Transcoding characteristics of Web images

Surendar Chandra
University of Georgia
Athens, GA

Ashish Gehani, Carla Schlatter Ellis, Amin Vahdat
Duke University
Durham, NC

DUKE Systems & Architecture

Proliferation of Multimedia

One size does not fit all

Media rich world

Image Transcoding

Foggy road with fall foliage canopy

Transliterate few bytes

Thumbnail
2KB

Transcoding as a Technique to Customize Images

- Network proxies to customize for client
  - GloMop (Fox, Brewer et al. - Berkeley)
  - Caubweb (Open Systems Group)
  - Chandra et. al.
  - Intel Quickweb
  - Oracle Portal-To-Go
  - IBM Websphere Transcoding Publisher
- Web servers to manage bandwidth consumption
  - Chandra et. al.
  - HP WebQoS
- Transcoding digital camera to manage storage and battery power
  - Chandra et. al.

Research Goal

1. Analyze characteristics of images accessed on the Internet
2. Analyze how various image transcoding techniques perform for these images (we focus on transcodings that yield file size savings)

Applicability:

Choose the right image transcodings that can save image file size for a large class of Internet images

Research Summary

- 80% of GIF images smaller than 6 KB
- 10% of GIF TrueImages make up 45.1% data
- Productive transcoding (50% file size saving for 50% of images)
  - Spatial geometry reduction by 4 along each axis
  - GIF⇒JPEG

- 60% of JPEG images smaller than 6 KB
- 35.6% of JPEG TrueImages make up 78% data
- Productive transcoding
  - JPEG compression metric
  - Spatial geometry reduction

Outline

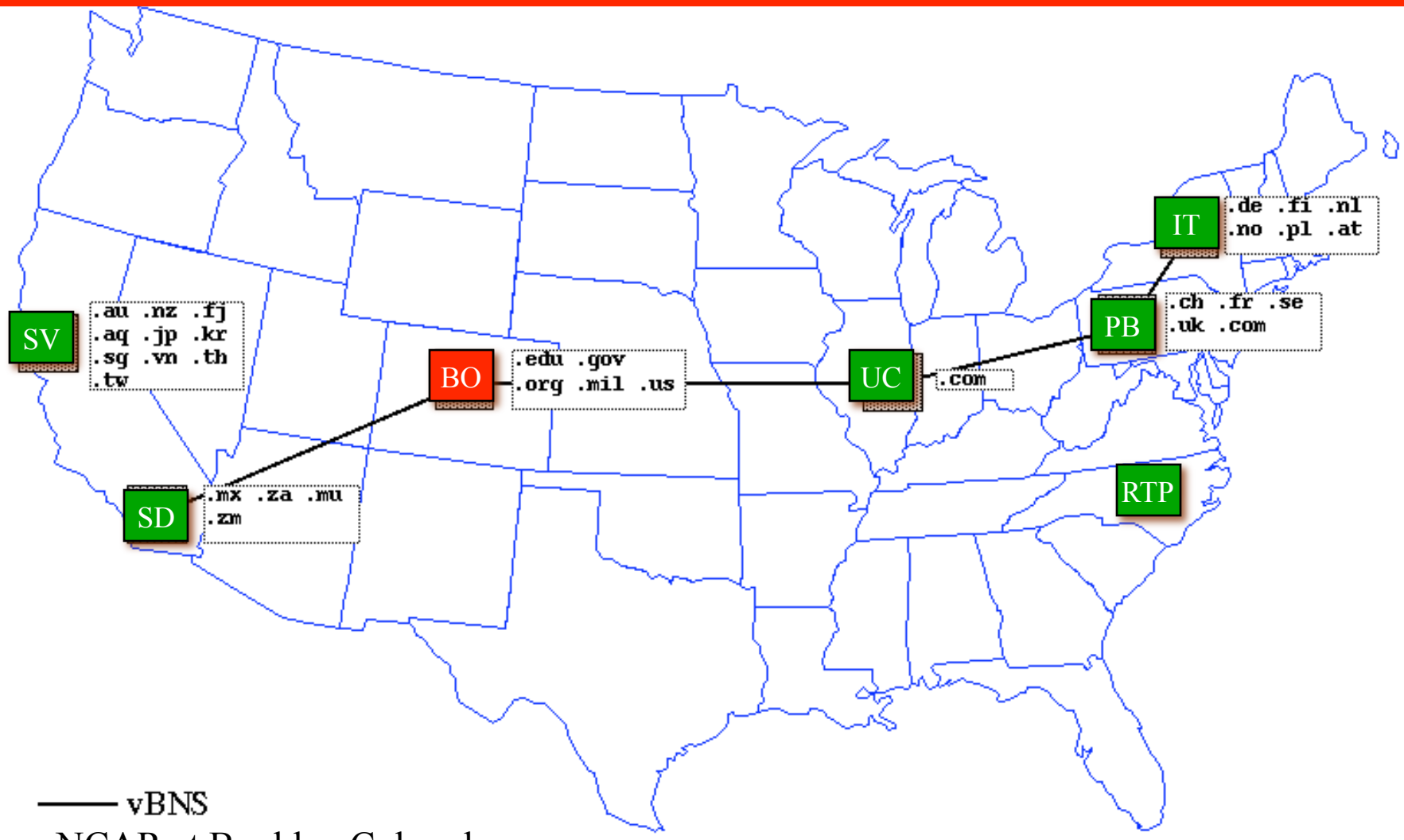
- Motivation
- **Experiments**
- **Results**
  - **Characterization of Images**
  - **Transcoding applicability**

Experiment Setup

- Ideal: Analyze actual images accessed on the Internet

- Practical: Analyze original images accessed via NLANR squid proxies

NLANR Squid Proxy Hierarchy



bo -- NCAR at Boulder, Colorado
pb -- PSC at Pittsburgh, Pennsylvania
sd -- SDSC at San Diego, California
sv -- NASA-Ames

uc -- NCSA at Urbana-Champaign, Illinois
pa -- Palo Alto, California
sj -- San Jose, California.
rtp -- Research Triangle Park, North Carolina

Image Collection

- Images in traces from NCAR proxy, Boulder, CO (Collected on Sep 27, 1999)

- 74.81% GIF, 24.41% JPEG, 0.78% other (incl. PNG)

- 97,900 GIF (402 MB) and 37,616 JPEG (440 MB)
- 48.61% (GIF bytes), 50.88% (JPEG bytes)

Experiment Methodology

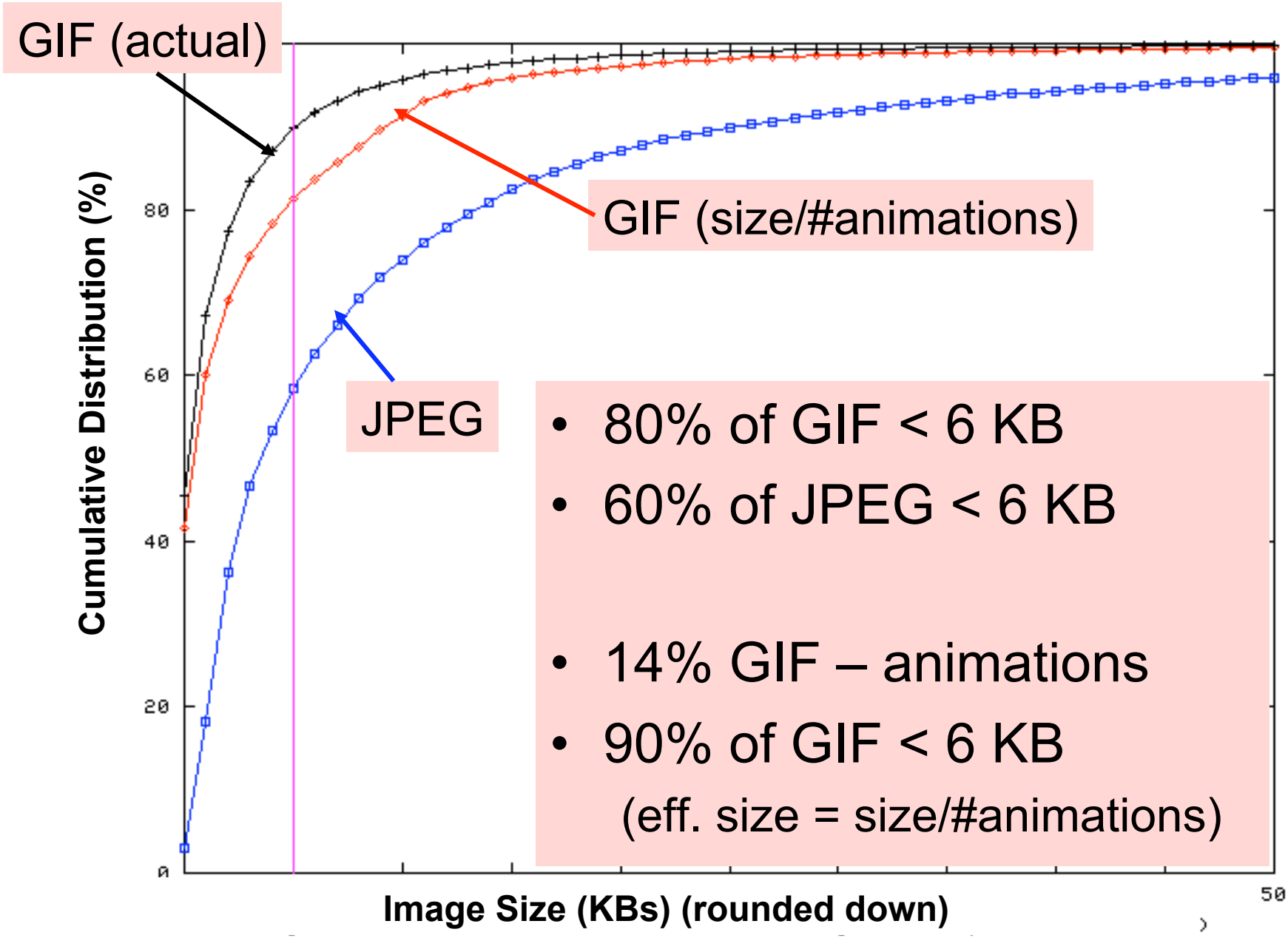
1. Static Image Characteristics
   - Image file size
   - Unique colors
   - Spatial geometry distribution
   - JPEG compression metric (Quality Factor)

2. Transcoding Image Characteristics
   - spatial geometry reduction (thumbnail)
   - color reduction*
   - format change (gif ⇒ jpeg, jpeg ⇒ gif*)
   - JPEG compression metric

Outline

- Motivation
- Experiments
- **Results**
  - **Characterization of Images**
  - **Transcoding applicability**

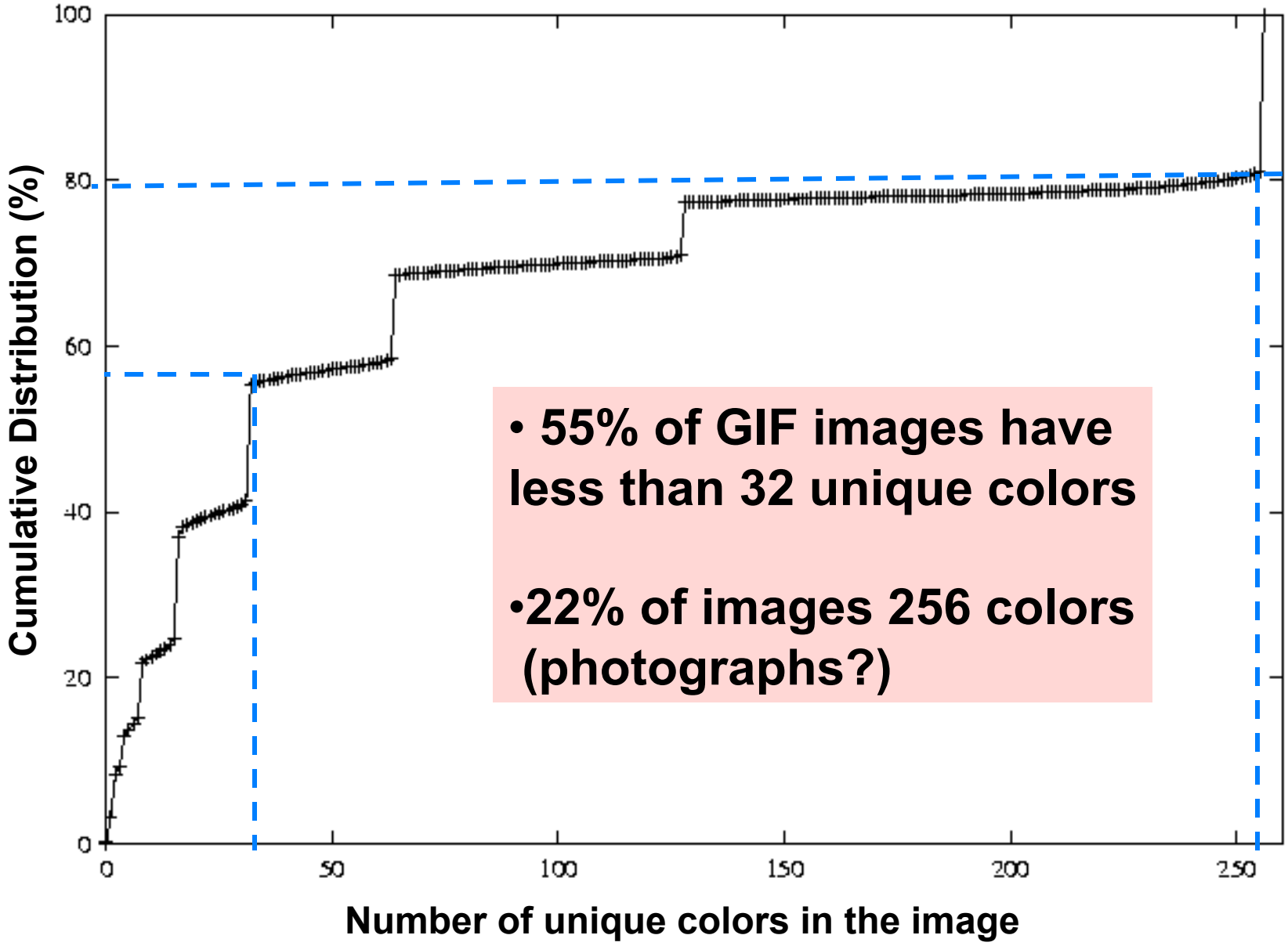
Image File Size



- 80% of GIF < 6 KB
- 60% of JPEG < 6 KB

- 14% GIF – animations
- 90% of GIF < 6 KB
  (eff. size = size/#animations)

Number of Unique Colors in GIF Images



- 55% of GIF images have less than 32 unique colors
- 22% of images 256 colors (photographs?)

Spatial Size Distribution (GIF)

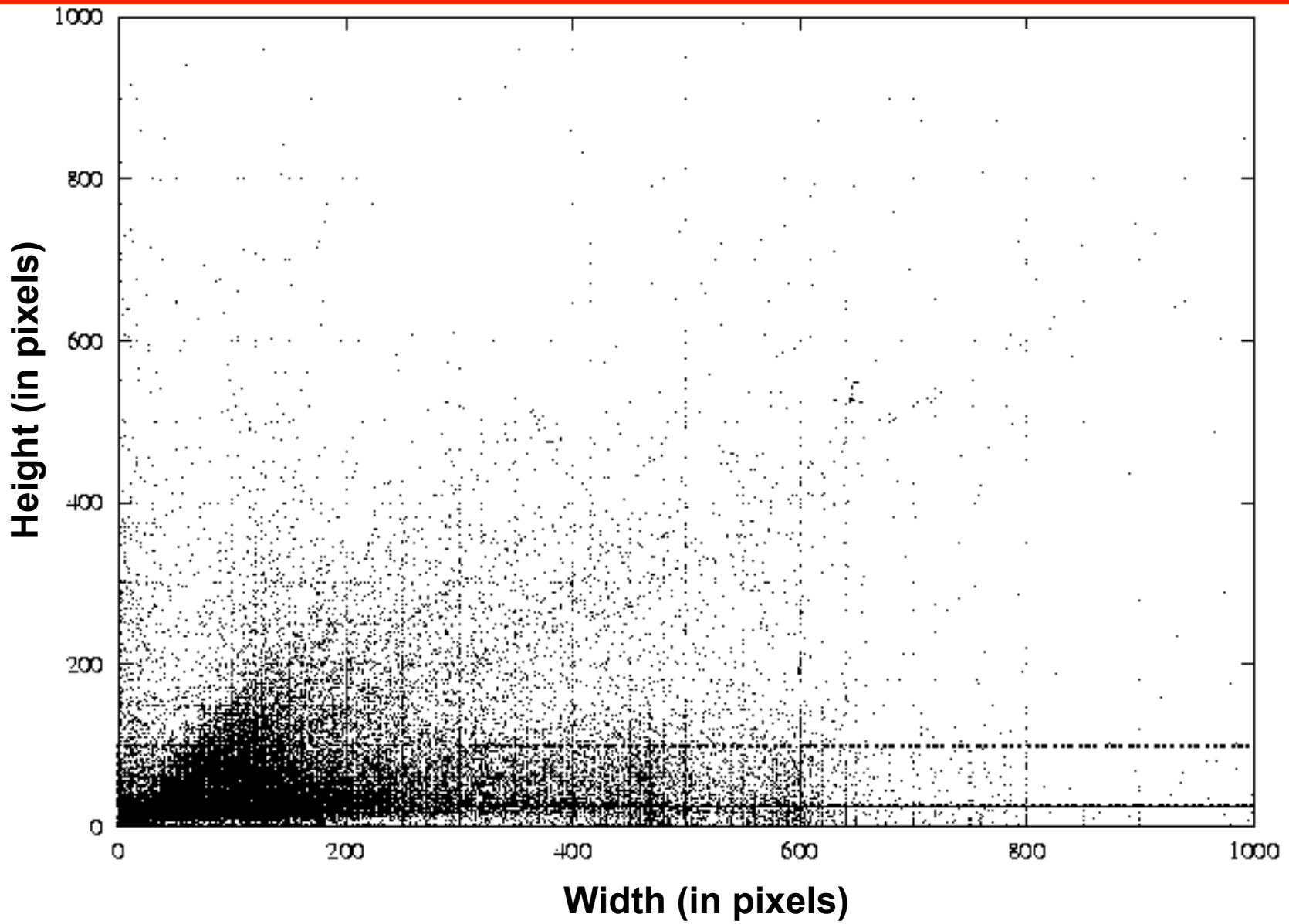
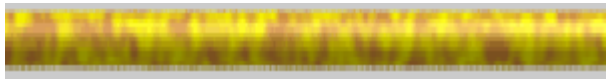


Image Categories

- Bullet - image geometry less than 25x25
  - HTML LI tag

- Lines - image width > 300, height < 25
  - HTML HR tag



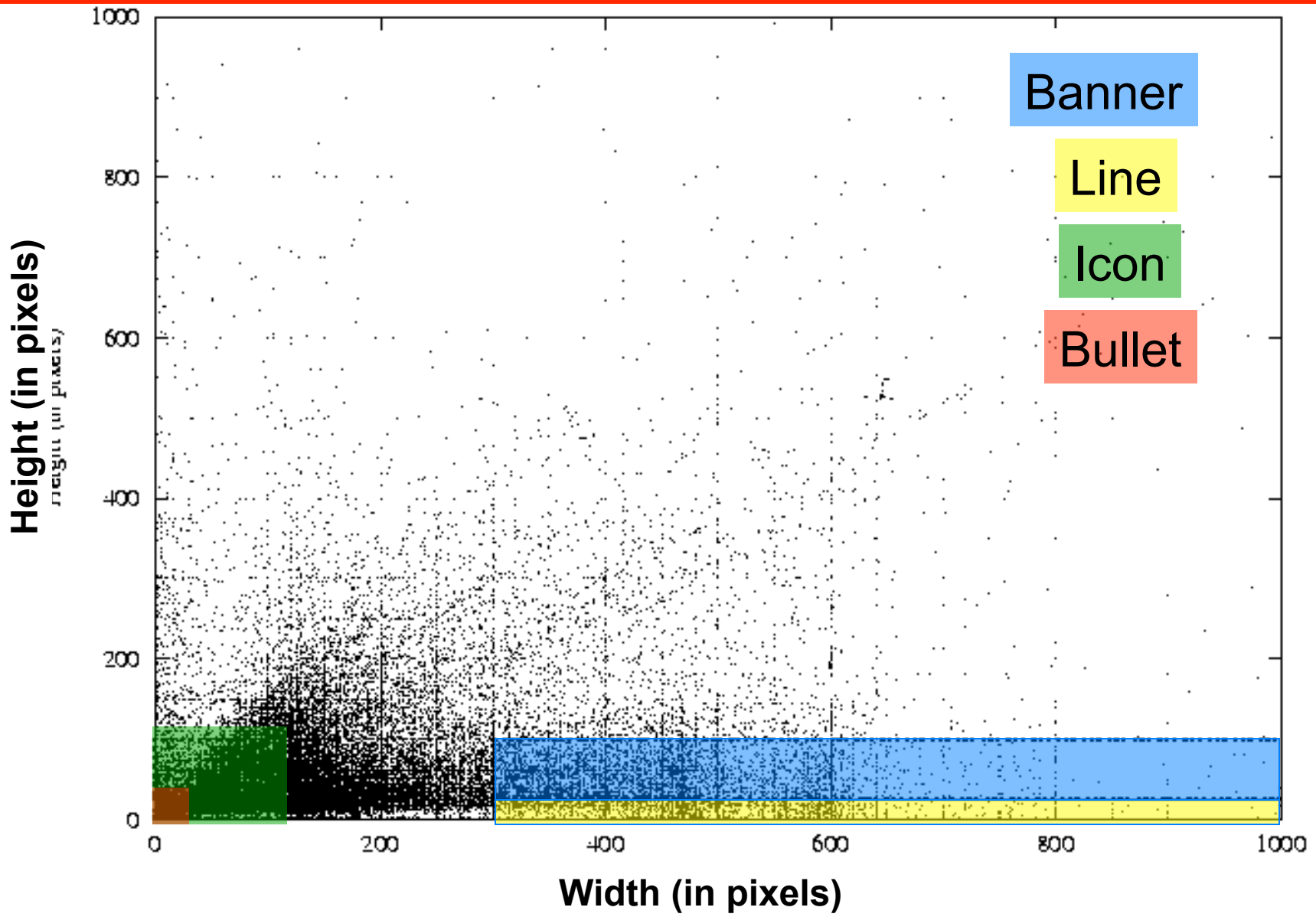
- Icons - between 25x25 and 100x100
  - HTML ALT text

- Banners - image width > 300, height between 25 and 100
  - HTML ALT text

Productive Transcoding

- at least 50% file saving for at least 50% of images

Spatial Size Distribution (GIF)



Spatial Size Distribution (JPEG)

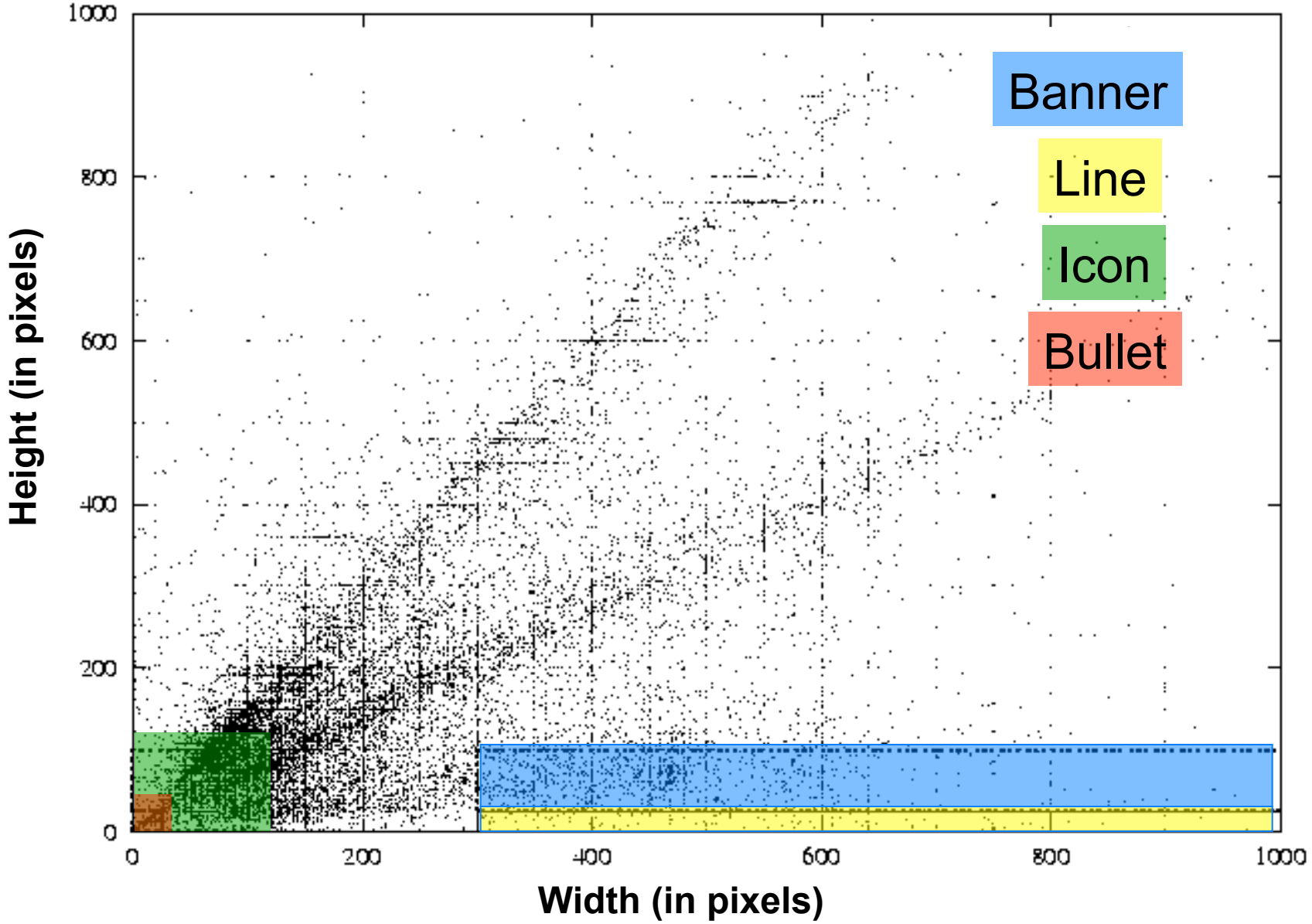
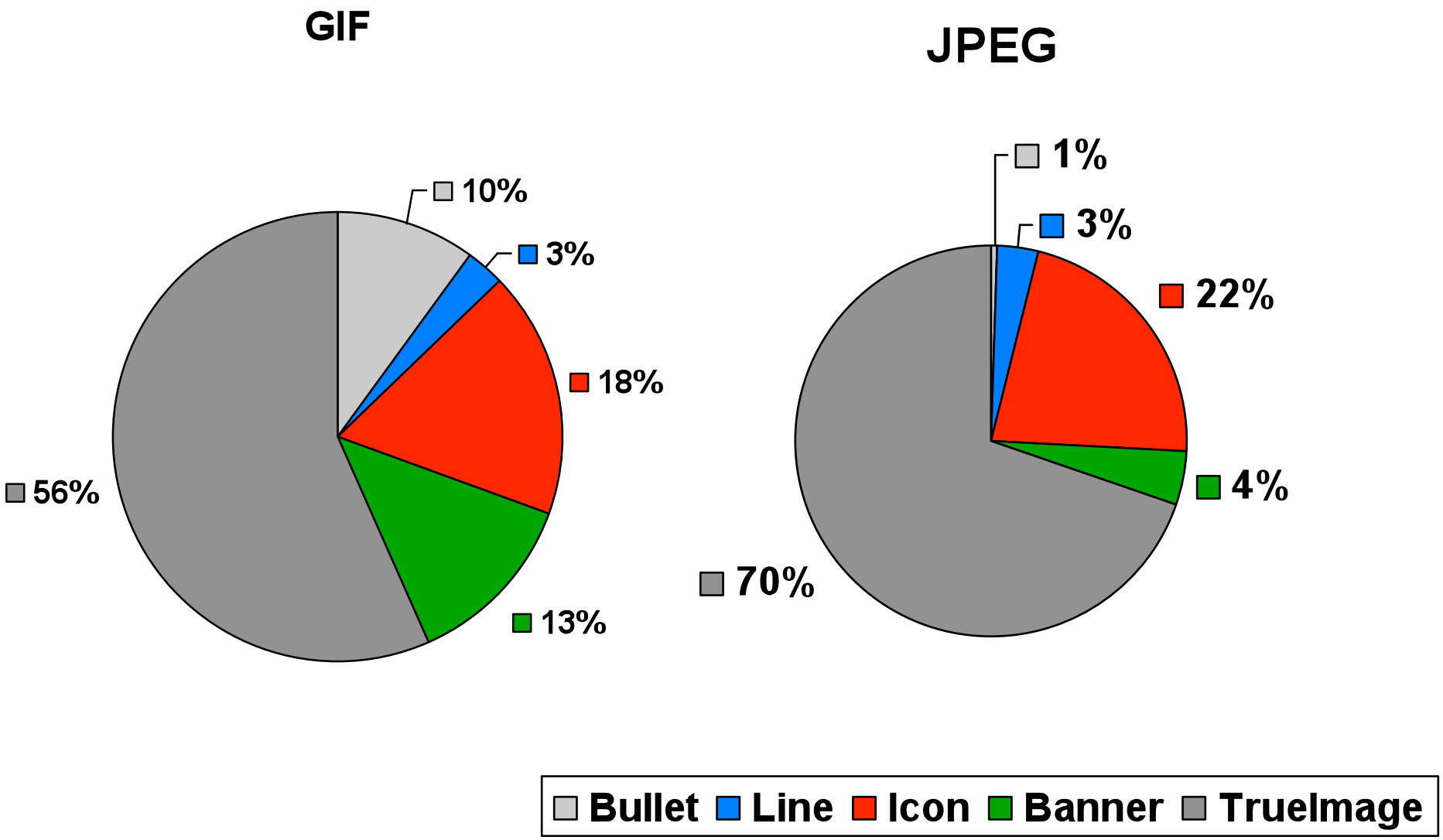


Image Category Distribution



JPEG Compression Metric

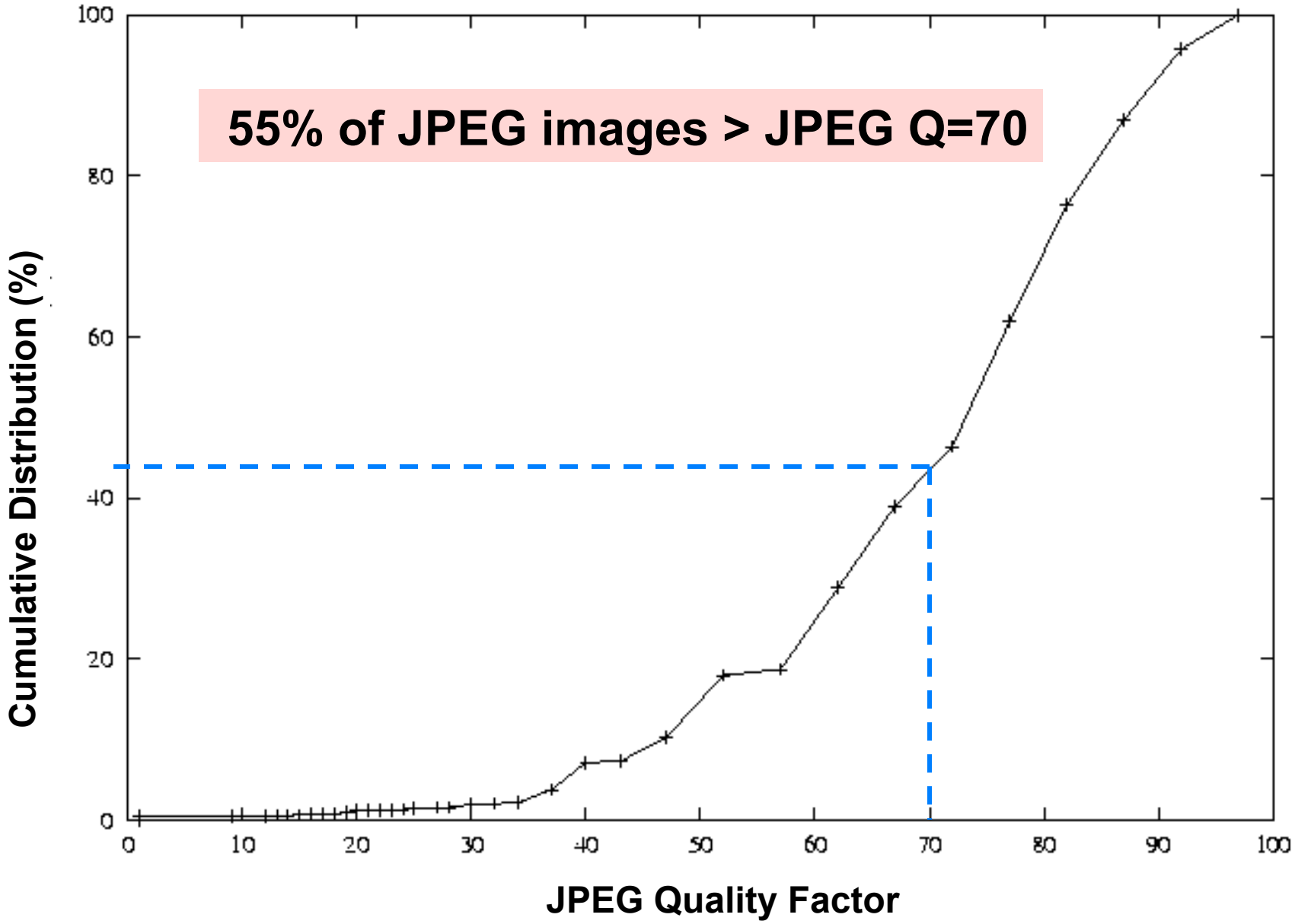
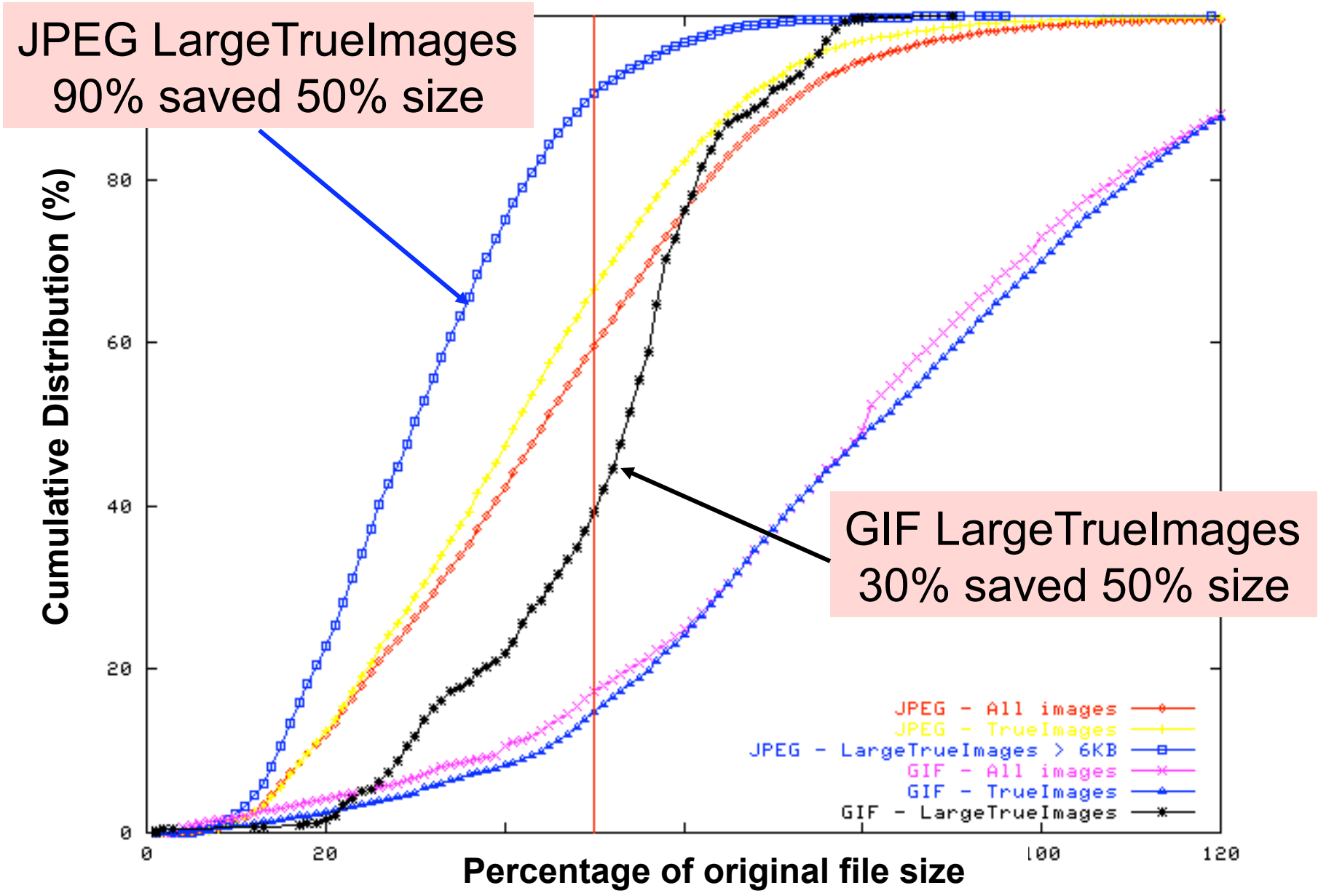


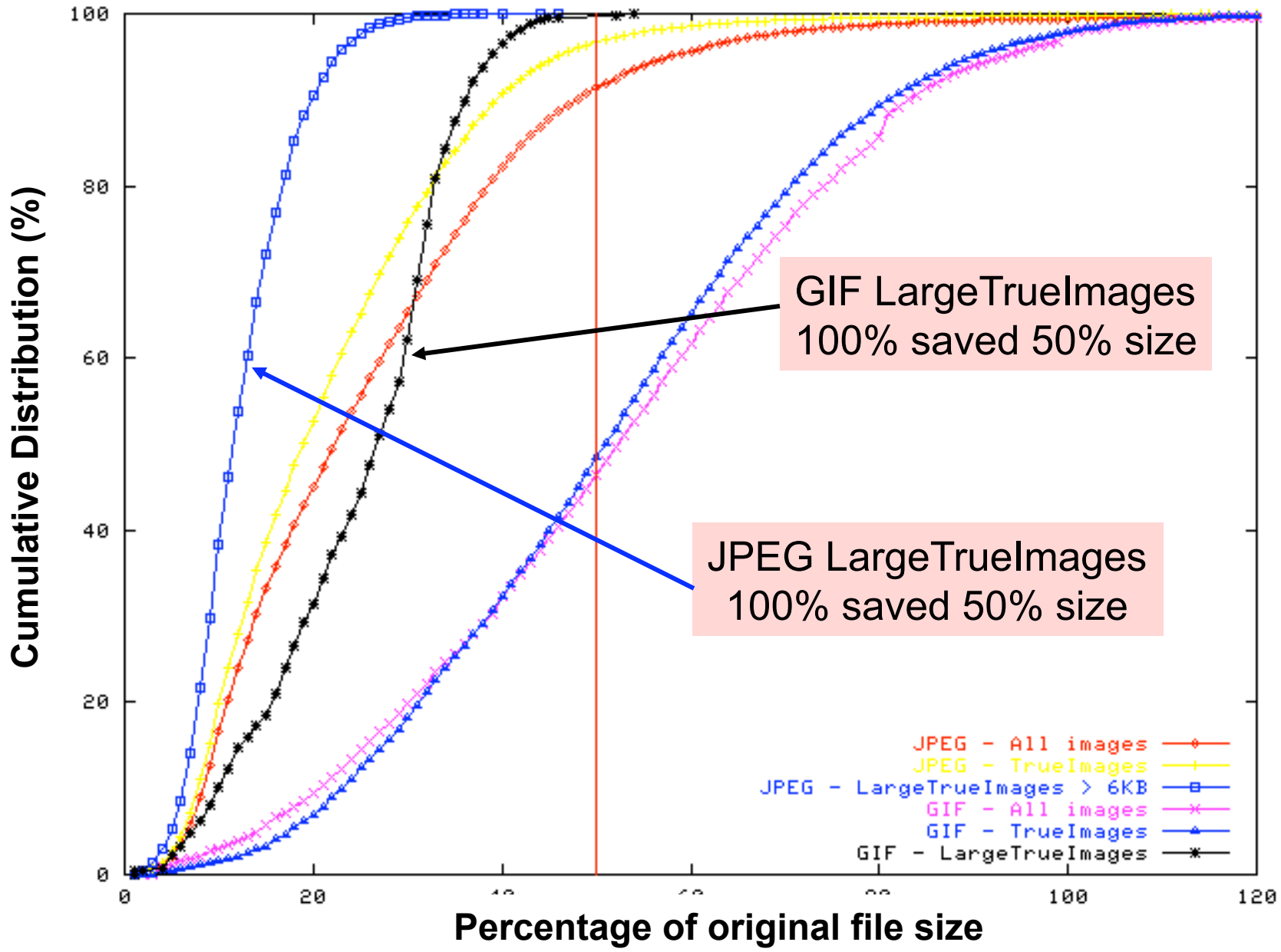
Image Transcoding Characteristics

- Transcodings analyzed:
  - spatial geometry reduction (thumbnail)
  - color reduction
  - format change (gif ⇒ jpeg, jpeg ⇒ gif)
  - JPEG compression metric

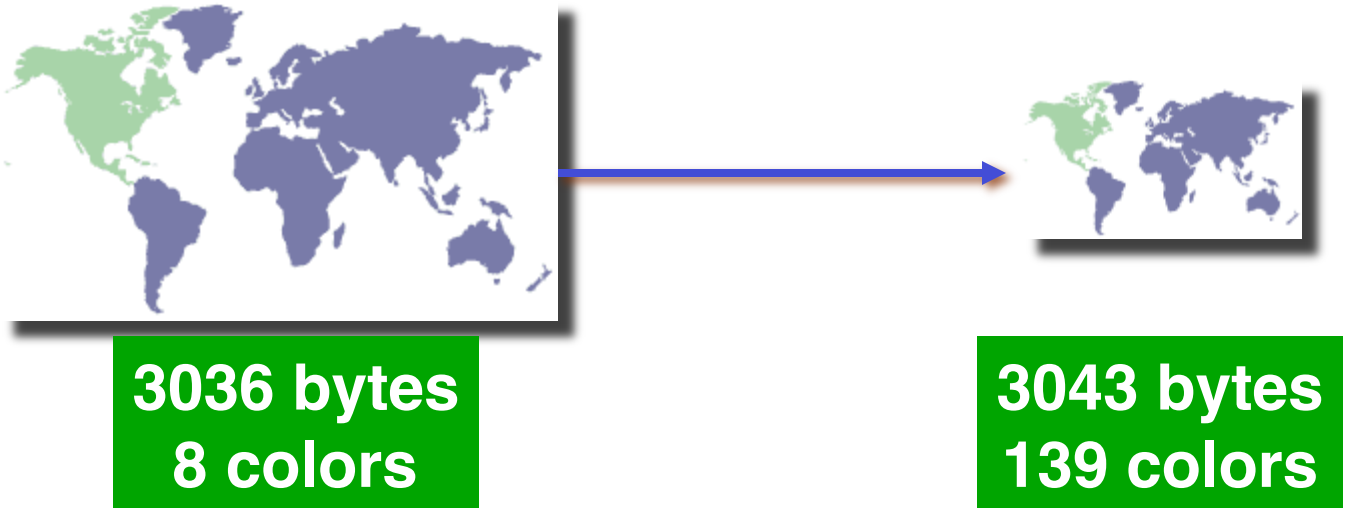
Spatial Geometry Reduction (by 2 along each axis)



Spatial Geometry Reduction (by 4 along each axis)

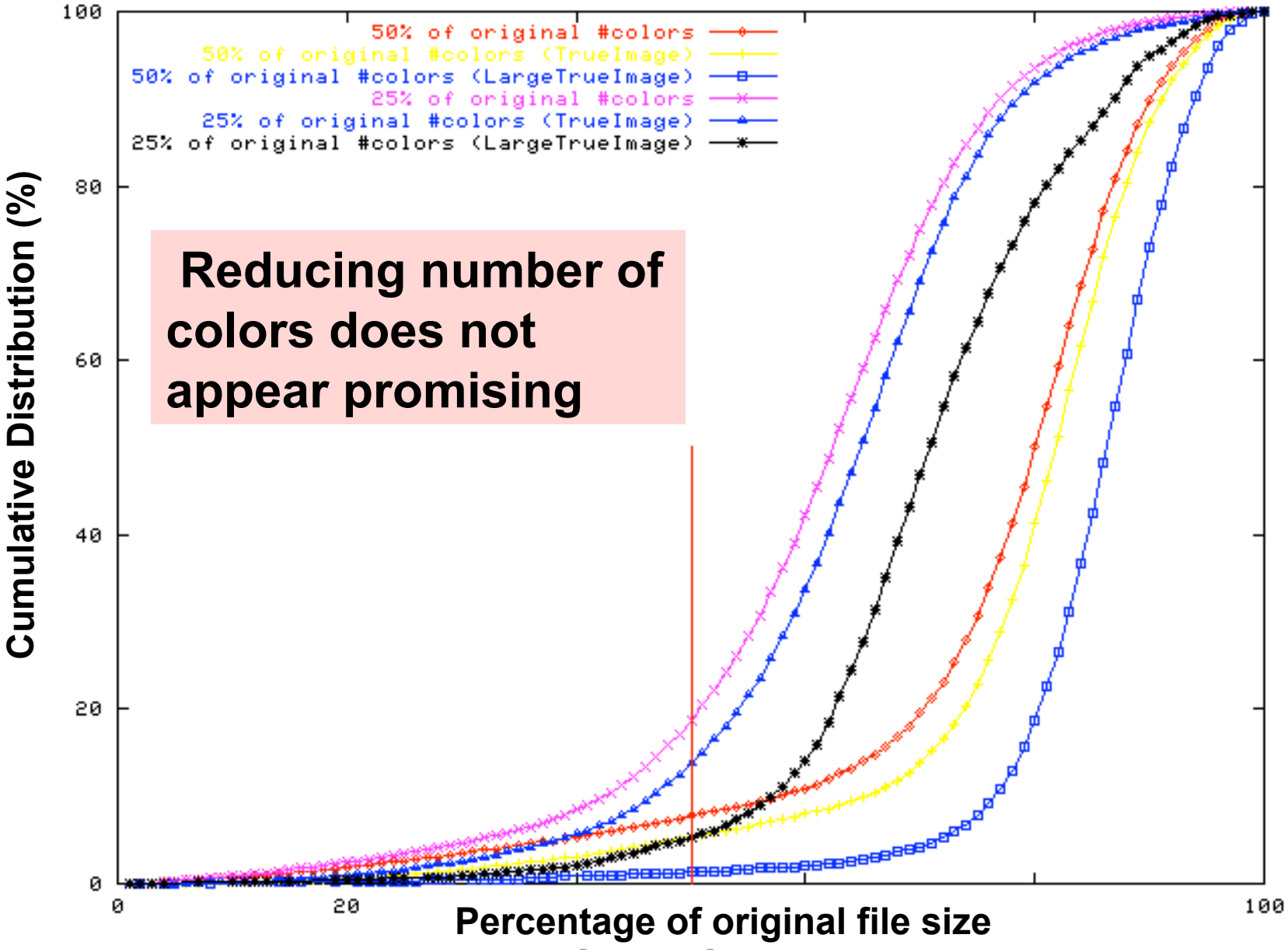


GIF Artifacts

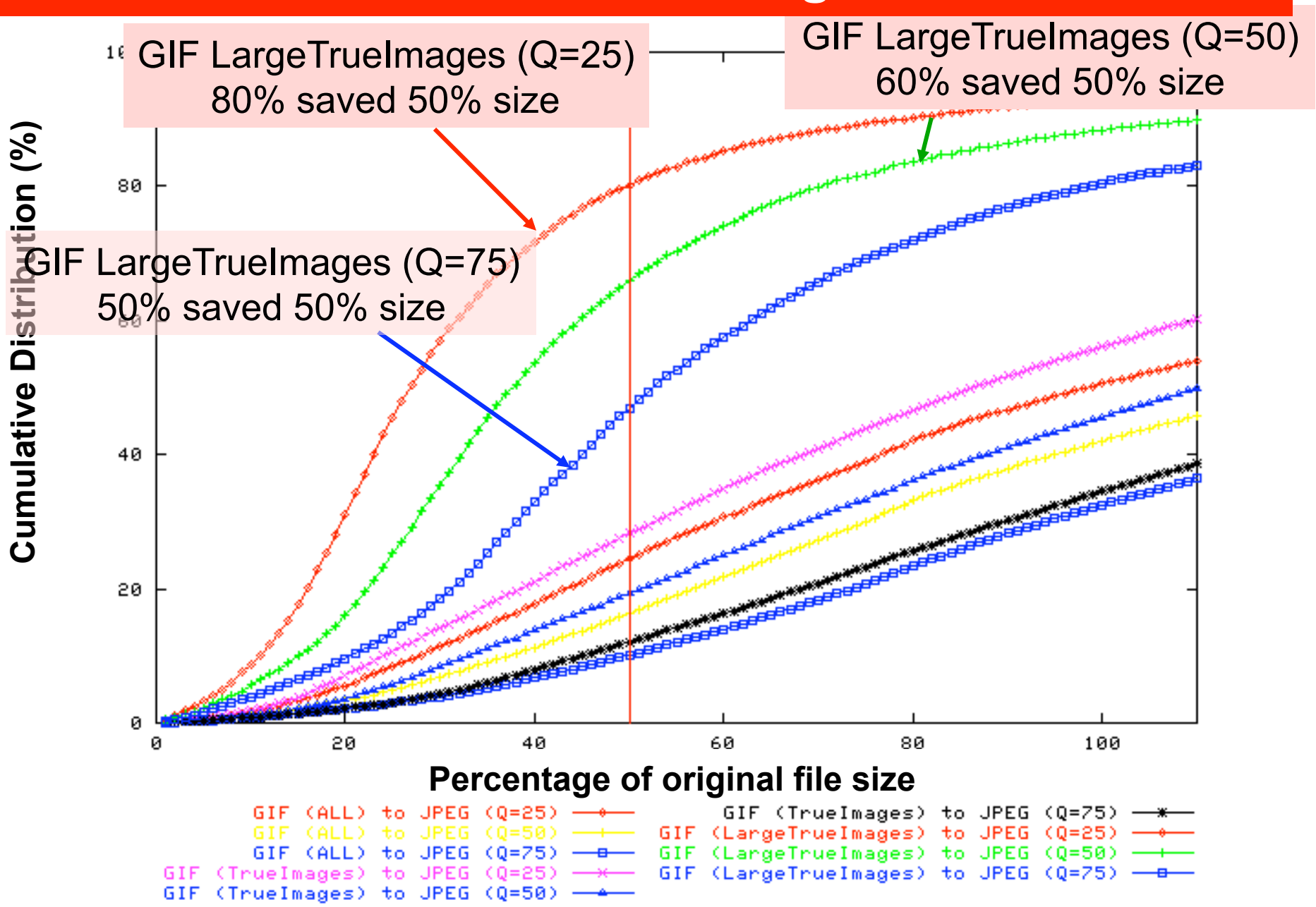


- GIF - variable bit coding
- Thumbnailing decreases relative frequency of popular colors
  - Unpopular colors use more bits

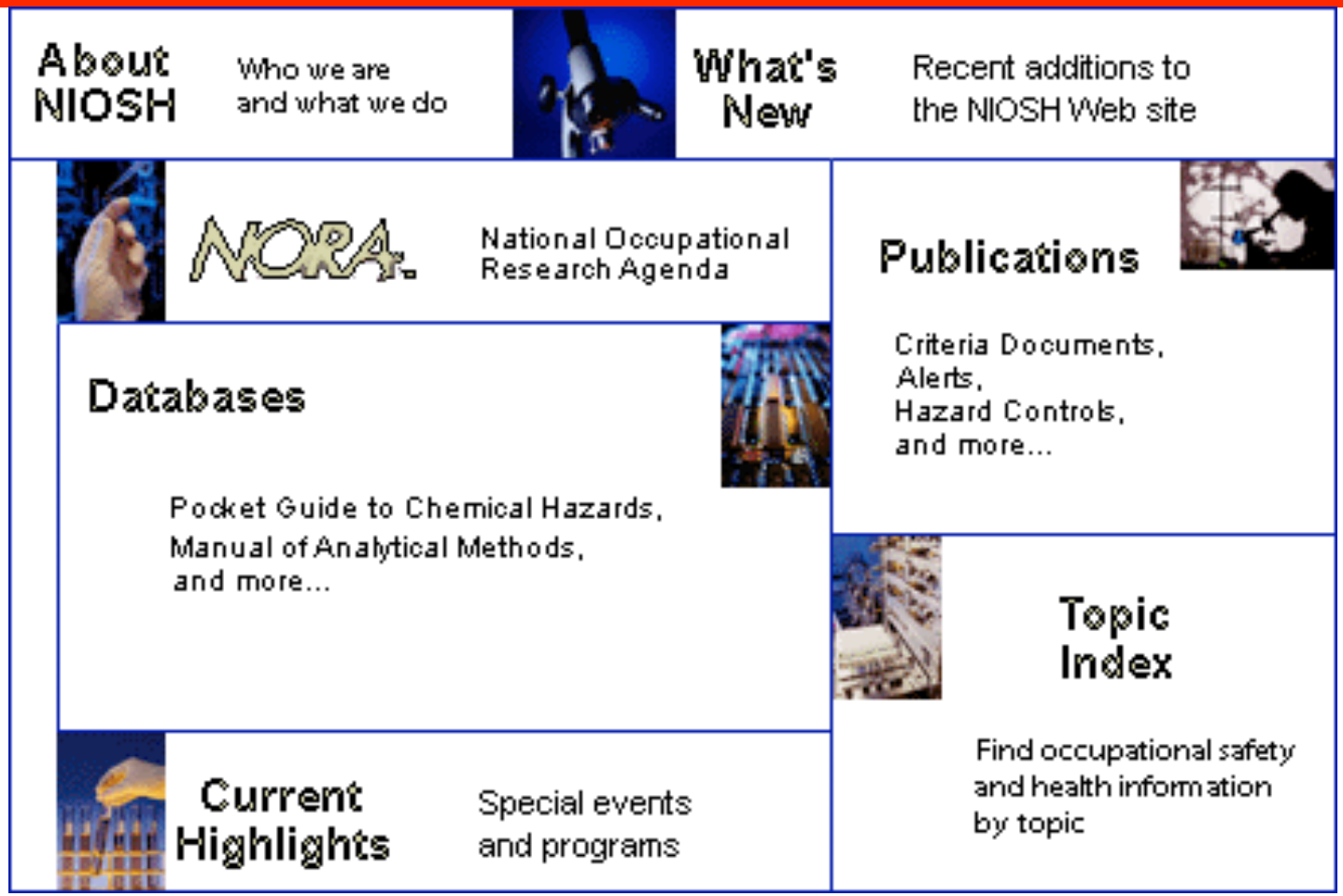
Reducing the Number of Colors (GIF)



GIF to JPEG images

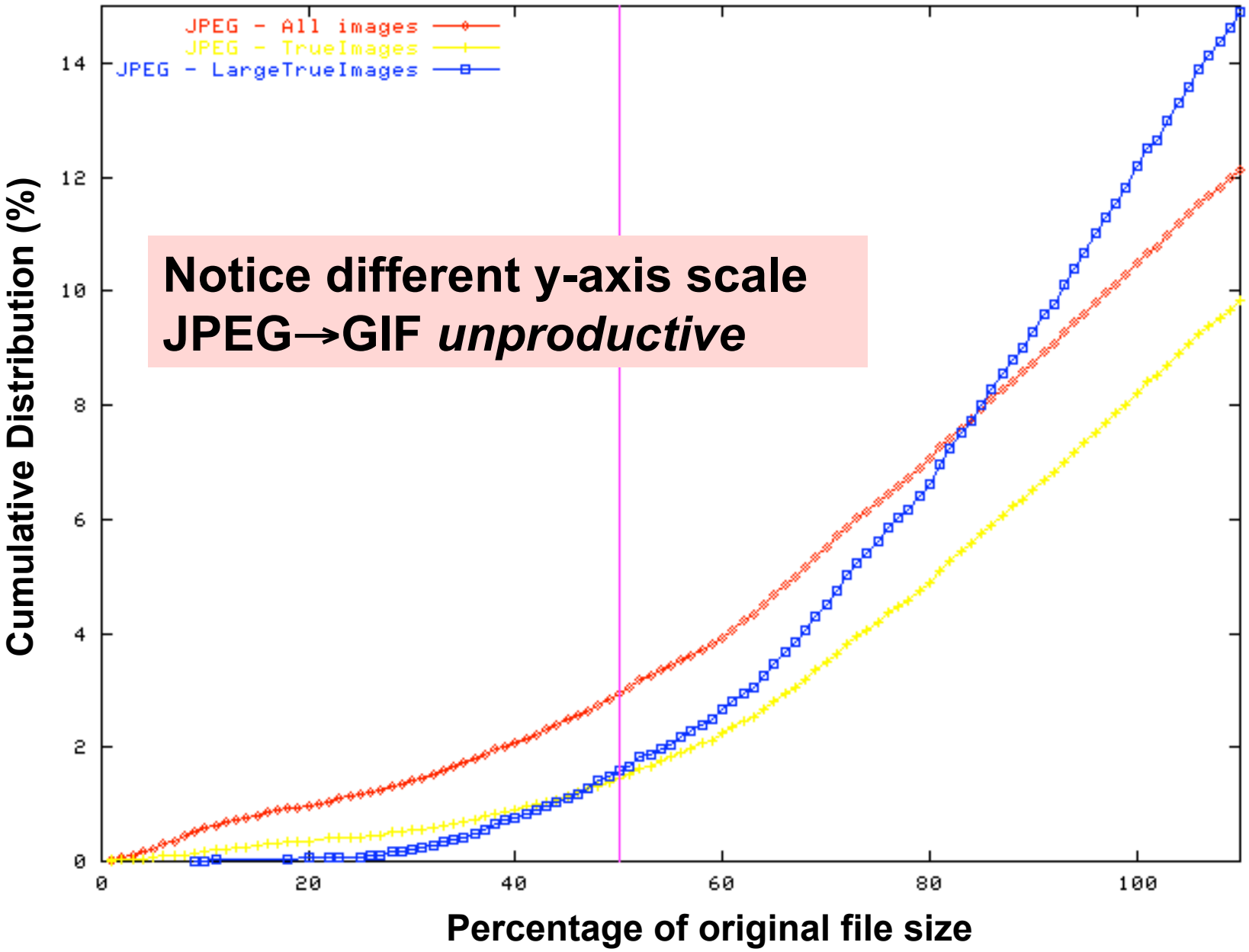


GIF artifacts



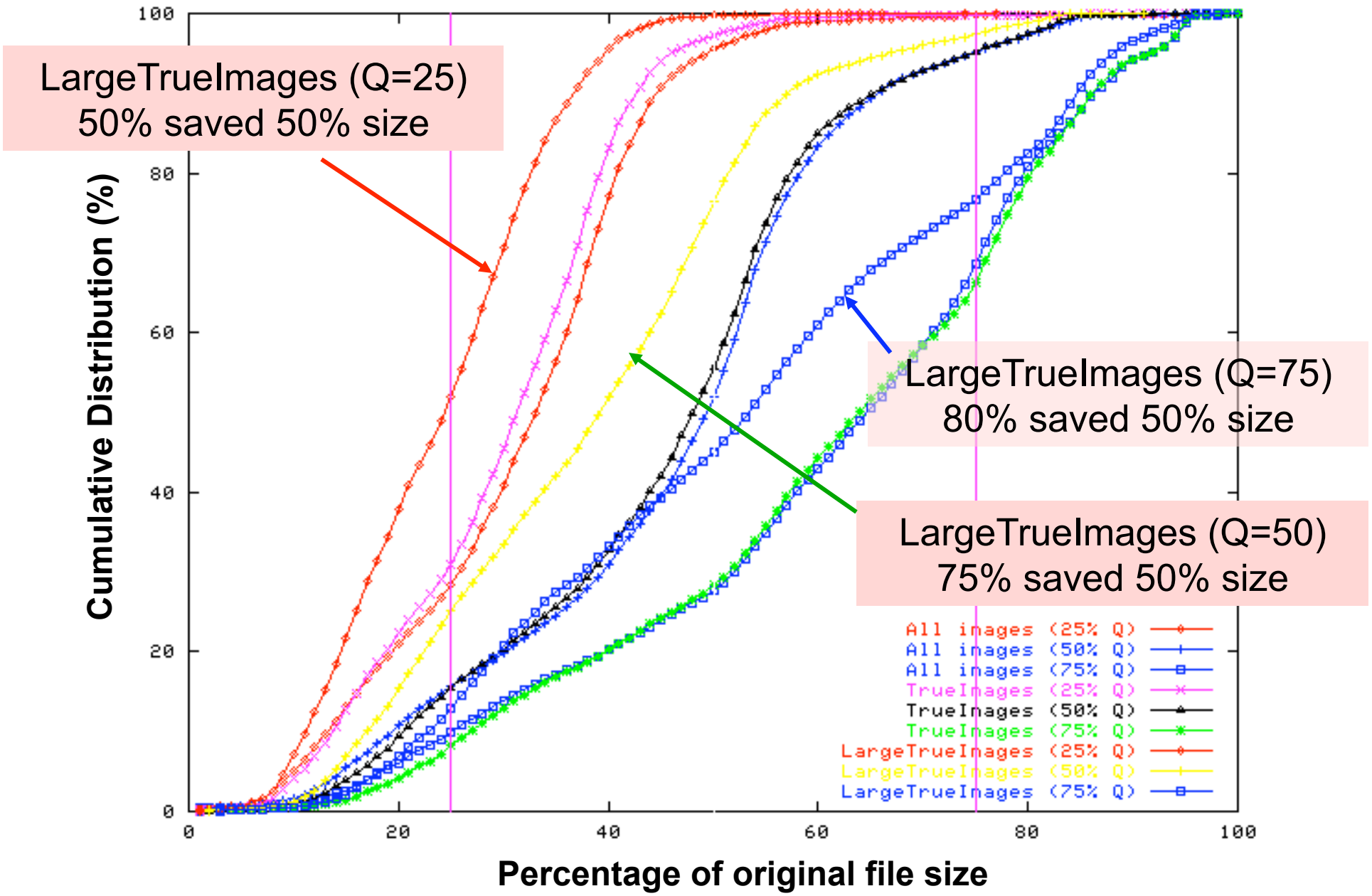
Some images are large, have 256 colors and still better encoded as GIF

JPEG to GIF



Notice different y-axis scale
JPEG→GIF *unproductive*

JPEG Quality Factor



Research Summary

- 80% of GIF images smaller than 6 KB
- 10% of GIF TrueImages make up 45.1% data
- Productive transcoding (50% file size saving for 50% of images)
  - Spatial geometry reduction by 4 along each axis
  - GIF->JPEG

- 60% of JPEG images smaller than 6 KB
- 35.6% of JPEG TrueImages make up 78% data
- Productive transcoding
  - JPEG compression metric
  - Spatial geometry reduction

Results useful in choosing image transcodings that can save image file size for a large class of Internet images