Midterm Review

Assignments 1,2: Lightbot 2.0
- RECURSION
- FUNCTIONS
- FUNCTIONAL ABSTRACTION

AIMS: Cover the Seven Big Ideas in computing
- As defined by the College Board for new AP test
  1. Computing is a creative human activity that enables innovation
  2. Abstraction is a way to understand and solve problems
  3. Data and information help to create knowledge
  4. Algorithms are tools for developing and expressing solutions to computational problems
  5. Programming is a creative process that produces computational artifacts
  6. Digital devices, systems, and the networks that interconnect them enable and foster computational approaches to solving problems
  7. Computing enables innovation in other fields, like sciences, engineering, humanities, etc.
  - We will have pre- and post-surveys
  - We want your feedback about what in computing you find exciting and interesting

Functions abstract by packaging Computation
- $F_1()$ packages actions: E.G. “process a riser”
Abstraction

- The word “abstraction” is used a lot in computing.
- Remember: it was one of the 7 big ideas.
- **Abstraction is a way to understand and solve problems.**
- As a general definition, abstraction eliminates details to focus on essential properties.
- The instruction example just given illustrates *functional abstraction* meaning that we have given a *name* to a *series of operations* that perform a coherent (and to us meaningful) activity; the name is the instruction, the series of operations are the bot’s actions to implement it.

Recursion in Everyday Life

- Mathematics: 0 … 9 are digits. A string of digits is a *digit* followed by a *string of digits*.

Abstraction …

- Formulating blocks of computation as a “concept” is **functional abstraction**
- What we did is important and we do it all the time in everyday life (otherwise we couldn’t deal with life’s complexity)...
  - We break a task down into (one or more) subtasks.
  - We solve a subtask using a sequence of instructions.
  - We put the solution into a function “package”, gave it a name, “process a riser,” and thus created a new thing, a concept, something we can talk about & use.
  - Then we used it to solve something more complicated … and probably repeat this approach at the next higher level.
- This lets us do more complicated things.

What controls when it happens?

- [Diagram of recursive behavior]

© 2010 Larry Snyder, CSE
Recursion is when

- A. You call the same function as many times as you want.
- B. A function calls itself
- C. You use functional abstraction.
- D. None of the above.

How do computers represent data?
How do computers compute?

Hollerith’s Tabulating Machine

To Process Data

- A mechanical machine can “read” a card with ... a “metal brush”
Sensing Punch Allows Some Action

- When the brush touches the source of current, the circuit closes,
- The electrical impulse can cause a mechanical action to happen that gives an instruction or records data

A General Idea

- Digital Information: Detecting the presence or absence of a phenomenon at a specific place and time: PandA
- Phenomena: light, magnetism, charge, mass, color, current, ...
- Detecting depends on phenomenon – but the result must be discrete: it was detected or not; there is no option for “sorta there”
- Place and time apply, but usually default to “obvious” values

Computing as part of our lives

- “Computing capacity is increasing at 58% annually, telecommunications at 28%, and storage at 23% per year. The former rate is approximately the rate of Moore’s Law, a doubling every 18 months. Communications are doubling every 34 months and storage every 40 months. Information has been expanding at this rate for the past decade.”

Increase of Computing Power
Big Data is here

The full scale of how much information we make is hard to appreciate. We humans collectively now have capacity to store approximately 300 exabytes of information. This is close the total amount of information stored in one person’s DNA. Or, as Hilbert puts it, it’s the equivalent of 80 Library of Alexandrias per person on the planet. And remember, the technium is doubling its capacity every year and a half, and your DNA is not. Broadcasting has grown at about the same speed as world’s GDP; but our information storage capacity has grown 4 times faster and telecommunication capacity has grown roughly 5 times faster than the world’s economic power.

The first electronic computers:

codebreaking during WW II

- http://www.youtube.com/watch?v=NbhbssXWDAE
- COLOSSUS in the U.K.
- ENIAC at Penn
- Female mathematicians were the first programmers!

McCulloch and Pitts 1943

- “A Logical Calculus of the Ideas Immanent in Nervous Activity”
- Simple model of human brain, neuron
- Influenced the design of the first computers
- Starts with Logic
- Ideas from logicians: Carnap, Russell and Whitehead.
McCulloch Pitts Neuron for AND

How the neuron computes logical AND

These are the fundamental units of computers

(NOT P) OR (NOT Q) vs. NOT (P AND Q)

- NOT (P ∧ 0) = NOT P ∨ NOT Q
- This is DeMorgan’s Law of Boolean Algebra
Truth Table for Not And (using 0 and 1)
- Not And = NAND

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>P AND Q</th>
<th>NOT (P AND Q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

(NOT P) OR (NOT Q) vs. NOT (P AND Q)

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>NOT P</th>
<th>NOT Q</th>
<th>P AND Q</th>
<th>NOT (P AND Q)</th>
<th>(NOT P) OR (NOT Q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

- NOT (P ∧ Q) = NOT P ∨ NOT Q
- This is DeMorgan’s Law of Boolean Algebra
- Boolean Algebra is “binary algebra”
- We showed the Law just with truth tables!
Consider two propositions, either of which may be true or false.

Exclusive-or is the relationship between them when JUST ONE OF THEM is true.

It EXCLUDES the case when both are true, so exclusive-or of the two is…

False when both are true or both are false, and true in the other two cases.

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>P XOR Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

What would you ever want XOR for anyway?

- Binary Addition!

What is binary addition?
What is binary addition? Let’s clicker this logic out!

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>S</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Transistors

A transistor is most like
A. A complete neuron calculating e.g. the AND function
B. A hole (or lack of) in a punched card
C. An entire punched card with multiple holes
D. A whole bank of light switches

Solid State Electronics

- A transistor is a switch: If the gate (black bar) is neutral, charge cannot pass; if gate is charged, the wires are connected

What else did we see binary representations for?
2/7/13

**With 8 places how many different letters?**

- **Leaf Node**

---

**UTF-8: All the alphabets in the world**

- Uniform Transformation Format: a variable-width encoding that can represent every character in the Unicode Character set
- 1,112,064 of them!!!
- UTF-8 is the dominant character encoding for the World-Wide Web, accounting for more than half of all Web pages.
- The Internet Engineering Task Force (IETF) requires all Internet protocols to identify the encoding used for character data
- The supported character encodings must include UTF-8.

---

**ASCII**

![ASCII Table](ASCII.png)

---

**RGB Color: [http://processing.org/learning/color/](http://processing.org/learning/color/)**

- Red + Green = Yellow
- Red + Blue = Purple
- Green + Blue = Cyan (blue-green)
- Red + Cyan + Blue = White
- no colors = Black
What is a color in processing?

- A sequence of 0's and 1's……
- Red, Green, Blue
- Each color element is represented by one BYTE (8 bits).
- \url{http://processing.org/learning/color/}
- So how many shades of Red (Green, Blue) are there?
  A. 8
  B. 16
  C. 256
  D. 512

Red or White?

This program draws a
A. red rectangle
B. white rectangle

\begin{verbatim}
rect(10,10,30,40);
fill(255,0,0); 
\end{verbatim}

Key ideas from last time

- By 1943, it had been shown that computers could be used for......???
  - Counting, Arithmetic
  - Translation of language (code breaking in WWII. BBC video)
  - Logic. McCulloch and Pitts, Neuron models for And/Or

Key ideas from last time

- Computational Thinking/programming:
  - Functional abstraction
  - Subroutine (F2 that does a part of the job)
  - Bottom up vs. top down programming

- Representation of data on computers:
  - Bits, bytes, ASCII, UTF
  - Presence or Absence of signal: PandA
  - Signal can be mechanical, electrical, color, light, blinking an eye, moving a finger, yes, no, true false
  - Hollerith tabulating machine: mechanical => electrical
**Total Information Awareness**

- Pharmacy purchases
- Elementary school attendance records
- Doctors' visits

Application: School absences and pharmacy records could lead CDC to identify bioterrorism

Assistant to Poindexter: “We were surprised that privacy was an issue”

The government can’t collect this information but we can pay Johns Hopkins to do it. They have a medical school.

**Data Aggregators: Computational Advertising**

is getting to be a big business
How did we get here?

What technological innovations have brought us where we are today?

**Integrated Circuits: Millions of logical gates**

- Cheaper fabrication, greater and greater complexity per inch
  => Miniaturization!!!!
  => Everyone can afford!!

**The Internet**

- Invented in 1969 for military purposes, it took almost 20 years to get out of the lab
- Communication by FTP (file transfer protocol).
- Ascii Terminal interface
- E.g. 1983 War Games Film
WWW + http: Early 90's changed everything

- WWW: An INTERFACE!!
- All computers “speak” a common language: hyper-text transfer protocol. HTTP
- Content points to other content
  - (Google page rank, later)
  - UTF ensures content of pages in any language can be displayed

So how did we get here?

- ICs =>
  - Personal
  - Miniaturization
- Internet
- WWW
- Digitization of Content
- Mobile
- The same technology that is incredibly useful and often fun! (so we like it), also affords ‘Big Brother’

Contrast Two Videos: Surreptitious, Not

- Star Wars Kid
  - A/K/A Ghyslain Raza
  - Just messing around in drama studio at school
  - Other kids found and uploaded
    - http://www.youtube.com/watch?v=HPPj6viiBmU
- Numa Numa Guy
  - A/K/A Gary Brolsma
    - http://www.youtube.com/watch?v=WgmgUHFhEARq&feature=fvw_re

Posting Audio/Video Online: Good or Evil?

Alexandra Wallace original, 2.52 minutes
Death Threats, Personal Info, Leave UCLA

Death Threats Persuade Alexandra Wallace To Leave UCLA

Over the weekend, Alexandra Wallace, a UCLA student that was arrested after uploading a racist rant against亚洲人 on YouTube shortly after the recent riots in Japan, apologized and announced that she would no longer be attending UCLA.

Her apology, and the news of her leaving the school, was published by The Daily Bruin, UCLA's online news source.

The full letter, as printed in The Daily Bruin, is as follows:

In an attempt to produce a harmless YouTube video, I have offended the UCLA community and the nation as a whole. I
acknowledge this mistake and the pain it caused anyone who was
affected by it. I am deeply sorry for my actions.

I would like to apologize to the UCLA community for my actions.

I am deeply sorry for my actions. I understand that my actions
have caused harm and that I have对不起 anyone who was affected
by my actions.

Alexandra Wallace

His personal details leaked by Anonymous

- Huffington Post Tony Bologna followup
- As an act of retribution, the hackers group Anonymous identified Bologna and leaked his personal information including his address, phone number, and names of known family members.

Recent Incidents: Cell phone videos

Occupy Wall Street

- pepper spray

NEW YORK (AP) — An internal New York Police Department review has found an official violated department guidelines when he used pepper spray on Occupy Wall Street protesters last month, a person with knowledge of the investigation said Tuesday.

Deputy Inspector Anthony Bologna faces discipline of a loss of 10 vacation days after the Sept. 24 incident near Union Square, shortly after the now-global protests began in a tiny private plaza in lower Manhattan, the person said. The person had direct knowledge of the review but was not authorized to speak publicly and spoke to The Associated Press on condition of anonymity.

Wikileaks – Changing democracy?

- Wikileaks Documentary.
- 55 minutes long.
- First five minutes, plus from time 30 min – 35 minutes in class
- Watch whole thing for homework!
Kinds of questions

- What do these examples show?
- Whose privacy was violated?

Wikileaks – Changing democracy?

- Wikileaks Documentary.
- 55 minutes long.
- First five minutes, plus from time 30 min – 35 minutes in class
- Watch whole thing for homework!
Why do you think I spent so much time on this? Digital Literacy…..“Digerati”

Who tries to protect your privacy?

- American Civil Liberties Union
- Electronic Privacy Information Center
- Electronic Frontier Foundation
- Center for Democracy and Technology
- Cycle of surprising privacy violation, lobbying for laws, new laws, but new technology or new twist, then cycle...

What are you supposed to learn?

- What is a digital footprint?
- What technology advances in the last ten years have made ‘Big Brother’ possible
- Which organizations try to protect your privacy.
- Why you should read the “Terms and Conditions” for every app you download
  - You never know what they might say. Example: Pulse App asks you to give permission for them to track every number you call
- Why you should consider what you put onto public sites like Facebook.
  - Are you sure your privacy settings are as you want?
- Name three recent cases of information going viral that could never have happened ten years ago.

Kinds of questions

- Name some new technologies that that make it easy to violate someone’s privacy
- Name some applications discussed in class that make it easy to violate someone’s privacy
- What do TIA and Google and Facebook have in common?
Processing …

- Processing is a language for programming graphical and image-based computations
  - More fun than programming an operating system
  - Easier to do because we “see” what’s happening
  - Immediate feedback => bottom up programming style

Processing is ALL Detail

- Notice everything!

```java
void setup() {
  size(400, 400);
  background(192, 64, 0);
  stroke(255, 255, 255);
}

void draw() {
  line(150, 158, 180, 408);
}
```

- Two Functions, One Common Form:
  - void <name> () {
  - } all symbols + placement, matter
- Every statement ends with a semicolon (;)
- The software colors text it understands – helpful
- Some functions include stuff inside parentheses; these are called arguments
  - If a function has arguments, each position has a specific meaning:
    - size(<width>, <height>);
    - stroke(<red value>, <green value>, <blue value>);
  - If your cursor is by a closing parenthesis or brace, the matching parenthesis or brace is highlighted
- Keywords are highlighted in blue
- Processing is case sensitive; notice!

Guest Lecture: Social media. Lifelogging. Digital Memory
Last week: First guest Lecture

- Guest Lecture Thursday: Lifelogging & Social Media
- Prof. Whittaker’s Home Page
- Testing on invited talks:
  - No homework on them but possible clicker questions or midterm questions that demonstrate
    - That you were there
    - That you were paying attention
    - That you understood the presentation
    - Come prepared to ask questions!

Clickers. Prof. Whittaker Guest Lecture.

- The main points of Professor Whittaker’s guest lecture included that:
  A. Current capacity for digital storage make it possible to really do Lifelogging
  B. Lifelogging is ethically and morally wrong and should be stopped
  C. To make Lifelogs useful, all we have to do is just record everything
  D. All of the above
  E. None of the above

Clickers. ChattyWeb and Piccy Web

- This table from Prof. Whittaker’s guest lecture shows:
  A. Best predictor of student’s grade was GPA coming into the class
  B. How much you used Chatty & Piccy significantly improved your grade.
  C. None of the above.

What does the graph show?

A. People often remember very little about an arbitrary day.
B. People who have been wearing a SenseCam cannot remember everything.
C. An interface that combines the information from both the SenseCam and GPS makes it possible to remember more about an arbitrary day
D. Surprisingly, when we are talking about improving human memory, an interface showing where you went in your day works better than an interface with information from the SenseCam.
E. All of the above
Huffman Coding, More bits and bytes.

Binary combinations, True/False possibilities

- One bit
  - 0
  - 1
- Two bits
  - 00
  - 01
  - 10
  - 11
- Three bits
  - 000
  - 001
  - 010
  - 011
  - 100
  - 101
  - 110
  - 111
- Four bits
  - 0000
  - 0001
  - 0010
  - 0011
  - 0100
  - 0101
  - 0110
  - 0111
  - 1000
  - 1001
  - 1010
  - 1011
  - 1100
  - 1101
  - 1110
  - 1111

Clickers: Which one represents different numbers?

A. 01 vs. 001
B. 11 vs. 110
C. 10 vs. 000010
D. 110 vs. 0110

What is binary addition? Let’s clicker this logic out!

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>S</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
What is the answer?

```
001111  
+110111
```

A. 0000110  
B. 1000110  
C. 1000111  
D. 1000101  
E. 0000101

---

UTF-8: All the alphabets in the world

- Uniform Transformation Format: a variable-width encoding that can represent every character in the Unicode Character set
- 1,112,064 of them!!!
- UTF-8 is the dominant character encoding for the World-Wide Web, accounting for more than half of all Web pages.
- The Internet Engineering Task Force (IETF) requires all Internet protocols to identify the encoding used for character data
- The supported character encodings must include UTF-8.

---

With 8 places how many different letters?

- Leaf Node

---

How many bits for all of Unicode?

There are 1,112,064 different Unicode characters. If a fixed bit format (like ascii with its 7 bits) were used, how many bits would you need for each character? (Hint: $2^{10} = 1024$)

A. 10  
B. 17  
C. 21  
D. 32  
E. 40
Coding can be used to do Compression

- **What is CODING?**
  - The conversion of one representation into another

- **What is COMPRESSION?**
  - Change the representation (digitization) in order to reduce size of data (number of bits needed to represent data)

- **Benefits**
  - **Reduce storage** needed
    - Consider growth of digitized data.
  - **Reduce** transmission cost / latency / bandwidth
  - When you have a 56K dialup modem, every savings in BITS counts, **SPEED**
    - Also consider telephone lines, texting

What makes it possible to do Compression?

- **IN OTHER WORDS: When is Coding USEFUL?**
  - **When there is Redundancy**
    - Recognize repeating patterns
    - Exploit using
      - Dictionary
      - Variable length encoding
  - When **Human perception is less sensitive** to some information
    - Can discard less important data

Huffman Code Algorithm Overview

- Order the symbols with least frequent first (will explain)
- Build a tree piece by piece…
- **Encoding**
  - Calculate frequency of symbols in the message, language
  - **JUST COUNT AND DIVIDE BY TOTAL NUMBER OF SYMBOLS**
  - Create binary tree representing “best” encoding
  - Use binary tree to encode compressed file
    - For each symbol, output path from root to leaf
    - Size of encoding = length of path
  - Save binary tree

Huffman Tree Construction 5

- A = 111
- C = 10
- E = 01
- I = 00
- H = 110
- 25
- 2
- 3
- 5
- 10
- 0
- 1
Huffman Coding

- How much wood could a woodchuck chuck?
- Peter Peter pumpkin eater
- She sells sea shells by the sea shore.
- Bananarama

Huffman Code Example

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Dog</th>
<th>Cat</th>
<th>Bird</th>
<th>Fish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>1/8</td>
<td>1/4</td>
<td>1/2</td>
<td>1/8</td>
</tr>
<tr>
<td>Original Encoding</td>
<td>00</td>
<td>01</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Huffman Encoding</td>
<td>110</td>
<td>10</td>
<td>0</td>
<td>111</td>
</tr>
</tbody>
</table>

How many bits are saved using the above Huffman coding for the sequence Dog Cat Bird Bird Bird?

A. 0     B. 1    C. 2     D. 3    E. 4

Huffman Code Algorithm Overview

- Order the symbols with least frequent first (will explain)
- Build a tree piece by piece…
- Encoding
  - Calculate frequency of symbols in the message, language
  - **JUST COUNT AND DIVIDE BY TOTAL NUMBER OF SYMBOLS**
  - Create binary tree representing “best” encoding
  - Use binary tree to encode compressed file
    - For each symbol, output path from root to leaf
    - Size of encoding = length of path
  - Save binary tree

Clicker: TO BE OR NOT TO BE

Which nodes can get combined next?

A. The brown node worth 2 (R, N) and the E node worth 2
B. The E node worth 2 and the B node worth 2
C. The brown node worth 2 (R, N) and the B node worth 2
D. All of the above
**Huffman Tree: TO BE OR NOT TO BE**

- **R** = 1
- **N** = 1
- **E** = 2
- **B** = 2
- **T** = 3
- **O** = 4

Which nodes can get combined next?

A. The brown node worth 4 (R, N, E) and the B node worth 2
B. The brown node worth 4 (R, N, E) and the O node worth 4
C. The B node worth 2 and the T node worth 3
D. All of the above

**Huffman Code Properties**

- Prefix code
  - No code is a prefix of another code
  - Example
    - Huffman("dog") ⇒ 01
    - Huffman("cat") ⇒ 011 // not legal prefix code
- Can stop as soon as complete code found
- No need for end-of-code marker
- Nondeterministic
  - Multiple Huffman coding possible for same input
  - If more than two trees with same minimal weight

**TO BE OR NOT TO BE**

How many bits would it take to store this message if every letter was represented with the same number of bits? You should first figure out how many bits it takes to represent 6 different values/letters.

- **N** = 1110
- **R** = 1111
- **E** = 110
- **B** = 01
- **O** = 10
- **T** = 00

A. 26
B. 32
C. 39
D. 48
E. 52

0010.01110.101111.11
101000.0010.01110
32 bits

**Huffman Tree: TO BE OR NOT TO BE**

- **R** = 1
- **N** = 1
- **E** = 2
- **B** = 2
- **T** = 3
- **O** = 4

Which nodes can get combined next?

A. The brown node worth 4 (R, N, E) and the B T node worth 5
B. The brown node worth 4 (R, N, E) and the O node worth 4
C. The B T node worth 5 and the O node worth 4
D. All of the above
No code is prefix of another

DECODING: Your turn

DECODING: Your turn

Steganography.
Digitize everything

What have we seen digital representations for?
- Colors
- Letters
- Logical Calculations
- Numbers

Compression Examples
- Tools
  - winzip, pkzip, compress, gzip
- Formats
  - Images
    - .jpg, .gif
  - Audio
    - .mp3, .wav
  - Video
    - mpeg1 (VCD), mpeg2 (DVD), mpeg4 (Divx)
  - General
    - .zip, .gz

Digitization (B2B Chapter 3)
- Two sides to it: The transformation of the original information can:
  - ADD TO IT
  - REMOVE SOMETHING FROM IT
  - DO BOTH => COMPLETE TRANSFORMATION OF ORIGINAL
- Consequence: WYSI NOT WYG: What You See (hear) on the screen is not What you Get
  - There can be stuff there you think you deleted or you didn’t know was there
  - The stuff that is there may have not be exactly what it looks like
Can You Believe Your Eyes?

Exposing Digital Forgeries in Complex Lighting Environments
M.K. Johnson and H. Farid

Make them do anything

Mixed reality: Mixture of real and animated
- Eagle snatches baby
- http://www.youtube.com/

Step 1: Reduce Bits of Guest
- We don’t need all of the bits in RGB to get a decent picture
Remember how colors work

Pure red is 255

\[
255 - 15 = 240 \\
00000111 - 00000000 = 11111111
\]

128 + 64 = 192

\[
10000000 + 01000000 = 11000000
\]

What bits matter most?

If you had to throw away some bits from each pixel of a message which should you throw away so that the remaining image was as close to the original as possible?

A. The rightmost (low order bits).
B. The leftmost (high order bits).
C. Doesn’t matter, just how many you throw is all that matters.

Step 2: Replace Bits In Host

- Put guest bits into right 2 bits of host

WYSI(not)WYG. PDF files

- Case of journalist Gioliana Sgrena (2005)
Document Metaphor?

- More than meets the eye.
- Nearly to impossible to retract/destroy once shared (or even when not shared)

- Also true for videos:

- Do you know what is really in the documents you share?

Recap – Ghosts in the Machine

- Meta data – what you see is less than what you get
- Steganography – hiding info in plain sight
- Erased/deleted data may still be around
  - On your disk drive
  - In the cloud
  - On your phone

Data That Just Won’t Go Away

- Disk Format
  - files = magazine article (continues on page x)
  - index/TOC
  - deleting just removes entry from index/TOC
- Deleting data in the Cloud?
- Data on your cell phone?
- Once a file is copied, it is hard to totally eradicate

UNCLASSIFIED

(U) At approximately 2030 hours, Major General approached Captain and asked him how he was doing and if Lieutenant Colonel had told him what was going on. Captain said no, but that he suspected it had something to do with the Italian journalist. Major General said “Yes, but it is best if no one knows.” Captain took this as an order from a General Officer not to pass that information on to anyone. (Annex 109C). Moreover, Major General did not intend for Captain to take any action whatsoever on that information. He only told Captain so that he would not be surprised when Ms. Sgrena arrived. (Annex 100C).