InFeRno - an Intelligent Framework for Recognizing Pornographic Web pages

S. Karavarsamis, N. Ntarmos and K. Blekas

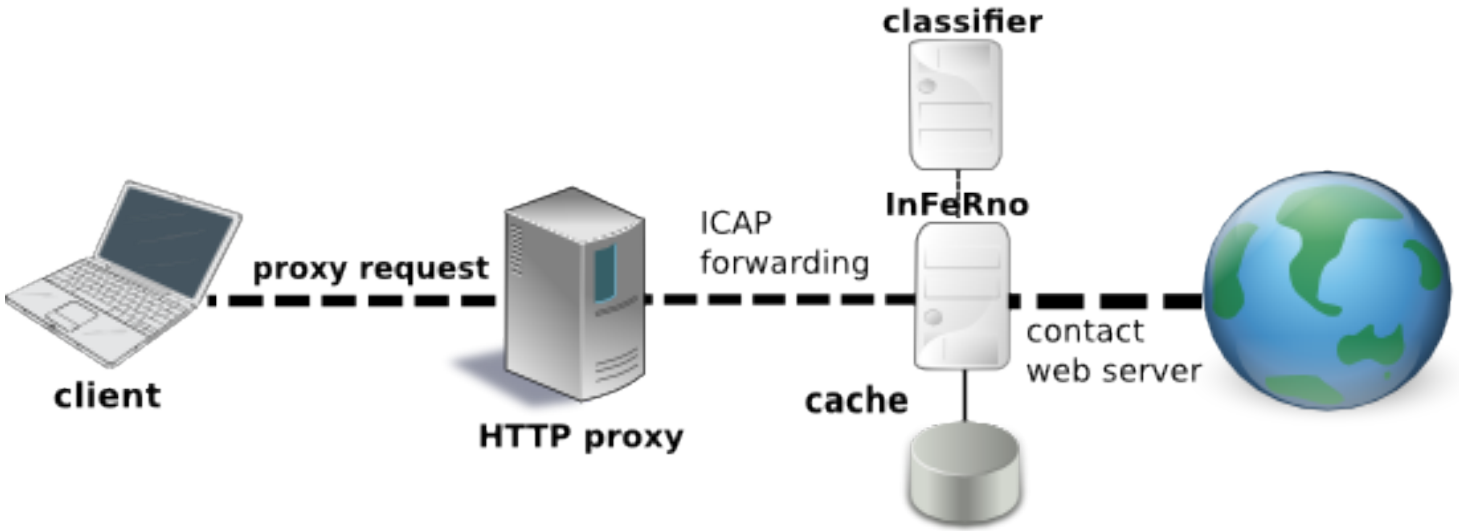
Department of Computer Science
University of Ioannina, Ioannina, Greece

E-mail: {cs061205, ntarmos, kblekas}@cs.uoi.gr

Problem preliminaries and intuition

- Motivation:
  1. Lots of "bad" pages out there
  2. Human-driven classification is highly impractical
  3. Need for an unobtrusive filtering method
- Main characteristics of our system:
  1. A minimal but powerful vector space
  2. An extra "bikini" class
  3. A highly accurate and fast classification scheme
  4. An implementation of the classifier as a standalone network service (ubiquity and ease of use)
  5. Integration of the classifier with off-the-shelf web proxy cache servers through an ICAP interface (can be transparently applied to whole networks)

InFeRno architecture

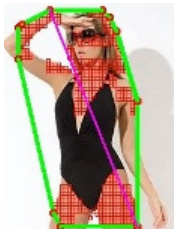


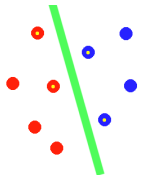
1. Implementation of the InFeRno core as an ICAP module (integrates well with most HTTP proxy servers)
2. Decoupled image classification and web page preprocessing (network I/O, image-score fusion)
3. Using a fast ISAM-based cache for fast I/O (classification lookups, updates, etc)
4. Flexible configuration (multiple classification & network parameters)

Classification System


original image


skin detection


contour extraction

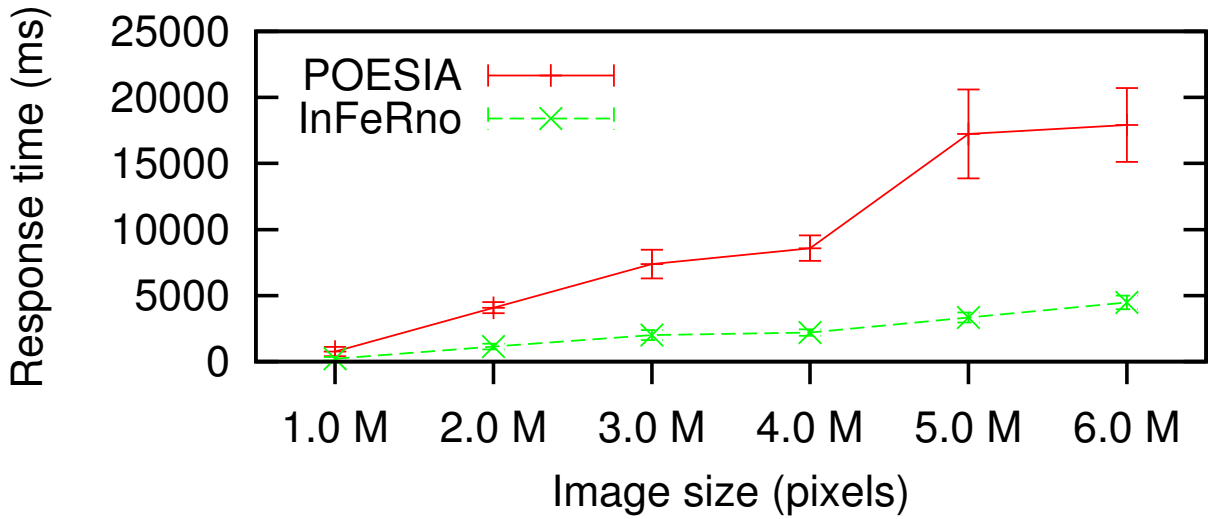

SVM

Three stages

1. Skin detection (rule-based)
2. Contour extraction (region splitting scheme)
3. Feature extraction and classification
   - Extracted 15 features: RGB color statistics, skin-to-nonskin ratio, contour orientation, Hu moments
   - SVM classifier with RBF kernel

Experimental results

- Training dataset: manually collected 680 bikini images, 660 porn images and 4260 benign images from the Web
- Comparison with the EU-funded *POESIA pornography elimination system*
- Results:



  - **4x speedup** improvement
  - High accuracy (comparable to POESIA for images and image-only pages)

Thank you for your attention! ☺