CSE 101

Computing Concepts and Competencies

Class Day 4 – Data Representation: Text, Image, Sound, Video

CSE 101
Day 4 Overview

- Homework Debrief
- Representing Text
- Representing Images
- Representing Sound
- Representing Video

Homework Debrief

- While you should go to helproom to discuss homework details, here are few major points.

- Excel Exercises
  - Exercise 9 – units spreadsheet - what were user mistakes in that sheet and what were system mistakes in that sheet?
  - Exercise 12 – what formula did you use to calculate highest unsigned integer

- Question 5 and tricksters
  - Part C – what happens with .2
  - Part D – what problem did you have with this part

Data Representation

- Recall that numbers larger than 1, numbers with fractional parts, or negative numbers require groupings of bits or binary digits

- Other types of data such as text characters, images, video, etc. are also represented in groupings of bits

Before we proceed, can anyone tell me what data the following number represents:

1001000

Data Representation
Text Characters

- Numerous character sets (and variations) have come and gone, but the computing industry seems to have settled on two common systems:
  - ASCII (American Standard Code for Information Interchange) –
    - typically a 7-bit representation system
    - extended ASCII uses an 8-bit representation
  - Unicode
    - most often found in the 16-bit version
    - or more recently with 32-bit variations

Given the above, how many distinct characters can you represent in ASCII? How many in Unicode (16-bits)?

Data Representation
Character Lookup Tables

- Bitstreams are matched to their character equivalent through look-up tables for the specific character set
  - This is an ASCII table

- Looking up
  - Example: 1001000

xxxyyyy		xxx							
		000	001	010	011	100	101	110	111
	0000	NUL	DLE	space	0	@	P	`	p
	0001	SOH	DC1	!	1	A	Q	a	q
	0010	STX	DC2	"	2	B	R	b	r
	0011	ETX	DC3	#	3	C	S	c	s
	0100	EOT	DC4	$	4	D	T	d	t
	0101	ENQ	NAK	%	5	E	U	e	u
y	0110	ACK	SYN	&	6	F	V	f	v
y	0111	BEL	ETB	'	7	G	W	g	w
y	1000	BS	CAN	(	8	H	X	h	x
y	1001	TAB	EM	)	9	I	Y	i	y
	1010	LF	SUB	*	:	J	Z	j	z
	1011	VT	ESQ	+	;	K	[	k	{
	1100	FF	FS	,	<	L	\	l	|
	1101	CR	GS	-	=	M	]	m	}
	1110	SO	RS	.	>	N	^	n	~
	1111	SI	US	/	?	O	_	o	DEL

Data Representation
Character Lookup Tables

- Let’s do a reverse lookup – find the binary equivalent for the letter “f”

- Looking up
  - Find “f” in the table
  - Find first 3 digits - 110
  - Find last 4 digits - 0110
  - 1100110

xxxyyyy		xxx							
		000	001	010	011	100	101	110	111
	0000	NUL	DLE	space	0	@	P	`	p
	0001	SOH	DC1	!	1	A	Q	a	q
	0010	STX	DC2	"	2	B	R	b	r
	0011	ETX	DC3	#	3	C	S	c	s
	0100	EOT	DC4	$	4	D	T	d	t
	0101	ENQ	NAK	%	5	E	U	e	u
y	0110	ACK	SYN	&	6	F	V	f	v
y	0111	BEL	ETB	'	7	G	W	g	w
y	1000	BS	CAN	(	8	H	X	h	x
y	1001	TAB	EM	)	9	I	Y	i	y
	1010	LF	SUB	*	:	J	Z	j	z
	1011	VT	ESQ	+	;	K	[	k	{
	1100	FF	FS	,	<	L	\	l	|
	1101	CR	GS	-	=	M	]	m	}
	1110	SO	RS	.	>	N	^	n	~
	1111	SI	US	/	?	O	_	o	DEL

Data Representation Conversion Exercise

- Convert the following to their corresponding values:
  - Binary → Character

    1000011111001110001
    0101000000011000110
    011110110001
  - Character → Binary

    goMSU

xxxyyyy		xxx							
		000	001	010	011	100	101	110	111
	0000	NUL	DLE	space	0	@	P	`	p
	0001	SOH	DC1	!	1	A	Q	a	q
	0010	STX	DC2	"	2	B	R	b	r
	0011	ETX	DC3	#	3	C	S	c	s
	0100	EOT	DC4	$	4	D	T	d	t
	0101	ENQ	NAK	%	5	E	U	e	u
y	0110	ACK	SYN	&	6	F	V	f	v
y	0111	BEL	ETB	'	7	G	W	g	w
y	1000	BS	CAN	(	8	H	X	h	x
y	1001	TAB	EM	)	9	I	Y	i	y
	1010	LF	SUB	*	:	J	Z	j	z
	1011	VT	ESQ	+	;	K	[	k	{
	1100	FF	FS	,	<	L	\	l	|
	1101	CR	GS	-	=	M	]	m	}
	1110	SO	RS	.	>	N	^	n	~
	1111	SI	US	/	?	O	_	o	DEL

Data Representation
Unicode / Chinese Character Set

	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0100111000000000	一	丁	丂	七	丄	丅	丆	万	丈	三	上	下	丌	不	与	丏
0100111000010000	丐	丑	丒	专	且	丕	世	丗	丘	丙	业	丛	东	丝	丞	丟
0100111000100000	丠	両	丢	丣	两	严	並	丧	丨	丩	个	丫	丬	中	丮	丯
0100111000110000	丰	丱	串	丳	临	丵	丶	丷	丸	丹	为	主	丼	丽	举	丿
0100111001000000	乀	乁	乂	乃	乄	久	乆	乇	么	义	乊	之	乌	乍	乎	乏
0100111001010000	乐	乑	乒	乓	乔	乕	乖	乗	乘	乙	乚	乛	乜	九	乞	也
0100111001100000	习	乡	乢	乣	乤	乥	书	乧	乨	乩	乪	乫	乬	乭	乮	乯
0100111001110000	买	乱	乲	乳	乴	乵	乶	乷	乸	乹	乺	乻	乼	乽	乾	乿
0100111010000000	亀	亁	亂	亃	亄	亅	了	亇	予	争	亊	事	二	亍	于	亏
0100111010010000	亐	云	互	亓	五	井	亖	亗	亘	亙	亚	些	亜	亝	亞	亟
0100111010100000	亠	亡	亢	亣	交	亥	亦	产	亨	亩	亪	享	京	亭	亮	亯
0100111010110000	亰	亱	亲	亳	亴	亵	亶	亷	亸	亹	人	亻	亼	亽	亾	亿
0100111011000000	什	仁	仂	仃	仄	仅	仆	仇	仈	仉	今	介	仌	仍	从	仏
0100111011010000	仐	仑	仒	仓	仔	仕	他	仗	付	仙	仚	仛	仜	仝	仞	仟
0100111011100000	仠	仡	仢	代	令	以	仦	仧	仨	仩	仪	仫	们	仭	仮	仯
0100111011110000	仰	仱	仲	仳	仴	仵	件	价	仸	仹	仺	任	仼	份	仾	仿

Lessons Learned

- Different character sets exist
  - implemented as look-up tables
  - no calculations – just find the value
- The more bits used, the more characters could be represented
  - slows down processing
  - allows to represent languages with more characters
  - allows to represent multiple languages

Images

- A photograph or drawing can be digitized by treating it as a series of colored dots (pixels)
  - Find a picture of La Grande Jatte by George Seurat
  - Each dot has a “position” designation
  - Each dot is assigned a binary number according to its color, e.g. white could be represented as 1111, green as 0100, etc.
  - Colors are translated to/from a bitstream via a lookup table like we saw with character lookups
  - A digital image is simply a list of color numbers for all the dots it contains

Bitmaps

In this example, the table consists of 16 colors only:

Black	**0000**	**Pink**	**1000**
Gray	**0001**	**Red**	**1001**
Silver	**0010**	**Yellow**	**1010**
Lime	**0011**	**Teal**	**1011**
Green	**0100**	**Aqua**	**1100**
Olive	**0101**	**Blue**	**1101**
Purple	**0110**	**Navy**	**1110**
Maroon	**0111**	**White**	**1111**

Parsons, June Jamrich, and Oja Dan, *Computer Concepts 2012*, Boston: Course Technology. 2012: 25

Exercise

- Copy the Excel file (pixelExample.xlsx) from today's course folder to your cse101 folder
- This sheet contains an image with cells containing color codes
- Use these color codes to put actual colors in the respective cells

Black	0000	Pink	1000
Gray	0001	Red	1001
Silver	0010	Yellow	1010
Lime	0011	Teal	1011
Green	0100	Aqua	1100
Olive	0101	Blue	1101
Purple	0110	Navy	1110
Maroon	0111	White	1111

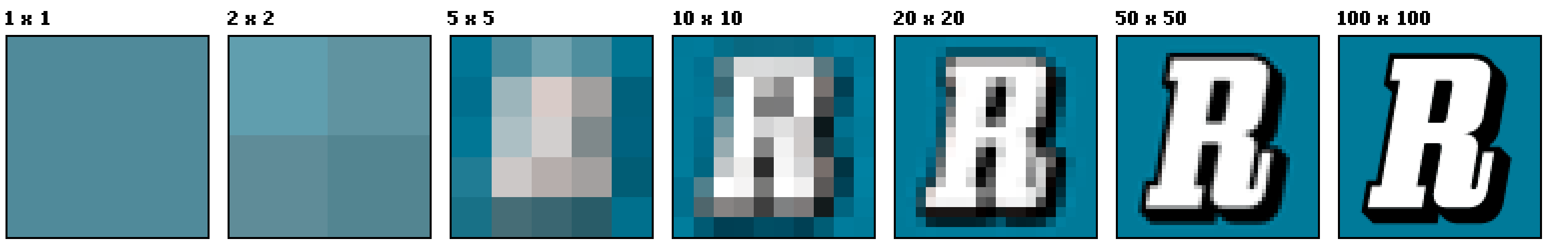
Image Representation: Resolution and Depth

- Resolution is the number of pixels per image
  - The more pixels, the better the resolution (and the larger image file size)

- Color depth is the number of bits used to represent a color
  - The more colors, the more realistic rendering (and the larger image files size)

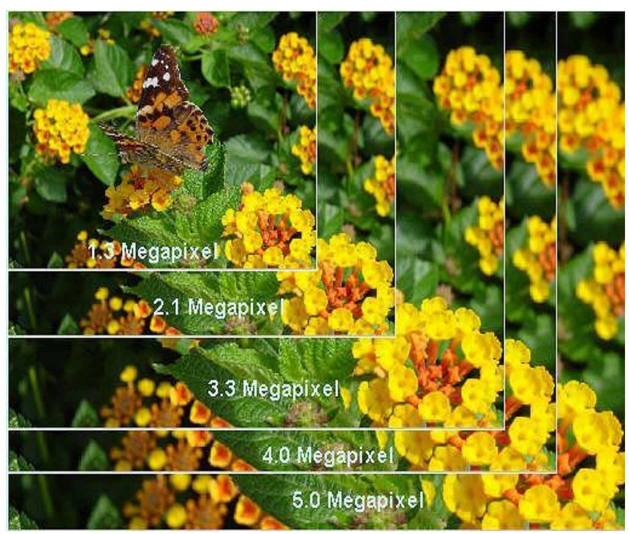
Examples of Image Resolution

Increasing resolution (more pixels), same size ➔ Better picture



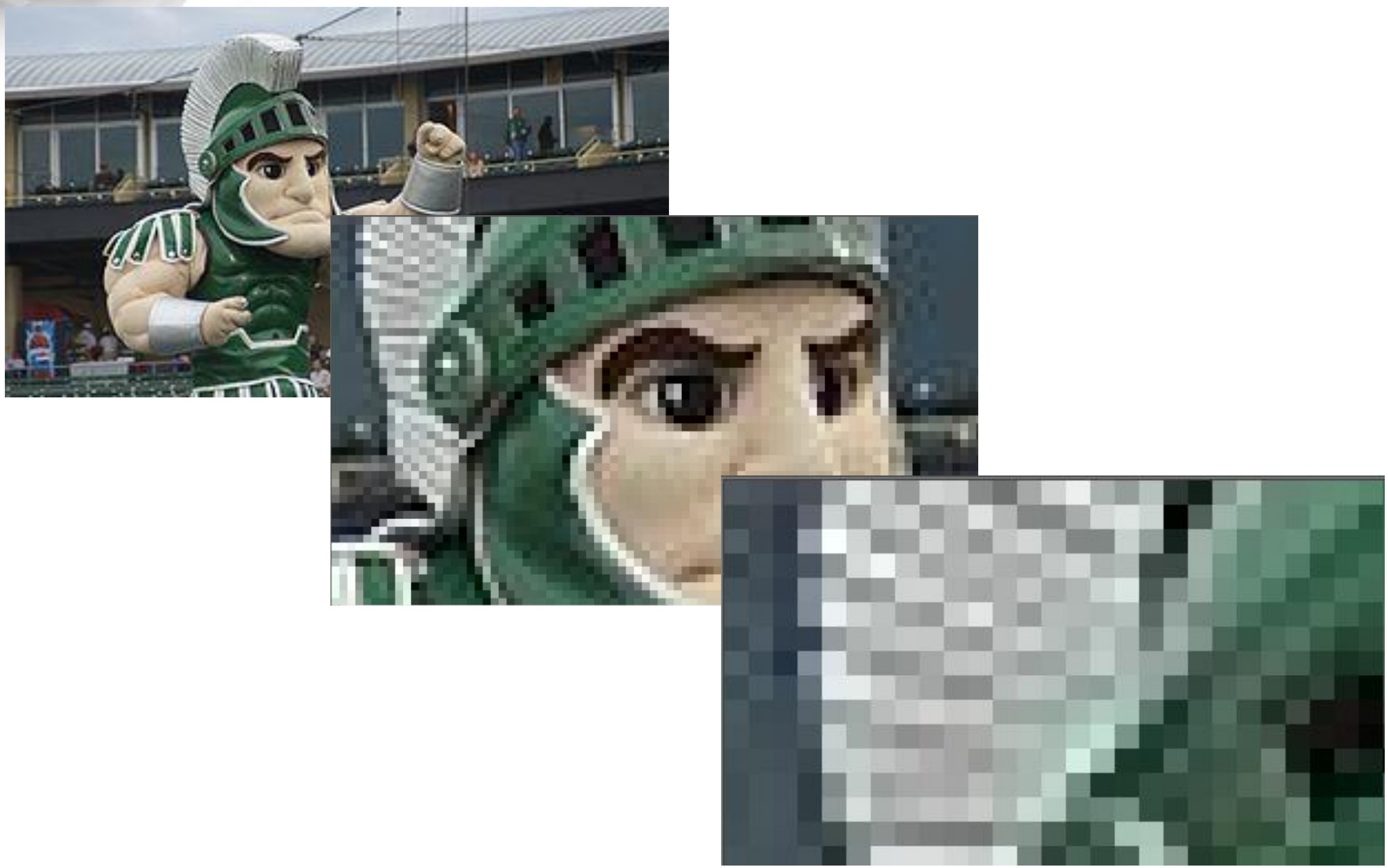
http://en.wikipedia.org/wiki/Image_resolution#Pixel_resolution

Increasing resolution, same picture quality➔ Larger picture size



http://www.vividlight.com/articles/3116.htm

Illustration of “Pixelation”



Color Depth

- Color depth is an additional factor in how good a picture looks and pertains to how many colors an image takes advantage of.

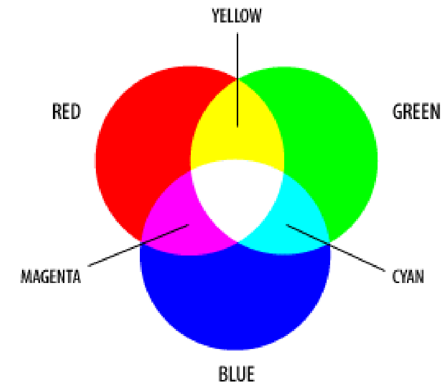
- Let’s look at the following reference:

  http://en.wikipedia.org/wiki/Color_depth

  – What do you see in pictures having a low color depth?

More about Pixels: Color Palette

- The codes used to render color are based on the computer version of primary colors – different intensity of each of these hues:
  - Red
  - Green
  - Blue



http://dba.med.sc.edu/price/irf/Adobe_tg/models/rgbcmy.html

Example

www.lunapic.com

boatSun.jpg

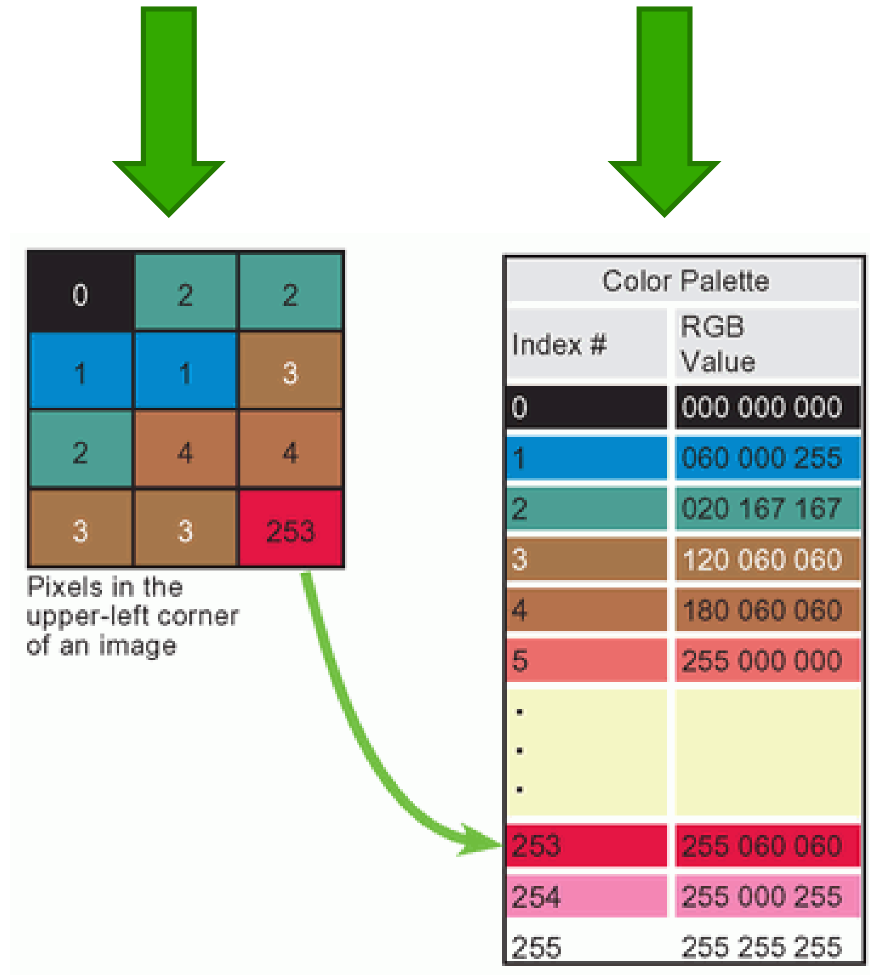
Reducing Size

- If we stored all pictures as bitmaps, our images would be … HUGE
  - height x width x depth

- Consider our smiley –
  - given that that picture consists of 10 rows and 9 columns
  - and that I would use 4 bits to represent color in each pixel in our file
  - what would be the number of bits that I need to represent smiley?
  - what if I change the number of bits to 24 because I want to use more colors per pixel than just 16?

Color Tables

- Color tables reduce the size of a picture by treating the picture as consisting of two entities
  - The color table which is the list of colors actually used in the picture
  - The picture itself, except each pixel instead of containing the full color code, contains a short index to the color table



Parsons, June Jamrich, and Oja Dan, *Computer Concepts 2012*, Boston: Course Technology. 2012: 437

How does it help?

- Full bitmap version of smiley (24 bits per pixel), we need
  - 10 rows x 9 columns x 24 bits = 2,160 bits

  since every 24-bit color needs to be saved for each pixel, the same color ends up stored multiple times (e.g. blue is saved twice, red 7x, yellow…)

- With color table approach, we store short indices for pixels, while the colors themselves are only stored once

How does it help?

10	10	10	10	10	10	10	10	10
10	10	10	10	10	10	10	10	10
10	10	01	10	10	10	01	10	10
10	10	10	10	10	10	10	10	10
10	10	10	10	10	10	10	10	10
10	11	10	10	10	10	10	11	10
10	10	11	10	10	10	11	10	10
10	10	10	11	11	11	10	10	10
10	10	10	10	10	10	10	10	10
10	10	10	10	10	10	10	10	10

color table

index	color	(color is still represented using RGB, meaning 24 bits per pixel)
10	yellow	(11111111, 11111111, 00000000)
11	red	(11111111, 00000000, 00000000)
01	blue	(00000000, 00000000, 11111111)

- After the table approach

  10 rows x 9 columns x 2 bits (for color index) = 180 bits

  + table size (3 colors x size of each color)

  size of each color = 2 bits for its index + 24 for the color itself

  180 + 78 = 258 bits

Sampling

- Instead of saving every pixel, save only some of them (according to some scheme – algorithm)
  - Split pixels into color and luminance
    - luminance – our perception of the amount of light reflected off a surface – which is not based on the actual amount of light, but rather on our comparison with the surface’s surroundings
  - Process the image in blocks (e.g. 8 x 8) based on their luminance and lose some of the color details

Lessons Learned

- File formats:
  - BMPs
    - all information pixel by pixel
    - large size
  - GIFs (and PNGs)
    - all information stored using color tables – does not lose details (lossless compression)
    - may become problematic when > 256 color used – one either need to adjust colors to 256 losing details or increased color table size may lead to size inflation
  - JPEGs
    - lower quality (lossy compression)
    - files created based on computation so the image loses quality even further each time it is processed

Image Comprehension

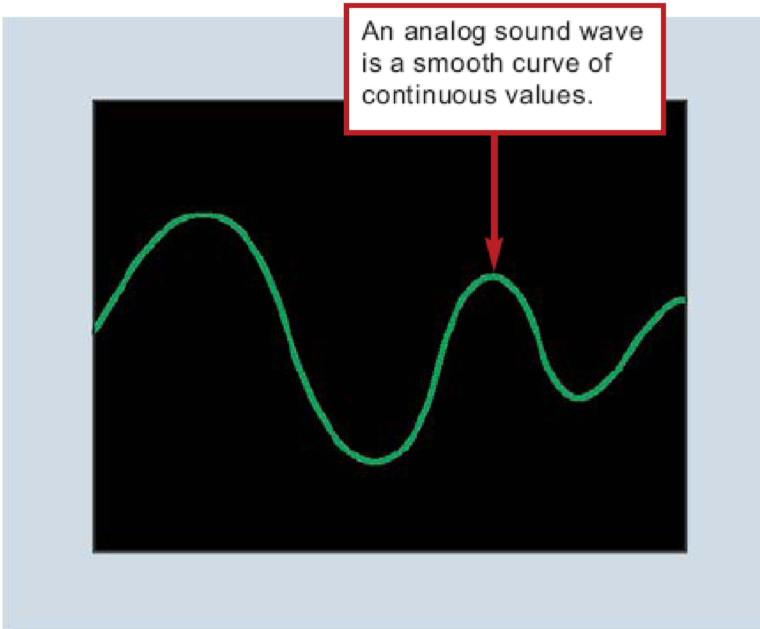
Computers can render an image provided they know the encoding scheme but do they understand what the image itself represents?

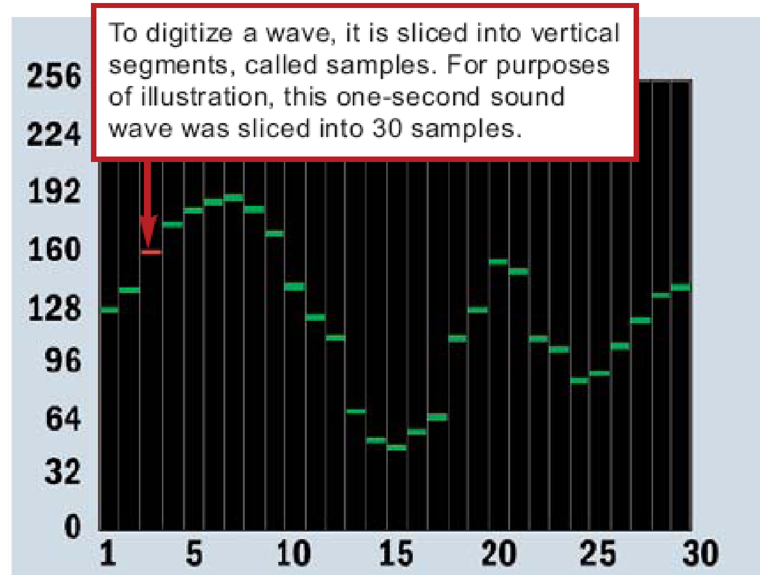
Computing Beyond Turing - Jeff Hawkins
http://www.youtube.com/watch?v=cCdbZqI1r7I
(minutes 13:02– 13:57)

Human Computation
http://www.youtube.com/watch?v=tx082gDwGcM
(minutes 1:36- 4:05)

Sound

- Sound – the wave of air produced by the vibration of matter
  - to represent sound we have to have a way of representing this wave
  - waves are analog, while computers are digital – measure the wave periodically (sampling) and record the measure as the binary number according to some agreed convention
  - again a computer uses a special table to store this information





wav (uncompressed), mp3 (compressed)

Sample	Sample Height (Decimal)	Sample Height (Binary)
1	130	10000010
2	140	1000110
3	160	10100000
4	175	10101111
5	185	10111001

The height of each sample is converted into a binary number and stored. The height of sample 3 is 160 (decimal), so it is stored as its binary equivalent—10100000.

Parsons, June Jamrich, and Oja Dan, *Computer Concepts 2012*, Boston: Course Technology. 2012: 422

Videos

- Illusion of movement – a series of still frames per second
  - 8-10 for old mechanical cameras
  - 24 for feature films
  - 30 for digital cameras

http://www.youtube.com/watch?v=-l-pLqBD58E

Videos

- Considerations – image perspective:
  - The size of the window
  - Frame rate
  - Compression
    - temporal compression - change only the information that changes from one frame to the next
    - spatial compression – remove redundant information within one frame (image compression)

Lessons Learned

- The ingenuity of a human mind required to create representational schemes for different types of media
- The power and limits of data representation
  - quality and accuracy vs size and speed
  - machine as the means to represent and transfer knowledge among humans

Homework - Reading

- Check course folder for today's homework/reading assignments
  - Current assessment policies document posted. You must understand these polcies before your quiz.
- Quiz Reminder:
  - M 9/15 or T 9/16 in your class section (no remote)
  - Covers days 01 through 03
  - Be sure to follow quiz ground rules
  - You should have all classwork and homework files with you in case any are needed

It's the end of class … do you know where your flash drive is? Don't forget to take it with you!