I.J. Information Technology and Computer Science, **2013, 04, 47-54**
Published Online March 2013 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2013.04.06

Detecting Pirated Movie's with Similarity Assessment Based on Earth Mover's Distance

Srinivas Baggam
Department of Computer Science and Engineering, M.V.G.R College of Engineering, India
srinio.b@mvgrce.edu.in

K.Venkata Rao
Department of Computer Science and Engineering, Vignan Institute of Information Technology, India
vrkoduganti@yahoo.co.uk

P. Suresh Varma
Department of Computer Science and Engineering, University College Adikavi Nannaya University, India
vermaps@yahoo.com

Abstract— Piracy is no new lesson to every user of technology. It is a giant existing in the industry of technology numbing the heights software is reaching in terms of efficiency and use in all the sectors to an extent where even the industry of enterprise, beaurocracy, technology etc. depend upon the use of software in various ways to contribute their needs and success. Dollars are running at stakes of any industry while the piracy is acting as a dominator of all illegal transactions indirectly by giving easy and such illegal access to many users. One such victim is the entertainment industry where movies are being pirated and accessed to users causing millions of dollars of loss to the entertainment industry. Many researchers and technical specialists have made their contributions to save technology. Our paper is another potential solution with a promise of accuracy rate in detecting the pirated movies. It is a known fact that primary detection is the source of prevention. Our paper puts forth a technique that efficiency detects piracy to reduce such illegal downloads using the efficient measures of signature generation, normalization in an effective manner.

Index Terms— Piracy, Download, Pirated Movies, Earth Movers Distance

I. Introduction

Movies are an indomitable form of entertainment and awareness of every kind. This has proved to be a point of platinum to the industry of enterprise, technology, cultural society etc… Software, being the concrete of reaching and applicability through many sectors has been the target of many profits and achievements leading to monetary and global development as well as a huge contributor to enhance the digitized content. Narrowing to the aspect of movies, this has been the ultimate target of various opponent users of technology. The piracy of movie through free downloads is a very well utilized mechanism from both the makers and the user's side. Various facilitators of such piracy have caused a loss in millions to the movies and respective producer industries. One such notorious facilitator is Pirate Bay set for illegal distribution of copyrighted material over peer-to-peer networks ranging over a large volume of material from movies, music, games, software, books to TV shows, in thirty four languages [2, 3, 5]. The above mentioned is just a copy of many other pirated developers out there who are hacking every penny of legal technology. The statistics themselves speak the wide spread piracy rate breaking the records of the original efforts and hours with the spat of losses of dollars affecting the market and the direction of software utilization. This paper speaks about a potential remedy to put a spat stop to the spread of piracy in the movie industry and its shambolic offence to the value, market and efforts of technology [1]. The various effective and comparative algorithms usage has enhanced the probability of accuracy despite a wide base to authenticate making this method a promising welcome to the need of the hour.

The rest of this paper is organized as follows: Section II. provides the review on related work. Section III. describes the Methodology of the proposed system. Section IV gives the experimental results and analysis based on the supervised and unsupervised sets. Anti-piracy movie system is discussed in section V. Discussions and Conclusion are given in the final section.

II. Literature Review

The level of complicacy is reaching heights with time from the viewpoint of the pirating movies and its respective arena with the gradual and consistent increase in technology giving a high temperature to outwit the opponent techniques with quality measures taken. Various researchers have contributed their aspects and theories to stop the piracy. Few such marked theories have emerged, mentioned in the following content. The early pioneers of techniques began from prevention of basic audio recordings. Tachibana et al proposed sonic watermarking to allow search for illegal recordings made available on the Internet by embedding a secret message into the audio signals. The most recordable and attractive characteristic of this sonic watermarking is that it is applicable even to unplugged live performances. Among certain other watermarking techniques proposed, Haitsma et al developed a video watermarking method for detection of illegal recordings but this technique stood its challenges because of the geometric distortion [10]. This problem was overcome by relying only on the time axis of the movie giving the benefit of such system allowing the identification of theater, presentation time, and other characteristics. This did not serve the right level of its purpose due to the constraints it carried with distortions that affected its accuracy rate. Gohshi et al. proposed a watermarking method that was designed to detect watermarks in unlawfully recorded film made from CRT screens [13].

A higher rate in accuracy was achieved through manual cancellation of geometric distortion of footage but this system had its backlogs in terms of the manual usage exposing it to the probability of error prone results. Later, Lubin et al. proposed a video watermarking method with a desire to strike at digital cinema applications [9]. This method included a scheme to cancel the geometric deformations but it had a higher rate of complicacy in its build and usage. All of these methods are able to distinguish illegal recordings made available on the Internet as well as effective in deterring the camcorder piracy [4]. They, however stand back in terms of specification of the recording locations or the places illegal recordings were made. There are other petty trials that worked to a noticeable extent like Movie Guard for webmasters or video distributors who want to protect Movie files from copyright infringement. This consists of two components. Movie Guard Video Producer that encrypts the respective Movie files in a way that they can only be played by the specified Movie Player. This program allows the creation of a movie playlist followed by encryption of that playlist with different parameters to enable various marketing uses and streaming video over the internet with copy protection [1, 13]. The second component is Movie Guard Video Player that is distributed or sold to customers along with the data files that are generated in the output directory to allow them access the respective video. It allows various options like changing volume etc…Now the movie can be saved in the installed native format only. This "allow save" option only works if the customer has registered with the institute. This is an advantage to use but this is hard to work only in shareware. Various such measures have embarked their way but rolled with their advantages and flaws.

III. Methodology

There are several anti-piracy movie systems used by the industry. They all are processed manually. On an average assumption that individuals would only choose the original version with a copy only under the assurance that he can easily and quickly find a high quality copy without any risk. This can lead us to an opinion that the demand for copies is affected by the availability and the risk associated when copying. In order to protect from such illegal copying, one of the potential measures used is the proposed system. To be a part of this respective security, the primary step for the producer or makers of films is to register their respective movies with the security system. On the view of the system, the methodology preferred from the view point of working the base of the proposed system is to begin and move forth with the three sections: the first one being, selection. Selection of the movies is very important since the efficiency of the trained system proved to be a very essential step for detection of piracy. The selection must be witty enough to grab the movies that are in demand like blockbuster movies or movies that are starred by famous actors since these movies have a high demand of view. The approach of selecting, searching, and downloads must be identical to the approach of an average Internet user thereby gaining a good predictability. Only in that respect viable conclusions can be made. Followed by the search is the second section: search. Once the movies to be searched are selected through many measures like entering the movie title into the ''Search for'' field or specifying the respective ''video'' option in the respective area or in the search options the file size can be specified based on which the required set is obtained. This allows us to focus our search on video files which are labeled with the requested movie or video title. The final ground work section leads: evaluation. On completion of the search process, evaluation is performed by filtering out fake movies like mismatched filename and title name taking into consideration file size, resolution, length, and file format followed by investigation of its playability etc…The three sections being performed conforms the groundwork on which detection and thereby prevention is performed through use of many techniques involving effective use of screenshots, EMD algorithm for validation and the justified respective procedure.

3.1 Earth Movers Distance

EMD comes from famous transportation problem. Earth Movers Distance is method to find the distance between multi-dimensional distributions in the given feature space [6, 7]. A signature is produced from the movie sample. Each signature contains set of features which we more often call it as ground distance factor (GDF). Let us consider there are m suppliers who supplies goods with some weight to the consumer he has. We represent Supplier set S as:

$$P = \{(P_1, Wp_1), (P_2, Wp_2), \ldots, (Pm, Wpm)\} \qquad (1)$$

Assume we have n consumers and every consumer consumes with a weight indicating the amount of weighted product he needs. We represent Consumer C as:

$$Q = \{(q_1, Wq_1), (q_2, Wq_2), \ldots\ldots, (qn, Wpn)\} \qquad (2)$$

Supplier wants to supply the products to the consumers. D can be represented in terms of distance between Supplier and Consumer, which is defined before calculating EMD. It can be formulated as:

$$D = [d_{ij}] \text{ where } 1 \leq i \leq m \text{ and } \quad 1 \leq j \leq n \qquad (3)$$

D is the ground distance matrix where d_{ij} is the ground distance between Supplier and consumer Pi and Qj. Supplier supplies the same product and consumer consumes the same product. To determine the overall cast we have to find the flow matrix F. which holds the factors of product to be moved from one supplier to one consumer.

$$F = [f_{ij}] \text{where } 1 \leq i \leq m \text{ and } \quad 1 \leq j \leq n \qquad (4)$$

The total transportation cost should be minimized and the product weight should be maximized when it is transported from supplier to the consumer. The total cost can be represented as:

$$Work(P, Q, \boldsymbol{F}) = \sum_{i=1}^{n}\sum_{j=1}^{m} fij.dij \qquad (5)$$

The F is calculated based on the following conditions.

$$fij \geq 0 \text{ where } 1 \leq i \leq m \text{ and } 1 \leq i \leq m \qquad (a)$$

$$\sum_{j=1}^{n} fij \leq Wpi, 1 \leq i \leq m \qquad (b)$$

$$\sum_{j=1}^{m} fij \leq Wqj, 1 \leq j \leq n \qquad (c)$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} fij = Min(\sum_{i=1}^{n} Wpi \sum_{j=1}^{m} Wqj) \qquad (d)$$

The above each constraint has a significance. The constraint (a) one allows moving supplies from P to Q and not in other direction. Next (b) and (c) conditions restricts the supplier supplies the product weight not more than the consumer consumable product. Next condition (d) imposes to move maximum amount of supply. We calculate F, the EMD is represented as:

$$EMD(P, Q) = \sum_{i=1}^{n}\sum_{j=1}^{m} fij.\ dij / \sum_{i=1}^{n}\sum_{j=1}^{m} fij \qquad (6)$$

EMD has great advantage when we compare multifeatured signatures.

3.2 Pre-processing Movie and Signature Generation

The successful and protected movies are obtained from the web followed by their respective signatures generated. The task of Movie Pre-processing consists of three steps: 1) First we obtain the movie video from web while in case of its unavailability; a sample of the movie is obtained. 2) Normalization is performed 3) Signature is generated from the downloaded movie or from sample video which contributes to the evaluation of similarity between the legitimate and suspected movie.

The screen shots are captured from the movie in jpeg format. Each image is further processed and normalized with size i.e. 100*100. In our approach each image will produce set of features which represents the signature. Pirated movies loose the clarity, when we capture the screen shots it maintains same poor clarity. We used Lanczos algorithm to calculate the resized image. It has great antialiasing properties and it is quite trivial to calculate in special domain. Fig. 1 shows original images and resized images. We took the samples from native movie.

3.3 Computing the similarity using Earth Mover's Distance

We use Earth Movers Distance to measure the similarity between the movies based on their signatures as given below. The distance matrix is defined in advance D= [dij] (1≤ i ≤ m, 1≤ j ≤ n).

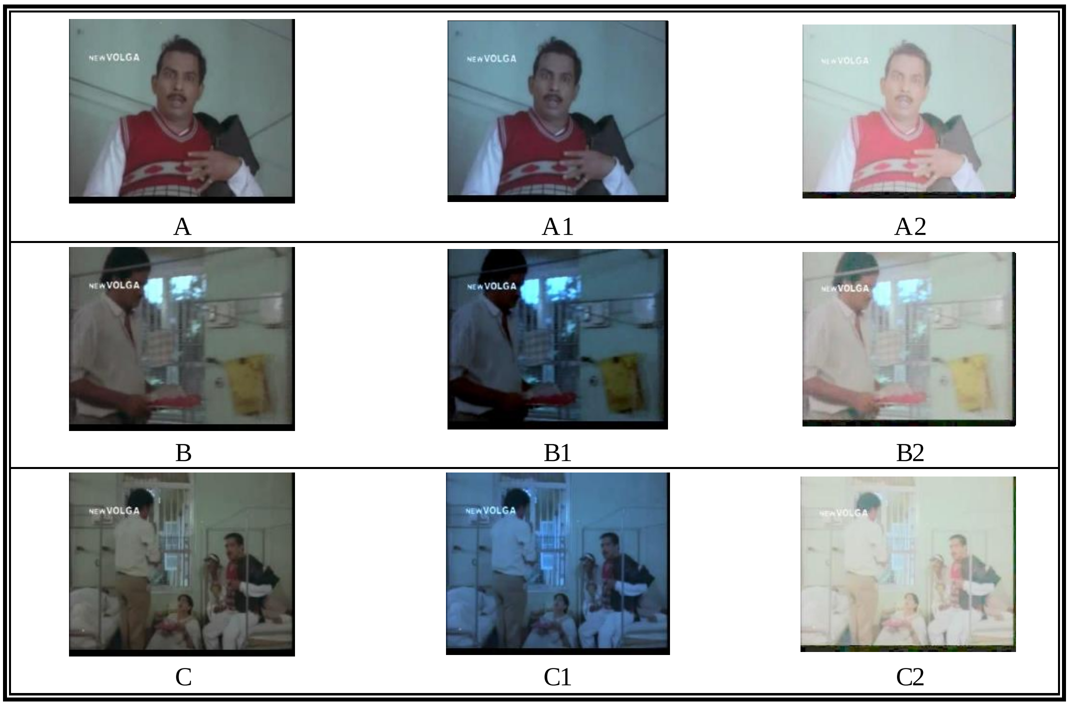

Fig. 1: movie samples with high and low quality

3.4 Image Euclidean Distance

We first calculate the Image Euclidian distance. In PQ dimensional Euclidian space, all the P by Q images can be represented. We denote the s1, s2, s3 …. Spq to form the coordinate system of the image space, where skq+1 links to an ideal point source with intensity at position (k,l). Thus the image f = (f1,f2,…..,fpq), here fkq+1 is the gray level at the (k,l)th pixel. We find the pixel level distance; here in this case one image is captured from high quality image space while the other is captured from suspected one which may be with low quality.

3.5 Classification

We use a special threshold value for each given screen to identify the movie is a pirated movie screen or not. Firstly, Threshold values for legitimate movie screens are calculated. When we suspect a movie, we calculate the extent of similarity with all the legitimate movies in order to differentiate it from movie legitimacy or piracy thereby being able to judge. The threshold value is represented as follows: $K = \{K_1, K_2, K_3, K_4, \ldots\ldots., K_{Nprotected}\}$, Where $K_i(1 \leq I \leq N_{protected})$. Here K denotes the threshold of the k^{th} Protected Movie. $K_i(1 \leq I \leq N_{protected})$ is defined as:

$$K_i = \text{argMin}_{k \epsilon VSVi}(\textit{Missclassification}(k)) -- \delta; \qquad (10)$$

Where Vs V_i is the previously generated Image set(i.e contains similar values of the previously tested pirated movie images) the i^{th} protected movie and *Missclassification*(k) is the number of misclassified movies in case we use k as the threshold.

There are two possible misclassifications:

1) False alarm: whenever the similarity is larger or equal to the threshold value k, this case of the movie is considered to be a non-pirated one. (this is false positive case).

2) Missing: when the similarity is less than k but, in fact, the movie is a pirated one (this is false negative case). k value should be selected in such a way that it is as small as possible , without increasing the misclassified number.

Each pirated movie entry in Vs V_i correlates to two accessory parameters, they are the false alarm number (FA) and false negative (FN). When we use EMD for training, the FA and FN represents the false alarm and false negative as a result it will return. We simply choose the value k that can minimize fa+fn.

IV. Experiments and Analysis

A 100 movie samples are taken from different torrents for the purpose of experimentation to show the efficiency of earth mover's distance in the detection of the pirated movies. We used torrentz search engine with 5 keywords as queries: Hindi, English, Telugu, Kannada, and Tamil movies. We use each term to collect 20 movies. Dead and duplicated url's are removed from the list. From the rest of the url's, 100 movies are collected. From the protected movie, it is observed that most of them are pirated.

4.1 Experiment Result Based on unsupervised learning Sets

Initially our anti-piracy movie system begins without any training. That means all the initial values and threshold values are also set to 0. The results are updated during training phase. When a suspected movie sample comes, the threshold value will be set automatically. The suspected pirated movie list is randomly ordered and given as inputs to the system, one after the other. Our system maintains the classification results in each process.

Table 1: Comparison of pirated Movie with legitimate copy

S.No	Language	Threshold Value	Nearest ones	Farthest ones
1	Language1	0.2779	2	47
2	Language1	0.9540	4	45
3	Language1	1.2250	8	41
4	Language2	1.4193	9	40
5	Language2	2.1171	5	44
6	Language2	3.3272	10	39
7	Language3	0.1779	6	43
8	Language3	0.5540	3	46
9	Language3	1.2340	5	44

4.2 Experiment Result Based on Training Set

In this case, we train our system with known data set and set the threshold value. We collected 100 movies, we select 12 pirated movies, combined all these to training dataset (95general + 5 protected + 12 pirated = 112 total dataset). Here we use the training dataset to calculate the threshold values for the 12 protected web pages. The result is shown in Table 3. We can observe that among the movies belonging to different languages, on comparison of values respectively the threshold values are ranging at the lowest approximately between 0 to values under 2 in the protected movie set in which about 100 trained movies are present.

Table 2: Detection of pirated movies(based on EMD) using supervised learning

S.No	Language	Threshold value	Nearest ones	Farthest ones
1	English	0.2569	4	45
2	English	0.8830	7	42
3	English	1.5434	8	41
4	Tamil	1.4545	3	46
5	Tamil	2.3434	5	44
6	Tamil	3.3434	6	43
7	Telugu	0.1818	12	36
8	Telugu	0.4540	8	41
9	Telugu	1.2340	7	42

We trained our system with 5 protected 12 pirated movies. The training result is listed in Table 2. We combined the legitimate movies with pirated ones. It can be observed from the above table that the movies when compared or referenced to the threshold value, the values that proved nearer to the threshold value is far-fetched from the values obtained farthest depicting based on the ground distance that protected movies can clearly be distinguished from the pirated ones through the distance based calculations obtained.

Table 3: Threshold values for protected movie set trained with 100 movies

S.No	*Protected Movie*	*Threshold values*
1	English	0.9540
2	Hindi	1.4193
3	Kannada	1.2723
4	Tamil	0.8308
5	Telugu	0.2779

In terms of time efficiency, the similarity assessed with the two movie samples are calculated in 0.05 seconds using standard machine with the following specification Intel core i3 processor, 2GB RAM.

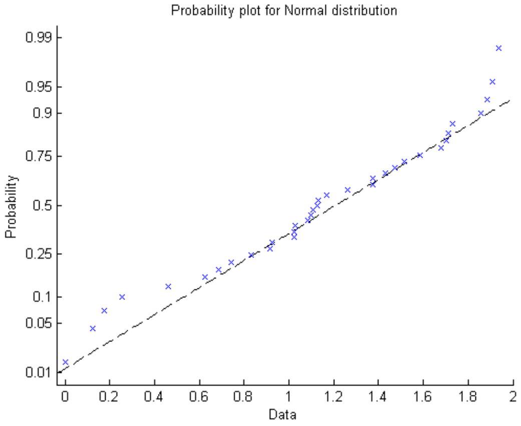

Fig 2: Correct detection of Movie

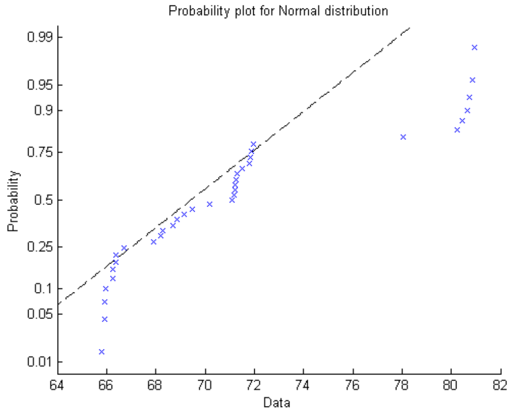

Fig. 3: Wrong detection of Movie

On correct observance of the graph, many values plotted on the respective required plot of normal distribution stand almost on the same line or nearest to it predicting the accuracy of the protected movie which can be observed from Fig. 2.

On the other hand if the graph in Fig. 3 shows the graph for wrong detection of movie. It can be observed that the points plotted on the graph stand very absurd and a little far from the respective line of probability plot for normal distribution. This depicts that the movie is a pirated one as it cannot match the accuracy measures or criteria to be proved as a protected movie.

V. Anti-Piracy Movie System

We developed anti-piracy movie system which automatically detects pirated movies over the internet by comparing their similarities to the protected movies. It is well known that all the pirated movies will be uploaded to file servers, torrents etc. which instigates further eagerness to download movies as they are easily accessible. We built an anti-piracy movie engine which detects such movies. Anti-piracy system updates its database with its threshold values. Fig. 4 shows the architecture of our system. Anti-piracy movie engine is a control panel for registering the legitimate movies which wants protection and maintains the pirated movie hosted url's for further reference.

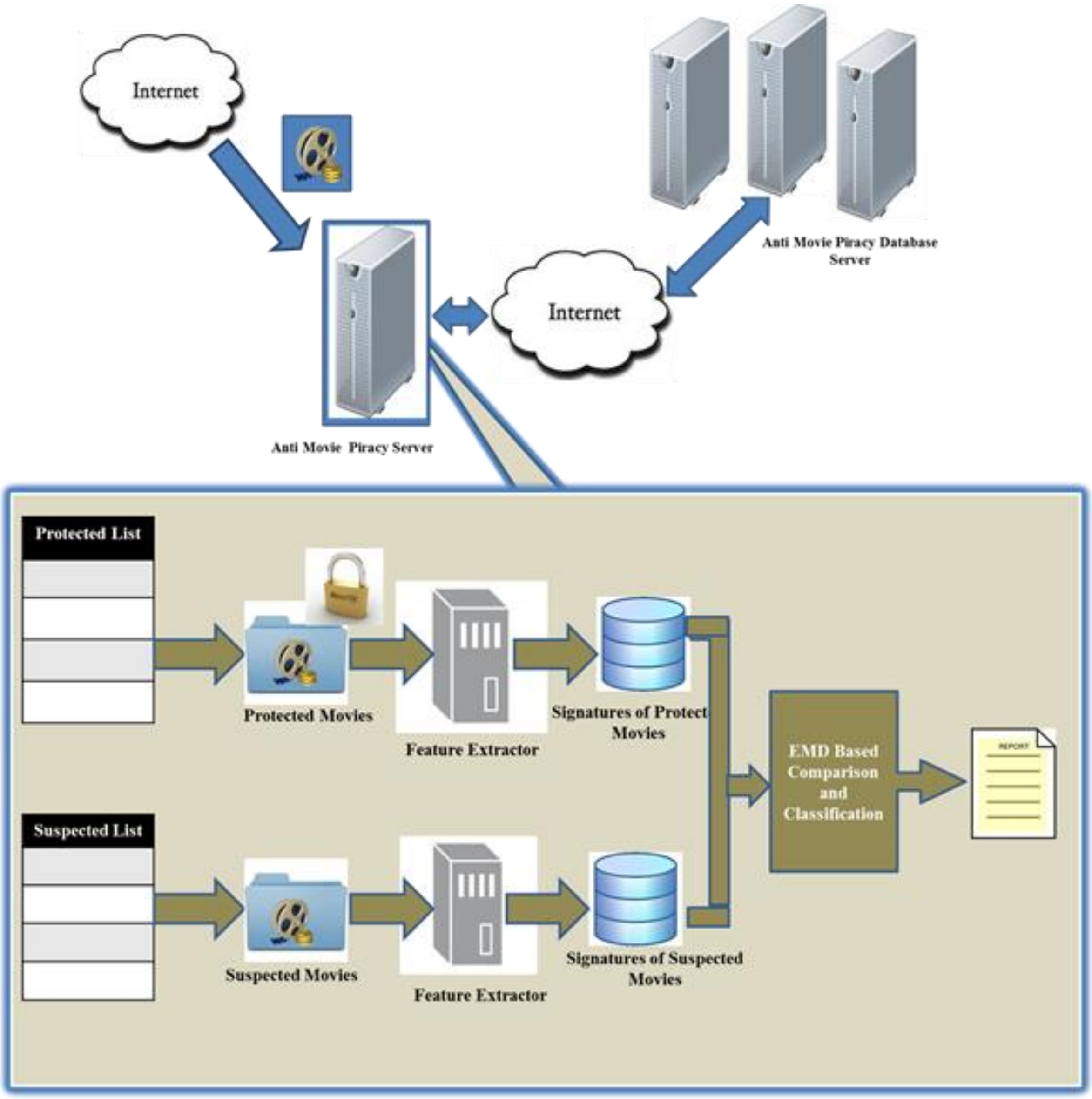

Fig. 4: Anti-Piracy System Architecture

The registered legitimate movies are pre-processed in advance while their corresponding signatures will be stored in the anti-piracy movie database; producer can specify set of features of his or her movie such as casting, location details and soon. So that the Anti-piracy movie engine takes features of casting and uses them as a note of reference when it gets a suspecting movie for the purpose of verification. In case of any similarity noticed, immediate intimation is performed. The system administrator can read the status report of any suspected movie and take the desired decision thereby reducing a huge amount of human effort in the identification issue.

VI. Discussions and Conclusion

An observance ranging from information to entertainment, movies, documentaries etc... Speak that many forms of video act as the lyceum of choice and time. Enhancing the rate of usage of original content can save millions of dollars as it is a known fact that one of the biggest contributors to economy is the entertainment and the box office industry. In Normative terms, by making the above system work, creativity can save its honor as every penny that comes will be a result of a effort of creativity, time and work. Pirating movies sounds easy but the mountain of work and illusion behind satisfying thousands of mind sets and making the content understandable to many varying people is anything but easy. Such an act is being folded into an easy torrent throwing away all the effort for a bin of piracy or easy access. This paper is a potential solution to save the normative side of entertainment industry and acts as a saving grace to many of those enterprises and sponsors who invest loads of dollars. Such a stopping spoon to differentiate pirated movies slows down the access to people eventually opening doors for original CDs and respective theatres that are a true payback for the makers of entertainment industry. It also enhances the technology by being flexible in use. Such an approach increases the reliability and also gives a helping hand to legal measures such as a strict door to use of piracy also makes implementation of legal measures against piracy.

References

[1] Alvisi, M., Argentesi, E., Carbonara, E., Piracy and Quality Choice in Monopolistic Markets, URL:http://www.serci.org/2002/Carbonara.pdf.

[2] Bhattacharjee, S., Gopal, R.D., et al, 2003. Digital music and online sharing: software piracy 2.0? Association for computing machinery. Communications of the ACM 46 (7), 107.

[3] Bhattarcharjee, S., Gopal, R. D., Lertwachara, K., & Marsden, J. R. (2006). Impact of legal threats on online music sharing activity: An analysis of music industry legal actions. Journal of Law and Economics, 49, 91–114.

[4] Das, Sanjukta. 2008. "Timing Movie Release on the Internet in the Context of Piracy." Journal of Organizational Computing and Electronic Commerce 18(4): 307. Mahwah: Oct 2008.

[5] D'Astous, F., Montpetit, D., 2005. Music piracy on the web – how effective are anti-piracy arguments? Evidence from the theory of planned behaviour. Journal of Consumer Policy 28, 289–310.

[6] E. Levina and P. Bickel, "The Earth Mover's Distance is the Mallows Distance: Some Insights from Statistics," Proc. IEEE Int'l Conf. Computer Vision, vol. 2, 2001.

[7] K. Grauman and T. Darrell, "Fast Contour Matching Using Approximate Earth Mover's Distance," Proc. 2004 IEEE CS Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 220-227, 2004

[8] Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211.

[9] J. Lubin, J. A. Bloom, and H. Cheng, "Robust, content-dependent, high-fidelity watermark for tracking in digital cinema," Proc. SPIE Security and Watermarking of Multimedia Contents V, vol. 5020, pp. 536–545, Jan. 2003.

[10] J. Haitsma and T. Kalker, "A watermarking scheme for digital cinema," in Proc. Int. Conf. Image Processing, Oct. 2001, vol. 2, pp. 487–489.

[11] MPAA Statistics. 2005. Cost of Movie Piracy. Retrieved from http://motionpictureassociation.org/researchStatistics.asp

[12] Prendergast, Gerard, Leung Hing Chuen, and Ian Phau. 2002. "Understanding Consumer Demand for Non-Deceptive Pirated Brands." Marketing Intelligence & Planning 20 (7): 405-16.

[13] S. Gohshi, H. Nakamura, H. Ito, R. Fujii, M. Suzuki, S. Takai, and Y.Tani, "A new watermark surviving after re-shooting the images displayed on a screen," KES (2), vol. 3682, pp. 1099–1107, 2005.

[14] Thong, James Y. L. and Chee-Sing Yap. 1998. "Testing an Ethical-Decision Making Theory: The Case of Softlifting." Journal of Management Information Systems 15 (1): 213-37.

[15] Wagner, Suzanne C. and Lawrence G. Sanders. 2001. "Considerations in Ethical Decision Making and Software Piracy." Journal of Business Ethics 29 (1): 1/2.

Srinivas Baggam: Assistant Professor of Computer Science and Engineering in MVGR College of Engineering. Received M.Tech in computer Science and engineering in 2008 from Acharya Nagarjuna University; he has two and half years of industry and four and half years of teaching experience.

Dr. Koduganti Venkata Rao: Professor of Computer science and Engineering in Vignan Institute of Information Technology and Vice-Principal. Received Ph.D in Computer Science and Engineering from Andhra University, M.Tech in (Computer Science and Technology) from Andhra University and M.Sc (computer science) from Nagarjuna University, 2008, 1999, 1994 respectively.

Dr. P. Suresh Varma: Professor of Computer science and Engineering in University College, Adikavi Nannaya University. Received Ph.D. in Computer Science and Engineering with specialization in Communication Networks from Acharya Nagarjuna University. M.Tech in (Computer Science & Science and Technology) from Andhra University in 1998. A.M.I.E (Computer Science and Engineering from Institute of Engineers. M.Sc (Nuclear Physics from Andhra University from 1993.

How to cite this paper: Srinivas Baggam, K.Venkata Rao, P. Suresh Varma,"Detecting Pirated Movie's with Similarity Assessment Based on Earth Mover's Distance", International Journal of Information Technology and Computer Science(IJITCS), vol.5, no.4, pp.47-54, 2013.DOI: 10.5815/ijitcs.2013.04.06