ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)
Vol. 2 Issue Special 1 Jan-March 2014
International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014)

Contour Point Face Recognized Gmail Accessor with Pattern Based Spam Detector

[I]E.Komathi, [II]A.Mani
[I,II]S.A Engineering College, Chennai, India

Abstract
Pattern recognition is used in security applications like biometric authentication, intrusion detection, and spam detection for identifying legitimate and malicious pattern class. Pattern classification shows that a there is a trade-off between classifier accuracy and security to achieve matching and reduce the storage for each e-mail content text which can be exploited for several purposes. In nature e-mail cannot fully catch the spams by comparing the keywords and URL located in the repository using top-down parsing search algorithm. If the tags are matched, the e-mail will be spam domain and URL will be stored for future rejection of spam e-mail. In the proposed system image sets are followed by the metric and the models are compared. A set of values which are available in a linear space are selected and converted to individual image set by a convex geometric region. The feature points set dissimilarity is measured by geometric distance between the convex models.

Keyword
Fractal code, Face recognition, Pattern recognition, Contour point, Spam Filter

I. Introduction

Electronic mail, most commonly referred to as email or e- mail. It is a method of exchanging digital messages from an sender to one or more receiver. Recent e-mail operates across the Internet or computer networks. In early email systems required that the sender and the recipient both be online at the same time, in common with instant messaging. Nowadays email systems are based on a store-forward model. Email servers accept, forward, deliver, and store up messages. Neither the users nor their computers are required to be online concurrently; they need to connect only temporarily, typically to an email server, for as long as it takes to send or receive messages.

Gmail is an open email service provided by Google. Users may access Gmail as protected webmail, as well as via POP3 or IMAP4 protocols. Gmail originally started as an invitation- only beta and it became available to the general public though still in beta status at that time. The service was upgraded from the beta status, along with the rest of the Google Apps suite.

An early storage capacity offer of 1 GB per user, Gmail drastically increased the webmail standard for free storage from the 2 to 4 MB its competitors such as Hotmail existing at that time. Individual Gmail messages, including files, images, may be up to 25 MB, which is larger than many new mail services support. Gmail has a search-oriented interface and a "chat view" same as Internet forum. Gmail is well-known by web developers for its pioneering use of Ajax. Gmail can run with Google GFE/2.0 on Linux. It is the most usually used web- based email provider with over 425 million active users worldwide.

A facial recognition system is a computer application for without human intervention identifying or verifying a person from a digital image or video frame from a video source. One way is by comparing selected facial features from the image and a facial database. It can be used in security systems and can be compared to many other biometrics such as fingerprint or eye iris recognition systems.

The human face is a very prosperous source of information that can be used to discover persons. This ability of recognition allows us to differentiate persons despite the facial similarity between them. Today, many researchers try to benefit from computer applications, which become usually used in face automatic recognition. After more than 30 years of research, we can categorize the dissimilar existing face recognition systems into three main approaches.

- Local approaches: which are based on the fact that the face contains parts that have a high discriminating power such as eyes, nose, mouth etc. To identify a person, we use either the blocks containing these regions or the geometric relationships between them.
- Global approaches: which treat the face as a whole object and use all the information included in it. Various methods have been proposed that include the use of Eigen faces, discrete cosine transform, and Gabor Wavelets.
- Hybrid approaches: The principle of these approaches is to imitate the human visual system, which uses the both local and global features to identify persons.

The amalgamation of these two methods has only one interest: to take advantage of the combined benefits of both approaches. In spite of the number of researchers and the proposed methods, a number of factors can significantly affect face recognition performances, such as the facade, the presence/absence of structural components, facial expressions, occlusion and clarification, variants. In order to encounter these factors and ensure a high recognition rate and a fast recognition time, we have used the fractal representation which exploits the inter-image resemblance. There are few that are related to this topic [face recognition using Iterated Function System (IFS) theory] [10].

II. Preliminaries

A. Pattern Cognition

Pattern recognition is a classification based on machine learning algorithms used in many security application network intrusion detection, bio-metric authentication system, spam filtering processing discriminate between "legitimate" and "malicious" of pattern matching (e.g., spam e-mails). Pattern recognition provides the pattern matching of URL and keywords. Pattern matching algorithm is regular expression. In which patterns are some sort of text data. Contrary pattern recognition, pattern matching algorithm is not a type of machine learning algorithm, also provides quality output to the sort by pattern recognition. This often gives rise to an arms race between the adversary and the classifier design.

A well-known common example of attacks pattern classifiers are: biometric traits can be to perform automate personal authentication

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

(finger print, face recognition, voice, hand geometry etc.) [1-2]. Each biometric based on pattern recognition systems to identify an unique legitimate technology. To modify-ing network packets belonging to intrusive network traffics to evade intrusion detection systems [3]; to manipulate the content of spam emails to get them past spam filters (e.g., by misspelling common spam words to avoid their detection) [4]–[6]. Adversarial scenarios can also occur in intelligent data analysis [7] and information retrieval [8]; e.g., a malicious webmaster may manipulate search engine rankings to artificially promote her web site.

B. Fractal Face Recognition

Humans can perform face recognition with relative ease under many varying conditions. It is desirable to have an automatic system that can perform that same task with similar speed and accuracy. As a result, automatic face recognition algorithms have been an active area of research for more than 20 years, although the performance has not reached that of humans.

Many algorithms have been proposed for face recognition, and one of the most popular and successful is based on the Eigen face method [8]. This method uses principal components analysis (PCA) to find a reduced basis for the sample space spanned by a set of training faces.

1. Spam Filtering

Assume that a classifier has to distinguish between legitimate and unsolicited mail emails on the basis of their textual content, and that the bag-of-words feature illustration has been chosen, with binary geographies representing the manifestation of a given set of words. This benevolent of classifier has been deliberated by several authors [6] and it is included in several real spam filters. There is no one exact algorithm for statistically defining whether or not a specified e-mail message is in fact a spam message. As discussed earlier, the most projecting approach to spam classification Involves the implementation of the Bayesian chain rule, also known as Bayesian spam filtering. The Bayesian algorithm is a high level algorithm for the creation and execution of a statistical spam e-mail classification system, based on clarifications from various implementations of the Bayesian classification system.

2. Biometric Authentication

Multimodal biometric systems for private identity recognition have expected abundant interest in the past few years. It has been publicized that combining statistics coming from dissimilar biometric behaviors can overcome the limits and the faintness inherent in every individual biometric, resulting in a higher accurateness. Moreover, it is frequently alleged that multimodal schemes also improve security against spoofing attacks, which contain of claiming a incorrect identity and stand up to at least one fake biometric trait to the system (e.g., a "gummy" fingerprint or a photograph of a user's face).

The purpose is that, to evade a multimodal system, one expects that the opposition should spoof all the matching biometric traits. In this application example, we show how the designer of a multimodal system can verify if this suggestion holds, before deploying the system, by pretending spoofing attacks against each of the matchers. To this end, we to some extent adventure the analysis in [1], [2].We consider a typical multimodal system, through up of a fingerprint and a face matcher, which works as follows.

The design level includes the acceptance of authorized users (clients): allusion templates of their biometric behaviors are put in storage into a database, laid-back with the analogous identities. During operation, each user provides the demanded biometric traits to the sensors, and entitlements the identity of a client. Then, each matcher compares the submitted trait with the template of the claimed identity, and provides a real-valued matching score: the higher the score, the higher the similarity. We denote the score of the fingerprint and the face matcher respectively as Xing and xface. Finally, the matching scores are combined through a proper fusion rule to decide whether the claimed identity is the user's identity (genuine user) or not (impostor).

C. Brute Force String Match Algorithm

String matching is something critical for database development and text processing software. Fortunately every up to date programming language and library is full of functions for string processing that help us in our day by day work. However is great to understand their principles. String match algorithms can be mainly divided into several categories.

Rabin-Karp Pseudo-Code pattern is M characters long hash_ p=hash value of pattern hash_t=hash value of first M letters in body of text
do
if (hash_p == hash_t)
brute force comparison of pattern and selected section of text
hash_t = hash value of next section of
text, one character over
while (end of text or
brute force comparison == true)

III. Related Work

A taxonomy of potential attacks against pattern classifiers was proposed in [11], [15], and subsequently extended in [14]. We will exploit it in our framework, as part of the definition of attack scenarios. The taxonomy is based on two main features: the kind of influence of attacks on the classifier, and the kind of security violation they cause.

The influence can be either causative, if it undermines the learning algorithm to cause subsequent misclassifications; or exploratory, if it exploits knowledge of the trained classifier to cause misclassifications, without affecting the learning algorithm. Thus, causative attacks may influence both training and testing data, or only training data, whereas exploratory attacks affect only testing data.

The security violation can be an integrity violation, if it allows the adversary to access the service or resource protected by the classifier; an availability violation, if it denies legitimate users access to it; or a privacy violation, if it allows the adversary to obtain confidential information from the classifier. Integrity violations result in misclassifying malicious samples as legitimate, while availability violations can also cause legitimate samples to be misclassified as malicious.

A third feature of the taxonomy is the specificity of an attack that ranges from targeted to indiscriminate, depending on whether the attack focuses on a single or few specific samples (e.g., a specific spam email misclassified as legitimate), or on a wider set of samples.

Limitations of classical performance are evaluation methods in adversarial classification. Classical performance evaluation methods, like k-fold cross validation and bootstrapping, aim

to estimate the performance that a classifier will exhibit during operation, by using data D collected during classifier design.

IV. Proposed System

In our proposed system, high effective authentication with the purpose of log on to the email service securely and efficient spamming are taken into consideration. Authentication in the form of fractal detection and recognition after contour detection of the face using the image of the user is introduced. Since fractal detection and recognition is a unique method to identify every human being, this concept is more effective in terms of authenticating into the service.

Pattern classifiers such as Keywords and URL's for data check, tag construction and keyword identity, automatic reading of mails is the concepts used in this system. Administrator of the email service uses the pattern classifiers and maintains a repository to filter out spam domains and keywords. Hence this perception spam's the frequent surplus mails from same domain with different mail ids.

Automatic reading of mails to examine the spammed keyword is an intriguing conception introduced in this system to overcome many flaws in case of spam filtering. Hence the authentication by means of fractal recognition and pattern classifier based spam filtering in the email service turn this proposed system more thriving and thus overcomes the drawbacks of existing system. Implementation is the stage of the project when the theoretical design is turned out into a working system.

The active contour method can be used to determine face features in a picture. This module is designed to check the input face using contour point facial recognition that is to be used as an authentication for the system. A programming interface has been designed to interact with the Gmail server. In this module, the connection establishment has been checked to proceed further. The keywords that are considered as spam on user's perspective are appended in the interface. A spam filter is a program that is used to detect unsolicited and unwanted email and prevent those messages from getting to a user's inbox.

Spam filter checks all incoming emails to your email accounts against mail filter rules. The face that has been recognized using the contour point facial recognition is used to authenticate into the Gmail via the programming interface. The user can utilize the smart camera's to recognize their face in order to authenticate into the system. The keywords that are considered as spam on user's perspective are appended in the interface. The spam keywords will be stored in repository. It can be used to reject the unwanted e-mails.

V. System Architecture

System architecture is the conceptual model that defines the structure and/or behavior of the system. It provides a way in which products can be procured, systems can be developed an architectural overview of the overall system. The system architecture for the proposed system is given in fig. 1.

The above architecture diagram Fig. 1 shows that the active contour method can be used to determine face features in a picture. To check the input face using contour point facial recognition that is to be used as an authentication for the system. The face that has been recognized using the contour point facial recognition is used to authenticate into the Gmail via the programming interface.

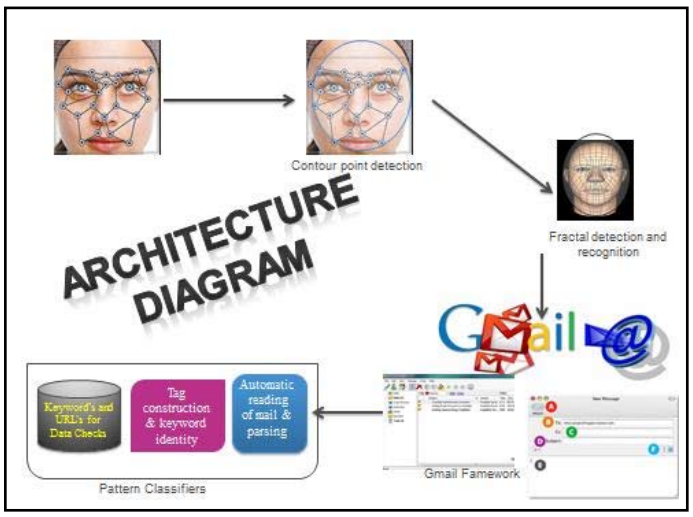

Fig. 1: Fractal Face Recognition

The user can utilize the smart camera's to recognize their face in order to authenticate into the system. The keywords that are considered as spam on user's perspective are appended in the interface. The spam keywords will be stored in repository. It can be used to reject the unwanted e-mails. The user can be able to view the Gmail inbox through the programming interface once after logging in into the system by facial recognition.

A spam filter is a program that is used to detect unsolicited and unwanted email and prevent those messages from getting to a user's inbox. Spam filter checks all incoming emails to your email accounts against mail filter rules.

VI. Conclusion

This paper proposed taxonomy of existing SPAM email countermeasures, and a brief description of the taxa we have proposed, and ascribe a number of existing SPAM filters to the various taxa. We have presented a wide range of the techniques that have been used or proposed for use to fight SPAM, and attempted to indicate which SPAM filters use which techniques.

These filters make assessments about the reputation of one or more of the participants (sender, recipient and intermediaries) in the email transaction. In this paper, we have introduced a new method for high effective authentication with the purpose of log on to the email service securely and efficient spamming. Pattern classifiers such as Keywords and URL's for data check, tag construction and keyword identity, automatic reading of mails is the concepts used in this system.

Spam filter checks all incoming emails to your email accounts against mail filter rules. The face that has been recognized using the contour point facial recognition is used to authenticate into the Gmail via the programming interface. The user can utilize the smart camera's to recognize their face in order to authenticate into the system.

References

[1] *R. N. Rodrigues, L. L. Ling, and V. Govindaraju, "Robustness of multimodal biometric fusion methods against spoof attacks," J.Vis. Lang. Comput., vol. 20, no. 3, pp. 169–179, 2009.*

[2] *P. Johnson, B. Tan, and S. Schuckers, "Multimodal fusion vulnerability to non-zero effort (spoof) imposters," in IEEE Int'l Workshop on Inf. Forensics and Security, 2010, pp.*

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

1–5.

[3] P. Fogla, M. Sharif, R. Perdisci, O. Kolesnikov, and W. Lee, "Polymorphic blending attacks," in Proc. 15th Conf. on USENIX Security Symp. CA, USA: USENIX Association, 2006.

[4] G. L. Wittel and S. F. Wu, "On attacking statistical spam filters," in 1st Conf. on Email and Anti-Spam, CA, USA, 2004.

[5] D. Lowd and C. Meek, "Good word attacks on statistical spam filters," in 2nd Conf. on Email and Anti-Spam, CA, USA, 2005.

[6] A. Kolcz and C. H. Teo, "Feature weighting for improved classifier robustness," in 6th Conf. on Email and Anti-Spam, CA, USA, 2009.

[7] D. B. Skillicorn, "Adversarial knowledge discovery," IEEE Intell.Syst., vol. 24, pp. 54–61, 2009.

[8] D. Fetterly, "Adversarial information retrieval: The manipulation of web content," ACM Computing Reviews, 2007.

[9] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification. Wiley-Interscience Publication, 2000.

[10] N. Dalvi, P. Domingos, Mausam, S. Sanghai, and D. Verma, "Adversarial classification," in 10th ACM SIGKDD Int'l Conf. on Knowl. Discovery and Data Mining, WA, USA, 2004, pp. 99–108.

[11] M. Barreno, B. Nelson, R. Sears, A. D. Joseph, and J. D. Tygar,"Can machine learning be secure?" in Proc. Symp. Inf., Computer and Commun. Sec. (ASIACCS). NY, USA: ACM, 2006, pp. 16–25.

[12] A. A. C´ardenas and J. S. Baras, "Evaluation of classifiers: Practical considerations for security applications," in AAAI Workshop on Evaluation Methods for Machine Learning, MA, USA, 2006.

[13] P. Laskov and R. Lippmann, "Machine learning in adversarial environments," Machine Learning, vol. 81, pp. 115–119, 2010.

[14] L. Huang, A. D. Joseph, B. Nelson, B. Rubinstein, and J. D. Tygar, "Adversarial machine learning," in 4th ACM Workshop on Artificial Intelligence and Security, IL, USA, 2011, pp. 43–57.

[15] M. Barreno, B. Nelson, A. Joseph, and J. Tygar, "The security of machine learning," Machine Learning, vol. 81, pp. 121–148, 2010.

[16] D. Lowd and C. Meek, "Adversarial learning," in Proc. 11th ACM SIGKDD Int'l Conf. on Knowl. Discovery and Data Mining, A. Press, Ed., IL, USA, 2005, pp. 641–647.

[17] P. Laskov and M. Kloft, "A framework for quantitative security analysis of machine learning," in Proc. 2nd ACM Workshop on Security and Artificial Intelligence. NY, USA: ACM, 2009, pp. 1–4.

[18] P. Laskov and R. Lippmann, Eds., NIPS Workshop on Machine Learning in Adversarial Environments for Computer Security, 2007. [Online]. Available: http://mls- nips07.first.fraunhofer.de/

[19] A. D. Joseph, P. Laskov, F. Roli, and D. Tygar, Eds., Dagstuhl Perspectives Workshop on Mach. Learning Methods for Computer Sec., 2012. [Online]. Available: http://www.dagstuhl.de/12371/

[20] A. M. Narasimhamurthy and L. I. Kuncheva, "A framework for generating data to simulate changing environments," in Artificial Intell. And Applications. IASTED/ACTA Press, 2007, pp. 415–420.

Ms. E. Komathi pursuing M.E CSE in S.A. Engineering College, Chennai. She has completed UG degree from Rajalakshmi Engineering College, Chennai. Her areas of interests are Data Structures, Network security and Mobile computing.

***Mr. A. Mani** is currently working as a Assistant Professor and Department of Computer Science Studies, S.A.Engineering College, Chennai, India. Him area of interests are DataBase Technology, Network Security and Data Structures.*