A Novel Framework for Video Piracy Detection

Harshala Gammulle Chamila Walgampaya Amalka Pinidiyaarachchi

Department of Statistics and Computer Science, Faculty of Science, University Of Peradeniya, Peradeniya, Sri Lanka.

Department of Engineering Mathematics, Faculty of Engineering, University Of Peradeniya, Peradeniya, Sri Lanka.

Department of Statistics and Computer Science, Faculty of Science, University Of Peradeniya, Peradeniya, Sri Lanka.

Abstract
The digital age has ushered in a plethora of ways for video recapture and video tampering. Subsequently, digital video forensics has become increasingly important, in which recaptured video detection is one branch. The applications are not limited for illegal video copies detection in professional cinematography and home entertainment, and surveillance video authentication in crime scene investigation, but also being able to detect recaptured videos will enhance the robot vision and add more intelligence to security systems such as face authentication systems, by enabling them to detect live scene from re-projected one. Furthermore embedded in web, monitoring systems may provide additional tools for protection and administration of video contents which would otherwise have cost thousands of man-hours for manual screening.

In this paper, an automated movie piracy detection mechanism based on multiple feature descriptors is proposed. The proposed method uses combinations of low-level features including amount of blur, noise, color moments and texture patterns of video frames. To demonstrate the accuracy and efficiency of the proposed method, we maintained a video dataset comprised of videos obtained at different resolutions and different shutter speeds. In order to compare our proposed method with existing state of the art, we used the same video database used in [22]. For practical purposes, videos in dataset is composed of different durations (from 30 seconds to 15 minutes approximately) and different categories including sports, educational, movies, TV commercials and animated. Deviated from [22] we have additionally included surveillance videos to the database as well. In order to obtain a recapture video database, videos were recaptured in an artificially lit room with fine tuned controllable settings. A special setup was used to ensure that recaptured videos are of high quality and they cannot be distinguished by naked eye. Extracted features are used to train different Support Vector Machines (SVMs) and a feed forward back propagation neural network. The experimental results show that our method uses a reduced number of feature dimensions and exhibits greater robustness as well as greater accuracy compared current state of the art [20] in identification of the recaptured videos. The method is capable to generalize the approach to both high quality videos as well as for the surveillance video sequences with low resolution. Therefore the proposed architecture provides an efficient and flexible solution for video piracy detection.

Keywords
Video recapture detection, Video piracy, Video forensics, Feature extraction.

1. Introduction
The advancement in the multimedia industry has caused the digital devices to replace their analog counterparts in all aspects. This fact is evadible even considering professional cinematography, home video, and surveillance cameras. The increasing number of multimedia sharing platforms has caused video sequences to be routinely acquired and uploaded for general diffusion on the Internet [2]. Motion picture piracy damages cinematographic industry by billionaire losses every year [23]. Illegal video distribution is reached mainly through Internet with peer-to-peer systems, user generated content and streaming. The next distribution source is with hard copies. It has been estimated that profit margins generated by trafficking DVDs illegal copies are greater than drugs trafficking gains [24]. Digital camcorders are used by pirates in movie theaters to obtain copies of reasonable quality that are subsequently sold on a black market and transcoded to low bit-rates for illegal distribution over the Internet. Camcorder theft is one of the biggest problems facing the film industry [25]. Illegal recordings from movies in the theater are the single largest source of fake DVDs sold on the street and unauthorized copies of movies distributed on the Internet [26]. With the aid of sophisticated color correction, noise reduction and anti blurring capabilities in state-of-the-art video processing applications, even general public is capable to produce such recaptured videos with considerable quality.

Figure 1 shows the general method of recapturing a video from a theater screen. Even though these recaptured videos are certainly not of the same quality as their subsequent DVD releases, increasingly compact and high resolution video recorders are affording better quality video recordings [2].

Figure 1: Recapturing movies from theater screen

Video forensics, especially video recapture detection is widely applicable in crime scene investigation [2]. In such applications it required to validate surveillance video sequences for its authenticity. This approach can be even extended into the field of robotics. Studies are being conducted to enhance robotic vision to provide additional information for a robot or an unmanned vehicle to distinguish a re-projection of an object from actual object [7].

Face authentication system has recently been adopted for access control on mobile devices such as laptop computers and smart phones. Such authentication system is designed for fast response time and often not equipped with sophisticated algorithms for verifying a live face. Based on recent studies [7], it is concluded that these authentication systems can be easily tampered with recaptures video streams. Therefore the need of determining the authenticity of a video sequence has become more urgent.

There have been an increasing number of techniques proposed in the expanding field of video forensics. Among these techniques, proposed methods generally fall into two broad categories: techniques originally developed for images and applied frame-wise to videos [2], and algorithms specifically tailored to video sequences [20].

In the proposed method we extend the general idea of image feature extraction proposed in [7, 9] for image recapture detection methods into videos. Irrespective to fact whether it is a recaptured video or recaptured image, the recaptured image or video sequence has set of distinct features when compared with its original one [7]. Low color saturation, blurriness or lack of sharpness and existence of noise and other artifacts, make the feature based recapture detection a possibility. By considering the chromatic, blurriness, texture and noise features for all the frames in the video sequence, the decision of classification of the video sequence in to the respective class (Original or Recaptured) is made. When comparing the feature vector dimensions in related image recapture detection methods, proposed method uses lesser number of feature dimensions. This dimensionality reduction results in significantly less computations and improved efficiency in overall system.

The rest of the paper is organized as follows. The paper first reviews the current approaches towards feature extraction methodologies for image and video recapture detection. In Section 3, our detection algorithm is presented. The evaluation results of the algorithm are presented in Section 4, while in Section 5 we draw our conclusions and discuss potential future avenues of research.

2. Literature survey

In [20] Wang et al. proposed a method for detecting recaptured videos by considering key point extraction with Scale Invariant Feature Transform (SIFT). For each video frame, the calculation of skew value based on key points has been done in order to determine the class of the video (i.e. original or recaptured). For each detected SIFT feature point the algorithm generates 128 dimensional feature vector. Therefore the computation of skew values for each of such feature point is computationally intensive.

When considering the related work done in the area of image recapture detection, a new face anti-spoofing approach [10] is proposed based on analysis of contrast and texture characteristics of captured and recaptured images. The approach utilizes the idea that recaptured images are low in contrast and artifacts such as texture patterns are introduced to the images due to the low resolution in recapturing devices. These assumptions generally do not hold when considering high quality recaptured videos produced by modern day high end video cameras.

The authors in [7] proposed a method for recaptured image detection based on some different physics based features such as surface gradient, contrast, Spatial distribution of specularity, background contextual information, etc. with a detailed analysis on each individual physics based feature.

Gray-level image noise level estimation algorithms are generally classifiable into patch-based and filter-based approaches. The authors in [19] proposed a filter-based noise level estimation method where the Laplacian operator has been used to suppress the image structure and to exclude pixels associated with edges. They have used an adaptive edge detection method. The main difficulty inherent in filter-based methods is that the difference between the original and filtered image is assumed to be the noise. But this assumption is not true for images with complex structures or details.

In [8], a patch-based algorithm is proposed in which an image is split into numerous patches. We can consider an image patch as a rectangular window in the image with size W × W. The patches with the small standard deviation among decomposed patches are call smooth patches. Those smooth patches have the least change of intensity. The intensity variation of a smooth patch is mainly caused by noise. The main issue of patch-based methods is how to identify the weak textured or smooth patches for various scenes in the presence of Gaussian noise. A novel algorithm is proposed by [12] to select weak texture patterns from a single noisy image based on the gradient of the patches and their statistics. After selecting weak texture

patches they have applied the Principle Component Analysis (PCA) to estimate noise levels.

The amount of blur is characterized by computing the average extent of the edges [17]. Based on a given edge detector, these metrics are sensitive not only to the threshold choice to classify the edge, but also to the presence of noise which can mislead the edge detection.

Basic gray-scale and rotation invariant texture classification has been initially addressed by the authors in [1, 21]. Both studies approached gray-scale invariance by assuming that the gray-scale transformation is a linear function. This is a somewhat strong simplification, which may limit the usefulness of the proposed methods. In [1], gray-scale invariance is realized by global normalization of the input image using histogram equalization. More computationally simple approach which is robust in terms of grayscale variations is proposed in [16]. They define a rotation invariant operator for detecting these fundamental properties of local image textures.

3. Methodology

3.1 Recapturing videos

To generate high quality recaptured videos we have set up a video recapturing environment, which contains large number of controllable settings including camera settings, display settings and environmental settings. A special attention was taken to create high quality recaptured video database which cannot be differentiated by naked human eye.

Recaptured videos were taken in 1/25 seconds, 1/60 seconds and 1/120 seconds shutter speeds. In order to identify the effect of focal length in recapture video detection, different distances between camera and display screen were considered. Cameras were placed at 100 cm, 200 cm and 250 cm from the display screen.

In order to consider recapturing from theater screen videos were recorder in a dark room, simulating a cinema ambience. Each video was projected in a white wall with a Panasonic LCD projector model PT-LB2. To demonstrate the effect of recapturing videos from other display medias we have recaptured videos from LCD screens as well. In this scenario the recapturing environment was artificially lit with cool white/daylight compact fluorescent lamps (CFL). Therefore the color temperatures in the video capturing devices were manually set to 5,000 K which is the default color temperature for fluorescent lamps. LCD screen brightness and contrast were calibrated using the built in calibration tool which comes along with the operating system. Two LCD screens were used for recapturing, one with resolution 1280 × 720 (progressive scan) and other with resolution 1920 × 1080 (progressive scan).

3.2 Low-level feature extraction

The low level feature descriptors used in this paper include blur amount, color moments, texture patterns (Local Binary Patterns) and noise levels.

3.2.1 Blurriness

The blurriness can be occurred due to three possible scenarios [9]. First, capture could be of low resolution. Second, the recaptured frame may be small and the display medium may have to be placed outside of the focus range due to a specific recaptured setting. And the third, if the end-user camera has a limited depth of field, the distant background may be blur, while the entire display medium is in focus. In this paper we used the concept of blur matrix which is introduced in [4].

A pixel in a color image is composed of luminance and chrominance components. Luminance contains the intensity or the gray value of the pixel where as chrominance component stores the color information. By the study conducted in [4], it is verified that the sharpness of an image is contained in its gray component. Therefore we estimate the blur annoyance only on the luminance component. Figure 2 shows the flow chart of the algorithm with the references to the following equations.

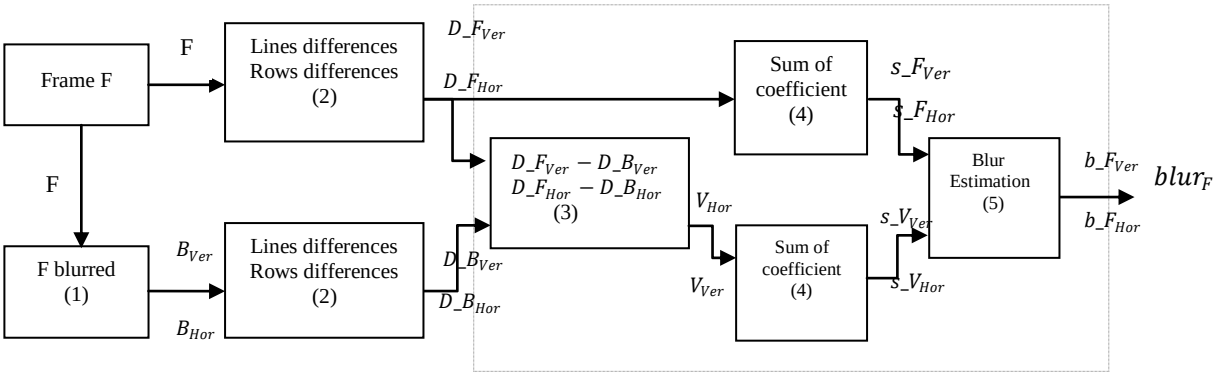


Figure 2: Flow chart of the blur estimation

Let F be the luminance component of a video frame of size of m × n pixels. To estimate the blur annoyance of F, blurred image B is obtained by blurring it. To model the blur effect and to create B_{Ver} and B_{Hor} vertical h_v and horizontal h_h, low pass filters are chosen.

$$h_v = \begin{bmatrix} \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{bmatrix}$$

$$h_h = \text{transpose}(h_v) = h_v' \quad (1)$$

$$B_{ver} = h_v * F \quad (2)$$

$$B_{Hor} = h_h * F \quad (3)$$

Then, in order to study the variations of the neighboring pixels, the absolute difference images D_F_{Ver}, D_F_{Hor}, D_B_{Ver} and D_B_{Hor} are computed.

$$D_{F_{Ver}(i,j)} = Abs\big(F(i,j) - F(i-1,j)\big) \quad (4)$$

for i=1 to m-1 , j=0 to n-1

$$D\_F_{Hor}(i,j) = Abs\big(F(i,j) - F(i,j-1)\big) \quad (5)$$

for j=1 to n-1 , i=0 to m-1

$$D_{B_{Ver}(i,j)} = Abs\big(B_{Ver}(i,j) - B_{Ver}(i-1,j)\big) \quad (6)$$

for i=1 to m-1 , j=0 to n-1

$$D\_B_{Hor}(i,j) = Abs\big(B_{Hor}(i,j) - B_{Hor}(i,j-1)\big) \quad (7)$$

for j=1 to n-1 , i=0 to m-1

Then the variation of the neighboring pixels after the blurring step is need to be analyzed. If this variation is high, the initial image or frame was sharp whereas if the variation is slight, the initial image or frame was already blurry. This variation is evaluated only on the absolute differences which have decreased.

$$V_{Ver} = Max(0, D\_F_{Ver}(i,j) - D\_B_{Ver}(i,j)) \quad (8)$$

for i=1 to m-1 , j=1 to n-1

$$V_{Hor} = Max(0, D\_F_{Hor}(i,j) - D\_B_{Hor}(i,j)) \quad (9)$$

for i=1 to m-1 , j=1 to n-1

Further, in order to compare the variations from the initial picture, sum of the coefficients of D_F_{Ver}, D_F_{Hor}, D_V_{Ver}, D_V_{Hor} are computed as follows.

$$s\_F_{Ver} = \sum_{i,j=1}^{m-1,n-1} D\_F_{Ver}(i,j) \quad (10)$$

$$s\_F_{Hor} = \sum_{i,j=1}^{m-1,n-1} D\_F_{Hor}(i,j) \quad (11)$$

$$s\_V_{Ver} = \sum_{i,j=1}^{m-1,n-1} D\_V_{Ver}(i,j) \quad (12)$$

$$s\_V_{Hor} = \sum_{i,j=1}^{m-1,n-1} D\_V_{Hor}(i,j) \quad (13)$$

Finally the result is normalized in a defined range from 0 to 1.

$$b\_F_{Ver} = \frac{s\_F_{Ver} - s\_V_{Ver}}{s\_F_{Ver}} \quad (14)$$

$$b\_F_{Hor} = \frac{s\_F_{Hor} - s\_V_{Hor}}{s\_F_{Hor}} \quad (15)$$

blur value which is more annoying among the vertical one and the horizontal one is selected as the final blur value.

$$blur\_F = Max(b\_F_{Ver}, b\_F_{Hor}) \quad (16)$$

With this method one dimensional feature vector is extracted.

3.2.2 Noise Levels

Noise levels are estimated based on weak texture patterns extracted from a single noise image. We have used the patch-based noise level estimation algorithm proposed in [12]. First the algorithm selects weak texture patterns based on the intensity variation of the patch and their statistics. The patches whose standard deviations of intensity close to the minimum standard deviation among decomposed patches are selected as weak texture patterns. Later the noise levels in image patches are estimated using PCA. The PCA technique estimates the dominant feature value on the weak textured patch dataset.

After decomposing the image into overlapping regions with distinctive features or texture patch, the image can be viewed as,

$$y_i = z_i + n_i \quad (17)$$

Where z_i is the original image patch with the i-th pixel at its center written in a vectorized format and y_i is the observed vectorized patch corrupted by zero-mean Gaussian noise vector ni. The goal of noise level estimation is to calculate the unknown standard deviation σ given only the observed noisy image [12]. In this study, we have selected the maximum eigenvalue of the image gradient covariance matrix as the metric for texture strength.

The gradient covariance matrix, C_y, for the image patch y is defined as:

$$C_y = G_y^T G_y \quad (18)$$

$$G_y = [D_h y \quad D_v y] \quad (19)$$

Where D_h and D_v respectively represent the matrices to represent horizontal and vertical derivative operators. Much information about the image patch can be reflected by the eigenvalue and eigenvector of the gradient covariance matrix.

$$C_y = V \begin{bmatrix} s_1^2 & 0 \\ 0 & s_2^2 \end{bmatrix} V^T \quad (20)$$

The maximum eigenvalue of the gradient covariance matrix s_1^2 reflects the strength of the dominant direction of that patch. The larger maximum eigenvalue reflects the higher noise level. With this method, a three dimensional feature vector is extracted. We estimate the maximum eigenvalue for each Red Blue and Green streams in the input video frame.

3.2.3 Color Moments

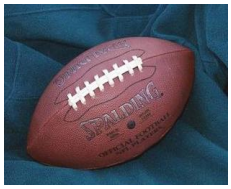
The majority of the color artifacts can be reduced by using the previously mentioned video recapturing setup with the high quality video camera and fine tuning the environmental settings. Still color of the finely recaptured

video can still become slightly different from its original video. Recaptured videos which are taken from an LCD screen are tampered with some blue tint when compared against its original video [7]. A frame from film dataset and its corresponding recaptured frame are shown in Figure 3. It is evident that recaptured frame is tampered with blue color tint.

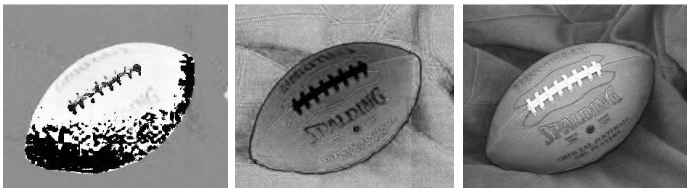
(a)Original video (b)Recaptured video

Figure 3: Original video and the recaptured video

The accuracy in classification applications using HSV (Hue, Saturation, Value) based features have accuracy which is 5.2% higher than the accuracy of RGB (Red, Green, Blue) based features [3]. A sample image and the separated hue saturation and value streams are shown in figure 4. As HSV color space separates the intensity value from its chromatic information, the changes in color features can be easily identified. Therefore, in this study only the HSV color moments are used. In order to increase the efficiency of the classifier, only mean and standard deviation of each stream is considered. Then, only six dimensional features will be extracted including means and standard deviations for each HSV component separately.



(a)Original image

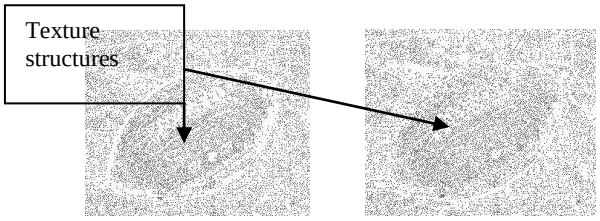


(b) Hue component (c) Saturation component (d) Value component

Figure 4: An RGB image and its corresponding HSV components

3.2.4 Texture Patterns (Local Binary Patterns)

Though the texture patterns can be easily observed on a poor quality video, it is generally impossible to detect them in a finely recaptured video. But complete elimination of texture patterns is also difficult. To capture texture patterns, features are extracted using Local Binary Patterns (LBP) [16], which is a widely used for texture analysis. LBP is a non-parametric descriptor, which efficiently summarizes the local structures of images by comparing each pixel with its neighboring pixels. Local binary patterns extracted for the original image in Figure 4 is shown in Figure 5. The local structures within red and green components are indicted in the figure.



(a)Local binary patterns in Red stream (b)Local binary patterns in green stream

Figure 5: Local binary patterns

3.3 Simulation of data losses

The utilization of a primary video copying detection method is not sufficient for a system where videos are exposed to severe and diverse attacks. Existing methods of video recapture identification schemes are robust against only some moderate attacks but they cannot detect video tampering process such as frame droppings, re-sampling, color space transformations and projective transformations. In order to come up with a classifier which is capable to detect such attacks, we manually adjusted 15 videos in the recaptured video database using the following guidelines.

1. Frame dropping: 10% of frames were dropped.
2. Frame rate: changed to 20 fps.
3. Color space transformations: from YCrCb to RGB.
4. Projective transformations: theta=2°
5. Compression techniques: changed from MPEG/ AVI/ MOV to MP4.

3.4 Classification

Both Feed forward neural networks and statistical classifiers such as Support vector machines (SVMs) are widely used in data mining literatures. As SVM has a theoretical guaranty regarding over fitting, flexible selection of kernels for non-linear correlation and classification of higher dimensional feature spaces, it has become increasingly popular choice among researcher [5]. Training the SVM requires the solution of very large quadratic programming (QP) optimization problem. Therefore training a SVM classifier is slow especially for large problems and SVM training algorithms are complex, subtle and sometimes difficult to implement.

Feed forward neural networks, with their ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques [11]. Among other advantages are the capability for adaptive learning, real time operations, fault tolerance via redundant information coding and high computation speed as a result of the parallel structure. But there are certain drawbacks as well. Errors of neural networks vary depending upon the architecture. Lengthy training times and the possibility of getting stuck on a local optimum are only a few of them.

The selection of classifier is subjective to the application. Therefore in the study we have trained a feed forward neural network model and two widely applied SVM models, which are C-SVM model and nu-SVM model.

In C-SVM training process involves the minimization of the error function:

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i \qquad (21)$$

Subject to the constraints,

$$y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i \; and \; \xi_i \geq 0 \, , i = 1, \ldots, N$$

where C is the capacity constant, w is the vector of coefficients, b is a constant, and ξ_i represents parameters for handling non-separable data (inputs). The index i labels the N training cases.

In contrast to classification C-SVM, the classification nu-SVM model minimizes the error function:

$$\frac{1}{2}w^T w - v\rho + \frac{1}{N}\sum_{i=1}^{N}\xi_i \qquad (22)$$

Subject to the constraints,

$$y_i(w^T\phi(x_i) + b) \geq \rho - \xi_i \, , \xi_i \geq 0 \, , i = 1, \ldots, N \; and \; \rho \geq 0$$

where w is the vector of coefficients, ρ is a constant, ξ_i represents parameters for handling non-separable data (inputs). The index i labels the N training cases.

4. Experimental Setup and Testing Results

4.1 Video dataset

We maintained an original video dataset and a recaptured video dataset. The original video dataset is composed by 50 open videos of different categories, that is, documental, TV commercials, animated, sports and movies with 1080dpi, 800dpi resolutions. This database exhibits a similar composition to the database used in [22]. Additionally, we included 21 low quality surveillance videos in 480dpi, 144dpi resolutions. All of them are in color without audio component. Table 1 enumerates the video dataset. Some video references are listed at the end of this document. Recaptured videos were obtained using three types of cameras (SONY HDR-CX240 full HD Camcorder, Canon Vixia HV30 HD, Nokia x3-02 5MP phone camera) with different focal lengths and shutter speeds.

4.2 Experiments

Fifty original videos and fifty recaptured videos are taken with different durations and frame rates. Generally the duration of a video is restricted to less than twenty minutes. For each video a pre specified percentage of frames are selected and feature extractions are performed only for these selected frames. A pre specified percentage is defined in order to avoid feature extractions of similar video frames and with that to make the process more efficient. Number of video frames and their respective resolutions are shown in Table 1. In proposed algorithm only sixteen dimensional feature vectors are computed. Some frames from our database is shown in Figure 6.

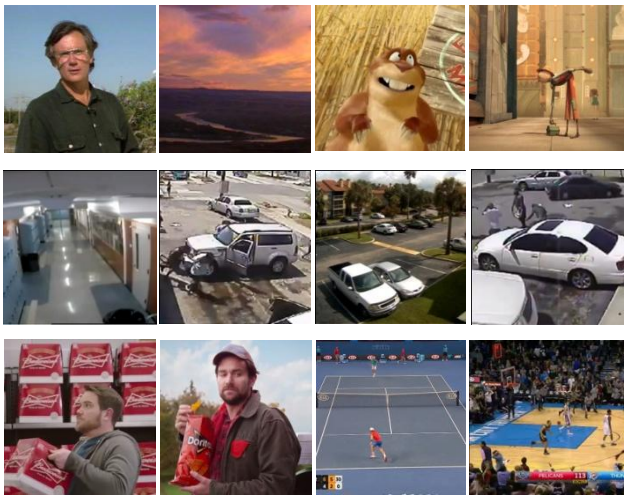


Figure 6: Sample frames from video database

As a preprocessing step entire data set is normalized to zero mean with standard deviation of one. In order to perform a quantitative validation on the proposed method, three experiments were conducted.

Table 1: Number of video frames extracted and their respective resolutions

		Number of video frames	
Video Type		**Original**	**Recaptured**
Movies (Resolution 1080dpi)		509	510
Documental (Resolution 800dpi)		132	124
TV commercials (Resolution 1080dpi)		201	204
Animated (Resolution 1080dpi)		98	112
Sports (Resolution 800dpi)		153	122
Surveillance videos	(480 dpi Resolution)	300	248
	(144 dpi Resolution)	220	270

In the first experiment we trained a feed forward back propagation neural network model for the overall dataset. In the next experiment we implemented both C-SVM and nu-SVM classifiers and evaluated their respective accuracies for each separate dataset and for the overall dataset.

In the final experiment we compare the proposed method against current state of the art.

4.2.1 Experiment 1

A neural network consisted with 3 layers, input layer with 16 neural, hidden layers with 30 neural and one output layer was used (Figure 7). For the overall dataset a mean square error of 0.67585 was observed. The performance plot indicating the mean square error against number epochs is shown in figure 8.

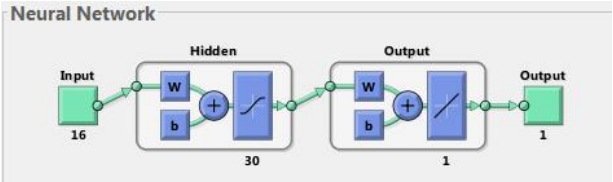


Figure 7: Neural network model

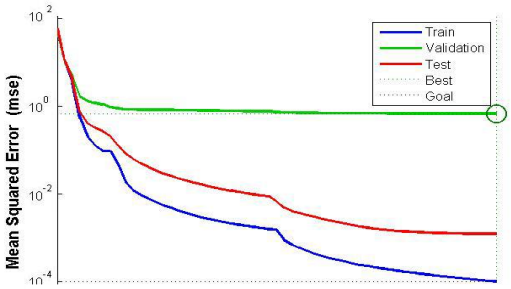


Figure 8: Mean square error VS epochs

70%, 15% and 15% of the dataset was allocated for the training; validation and testing respectively which results 2062 frames for training, 442 frames for validation and remain 442 frames for testing. Regression plots obtained for the final neural training process are shown in Figure 9. First graph refers to the training process and the regression value (R) is 0.96464, second and third graphs refer to the validation and testing where R values are 0.94577 and 0.96462. Therefore final R value is 0.96174.

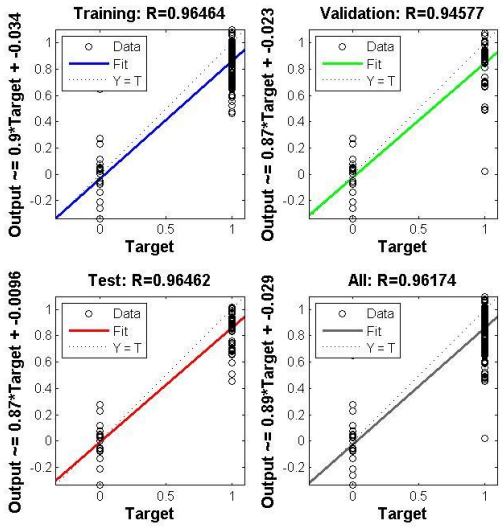


Figure 9: Regression Plots

The dashed line in each plot represents the perfect result – outputs = targets. The solid line represents the best fit linear regression line between outputs and targets. The overall R value of 0.96174 is an indication of good linear relationship between the outputs and targets. The validation and test results with R values around 0.95 verify the above comment.

4.2.2 Experiment 2

The obtained results for both C-SVM and nu-SVM classifier models are shown in Tables 2 and 3. The accuracy of the trained classifier is tested with 66% of the dataset for training together with 33% for testing and also with 5 step cross validation.

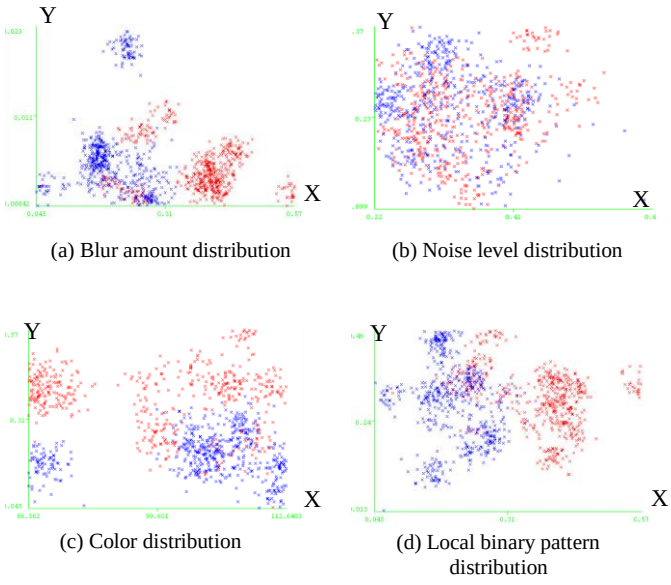
Table 2: The accuracy obtained with different classifiers when testing 33% of data after training classifiers with 66% of data for different datasets separately

Dataset	Number of frames	Accuracy (in %)	
		nu-SVM	C-SVM
Movies (Resolution 1080dpi)	1019	94.958	85.599
Documental (Resolution 800dpi)	202	89.278	91.952
TV commercials (Resolution 1080dpi)	320	87.509	88.181
Animated (Resolution 1080dpi)	155	92.031	89.735
Sports (Resolution 800dpi)	212	85.419	86.282
Surveillance videos(480 dpi Resolution)	548	90.531	82.358
Surveillance videos(144 dpi Resolution)	490	91.167	88.379
Overall	**2946**	**90.127**	**87.498**

Table 3: The accuracy obtained with different classifiers with 5 step cross validation for different datasets separately

Dataset	Number of frames	Accuracy (in %)	
		nu-SVM	C-SVM
Movies (Resolution 1080dpi)	1019	95.023	92.952
Documental (Resolution 800dpi)	202	94.222	93.342
TV commercials (Resolution 1080dpi)	320	91.593	84.105
Animated (Resolution 1080dpi)	155	89.192	86.427
Sports (Resolution 800dpi)	212	91.782	90.183
Surveillance videos(480 dpi Resolution)	548	90.114	89.565
Surveillance videos(144 dpi Resolution)	490	87.498	88.299
Overall	**2946**	**91.346**	**89.267**

We linearly project each type of the features to a 2 dimensional space through Fisher discriminant analysis. Figure 10 shows the distribution of the six types of features described in Section 3. From the Figure 10, it is observed that color features, local binary patterns and blur amount distribution are the most effective features for our dataset, while noise level shows less effective random distribution.



(a) Blur amount distribution (b) Noise level distribution

(c) Color distribution (d) Local binary pattern distribution

Figure 10: 2D projection of the feature distribution for the "movie dataset ". Original videos (red) and the recaptured videos (blue)

4.2.3 Experiment 3

In order to compare the accuracy and efficiency of the proposed method against existing ones, we implemented the method proposed by Wang et. al [20] in our platform and evaluated False Alarm Rate (FAR) and execution time for a random sample of videos from our database. Same videos were assigned to the proposed algorithm and parameters were evaluated. We ran the experiment 100 times on a computer with Intel Core i5 processor, 4GB RAM and Windows 7 operating system. Table 4 contains the values for evaluation matrices measured for six sample videos.

$$\text{False Alarm Rate (FAR)} = \frac{\text{Number of incorrect classifications}}{\text{Number of trials x Number of frames in the database}} \quad (23)$$

Table 4: Evaluation matrices comparison

Seq.	# Fr.	Average Execution time (in seconds)		Average False-Alarm rate	
		Algorithm in [20]	Proposed	Algorithm in [20]	Proposed
1	21	0.7	0.5	0.43	0.062
2	62	3.5	3.1	0.14	0.069
3	55	2.2	2.3	0.21	0.011
4	46	1.8	1.3	0.133	0.053
5	40	0.4	0.8	0.173	0.067
6	95	4.9	5	0.24	0.098
avg	**53**	**2.2**	**2.1**	**0.221**	**0.060**

5. Conclusion

In this paper, we proposed a video recapture detection method based on multiple feature descriptors. The effectiveness of proposed method is demonstrated using same video datasets used by Zavaleta et. al [22] in their study. We tested our algorithm on both high quality videos as well as for videos with lower resolution covering all aspects that are generally found in multimedia industry. Through a proper setup of the video recapturing environment and by fine tuning the controllable settings, we have recaptured the videos displayed on different types of screens with reasonably higher quality where the videos cannot be classified with the naked human eye. Apart from general camera recordings we have also considered other types of attacks in video piracy, such as data losses: including frame dropping, bit rate change, frame rate change, compression and visual transformations: including cropping, projective transformations and color space transformations. Our proposed features capture the textured patterns, the loss-of-fine details characteristics, introduction of noise and the color anomalies introduced in the video recapturing process. Experimental results prove that the extracted features to be highly effective while keeping a much lighter weight dimension.

The extracted feature vectors are used to train different SVMs as well as a feed forward back propagation neural network. The experimental results suggest that feed forward neural network with 30 hidden layers exhibits

significantly higher accuracy than both nu-SVM and C-SVM classifiers. The overall accuracy for both nu-SVM and C-SVM models are around 90% where as the accuracy of the feed forward neural network model is around 95%.

Based on our comparisons, proposed method can be considered as a better approach compared to current state of the art [20] due to its significantly less feature dimensions and higher performance compared to existing methods. Therefore a significant improvement with respect to both accuracy and efficiency in identification of recaptured videos is observed.

Further analysis could be devoted to identification of source digital camcorder based on the extracted feature vectors. Another interesting direction is the multimodal analysis of video and audio streams to further increase the detector robustness. A special attention would be paid for identification of synthetic distortions.

6. References

[1] Chen L. , Kundu A. 1994. Rotation and Gray Scale Transform Invariant Texture Identification Using Wavelet Decomposition and Hidden Markov Model, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16 (2), Feb 1994, 208-214.

[2] Chen M., Fridrich J., Goljan M., Lukáš J. 2007. Source Digital Camcorder Identification Using Sensor Photo Response Non-Uniformity, *Proceedings of SPIE Electronic Imaging, Photonics West*,2007.

[3] Chen W., Shi Y. Q., Xuan G. 2007. Identifying Computer Graphics using HSV Color Model and Statistical Moments of Characteristic Functions, *Multimedia and Expo IEEE International Conference* , 2007.

[4] Crété-Roffet F., Dolmiere T., Ladret P., Nicolas M. 2007. The Blur Effect- Perception and Estimation with a New No-Reference Perceptual Blur Metric, *GRENOBLE In SPIE proceedings - SPIE Electronic Imaging Symposium Conf Human Vision and Electronic Imaging*, 2007.

[5] Cristianini N., Taylor J.S. 2000. Support Vector Machinesand other kernel-based learning methods. *Cambridge University Press*, 2000.

[6] Dehnie S., H. T. Sencar, Memon N. D. 2006. Digital Image Forensics for Identifying Computer Generated and Digital Camera Images. *ICIP* 2006: 2313-2316.

[7] Gao X., Ng T., Qiu B., Chang S. 2010. Single-view Recaptured Image Detection Based on Physicsbased Features, *IEEE International Conference on Multimedia and Expo (ICME)*, 2010.

[8] Hyuk S. D., Hong P. R., Joon Y. S., Han J.J, 2005. Block-based noise estimation using adaptive Gaussian filtering. *IEEE Transactions on Consumer Electronics*, Feb. 2005, vol. 51, pp. 218–226.

[9] Ke Y., Shan Q., Qin F., Min W. 2013. Image Recapture Detection Using Multiple Features, *International Journal of Multimedia and Ubiquitous Engineering*, 2013.

[10] Kose N. , Dugelay L. 2012. Classification of Captured and Recaptured Images to Detect Photograph Spoofing, 1st *International Conference on Informatics, Electronics and Vision*, 2012.

[11] Kustrin S.A , Beresford R. 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, 2000.

[12] Liu X., Tanaka M. and Okutomi M. 2012. Noise Level Estimation Using Weak Textured Patches of a Single Noisy Image, *IEEE International Conference on Image Processing (ICIP)*, 2012.

[13] Lukas J., Fridrich J., and Goljan M. 2006. Detecting Digital Image Forgeries Using Sensor Pattern Noise, *Proceedings of SPIE Electronic Imaging, Security, Steganography, and Watermarking*, January 2006, pp. 16-19.

[14] Lukas J., Fridrich J., and Goljan M. 2005. Determining digital image origin using sensor imperfections, *SPIE Electronic Imaging, January*, 2005, pp. 249–260.

[15] Milani S., Bestagini P., Tagliasacchi M., and Tubaro S. 2012. Multiple compression detection for video sequences, *IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, 2012, pp. 112 – 117.

[16] Ojala T., Pietikäinen M. Mäenpää T. 2000. Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns, *Lecture Notes in Computer Science Volume 1842*, 2000, pp 404-420.

[17] Ong E., Lin W., Lu Z., Yang X., Yao S., Pan F., Jiang L., Moschetti F. 2003. A no-reference quality metric for measuring image blur, *Seventh International Symposium on Signal Processing and Its Applications*, July 2003.

[18] Schwartz W.R., Kembhavi A. , Harwood D. and Davis L. S. 2009. Human Detection Using Partial Least Squares Analysis, *Proceeding of IEEE 12th International Conference on Computer Vision*, 2009.

[19] Tai S.C., Yang S. 2008. A fast method for image noise estimation using laplacian operator and adaptive edge detection, 3rd *International Symposium on Communications, Control and Signal Processing*, March 2008, pp. 1077–1081.

[20] Wang W. and Farid H. 2008. Detecting Re-Projected Video, *Springer, Berlin Heidelberg*, 2008.

[21] Wu R. , Wei C. 1996. Rotation and Gray-Scale Transform Invariant Texture Classification Using Spiral Resampling, Subband Decomposition and Hidden Markov Model, *IEEE Trans. Image Processing*, 1996.

[22] Zavaleta J., Feregrino C. 2014 Content Multimodal Based Video Copy Detection Method for Streaming Applications. *Technical Report No. CCC-14-001, National Institute of Astrophysics, Optics and Electronics (INAOE)* , January 2014.

[23] WIPO, "Which products are typically affected? on Program Activities," 2013. [Online]. Available:http://www.wipo.int/enforcement/es/faq/counterfeiting/faq03.html. [Accessed 14 August 2014]

[24] B. Monnet and P. Véry, Les nouveaux pirates de l'entreprise. Mafias et terrorisme, Paris: CNRS, 2010.

[25] MPAA-Types of content theft, "Motion Picture Assosiation of America," 2013. [Online]. Available:http://www.mpaa.org/contentprotection/types-of-content-theft. [Accessed 14 August 2014]

[26] MPAA-Camcorder laws, "Motion Picture Assosiation of America," 2013. [Online]. Available: http://www.mpaa.org/contentprotection/camcorder-laws. [Accessed 14 August 2014].

Harshala Gammulle is currently an undergraduate student following a BSc computer science special degree program in Faculty of science, University of Peradeniya Sir Lanka. Her research interest include, Digital forensics, Artificial intelligence and image processing.

Dr Chamila Walgampaya is currently a lecturer in the Department of Engineering Mathematics, University of Peradeniya.He earned his Ph.D. in August 2011 from the School of Engineering at the University of Louisville. His research interests include Click fraud mining, Automatic web robots and agents, Data and evidence fusion.

Dr. (Mrs.) Amalka J. Pinidiyaarachchi currently a lecturer in the Department of Statistics and Computer Science University of Peradeniya. She obtained her PhD from Uppsala University Sweden (2009) and her BSc from University of Peradeniya (2001). Her research experties include Biomedical engineering, Cell image analysis and Coarse to fine search in object recognition.

Appendix

Some Video Dataset References

Table 5: Video references

Category	URL
Documental	http://www.open-video.org/details.php?videoid=346
	http://www.open-video.org/details.php?videoid=348
	http://www.open-video.org/details.php?videoid=351
TV Commercials	http://www.youtube.com/watch?v=4KEBw6opgVk
	http://www.youtube.com/watch?v=k_c_zNw2tBQ
	http://www.youtube.com/watch?v=2SXOfIKtJIk
	http://www.youtube.com/watch?v=Fo31riY3mzM
	http://www.youtube.com/watch?v=36kHzCCJkeM
Animated	http://www.bigbuckbunny.org/index.php/download
	http://www.youtube.com/watch?v=lUtnas5ScSE
	http://www.youtube.com/watch?v=oxtP3wxXlTA
Sports	http://www.youtube.com/watch?v=kmar9bLehVY
	http://www.youtube.com/watch?v=oyxhHkOel2I
	https://www.youtube.com/watch?v=3xZo77kkfbk
Surveillance videos	https://www.youtube.com/watch?v=7qhWzKJuras
	https://www.youtube.com/watch?v=15vqUf6H-po
	https://www.youtube.com/watch?v=cAKc5VPfQ7Q