A NOVEL SOURCE MPEG-2 VIDEO IDENTIFICATION ALGORITHM

YUTING SU*, JUNYU XU†, BO DONG‡ and JING ZHANG§

School of Electronic Information Engineering
Tianjin University, Tianjin 300072, P. R. China
*ysu@tju.edu.cn
†xujunyu@tju.edu.cn
‡dongbo@tju.edu.cn
§zhangjing@tju.edu.cn

QINGZHONG LIU
Department of Computer Science
Sam Houston State University
Huntsville, TX 77341, USA
liu@shsu.edu

With the availability of powerful multimedia editing software, all types of personalized image and video resources are available in networks. Multimedia forensics technology has become a new topic in the field of information security. In this paper, a new source video system identification algorithm is proposed based on the features in the video stream; it takes full advantage of the different characteristics in the rate control module and the motion prediction module, which are two open parts in the MPEG-2 video compression standard, and combines a support vector machine classifier to build an intelligent computing system for video source identification. The experiments show this proposed algorithm can effectively identify video streams that come from a number of video coding systems.

Keywords: Source identification; digital video forensics; motion vector; bit-stream.

1. Introduction

With the tremendous growth of high-quality digital cameras and easy-to-use multimedia editing software, it is becoming increasingly easier to tamper with digital multimedia resources. Earlier, most of these modification actions only served the purpose of entertainment, but recently some malicious tampering and falsification of digital images and videos have begun to impact the bottom line of laws and morals. People have started to doubt the factuality and authority of every digital image they acquire from the network, because many events on false news and scientific data have emerged. Compared with digital images, the fabrication of digital video is more effective and destructive, because the videos in the television broadcast convey...
real-world scenes more vividly and credibly. In 2007, a television station in the Czech Republic played a video section without careful censorship, in fact, it was made by combining a local scene in Bohemia with a nuclear mushroom cloud. Thousands of audiences watched this terrible scene in their country, which almost led to social panic. On the other hand, with the development of video sharing websites, such as YouTube, various kinds of individual videos have begun to spread on the Internet quickly. Different from traditional videos, these online video resources are more personalized and less trustworthy. How to effectively monitor and manage these media resources has become a key to maintain a healthy and stable development for the information industry, so the passive blind forensics technology on digital media resources is becoming a hot issue in the field of information security techniques. The text and images are still major media resources on the internet, but accompanied by the expansion of network bandwidth, digital video resources are gradually becoming more popular. At the same time, with the availability of powerful editing software, amateurs can easily manipulate the video sections. In the future, video forensics will play an important role in the multimedia security management.

The current research on passive blind forensics technology focuses on three key issues — digital media source identification, discrimination of synthetic media and media forgery detection. Digital media source identification is the first step in the media forensics process, comprising two steps — source model identification and individual source identification, whose purpose is to identify and provide evidence of the digital media collection, processing and output of equipments (such as digital cameras, camcorders, scanners, printers, etc.). The simplest approach is to examine an image or video file’s header. In some standard multimedia format files, system information, such as types of digital cameras or video recorders, compression codec mode, date and time of media, can be found in their header, but this information may be maliciously altered or discarded during files edited. Another robust approach is based on the inherent characteristics of specific equipment and the statistics characteristics of output images or video data. Researchers tried to use some detail traits in the image processing for identifying the source camera, which include the camera lens distortion, the sensor-based defects, the specific periodic correlations among pixel values introduced by different interpolation methods from camera models, the noise pattern of each CCD sensor, and so on. Existing source camera identification algorithms can effectively distinguish different brands and models of cameras, but the identification of individual source cameras that belong to the same model remains a significant challenge in this field.

In the digital video forensic techniques, some researchers have extended image detection algorithms to video forensics applications. Houten et al. introduced the Photo Response Non-Uniformity (PRNU) as a unique sensor noise pattern to identify the source video cameras; Wang et al. exploited the static and temporal artifacts when a video sequence was subjected to double MPEG compression. In this scheme, a video sequence was divided into a set of still images, and an intra frame in double MPEG compression stream was viewed as an image which was subjected to
double JPEG compression, thus a double JPEG compression detection algorithm mentioned above could be directly extended to video coding system. Other researchers tried to use some special features in the video system, such as temporal artifacts, de-interlacing correlations, to detect different video tampering or editing. In the above video forensic algorithms, the video coding system is often assumed as a known condition, but in practical applications, even given the same input sequences and the same restrictive conditions, different video encoders may still cause significant differences in statistical characters and the objective quality of reconstructed frames. An instance is shown in Fig. 1, where the same test sequence is encoded by three different MPEG-2 encoders and there are obvious differences in the values of peak signal-to-noise ratio (PSNR) of corresponding reconstructed frames. Because estimation of the model and system parameters in the corresponding video encoding system will directly affect the performance of forgery detection algorithms, video source identification technique is a key task before realizing the video forensics algorithms.

In this paper, we propose a source identification system for the video streams compressed by a number of video coding systems based on the features extracted from bit stream and motion vectors. The rest of this paper is organized as follows. In Sec. 2, a brief introduction on digital video compression system is given, and a novel rate control algorithm and a novel motion vector estimation algorithm in the Test Model 5 (TM5) are taken as an example to illustrate the characteristics of rate control strategy and motion vector estimation algorithm in MPEG-2 system. Section 3 proposes a source video identification algorithm. Experimental results are shown in Sec. 4, followed by conclusions and the tips for future study in Sec. 5.

![Fig. 1. The PSNR for coded sequence “Schumacher” using Test Model 5 (TM5), the default MPEG-2 encoder in Adobe Premiere Pro 2.0 (Premiere), and Cinema.Craft.Encoder 2.7 (CCE), respectively.](image-url)
2. Video Compression Standard and Open Components in the Coding System

In the 1980s, the development and maturity of data compression technology makes it possible to use digital video systems for various telecommunication applications, such as digital TV broadcast, teleconferencing, and so on. At the same time, in order to meet the requirements of large-scale industrial production, the corresponding international organizations (such as ITU, ISO/IEC) launched a variety of video compression international standards for different applications, such as MPEG-2, MPEG-4, and so on. Due to the fact that MPEG-2 video coding standard is widely adopted by most digital videos and supported by most video encoding software, we select MPEG-2 video sequences for source identification study.

2.1. Video compression coding system

Most of the existing video compression systems adopt the hybrid coding structure, which integrates the three classical compression techniques: prediction coding, transform coding and entropy coding, as shown in Fig. 2. As an example of digital video system, the general MPEG-2 encoder will be briefly introduced as follows.

A MPEG-2 system defines three frame types in terms of temporal processing. They are \( I \) (intra-frame coded) frame, \( P \) (forward predictive coded) frame and \( B \) (bidirectionally predictive coded) frame. An \( I \)-frame goes through the discrete cosine transform (DCT) for the reduction of spatial redundancy, and the coefficients are quantized (Q) according to the characteristics of the human visual system, followed by variable length coding (VLC), which is a standard entropy coding algorithm. Next, it passes the buffer to ensure a constant bit-rate. Since the \( I \)-frame is used for predictive coding of \( P \) or \( B \)-frames, after quantization, it goes through dequantization \((Q^{-1})\) and an inverse DCT \((\text{DCT}^{-1})\), and eventually the original frame is reconstructed. That frame is then placed in frame store (FS) so that the future \( P \) and \( B \)-frames can use it for motion compensation (MC) to reduce the temporal redundancy.

![Fig. 2. Hybrid coding system structure diagram.](image-url)
The same general process is used for $P$ and $B$-frames, except that $B$-frames cannot be referred to, so only MC is used and the frames are not kept in FS. This process is looped until all frames are converted into a bit stream. The opposite process occurs at the MPEG-2 decoder.

The MPEG-2 standard provides a compressed stream format and syntax, and prescribes a standard MPEG-2 decoder system and related parameters, but it does not describe some specific technical details, such as rate control module and motion estimation module. It allows sufficient flexibility to introduce novel technologies in order to constantly improve the performance of the whole video compression system. IT manufacturers can use these open components to add their three-party schemes, and then create unique video systems with their own characteristics that can be seen as the special labels of their video systems. If the features in these open components can be extracted, we can find a way to identify the different video sources.

### 2.2. Rate control module in video system

When the variable-rate bit stream needs to be transmitted over a fixed-rate channel, a channel buffer is usually used to smooth out the bit stream. In order to prevent the channel buffer from overflowing and underflowing, rate control is an indispensable module for the coding system, which utilizes a feedback mechanism to control the coding parameters, such as quantization and the frame quality. It can temporarily and dynamically increase or decrease the coding length of subsequent frames to maintain a certain average rate dictated by the channel parameters.

Similar to other video compression standards, MPEG-2 does not mandate how to implement rate control. In the Test Model 5 of MPEG-2, a rate control method is described, which consists of three steps: bit allocation, rate control and modulation. Based on the bits of the previous frames, bit allocation assigns a target number of bits to each frame. To control the output bit-rate, the quantization parameter is adjusted at each macroblock according to the channel buffer fullness. To address the spatial content variation of a frame, the quantization parameter is further modulated by the macroblock spatial activity measure. This algorithm\textsuperscript{15} is simple but has some problems; some researchers have designed some advanced rate control algorithms.\textsuperscript{2,28} Because of the diversity of rate control algorithms, we can detect the special marks in the bit stream produced by different control algorithms, and thus design the corresponding classification or identification system.

### 2.3. Motion estimation module in video system

Motion estimation is the core module in the inter-frame prediction technique, and in most cases motion estimation occupies roughly 70% of the computational load on the video encoder. Most video compression standards take the block-matching algorithm, which estimates motion vectors on the block-by-block basis, but they only define the fundamental unit of block-matching (e.g. $16 \times 16$ pixels or $8 \times 8$ pixels), the upper limit of search range (e.g. 1024 for MPEG-1, and 2048 for MPEG-2), the
mode of compression coding and so on. The matching criterion, the search path and other specific technical details in the encoder do not conduct mandatory provision.

The traditional motion estimation algorithm uses an exhaustive search (ES), where every possible integer displacement within a presumed square search region is searched with a very high computation cost. Different manufacturers always develop some fast motion estimation algorithms to suit their specific applications. In the Test Model 5 of MPEG-2, a motion estimation algorithm is described, which consists of two steps: full pixel accuracy and half pixel accuracy. The first step calculates absolute difference between current macroblock and each macroblock in the reference region that is determined by horizontal and vertical search ranges in the original frame, and then the horizontal and vertical coordinates of the minimum errors are obtained by comparison. The second step uses the result of the first step to find the reference macroblock in the local decoded frame and then interpolate the reference macroblock to get half-pixel accuracy by using the same method in the first step. In many improved algorithms, optimizing the search path and matching criteria are the main approaches to improve the coding efficiency of the whole system. Figure 3 shows the motion vectors of each macroblock in one P frame encoded by two MPEG-2 encoders. The results of different motion estimation algorithms are obviously different.

3. Source Video Identification Scheme

As the core module in the digital video coding system, video compression algorithm will directly affect the statistical distribution of encoded streams and the quality of the reconstructed frames. Due to the characteristics in the MPEG-2 video compression standard, our proposed source video identification scheme will extract the features in these open modules in the coding system, and combine a support vector

![Motion vectors](image1)

(a) Motion vectors calculated by Adobe Premiere Pro 2.0

![Motion vectors](image2)

(b) Motion vectors calculated by TM5

Fig. 3. The motion vectors of the same frame encoded by two MPEG-2 encoders.
machine for multiclass classification to identify the source of video resources. In our source identification scheme, a MPEG-2 coded stream is analyzed firstly, some elementary coding parameters of video stream, such as image resolution, average output bit-rate, frame rate, GOP structure, and VBV buffer size, are extracted. By applying SVMs to these features, an intelligent identification system is built to determine the source of MPEG-2 video stream.

3.1. Feature extraction

As discussed above, rate control and motion estimation are the two main open modules in the MPEG-2 coding standard. The aspects that they finally express in the encoded stream are bit-rate, quantization factors and motion vectors. Thus, we extract the features from these three aspects, described as follows.

(a) Bit-rate features

Bit allocation is always the first step in the rate control scheme, which can reflect the different thread of designers. Before a frame is encoded, a rate control scheme will budget the number of bits for each frame according to many factors, such as the frame type, the number of bits of previous frames, buffer state, complexity of current frame and so on. Some encoders tend to improve the quality of spatial resolution, thus they may increase the number of bits for I frames (\(NB_I\)). But if the encoders incline to assure the quality of temporal resolution, they will increase the number of bits for P frames (\(NB_P\)) and B frames (\(NB_B\)) in order to guarantee the continuity of video content. On the other hand, some simple or real time video systems scarcely ever change bit allocation between adjacent B frames, while other complex encoders may introduce some good fine adjustment strategies into adjacent B frames. Based on these analyses, four kinds of features of bit-rate are defined:

(1) \(M, N\), are the number of P frames and B frames in a group of pictures (GOP) respectively.

(2) \(RPI\), is the ratio of the average number of bits of P frames to the number of bits of I frame in a GOP.

\[
RPI = \frac{\frac{1}{M} \sum_{i=1}^{M} NB_P}{NB_I}
\]  

(3) \(RAP, RDP\), are the average and variance of the relative difference between adjacent P frames in a GOP.

\[
RAP = \frac{1}{M-1} \sum_{j=1}^{M-1} D_P(j)
\]  

\[
RDP = \frac{1}{M-1} \sum_{j=1}^{M-1} (D_P(j) - RAP)^2
\]
where $D_P(j)$ is the relative difference between adjacent $P$ frames, defined as (4).

$$D_P(j) = \frac{|NB_P(j + 1) - NB_P(j)|}{NB_P(j)} \quad j = 1, 2, \ldots, M - 1$$

(4) $RA_B, RD_B$, are the average and variance of the relative difference between adjacent two $B$ frames in a GOP.

$$RA_B = \frac{2}{N} \sum_{j=1}^{N} D_B(j)$$

(5)

$$RD_B = \frac{2}{N} \sum_{j=1}^{N} (D_B(j) - RA_B)^2$$

(6)

Where $D_B(j)$ is the relative difference, defined as (7).

$$D_B(j) = \frac{|NB_B(2j) - NB_B(2j - 1)|}{NB_B(2j - 1)} \quad j = 1, 2, \ldots, \frac{N}{2}$$

(7)

(b) Quantization factor features

In most existing coding systems, the core of different rate control algorithms is based on the same principle that they control bit stream by adjusting three parameters: quantization parameter, frame rate and coding mode for some blocks in inter-frame. The last two are often used to deal with abnormal conditions, such as underflowing or overflowing in the buffer. Adjusting quantization parameters is the main way to realize their own rate control targets. Control rule of quantization parameters can also reflect the features of different rate control schemes inversely. For example, to ensure that all reconstructed macroblocks have similar quality in each frame, some rate control algorithms minimize the modification of quantization parameters in the same frame. To output the smooth and steady bit stream, some rate control algorithms introduce fine adjustment schemes to change quantization parameters according to the state of buffer, the number of bits of previous macroblocks and so on. Therefore, we define some features from quantization parameters as the basis for the identification of different video systems.

1. $QM_k, k \in \{I, P, B\}$, is the maximum number of successive macroblocks with the same quantization parameter in a frame of type $k$, with the order from left to right and from top to bottom.
2. $QA_k, QV_k, k \in \{I, P, B\}$, are the average and variance of the number of successive macroblocks with the same quantization parameter in a frame of type $k$ with the above order.
3. $QMD_k, k \in \{I, P, B\}$, is the maximum difference value of quantization parameters between adjacent macroblocks in a frame of type $k$.
4. $QAD_k, k \in \{I, P, B\}$, is the average difference value of quantization parameters between adjacent macroblocks in a frame of type $k$. 
(c) *Motion vectors features*

MPEG-2 encoder designers have introduced different motion estimation algorithms according to actual application needs. For example, although most video compression standards also define the maximal size of motion vectors, the actual system usually uses a small search window for reducing the computation, especially in some real-time hardware coding system. On the other hand, in the motion estimation algorithm, a threshold value will be set to judge whether the current block is coded as a still block or not. The threshold value is always determined by system performance requirements. Some real-time coding systems always select a big value to shorten the coding time, while some others select a small one in order to make full use of prediction coding and improve the coding efficiency. We can also evaluate the performance of motion estimation algorithms and the match criteria to discriminate encoders. Thus, four types of features are defined as follows.

1. **$MX$** and **$MY$**, are the maximal horizontal and vertical sizes of the searching window.

2. **$MZ$**, is the still block feature which can reflect the decision threshold in the algorithm not to search a reference macroblock.

$$MZ = \frac{MM + MS}{2}$$  

Here, **$MM$** is the minimal difference between the moving macroblock in inter frame and the macroblock in reference frame at the same position, and similarly, **$MS$** is the maximal difference between still macroblock in inter frame and the macroblock in reference at the same position. They are defined as (9) and (10).

$$MM = \min_n \left( \sum_{x=1}^{8} \sum_{y=1}^{8} |X_M(x, y; n) - X^R_M(x, y; n)| \right) \quad n = 1, 2, \ldots$$ (9)

$$MS = \max_m \left( \sum_{x=1}^{8} \sum_{y=1}^{8} |X_S(x, y; m) - X^R_S(x, y; m)| \right) \quad m = 1, 2, \ldots$$ (10)

where $X_M(x, y; n)$ is the pixel value at location $(x, y)$ in the $n$th moving macroblock of the current frame and $X^R_M(x, y; n)$ is that of the reference frame in the same position. Similarly, $X_S(x, y; m)$ and $X^R_S(x, y; m)$ are also the pixel values in the $m$th still macroblock.

3. **$MAX_k$, $MDX_k$, $MAY_k$, $MDY_k$, $k \in \{P, B\}$** are the average and variance of the horizontal and vertical relative difference between the current motion vector $MV(x, y)$ and re-estimated motion vector $MV_0(x, y)$, respectively.

$$MAX_k = \frac{1}{n} \sum_x \sum_y F_H(k; x, y)$$ (11)

$$MDX_k = \frac{1}{n} \sum_x \sum_y (F_H(k; x, y) - MAX_k)^2$$ (12)
\[
MAY_k = \frac{1}{n} \sum_{x} \sum_{y} F_V(k; x, y) 
\]

\[
MDY_k = \frac{1}{n} \sum_{x} \sum_{y} (F_V(k; x, y) - MAY_k)^2 
\]

where \(F_H(k; x, y), F_V(k; x, y)\) are the horizontal and vertical relative difference at position \((x, y)\) in an inter frame of type \(K\), defined as follows.

\[
F_H(k; x, y) = \frac{MVH(k; x, y) - MVH_0(k; x, y)}{MVH_0(k; x, y)} 
\]

\[
F_V(k; x, y) = \frac{MVV(k; x, y) - MVV_0(k; x, y)}{MVV_0(k; x, y)} 
\]

where \(MVH(k; x, y)\) and \(MVV(k; x, y)\) are the horizontal and vertical components of \(MV\) at location \((x, y)\) in a frame of type \(k\), and \(MVH_0(k; x, y)\) and \(MVV_0(k; x, y)\) are those of \(MV_0\). In order to evaluate the performance of motion estimation algorithms, we utilize a full search algorithm which is similar to the scheme in the TM5, to re-estimate and get the optimal motion vector \(MV_0(x, y)\) for each motion block in the reconstructed frames. Finally, we use the distance between two types of motion vectors as a feature to distinguish different encoders.

\(4\) \(MC\), matching criterion feature is defined as (17).

\[
MC = \frac{1}{m} \sum_{x} \sum_{y} R_m(x, y) 
\]

where \(R_m(x, y)\) is the binary representation of difference in blocks in the \(m\)th \(P\) frame.

\[
R(x, y) = \begin{cases} 
1 & \text{if } \min_{i,j}(MAE(i + MV_h, j + MV_v)) \\
= MAE(MV_h, MV_v) & i, j = -1, 0, 1 \\
0 & \text{otherwise}.
\end{cases} \quad (18)
\]

Calculate Mean Absolute Error (MAE) between the current block and these blocks which surround the reference block to obtain the distortion measure factor \(R(x, y)\). \(MC\) can reflect which distortion measure the encoder takes.

\(3.2.\) \textit{SVM classification and vote decision}

Support vector machine (SVM) is a statistical classification method proposed by Vapnik in 1995,\(^{27}\) whose basic idea is to map data into a higher dimensional space and find a separating hyperplane with the maximal margin. Given a labeled training set of two classes:

\[
S = \{(x_i, y_i) | x_i \in \mathbb{R}^n, y_i \in \{-1, 1\}, i = 1, 2, \ldots, m\} 
\]
where $x_i$ is a training vector and $y_i$ is the class label, SVM solves a quadratic optimization problem:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{m} \xi_i,$$

subject to $y_i (w^T \varphi(x_i) + b) \geq 1 - \xi_i$, $\xi_i \geq 0$, $i = 1, \ldots, m$

(20)

where training data are mapped into a higher dimensional space by the function $\varphi$, and $C$ is a penalty parameter on the training error. For a testing instance $x$, the decision function (predictor) is

$$f(x) = \text{sgn}(w^T \varphi(x) + b)$$

(21)

Practically, the kernel function $K(x, x') = \varphi(x)^T \varphi(x')$ is just needed to train the SVM. The RBF kernel is used in the experiment:

$$K(x, x') = \exp(-\gamma \|x - x'\|^2)$$

(22)

The hyper-parameter pair $(C, \gamma)$ is determined as

$$\arg \min_{(C,\gamma) \in G} \text{Error}(C, \gamma)$$

(23)

where $\text{Error}(C, \gamma)$ is the error estimated by ten-fold cross-validation and $G$ is a multiplicative grid $G = \{(2^i, 2^j)|i, j \in \mathbb{Z}\}$. To overcome the problem that the set $G$ is unbounded, we exploit the fact that, the error surface of SVM estimated with cross-validation is convex for most practical problems. After the initial search whose range was chosen as a common one to all SVMs, we checked if the point with the least estimated error was at the boundary of the grid. If so, a larger search continued in the direction perpendicular to the boundary where the best point was laid until the best one was found within the explored grid. With this method, a small distance between the best point and the optimal point was ensured. Before training, all of the elements of the feature vector are scaled to the interval $[-1, 1]$ so as to be trained under the condition of normalization.

One-against-one algorithm\textsuperscript{8,13} was used here in order to implement multiclass classification, whose basic idea is to construct one binary classifier for every pair of classes. A testing instance $x$ is labeled according to the well-known strategy max-wins.

In this experiment, our implementation of classifier is based upon SVM with a RBF kernel, which is trained with $N$ classes training data based on the combined features of bit stream, quantization parameters and motion vectors.

4. Experimental Results and Analysis

In this section, eight MPEG-2 encoders are selected as subjects, including DV-Canon FS10E, DV-Sony HDR-XR500E, the default MPEG-2 encoder in Adobe Premiere
Pro 2.0, ImToo MPEG Encoder (ImToo) 5.1, ImToo Video Encoder (WinAVI) 9.0, the MPEG-2 encoder in Nero Ultra Edition (Nero) 8.0, and Test Model 5 (TM5) of MPEG-2. The first two are the popular DV in the market and the last six encoders are the popular MPEG-2 software encoders on the internet. 800 test sequences are obtained in two ways. 100 YUV sequences, containing 20 standard test sequences from Video Quality Experts Group (VQEG), and the others from high definition DVD, have been encoded respectively by the six software MPEG-2 encoders under the same basic coding condition that is output resolution \((720 \times 576)\), const bit-rate (6 Mbits/s), and frame rate (25 f/s). On the other hand, we use each DV to record 100 nature video clips with length of 300 frames in our campus, which are initially captured in MPEG-2 format at 6 Mbps. In total 800 MPEG-2 video streams, each GOP is defined as a sample. The MPEG-2 software encoder may introduce abnormal control mode, such as skipping frame, to maintain the steady output of bit-rate when the buffer meets overflowing or underflowing. Our stream analyzer will delete these abnormal GOPs according to buffer status. In all, we prepare about 19500 samples for SVM training and testing. In our experiments, the ratio of training samples to testing samples is 1:1.

4.1. The parameters setting in MPEG-2 encoders

Compared with source camera identification, video encoders have more optional parameters, especially in software coding system. In the digital video camera, the most important parameters are resolution and record mode (i.e. the output bit-rate), and the other parameters, such as automatic exposure, white balance, flash mode and focus, just only affect the subjective display effects of video resources. Thus, the standard definition video format is selected as our encoding mode, where resolution is \(720 \times 576\) and output bit-rate is 6 Mbps, and other parameters are all as default. In the MPEG-2 software encoders, most of parameters are set similar to the initialization value of MPEG-2 encoders in DVs, as shown in Table 1, and the other parameters are set as the system default values.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV standard</td>
<td>PAL (720 × 576)</td>
</tr>
<tr>
<td>Frame Rate (f/s)</td>
<td>25</td>
</tr>
<tr>
<td>Pixel aspect ratio</td>
<td>4:3</td>
</tr>
<tr>
<td>Profile</td>
<td>Main</td>
</tr>
<tr>
<td>Level</td>
<td>Main</td>
</tr>
<tr>
<td>Bit rate (Mbps)</td>
<td>6</td>
</tr>
<tr>
<td>GOP Setting</td>
<td>(M = 3, N = 12)</td>
</tr>
<tr>
<td>VBV buffer size (16kbit)</td>
<td>112</td>
</tr>
</tbody>
</table>
4.2. Video source identification results

Performance of the identification algorithm is measured in terms of recall and precision which are defined as follows.

\[
\text{Precision} = \frac{C}{C + F}, \quad (24)
\]

\[
\text{Recall} = \frac{C}{C + M}, \quad (25)
\]

where \(C\) is the number of correctly detected caption frames or characters, \(F\) is the number of false alarms, and \(M\) is the number of misses.

Tables 2 and 3 respectively give the identification performance and the confusion matrix of our proposed algorithm at constant bit-rate 6 Mbits/s. The results show that our identification algorithm can effectively distinguish different sources of video streams. During in-depth comparison of identification performance of different MPEG-2 encoders, we find that, when some MPEG-2 encoders introduce special schemes or settings to meet system requirements, such as real-time or limited size of buffer, these detection precision and recall are very good; while the MPEG-2 software encoders adopt the similar algorithm in the open modules, these precision and recall decrease slightly.

Table 2. Performance of the proposed identification method at 6 Mbits/s.

<table>
<thead>
<tr>
<th>MPEG2-Encoder</th>
<th>Precision%</th>
<th>Recall%</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM5</td>
<td>91.30</td>
<td>92.40</td>
</tr>
<tr>
<td>Premiere</td>
<td>86.73</td>
<td>82.72</td>
</tr>
<tr>
<td>CCE</td>
<td>89.99</td>
<td>94.56</td>
</tr>
<tr>
<td>ImToo</td>
<td>98.31</td>
<td>88.24</td>
</tr>
<tr>
<td>WinAVI</td>
<td>88.81</td>
<td>93.36</td>
</tr>
<tr>
<td>Nero</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Canon</td>
<td>97.52</td>
<td>95.92</td>
</tr>
<tr>
<td>Sony</td>
<td>93.48</td>
<td>95.24</td>
</tr>
</tbody>
</table>

Table 3. Confusion matrix of the proposed identification method at 6 Mbits/s.

<table>
<thead>
<tr>
<th>Predicted</th>
<th>TM5%</th>
<th>Premiere%</th>
<th>CCE%</th>
<th>ImToo%</th>
<th>WinAVI%</th>
<th>Nero%</th>
<th>Canon%</th>
<th>Sony%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TM5</td>
<td>92.40</td>
<td>5.84</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.76</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Premiere</td>
<td>3.76</td>
<td>85.52</td>
<td>7.68</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3.04</td>
</tr>
<tr>
<td>CCE</td>
<td>0.00</td>
<td>3.60</td>
<td>94.56</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.84</td>
</tr>
<tr>
<td>ImToo</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>88.24</td>
<td>11.76</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>WinAVI</td>
<td>3.68</td>
<td>0.00</td>
<td>1.44</td>
<td>1.52</td>
<td>93.36</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Nero</td>
<td>0.00</td>
<td>0.00</td>
<td>0.72</td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Canon</td>
<td>0.00</td>
<td>1.60</td>
<td>0.68</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>95.92</td>
<td>1.76</td>
</tr>
<tr>
<td>Sony</td>
<td>1.36</td>
<td>2.04</td>
<td>0.68</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.68</td>
<td>95.24</td>
</tr>
</tbody>
</table>
To verify the detection performance under different parameter settings in the MPEG-2 coding systems, the bit rate at Table 1 is adjusted to 9 Mbps, and the detection performance is shown in Tables 4 and 5. In comparison with the experimental results at 6 Mbps, the results at 9 Mbps are not as good but still very promising. In our opinion, when the output bit-rate increases, some constraints in the rate control and motion estimation algorithms are relaxed, weakening the discriminative powers of some features. Thus, the performance in detecting these MPEG-2 encoders declines. On average, our detection system still delivers good performance.

### 4.3. Discussion

In the Ref. 26, the Photo Response Non-Uniformity (PRNU) is used to identify the source video cameras with the same video codec, but the codec and their parameters are assumed to be known conditions, which may not be suitable in the identification of different video coding systems. Overcoming the weakness, our proposed algorithm can effectively identify the video streams that come from a number of video coding systems.

In our identification technique, all features are extracted from the open modules in the MPEG-2 video coding systems. In these open components, the codec designers

<p>| Table 4. Performance of the proposed identification method at 9 Mbits/s. |
|---------------------------------|-------------|-------------|</p>
<table>
<thead>
<tr>
<th>MPEG2-Encoder</th>
<th>Precision%</th>
<th>Recall%</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM5</td>
<td>88.46</td>
<td>88.96</td>
</tr>
<tr>
<td>Premiere</td>
<td>85.76</td>
<td>85.28</td>
</tr>
<tr>
<td>CCE</td>
<td>86.93</td>
<td>91.52</td>
</tr>
<tr>
<td>ImToo</td>
<td>93.69</td>
<td>87.84</td>
</tr>
<tr>
<td>WinAVI</td>
<td>86.39</td>
<td>90.40</td>
</tr>
<tr>
<td>Nero</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Canon</td>
<td>93.15</td>
<td>91.36</td>
</tr>
<tr>
<td>Sony</td>
<td>95.03</td>
<td>93.36</td>
</tr>
</tbody>
</table>

<p>| Table 5. Confusion matrix of the proposed identification method at 6 Mbits/s. |
|---------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|</p>
<table>
<thead>
<tr>
<th>Predicted</th>
<th>Actual</th>
<th>TM5%</th>
<th>Premiere%</th>
<th>CCE%</th>
<th>ImToo%</th>
<th>WinAVI%</th>
<th>Nero%</th>
<th>Canon%</th>
<th>Sony%</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM5</td>
<td>88.96</td>
<td>3.76</td>
<td>1.76</td>
<td>1.84</td>
<td>1.92</td>
<td>0.00</td>
<td>1.76</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Premiere</td>
<td>4.08</td>
<td>85.28</td>
<td>7.36</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.68</td>
<td>1.60</td>
</tr>
<tr>
<td>CCE</td>
<td>0.00</td>
<td>3.76</td>
<td>91.52</td>
<td>0.00</td>
<td>1.76</td>
<td>0.00</td>
<td>1.52</td>
<td>0.00</td>
<td>1.44</td>
</tr>
<tr>
<td>ImToo</td>
<td>1.60</td>
<td>0.00</td>
<td>0.00</td>
<td>87.84</td>
<td>10.56</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>WinAVI</td>
<td>4.08</td>
<td>0.00</td>
<td>1.44</td>
<td>4.08</td>
<td>90.40</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Nero</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Canon</td>
<td>1.84</td>
<td>3.36</td>
<td>1.60</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>91.36</td>
<td>1.84</td>
<td></td>
</tr>
<tr>
<td>Sony</td>
<td>0.00</td>
<td>3.28</td>
<td>1.60</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.76</td>
<td>93.36</td>
</tr>
</tbody>
</table>
introduced their unique and private algorithms to improve the compression performance. Apart from a few open source MPEG-2 codec (such as TM5), we cannot find the design details of these algorithms. Therefore, most of the features are obtained directly by reverse analyzing and summarizing a variety of different source video streams.

It is worthwhile to note that the classification performance of each kind of feature is not the same for the identification of different video coding systems with different activities in video streams. In some video shots with very low activity, the number of motion vectors is too few to describe the characteristics of different encoders, but the variance of bit-stream is relatively steady, so the discriminative power of bit-stream features is much higher than that of motion vector features. On the other hand, in some high activity video shots, the bit-stream features are unstable, but the motion vector features can be used as important evidence in identifying the video coders. Combining these features, it can effectively improve the classification accuracy and expand the adaptability of our detection scheme.

5. Conclusion and Future Work

Digital video camera system is similar to digital image acquisition, which consists of optical acquisition and digital processing. Because of the huge volume of digital video data, the compression coding is indispensable for digital processing. In the meantime, most of the characters on the front-end camera system may be covered up. So the essence of source video identification is to identify the encoding algorithm in the coding system. We first recognize different coding systems and classify the type of coding system, and then build a respective model for each coding system for detection.

In this paper, we have proposed a video source identification algorithm based on features in the MPEG-2 encoded stream, which had been verified by using eight types of coding system, and provided us a clue to resolve the above problem. In the open modules of the video compression standard, each applied video system may design some particular scheme to improve the performance of the whole system, so that we can extract some features in these modules to identify the video system.

In the future, we will find other kinds of fine features in open components to increasingly improve the performance of identification algorithm. We will collect many other kinds of video compression systems, and set up the feature database for each one, in order to increase the number of video systems which can be identified by our algorithm. Finally, we will analyze in-depth the distinction among different video systems to provide reliable information for double compression detection techniques.

Feature selection is an important issue in classification. Some good feature selection algorithms have been proposed to deal with the high dimensionality in bioinformatics study. Some of them have been adopted in steganalysis to improve the detection performance. In our future study, the detection performance under different feature selection methods will be conducted in order to produce an optimal identification system.
Acknowledgments

This work was supported in part by the National High Technology Research and Development Program of China (“863” program, No.2006AA01Z407), in part by the Tianjin Municipal Natural Science Foundation (No.07JCZDC05800). All authors are very grateful to anonymous reviewers and the Editor for their insightful comments and very helpful suggestions.

References


33. Q. Liu, A. Sung and M. Qiao, Neighboring joint density based JPEG steganalysis, *ACM Transactions on Intelligent Systems and Technology*, in press.
Yuting Su received the B.Sc., M.Sc. and Ph.D. degrees in electronic engineering from Tianjin University, in 1995, 1998 and 2001, respectively. Currently, as an associate professor, he has been with the School of Electronic Information Engineering in Tianjin University, China.

His research interests include digital video coding, digital watermarking and data hiding, multimedia forensics, and multimedia retrieval.

Junyu Xu received the B.S. degree in electronic engineering and the M.S. degree in signal and information processing from Tianjin University, Tianjin, China, in 2008 and 2010, respectively. Currently, he is pursuing the Ph.D. degree in information and communication engineering from Tianjin University.

His research interests include video processing and multimedia forensics.

Bo Dong received the B.S. degree in electronic information engineering from Tianjin University, Tianjin, China, in 2009. Now he is pursuing the M.E. degree in information and communication engineering in Tianjin University.

His research interests focus on digital video forensics and digital video processing.

Jing Zhang received the B.E. and the M.E. degrees from Sichuan University, Chengdu, China, in 1994 and 1997 respectively, both in measurement & control, and the Ph.D. degree from Tianjin University, Tianjin, China, in 2002, in electrical engineering.

From September 2002 to September 2004, she was with Tianjin University as a post-doc in electrical science and technology. She joined the faculty at School of Electronic Information Engineering, Tianjin University, in September 2004.

Her current research interests include image and video processing, digital forensics, information hiding and digital watermarking.

Qingzhong Liu received his B.E. degree in material science and engineering in 1993 from Northwestern Polytechnic University, China, M.E. degree in computer science and engineering in 1997 from Sichuan University, China, and Ph.D. degree in computer science in 2007 from New Mexico Institute of Mining and Technology, a Center of Academic Excellence in Information Assurance Education and Research, designated by U.S. National Security Agency and Department of Homeland Security, wherein he worked as a senior research scientist and adjunct faculty for three years after he obtained his Ph.D. degree. He is currently a tenure-track assistant professor at the Department of Computer Science of the Sam Houston State University.

His research interests include data mining, pattern recognition, image processing and multimedia computing, information security, digital forensics and bioinformatics.