

# Image similarity for detecting image-based spam and phishing attacks

## Background

An image-based spam message is a spam message in which the only content is an image. A typical spam image has several variants in which the image is altered without changing the relevant visual content. Conventional signature algorithms are easily defeated by these alterations as developing unique signatures for each variant is impractical. A similar problem arises in the detection of phishing attacks, which often use visually similar scalings or saltings of corporate logos to evade anti-phishing software. These problems are assuming increasing importance due to the growth in image-based spam and phishing attacks.

## The Problem

Symantec is investigating the use of image similarity algorithms to identify the image variants described above. For useful background on techniques for comparing images, see [2] or [1]. The goal of the RIPS project will be to evaluate, compare, and optimize image similarity techniques for detecting image-based spam. Techniques will be evaluated for performance and accuracy on large “real world” data sets collected by Symantec’s worldwide probe network.

Candidate techniques include Tamura features (Section 2.2 of [2]), simultaneous Auto-regressive models (Section 2.2 of [2]), color coherency vectors, Correlograms, Wavelet-transform coefficients, R-trees (Section 3.2 of [2]), self-organizing maps (Section 3.2 of [2]), and Monte Carlo techniques.

## References

- [1] David Feng, W.C. Siu, and Hong J. Zhang, editors. *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*. Signals and Communications Technology. Springer, 2003.
- [2] Y. Rui, T. Huang, and S. Chang. Image retrieval: current techniques, promising directions and open issues, April 1999.