Using JPEG Quantization Tables to Identify Imagery Processed by Software

By

Jesse Kornblum

JPEG Quantization Tables

Jesse Kornblum

Overview

- Motivation
- Everything You Always Wanted to Know about JPEGs But Were Afraid to Ask
- Quantization Tables
- Types of Tables
- Calvin
- Future Work

Motivation

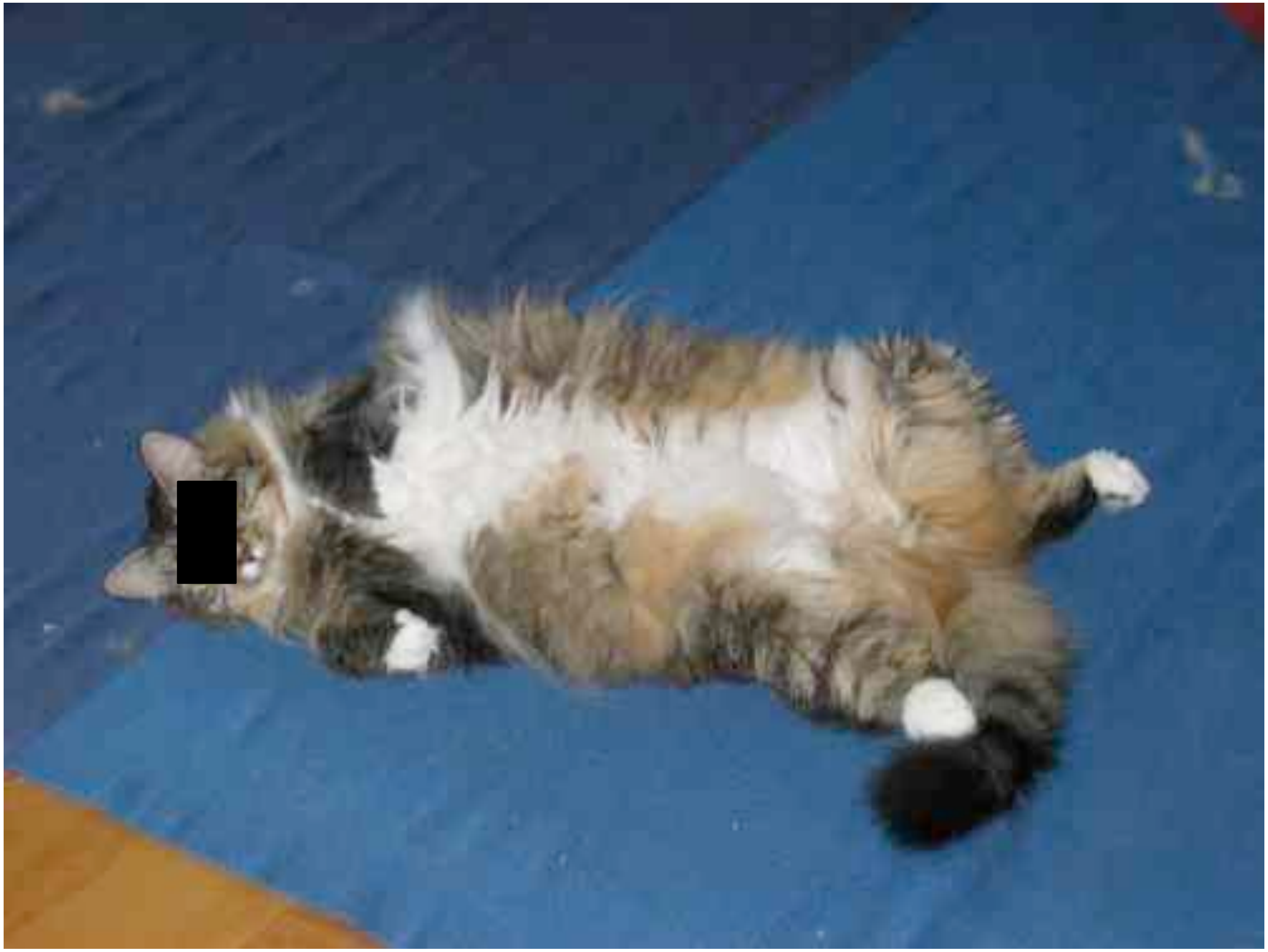
- *Ashcroft v. Free Speech Coalition*, 2002
- Cases now have hundreds of thousands of images
- Only a few needed to convict
  - Must be real pictures
- Need to find the real pictures
  - Not as easy as you'd think

Real Picture

WARNING:

EXPLICIT IMAGERY

Real Picture



Real Picture

Computer Generated Image

DC3

Pixar Kasuma 2006

All About JPEGs

- JPEG Compression
  - Lossy compression
- Six step process
  - Color space transform RGB to YCbCr
  - Downsampling
  - Block Splitting
  - Discrete Cosine Transform
  - Quantization (where the magic happens)
  - Encoding (lossless compression)

Quantization Tables

- Table used to control lossy compression
- Up to four sets of tables
  - 64 values in each table
- Value for each pixel is divided by a table value
  - Decimals thrown away
  - Decimal loss leads to image quality loss

- 124 / 50 --> 2
- When decompressed 2*50 = 100

Quantization Tables

2	1	1	1	1	1	2	1
1	1	2	2	2	2	2	4
3	2	2	2	2	5	4	4
3	4	6	5	6	6	6	5
6	6	6	7	9	8	6	7
9	7	6	6	8	11	8	9
10	10	10	10	10	6	8	11
12	11	10	12	9	10	10	10

Quantization Tables

- Higher numbers mean lower quality image
- Lower numbers mean higher quality image
- Best images have tables of all ones
  - No compression

Quantization Calculations

- Original value = 124

- Table value of 1 --> 124 --> 124
- Table value of 10 --> 12 --> 120
- Table value of 20 --> 6 --> 120
- Table value of 50 --> 2 --> 100
- Table value of 75 --> 1 --> 75

DC3

Making Tables

- Independent JPEG Group (IJG) Tables
  - Last updated 1998
- Scaling method uses quality factor Q
- Q can be between 1 and 100
- S = (Q < 50) ? 5000/Q : 200 – 2Q
- $T_s[i] = (S * T_b[i] + 50) / 100$
- Integer math
  - No decimals, information lost
- Scaling with Q=50 means no change

IJG Standard Table

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

IJG Standard Table, Q=80

6	4	4	6	10	16	20	24
5	5	6	8	10	23	24	22
6	5	6	10	16	23	28	22
6	7	9	12	20	35	32	25
7	9	15	22	27	44	41	31
10	14	22	26	32	42	45	37
20	26	31	35	41	48	48	40
29	37	38	39	45	40	41	40

IJG Standard Tables

- Most software uses IJG Standard Tables
- libjpeg is free and easy to use
  - Programmers are lazy
- Allows user to specify quality setting Q
- Examples:
  - The Gimp
  - Microsoft Paint
  - Infranview
  - Some camera phones

Extended IJG Tables

- Three tables instead of two
- The third is a duplicate of the second

A B

A B B

Adobe Photoshop

- Adobe Photoshop uses its own quantization tables
- Users select one of 12 quality settings
- Table depends only on quality setting
  - Does not consider image

Categorizing Quantization Tables

- All ones
  - No data
- Standard Tables
  - Two IJG
- Extended Standard Tables
  - Three IJG
- Custom Fixed Tables
  - Adobe Photoshop
- Custom Adaptive Tables

Custom Adaptive Tables

- Table is computed on the fly
- Usually based on image being processed
- Most cameras do this
  - Most vendors have patents on quantization table construction

Digital Ballistics

- Match images back to the device that created them
  - Match to *individual* device
  - Match to *type of* device

Digital Ballistics

- Match to individual devices
  - Depends on small imperfections in lens, sensor
  - Requires lots of images from each camera
  - Beyond the scope of this presentation

Digital Ballistics

- Match to type of device
  - Possible to identify IJG tables
    - Except when adaptive makes these by accident
  - Possible to identify Photoshop tables
    - But could, in theory, be adaptive tables
  - Possible to identify adaptive tables
    - But could be either hardware or software
- In all cases, may only be last device to process

Digital Ballistics

- Set of known quantization tables
  - 99 Standard Tables
  - 99 Extended Standard Tables
  - Tables from Adobe Photoshop
- Compare each unknown images to set of known
  - Matches are *most likely last processed by* software

Calvin

- Col. Calvin Goddard, 1891-1955
  - Founded firearms identification
  - Identified weapons used by Al Capone in St. Valentine's Day Massacre

Picture courtesy FBI, http://www.fbi.gov/hq/lab/labdedication/labstory.htm

Calvin

- **By default, displays filenames not matched (e.g. possible photographs)**

```
C:\> calvin *.jpg
C:\kitty-pr0n.jpg
```

Calvin

- Can display results for all files

```
C:\> calvin -vv *.jpg
C:\from-gimp.jpg: Standard Tables, Q=80
C:\kitty-pr0n.jpg: possible hardware
```

Calvin

DC3

- Can dump tables from an image

```
C:\> calvin -g kitty-pr0n.jpg
C:\kitty-pr0n.jpg
5,4,2,6,7,2,4,5,2,10,3,6,3,6,4,2,2,11,7,3,9,6,4,6,7,4,5,6
6,3,6,4,2,3,5,10,4,6,9,7,5,3,8,6,4,6,3,1,6,8,5,3,3,6,8,4,1,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
```

Calvin

- Can use signatures on next run

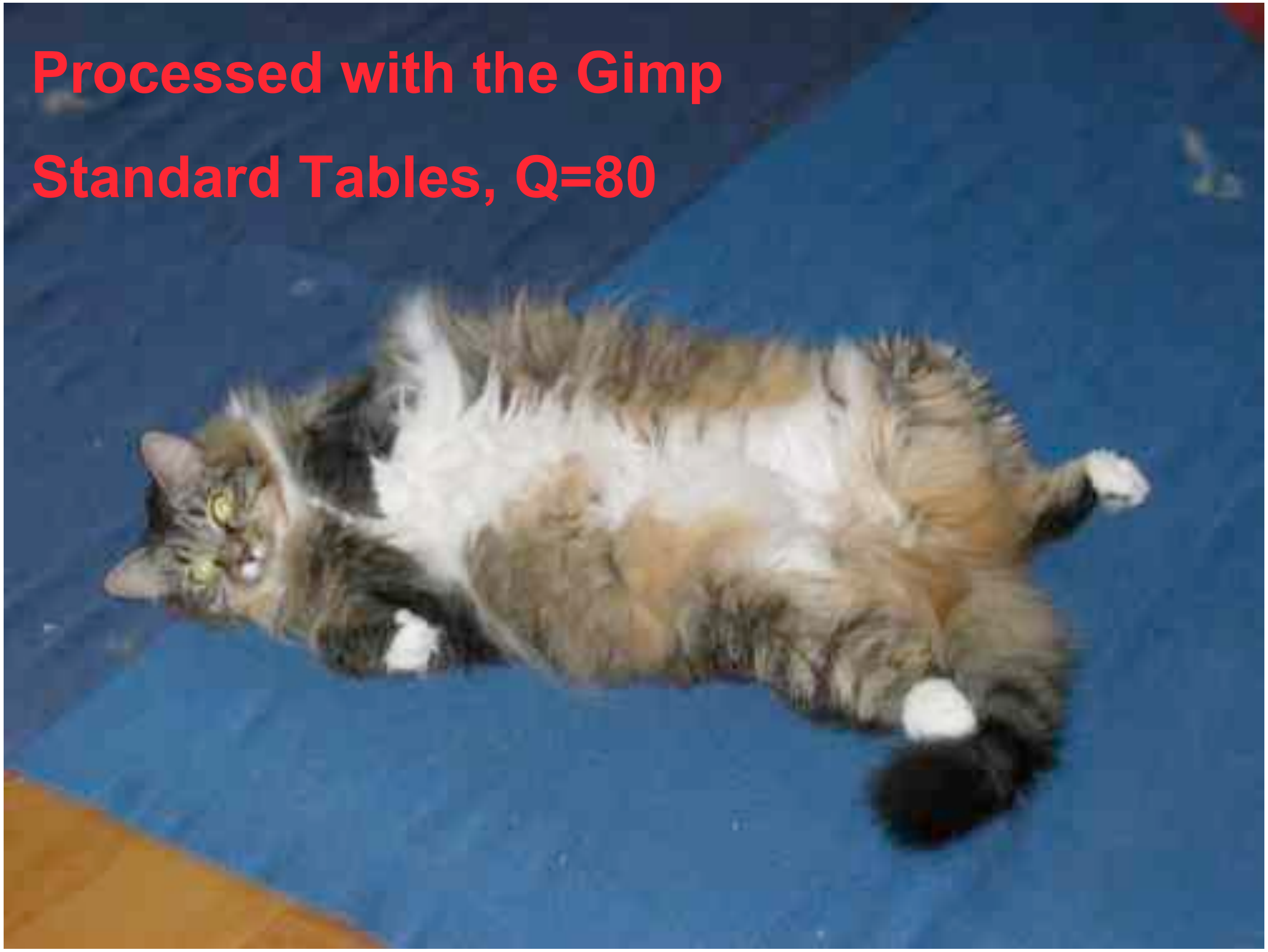
```
C:\> calvin -g kitty-pr0n.jpg > sigs.txt
C:\> calvin -a sigs.txt -vv d:\unknown\*.jpg
D:\unknown\also-kitty-pr0n.jpg: kitty-pr0n.jpg
```

Digital Ballistics

- This is just one step in the process
- Image from camera processed in Photoshop
  - According to Calvin, is from Photoshop
  - But image contains EXIF and other metadata
  - Clues in the image itself (e.g. presence of skin tones)

Digital Ballistics

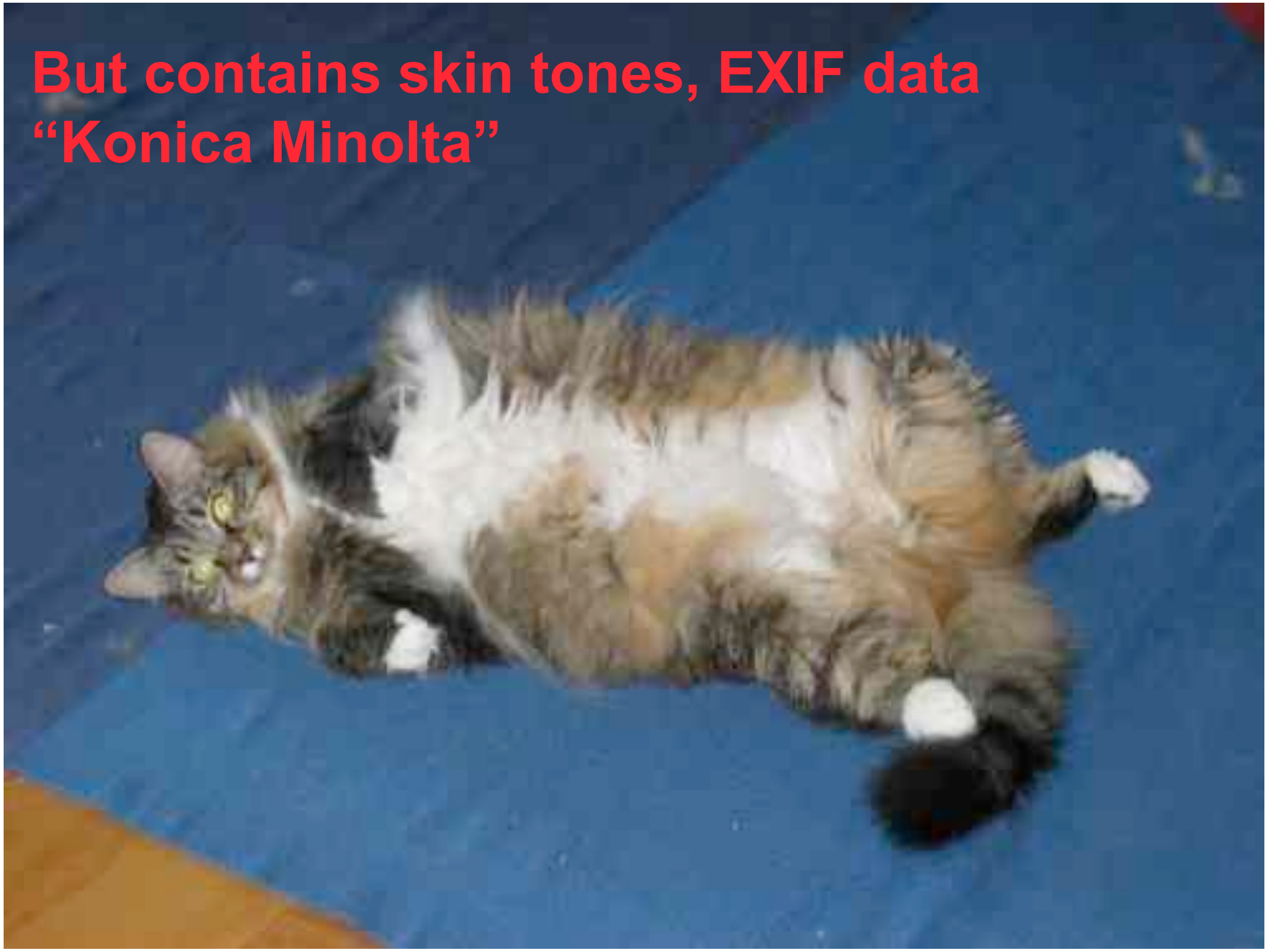
DC3



Processed with the Gimp

Standard Tables, Q=80

Digital Ballistics



But contains skin tones, EXIF data “Konica Minolta”

Digital Ballistics

- Best used as part of a larger system
- DC3 VISION system

Acknowledgements

- Imagery provided by FBI, Pisan Kaewma
- libjpeg: http://www.ijg.org/
- No animals were harmed in the making of this presentation

Department of Defense Cyber Crime Center

DC3

Jesse Kornblum