International Journal of Pure and Applied Mathematics
Volume 119 No. 15 2018, 325-346
ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue
http://www.acadpubl.eu/hub/

Analysis Study of Spam Image-based Emails Filtering Techniques

Yasmine K. Zamel*, Suhad Ahmed Ali and Mohammed Abdullah Naser

Department of Computer Science, College of Science for Women, University of Babylon, Babylon, Iraq

Yasakhalid88@yahoo.com, wsci.suhad.ahmed@uobabylon.edu.iq, and wsci.mohammed.abud@uobabylon.edu.iq

Abstract

Solving Email has become increasingly important and widespread method of communication because of its time speed and cheap. It is used not only for personal contacts but also for business. The number of email messages received per day may be thousands for companies. The spam emails are mainly an illegal form where can reach to thousands of users easily. This type of emails has become a serious problem.

The major problems are concerning with spam for the user is time and space waste; also, Messages can contain things that are offensive to people and can cause general psychological discomfort. A large amount of spam can disable unsolicited email, and personal messages can easily be lost. So there is an urgent need to control the growing spam flood.To resolve the problem there is a growing need for emails classification method which may be based on content, text properties and features.

Recently spammers have adopted a new style of spam; that is the image spam trick to make the analysis of emails' body text inefficient. Image Spam is an e-mail solicitation that uses graphical images of text to avoid text-based filtering techniques, the ambiguity to prevent recognition by optical character recognition (OCR)tools.

In this paper, we will provide a comprehensive survey and a complete classification of the types of spam image based e-mail, the most important tricks used, and the prominent filtering techniques used with these types of unwanted messages. Also, review of previous works and various problems and challenges faced, in addition, to present the standards metrics to evaluate the performance of filtering techniques.

Keywords: Spam Image, Filtering Techniques.

1. Introduction

The growing use of internet has promoted an easy and fast way of e-communication, the well-known example for this is e-mail. Nowadays, sending and receiving the e-mail as a means of communication is very popularly used [1]. Email benefit is exceptionally modest, no take time, a client can transmit and receive continuously, only in real time, best for official interchanges, evacuates time-zone restrictions, and so on; next to the all specified focal points, the issue of unwanted [2]. As indicated by numerous review reports, the measure of Unwanted mails, As called spam emails, spam mails are the emails sent by some unknown sender just to hamper the development of Internet e.g. advertisement and many more.

The continuous growth of spam mails that is, the bulk delivery of unimportant emails, mainly of commercial nature or with unpleasant content. This can become the main problem of the email service for Internet service provider(ISP) [1]. Since they make tremendous misfortunes for the Institutions, begin from bandwidth utilization, mail server preparing load, to client's profitability because of time spent to find and managing spam email sender [3]. These spams message not just increase the system correspondence and memory space, yet can likewise be utilized for some assault. This assault can be utilized to wreck client's data or to take important personal information like account no. and passwords [4,1].

The recent study survey of the email server has reported 60% of all email traffic is spam, so the need to develop an anti-spam mail filter is obligatory. Current spam mail filters are made to detect various features of spam mail. Especially, text categorization technique is used for filtering the spam mails. However, spammers introduced the new technique of embedding the spam mails in the attached image in the mail [1], experience image spams as another danger; so far this is the most modern sort of spam email. It is a stellar essential one since it makes message intriguing for the client and difficult for recognize. In the earlier year, spammer put the messages in email's body; it is a customary path for spammers. There are a wide number of text-based antispam filters that defeat the conventional method for spamming [2]. Today spammers evade images dodge image in three methods, implanting in email body appending an image document to messages,messages, putting a hyperlink in the email's body and the objective message is conveyed by the image. Text-based spam filtering techniques not success and careless in recent to recognize the new spammers' methods image spams are rich in substance and assortment of kind and image spam might be a one image or it can contain a many images inside a one image, so image spam filters apply distinctive and different technique on their filters to have the capacity to recognize image spam planning to fight the image spammers. The most straightforward strategy is to utilize Optical Character Recognition(OCR) tool to find text and extracted from image spam and investigating the text. Some of them are utilizing feature extraction techniques from image to analyze it [2]. Spam images are represented in Fig1.

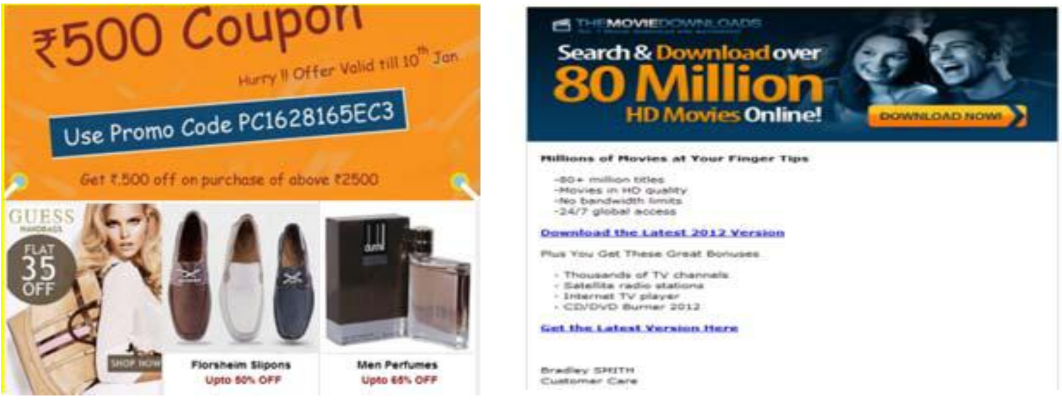


Figure 1: Examples of spam images.

The rest of the paper is organized as follows. In section 2, a brief review of present related works. Section 3 provides a general overview of the spam image types and filtering techniques. Section 4 reviews emerging approaches to spam filtering and discuss the challenges that these approaches present. In section 5 presents the important measures of performance evaluation. Section 6 concludes some conclusions. Finally, Section 7 lists the references of this paper.

2. Literature Survey

There are two main approaches to filter emails based on spam image, first those low-level based methods, and the second is based on OCR. So we will review a number of important and modern research that belong to both types.

2.1 Based on low-level image feature.

Low-level image features; sometimes called visual features, consist of several features, such as texture, color, shape, and appearance.Below several proposed works that operate according to this type of ways.

Rui Chan [5-2017].The proposed system includes three-layer spam filtering. Spam is filtered by analyzing both the header and the image.The structure of the model explicates carefully the idea of the design and many technologies related to the model. Experimental results show that this system has a satisfactory filtering effect.
Monireh Sadat Hosseinia et. al [2015] Suggested method, image texture feature was used to classify the spam image. The gray level co-occurrence matrix has been applied to each image. The properties obtained are 22 features and then the k-nearest neighbor classifier and naive Bayesian are used to evaluate the images obtained from the both of works database Dredze and Image Spam Hunter[6].
T. Kumaresan et. al [7-2015] suggested a scheme which extracts the features especially low-level features (like metadata and histogram features of images). An SVM classifier with kernel function is used to identify a spam image based on extracted features, the accuracy of this method 90 %, but the time complexity still is a problem in this work.

Jianyi Wang et. al [8- 2014] proposed an approach that was based on combines the characteristics of spam images with the corner point density to detect. The general idea of the algorithm is based on the corner proportion of the images to judge if it is a spam or not spam.

Minal Kamble and L. G. Malik [9 - 2012] design a spam detection system. The system will analyze the image content based on color histogram features for feature extraction, using PCA to selection feature and by using SVM to classify the embedded image as spam or legitimate hence classify the email accordingly.

Basheer Al-Duwairi et. al [10-2012] suggested a way to filter unwanted images that based on image texture analysis and used low-level features of the image to characterize them. They have been evaluated for performance by a collection of classifiers and these works are for the most important. These classifiers are C4.5 decision tree, support vector machine, multilayer perception, naïve bays, Bayesian network. The experimental results show that the random forest classifier outperforms all other classifiers has 98.6% an average precision, recall, accuracy, and f-measure.

M.Soranamageswari et. al [11-2011].The proposed method for filtering spam images in email is by using a gradient histogram to classify images as ham or undesirable images.

2.2 Based on OCR image feature.

OCR can be defined as an electronic or translating the mechanical nature of scanned images from handwritten, written or printed texts into machine-encoded text.There are many research works that have been proposed in this regard, such as:

V. Sathiya et. al [12-2011]. The suggested method is to extract the text from the image by using OCR tool. E-mails in this work are divided into three models text character, special character and image feature, which are evaluated for these models as image and text classification models, the corpus of image spam dataset used for image classification models and corpus of vocal spam used for text classification models After that, mail will classify as certified E-mail and uncertified E-mail

Pattaraporn Klangpraphant et. al [13-2010]. The suggested way called as Partial Image Spam Inspector (PIMSI). This method can detect a set of image spam e-mail, the method comprises on spam that consists from texts and images in the e-mail system.The important feature of this technique is that it will sensitively protect the distribution of all messages which have a partial similarity of spam e-mail. As a result, users get rid of spam in the mailbox.

Giorgio Fumera et. al [14-2006]. The proposed approach used depends on the use of OCR tools to extract the hidden text in the image The effectiveness of this approach is evaluated depended on two large corpora of spam e-mails and used SVM classifier for classification.

2.3 Based on two types (low level and OCR image feature)

The following works use both ways that mentioned above:

Nisha D. Chopra et. al [1-2015]. In this paper the researcher has used two methods first method using OCR tool for separating text from the image, the second method is used a Bayesian algorithm to detect the words in the mail are spam or not spam.

Meghali Das et. al [16- 2014] propose a method that based on analyzing the image that contains only a text region. Then classify the embedded image as spam or legitimate accordingly, the dataset used Dredze.

Anand GuptaIn et. al [15-2012] present two methodologies for addressing image spam problems, methodology(1) extracts low-level features while methodology(2) extracts metadata features of the image. Both methods are applied only to the images that contain only text regions. In both methods, text regions are extracted and then remove noise from these regions in order to input to OCR system.

The below table gives more details on all the works being reviewed:

Table 1: The related works with some details.

Related work	Public year	Techniques used to filter/classify images spam	The accuracy rate
[5]	2017	Multi-layer algorithm	96.2%
[6]	2015	k-nearest neighbor classifier (KNN) and naive Bayesian (NB)	-Acc1=91/41 knn,75/49 NB -Acc2=93/74 knn,99/19 NB
[7]	2015	Support Vector Machine and Particle Swarm Optimization.	90%
[8]	2014	Thresholding	91.3%
[9]	2012	SVM Classifier	98%
[10]	2012	They are: Support Vector Machine, C4.5 Decision Tree, Bayesian Network, Multilayer Perception, Naïve Bays and Random Forest	98.6%.
[11]	2011	gradient histogram	93.7%
[12]	2011	OCR	Not defined
[13]	2010	OCR	Not defined
[14]	2006	OCR	Not defined
[1]	2015	OCR and Bayesian Algorithm	Not defined
[16]	2014	Content analysis	Not defined
[15]	2012	OCR	- methodology 1 is 0.92 - methodology 2 is 93.3%.

3. An overview of spam Image

In this section, we address several important concepts, including:

3.1 Types and aims of spam images

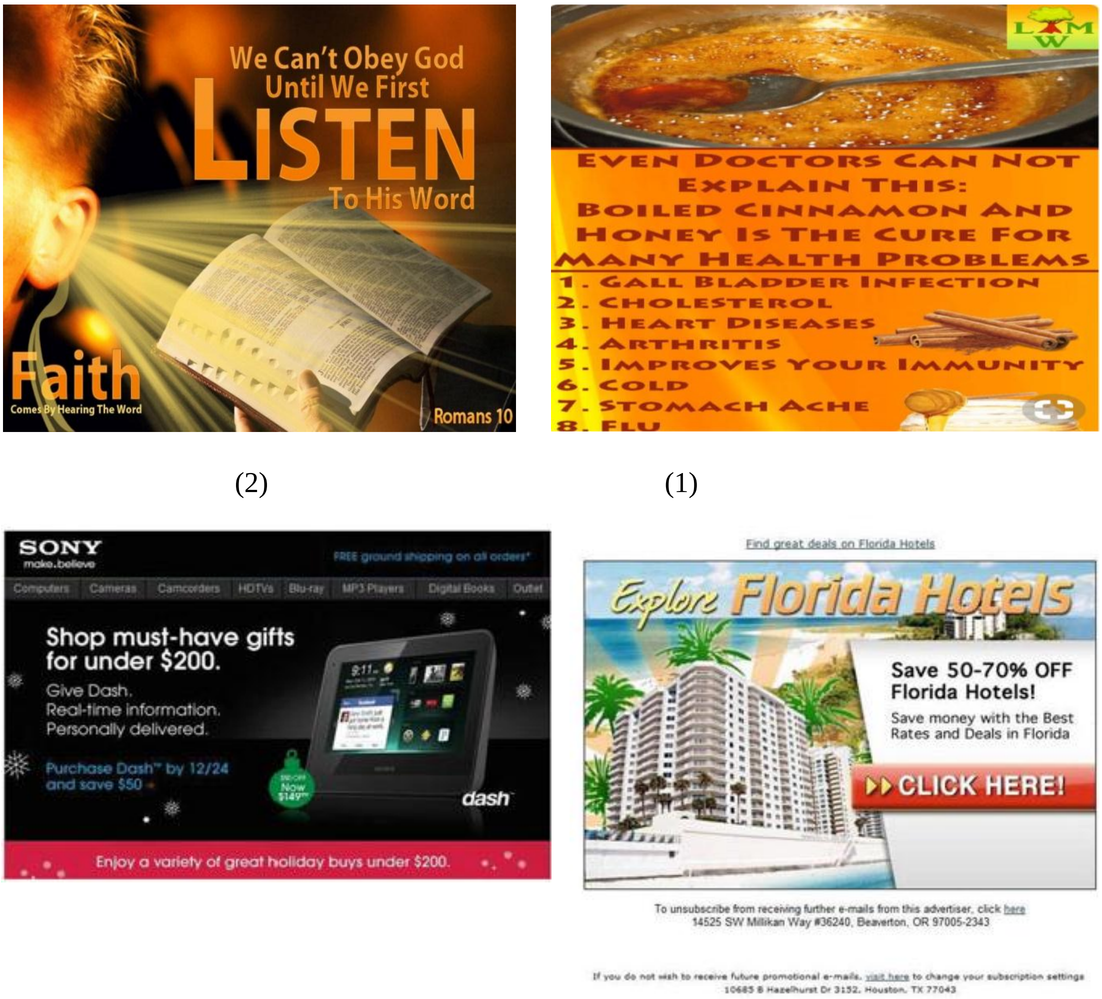
Unsolicited image messages are displayed in several formats. At most, spam images are divided into several categories or branches, the first one placing the image in a format that displays the target of the spammer, URL address is put in the image, spammer makes the image with a dazzling attractive to persuade the user for visiting the URL address.

The second type is image spam content is of all substance which is utilized as a part of text-based spam email. One might say that the utilized image is a screenshot of the typical text-based spam email. All objectives are displayed in the image with all subtle elements that spammer needs to share for clients. For instance, if spammer needs to demonstrate advertisements in picture spam, it can contain item name, the portrayal of the item, maker name, address, phone number, and so forth.

The third type is the unwanted image can be embedded in a hyperlink format after clicking by the user on the Hyperlink can be moved to the link that represents the target of the spammer was a virus or an advertisement for a particular product or otherwise[25,10].

On the other hand, most of the spamming targets in image spam e-mail are advertisements and the various ads are described below [2, 3, 17]:-

1- Adult: means the services or products of persons who are older than the age of 18, Email attacks of this kind are not positives or improper.
2- Inancial: Advertisements of unsolicited e-mail may be advertisements for financial offers such as investment offers or offers for the purpose of theft, fraud, etc…
3- Products: Advertising for specific products and services that are clothing or make-up and so on.
4- Internet: includes advertising on the Internet or computer services, for example, online shopping, and web design or may include attacks to email.
5- Leisure: Advertising, for specific competitions or holiday casinos as well as attacks may include e-mail.
6- Health: includes the announcement of services related to health, such as advertising for nutritional supplements, advertising for the treatment of herbal medicines or other things related to health.
7- Political: E-mail attacks that include a request for donations for a political party or party or a propaganda campaign for a political person that was for the purpose of elections or donations to a particular party.
8- Education: Email attacks are related to education such as distance learning through e-learning.
9- Spiritual: Email attacks include information such as spiritual, religious issues, psychology, and astrology.


We Can't Obey God
Until We First
LISTEN
To His Word
Faith
Comes By Hearing The Word
Romans 10
LMW
Even Doctors Can Not Explain This:
Boiled Cinnamon And Honey Is The Cure For Many Health Problems
1. Gall Bladder Infection
2. Cholesterol
3. Heart Diseases
4. Arthritis
5. Improves Your Immunity
6. Cold
7. Stomach Ache
8. Flu
SONY
make.believe
FREE ground shipping on all orders*
Computers
Cameras
Camcorders
HDTVs
Blu-ray
MP3 Players
Digital Books
Outlet
Shop must-have gifts for under $200.
Give Dash.
Real-time information.
Personally delivered.
Purchase Dash™ by 12/24 and save $50
$50 OFF Now $149
dash
Enjoy a variety of great holiday buys under $200.
Find great deals on Florida Hotels
Explore Florida Hotels
Save 50-70% OFF Florida Hotels!
Save money with the Best Rates and Deals in Florida
CLICK HERE!
To unsubscribe from receiving further e-mails from this advertiser, click here
14525 SW Millikan Way #36240, Beaverton, OR 97005-2343
If you do not wish to receive future promotional e-mails, visit here to change your subscription settings
10685 B Hazelhurst Dr 3152, Houston, TX 77043

(2) (1)

(3) **(4)**

Figure (2) Examples of spam images such as (1) Spiritual (2) health (3) financial (4) leisure.

3-2 Spam image tricks

There are many attempts that spammer uses to make filtering - based methods unable to detect their tricks. Every time the spammer tries to use new tricks and researchers are trying to use a new way to reduce them for the purpose of restoring efficiency and continue this story, the tricks used for this time classified into three sections of ordinary tricks, bold trick and proposed new tricks: [2].

3.2.1 Ordinary trick

The oldest trick in this area is obfuscation, for example, the rotation of the text and the blurring of the lines of text.

Text only is another kind of tricks in this area some trick are made up of text and this type is very expensive.

The trick of multiple sections is to split this image into multiple sections this trick makes it difficult to rearrange or compile the text.

The trick of the handwritten images using the handwriting It is an attempt to prevent the OCR tools from recognizing it.

The wild background is used for blurry and mysterious backgrounds to prevent OCR tool from characterizing it.

The trick of Geometric variance is that every time the font color is changed and the background is changed [2,18,17].

3.2.2 Focus on bold trick

The bold tricks that have paid others less attention to them:

The trick of the cartoon is a strange trick and uses the cartoons to prevent filters spam images of their discovery.

The re-coloring trick uses the same image with the same objects but different colors.

Obfuscate and curtail URLs trick eliminate URLs and cut URLs from some of the URL filters, some URLs are identified for filters related to their history of action. Curtail URLs service let the spammer evade such filters by a different URL readdressed users from spam sites.

Forge header information trick spammers only falsify information in the head part and thus obtain unreliable information in the head.

Template-Driven Generate spam mail in predefined syntax.

Emerge new source at present there are companies specializing in spamming and it is possible to identify the sender easily and prevent it.

People’s computers the goal of the spammer in this type of trick is to infect the computer with people. Harvest email addresses there is a ration of spam bots over the WWW to gather email addresses and find new destinations to bout[2,17] .

3.2.3 New and effective trick

You can unsubscribe! , Spammers do anything to keep in touch with the user. Natural image to deceive filters, spam filters use natural images to achieve their goal. Patchy font trick In this trick, each character is made out of numerous stuff and text image spammers use the alphabets in this trick to draw their own trick[2].

4. Spam Image Filtering Techniques

Always, methods to recognize spam image are separated into three classifications: [6]

4.1 Header based techniques

At present email customers frequently conceal the headers from the client. This is why most people have not seen their e-mail header. However, the header is delivered with the contents of the e-mail. Most e-mail messages provide the option of "Enable e-mail address display or disable. The idea of this method is to analyze only the waste part of the email course.The header of e-mail consists of many fields that provide a useful information margin. A portion of the header fields are utilized as a part of training and they are depicted below [2, 6]:

Header fields	Description
Sender IP:	Each PC associated with the Internet is allocated a one of a kind numerical code known as an IP address (Internet Protocol Address). IP tends to distribution is two writes static and dynamic. Sender IP address demonstrates address of the source PC from which the email was sent. It can tell where the email was sent from and perhaps who sent it.
Sender email address:	Sender email address is a sure string indicates the sender email account.
Precedence:	Email priority is a mark in email header part that shows email kind; messages can be conveyed to the rundown, mass mail, garbage mail or custom. Numerous hostile to spam frameworks in both email servers and email customers of end client focus on priority mark to oversee email messages wellness.
List-Help:	List-Help shows an email URL and additionally web URL to get help for a rundown. It might likewise incorporate a few "remarks". The List-Help additionally considered as the most critical field in email header part.
Errors-To:	Errors-To shows the address which blunder notices are to be sent. In other hand, it's demand to get conveyance notices. In the event that mistakes happen anyplace amid preparing, this header field will cause blunder messages to go to the recorded locations. Blunders To handle will leave in a future discharge
Sender:	Sender determines the letterbox of the specialist in charge of transmission of the email message.
In-Reply-To:	In-Reply-To indicates message identifier of the first message to which the present message is an answer to.
X-Mailer:	X-Mailer gives data about the customer programming of the originator in other hand portray the product used to making and sending the message. For example, if you send the email utilizing Outlook, the X-Mailer header field says Outlook and it's adaptation.
X-Priority:	X-Priority determines the messages need. The field is given a numerical estimation of 1 through 5:Values: 1 (Highest), 2 (High), 3 (Normal), 4 (Low), 5 (Lowest). 3 (Normal) is a default if the field is discarded.
X-Mime OLE:	X-Mime OLE included by Microsoft Outlook and perhaps other Microsoft programming.
Delivered-To:	Delivered-To determines a beneficiary address
Content-Transfer-Encoding:	Coding technique utilized as a part of a MIME message body part.
Message-ID:	Message-identity field is a completely unique identifier for every specific version of the e-mail message. it's far like the tracing id of an express postal mail. Message-id is composed by the call of the server that assigned the id and a unique string.
Date:	Specifies the date and time at which the author of the message indicated that the message became complete and prepared to go into the mail delivery machine. The time while the message was written (or submitted)
From:	Designates the author of the message, the mailbox of the person or system responsible for the indictment of the message.
To:	Contains the address of the primary recipient of the message
Received:	It includes information about receipt of the current message by a mail transfer agent on [2].

4.2 Content-based techniques

Content-based strategies using feature extraction and image content analysis. This kind of filters to analyze and study about the image substance and features, for example, shading, edge, surface, and so on are separated from the image that communicates the general attributes of image spam [6].We will follow it in detail in section 5.

4.3 OCR based techniques

These techniques are using Optical character recognition (OCR) tool to extract the text embedding into the image [6]. OCR is the mechanical or electronic interpretation of checked images of manually written, typewritten or printed content into machine-encoded content. It is generally utilized to change over books and reports into electronic documents, to modernize a record-keeping framework in an office, or to distribute the content on a site. OCR able find the text and edit it, look for a word or a certain phrase, store more tightly, show or print a duplicate free of scanning artifacts, and apply methods, for example, machine interpretation, text to speech and text mining to it [12,17]. Subsequent to separating text to investigate it pays to discover words or sentences that are related to the spam images.After separating the text from the image using OCR at this point it is verified that the text is ham or spam [6]. OCR is a field of research in pattern recognition, artificial intelligence, and computer vision. OCR programming application is fit for perusing highly contrasting pixels on any image and can recognize the exact alpha character or numeric number. In this method, the most recent OCR tool is very helpful when we require coding our lawful papers. At this point when the individuals check any archive, it's put in the form of an image, which can't be altered. While OCR has made it conceivable to scan any printed archive and migrate it into word - handling programming, for example, such as MS word where we can easily edit it as per our need. OCR is an advanced technology that enables individuals to change diverse sorts of reports including examined paper archives, PDF documents or pictures taking by an advanced camera into editable and accessible information too. Utilizing OCR the framework can detecting the static state of the character. It is important to comprehend that OCR method is a fundamental way additionally utilized in advanced scanning applications. Figure 3 indicates how when Figure 2 is divided into content and image utilizing OCR [12]. Mostly this strategy was effective, yet as of late most producers utilize diverse muddling methods that Prevent vision the spam image that’s why it caused the need for spam filters, inefficiency. Initial studies have shown that OCR is the best technique to filter spam images in the email, however but several reasons led to their revision. In the first place, arithmetic is very expensive when detection image spam. Although OCR has been able to overcome some tricks successfully, success in discovering some tricks is very difficult. "Because the spammer was able to overcome the optical characterization using some tricks, optical characterization cannot characterize the characters, which led to the need to update the definition The optical characters to be able to distinguish the tricks of the splash update this lead to an increased cost. In 2005 and previously, Although the spammer did not use the jamming techniques for unwanted images during the years before, the applications of optical character recognition using the technique of detecting unwanted images muddled although not used by the spammer [6]. The OCR based calculation experiences the challenges. The location of the text and recognize it is not exactly [19].

- They can't recognize images covered by CAPTCHA or like methods [19,16].

- They are computationally costly and along these lines can't work on heavily loaded email servers[16,6,19].
- OCR mistakes are one of the downsides of this sort of filters, particularly when spammers prevent the substance of the image by including noise, dots, changing the foundation hues and turning images, which influences the proficiency of OCR text extraction, this reality has prompted different procedures such as used Other techniques[10].
- Spammer one is confusing text to prevent OCR tools from detecting it [6].

5. Content-based Spam Image Filtering Methods

In this section, we address several important concepts are related to content-based spam image filtering methods, including:

5.1 Definition

This technique depends on the analysis and extraction of a variety of properties of the image and these characteristics represent the characteristics of the image undesirable [10]. Analyze content and features such as color, edge, texture and other general properties of spam image [6], the body of the e-mail look for certain properties that are used by the sender of the spam. Email can be text or images or text and image together. Content-based filtering methods always depended on both types of content and are representative of the popular and primary ways to get rid of spam and always send spammers tend to use new techniques to escape detection, these filters can be passed by the complex hacker and reach the user's mail [2].

5.2 Types of Content-Based Techniques

Content-based advancements are classified into three fundamental branches [6, 10, 19, 17]:

5.2.1 Low-level image features

The classification of the image in this technique depends on a set of attributes (features) or characteristics extracted from the image. The classification procedure depends on these selected features [10]. Low-level may be called visual features to represent the image there are many visual features such as color features, formatting features and textual features [21] in their research they relied on a number of properties in the extraction of properties from the image. These characteristics represent the visual characteristics. These characteristics are contrast, entropy, energy, and homogenate techniques based on the use of low-level properties. This technique can be separated into two type: image classification techniques and near-duplicate detection techniques [17]:

- **Image classification techniques**

Many authors have proposed to detect spam images by using a discriminatory method. A collection of low-level features is extracted from the image and the classifier is trained to create vector features for a group of images to legitimate and ham images. The main property of the different techniques is the selected feature, often Based on some assumptions about the characteristics that recognize spam image from ham image that use from image analysis, the fact about image classifier it apply algorithm infer and recognize new types of spam image, depending on choice feature suitable for training [20]. There is second feature classifier ordinary, some work utilizes classifier type two class, other utilize one class (i.e., training on the only spam image)while we discovered only one work in which a multi-class approach is proposed in[17]. Propose approach which utilizes basic edge-based features to represent too significant shape properties of images and accepted support vector machine to carry out image classification based on the features extracted [4].

- **Near-duplicate detection techniques**

Spam images are may be created from a shared template and randomized to sidestep the filters that based on the signature. In addition , the spam images sent in groups to several clients. Along these lines, images created from a similar format are outwardly comparable (near- duplicate)[10].

The near-duplicate detection technique is equivalent to content-based image retrieval techniques, which work to find the images that "look alike" a query image and can be recognized by comparing it with a database of known spam images [17]. The limits of the Near-duplicate detection solutions only similar images to the known images of the models as spam images, this achieves a low false positive rate. And by carefully maintaining the size of the template set when comparing the Near-duplicate detection solutions with image classification, the first less time processing in the testing phase.

Because Near-duplicate depended on similarities, In any case, near-duplicate detection solutions can just recognize spam images like the images in the template set. Without Every time updates to the template set, new kinds of spam images can pass the spam filtering system [20].

The features suggested in [17] depended on the presumption that spam images are frequently artificially generated, and are identified with spatial data (pixel coordinates), texture and color. Their circulation was approximated with Gaussian mixture models. The separation between images was estimated utilizing the (Jensen-Shannon divergence), given that it is symmetric, unlike commonly used measures of distance between probability distributions like images are processed in two stage. First, a very fast comparison is made between template images: file size, image width and height, bit depth, and aspect ratio and file properties of the input and If they are deemed similar, another examination is done carried out on gray-level or color image histogram, using measures such that Euclidean distance and histogram intersection. This technique is claimed to be fast, although it is not reported the processing time [17].

Ref no.	Arther(s)	Public year	proposed method	Technique used for image spam filtering/classification	Features used for classification			best Performance obtained	Filtering /Classification Results	Dataset used
					Image feature					
					Low level	High level	texture			
[10]	Basheer Al-Duwairi*, Ismail Khater† and Omar Al-Jarrah	2012	Image Texture Analysis-Based Image Spam Filtering (ITA-ISF)	C4.5 Decision Tree (DT),Support Vector Machine (SVM), Multilayer Perception (MP), Naïve Bays (NB), Bayesian Network (BN), and Random Forest(RF)	√	X	√	Accuracy	98.6%.	Dredze data set and Hunter ISH data set
[9]	Minal Kamble, Dr. L. G. Malik	2012	PCA has been proposed for feature selection	SVM Classifier	X	X	√	Accuracy	98%	Spam image taken from the website Giorgio Fumera's Group natural images are taken from flickr's website
[8]	Jianyi Wang and Kazuki Katagishi	2014	Image Content-Based "Email Spam Image" Filtering	Thresholding	X	X	√	0.01 as threshold	91.3%	999 pieces spam 453 pieces of non-spam images

[11]	M.Soranamages wari, Dr.C.Meena	2011	Gradient histogram based feature point extraction	feed forward back propagation neural network (BPNN) model.	√	X	√	Accuracy	93.7%	Giorgio Fumera's group
[16]	Meghali Das1 and Vijay Prasad2	2014	Analysis of an image spam in email based in content analysis	SVM classifier	√	X	√	Accuracy	81.40%	Dredze dataset

[21]	Anand Gupta1, Chhavi Singhal2 and Somya Aggarwal1	2010	a new approach based on Base64 encoding of image files and *n*-gram technique for feature extraction	SVM Classifier	√	X	√	accuracy, precision, and recall *(F1- measure)*	.٪98.4	Dredze dataset
[14]	T. Kumaresan .et.al	2006	On The Analysis Of Text Information Embedded Into Images	Two classifier 1- SVM 2- Thresholding SVM classifier before thresholding	X	X	X	accuracy	Not define	Personal corpus and the publicly available Spam Archive corpus
[24]	Congfu Xu1, Kevin Chiew2, Yafang Chen1, and Juxin Liu1	2011	Fusion of Text and Image Features	SVM Classifier	√	X	√	Accuracy Precision Recall non-spam recall	98.205%	Not define
[15]	Anand Gupta1, Chhavi Singhal2 and Somya Aggarwal1	2012	Using low level & metadata feature	threshold	√	√	X	Accuracy	93.3%	Dataset uses depends on the type of spam images, i.e. whether it contains only text, text and images or images only
[7]	T. Kumaresan .et.al	2015	using Support Vector Machine and Particle Swarm Optimization	using Support Vector Machine and Particle Swarm Optimization	√	√	X	Accuracy	90%	Personal Ham and personal spam subset

5.2.2 High level based technique.

High-level image features include the data in the image header, for example[3, 21]:

- **Image dimensions:** This feature includes width and height of an image which is included in the image file header.
- **The format of the Image file**: This feature included different image file formats such as GIF, JEPG, PNG, and BMP.
- **Image file size:** this feature compute the size of a file.
- **Aspect ratio**: defined as the ratio between the width (w) to the height (h) of an image.
- **Compression:** defined as the ratio between the area of an image to file size ofimage.

5.2.3 Texture technique

This type work depends on the textual features . The textual featues include the foolowing [21].

- Text Length: the characters number in the entire text.
- Words Number: the words number in the text.
- Ambiguity: n1=n2, where n1 is the number of special characters and n2 the number of normal characters.
- Correctness: Nn=Ns, where Nn is the number of words that include normal character, and Ns the number of words that include the special character.
- Special Length: the maximum length of a continuous sequence of special character.
- Special Distance: represented by two special characters the maximum distance between them.

TABLE (2): COMPARISON OF DIFFERENT SPAM TECHNIQUES

Approach	Advantage	Disadvantage
High-level method	They are quick when compared the features of the low-level image, in light of the fact that the file properties of the file can be checked in a single pass.	The features of the high-level image, not highly accurate for image classification as spam or not spam
Low-level method	features of the low-level acquaint better understanding of the order and give high classification accuracy quick and it is successful in light of the fact that it doesn't depend on extracting content and investigating email content [19]	Low level has the disadvantage of low recall proportion and furthermore, time escalated [19].
Texture method	Image spam discovery methods in light of the textual features give great precision in identification image spam	take time in extracting text content this represented the disadvantage for the texture method

6. Performance Evaluation Metrics

In order to calculate and evaluate performance, there are standards metric are accuracy, precision, recall and F-measure which can be defined as follows [1, 6]:

Measure	Defined as	What it means
Accuracy	$\frac{\mathrm{TP}\ +\mathrm{TN}}{\mathrm{FN}+\mathrm{FP}+\mathrm{TN}+\mathrm{TP}}$	Percentage of correct predictions [22]
Precision	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$	It is the correct level of expectations [10]
Recall	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$	it is the completeness of the retrieval process (possible recovery of positive examples)[10]
F-measure	$\frac{2 * \mathrm{Precision\ x\ Recall}}{\mathrm{Precision}\ +\ \mathrm{Recall}}$	It is a weighted average of precision and recall [10]

Where FP, FN, TP, TN are defined as follows[10,22,6].

- False Positive (FP): It means the number of ham messages (image based e_mail) that are classified incorrectly.
- False Negative (FN): It means the number of spam messages (image based e_mail) that are classified incorrectly...
- True Positive (TP): It means the correct classification of spam messages (image based e_mail).
- True Negative (TN): It means the correct classification of ham messages (image based e_mail).

7. Conclusions

Several important conclusions have been obtained from this study as follows:

1- There is a direct correlation between the growing use of the Internet and the rise in spam image. Also, the main problem caused by this type of unwanted messages is a loss of productivity and IT resource depletion.

2- There is a clear evolution in procedures of spammers to adapt to all challenges, so there must be new techniques to filter unwanted images that are adaptable to future tricks of spammers.

3- Most spam image filtering techniques include classifiers based on machine learning or pattern matching techniques. These techniques can be either supervised or semisupervised.

4- Usually, the filtering techniques that based on low-level features images give better results than high-level image features. However, it is preferable to choose a suitable combination set of features and appropriate filtering algorithms.

5- This study will form the basis for our future work in filtering spam images using one of machine learning technique. It will also help define the desired line in the implementation of a new framework for effective protection against these unwanted images. Also may present useful notes to develop new filtering techniques to filter content in multimedia data, especially in conflict environments.

6- Through this study, we did not find that there is a filtering approach that gives 100% results.

References

[1] Chopra, Nisha D., and K. P. Gaikwad (2015). "Image and text spam mail filtering." *Int. J. Comput. Technol. Electron. Eng (IJCTEE)* 5, no. 3.

[2] Attar A, Rad RM, Atani RE (2013). A survey of image spamming and filtering techniques. Artificial Intelligence Review. 40(1):71-105.

[3] Khanum MA, Ketari LM (2012). Trends in Combating Image Spam E-mails. arXiv preprint arXiv:1212.1763.

[4] Ravikumar K, Gandhimathi P. A (2014) Review on Different Spam Detection Approaches.

[5] Chang R. (2017) Application of multi-layer algorithm on image spam filtering. Modern Physics Letters B. (19-21):1740030.

[6] sadat Hosseini M, Rahmati M. (2015)A Method for Image Spam Detection Using Texture Features.

[7] Kumaresan, T., Sanjushree, S., Suhasini, K. and Palanisamy, C., (2015). Image spam filtering using support vector machine and particle swarm optimization. *Int. J. Comput. Appl*, *1*, pp.17-21.

[8] Wang, Jianyi, and Kazuki Katagishi (2014) "Image Content-Based" Email Spam Image" Filtering." *Journal of Advances in Computer Networks* 2, no. 2: 110-114.

[9] Minal Kamble and Dr. L. G. Malik(2012) " Detecting Image Spam Using Principal Component Analysis & SVM Classifier" , *International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555, Vol. 2, No.6.*

[10] Al-Duwairi, Basheer, Ismail Khater, and Omar Al-Jarrah (2012). "Detecting image spam using image texture features." *International Journal for Information Security Research (IJISR)* 2, no. 3/4, pp. 344-35

[11] Soranamageswari, M., and C. Meena (2011). "A novel approach towards image spam classification." *International Journal of Computer Theory and Engineering* 3, no. 1.

[12] Sathiya, V., M. Divakar, and T. S. Sumi (2011). "Partial Image Spam E-Mail Detection Using OCR." *International Journal of Engineering Trends and Technology* 1, no. 1, pp.55-59.

[13] Klangpraphant, Pattaraporn, and Pattarasinee Bhattarakosol (2010). "PIMSI: A partial image SPAM inspector." In *Future Information Technology (FutureTech), 2010 5th International Conference on*, pp. 1-6. IEEE.

[14] Fumera, Giorgio, Ignazio Pillai, and Fabio Roli (2006). "Spam filtering based on the analysis of text information embedded into images." *Journal of Machine Learning Research* 7, pp. 2699-2720.

[15] Gupta, Anand, Chhavi Singhal, and Somya Aggarwal (2012). "Identification of image spam by using low level & metadata features." *International Journal of Network Security & ITS Applications* 4, no. 2.

[16] Das, Meghali, and Vijay Prasad (2014). "Analysis of an Image Spam in Email Based on Content Analysis." *International Journal on Natural Language Computing (IJNLC)* 3, no. 3, pp. 129-140.

[17] Biggio, Battista, Giorgio Fumera, Ignazio Pillai, and Fabio Roli (2011). "A survey and experimental evaluation of image spam filtering techniques." *Pattern Recognition Letters* 32, no. 10, pp.1436-1446.

[18] Biggio, Battista, Giorgio Fumera, Ignazio Pillai, and Fabio Roli (2007). "Image spam filtering using visual information." In *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, pp. 105-110. IEEE.

[19] Ketari, Lamia Mohammed, Munesh Chandra, and Mohammadi Akheela Khanum (2012). "A study of image spam filtering techniques." In *Computational Intelligence and Communication Networks (CICN), 2012 Fourth International Conference on*, pp. 245-250. IEEE.

[20] Liu, Tzong-Jye, Cheng-Nan Wu, Chia-Lin Lee, and Ching-Wen Chen (2014). "A self-adaptable image spam filtering system." *Journal of the Chinese Institute of Engineers* 37, no. 4, pp. 517-528.

[21] Xu, Congfu, Yafang Chen, and Kevin Chiew. (2010) "An approach to image spam filtering based on base64 encoding and N-Gram feature extraction." In *Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on*, vol. 1, pp. 171-177. IEEE.

[22] Kumar, N. S., D. P. Rana, and R. G. Mehta (2012). "Detecting e-mail spam using spam word associations." *International Journal of Emerging Technology and Advanced Engineering* 2, no. 4, pp. 222-226.

[23] Foqaha, Mohammed Awad1and Monir.(2016) "EMAIL SPAM CLASSIFICATION USING HYBRID APPROACH OF RBF NEURAL NETWORK AND PARTICLE SWARM OPTIMIZATION.".

[24] Xu, C., Chiew, K., Chen, Y., & Liu, J. (2011). Fusion of text and image features: A new approach to image spam filtering. In *Practical Applications of Intelligent Systems* (pp. 129-140). Springer, Berlin, Heidelberg.

[25]. D.Sasikala, R.Roshiniya, Sarishnaratnakaran, Tapati Deb," Texture Analysis Of Plaque In Carotid Artery", *International Journal Of Innovations In Scientificand Engineering Research(Ijiser)*, Vol 4 Issue 2 Feb 2017, Pp.66-70.