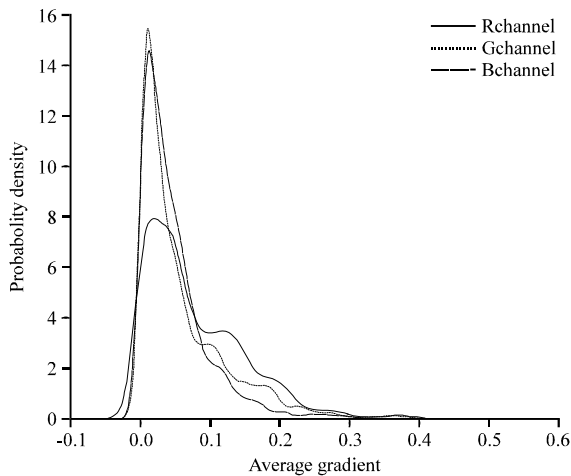


Fig. 7: Average gradients of RGB colour channel

where,  $\Psi$  is the digamma function, i.e.,  $\Psi(z) = \Gamma'(z)$  for  $\beta > 0$  then (3) has a unique, real and positive solution as:

$$\hat{\alpha} = \left( \frac{\beta}{N} \sum_{i=1}^N |x_i|^\beta \right)^{1/\beta} \quad (5)$$

Substitute this into Eq. 4, the shape parameter is the solution of the following transcendental equation:

$$1 + \frac{\Psi(1/\hat{\beta})}{\hat{\beta}} - \frac{\sum_{i=1}^N x_i^\beta \log|x_i|}{\sum_{i=1}^N x_i^\beta} + \frac{\log\left(\frac{\beta}{N} \sum_{i=1}^N |x_i|^\beta\right)}{\hat{\beta}} = 0 \quad (6)$$

Which can be solved numerically, maximum likelihood  $\hat{\beta}$  of shape parameter  $\beta$  is figured out by using the Newton-Raphson iterative procedure with the initial guess from the moment method.

### GRADIENT INFORMATION

Average gradient is the accumulation of image local luminance contrast, generally, the larger gradient value is, the stronger contrast among local parts exists and the clearer the image is, in a word, average gradient sensitively reflect the details of image. The rows and columns of the image  $f(x,y)$  are described as row and col and the expression is:

$$\bar{g} = \frac{1}{(\text{row}-1)(\text{col}-1)} \times \sum_{x=1}^{\text{row}-1} \sum_{y=1}^{\text{col}-1} \sqrt{\frac{\partial f(x,y)^2}{\partial x} + \frac{\partial f(x,y)^2}{\partial y}} \quad (7)$$

Figure 6 presents gradients of the frames both in X-direction and Y-direction. Original frames are listed in column (a), with their gradients in X-direction showed in column (b) and Y-direction showed in column (c). It is apparent that accumulation of local luminance contrast successfully reflects the details of frames.

Let  $z = (z_1, \dots, z_k)$  ( $k$  indicates the number of gradient values) indicate the gradient components of AWOB.

As showed in Fig. 7, average gradients of RGB colour channel are computed, respectively to reflect edge intensities, which present the same trend. It could be seen that edge intensities are fitted into Rayleigh distribution (Jing *et al.*, 2011), probability function of which is described in formula (8):

$$f(z) = \frac{z}{\sigma^2} \exp\left(-\frac{z^2}{2\sigma^2}\right), z \geq 0 \quad (8)$$

And the parameters of Rayleigh distribution are estimated by MLE:

$$\delta = \sqrt{\frac{\sum_{i=1}^k z_i^2}{2k}} \quad (9)$$

### THE CLASSIFICATION BASED ON SVM

Since SVM (Hung and Liao, 2008) is a kind of machine learning algorithm based on statistical theory, it has special advantages in solving small samples, non-linear and high dimensional pattern recognition problem. The basic idea behind SVM is to map sample space into a high or infinite dimension space (Hilbert space) by nonlinear mapping  $p$ . The non-linear problem in the original sample space will be converted into a linear problem in feature space.

Assuming that those videos with natural objects are positive samples, and video sequences with object-based forgery are negative samples, there will be four categories of possible judgments for a two-class (positive and negative) classification problem: (1) True Positive (TP): positive samples are predicted as positive; (2) False Negative (FN): positive samples are predicted as negative; (3) False Positive (FP): negative samples are predicted as positive; (4) True Negative (TN): negative samples are predicted as negative. The performance of classifier can be described by the following metrics, such as Accuracy, Precision and Recall. Their definitions are as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (10)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (12)$$

Two methods are used for the performance evaluation of classifier. One is True Positive Rate (TPR) and False Positive Rate (FPR) and the other is ROC (Receiver Operating Characteristic), which describes the trade-off between TP and FP. The AUC (Area Under ROC Curve) can reflect how well classifier works. The larger AUC is, the better performance the classifier has:

$$\text{TPR} = \text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

$$\text{FPR} = \frac{FP}{FP + TN} \quad (14)$$

ROC curve is utilized to describe the trade-off between TP and FP, which shows the performance of

classifier and AUC marks how classifier works, the larger value is the better performance the classifier will demonstrate.

### EXPERIMENTAL RESULTS AND DISCUSSIONS

The test videos are directly taken from Shih *et al.* (2006) without processing. Small sample data set are obtained by the object-based forgery and manipulation algorithm. 9 typical static video sequences are chosen. Among them, four are natural videos and the rest five samples are forged. Their compressed file formats are AVI and WMV, respectively. The resolution of video frame is 320\*240. The experiment is performed with Matlab R2009a. The PC configurations are as follows: Inter (R) Core (TM) i3 CPU, 2.53 GHz, RAM 2 GB. Fifty video frames are selected from every video sequence and given the frame image, as shown in Fig. 1, details of the forensics scheme is described as follows:

- Step 1:** Video object of original frames is obtained by background subtraction and so is the contour of its AWOB, which will be the target of the experiments
- Step 2:** Detail subband coefficients of the contourlet transform are computed by 3 layer decomposition of NSCT, and fitted into General Gaussian Distribution model, then the corresponding parameters are estimated by MLE
- Step 3:** Edge intensities are analysed, after fitting into Rayleigh distribution, the parameters are estimated for RGB channels, respectively, which will be the next 3 feature vectors
- Step 4:** Different weights are assigned with these features to form higher dimensional vector, which are obtained by training with step size gradually increased from 0.5 to 0.9, Each group of weights are utilized for feature combination (Sasikala and Kumaravel, 2007) and the group with the highest accuracy of classification is selected
- Step 5:** Every component in the feature vector is normalized into (0, 1) on the basis of the following mapping formula so that every component will play a balanced role:

$$f; x \rightarrow x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (15)$$

where  $x$ ,  $x'$  represent sample features before and after normalization, respectively.

Table 1: Classification results of extracted features

Features	Accuracy Precision		TPR (%)	FPR (%)	AUC (%)
	(%)	(%)			
Mean of NSCT coefficients	88.73	84.85	90.32	16.13	0.9194
Variance of NSCT coefficients	94.37	93.55	93.46	6.46	0.9742
R channel average gradient	98.04	96.30	97.27	2.79	0.9947
G channel average gradient	99.13	98.15	99.09	0.83	0.9997
B channel average gradient	94.71	94.55	91.82	3.57	0.9921
Features combination	96.83	96.28	96.43	1.18	0.9956

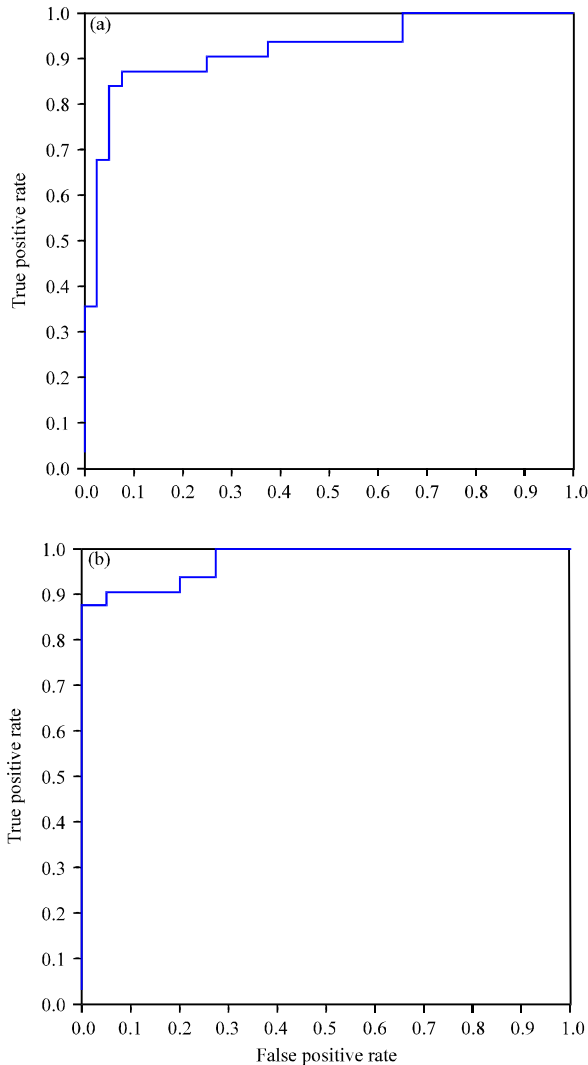


Fig. 8(a-b): ROC curves of different circumstances, (a) AUC = 0.91935 and (b) AUC = 0.97419

**Step 6:** The library libsvm is utilized, where Radial Basis Function (RBF) is adopted as the kernel function. By cross validation, the penalty parameter  $c$  and RBF kernel function parameter  $g$  are obtained. To achieve better classification, the above processes are repeated for 10 times to obtain the average results

The classification performances of five features and their combination are summarized in Table 1. From it, it is apparent that various features show different effects on the classification of real videos and forged ones, with the highest accuracy of 99.13% (result of G channel average gradient) and the lowest accuracy of 88.73% (result of means of NSCT coefficients) though they all have accuracies above 90%. Moreover, gradient features are relatively more effective than those NSCT coefficients features. After feature combination, both Accuracy and Precision reach stable result for over 96%.

Figure 8 contains the ROC curves when the classification accuracies are 88.73 and 94.37%, respectively. Apparently, the right figure shows better performance because ROC is closer to the upper left of the coordinate, the results of AUC are 0.9194 on the left side and 0.9742 on the right.

**CONCLUSION**

A passive forensic method for object-based video forgery is proposed, which utilizes the statistical features of object contour and its surrounding area. By using non-subsampled contourlet transform coefficients in small-scale and gradient information of each colour channel, a classification of natural objects and forged ones is achieved by SVM. The experimental results show that the proposed approach achieves desirable forensics results. However, due to the diversity and content complexity of digital videos, it still needs the support of the sample database in the training process. Future investigation will be searching for features that not too heavily-dependent on the training samples. Furthermore, a comprehensive data set similar to Columbia Image Splicing Detection Evaluation Dataset (DVMM, 2008) is also needed for video forensic.

**ACKNOWLEDGMENTS**

This work was supported by National Natural Science Foundation of China (Grant No. 61072122), Special Pro-phase Project on National Basic Research Program of China (Grant No. 2010CB334706) and Program for New Century Excellent Talents in University (Grant No. NCET-11-0134).

**REFERENCES**

Abboud, I., 2006. Deblocking in BDCT image and video coding using a simple and effective method. Inform. Technol. J., 5: 422-426.

- Abdulfetah, A.A., X. Sun and H. Yang, 2010. Robust adaptive video watermarking scheme using visual models in DWT domain. *Inform. Technol. J.*, 9: 1409-1414.
- Bing, C.W., Y.G. Bo and C. Richao, 2011. Digital video passive forensics for authenticity and source. *J. Commun.*, 6: 177-183.
- DVMM, 2008. Columbia image splicing detection evaluation dataset. The DVMM Laboratory of Columbia University. <http://www.ee.columbia.edu/ln/dvmm/downloads/AuthSplicedDataSet/AuthSplicedDataSet.htm>.
- Do, M.N and M. Vetterli, 2002. Contourlets: A directional multiresolution image representation. *Image Process.*, 1: I-357-I-360.
- Hung, Y.H. and Y.S. Liao, 2008. Applying PCA and fixed size LS-SVM method for large scale classification problems. *Inform. Technol. J.*, 7: 890-896.
- Jin, C. and J. Peng, 2006. A robust wavelet-based blind digital watermarking algorithm. *Inform. Technol. J.*, 5: 358-363.
- Jing, W.L., R.H. Wang and X.L. Xu, 2011. The statistical analysis of rayleigh distribution under the full sample. *Sci. Technol. Eng.*, 11: 2551-2553.
- Jodoin, P., M. Mignotte and J. Konrad, 2007. Statistical background subtraction methods using spatial cues. *Circuits Syst. Video Technol.*, 17: 1758-1764.
- Kobayashi, M., T. Okabe and Y. Sato, 2010. Detecting forgery from static-scene video based on inconsistency in noise level functions. *Forensics Secur.*, 5: 883-892.
- Lee, J.Y., 2010. Parameter estimation of the extended generalized gaussian family distributions using maximum likelihood scheme. *Inform. Technol. J.*, 9: 61-66.
- Luo, W.Q., M. Wu and J.W. Huang, 2008. MPEG recompression detection based on block artifacts. *Proceedings of the SPIE on Security, Forensics, Steganography and Watermarking of Multimedia Contents X*, January 28-30, 2008, San Jose, CA., USA., pp: 1-12.
- Renuga, R. and S. Sadasivam, 2009. Data discovery in grid using content based searching technique. *Inform. Technol. J.*, 8: 71-76.
- Sasikala, M. and N. Kumaravel, 2007. A comparative analysis of feature based image fusion methods. *Inform. Technol. J.*, 6: 1224-1230.
- Shih, T.K., N.C. Tang, W.S. Yeh, T.J. Chen and W. Lee, 2006. Video inpainting and implant via diversified temporal continuations. *Proceedings of the 14th Annual ACM International Conference on Multimedia*, October 23-27, 2006, Santa Barbara, CA., USA., pp: 133-136.
- Shitong, W., L. Yueyang, F.L. Chung and C. Shu, 2005. Iterative self-adaptive filtering algorithm for reducing impulsive noise in color images. *Inform. Technol. J.*, 4: 456-461.
- Su, Y.T., J. Zhang and J. Liu, 2009. Exposing digital video forgery by detecting motion-compensated edge artifact. *Proceedings of the International Conference on Intelligence and Software Engineering*, December 11-13, 2009, Wuhan, China, pp: 1-4.
- 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.
- Wang, W.H. and H. Farid, 2007. Exposing digital forgeries in video by detecting duplication. *Proceedings of ACM Multimedia and Security Workshop*, September 20-21, 2007, Dallas, TX., USA., pp: 35-42.
- Xiang-Wei, L., Z. Ming-Xin, Z. Geng-Lie, Z. Ya-Lin and Z. Shuang-Ping, 2009. Effective video analysis preprocessing algorithm based on rough sets in compressed domain. *Res. J. Inform. Technol.*, 1: 51-56.
- Xin, G., B. Zou, J. Li and Y. Liang, 2011. Multi-focus image fusion based on the nonsubsampling contourlet transform and dual-layer PCNN model. *Inform. Technol. J.*, 10: 1138-1149.
- Xue, L.X., Y.L. Luo and Z.C. Wang, 2011. Detection algorithm of adaptive moving objects based on frame difference method. *Appl. Res. Comput.*, 28: 1551-1553.
- Yao, L., D.X. Li and M. Zhang, 2012. Temporally consistent depth maps recovery from stereo videos. *Inform. Technol. J.*, 11: 30-39.
- Zhang, J., T. Suy and M.Y. Zhang, 2009. Exposing digital video forgery by ghost shadow artifact. *Proceedings of the 1st ACM Workshop on Multimedia in Forensics*, October 19-24, 2009, Beijing, China, pp: 49-54.