Effective Content Tracking for Digital Rights Management in Digital Libraries

Jen-Hao Hsiao, Cheng-Hung Li, Chih-Yi Chiu, Jenq-Haur Wang, Chu-Song Chen, and Lee-Feng Chien

Institute of Information Science, Academia Sinica, Taipei, Taiwan
{jenhao, chli, cychiu, jhwang, song, lfchien}@iis.sinica.edu.tw

Abstract. A usual way for content protection of digital libraries is to use digital watermarks and a DRM-based access-control environment. These methods, however, have limitations. Digital watermarks embedded in digital content could be removed by malicious users via post-processing, whereas DRM-based access-control solutions could be hacked. In this paper, we introduce a content tracking mechanism that we have built for multimedia-content near-replica detection as the second line of defense. The integrated framework aims to detect unlawful copyright infringements on the Internet, and combines the strengths of static rights enforcement and dynamic illegal content tracking. The issues of accuracy and huge computation cost in copy detection have been addressed by the introduced content-based techniques. Our experiments demonstrate the efficacy of proposed copy detector.

1 Introduction

Protection of the copyrights and revenues of content owners in digital libraries has become increasingly important in recent years. Since digital content differs from objects in real world, it can be easily copied, altered, and distributed to a large number of recipients. This almost certainly causes copyright infringement and therefore revenue losses to content owners. The National Digital Archives Program (NDAP) of Taiwan has amassed a rich collection of cultural and historical artifacts. These assets have been digitized to enhance their preservation, and make them more accessible to users. The metadata and digital content storage systems are called archival systems, and – like other types of digital content – they too face the problem of piracy. Thus, content holders are sometimes unwilling to release digital content, because their intellectual property rights could be infringed.

To prevent the abuse of digital content, a number of approaches have been proposed. Digital watermarking is the most widely used form of copy protection. A digital watermark, which is an identification code that carries information about the copyright owner, is invisible and permanently embedded in digital data for copyright protection, proof of ownership, and integrity checks of digital content. It can also provide evidence of copyright infringement. Though useful, watermark-based protection systems have some significant limitations. First, watermarking could degrade the quality of digital content. Second, embedded watermarks are not expected to survive

J. Gonzalo et al. (Eds.): ECDL 2006, LNCS 4172, pp. 415 – 425, 2006.

under several kinds of attack. In practice, although many techniques have been proposed, watermark-based techniques are not robust enough to prevent malicious users removing watermark via post-processing.

The Digital Rights Management (DRM) system is another popular method for protecting high-value digital assets. DRM is a protocol of hardware and software services and technologies that governs the authorized use of digital content and manages its use throughout the entire life-cycle of the content (as defined by IDC [3]). The primary objective of DRM is to build a DRE (digital right enforcement) environment that only allows access to protected content under the conditions specified by the content owner. Many DRM and DRE frameworks [3][14][16][17] have been proposed in recent years. Although these architectures provide a way to construct a copyright protection environment, the security of digital content is not fully addressed. For example, in the area of rights enforcement, authorized users could still distribute digital assets easily after they pass the identity authentication process. Hence, how to enforce the usage rules and protect content owners' property rights after digital content have been released is still a challenging aspect of DRM research.

Recently, the concept of content-based copy detection has been proposed as a complementary solution for traditional DRM systems. The idea is that, instead of hiding additional information in the digital content (such as digital images and videos) for copy detection, the content itself can be employed. A content-based copy detection system works as follows. It starts by extracting features from the original content, and compare them with features extracted from a suspicious to determine whether the latter is a copy. Content-based copy detection itself can be used to identify illegal copies, or it can be used to complement digital watermarking techniques.

In the past, existed content-based image copy detection techniques [4][11][13] emphasize on finding unique image features with good performance that could resist a variety of image attacks, but finding a globally effective feature is difficult, and in many situations, domain dependent. Hence, the accuracy of image copy detectors is still restricted.

With respect to video copy detection, most approaches [7][8][12][19] employ high-cost computation techniques to match videos, whereby a fix-sized window that slides frame by frame is used to detect copies. However, the window cannot handle some temporal variations, e.g., fast and slow motion. These drawbacks inevitably impede the practicability of the system.

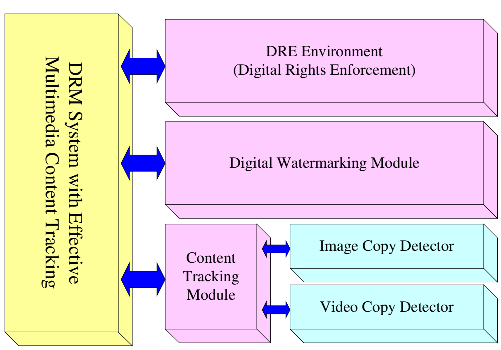

Fig. 1. Overview of the proposed system

As current copy protection technologies have certain limitation, in this paper, we seek to address the problem by introducing a novel architecture that integrates a DRM system with an effective content tracking mechanism to discourage attackers and further strengthen the proposed system's security. The remainder of this paper is organized as follows. Section 2 introduces our main framework. Sections 3 and 4 describe the proposed content-based copy detection methods in detail. In Section 5, we present the experiment results and demonstrate the effectiveness of content tracking mechanism. Our conclusions are presented in Section 6.

2 Effective Content Tracking for DRM system

Most of the valuable digital content to be protected in archival systems consists of multimedia objects, such as digital images and videos. Figure 1 gives an overview of the proposed DRM system, which consists of three building blocks: (1) the DRE (Digital Rights Enforcement) Environment, (2) the Digital Watermark Module, and (3) the Content Tracking Module. First, the system packages the content to be protected in a secure manner, and the DRE environment ensures that the usage rules are enforced. We use a wrapper-based DRE technique [10] to protect the digital rights. When a user downloads digital content from the network and views it on a player (e.g., a browser), the wrapper automatically monitors the user's behavior. If the rules are violated, or the user refuses to be monitored by the wrapper, the content is rendered unavailable. The second component, the digital watermarking module, can embed an invisible digital watermark into digital content. If necessary, the content holder can extract the watermark to prove ownership if there is copyright infringement. The third component, the content tracking module, can be regarded as the second line of defense. It is composed of two key kernels: an image copy detector and a video copy detector, which can determine whether or not suspicious digital content is copyrighted (registered). By integrating a web crawler with the content tracking module, illegal use of digital content on the Internet can be detected automatically.

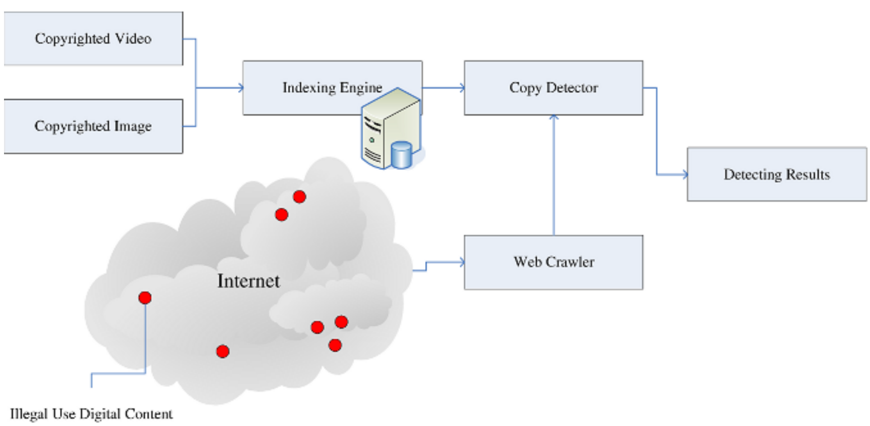

Fig. 2. Block diagram of copy detection

As shown in Figure 2, the content module first registers the image/video with the database. Only feature vectors are stored in the database in order to accelerate the detection process and reduce the amount of storage space required. The image/video copy detector then conducts a matching process to determine whether the suspicious digital content grabbed by web crawler is copyrighted.

3 Image Copy Detection

Previous researchers have tried to find an image feature that can be employed universally for copy detection. Various features have been studied, for example, local [1][18], global [2][11], DCT-based [2], wavelet-based [4][13], geometrically variant [2][4], and invariant [1][13][18]. Obviously, the accuracy of existing copy detectors relies heavily on the robustness of the feature used, and on a suitable threshold that can balance the false rejection and false acceptance rates. However, as we know that it is difficult to find a unique feature that is invariant to various kinds of attack. Another limitation of existing approaches is that they lack a mechanism to exploit useful priori information, such as possible attack models, to boost the copy detection performance – even when such information is easy to generate or acquire.

Hence, instead of extracting the feature vector from a copyrighted image, we use virtual attacks as prior guidance to conduct a new copy-detection framework[9]. Typical attacks considered in our approach include signal-processing attacks, geometric attacks, and image-compression attacks. By applying the attacks to a copyrighted image, a set of novel images can be generated. Both the copyrighted and novel images are processed by extracting their features, where the features extracted from the former and the latter are referred to as the *original* and *extended features* in our framework, respectively. Figure 3 shows the concept of copy detection in a 2-dimensional space. In Figure 3(a), I denotes the feature vector of a copyrighted image, and A, B, and C are the copyrighted images under some malicious attacks. The radius of cluster ε denotes the error tolerance for copy detection in the feature space, which is decided by a predefined threshold. It often occurs that some attack, say A, can be successfully resisted, but the others more severe ones B and C cannot be detected since they are far away from A in the feature space. In our experience, this problem is difficult to solve in practice by simply changing the features being used. Figure 3(b) shows the concept of using *EFS* (extended feature set) to enhance the performance of copy detection, where the gray points denote the extended features. In this case, the problem can be solved by grouping features so that the modified images A, B, and C can be identified correctly.

Although modeling copy detection as a one-class classification problem is likely to boost the system's performance, many empirical studies of pattern classification reveal that the classifier can be trained better if much more prior knowledge is given. In particular, if some negative examples are available, using them would help build a better classifier than using only positive examples. Therefore, in our approach, not only positive examples (where they are mainly extended features), but also negative examples are used. The negative examples are easy to acquire or generate; for example, they can be obtained from the Internet. Also, a registered image can serve as a negative example of another registered image. Our framework transforms the

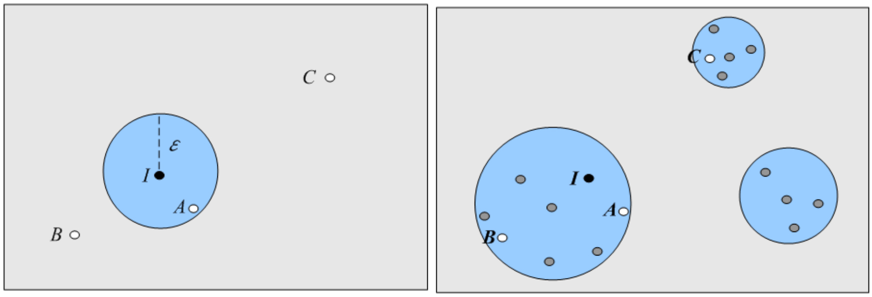

Fig. 3. (a) A typical image copy detection algorithm. (b) Using EFS to solve the problem in (a).

copy-detection problem into a two-class classification problem. We demonstrate by experiments that our approach generally outperforms the conventional technique when the same feature space is employed.

A popular method for solving the two-classification problem is based on GMM (Gaussian Mixture Model), defined as:

$$f_k(x \mid \theta) = \sum_{j=1}^{k} w_j g(x \mid \lambda_j),$$

where $g(x|\lambda_j)$ is a multivariate Gaussian distribution, $\lambda = (u,\Sigma)$ is the Gaussian component parameter set, w_j is the weight of jth component, k is the number of Gaussian components, and $\theta=\{w_j, \lambda_j \mid j = 1,2,\ldots,k\}$ is the model's parameter set.

To learn the GMM model for each class, we apply the expectation-maximization (EM) algorithm that can converge to a maximum likelihood estimation of the parameter set. The selection cluster number k is a critical factor in training a GMM [6]. Since we have prior categorical knowledge about our training data, the number of clusters can be set, in advance, as the number of attacks we would like to model. To improve the accuracy, k can also be assigned automatically by maximizing the log-likelihood of the training samples, and estimated via cross-validation. In our approach, we initially set k as the category number, and continue adding clusters until the log-likelihood either (1) starts to decline or (2) keep on increasing but with an amount less than a specific threshold. In Section 5, we conduct some experiments to examine the performance of the proposed framework when a Gaussian mixture classifier is used.

4 Video Copy Detection

The problem definition of video copy detection is to determine if a given video clip (query) appears in another video clip (target) which is doubtful. However, if it does appear, we need to determine its location. The proposed video copy detection module is responsible for three steps: key frame extraction, candidate clip selection, and sequence matching. Suppose that Q_V and T_V are the query and target video clips, respectively. Q_V is represented as $\{q_{vj} \mid j = 1, 2, \ldots, N\}$, and T_V as $\{t_{vi} \mid i = 1, 2, \ldots, M\}$, where M and N are the number of frames, $M >> N$ and t_{vi} and q_{vj} are the ordinal signatures of the corresponding frames. The details of the ordinal signature are as follows: A video frame is partitioned into $n_x \times n_y$ blocks and the average luminance level in each block is computed. In our case, we utilize 3×3 block of each frame for ordinal signature extraction. Then we sort the set of average intensities in ascending order and a rank is assigned to each block. The ranked $n_x \times n_y$ dimensional sequence is then generated [7][8][12][19]. Thus a video frame is partitioned into 3×3 blocks , as its ordinal signature a 3×3 matrix. We then reshape the matrix to a 9×1 vector. Based on the steps mentioned above, the task of copy detection is to find the subsequences from T_V, whose signature series are similar to those of Q_V .

The first step is to extract key frames from video clips. In addition to reduce the storage and computation costs, it can moderate the effects of temporal variations. Let us take the target clip T_V as an example. In order to search the peak or foot of a sequence, we define a 9×9 Laplacian of a Gaussian filter F, which is often used to calculate second order derivatives in a signal:

$$F(x, y) = -\frac{1}{\pi\sigma^4}\left|1 - \frac{x^2+y^2}{2\sigma^2}\right| e^{\frac{x^2+y^2}{2\sigma^2}},$$

The second order derivatives reveal signal transitions, which can be chosen as key frames.

We then convolute F and T_V to obtain a vector A and find the local extreme on A, as shown in Figure 4. The extracted key frames are denoted as $T_K = \{t_{k1}, t_{k2}, \ldots, t_{km}\}$. For the query clip Q_V, we repeat the above procedure to extract Q_V*'s* key frame sequence $Q_K = \{q_{k1}, q_{k2}, \ldots, q_{kn}\}$.

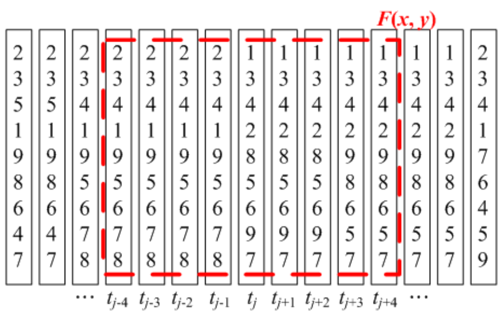

Fig. 4. The convolution of the filter F and the target T. The dash square indicates the range of F, and t_j is the ordinal signature of the j-th frame in T.

After the key frames has been extracted, the key frame sequence of T_K is still very long. To avoid an exhaustive search of the long sequence, we roughly scan T_K to find subsequences that may be copies of Q_K. First we search for the start and end indices of candidates CI_{start} and CI_{end} in *TK*. These candidates are frames that are similar to the first and end frames of Q_K (i.e., q_{k1}). Then we scan the second candidate lists CI_{start} and CI_{end}. A subsequence $C = \{t_{ks}, t_{ks+1}, \ldots, t_{ke}\}$ in T_K is reported as a candidate clip according to following conditions: First, keep the order of the start and end candidates. Second, select the smallest frame set from the candidate combinations. Third, filter out clips that are too long or too short.

Finally, the sequence matching be processed to compute which clip is similar as copy, hence the Dynamic Time Warping (DTW) algorithm is applied to compute the similarity between the query example Q_K and the candidate clip C. Since DTW can compensate for differences in length, it is suitable for dealing with video temporal variations in videos. We define the following distance function:

$$dist(Q_K, C) = cost(n, l),$$

where n and l are the frame number of Q_K and C respectively, and $cost(n, l)$ is a recursive function:

$$cost(1,1) = \|qk_1 - tk_1\|,$$
$$cost(n,l) = \|qk_n - tk_l\| + min\{cost(n-1,l), cost(n,l-1), cost(n-1,l-1)\},$$

where $\|l - n\| \leq$ the maximum warping distance, which is normalized to determine whether Y is a copy of Q_K.

5 Experiment Results

In this section, we conduct two experiments for evaluating the performance of content-based copy detection. We divide the experiments into two cases. In the first case, the detection results of image content tracking are presented; while in the second, video data from the National Digital Architecture Program in Taiwan is used to verify the effectiveness of our method.

5.1 Image Detection Results

We took Kim's approach – DCT ordinal measure [11] as the basis for comparison. In this approach, an input image is divided into 8×8 equal-sized sub-images. Only AC coefficients of the 8×8 DCT coefficients are used as the ordinal measure. We thus generated a 63-dimensional image feature vector.

In the first test, one hundred copyrighted images were registered in the database and used as queries to determine how many modified versions could be detected successfully. A standard benchmark, Stirmark 4.0 [15], was used to generate novel testing data. The image replicas were randomly generated by StirMark 4.0 with 7 categories of pre-learned image attacks (convolution filtering; cropping; JPEG; median filtering; noise adding; scaling; and rotation), and 6 categories of novel attacks, including affine transformation, self-similarity, removal of lines, PSNR, rotation+rescaling (abbreviated as RRS), and rotation+cropping (abbreviated as RC). We also generated 124 near-replicas for each copyrighted image. In addition to image replicas, the testing data also contained 15,000 randomly picked unrelated images, giving a total of 27,400 images for testing.

To evaluate the performance, the precision rate, recall rate, and F-measure are used:

$$F-Measure = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}}.$$

The results in Table 1 show that our framework outperforms that of the DCT ordinal measure. The EFS for the Gaussian mixture model achieve very high precision and recall rate of 96.56% and 93.54% respectively, while the F-measure is 95.03.

Table 1. Average precision and recall rates by using extended features and pure DCT ordinal measures. The response time consists of both the feature extraction and classification times.

Algorithm	Avg. Precision	Avg. Recall	F-Measure
DCT ordinal measures	93.13%	54.79%	68.99
Gaussian Mixture Classifier with EFS	96.56%	93.54%	95.03

Fig. 5. Three color images in the digital museum (512*512 pixels): a container, a rare book and a painting were chosen in the second experiment

Table 2. Recognition rates of the Gaussian Mixture Classifier (including novel attacks): The first column indicates whether the type of image attack was pre-learned, while the second column shows the attack model and how many times it was applied. For example, " Noise * 12" means that the noise attack was applied to the image 12 times. In the other columns, "m(n)" indicates that the number of image replicas successfully detected by our Gaussian mixture method and by the pure DCT ordinal measures method was m and n respectively.

Pre-learned	Testing Item	Container	Rare Book	Painting
∨	Convolution Filtering * 2	2(2)	2(2)	2(2)
∨	JPEG * 14	14(14)	14(14)	14(14)
∨	Median Filtering * 4	4(4)	4(4)	4(4)
∨	Noise * 12	12(10)	12(9)	12(8)
	Self-Similarities * 3	3(3)	3(3)	3(3)
	PSNR * 10	10(10)	10(10)	10(10)
∨	Scaling * 10	10(10)	10(10)	10(10)
∨	Cropping * 13	9(1)	8(2)	11(0)
∨	Rotation * 18	17(0)	16(0)	18(0)
	Affine * 8	7(7)	8(6)	7(6)
	Removing Lines * 10	10(8)	10(8)	10(8)
	RRS * 10	9(1)	9(1)	9(0)
	RC * 10	9(1)	9(1)	9(0)
Recognition Rate (DCT ordinal measures)		(71+70+65) / (124*3) = 55.38%		
Recognition Rate (Gaussian Mixture Classifier with EFS)		(116+115+119) / (124*3) = 94.07%		

The above experiment shows the overall performance of our method. To test the robustness against different attacks, we conducted another smaller-scale experiment in which only the three images shown in Figure 5 were used. This allowed us to show the comparisons of the performance of EFS with conventional copy detection method in more detail. The results are summarized in Table 2. We also applied some novel attacks (i.e., attacks not modeled in the training phase) to examine the performance of our approach. The results show that the images' resistance to geometric attacks (cropping, rotation, scaling) was significantly enhanced by our approach; on average, more than half the manipulated geometric images were correctly identified in the experiment. In Table 2, the first column indicates whether the image attacks were prelearned. Clearly, for those novel attacks we did not model in advance, our approach still achieves an acceptable performance and outperforms the pure DCT ordinal measure method.

5.2 Video Detection Results

We experimented with approximately 106,333 frames of video data from the NDAP's digital video library of social culture in Taiwan. The format of the videos is MPEG-1 NTSC, for which the resolution is 352×240 and frame rate is 29.97 fps. To test the performance of the proposed approach, the video data was modified to generate eight copies for brightness, histogram equalization, changing the resolution to 176×120, changing the frame rate to 15 fps and 10 fps, slow motion (0.5×), fast motion (2×), and hybrid modification (changing to 176×120 resolution, 10 fps, and 2× fast motion). We randomly selected 100 video clips (100×1000 frames in total) as query clips for each type of copy. Hence there are 800 queries in the experiment to verify the track performance in our video copy detection module.

Table 3. The F-measure of brightness, equalization, and frame size changing (spatial variations), and frame rate changing, slow and fast motion (temporal variations) copy in Hua's, Kim's and our proposed approach

	Brightness	**Equalization**	**176×120**	**10fps**	**15fps**	**0.5×**	**2×**	**Hybrid**
Hua	89.98	94.87	90.13	94.25	96.01	53.27	75.94	65.52
Kim	93.61	95.89	93.14	76.54	85.90	25.86	43.30	40.27
Proposed	94.26	96.19	94.24	93.87	95.60	83.55	94.06	83.38

We compared the results of the proposed approach with Hua's [7] and Kim's [11] approaches using F-measure. Table 3 shows the F-measure of all cases, and our approach outperforms the other two greatly. According to the experiment results, we see that our method performs slightly better than Hua's and Kim's for spatial variation attacks such as brightness, equalization and frame size change. For frame rate changes, our method performs better than Kim's but slight worse than Hua's. However, our method achieves a far better performance (average F-measure is 88.81) for the attacks of fast and slow motion than those of the others (average F-measure is 64.61 and 34.58). To conclude, our method has better performance in overall for the hybrid case, and is effective not only for spatial-variation but also temporal-variation attacks.

6 Conclusions

Protecting digital content presents serious technical challenges that the existing approaches have not overcome. The integrated framework presented in this paper provides a solution for digital content protection of digital libraries. With the wrapper-based DRE technique, a digital rights enforcement environment can be built to maintain the usage rules of digital content. With the help of such content tracking mechanism, pirated digital content altered from original images and videos can be effectively identified. Also, the introduced copy detection techniques have been demonstrated to be more accurate than traditional approaches. By employing such a complementary design, the abuse of valuable digital content can be prevented, and further reduce the copyright infringements.

Acknowledgement

This work was partially supported by National Science Council, Taiwan, under Grants NSC94-2422-H-001-006, NSC94-2422-H-001-007, NSC95-2422-H-001-007, and NSC95-2422-H-001-008.

References

1. S.-A. Berrani, L. Amsaleg, and P. Gros, "Robust content-based image searches for copyright protection," *Proceedings of the 1st ACM international workshop on Multimedia databases*, pp. 70-77, 2003.
2. D.-N. Bhat and S.-K. Nayar, "Ordinal measures for image correspondence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 20 , Issue 4, pp. 415-423, 1998.
3. C.P. Bogdan et al., "A DRM security architecture for home networks," *Proceedings of the 4th ACM workshop on Digital rights management*, pp 1-10, 2004.
4. E.-Y. Chang, J.-Z. Wang, C. Li, and G.. Wiederhold, "RIME: a replicated image detector for the world-wide-web," *Proceedings of the SPIE Multimedia Storage and Archiving Systems*, San Jose, CA, November 1998.
5. J. Duhl, S. Kevorkian, "Understanding DRM Systems," *An IDC Research White Paper*, 2001.
6. M.A.F. Figueiredo and A.-K. Jain, "Unsupervised learning of Finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 24 Issue: 3, pp. 381–396, March 2002.
7. A. Hampapur, K.-H. Hyun, and R. M. Bolle, "Comparison of sequence matching techniques for video copy detection," *The SPIE Conference on Storage and Retrieval for Media Databases*, pp. 194-201, 2002.
8. X. S. Hua, X. Chen, H. J. Zhang, "Robust video signature based on ordinal measure," *International Conference on Image Processing*, 2004.
9. J.-H. Hsiao, C.-S. Chen, L.-F. Chien, and M.-S. Chen, "Image Copy Detection via Grouping in Feature Space Based on Virtual Prior Attacks," *Proceeding of International Conference on Image Processing*, Atlanta, GA, USA, 2006.

10. J.-H. Hsiao, J.-H. Wang, M.-S. Chen, C.-S. Chen and L.-F. Chien, "Constructing a Wrapper-Based DRM System for Digital Content Protection in Digital Libraries," *Proceedings of the 8th International Conference on Asian Digital Libraries*, ICADL 2005, pp. 375-379, 2005.
11. C. Kim, "Content-based Image Copy Detection," *Signal Processing: Image Communication*, Vol 18, pp. 169-184, 2003.
12. C. Kim and B. Vasudev, "Spatiotemporal sequence matching for efficient video copy detection," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 15, No. 1, pp. 127-132, 2005
13. C.-S. Lu, C.-Y. Hsu, S.-W. Sun, and P.-C. Chang, "Robust mesh-based hashing for copy detection and tracing of images," *Proceedings of IEEE International Conference on Multimedia and Expo*, Volume 1, pp. 731-734, June 2004.
14. T. Nicolakis, C.E. Pizano, B. Prumo and M. Webb, "Protecting Digital Archives at the Greek Orthodox Archdiocese of America," *Proceedings of the 2003 ACM workshop on Digital rights management*, October 2003.
15. F. Petitcolas, "Watermarking Schemes Evaluation," *IEEE Signal Processing Magazine*, Volume 17, No. 5, pp. 58-64, 2000.
16. A.J. Pramod and L.H. Gregory, "DRM as a layered system," *Proceedings of the 4th ACM workshop on Digital rights management*, pp 11-21, 2004.
17. G. Susanne, "A Sample DRM System," *Lecture Notes in Computer Science*, Volume 2770, pp. 150-161, 2003.
18. K. Yan, R. Sukthankar, and L. Huston, "Efficient Near-duplicate Detection and Subimage Retrieval," *Proceedings of the 12th annual ACM international conference on Multimedia*, session 13, pp. 869-876, 2004.
19. J. Yuan, Q. Tian, and S. Ranganath, "Fast and robust search method for short video clips from large video collection," *International Conference on Pattern Recognition*, Cambridge, UK, 2004.