Multimodal Periodicity Analysis for Illicit Content Detection in Videos

N. Rea [1]*, G. Lacey[1], C. Lambe[2] and R. Dahyot[1]†

[1] School of Computer Science and Statistics, Trinity College Dublin, Ireland.
[2] Pixalert, Digital Hub, 10-13 Thomas Street, Dublin 8, Ireland.

Keywords: Multimodal retrieval; Illicit content detection; Periodicity; Segmentation.

Abstract

In this paper, we propose a multimodal approach to illicit content detection in videos. Distribution of pornographic material over computer networks has been taking place since the inception of the internet. Until recently, most of the research focuses on illicit content detection in images and text. Typically it involves robust skin detection, texture characterisation, shape modelling and keyword filtering. Video however provides the opportunity of exploiting supplementary features including audio and motion for additional confidence in the classification process. This work investigates the use of visual motion information and periodicity in the audio stream for illicit content detection in videos.

1 Introduction

The rise of the digital era associated with powerful modern communication technologies has facilitated the creation and distribution of multimedia data such as text, images or more recently videos. In particular, pornographic materials have largely benefited from this revolution [6]. This kind of material is illegal on the workplace and therefore refered as illicit in this article. Since then, the issue of filtering "illicit content" has been one of major concerns since the introduction of the web in the early 1990's. The most obvious concern is that of morality and the situation which parents and members of educational institutes are experiencing in trying to protect children from such content. Furthermore, there are political consequences such as employers running the risk of legal liability as part of sexual harassment suits for not providing sufficient e-mail filtering [22]. This has resulted in several commercial products being made available. Pixalert's 'Auditor' and 'Monitor'[1], FutureSoft's 'DynaComm i:scan' [2] and Hyperdyne Software's 'Snitch'[3] all provide image and text filtering for remote scanning of e-mail, hard disks and peripheral storage devices (e.g. USB memory keys).

There has been much research conducted into content-based analysis of illicit images [13, 19, 5, 30, 4, 2]. It appears however, that none have yet addressed the problem in video. Faster networks coupled with advanced compression techniques and the relative inexpensiveness of storage and video capturing devices have made illicit multimedia widely available. In tandem with this proliferation of illicit content there has been a massive upsurge in quantity of video being produced by home users in the form of home movies. Sport footage is extensively available, movies, TV shows and general inoffensive multimedia content can also be downloaded from the Web (e.g. YouTube, Video Google), sent as e-mail attachments or file-shared on P2P networks. There is therefore a need to be able to *effectively* and *efficiently* disambiguate between safe and indecent media streams. The same, extensively researched image based methods could be applied to individual keyframes selected at particular time intervals, but video offers more features to aid with classification. The features used should be easily extracted from the footage with the expectation of being able to filter video while it is being played back or in faster than real time for offline scanning.

We propose the use of multimodal information extracted from the audio-visual stream as pornographic material can exhibit characteristic motion and audio patterns. This paper concerns itself with extracting these features and exploring models for classification. Existing colour based approaches are also exploited. Section 2 provides a review of current techniques for illicit content detection. Section 4 and 3 discusses the audio and visual features respectively. The paper finishes with conclusions and future directions.

2 State of the art

As discussed in the introduction, all papers in this domain have only considered image and text modes. This section reviews some of these works.

In [20], a machine learning approach is used to design a Intelligent Classification Engine to classify webpages as pornographic or nonpornographic based on the analysis of the displayed and non-displayed text content. Alternatively statistical methods such as Bayesian classification [15] can also be used. However such approaches require an extensive learning when it comes to deal with multilingual web documents. As a major advantage, image analysis techniques for illicit content detection does not suffer from this drawback

*niall.rea@gmail.com
†Rozenn.Dahyot@cs.tcd.ie

[1] `http://www.pixalert.com/product/product.htm`
[2] `http://www.futuresoft.com/documentation/dciscan/imagerecognition.pdf`
[3] `http://www.hyperdynesoftware.com/clean-porn.html`

as image understanding or interpretation is not language dependent.

Many illicit content detection systems in images rely on skin processing as a first step [30, 13, 18, 11]. Parametric models using mixtures of Gaussians [23] have been shown to give good results but are computationally complex since a likelihood must be evaluated on each Gaussian for each pixel in the image. Jones et al. [18] use a non-parametric approach by building global skin and non-skin histograms from random trawls of images on the web. The choice of colour space for skin modelling has also been an area of much discussion. Log-opponent colour space [3] is said to better model the human vision system, while normalised RGB space is invariant to different luminance conditions [25]. Skin tone detection has also been undertaken by Jedynak et al. [17] where Hidden Markov Models and colour gradients are used to enforce smoothness by providing spatial pixel dependencies. Spatial consideration of skin patches is further researched by Sebe et al. [26] and a framework for training using labelled as well as unlabelled data is achieved using Bayesian networks for classification.

Various information can be extracted from the skin segmentation stage such as the relative position and size of the skin blobs [11]. More high-level process such as face detection can also be used to help in the classification of the images as pornographic or nonpornographic. Image feature classification is, as for text analysis, performed using machine learning methods [11] or statistical decision [2].

As illicit content detection in videos seems to not have yet been addressed in the scientific literature, we propose first in section 3, to improve skin segmentation using motion information. In addition to visual information, video usually also contains audio data. An original illicit content detection in the audio stream is then proposed in section 4.

3 Visual Features for Illicit content detection in Video streams

Videos containing illicit content all exhibit similar characteristics in terms of colour and motion features. As discussed in section 2, most, if not all content based analysis approaches for illicit image detection rely on colour, texture and certain geometrical features. Exploiting motion should provide additional confidence for detecting such content in video. In this section we detail the colour feature for skin detection which is subsequently refined using MPEG motion vectors.

3.1 Colour information

A variety of commercial products including those by Pixalert, FutureSoft and Hyperdyne Software provide video filtering, but for the most part they simply apply the same image filtering techniques to keyframes selected at regular time intervals. There can be a number of issues arising from using only colour and texture based segmentation in video:

- Colour bleed: VHS tapes operate at low resolution and the physical aspect of the media lends itself to colour bleeding. Spatial subsampling can also result in colour bleed as can the analog to digital transfer.
- Bad lighting: If a scene is badly set up in terms of the lighting (e.g. using direct instead of diffuse lighting) object reflections can occur. This also might result in white objects appearing as a pink hue.
- Colour contrast: Occasionally the background colour distribution might lie within the range of values of a skin filter.
- Red-shift: This is a phenomena in older cameras and results in a spectral shift of colours on film.

Since skin regions occupy a relatively narrow range in the colour spectrum they provide a significant clue to the presence of illicit content.

3.2 Detecting skin regions

Skin and non-skin reference histograms were obtained from the open-source filtering Poseia project[4]. The project has compiled 32^3 3D RGB histograms from the Compaq database. A pixel z at location (i, j) in an image X can be classified as skin using the Bayesian formulation in equation 1 which results in a skin probability map $X_p(i, j)$.

$$\begin{aligned} P(\text{skin}|z) &= \frac{P(z|\text{skin})P(\text{skin})}{P(z)} \\ &= \frac{P(z|\text{skin})P(\text{skin})}{P(z|\text{skin})P(\text{skin})+P(z|\text{nonskin})P(\text{nonskin})} \end{aligned} \quad (1)$$

While this formulation treats pixels independently, it is a sufficient model for the initial skin segmentation.

A skin binary map $X_b(i, j)$ is then generated according to equation 2.

$$X_b(i,j) = \begin{cases} 1 & \text{if } X_p(i,j) \geq 0.7 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

[4] `http://www.poesia-filter.org/`

3.3 Motion information

Tracking algorithms can be computationally intensive, so we would like to exploit features which are at hand to allow for real time processing. In the case of this work we are not interested in *explicitly* tracking individuals or body parts, but monitoring regions for "homogeneous motion". We exploit the MPEG motion vector field which is available from the decoding process and while it is generally accepted that block matching motion estimation techniques such as those used by MPEG are insufficient for tracking it is suitable for our purpose.

We use a simple model to characterise the motion in illicit video. We assume that the scene involves only two distinct types of motion: a local homogeneous foreground motion and a global homogeneous background motion. Obviously the level of zoom/close-up will result in a certain ambiguity, however it is assumed that the region comprising a higher amount of skin to non-skin pixel ratio is the foreground.

Motion vectors are firstly extracted from the video stream. These vectors are then compensated for global motion and clustered using k-means clustering for segmentation.

3.3.1 Extracting MPEG motion vectors

In order to extract motion vectors from a variety of motion compensated compressed video sequences (including MPEG-1/2 and 4), a modified version of the open source decoder ffdshow[5] was used in a Microsoft Directshow architecture. With supplementary custom Directshow filters, the motion vectors could be extracted from the partially decoded sequence in faster than real time.

3.3.2 Global motion compensation

Occasionally the footage experiences some camera pan and zoom. We compensate for this motion using the method in Coudray et al. [7]. A 4-parameter motion model (equation 3, where $V(x,y)$ are the MPEG motion vectors at macroblock locations (x,y)) was deemed to be sufficient, as these videos do not generally exhibit complicated camera movements. This also keeps computational complexity to a minimum.

$$V(x,y) = Z\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \tag{3}$$

[5] http://ffdshow.sourceforge.net/tikiwiki/tiki-view_articles.php

Zoom in the 4 parameter model is calculated according to

$$Z = \frac{\partial V_x}{\partial x} = \frac{\partial V_y}{\partial y} \tag{4}$$

And x and y translations t_x and t_y are therefore calculated as

$$\begin{pmatrix} t_x \\ t_y \end{pmatrix} = V(x,y) - Z\begin{pmatrix} x \\ y \end{pmatrix} \tag{5}$$

3.3.3 Rejecting vectors

Since the foreground is known to contain skin regions, only vectors in non-skin regions are considered to be attributed to global motion (i.e. macroblocks which contain less than 30% skin pixels). Furthermore, block based motion estimators such as those used for MPEG are prone to errors in regions of low texture. Similar to [7], the DCT coefficients from the MPEG stream are exploited to determine a measure of texture in each macroblock. The vectors attached to those blocks with low texture [14] are not used in the approximation of the global motion field. The remaining vectors are accumulated in a 2D motion vector histogram, the mode of which is deemed to be the global translational motion.

3.3.4 3D vector median filtering

Anomalous motion vectors can result during the MPEG encoding process. The vector field is therefore filtered using the ML3D filter outlined in Alp et al. [1]. Firstly, the vector median of the diagonals of a window of 3x3 vectors in the current frame and the centre vector values from the same window location in the previous and next frame are computed. Secondly the vector median of the vertical and horizontals through the centre of the window and the centre vector values from the same window location in the previous and next frame are once again calculated. The output of the ML3D filter is the vector median of the previous operations.

3.3.5 Motion segmentation

Once the vectors have been compensated for global motion, they are clustered using k-means. K-means is used since it is a computationally efficient clustering algorithm and gives satisfactory results compared to the watershed segmentation used by [8] and KM-EAC [16]. Techniques such as normalized cuts would also be too excessive computationally.

As discussed previously, we only assume a single foreground homogeneous motion and one background motion, so only two clusters are needed to completely characterise the motion in a scene. An initial guess is used as the centroid for the first iteration of the first frame. Since there should not be much deviation in vector field in consecutive frames, the centroids from the previous frame are used as in the first iteration of the current frame. This results in a motion segmented binary map $X_k(i,j)$.

The mean motion vector at frame f, V_f, occurring in the binary 'and' result ($X_a(i,j)$ equation 6) of $X_k(i,j)$ and $X_b(i,j)$, is computed where M, N are the number of vertical and horizontal motion vectors respectively (equation 7).

$$X_a(i,j) = X_k(i,j) \wedge X_b(i,j) \tag{6}$$

$$V_f(x,y) = \frac{1}{NM} \sum_{x \in X_a}^{N} \sum_{y \in X_a}^{M} V(x,y) \tag{7}$$

Figure 1 shows results of background-foreground segmentation using motion vectors depending of the type of the frame (B, P or I) in the MPEG stream. Best segmentation is obtained using the first B frame in each group of pictures (GOP). A GOP is a fundamental unit of 12 encoded images in MPEG-1/2 which uses the I-frame as an anchor from which all frames in the GOP are related. I- and P-frames are also known as reference frames.

Figure 2 shows the binary skin image, the motion compensated segmentation with overlaid motion vectors and the binary 'and' result $X_a(i,j)$ for a still from *When Harry met Sally*. Using the motion information helps to segment relevant skin region with higher accuracy. Since skin detection is a basic preliminary step into most illicit image detection softwares, using this approach in videos to improve skin region segmentation will improve current methods where videos are treated as independent set of images.

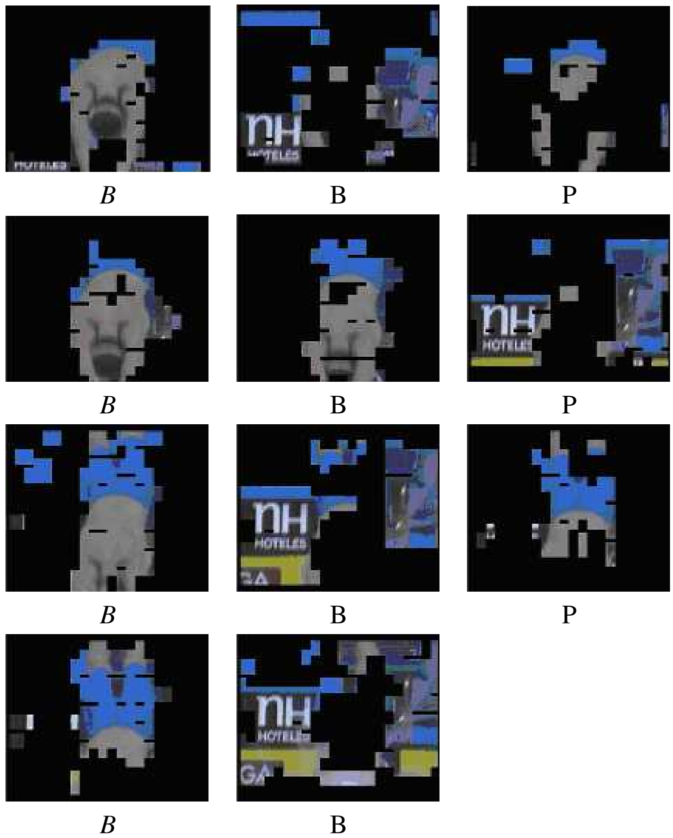


Figure 1: Sequence of decoded MPEG-1 frames which are motion segmented using k-means clustering of the motion vectors [24]. The best segmentation is obtained for the *B*-frames.

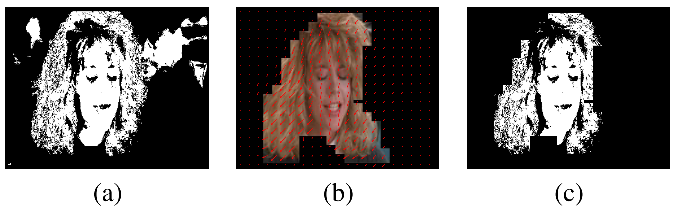


Figure 2: (a) Binary map of the skin segmentation; (b) Motion segmentation with overlaid motion vectors; (c) Binary 'and' of motion and skin segmentations

4 Audio features for Illicit Content detection

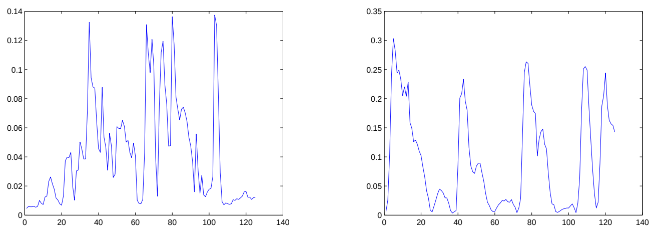
The approach presented here is motivated by the idea that, even when not watching the video content from a multimedia stream, the nature of the stream can still be understood from the audio information alone. Examples of applications can be found in sport video indexing for instance, where the recognisable sounds of racket hits are easy clues to segment playing periods in tennis videos [10].

We consider illicit video material, and in particular, we aim at detecting pornographic content. In addition to visual information, the audio data also presents valuable clues such as periodic sounds generated by the specific activities of the actors.

For illustration, we have chosen the famous scene from the movie *When Harry met Sally* (Sally's simulation of an orgasm, which is a series of moans and screams). The scene starts with a conversation between Sally and Harry.

4.1 Energy of the signal

The loudness of the audio signal is computed on a temporal window of $0.04s$ (duration of a 25fps video frame). For analysis of periodic patterns, we consider a 5 second period corresponding to 125 measurements of volume. Figure 3 presents two 5 second periods.

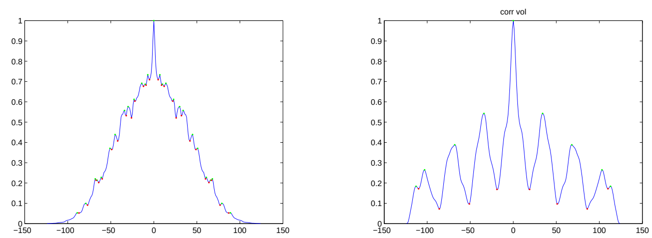


(a) (b)

Figure 3: Audio energy computed over $5s$ when Sally talks to Harry (a), and when Sally is simulating (b).

4.2 Autocorrelation

A typical clue to detect illicit material is to look at periodic patterns in the audio file. Figure 3 confirms that a periodic pattern is exhibited during the *illicit* extract (b) more so than during the conversation (a). Periodicity in the signal is usually analysed by autocorrelation, circular correlation or periodogram [29, 27]. We simply choose here to use the autocorrelation of the energy. The corresponding autocorrelation for the two signals in figure 3 is given in figure 4. Peaks appearing in (b) show that the signal is periodic.



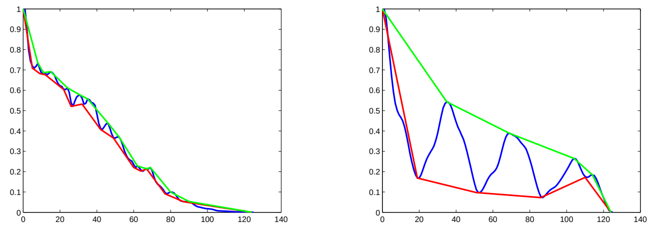
(a) (b)

Figure 4: Autocorrelation of the energy in the audio data with their maxima (green dots) and minima (red dots).

4.3 Measure of Audio periodicity

Previous works have already studied periodicity in signals [29, 27]. We aim here to define a measure to discriminate autocorrelations of classes similar to (a) and (b) (cf. figure 4). As a measure of periodicity, we simply propose to compute the difference between the surface defined by the minimas and the maximas of the autocorrelation. This is illustrated in figure 5 for the same audio extracts (a) and (b).

Figure 6 shows this periodicity measure during the whole scene of *When Harry met Sally*. The measure is low at the start as only a conversation occurs between the two main characters. Then starting at 95 seconds, the periodic pattern begins. In this case, periodic moaning and screaming appears on the audio data. By the end of the scene, standard conversation takes place again and the measure of periodicity decreases.



(a) (b)

Figure 5: The measure of periodicity on the half-autocorrelation is computed by the surface between the green curve (defined by the maxima's in figure 4) and the red curve (defined by the minima's in figure 4).

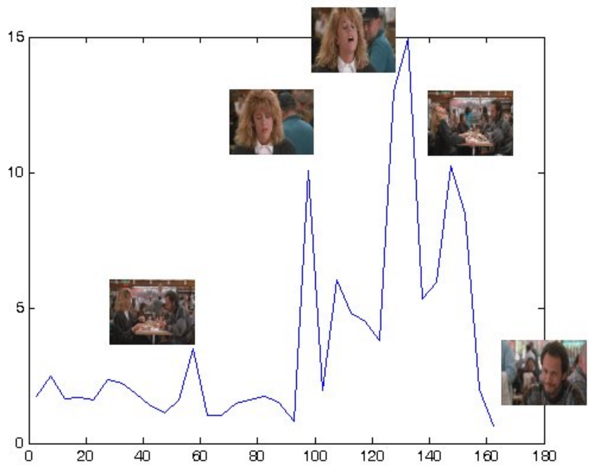


Figure 6: Measure of periodicity in the scene of *When Harry met Sally* w.r.t the time (in seconds).

4.4 Remarks

Periodic sounds in illicit videos are not only voices but also sounds related to the periodic activities of the characters. As an advantage, the energy of the audio signal taken here does not discriminate between voices and those other specific sounds. On the other hand, other non-illicit audio signals can have an energy with similar periodic behaviour (tennis rallies, music beat, etc.). However the video features can removes ambiguities. It is worth stressing that, by its simplicity, the audio processing method proposed here is performed faster than real time.

Speech recognition systems associated with text analysis for illicit content detection could also be a possible solution to this problem. However such an approach would be language dependent and more computer intensive. Our current solution does not have those drawbacks, and follows the intuition that illicit materials can be easily identified in listening without having to understand the language.

4.5 Audio results

The detection of illicit audio material is performed on the measure of periodicity previously defined. Values greater than a specific threshold indicates illicit audio scenes, and the other values correspond to non-illicit materials. The threshold has been chosen equal to 4 and performs a perfect segmentation in the scene of *When Harry met Sally* (cf. figure 6).

This method has been assessed first on non-illicit materials (20 minutes of extracts from movies and music videos) to evaluate the false alarm rate of the method. Various audio sources was used (music, speech, explosion, scream etc.), and in all those, the false alarm rate is rather low at 2%.

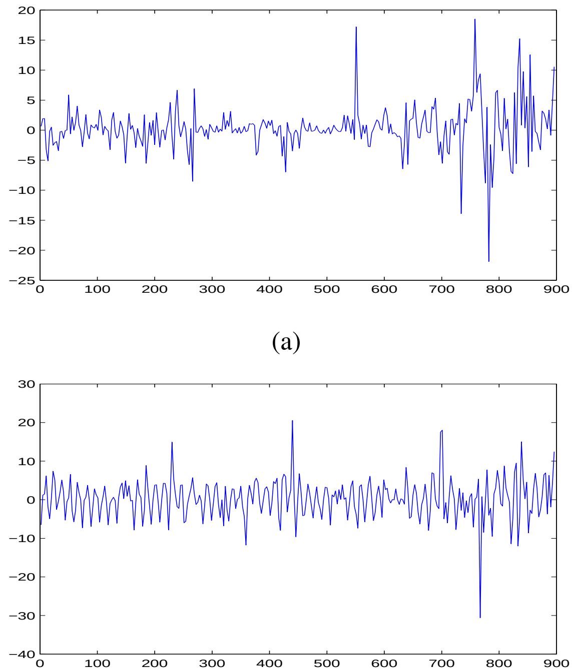
The detection rate is more difficult to assess as periodic sounds do not occur all the time in the audio stream. Ten minutes of eight different extracts of illicit materials showing periodic sounds have been used. Five extracts corresponding to 9 minutes of the test have been properly detected. Three short extracts (representing 1 minute of recording) are missed. On those three files, a mixture of sounds is occurring (speech or music) masking the relevant periodicity on the loudness feature.

5 Current investigations on motion periodicity

Periodic sounds in illicit materials are usually correlated to periodic motion occurring in the visual stream. Detecting periodic motion behaviour has become increasingly popular for retrieval in video [9, 21, 12]. Cutler [9] introduces similarity plot to illustrate the behaviour of natural periodic motion. These similarity plots expose a confusion matrix of the cross-correlation of segmented dynamic objects over a particular duration. Liu et al. [21] linearly combine similarity plots described above with the well known Motion History Image and Motion Energy Image along with an auto-regressive trajectory model to retrieve similar sequences from a repository of videos. Sport is the consideration of Fangxiang et al. [12]. Motion features are used in a neural network classifier to distinguish various sports events with characteristic periodic and non-periodic motions (e.g. sprinting, canoeing, power-lifting).

An interesting approach is proposed in [28] where first, local and global motions are separated in videos. Local motion periodicity, measured with the autocorrelation of motion energy, provides a relevant clues to classifying different type of strokes in swimming sport videos.

In [28], local motion is estimated from raw video data and a robust mean is computed on regions of interest. We propose to use our motion field extracted from MPEG on skin regions and to compute the average motion on those regions of interest. Since the MPEG *B*-frames are giving the best segmentation and the best estimated motion field (cf. section 3.3.5), only these are used to estimate the mean motion vector of the foreground object. Figure 7 shows the average motion on skin region with respect to time. Some periodic behaviour seems to appear in the vertical direction in this illicit video. Figure 8 illustrates the

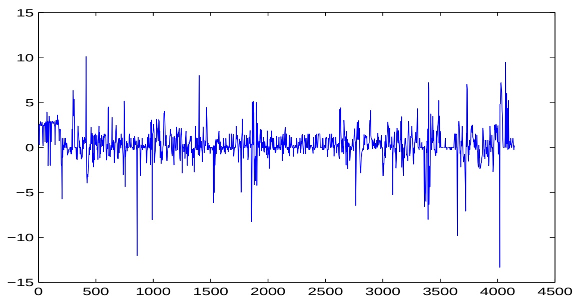


(b)

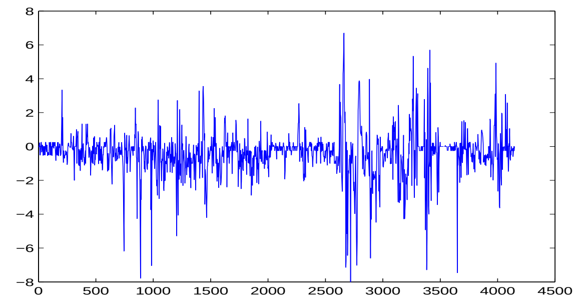
Figure 7: Mean of motion vectors in skin regions w.r.t. frame number from an illicit source (*B*-frames): (a) horizontal projection, (b) vertical projection.

vertical and horizontal mean motion vector of the B-frames in the "simulation scene" of *When Harry met Sally*. No periodic pattern emerges.

The method presented in section 4 for measuring the periodicity of the energy of the audio data, is currently investigated for motion periodicity detection. Preliminary results are shown in figure 9. As observed in section 4, the area in between the curves defined by the minima (red) and the maxima (green) is maximal in illicit material where periodicity occurs.



(a)

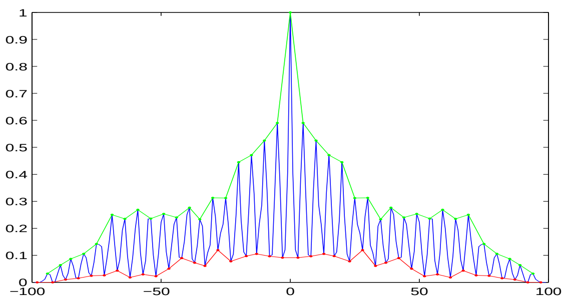


(b)

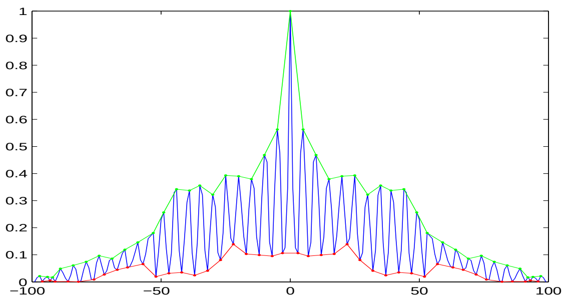
Figure 8: Mean of motion vectors in skin regions for the simulation scene of *When Harry met Sally* (*B*-frames): (a) horizontal projection, (b) vertical projection.

6 Conclusions and Future work

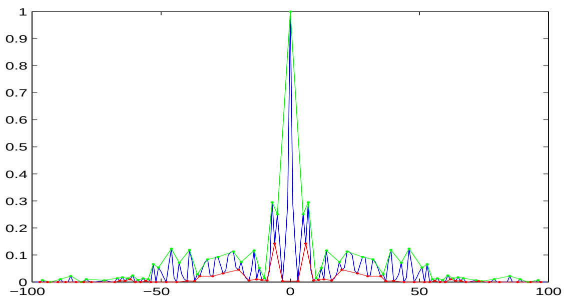
This paper has presented audio and visual features for the purpose of identifying videos containing illicit content. As an improvement to illicit content detection softwares for images, we have proposed to use motion information directly extracted from MPEG coding to improve the segmentation of skin regions. Moreover, we have presented an original audio feature to detect pornographic audio material. It is mainly based on the occurrence of periodic patterns in the energy of the audio stream. Both our skin segmentation using colour and motion and audio analysis are performed in real time. It is anticipated that when fused, these features will provide a powerful method for filtering video in real time. This could be built as a plug-in for a media player for the purpose of parental control or for corporations who wish to protect themselves from possible legal action. Fast computing is indeed an important feature for applications such as online illicit content filtering for IPTV or webcam broadcastings. Future work will look at assessing motion periodicity as a clue for illicit material detection and at merging the multimodal approaches presented in this work.



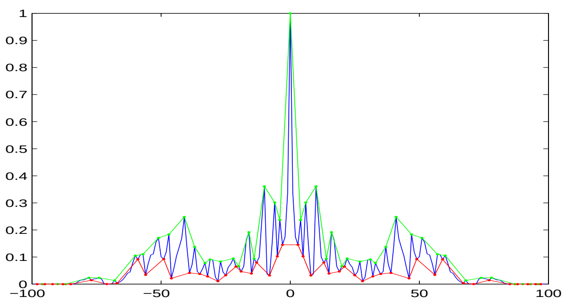
(a)



(b)



(c)



(d)

Figure 9: Autocorrelation of the mean motion in the vertical direction on skin region computed on sequences of 100 successive *B*-frames in illicit materials ((a) and (b)) and in the non-illicit scene *When Harry met Sally* ((c)and (d)).

Acknowledgements

This work has been funded by the European Network of Excellence on *Multimedia Understanding through Semantics,* *Computation and Learning*, MUSCLE FP6-5077-52 `http://www.muscle-noe.org/`, and by Enterprise Ireland, IP 2006/334/Y.

References

[1] B. Alp, P. Haavisto, T. Jarske, K. Oistamo, and Y. Neuvo. Median based algorithms for image sequence processing. In *Visual Communications and Image Processing*, pages 122–134, 1990.

[2] W. Arentz and B. Olstad. Classifying offensive sites based on image content. *Computer Vision and Image Understanding*, 94:295–310, 2004.

[3] J. Berens and G.D. Finlayson. Log-opponent chromaticity coding of colour space. In *International Conference on Pattern Recognition*, volume 1, pages 206–211, 2000.

[4] A. Bosson, G. Cawley, Y. Chan, and R. Harvey. Nonretrieval: blocking pornographic images. In *CIVR*, pages 50–60, 2002.

[5] Yi Chan, Richard Harvey, and J. Andrew Bangham. Using colour features to block dubious images. In *European Signal Processing Conference (EUSIPCO)*, 2000.

[6] J. Coopersmith. Pornography, videotape, and the internet. *IEEE Technology and Society Magazine*, pages 27–34, Spring 2000.

[7] R. Coudray and B. Besserer. Global motion estimation for MPEG-encoded streams. In *International Conference on Image Processing*, pages 3411–3414, 2004.

[8] R. Coudray and B. Besserer. Motion Based Segmentation using MPEG Streams and Watershed Method. In *International Symposium on Visual Computing*, pages 729–736, 2005.

[9] R. Cutler and L.S. Davis. Robust real-time periodic motion detection, analysis, and applications. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 22(8):781–796, August 2000.

[10] R. Dahyot, A. C. Kokaram, N. Rea, and H. Denman. Joint audio visual retrieval for tennis broadcasts. In *IEEE proceedings of the International Conference on Acoustics Speech and Signal Processing (ICASSP)*, Hong Kong, April 2003.

[11] J. Ruiz del Solar, V. Casta neda, R. Verschae, R. Baeza-Yates, and F. Ortiz. Characterizing objectionable image content (pornography and nude images) of specific web segments: Chile as a case study. In *IEEE Third Latin American Web Congress (LA-WEB)*, 2005.

[12] C. Fangxiang, W. Christmas, and J. Kittler. Periodic human motion description for sports video databases. In *International Conference on Pattern Recognition*, volume 3, pages 870 – 873, 2004.

[13] Margaret M. Fleck, David A. Forsyth, and Chris Bregler. Finding Naked People. In *European Conference on Computer Vision (2)*, pages 593–602, 1996.

[14] Richard E. Frye and Robert S. Ledley. Texture discrimination using discrete cosine transformation shift-insensitive (dctsis) descriptors. *Pattern Recognition*, 33:1585–1598, 2000.

[15] W. H. Ho and P. A. Watters. Statistical and structural approaches to filtering internet pornography. In *IEEE International Conference on Systems, Man and Cybernetics*, pages 696–702, 2004.

[16] A. Jain and A. Fred. Evidence accumulation clustering based on the k-means algorithm. In *Structural, Syntactic, and Statistical Pattern Recognition (SSPR)*, pages 442–451, 2002.

[17] Bruno Jedynak, Huicheng Zheng, and Mohamed Daoudi. Maximum entropy models for skin detection. In *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, volume 2683, pages 180–193, 2003.

[18] M. Jones and J. Rehg. Statistical color models with application to skin detection. In *International Conference on Computer Vision and Pattern Recognition*, pages 274–280, 1999.

[19] Michael J. Jones and James M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1):81–96, 2002.

[20] P. Y. Lee, S. C. Hui, and A. C. M. Fong. Neural network for web content filtering. *IEEE Intelligent System*, pages 48–57, 2002.

[21] Che-Bin Liu and Narendra Ahuja. Motion based retrieval of dynamic objects in videos. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 288–291, 2004.

[22] Declan McCullagh. Porn spam–legal minefield for employers. Website, 2003. `http://news.zdnet.com/2100-1009_22-995658.html`.

[23] S.J. McKenna, S. Gong, and Y. Raja. Modelling facial colour and identity with gaussian mixtures. *Pattern Recognition*, 31(12):1883–1892, December 1998.

[24] N. Rea, R. Dahyot, C. Lambe, and G. Lacey. Detecting illicit content in video streams. Technical Report IP/2006/0334, Final Technical Report for Innovation Partnerships, Enterprise Ireland, 2006.

[25] B. Schiele and A. Waibel. Gaze tracking based on face-color. In *Workshop on Automatic Face and Gesture Recognition*, pages 344–349, 1995.

[26] Nicu Sebe, Ira Cohen, Thomas S. Huang, and Theo Gevers. Skin detection: A bayesian network approach. In *International Conference on Pattern Recognition*, volume 2, pages 903–906, 2004.

[27] W. A. Sethares and T. W. Staley. Periodicity transforms. *IEEE transactions on Signal Processing*, 47(11), November 1999.

[28] Xiaofeng Tong, Lingyu Duan, Changsheng Xu, Qi Tian, Hanqing Lu, Jinjun Wang, and J.S Jin. Periodicity detection of local motion. In *International Conference Multimedia and Expo (ICME)*, July 2005.

[29] M. Vlachos, P. Yu, and V. Castelli. On periodicity detection and structural periodic similarity. In *SIAM International Conference on Data Mining*, 2005.

[30] J. Ze Wang, J. Li, G. Wiederhold, and O. Firschein. System for screening objectionable images using Daubechies' wavelets and color histograms. In *International Workshop on Interactive Distributed Multimedia Systems and Telecommunication Services*, page 2030, 1997.