USER-GENERATED PORNOGRAPHIC VIDEO DETECTION USING SHOT-BASED SENSOR PATTERN NOISE

Dae-Jin Jung*, Dai-Kyung Hyun*, Seung-Jin Ryu* and Heung-Kyu Lee**
Department of Computer Science* and Division of Web Science and Technology**
Korea Advanced Institute of Science and Technology
291 Daehak-ro, Yuseong-gu
Daejeon, Republic of Korea
{djjung, dkhyun, sjryu, hklee}@mmc.kaist.ac.kr

ABSTRACT
Illegal distribution of user-generated pornography (UGP) videos raises lots of negative aspects associated with the digital contents. This paper proposes a first forensic technique to detect UGP video. To discriminate UGP videos, we exploit shot based sensor pattern noise (SPN) from an image under investigation. By thresholding average peak-to-correlation (PCE) values from the shot based SPNs, the video is decided as a UGP video. Preliminary experiments, which consist of UGP videos, normal videos, and Japanese commercial pornographic videos, indicate adequate performance of the proposed technique.

KEY WORDS
pornography detection, pornographic video detection, shot-based SPN, sensor pattern noise

1 Introduction

With the advent of highly sophisticated IT technology, various types of digital contents are generated and shared through the Internet. These digital contents contribute to the development of various industries such as education, broadcasting, movie, and so on. At the same time, the very nature of digital contents which can be easily generated and distributed raises lots of negative aspects associated with the digital contents. One of the most significant problems is indiscriminate and rapid distribution of illegal pornographic videos. The pornographic contents can encourage wrong cognition for sex and the copycat sexual crimes. Furthermore, disclosed privacy problems can be caused as well by illegally distributed pornographic videos.

Of course, the distribution of pornographic contents is just not a recent problem. In India, Kama Sutra (which is hardly classified as a pornographic content but a sex manual) was composed between 400 BCE and 200 CE. In Japan, wood sculptures were made to describe the sexual intercourses from 1600 CE. In those ages, the amount of pornographic contents was not too much so that it was hard to distribute or access pornographic contents. Although the production of analog pornographic videos causes various issues, the range of dissemination was still not wide. In 1990s, the advent of the Internet accelerated the distribution of digitalized pornographic contents. Nowadays, with diverse and real-time distributing path, various types of pornographic contents, including both legal and illegal, are distributed so that lots of social issues are daily reported.

Among them, the distribution of user-generated pornography (UGP) videos causes more harmful effects on society. The UGP videos inherently include a privacy problem. Moreover, most of child pornography is user-generated. Although the importance of detecting the user-generated pornography is increased, to the best of our knowledge, researches focused onto the UGP rarely exist.

Most works related to detection of pornographic contents have concentrated on the image domain. It is reasonable because a pornographic image contains a lot of skin-colored pixels. Therefore, several researches approached the problem based on skin color model. Jones and Regh computed five different features based on their skin color histogram model [1]. Rowley *et al.* generated both a skin color map for the image and a connected components structure to extract 19 different features for classification [2]. To decrease the false positive ratio caused by close-up face images, Choi *et al.* used MPEG-7 visual descriptors [3]. Another researches focus on the shape of images. They also rely on skin-colored pixels to a certain degree. Hu *et al.* obtained an outline of the largest skin-colored object by searching skin-colored blocks for feature extraction. Bosson *et al.* found initial skin segments from a HSV histogram, and then expanded them to have a large skin region. [4]. Lopes *et al.* considered SIFT key-points as a local feature descriptor [5]. Most of the methods trained and tested support vector machine for classification.

Many works to detect pornographic video were also proposed. Kim *et al.* filtered obscene frames by revealing motion, color, texture, and shape [6]. Then extract skin region and match them to samples in a database. Rea *et al.*proposed to use both visual information and audio information [7]. In their method, motion and sound information were extracted from the segment of skin-region and periodic patterns in audio stream, respectively. . Motion information is extracted for skin-region segmentation and periodic patterns in audio stream are analyzed. Tong *et al.* adopted periodicity of motion vector to classify pornographic videos [8] and Jansohn *et al.* combined image features with motion information by their classification scores [9]. Nevertheless, these works are not adequate to discriminate UGP videos because their target is just pornography videos.

In this paper, we propose an UGP video detection method based on shot-based sensor pattern noise (SPN). SPN is utilized as a statistical feature because it appropriately uncovers the properties of UGP video (See Sec. 2 for more detailed properties.) By adopting the shot based SPN, the proposed method is able to discriminate UGP videos from normal videos.

The rest of this paper is structured as follow. The differences between normal videos and pornographic videos are analyzed in Sec. 2. The detail of the proposed method is explained in Sec. 3. Experimental results are exhibited in Sec. 4 and Sec.5 concludes.

2 Differences between normal and user-generated pornographic videos

In this section, we describe the differences between normal videos (movies, TV shows, commercial pornographic videos, etc.) and UGP videos. The differences are caused by following factors:

- **The number and the length of shots:** In filmmaking, a shot is a series of frames, which runs for an uninterrupted period of time. Shots are filmed (recorded) with a single camera and can be any duration. For smooth story-telling and causing extreme concentration, frequent shot changes are essential in normal videos. In normal videos, frequent shot changes occur so that the number of frames in each shot is small. In UGP videos, shot is rarely changed and correspondingly enormous frames exist in each shot. The purpose of the pornographic videos is describing sexual intercourses in detail, instead of storytelling, so that they do not need many shot changes.

- **Recording devices:** Usually normal videos are recorded by both analog cameras and digital camcorders. Although digital camcorders can provide many benefits such as cheap price including reducing analog film cost, easiness of editing & CG processing, various built-in functions, and etc., analog film cameras are dominating many industries because of good characteristics of analog film (e. g. high resolution, soft shades of color, and etc.). On the contrary, UGP videos are normally recorded by digital camcorders. Most people do not have enough budgets to buy expensive analog film cameras. In contrast, many digitally recordable devices (e. g. digital camcorder, smartphones, etc.) are cheap and easily usable. Therefore, illegal pornographic videos such as hidden camera pornographic videos or amateur pornographic videos are recorded and distributed by digital camcorders.

- **The number of cameras used in recording:** Creating normal videos, including commercial pornography, need multiple cameras for recording various scenes and shots. As a result, each shot from the normal videos would be recorded by different cameras with high probability. On the other hand, user generated pornographic videos are normally generated with only a single digital camcorder or smart phone as we stated above.

- **Post-processing:** Normal videos undergo huge amount of post-processing for inserting CGs or harmonizing each shot from different cameras. On the contrary, user-generated pornographic videos hardly undergo complex post-processing except simpler post-processing such as recompression or resizing.

From the above observation, we can expect that SPNs of normal and UGP pornographic videos would be different. The SPN (also known as PRNU) is pixel variation under illumination. It was first proposed by Lukas *et al.* to identify the source digital camera [10]. Every digital camera has a charge coupled device (CCD) or a complementary metal-oxide-semiconductor sensor (CMOS), and the sensor translate the photoelectron to an electrical signal using photoelectric effect. The power of the electrical signal is affected by sensitivity to the light of the sensor. However, sensor imperfection which is introduced in manufacturing process makes the sensor have different sensitivities pixel by pixel. This causes SPN in recorded images and video frames. Therefore, the SPNs can be used to identify source digital cameras or camcorders. By adopting the above enumerated factors and the characteristics of SPN, we finally infer specific properties for the UGP video detection as follows:

- The shot-based SPN estimated from normal videos would not be correlated with each other. First, SPN estimated from analog film videos does not have unique pattern because analog film camera does not use any digital sensors. Thus, SPN estimated from analog film videos cannot be used for identifying source analog film cameras. Second, different source cameras used in normal videos generate different SPNs. Thus, SPNs estimated from each shot would be uncorrelated with the others with high probability. Although shots are fully recorded by digital camcorders, each shot have different source digital camcorders. It results in low correlation between each shot-based SPN. Third, even if small number (2 or 3) of digital camcorders are used in recording shots in normal videos, heavy post processing affect the quality of the estimated SPN. The compromised shot-based SPN cannot be correlated with other shot-based SPN. Finally, the small number of frames in a shot is not enough to estimate accurate SPN. This results in damaged SPN as well.

- On the contrary, shot-based SPN estimated from a part of UGP video is correlated with the other SPN with high probability because every shot is recorded by a single digital camcorder (including smartphone). Moreover, the quality of SPN from UGP videos is relatively higher than those of normal videos. They are rarely affected by post-processing. In addition, they are constructed with the large number of frames in each shot.

Therefore, by exploiting these properties, we can discriminate UGP videos from normal videos.

3 Proposed Method

We propose a method that discriminates UGP videos from normal videos. Fig. 1 depicts the process of the proposed method. The proposed method consists of three steps. First, a target video is divided into many shots in the shot-division step. In the SPN estimation step, shot-based SPN is estimated from the divided shots. After that, peak-to-correlation energy (PCE) values between shot-based SPN. By examining PCE values calculated from each shot-based SPN, we can discriminate UGP videos from normal videos in the decision step.

3.1 Shot-division Step

First, a video in investigation is divided into multiple shots by utilizing a proper shot change detector. The shots which have information about only a single source camera are required. If we fail to obtain them, several problems appear. If many shots are declared as a single shot by a shot change detector, SPN estimated from that shot is not pure SPN from a single source digital camera but damaged or mixed SPN from plural cameras. It causes the increase of false positive rate in SPN comparison. On the other hand, if a shot is declared as many shots, it can give meaningless duplicated SPN and it turn the original shot into useless shots for SPN estimation.

Numerous shot change detection methods have been reported. Among them, a histogram comparison method is adopted in our method because it is of good performance and relatively fast [11]. Let $H_i(j)$ denotes a histogram value for ith frame, where j is one of G possible gray levels and SD_i is the sum of absolute differences of histogram values between ith frame and $(i+1)$th frame. Then the sum of absolute differences of histogram values, SD_i, is given by the following formula:

$$SD_i = \sum_{j=1}^{G} |H_i(j) - H_{i+1}(j)| \tag{1}$$

SD_i cannot be used directly for shot change detection because SD_i would be greater when the frame size grows. Thus, we use NSD_i in our method that is the normalized version of SD_i by frame size. The shot change is declared when NSD_i is larger than a given threshold. If kth shot change is declared, the successive frames between $(k-1)$th declared frame and kth declared frame are considered as kth shot.

3.2 SPN Estimation Step

SPN had been proposed to identify the source digital camera and after that, the source digital camcorder identification method using SPN was proposed [12]. SPN is considered as a fingerprint of a digital camera (camcorder). Shot-based SPN can be estimated from the frames of a shot. The SPN model for digital camcorders is as follow:

$$\mathbf{I} = g^{\gamma} \cdot [(\mathbf{1} + \mathbf{K})\mathbf{Y} + \mathbf{\Lambda} + \mathbf{\Theta}_s + \mathbf{\Theta}_r]^{\gamma} + \mathbf{\Theta}_q \tag{2}$$

where $\mathbf{I}$ denotes the sensor output compromised by numerous in-camcorder processing, g does the color channel gain, γ is the gamma correction factor, $\mathbf{K}$ is SPN multiplicative factor, Y is the light intensity, and $\mathbf{\Lambda}$, $\mathbf{\Theta}_s$, $\mathbf{\Theta}_r$, $\mathbf{\Theta}_q$ denote dark current, shot noise, read-out noise, and quantization noise, respectively. Using first order Taylor expansion, simple form of this model can be obtained:

$$\mathbf{I} = \mathbf{I}^{(0)} + \gamma\mathbf{I}^{(0)}\mathbf{K} + \mathbf{\Theta} \tag{3}$$

Here, $\mathbf{I}^{(0)}$ is the noise-free sensor output from one channel before demosaicing is applied. $\mathbf{\Theta}$ is a noise component including enumerated noises above.

We use simplified model in Eq (3) to estimate SPN from each shot. To obtain noise which has less influence of the noise-free frame $\mathbf{I}^{(0)}$, an estimate $\hat{\mathbf{I}}^{(0)}$ of $\mathbf{I}^{(0)}$ is subtracted from both sides of Eq (3). $\hat{\mathbf{I}}^{(0)}$ is estimated by using denoising filter which is a wavelet based filter [13].

$$\mathbf{W} = \mathbf{I} - \hat{\mathbf{I}}^{(0)} = \mathbf{IK} + (\mathbf{I}^{(0)} - \hat{\mathbf{I}}^{(0)}) + [(\mathbf{I}^{(0)} - \mathbf{I})\mathbf{K}] + \mathbf{\Theta} \tag{4}$$

PRNU factor $\mathbf{K}$ can be estimated by using Maximum Likelihood Estimation (MLE) method as follow:

$$\gamma\hat{\mathbf{K}} = \frac{\sum_{k=1}^{N} \mathbf{W}_k \hat{\mathbf{I}}_k^{(0)}}{\sum_{k=1}^{N} (\hat{\mathbf{I}}_k^{(0)})^2} \tag{5}$$

where $\mathbf{W}_k$ is noise residual of kth frame. Notice that equations in this section are element-wise.

In video compression process, frames are compromised by huge codec noise unlike image compression process. Those noises are block artifacts which are caused by DPCM-block DCT transform [12]. These block artifacts usually have higher power than SPN. They can cause false positive correlations in comparing SPN which are from different digital camcorders. The codec noise shall be suppressed for better performance. To do this, denoising filter is used. Wiener filter in frequency domain is adopted in our method to suppress the codec noise [14].

To decide whether two shots are recorded by same digital camcorder or not, shot-based SPN from those shots are tested by calculating PCE value. PCE is a robust measurement of how two quantities are correlated. To calculate

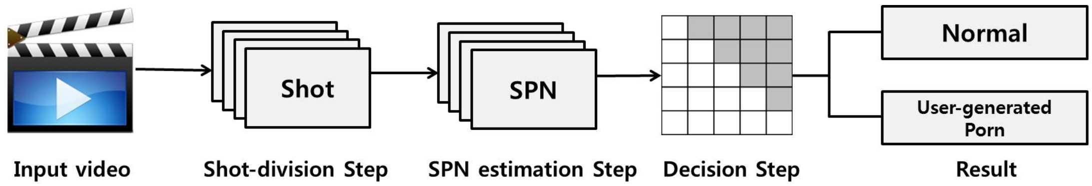


Figure 1. An overview of proposed UGP videos detection

PCE, we need to calculate normalized correlation first:

$$NCC[\mathbf{X},\mathbf{Y}] = \frac{(\mathbf{X}-\overline{\mathbf{X}}) * (\mathbf{Y}-\overline{\mathbf{Y}})}{\|\mathbf{X}-\overline{\mathbf{X}}\|\|\mathbf{Y}-\overline{\mathbf{Y}}\|} \quad (6)$$

where $\mathbf{X}$, $\mathbf{Y}$ are SPN, $\overline{\mathbf{X}}$ is mean of $\mathbf{X}$, $\mathbf{X} * \mathbf{Y}$ is dot product between $\mathbf{X}$ and $\mathbf{Y}$. $\|\mathbf{X}\|$ denotes the norm of $\mathbf{X}$. Finally, PCE value is calculated as follow [15]:

$$PCE[\mathbf{X},\mathbf{Y}] = \frac{|NCC[\mathbf{X},\mathbf{Y}](u_{peak}, v_{peak})|^2}{\mathbf{E}_{NCC[\mathbf{X},\mathbf{Y}]}} \quad (7)$$

where (u_{peak}, v_{peak}) denotes the peak location in the correlation plane $NCC[\mathbf{X},\mathbf{Y}]$ and $\mathbf{E}_{NCC[\mathbf{X},\mathbf{Y}]}$ does the correlation plane energy of $NCC[\mathbf{X},\mathbf{Y}]$. If we obtain higher PCE value than a given threshold from two shot-based SPN, we decide that those shots are recorded by same digital camcorder.

3.3 Decision Step

To decide whether a video under investigation is an UGP video or not, we investigate every shot-based SPN estimated from the video. First, an $N \times N$ PCE value matrix (M) is calculated to represent relationship between each shot. Let the X_i be the SPN estimated from ith shot, where $i = 1, ..., N$ and N is the number of shots divided by shot detector. The $(i, j)$the element of M which is $PCE[X_i, X_j]$ represents the relationship between shot i and shot j. If the value of $PCE[X_i, X_j]$ is high enough, those two shots (i, j) are considered that they are recorded by same digital camcorder. We do not need to calculate N^2 PCE values in matrix M, but $\frac{N*(N-1)}{2}$ PCE vales to fill the elements of upper triangular matrix (excluding diagonal elements) in M for avoiding duplications. We set a measurement as the average value of $\frac{N*(N-1)}{2}$ PCE values. If the average value is high enough, we declare the suspicious video as an UGP video.

4 Experimental Results

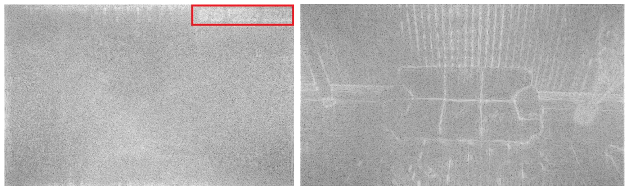
This section presents the performance of the proposed UGP video detector. We prepared 20 movies (including soap dramas), 20 commercial Japanese pornographic videos, and 50 UGP videos. 20 commercial Japanese pornographic videos were selected from top twenty ranked videos in DMM site (Japanese DVD rank site). In contrast, 50 UGP videos were collected by a random crawl over unprotected pornographic websites. The resolutions of normal and UGP videos varied from 720x480 to 1280x720 and 520x420 to 720x480, respectively. Although we could not find out the main recording cameras for 20 Japanese pornographic videos, the main cameras for remaining 20 movies from www.IMDB.com are depicted by 1. Seven normal videos, mainly recorded by digital cameras, are included for testing our assumptions in Sec. 2.

Movie	Main Camera (Digital/Analog)
1	Sony PMW-F3 (Digital)
2	Sony PMW-F3 (Digital)
3	Sony PMW-F3 (Digital)
4	Sony PMW-F3 (Digital)
5	Red One (Digital)
6	Red One (Digital)
7	Red One (Digital)
8	Panavision camera (Analog)
9	Panavision Panaflex Platinum (Analog)
10-20	unknown (unknown)

Table 1. Main camera names and their types which were used in recording normal videos.

In the shot-division step, each color frame was first converted into a gray frame to build a histogram. After that, inappropriate shots to estimate SPN were excluded. More specifically, shots with the small number of frames were excluded since the SPN estimated from those shots is unlikely to identify a specific digital camera. Therefore, the minimum number of frames in a shot was set as 200. If too small number of shots having enough number of frames were selected by shot change detector, they were divided into sub-shots.

With the detected shots, two types of possible false positive correlation in SPN comparison exist. The examples are shown in Fig. 2. Logos are usually positioned in same spot and leave strong edges on frames. These edges survive even after denoising process of SPN estimation step. As a result, they cause high correlation in SPN comparison even though shots are not recorded by the same digital camera. To avoid this problem, we cut-off 15 percent of top, bottom, left, and right regions because logos are usually position at those areas. Static shots (including



(a) (b)

Figure 2. Examples of false positive correlations : (a) Logos and advertisements (an advertisement in the example is located in the top-right box), (b) Successive static shots causing high energy edges

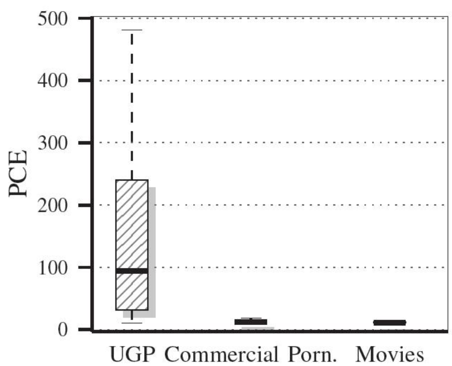


Figure 3. Boxplots of average PCE values for each set of videos

survived logos from previous processing) also cause false positive correlations. If the shooting angle of the camera does not change, recorded shots have similar backgrounds. They also survive against denoising process and cause false positive correlations by the strong energy of edges. To resolve this problem, too high PCE values were skipped in the decision step. We set the threshold for high PCE values as 1000 empirically.

Under the environment, we measured the performance of the proposed method. At first, we calculated average PCE values from each video. Boxplots in Fig. 3 presents the distribution of average PCE values. As shown in Fig. 3, UGP videos revealed higher average PCE values than the others. Table 2. depicts statistical values for average PCE values specifically. Higher mean and median values from UGP videos proved the discrimination performance of the proposed method. However, higher standard deviation and wide inter quartile range (IQR) of average PCE values from UGP videos also brought about false positive errors. In Fig. 4., we report the false positive error in detail. At first, discrimination between UGP and commercial pornographic videos were examined. Simultaneously, ROC curves from UGP and movies were generated. Experiments exhibited higher true positive rate, which decide UGP videos as UGP videos, against movies set when the false positive rate was set as zero. This is reasonable because commercial pornographic videos are generally recorded with static background and smaller number of camcorders.

	UGP	Commercial Porn	Movie
Mean	142.4	13.1	11.3
Std	146.9	3.0	0.6
Median	94.2	11.4	11.1
IQR	208.3	4.9	0.8

Table 2. Statistical values for average PCE values

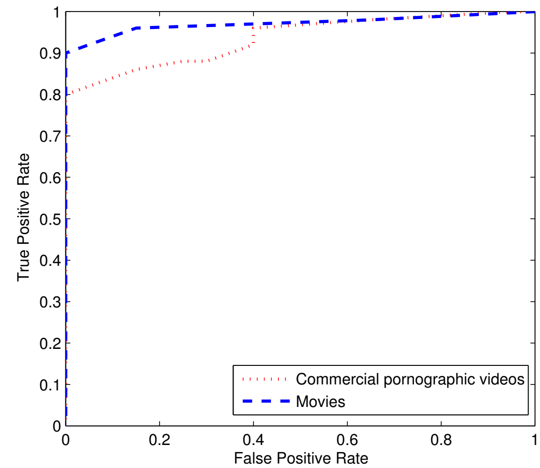


Figure 4. ROC curves : UGP and commercial pornographic videos test (blue dashed line), UGP and movies test (red dotted line)

We further analyze the reason of false negative rate. There might be several factors which affect to low average PCE values in UGP videos. First, several shots from UGP shots were recorded in dark areas. They are unlikely to contain SPN because SPN is generated by sensitivity against light. Second, several UGP videos were recorded by multiple cameras so that few shots were recorded from the same source camcorder. At last, several shot-based SPNs were compromised by heavy post-processing (e. g. censoring faces, advertisements insertion, and heavy recompression).

5 Conclusion

Forensic examination of pornographic videos encounters indiscriminate distribution of user generated pornography (UGP) videos. In this paper, we have investigated to automatically discriminate UGP videos from normal videos.

Our key contribution lies in the fact that the proposed method is the first approach to detect UGP videos and utilizing sensor pattern noise (SPN), the unique fingerprint of digital image sensor, in pornography detection. Our proposed method is based on sensor pattern noise (SPN) which is unique fingerprint of digital image sensors. The proposed method consists of 3 steps. First, a video under investigation is divided into many shots. Afterwards, SPN estimated from the divided shots. Finally, by averaging PCE values among the SPNs, the suspicious video is decided as a normal video or an UGP video. Experimental results show the performance of the proposed approach. About 90% and 80% of true positive detection rate (which appropriately detects UGP) without false positive error against movies and commercial pornographic videos, respectively, indicate the adequate performance of the proposed method. It is promising result considering the impact of UGP videos. Moreover, better performance is expected by combining existing video pornography detection methods. As to the limitations, we note that the proposed detector is weak against such as heavy re-compression. Suppressing shot-edges with high energy to utilize every unimpaired shot will be considered as well.

6 Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. 2012R1A2A1A05026327) and WCU (World Class University) program (Project No: R31-30007)

References

[1] Jones MJ, Rehg JM, Statistical color models with application to skin detection. *Int. J. Comput. Vision*, 46(1), 2002, 81-96.

[2] Rowley HA, Jing Y, Baluja S, Large scale image-based adult-content filtering. *Proceedings of the 1st international conference on computer vision theory*, 2006, 290-296.

[3] Choi B, Chung B, Ryou J, Adult image detection using bayesian decision rule weighted by svm probability. *Proceedings of the 2009 4th international conference on computer sciences and convergence information technology, ICCIT*, 2009, 659-662.

[4] Bosson A, Cawley G, Chan Y, Harvey R, Non-retrieval: blocking pornographic images. *Proceedings of the international conference on image and video retrieval*, 2002, 50-60.

[5] Lopes APB, de Avila SEF, Peixoto ANA, Oliveira RS, de A. Arajo A, A bag-of-features approach based on hue-sift descriptor for nude detection. *Proceedings of the 17th European Signal Processing Conference. EUSIPCO*, 2009, 1552-1556.

[6] C.-Y. Kim, O.-J. Kwon, W.-G. Kim, and S.-R. Choi., Automatic System for Filtering Obscene Video. *International Conference on Advanced Communications Technology*, 2008, 2, 1435–1438.

[7] N. Rea, G. Lacey, C. Lambe, and R. Dahyot., Multimodal Periodicity Analysis for Illicit Content Detection in Videos. *Conference on Visual Media Production*, 2006, 106–114.

[8] X. Tong, L. Duan, C. Xu, Q. Tian, Hanqing L., J. Wang, and J. Jin., Periodicity Detection of Local Motion. *IEEE International Conference on Multimedia and Expo*, 2005, 650–653.

[9] C. Jansohn, A. Ulges, T. M. Breuel, Detecting Pornographic Video Content by Combining Image Features with Motion Information. *ACM international conference on Multimedia*, 2009, 601–604.

[10] Lukas, J., Fridrich, J., Goljan, M., Digital camera identification from sensor pattern noise. *IEEE Trans. Information Forensics and Security*, 2006, 1(2), 205-214.

[11] Zhang, H.J., Kankanhalli, A., Smoliar, S.W., Automatic partitioning of full-motion video. *Journal of Multimedia Systems*, 1993, 1, 10–28.

[12] Chen, M., Fridrich, J., Goljan, M., Lukas, J., Source Digital Camcorder Identification Using Sensor Photo Response Non-Uniformity. *The International Society for Optical Engineering (SPIE)*, 2007, 6505, 65051G.

[13] Mzhqak, M.K., Kozintsev, I., Ramchandran, K., Spatially Adaptive Statistical Modeling of Wavelet Image Coefficients and its Application to Denoising. *Acoustics, Speech, and Signal Processing (ICASSP)*, 1999, 6, 3253-3256.

[14] Chen, M., Fridrich, J., Goljan, M., Lukas, J., Digital Imaging Sensor Identification (Further Study). *The International Society for Optical Engineering (SPIE)*, 2007, 6505, 0P–0Q.

[15] Kumar, B.V.K.V., Hassebrook, L.,Performance measures for correlation filters. *Applied Optics*, 1990, 29, 2997-3006.