Computing, Information Systems, Development Informatics & Allied Research Journal
Vol. 15, No. 1, 2024 - www.isteams.net/cisdijournal.net

Web Framework Content Based Filtering for Picture Sharing

[1]Ehigiator Egho-Promise & [2]Mensah Sitti
[1]Faculty of CreaTech, Department of ICT, City of Oxford College & University Centre, Oxford, UK
[2]Dept. Computer Science and Engineering, University of Mines and Technology, Tarkwa, Ghana
E-mails: eghopromise@yahoo.com; msitti@umat.edu.gh

ABSTRACT

Information technology has greatly impacted the multimedia industry, particularly in the areas of video and audio content. There are various applications available for managing individual videos and converting them into formats suitable for online sharing. With the widespread use of mobile phones and computers, capturing and storing pictures and videos has become a common practice. These files can be easily shared on different networks, including popular platforms like Facebook, YouTube, and WhatsApp. The rise of web 2.0 has led to an increase in file sharing, including explicit content, as uploads are often not properly filtered. Social media platforms play a significant role in facilitating real-time video sharing, but they also host unwanted content such as violent and pornographic videos. This research aims to develop a web framework application that filters images and videos before they are uploaded to interconnected networks, ensuring that unwanted content is detected and flagged. Users will not be blocked from sharing content, but the system will help identify and address inappropriate material.
Keywords: Picture Sharing, Web Framework, Content-Based Filtering, Online Security .

CISDI Journal Reference Format
Ehigiator Egho-Promise & Mensah Sitti (2024 Web Framework Content Based Filtering for Picture Sharing. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 15 No 1, Pp 29--46. dx.doi.org/10.22624/AIMS/CISDI/V15N1P3. Available online at www.isteams.net/cisdijournal

1. INTRODUCTION

Due to its difficulty, there exists no true solution to the use of high-level semantic concepts for content-based classification and retrieval of visual files like images and videos. This is as a result of the semantic gap between the low-level concepts like pixels, frames, etc. and the high-level concepts that are involved. The following achievements have contributed greatly to closing the semantic gap. Local features [24] and representations based on codebooks [5] when combined, use the discriminating power of the former and the generalization abilities of the latter to perform high-level semantic tasks. Another achievement is the establishment of local features that describe motion because of the dual nature of the media (spatial and temporal) Specifically for video, the introduction of “motion-aware” local features [6] [13] [14] [15] [16] Also, the advent of machine learning algorithms example SVM [23] has led to an effective framework for complex classification tasks. The proposed scheme is simple, based on voting algorithms, a very good technique which has previously been utilized in tasks ranging from object detection [7], parameter estimation, and video classification. The scheme combines data from classifiers utilizing different features computed over different elements of the video, to generate a decision whether a videos content are unwanted/explicit or not (e.g. gore, nudity).

The crux is to retrieve the label given to video by several classifiers. These are counted and the majority vote determines the video's classification. The two main contributions to the project are: Firstly, the thorough evaluation of several combinations of descriptors on three applicative scenarios, which have consistently indicated that a representation based on spatiotemporal bags of features is more selective than all other possible choices. This is important because the current efforts still depend heavily on static features for video classification, example, majority of the methods for detecting violence depend on blood colour detection. The project also seeks to prove that the technique of majority voting accomplishes the difficult goal of filtering social media content. The architecture can be adapted to other high-level semantic tasks.

1.1 Problem Statement

The internet is inherently free of mediation. Although this quality is useful for sharing information from where all other forms of information have been suppressed, it also brings about the danger of allowing upload of explicit content like pornography and violence. Since impressionable, young children also use the internet the presence of such unwanted media could guide their thoughts dangerously. These unwanted media may be there without the knowledge of a website's administrators, the only prerequisite is that uploading of media is allowed on a website. Where the harm is more is when these media find their way on to different social media platforms. In essence, there is a need to establish filtering mechanisms on websites so that explicit media cannot be uploaded, so that young children and people who do not wish to see them, do not.

1.2 The Proposed Scheme

The proposed scheme is very simple and works by extracting elements from the video (shots, frames, key frames, etc.), extracting features from those elements (global features, bags of visual features based on local features, statistics, etc.) and training different classifiers for each type of feature used. In the classification phase, the classifier opinion is asked for each individual video element, and the final decision is reached by majority voting.

Pre-processing (video element and feature extraction) step

1. The elements of each video are extracted (shots, frames, key frames, etc.);
2. The features are extracted from the appropriate elements of the video. Those may be visual features, statistics, etc.

Training step

1. A SVM classifier is created for each type of feature. In our work, a linear kernel (which in preliminary tests, has offered the best results) is used;
2. Each classifier is trained with the corresponding features. Care is taken to balance the classes (positive and negative) so each is given roughly the same number of training samples at this step;

Classification step

1. Each SVM classifier is asked about each single feature concerning all elements of the video related to that feature (i.e., if a feature is computed over key frames, there will be a feature available for every key frame, and the corresponding classifier will be asked once for each one of those features);
2. Every time it is enquired, a SVM classifier casts a vote: positive (the video is “unwanted”) or negative (the video is “ok”);

3. Those votes are counted for all classifiers on all features concerning the video. The majority label is given to the video.

1.3 Objectives

To design and implement a web framework content-based filtering for video sharing in any social media that allows the sharing of videos.

1. To provide a platform of decency and free from pornographic materials.
2. To promote morality on the social media.
3. To prevent the stimulation of violent in viewers after watching violent scenes.

1.4 Significance of the study

The development of this project will be of tremendous support to all websites that do not want to allow unwanted pictures and videos to upload to their database. Indeed, the website with this web framework, can allow all Christians and Non-Christians to visit the site without feeling guilty of accidentally coming across incisive unwanted videos. It will be exciting to have such a project that could scrutinize the content of a video to detect that the video has insightful contents.

2. LITERATURE REVIEW

2.1 Background Theory of Web Framework Content Based Filtering for Picture Sharing

First Testbed Application: Pornography Detection

Pornography has become common on the web. The reason behind this is the fact that sites with this content tend to generate a lot of traffic which when coupled with ads on the pages, results in substantial income for such webpages. Pornography is not easily defined. Nudity, skin exposure define pornography but so too do other activities like sports, so solutions that depend on these attributes [2] , Pornography may be printed or visual material containing the explicit description or display of sexual organs or activity, intended to stimulate sexual excitement. The definition raises several obstacles. Firstly, what is the level of explicitness that qualifies a file as pornographic. Division of classes [1] is used by some solutions but this results in an ambiguous definition, and also results in complex classification task.

Test Database

The evaluation task is pretty simple, only two classes are generated (porn and non-porn). The two classes are however judiciously constructed to represent the diversity found in porn and non-porn media. In the pornographic class, a wide range of 400 samples which cover different genres and ethnicities have been used. As for the non-pornographic class, there are two subclasses easy and difficult. The easy subclass contains 200 samples that are relatively less complex to distinguish from porn, such as workers in full work clothing while the difficult subclass contains 200 samples that are relatively more complex to distinguish from porn, such as swimming, etc.

2.2 Previous Related Works and Recent Trends in Web Framework Content Based Filtering for Picture Sharing

2.2.1 Video Feature Extraction

Semantic classification of visual documents has only become feasible after the emergence of effective feature extraction algorithms. Many of those may be applied to video, some being just still-image descriptors of individual frames, others being specially conceived to take into account the spatiotemporal nature of the moving image. Though global image descriptors may be employed to characterize video frames, in the recent years a great deal of interest has been directed to local descriptors. Those are associated to different features of the image (regions, edges or small patches around points of interest) and have been shown to provide great robustness and discriminating power [17], [18], [24], [25], [26], [27].

The most popular local descriptor, SIFT [7], is both a point of interest detector, based on differences of Gaussians and a local descriptor, based on the orientations of grayscale gradients. Using SIFT, visual content is represented by a set of scale and rotation invariant descriptors, which provides a characterization of local shapes. The generated descriptors allow for adequate levels of affine, viewpoint and illumination invariance. Since colour information is considered important for many tasks (e.g., nude detection), colour extensions of SIFT have been proposed [9] [21] For example, a SIFT descriptor adapted to carry hue information (aptly named HueSIFT) had been proposed [9] It provides colour distinctiveness in addition to shape distinctiveness.

Intuition tells us that temporal information should be of prominent importance for recognition tasks in videos, for being likely to indicate interesting patterns of motion. Considering that, a few local features detectors and descriptors have been proposed, take into account the temporal nature of video [6] [13] [14] [15] [16] For example, STIP [6] is designed as a differential operator, simultaneously considering extrema over spatial and temporal scales that correspond to particular patterns of events in specific locations. It extends the Harris corner detector [8], to the temporal domain, finding interest points as moving corner changes direction across a sequence — if the corner movement is constant, no interest point is detected. This allows detecting noteworthy “events” on the video sequence.

2.2.2 Codebooks of Visual Features

The discriminating power of local descriptors is extremely advantageous when matching objects in scenes or retrieving specific target documents. However, when considering high-level semantic categories, it quickly becomes an obstacle, since the ability to generalize becomes then essential. A solution to this problem is to quantize the description spaces by using codebooks of local descriptors, in a technique sometimes named “visual dictionary”. The visual dictionary is nothing more than a representation which splits the descriptor space into multiple regions, usually by employing non-supervised learning techniques, like clustering. Each region becomes then a “visual word”.

The idea is that different regions of the description space will become associated to different semantic concepts, for example, parts of the human body, corners of furniture, vegetation, clear sky, clouds, features of buildings, etc. Once the codebook is obtained, description is greatly simplified, since it is no longer based on the exact value of descriptors, but only on their associated “words”. The condensed description may be, for example, a histogram or simply a set of the words the video contains. This has two advantages: the rougher description is better adapted to complex semantics; and the computational burthen is alleviated, since algorithms now operate on a single summarized description, instead of a myriad of individual local descriptors.

Building the dictionary requires the quantization of the description space, which can be obtained by a clustering algorithm. However, state-of-the-art clustering methods are seldom (if ever) conceived for the needs of visual dictionary construction: high-dimensional spaces, large datasets and a large number of clusters. The commonest choice found in the literature is a combination of aggressive sub-sampling of the dataset, dimensionality reduction using PCA (Principal Component Analysis), and clustering using a simple or hierarchical k-means algorithm with Euclidean distance. This typical choice however, may be considerable faulty on several grounds [22] and the design of good methods for visual dictionary creation is an active, open area of inquire.

In addition to moderating the discriminating power of descriptors, the dictionaries allow adapting to visual documents techniques formerly available only to textual data. Among those borrowings, one of the most successful has been the technique of bags of words (which considers textual documents simply as sets of words, ignoring any inherent structure). The equivalent in the CBIR universe has been called bags of visual words, bags of features or bags of visual features, sometimes abbreviated as BoVF.

The straightforward extension of BoVF to video uses individual frame images (or selected keyframes). This allows representing semantic concepts which are independent from motion. However, previous works in human annotation of video databases [40] indicate that even for humans, many important concepts can only be adequately apprehended by taking into account the temporal aspects of video. Therefore, an interesting possibility is making codebooks of space-time local descriptors [6] which take into account the dynamic aspects of video. In this work the performance of both static and "motion-aware" bags of features are utilized.

2.3 Drawbacks, Problems and Shortcoming of Related Works

The previous related works have considered various based papers which address a specific area of problem. It indicated problem when tested with filtering unwanted scenes in video content. In view of this project, BoVF(Bag of Visual Features) is to be used to solve the problem.

2.4 What Makes This Project Better

The idea of preventing a website to store unwanted scenes such as pornographic clips, violent actions, is laudable and very supportive to instilling discipline and morality in humanity. The BoVF is an approach that appropriately filters effectively the unwanted scene in videos. Most of the social networks have been restricting some users from accessing their account platforms because of disrespectful pictures and videos that have landed on the global platform.

The implementation of this web framework can avoid users uploading unwanted scenes unto social media. It is a new emerging technology, content-based filtering, that researchers are digitizing various application to avoid or restricting uploading of files unto web servers. Similar knowledge of content-based applications has been implemented in the areas of sending messages, files, documents, etc. However, the sensitive information has been encrypted with various algorithms to hide its contents from scrutinizing to avoid revealing its contents to unscrupulous people (hackers).

At times unscrupulous people gain access to explicit media of other people and use this to extort funds from their targets who often agree because the consequence of not accepting can be dare, such as the explicit picture being shared on web platforms. This project aims to stem the tide of such nefarious activities.

3. SYSTEM DESIGN AND IMPLEMENTATION

3.1 Proposed Work

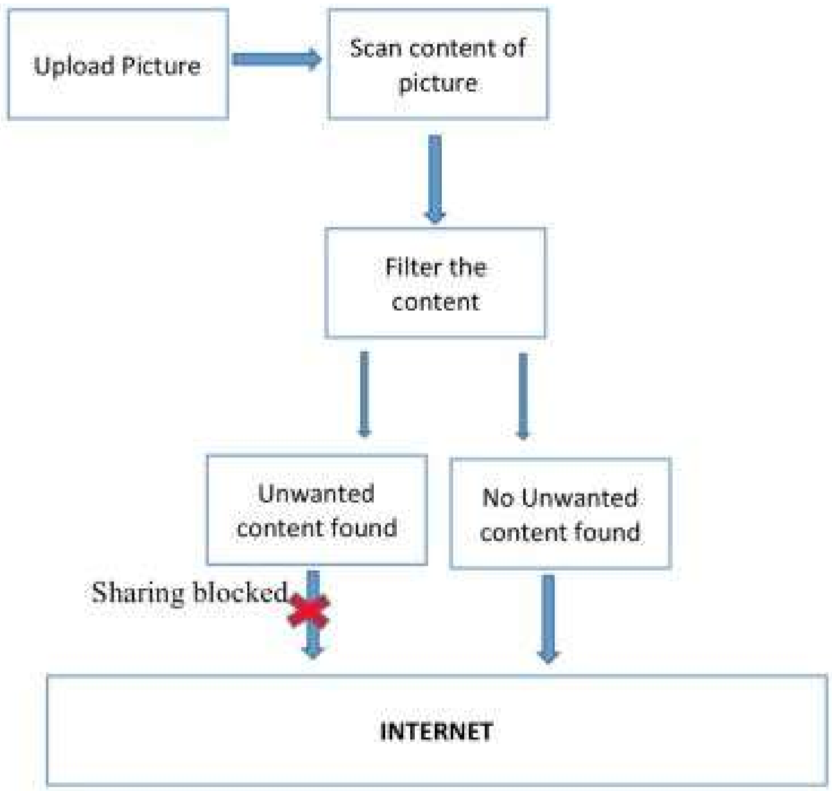

Figure 1. Proposed work

Upload Media
The user attempts to upload his/her media. The media is kept in a temporary location for action on it.

Scan Content
The stored image is scanned by extracting its features using SIFT (Scale-invariant Feature Transform) and generating BoVF (Bag of Visual Features)

Filter Content
The extracted visual features of the scanned image are compared to those of the classification classes.

Unwanted Content Found
After filtering the media, unwanted content if available will be detected in the media.

Block Upload
If unwanted content is found in the media, the upload is blocked.

3.2 Design of The Proposed System

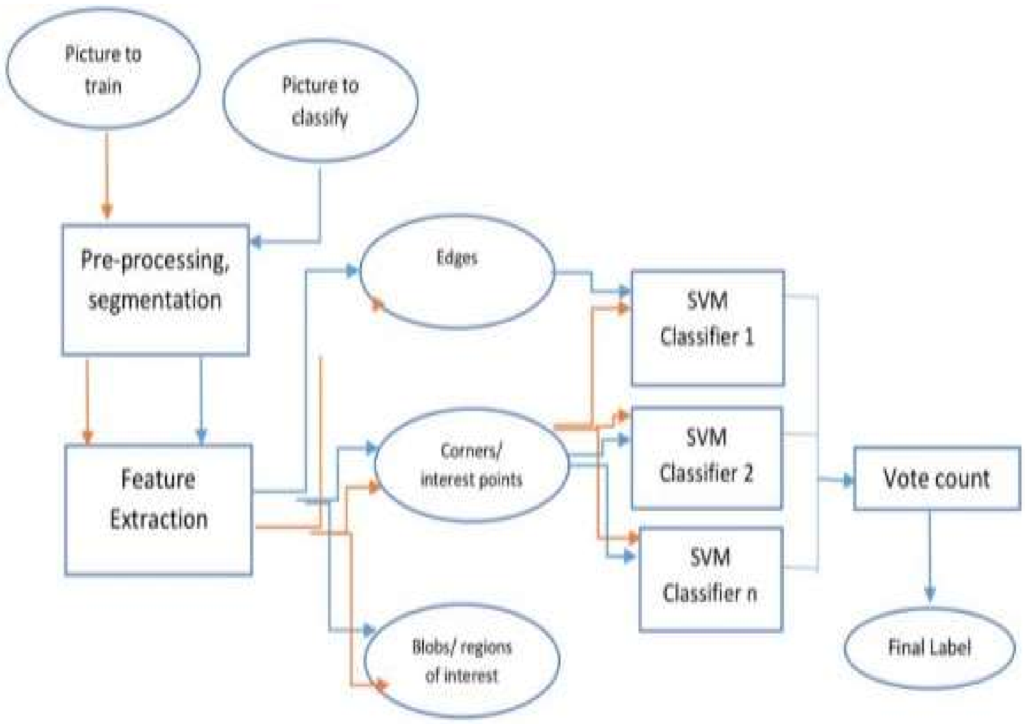

Figure 2. Architectural Diagram of the proposed system

3.3 Components of the Proposed System

Picture to train
Uploading pictures that will be trained and stored to be used to compare any other video that will be uploaded.

Picture to Classify
This is the picture to upload to test if the application will filter the content to ascertain if the application will be apple to predict the nature of the content.

Pre-processing and segmentation: This does the process and segmenting the content.
Feature Extraction
This extracts the necessary features for the identification if is an unwanted scene or not.

Feature Type
This describes of the type of scene extracted from the picture.

SVM Classifiers
These categorize files under violent picture or picture, or pornographic picture

Vote Count
This indicates the total number of histogram colour representation for segmented trained section of the picture.

Final Label
This label the final classification as picture with pornographic contents or violent scene.

3.4 Methodology
The designing will follow the concept of software development life cycle, and the waterfall model as the benchmark for the visual programming. This project considers the implementation of content-based filtering using a web-based application technology using to generate the web pages:

- HTML: To build main framework of the web page.
- CSS: To design the layout of the web pages.
- JavaScript/JQuery: To program the behaviour of the web pages.
- PHP: For capturing, analysing and manipulating pixel values of images to be filtered.
- MySQL: To control database interactions and activities.

3.5 Algorithms

Experimental setup
In this experiment, the evidence fusion scheme has been parameterized as follows:

1. The two classes considered were stuffing (positive) and legitimate (negative).
2. The video elements considered were the entire video and key frames extracted using a State-of-the-art static video summarization method [19].
3. These features were extracted for each key frame:
    a) Global Features:
    b) Color Histogram: a normalized 256-bin RGB color histogram
    c) Hue Histogram: a normalized 256-bin histogram of the hue component of color;
    d) Zernike Moments: the 10 first Zernike moments;
    e) Local Features:
    f) SIFT-BoVF: a 5000-bin normalized BoVF using the SIFT descriptor [7];
    g) PCA-SIFT-BoVF: a 5000-bin normalized BoVF using the PCA-SIFT descriptor [20].
4. These features were computed for the entire video:
    a. Local Features:
    b. STIP-BoVF: a 5000-bin normalized BoVF using the STIP descriptor (Laptev, 2005);
    c. Statistics extracted by the video summarization method
    d. The number of key frames in the video
    e. The ratio between the number of extracted key frames and the original number of frames.

The key frame statistics (used in step 4) are intended as a rough measure of the complexity of the video. Since the approaches found in literature [28] [29] are not comparable to ours — one is concerned with the classification of users; the other does not address the concept of ballot stuffing. Instead, I have evaluated how different feature choices affected the result. The experimental design was a classical 5-fold cross-validation, generating approximately 800 videos for training and 200 for testing on each fold.

3.6 Data Flow Diagram of the Proposed System

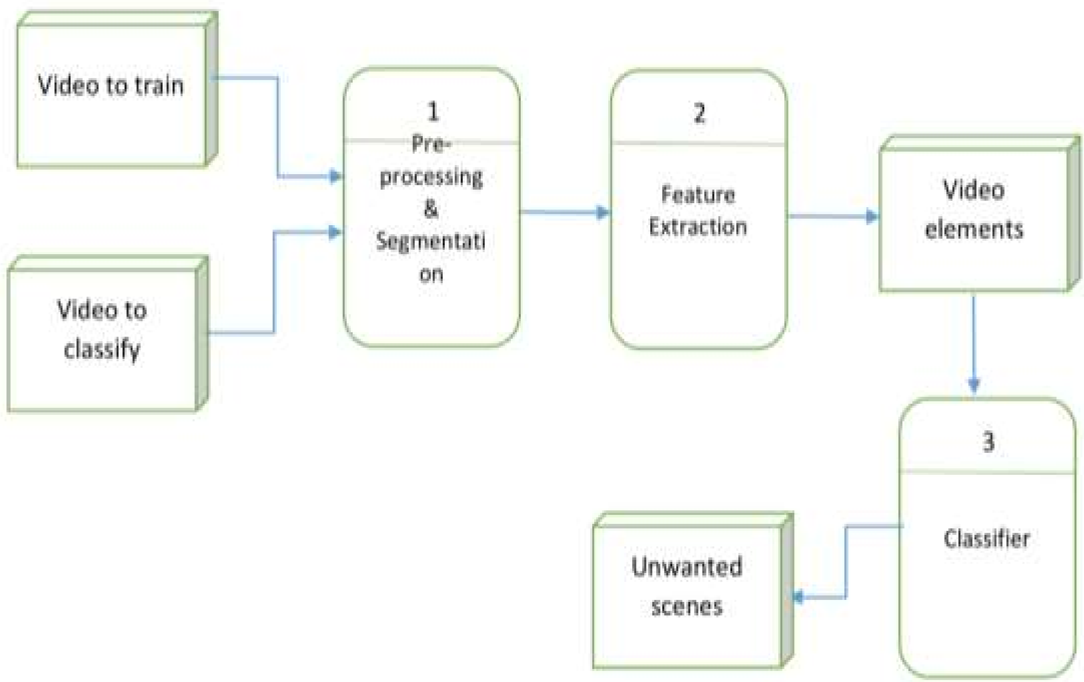

Figure 3. Data Flow Diagram of the proposed system

Video or picture to train/Video or picture to classify/Pre-processing & Segmentation and Feature Extraction: The elements of each video are extracted (shots, frames, keyframes, etc.); The features are extracted from the appropriate elements of the video. Those may be visual features, statistics, etc. A SVM classifier is created for each type of feature. In this work, a linear kernel is used.

Video elements/ classifier/ Unwanted scenes: Each SVM classifier is asked about each single feature concerning all elements of the video related to that feature (i.e., if a feature is computed over keyframes, there will be a feature available for every keyframe, and the corresponding classifier will be asked once for each one of those features); Every time it is enquired, a SVM classifier casts a vote: positive (the video is “unwanted”) or negative (the video is “ok”); Those votes are counted for all classifiers on all features concerning the video. The majority label is given to the video.

4. TESTING AND EXPERIMENTAL RESULTS

4.1 Domain and Platform

This project is about filtering the content of videos uploaded unto the website. Videos are frames put together. The application scans and filters the frames streamed from videos. The web-based application considers PHP/MySQL as the platform for the development of the application. The XAMPP server allows the localhost configuration for a standalone machine. The localhost does not need internet connectivity; therefore, it is tested locally on the XAMPP server installed on the local machine.

The maximum size that can be uploaded can be resized in the PHP configuration files. In storing the images, a database is needed to keep the records so MySQL is also used, which is part of the XAMPP server.

4.2 Testing Methodology
The testing of the application was to ensure that it works purposely for what the objectives have been set. This application was designed to filter unwanted images and videos unto the website, so images that are not accepted to be watched are used to test the application.

4.3 Test Method
The quality of the application was satisfied through the use of the below test methods, to ensure that the application has the standard quality to be hosted for global use. In view of the testing, the application was tested by applying system test, performance test, stress and volume test, and recovery test for the database.

4.4 System Test
In the development of the application, each module was considered to ensure the errors are properly fixed since one unit can impact negatively on the entire application thereby making the application redundant. The modules were then integrated and tested to ensure that the application meets its standard to be hosted. The unit and integration tests were successful so the system test ensures that the application could be tested for its performance.

4.5 Performance Test
The application performance tests were based on the number of videos and images uploaded to identify if the application could measure the percentage in which it could detect an image does not qualify to be uploaded unto the website. The following performance metrics were considered in the testing: Performance metrics: The percentage of scanning images was considered

Table 1. This Shows the Various Percentages of Files Uploaded

Image size	Dimension(pixels)	Type	Percentage %	Action
27.5KB	275x206	JPG	40.44	Blocked
19.0KB	366x206	JPG	84.14	Blocked
19.3KB	206x309	JPG	43.87	Blocked
450KB	387x206	GIF	0.0	None

In table 1, it indicated that the application was able to scan through jpg files and blocked all of them with their respective percentages. The image with image size 19.0 recorded 84.14 percent, which shows that the image uploaded had more naked part.

4.6 Stress and Volume Test
The application accepts videos and images so the server will be very heavy. The details of every video and image uploaded will also be stored in the database. The test was conducted on the processing capacity of the site when numerous videos and images are being uploaded simultaneously. The application was able to withstand the huge uploads of files, and it also function very well.

4.7 Recovery Test
The database and directory that store the videos and files will be backed up periodically so that if there is any damage to the server, the data will not be lost.

4.8 Snapshots of Test Results

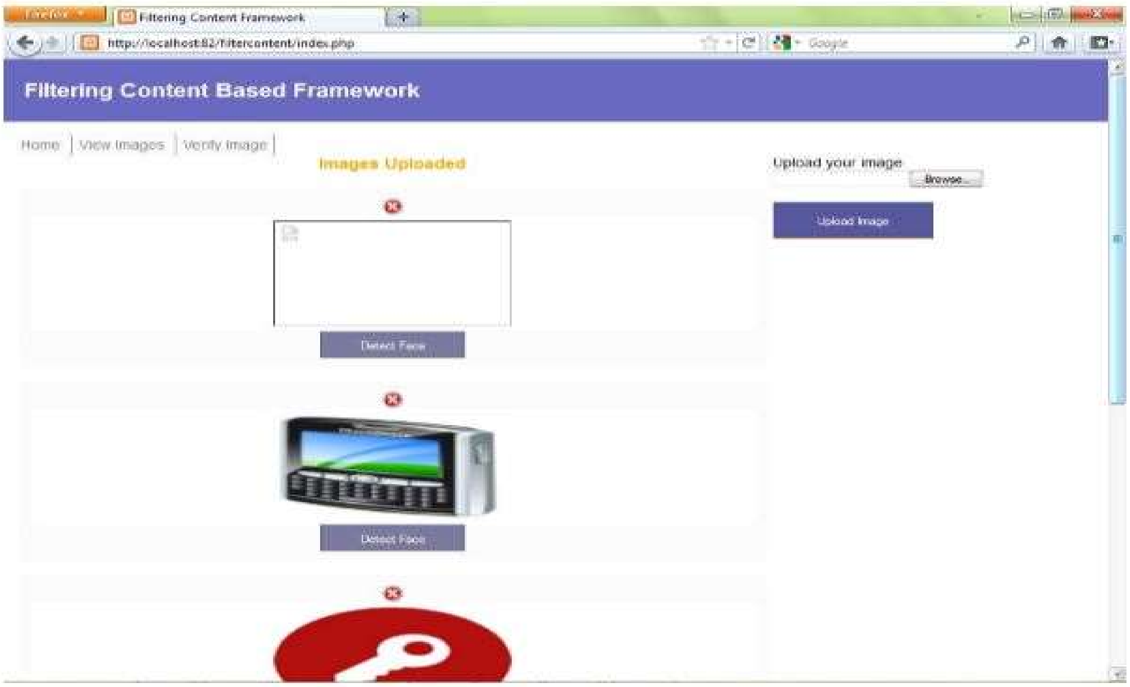

Figure 4. Showing the home page of the video and image filter framework

Figure 5. Image upload

Figure 5: Shows the uploaded image and the percentage indicated as 48.00%, Here, the image was blocked by the application after trying to upload

Figure 6. image upload
Figure 6: Shows another uploaded image and the percentage indicated as 84.14%

Figure 7. Image Upload
Figure 7 Showing another image uploaded with 53.42%

Figure 8. Image upload

Figure 8 Shows another image trying to be uploaded and the application blocked based on the 46.87%.

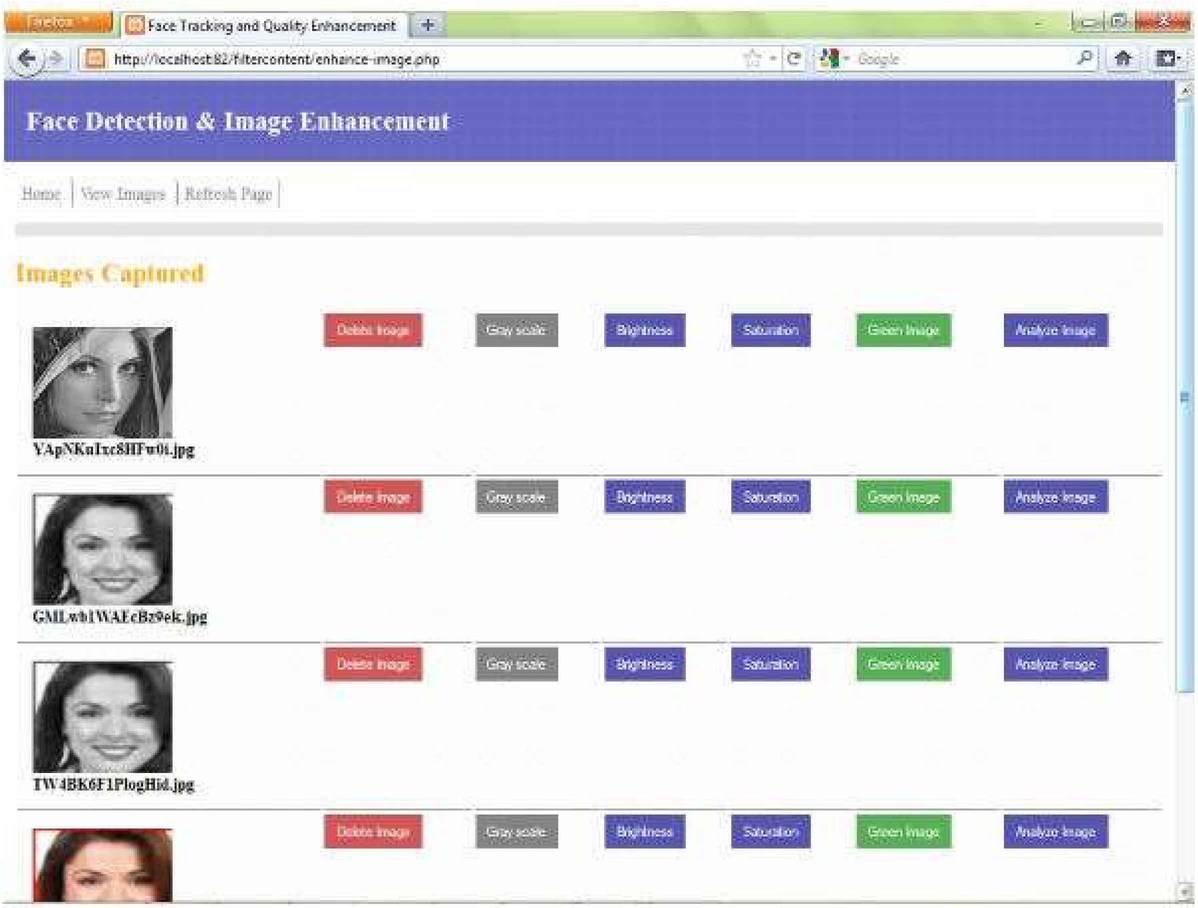

Figure 9.Shows all the uploaded images and can now undergo manipulation

4.9 Findings

The application designed was tested by a group of people who were made to upload any image that they find, including pornographic pictures. The pictures uploaded by the people; somewhere blocked others were uploaded successfully unto the site. The following were the comments that they made: First person: "I think the social media can employ this web framework and integrate it into their application to screen all the videos and images that are uploaded unto the social media page. This will at least reduce the unwanted pictures on the social media." Second person: "I am really interested in this application and I will integrate it to my company website so that anyone who wants to upload unwanted videos or pictures can be prevented. I will like to recommend to all the social media to prevent people from uploading unwanted images."

4.10 Graphs and Evaluations

The user acceptance indicates the number of users who tested the application and recommended to be used by social media to identify any unwanted videos or images that are uploaded unto the server. The users were identified whether they would like to have this added to their websites to screen any video or images, and the Figure 10 and Table 2 show the chart and the table with percentages respectively. In the findings some of the users made their comments on the application.

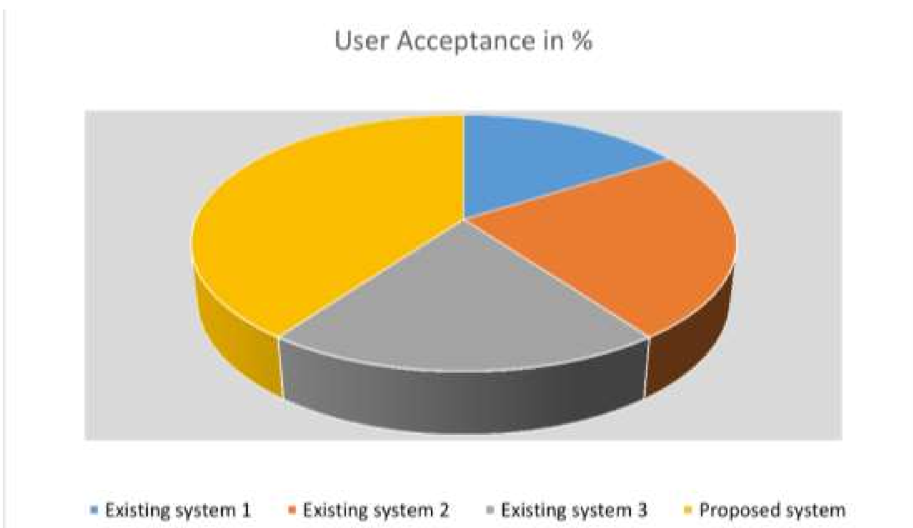

Figure 10; The chart showing the user acceptance

Table 2.The table showing the percentages of the acceptance of the proposed system over the existing system by the users.

System	User Acceptance in %
Existing system 1	16
Existing system 2	24
Existing system 3	20
Proposed system	40
Total	100

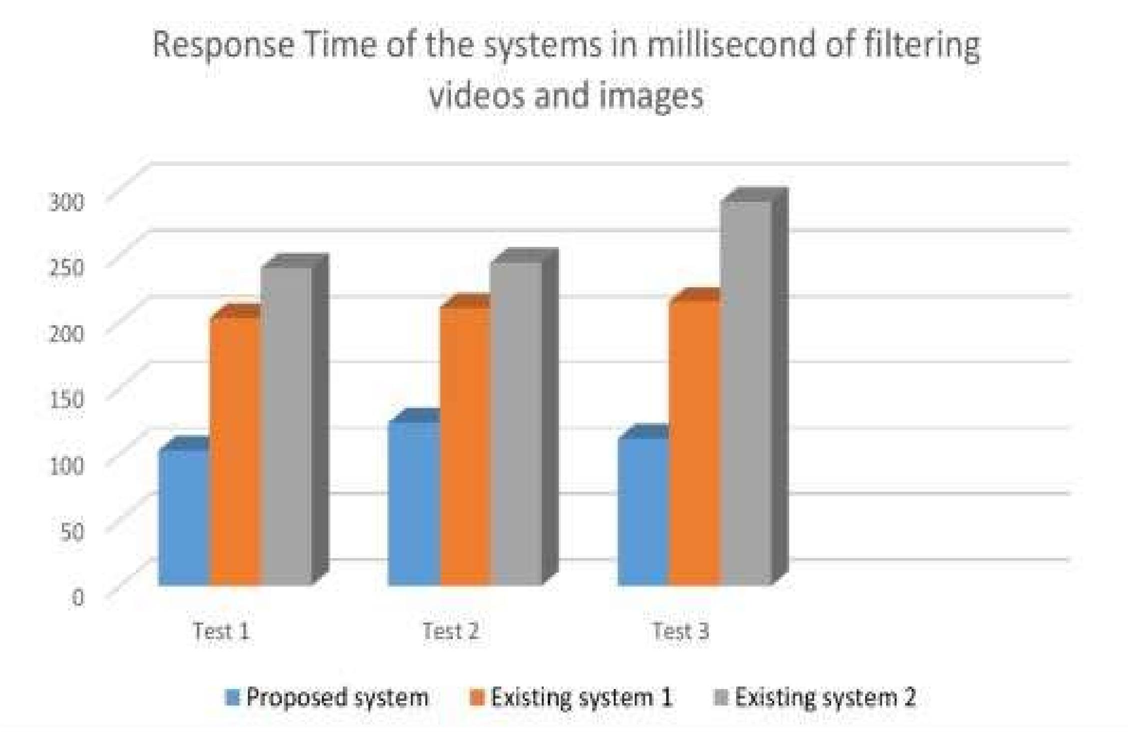

Figure 11. The response time of the systems in millisecond of filtering videos and images.

In Figure 11, it shows that the proposed system responded to the user during uploading and filtering of the videos and pictures.

Table 3: The values showing the various response time of the filtering of videos and images in millisecond.

Test	Proposed system	Existing system 1	Existing system 2
Test 1	102.45	202.76	240.68
Test 2	123.43	210.58	244.87
Test 3	111.54	215.34	290.68

Table 4: Advantages of the proposed system over the existing system

Feature	Proposed System	Existing System
Site protection	It protects the site from accepting unwanted pictures and videos.	It does not protect the site from accepting unwanted pictures and videos.
Integrity	This site ensures the integrity of the company since no illicit picture or video can be uploaded unto the server.	Illicit picture which damages the company's integrity can be uploaded. This is so because there are no checks.
Moral life	It promotes moral life on the site.	It does not promote moral life on the site.

5. SUMMARY AND CONCLUSION

5.1 Summary

The development of this application was based on filtering the content of a video or image before it can allow you to upload, and manipulate and then send. The application can be employed by any website for the protection of their website.

5.2 Conclusion

The application designed will be effective in promoting and protecting social media contents. Any company with a website that allow images to upload can encounter a problem of receiving any image from the users. Sometimes, the company may not be aware of the images in the server, so to prevent unaware of displaying pornography films and images, it is therefore good to integrate the website with an application that will prevent people from uploading unwanted images.
The filtering techniques determine the content based on the number of pixels on the frame. A particular pixel can fall within a certain colour that is used to describe the colour of the person. In most cases, every human being has different body texture and colour. The application considers the texture and the colour within a certain designed colour combination as considered in image processing. This made the application to be able to detect a video or image with pornographic content.

REFERENCES

[1] T. Deselaers, L. Pimenidis and H. Hey. "bag-Of-Visual-Words Model for Adult Image Classification and Filtering", In: Proceedings of the International Conference on Pattern Recognition (ICPR '08), PP. 1-4, 2008.

[2] W. Kelly, A. Donnellan, D. Molloy. "Screening for Objectionable Images: A Review of Skin Detection Techniques", In: International Machine Vision and Image Processing Conference (IMVIP'08), pp. 151-158, 2008.

[3] W. Kim, S.J. Yoo, J-s. Kim, T.Y. Nam, and K. Yoon. "Detecting Adult Images Using Seven MPEG-7 Visual Descriptors", In: Web and Communication Technologies and Internet-Related Social Issues – HSI 2005, PP. 336-339, 2005.

[4] A. P. B. Lopes, S. E. F. de Avila, A. N. A. Peixoto, R. S. Oliveira, and A. de A. Araújo. "A Bag-of-Features Approach Based on Hue-SIFT Descriptor for Nude Detection", In: European Signal Processing Conference (EUSIPCO'09), pp. 1552- 1556, 2009.

[5] J. Sivic and A. Zisserman. "Video Google: A Text Retrieval Approach to Object Matching in Videos", In: Proceedings of the IEEE International Conference on Computer Vision (ICCV'03), pp. 1470-1477, 2003.

[6] I. Laptev. "On Space-Time Interest Points", In: International Journal of Computer Vision (IJCV'05), vol 64, number 2/3, p.107-123, 2005.

[7] D. G. Lowe. "Distinctive Image Features from Scale-Invariant Keypoints", In: International Journal of Computer Vision (IJCV'04), vol. 60, no. 2, pp. 91-110, 2004.

[8] C. Harris and M. Stephens. "A Combined Corner and Edge Detector", In: Alvey Vision Conference, pp. 147-152, 1998.

[9] K. E. A. van de Sande, T. Gevers and C. G. M. Snoek. "Evaluating Color Descriptors for Object and Scene Recognition", In: IEEE Transactions on Pattern Analysis and Machine Intelligence (in press), 2010.

[10] A. P. B. Lopes, S. E. F. Avila, A. N. A. Peixoto, R. S. Oliveira, M. M. Coelho and A. A. Araújo. "Nude Detection in Video Using Bag-of-Visual-Features", In: Proceedings of the 22th Brazilian Symposium on Computer Graphics and Image (SIBGRAPI), pp. 224-231, 2009.
[11] Y. G. Jiang, C. W. Ngo and J. Yang. "Towards Optimal Bag-of-Features for Object Categorization and Semantic Video Retrieval", In: Proceedings of the 6th ACM International Conference on Image and Video Retrieval (CIVR '07), pp. 494- 501, 2007.
[12] J. Yang, Y.-G. Jiang, A. G. Hauptmann and C.-W. Ngo. "Evaluating Bag-of-Visual-Words Representations in Scene Classification", In: Proceedings of the International Workshop on Multimedia Information Retrieval (MIR'07), pp. 197-206, 2007.
[13] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior Recognition via Sparse Spatio-Temporal Features", In: Proceedings of the 14th International Conference on Computer Communications and Networks (ICCCN'05), pp. 65-72, 2005.
[14] H. Ning, Y. Hu and T. Huang. "Searching Human Behaviors Using Spatial-Temporal Words", In: Proceedings of the IEEE International Conference on Image Processing (ICIP'07), pp. 337-340, 2007.
[15] J. C. Niebles, H. Wang, and L. Fei-Fei, "Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words", In: International Journal of Computer Vision (IJCV'08), vol. 79, no. 3, pp. 299-318, 2008.
[16] Y. Ke, R. Sukthankar, and M. Hebert, "Spatio-Temporal Shape and Flow Correlation for Action Recognition", In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'07), pp. 1-8, 2007.
[17] K. Mikolajczyk and C. Schmid. "An Affine Invariant Interest Point Detector", In: Proceedings of the 7th European Conference on Computer Vision-Part I (ECCV'02), vol. 2350 of Lecture Notes in Computer Science, Springer Verlag, Berlin, Copenhagen, Denmark, pp. 128-142, 2002.
[18] T. Tuytelaars and L. Van Gool. "Wide Baseline Stereo Matching Based on Local, Affinely Invariant Regions", In: British Machine Vision Conference (BMVC'00), pp. 412-425, 2000. [19] S. E. F. Avila, A. Luz Jr. and A. A. Araújo. "VSUMM: A Simple and Efficient Approach for Automatic Video Summarization", In: International Conference on Systems, Signals and Image Processing (IWSSIP'08), pp. 449-452, 2008.
[20] Y. Ke and R. Sukthankar. "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors". In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04), vol. 2, pp. 506-513, 2004.
[21] A. E. Abdel-Hakim and A. A. Farag. "CSIFT: A SIFT Descriptor with Color Invariant Characteristics", In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), pp. 1978-1983, 2006.
[22] F. Jurie and B. Triggs. "Creating Efficient Codebooks for Visual Recognition", In: Proceedings of the IEEE International Conference on Computer Vision (ICCV'05), vol. 1, pp. 604-610, 2005.
[23] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines", 2001, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[24] T. Tuytelaars and K. Mikolajczyk. "Local Invariant Feature Detectors: A Survey", In: Foundations and Trends in Computer Graphics and Vision. vol. 3, no. 3, pp.177-280, 2008.
[25] K. Mikolajczyk and C. Schmid. "Indexing Based on Scale Invariant Interest Points", In: Proceedings of the IEEE International Conference on Computer Vision (ICCV'01), pp. 525-531, 2001.
[26] C. Schuldt, I. Laptev and B. Caputo. "Recognizing Human Actions: A Local SVM Approach", In: Proceedings of the International Conference on Pattern Recognition (ICPR'04), pp. III: 32-36, 2004.

[27] K. Mikolajczyk and C. Schmid. “A Performance Evaluation of Local Descriptors”, In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, n. 10, pp. 1615-1630, 2005.
[28] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida and M. Gonçalves, “Detecting Spammers and Content Promoters in Online Video Social Networks”, In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 620-627, 2009.
[29] R. Crane and D. Sornette, “Robust Dynamic Classes Revealed by Measuring the Response Function of a Social System”, In: Proceedings of the National Academy of Sciences, 105(41):15649-15653, 2008.
[30] J. Demšar. “Statistical Comparisons of Classifiers over Multiple Data Sets”. In: The Journal of Machine Learning Research, vol. 7, 1-30, 2006.