CSCI 497P/597P: Computer Vision

	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT
80	102	115	127	119	118	110	111	107	110	111	115	117	120	124	122	119	120	120
81	102	115	126	118	117	109	110	106	110	111	118	119	122	126	124	121	122	124
82	101	105	112	103	108	108	109	110	107	108	113	115	118	121	117	115	113	103
83	101	105	114	105	112	112	113	116	115	116	121	123	126	129	125	123	123	114
84	99	103	113	104	111	111	112	116	117	118	123	125	128	131	127	125	125	118
85	118	123	113	113	109	116	118	120	120	120	120	121	124	125	123	121	117	113
86	120	125	115	115	113	122	124	128	128	128	128	129	129	130	131	129	125	123
87	119	124	114	114	114	122	124	130	130	130	130	131	132	133	133	131	127	125
88	112	113	112	111	115	128	130	126	127	123	122	124	125	124	121	117	117	116
89	114	115	114	115	121	134	135	134	135	131	130	132	130	129	129	125	125	126
90	113	114	113	114	121	134	138	136	137	133	132	134	133	132	131	127	127	128
91	112	110	104	114	129	135	139	138	131	129	126	123	123	123	122	121	123	119
92	114	112	108	118	135	140	144	143	139	137	134	131	131	131	130	129	128	127
93	113	111	107	117	135	143	147	146	141	139	136	133	133	133	132	131	131	130
94	119	111	116	131	138	145	144	137	130	128	125	123	116	119	117	118	122	120
95	121	115	120	135	143	150	149	142	138	136	133	131	124	127	125	126	130	128
96	120	114	119	136	146	153	152	145	140	138	135	133	126	129	127	128	132	131
97	117	112	131	141	147	149	141	137	133	131	126	121	122	115	113	118	123	123
98	121	118	137	146	152	154	149	145	141	139	136	131	132	125	123	128	133	134
99	120	116	137	149	155	157	151	147	143	141	137	132	131	124	122	127	134	136
100	100	124	143	145	146	147	142	138	134	130	124	124	126	118	115	118	115	110
101	106	130	148	150	151	152	150	146	142	140	134	134	136	130	127	130	126	121
102	104	130	151	153	154	155	152	148	144	141	135	135	135	128	125	130	128	125
103	95	132	145	145	147	146	140	134	132	130	126	123	116	110	109	113	105	108
104	104	141	153	153	153	152	145	142	140	140	136	135	128	122	121	124	119	121
105	103	140	155	155	153	152	148	144	142	141	135	133	126	120	121	126	122	127
106	109	131	140	145	144	143	142	134	134	129	120	109	106	106	112	108	106	106

Lecture 1: Course Overview; Images

Norms

Goals Slides

- A slide like this will appear at the beginning of each lecture.
- This is my attempt to be transparent about what I want you to get out of the lecture.
- This can be helpful for you (when studying for exams) as well as me (when writing exams).
- When I introduce key terminology, I (try to) highlight it in blue like this.

Goals

- Understand how vision tasks live on a spectrum from "low level" to "high level" vision.
- Appreciate why computer vision is hard.
- Understand how images are represented:
  - On a computer
  - In math
- Be able to write down mathematical image transformations that perform simple photometric or geometric manipulations of images functions.
- Know what is meant by image noise.

Course Overview

What does this course cover?

Visual Perception

What can you tell me about this image? or,
What might you want to know about this image?

send suggestions to #general

Levels of Vision

What can you tell me about this image? or,
What might you want to know about this image?

What would a noise-free image look like?
Where are the edges?
Where are there straight lines?
Which patches of this image are distinctive?
How would the unblurred background appear?
Is this the same scene as another image?
How far from the camera is each point?
What is the 3D shape of the subject?
Is the image level? Which way is up?
Which groups of pixels "belong" together?
Which pixels belong to the subject?
What is the subject?
What breed of dog is the subject?
What is the subject's emotional state?
Is the subject a Very Good Boi?

Levels of Vision

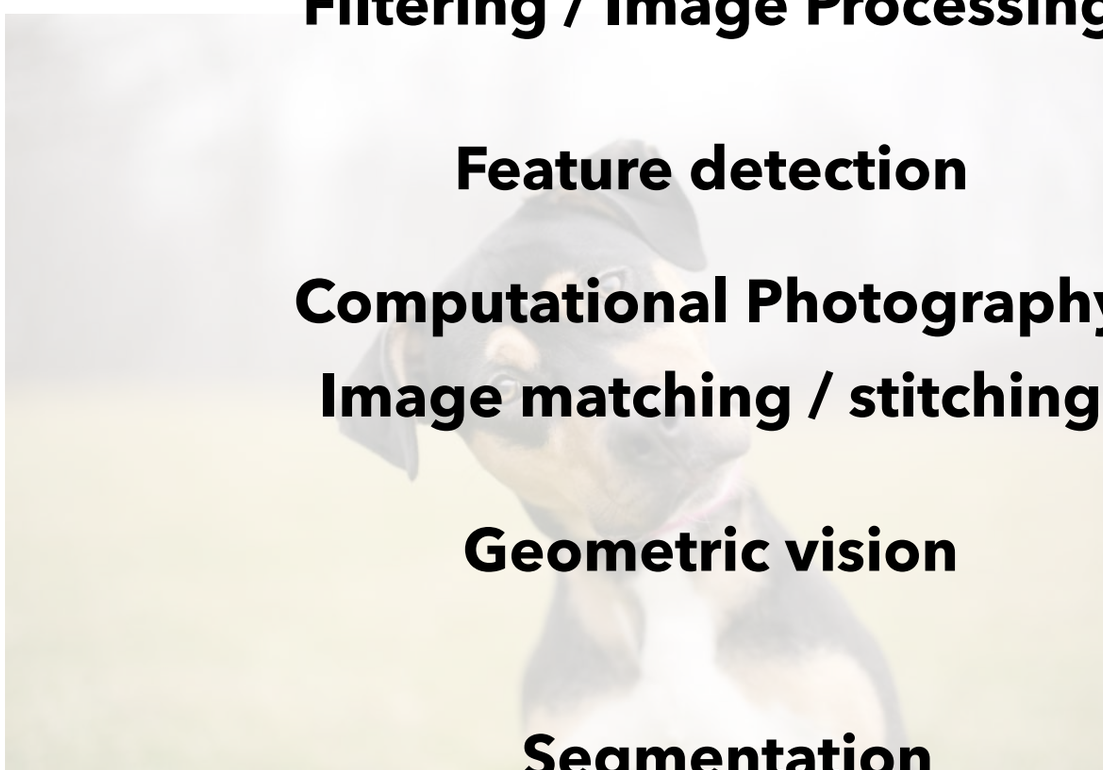
What can you tell me about this image? or,
What might you want to know about this image?

"low-level"

What would a noise-free image look like?
Where are the edges?
Where are there straight lines?
Which patches of this image are distinctive?
How would the unblurred background appear?
Is this the same scene as another image?
How far from the camera is each point?
What is the 3D shape of the subject?
Is the image level? Which way is up?
Which groups of pixels "belong" together?
Which pixels belong to the subject?
What is the subject?
What breed of dog is the subject?
What is the subject's emotional state?
Is the subject a Very Good Boi?

"high-level"

Levels of Vision



Subfield/topic	*Example(s)*
	"low-level"
Filtering / Image Processing	What would a noise-free image look like? Where are the edges?
Feature detection	Where are there straight lines? Which patches of this image are distinctive?
Computational Photography	How would the unblurred background appear?
Image matching / stitching	Is this the same scene as another image?
Geometric vision	How far from the camera is each point? What is the 3D shape of the subject? Is the image level? Which way is up?
Segmentation	Which groups of pixels "belong" together? Which pixels belong to the subject?
Semantic understanding	What is the subject? What breed of dog is the subject? What is the dog's emotional state?
	"high-level"

In this course...

"low-level"

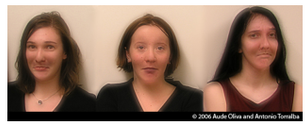
What would a noise-free image look like?
Where are the edges?
Where are there straight lines?
Which patches of this image are distinctive?
How would the unblurred background appear?
Is this the same scene as another image?
How far from the camera is each point?
What is the 3D shape of the subject?
Is the image level? Which way is up?
Which groups of pixels "belong" together?
Which pixels belong to the subject?
What is the subject?
What breed of dog is the subject?
What is the dog's emotional state?

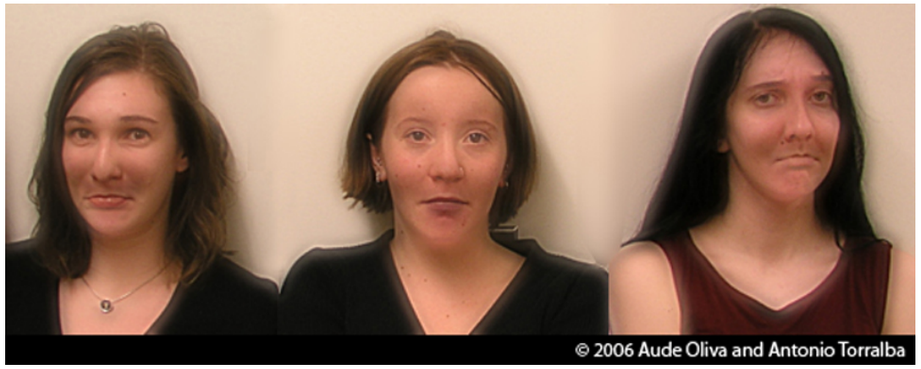
"high-level"

Course Overview

5ish major topics:

- Image filtering





- Feature detection
- Panorama stitching
- Depth estimation via stereo vision
- Image recognition via deep learning

"low-level"

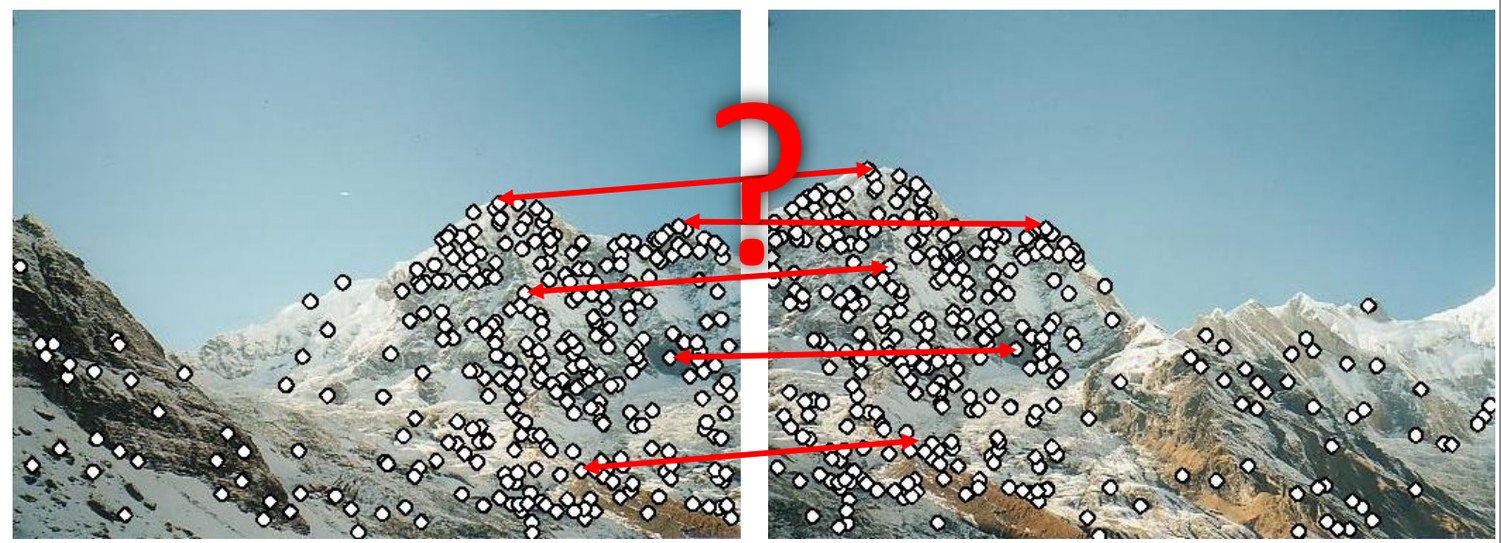
"high-level"

Course Overview

5ish major topics:

"low-level"

- Image filtering
- Feature detection
- Panorama stitching



- Depth estimation via stereo vision
- Image recognition via deep learning

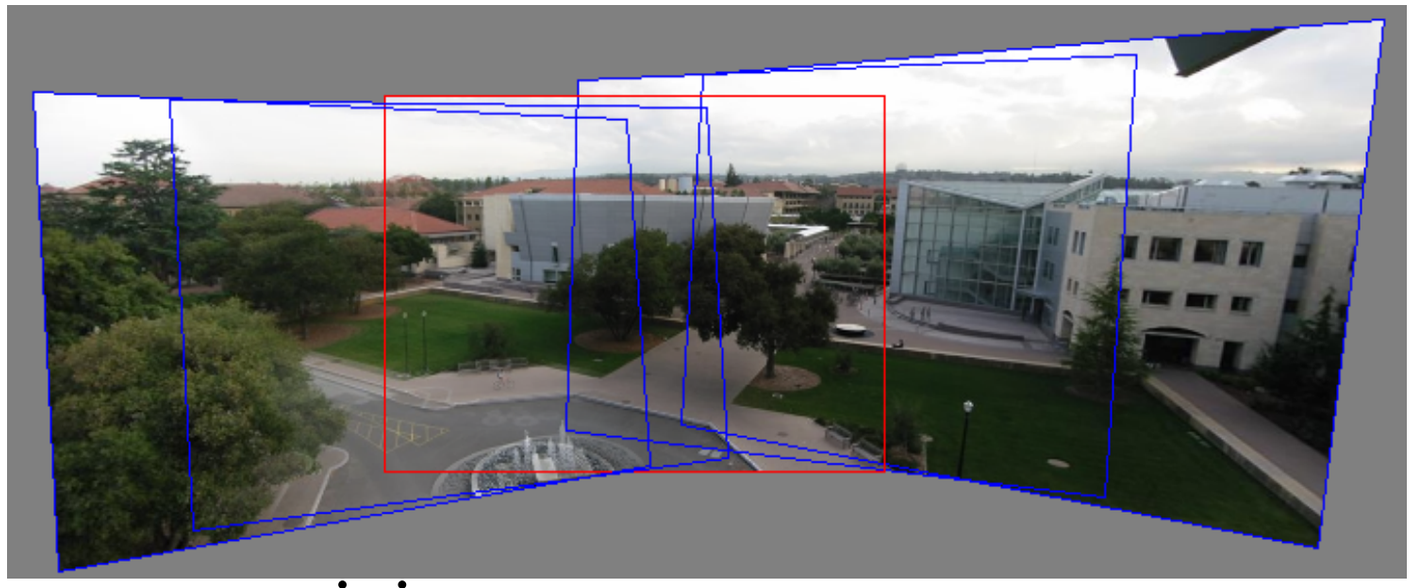
"high-level"

Course Overview

5ish major topics:

"low-level"

- Image filtering
- Feature detection
- Panorama stitching



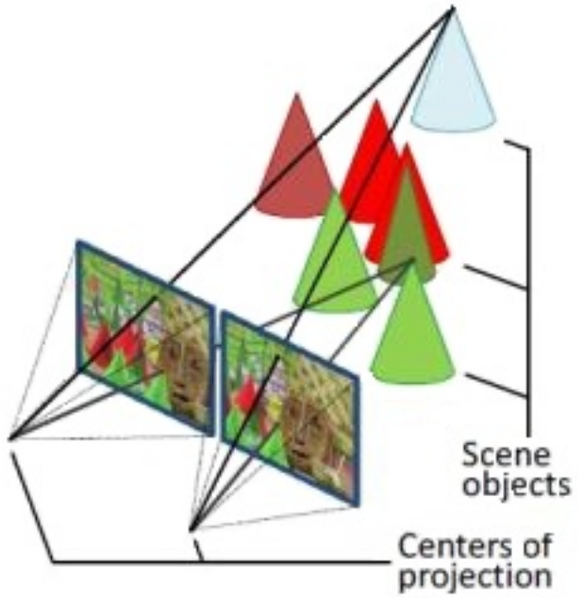
- Depth estimation via stereo vision
- Image recognition via deep learning

"high-level"

Course Overview

5ish major topics:

- Image filtering
- Feature detection
- Panorama stitching
- Depth estimation via stereo vision
- Image recognition via deep learning



"low-level"

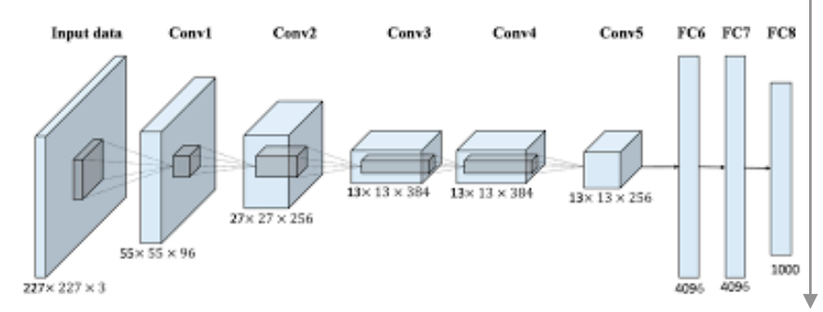
"high-level"

Course Overview

5ish major topics:

- Image filtering
- Feature detection
- Panorama stitching
- Depth estimation via stereo vision
- Image recognition via deep learning

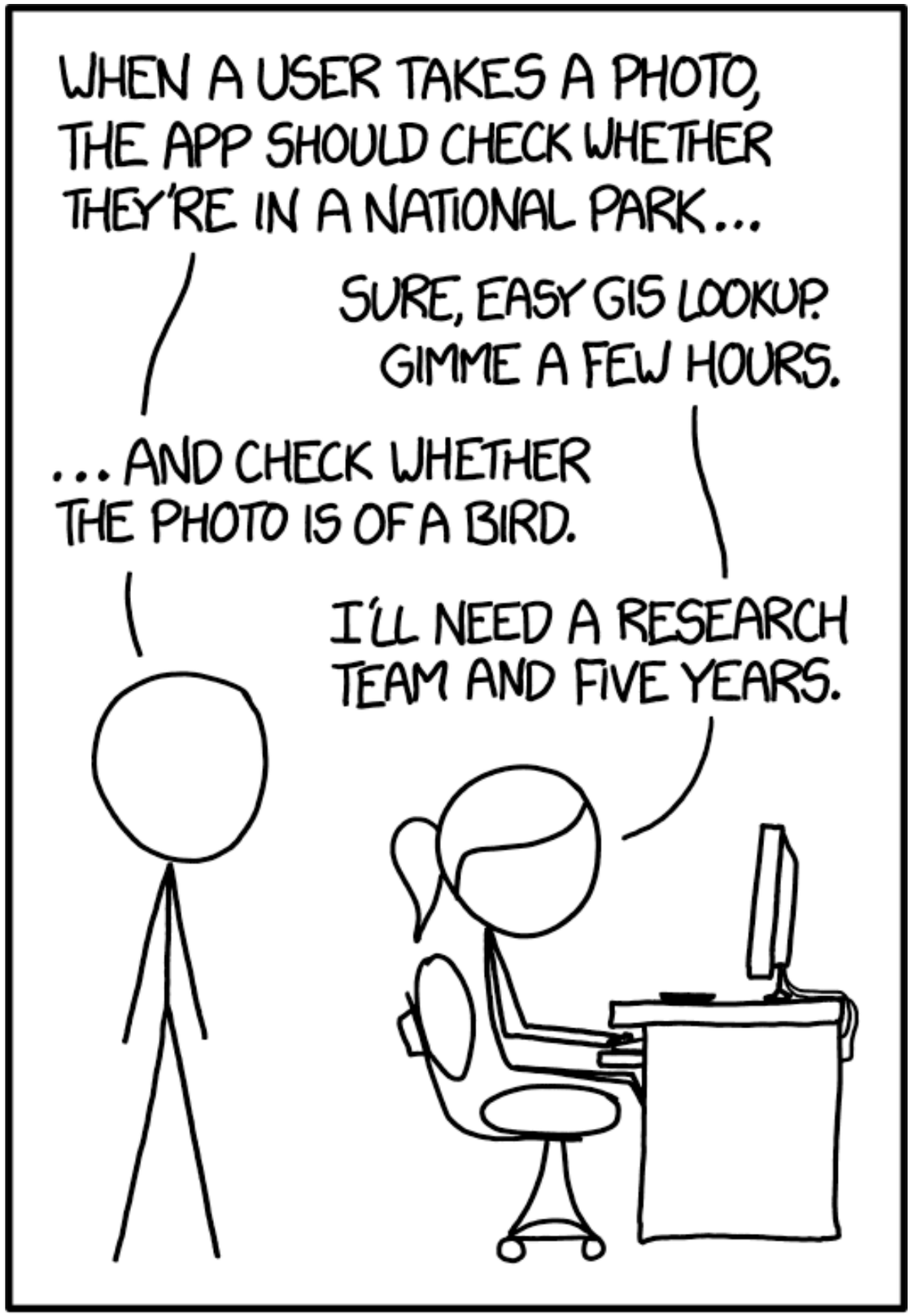
"low-level"



"high-level"

Vision is hard?

What would a noise-free image look like?
Where are the edges?
Where are there straight lines?
Which patches of this image are distinctive?
How would the unblurred background appear?
Is this the same scene as another image?
How far from the camera is each point?
What is the 3D shape of the subject?
Is the image level? Which way is up?
Which groups of pixels "belong" together?
Which pixels belong to the subject?
What is the subject?
What breed of dog is the subject?
What is the dog's emotional state?



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

xkcd

9/24/2014

Posted on October 20, 2014 by Rob Hess, Clayton Mellina, and Friends

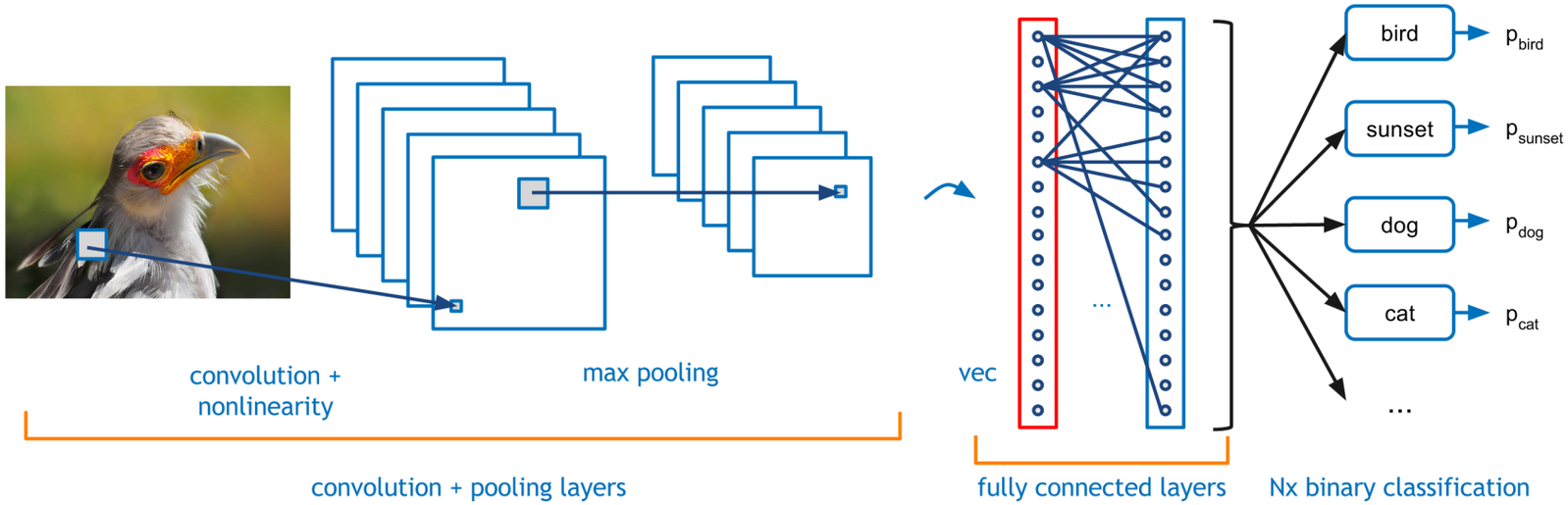
← Previous Next →

Introducing: Flickr PARK or BIRD

Zion National Park Utah by Les Haines (cc) BY

OR

Secretary Bird by Bill Gracey (cc) BY-NC-ND



flickr

10/20/2014

Why is vision hard?

Why is vision hard?

Viewpoint variation

Why is vision hard?

Viewpoint variation

Scale

Why is vision hard?

Viewpoint variation

Illumination

Scale

Why is vision hard?

Why is vision hard?

Intra-class variation

Why is vision hard?

Intra-class variation

Motion (Source: S. Lazebnik)

Why is vision hard?

Intra-class variation

Motion (Source: S. Lazebnik)

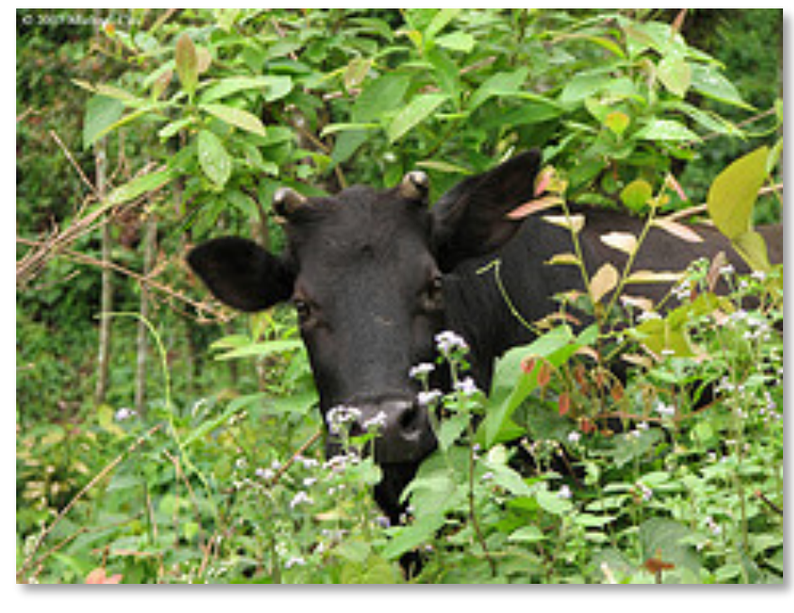
Occlusion

Why is vision hard?

Intra-class variation

Motion (Source: S. Lazebnik)

Background clutter



Occlusion

Why is vision hard?

The state of Computer Vision and AI: we are really, really far.

Oct 22, 2012

The picture above is funny.

But for me it is also one of those examples that make me sad about the outlook for AI and for Computer Vision. What would it take for a computer to understand this image as you or I do? I challenge you to think explicitly of all the pieces of knowledge that have to fall in place for it to make sense. Here is my short attempt:

- You recognize it is an image of a bunch of people and you understand they are in a hallway
- You recognize that there are 3 mirrors in the scene so some of those people are "fake" replicas from different viewpoints.
- You recognize Obama from the few pixels that make up his face. It helps that he is in his suit and that he is surrounded by other people with suits.
- You recognize that there's a person standing on a scale, even though the scale occupies only very few white pixels that blend with the background. But, you've used the person's pose and knowledge of how people interact with objects to figure it out.
- You recognize that Obama has his foot positioned just slightly on top of the scale. Notice the language I'm using: It is in terms of the 3D structure of the scene, not the position of the leg in the 2D coordinate system of the image.
- You know how physics works: Obama is leaning in on the scale, which applies a force on it. Scale measures force that is applied on it, that's how it works => it will over-estimate the weight of the person standing on it.
- The person measuring his weight is not aware of Obama doing this. You derive this because you know his pose, you understand that the field of view of a person is finite, and you understand that he is not very likely to sense the slight push of Obama's foot.
- You understand that people are self-conscious about their weight. You also understand that he is reading off the scale measurement, and that shortly the over-estimated weight will confuse him because it will probably be much higher than what he expects. In other words, you reason about implications of the events that are about to unfold seconds after this photo was taken, and especially about the thoughts and how they will develop inside people's heads. You also reason about what pieces of information are available to people.
- There are people in the back who find the person's imminent confusion funny. In other words you are reasoning about state of mind of people, and their view of the state of mind of another person. That's getting frighteningly meta.
- Finally, the fact that the perpetrator here is the president makes it maybe even a little more funnier. You understand what actions are more or less likely to be undertaken by different people based on their status and identity.

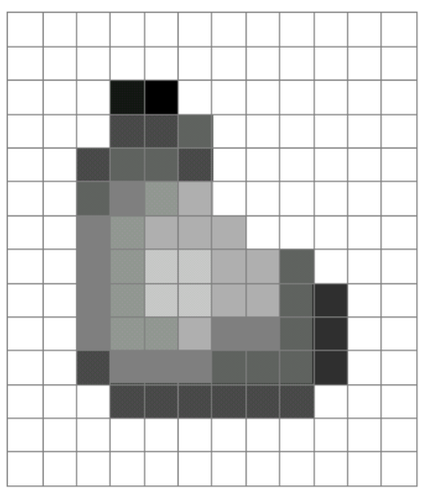
What is an image?

Computationally speaking...

A grayscale image is a 2D array of numbers.

What is an image?

Computationally speaking...



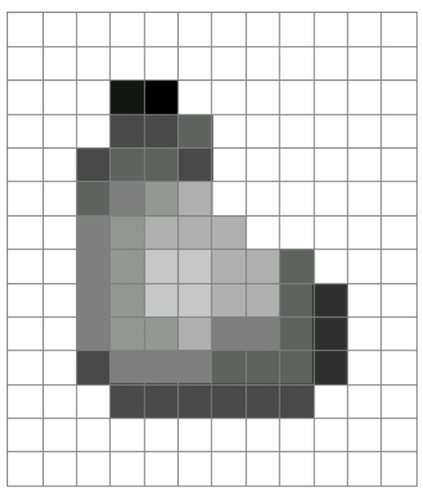
=

255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	20	0	255	255	255	255	255	255	255
255	255	255	75	75	75	255	255	255	255	255	255
255	255	75	95	95	75	255	255	255	255	255	255
255	255	96	127	145	175	255	255	255	255	255	255
255	255	127	145	175	175	175	255	255	255	255	255
255	255	127	145	200	200	175	175	95	255	255	255
255	255	127	145	200	200	175	175	95	47	255	255
255	255	127	145	145	175	127	127	95	47	255	255
255	255	74	127	127	127	95	95	95	47	255	255
255	255	255	74	74	74	74	74	74	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255

A grayscale image is a 2D array of numbers.

What is an image?

Computationally speaking...



=

255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	20	0	255	255	255	255	255	255	255
255	255	255	75	75	75	255	255	255	255	255	255
255	255	75	95	95	75	255	255	255	255	255	255
255	255	96	127	145	175	255	255	255	255	255	255
255	255	127	145	175	175	175	255	255	255	255	255
255	255	127	145	200	200	175	175	95	255	255	255
255	255	127	145	200	200	175	175	95	47	255	255
255	255	127	145	145	175	127	127	95	47	255	255
255	255	74	127	127	127	95	95	95	47	255	255
255	255	255	74	74	74	74	74	74	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255

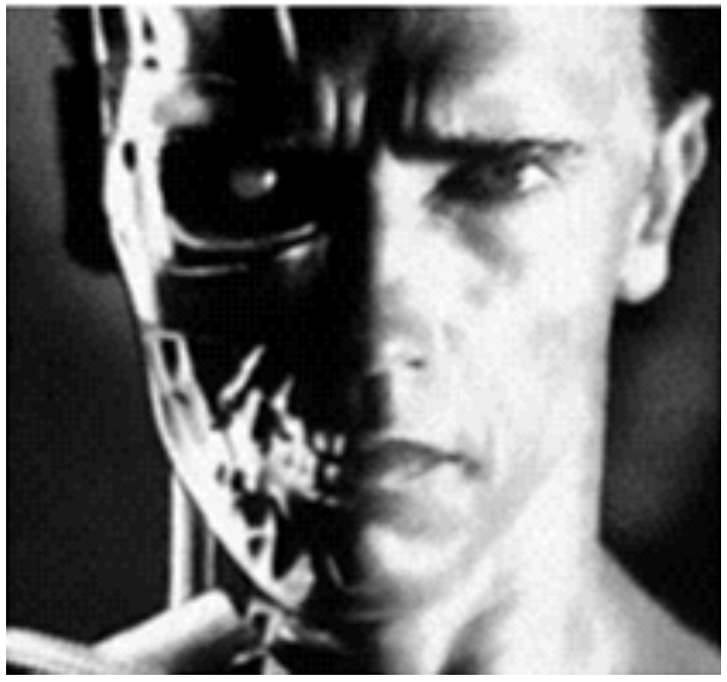
usually one byte per pixel

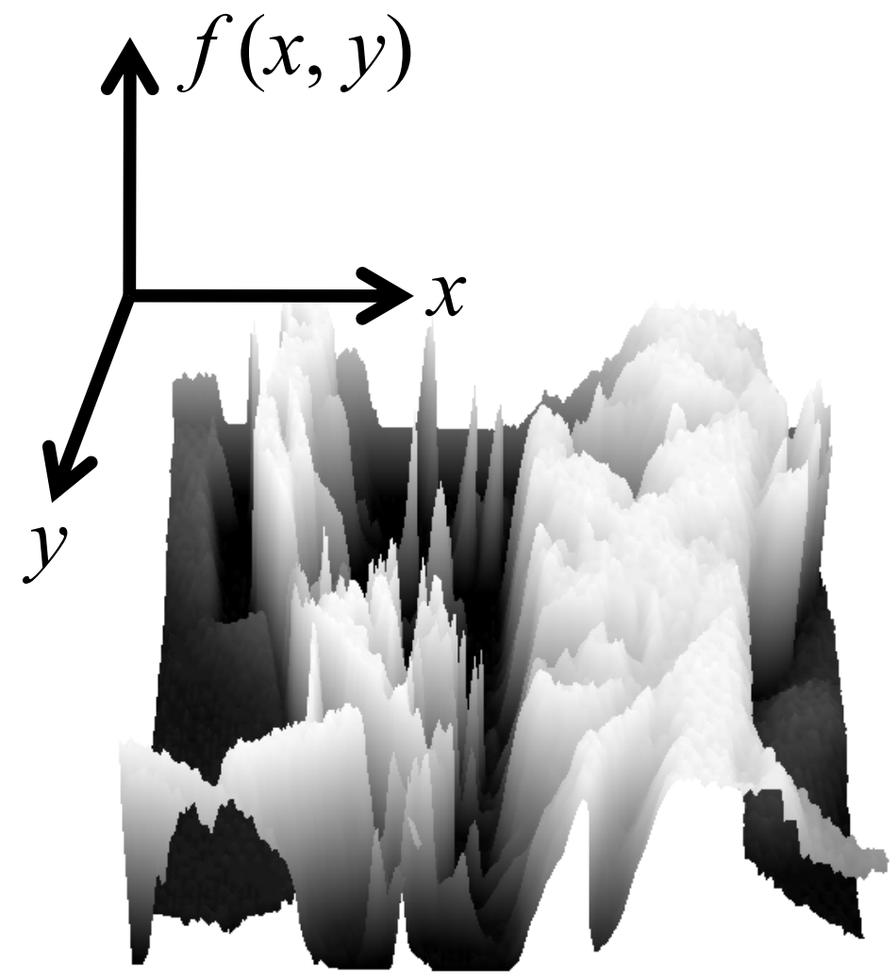
A grayscale image is a 2D array of numbers.

What is an image?

Mathematically speaking...

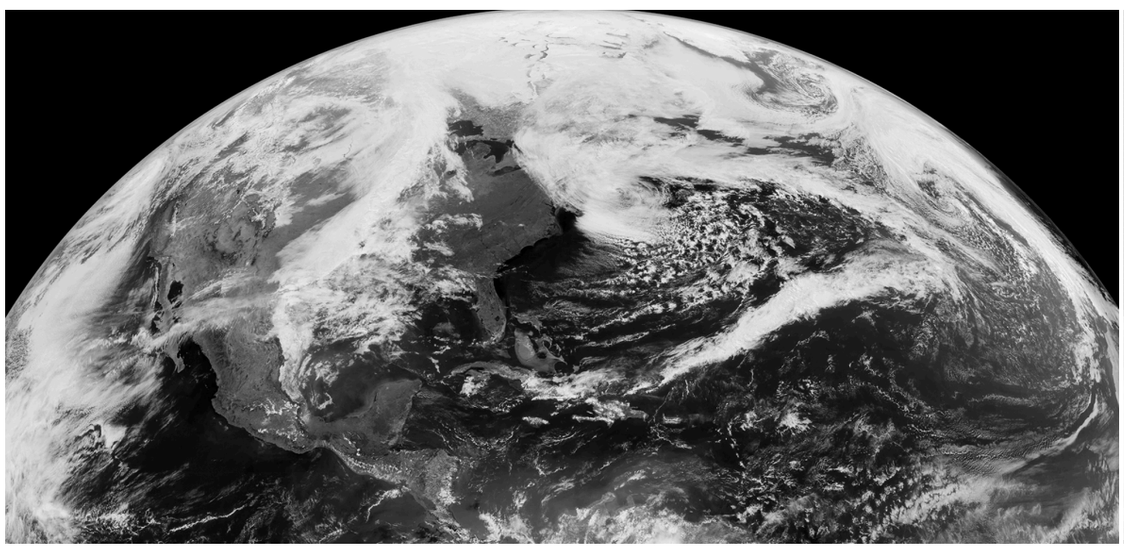
A grayscale image is a function f, from $\mathbb{R}^2 \rightarrow \mathbb{R}$



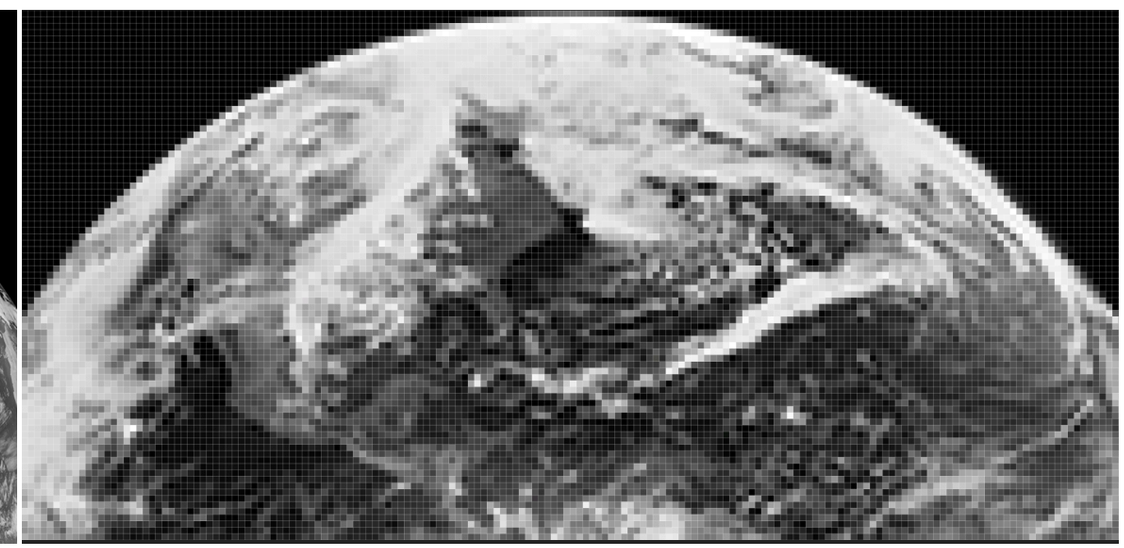


What's the difference?

The computational representation is a **sampled** version of the (ideal) mathematical representation.



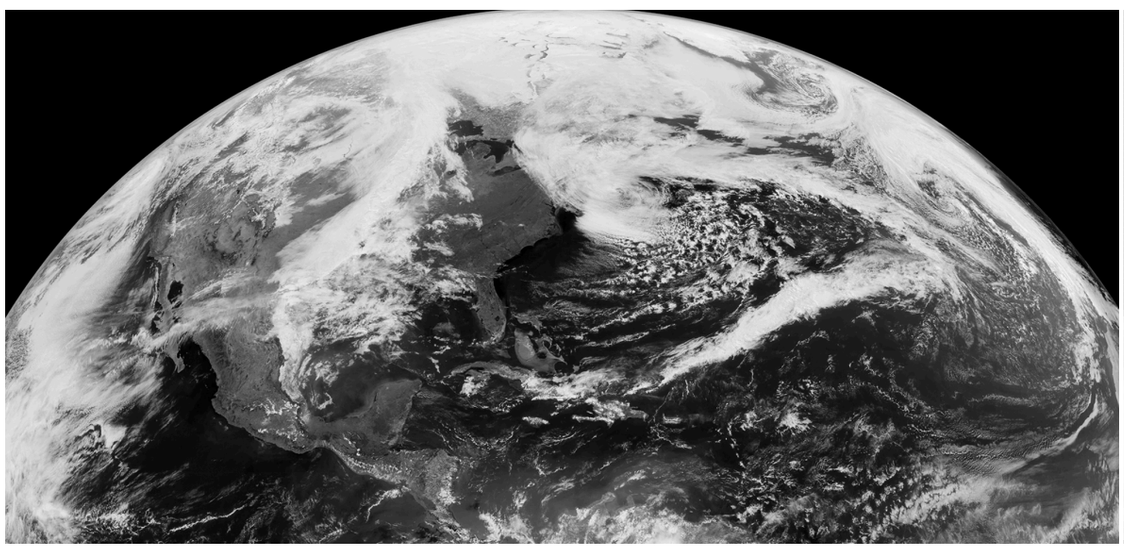
(ideally) continuous



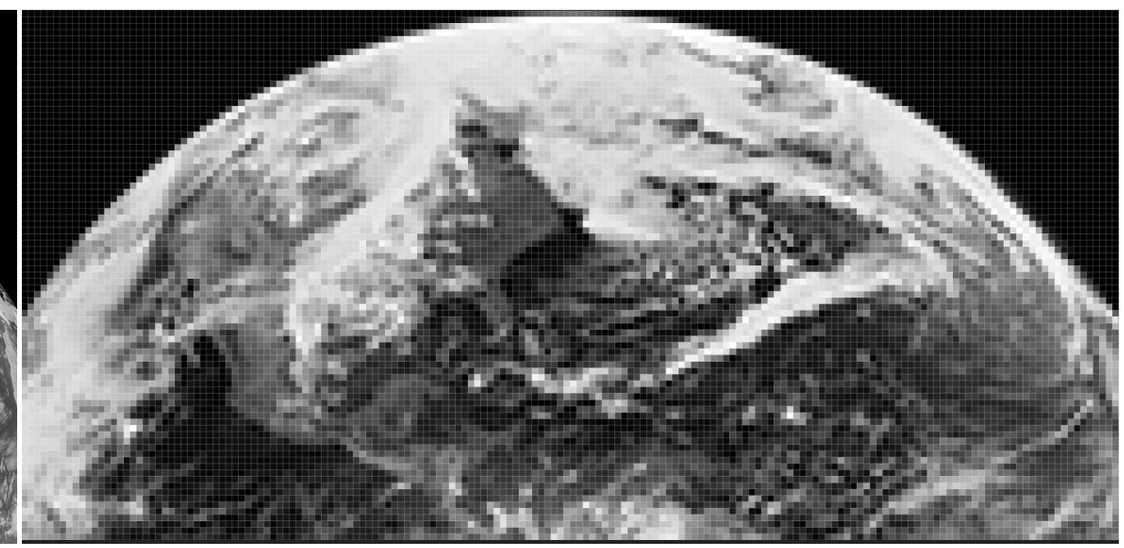
step function

What's the difference?

The computational representation is a **sampled** version of the (ideal) mathematical representation.



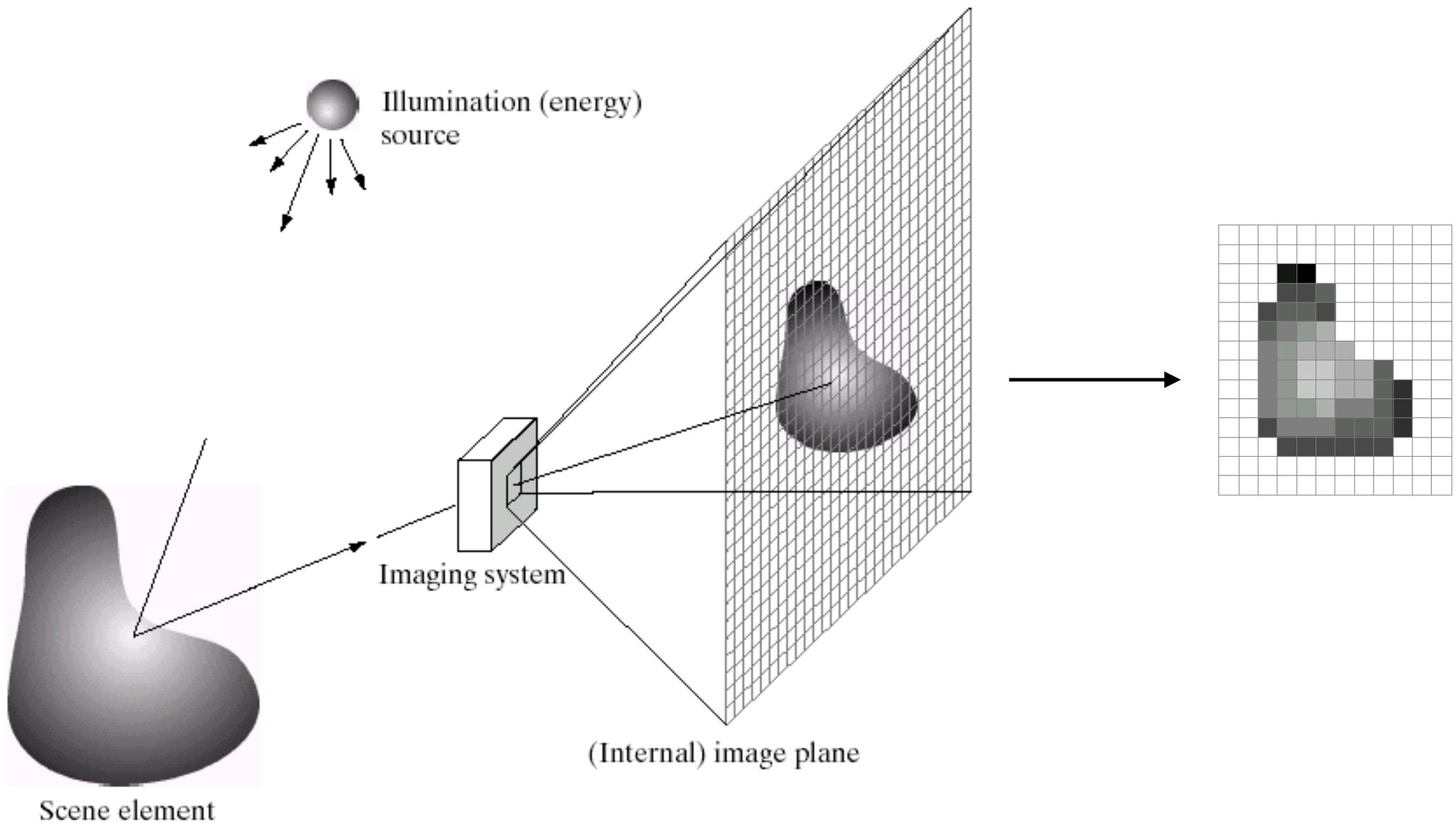
(ideally) continuous



step function

(we can also still write a step function that represents the sampled version)

How do images happen?

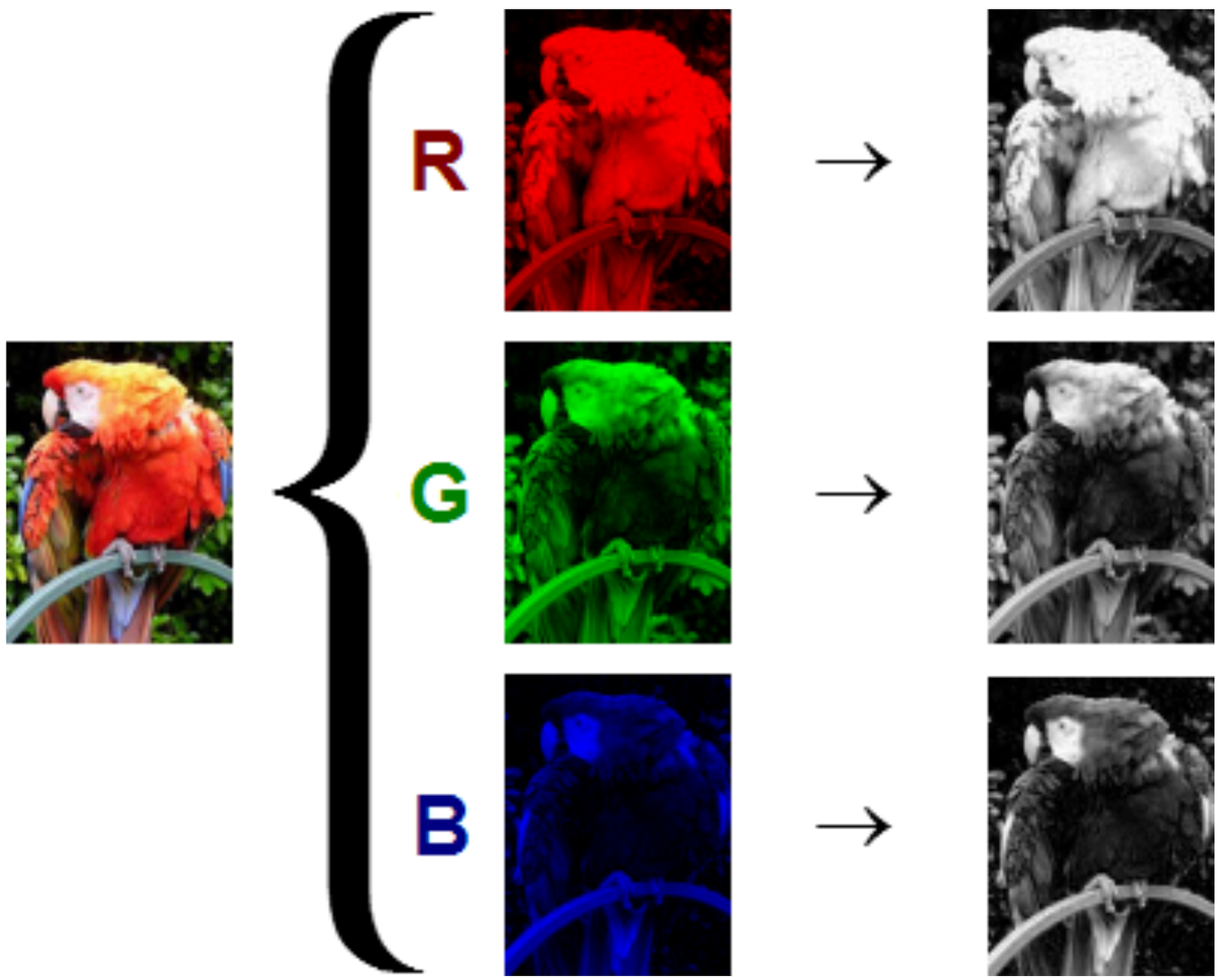


more on this process later...

What is a color image?

3 grayscale images, each representing one color channel

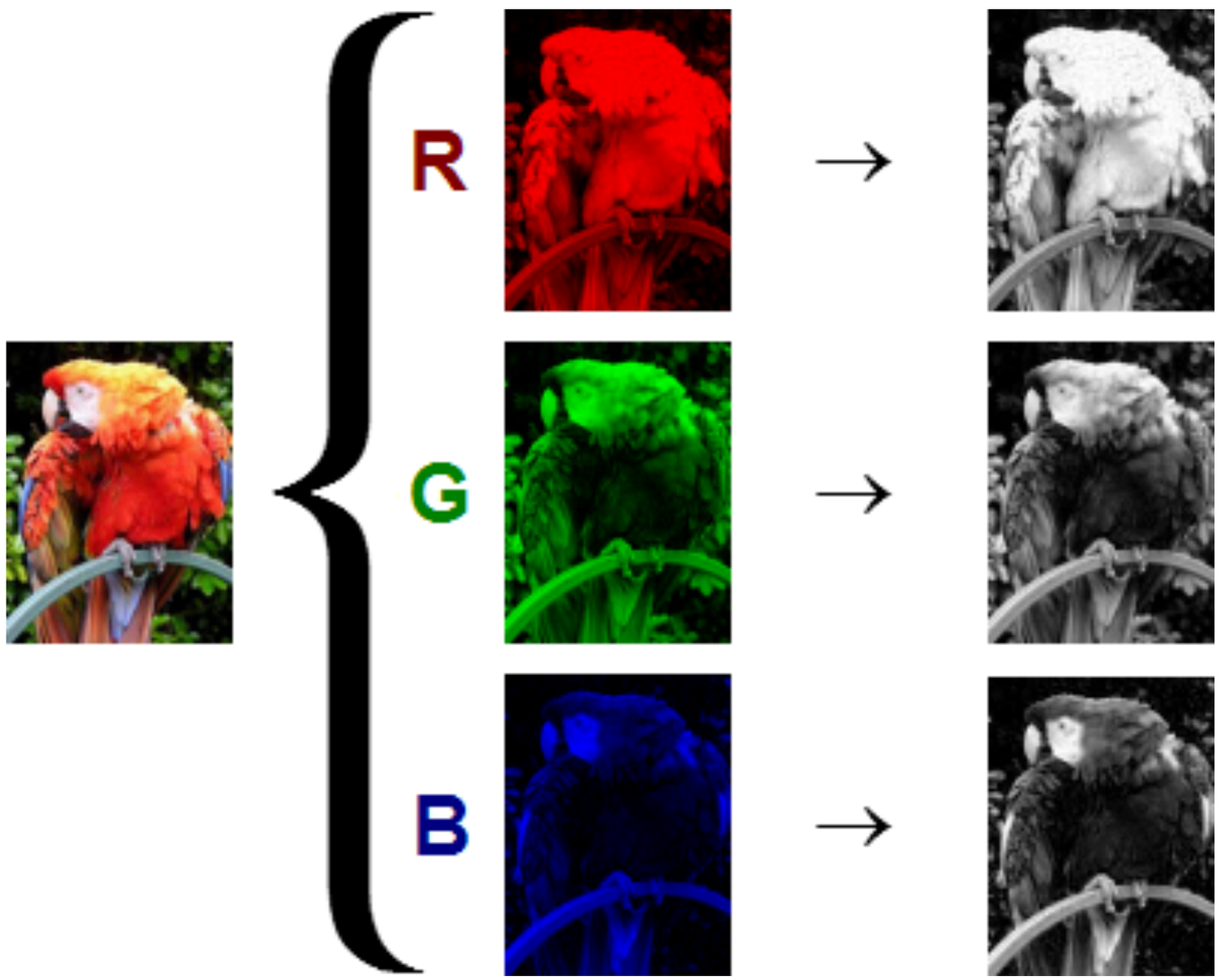
Often red, green and blue (RGB)



What is a color image?

3 grayscale images, each representing one color channel

Often red, green and blue (RGB)

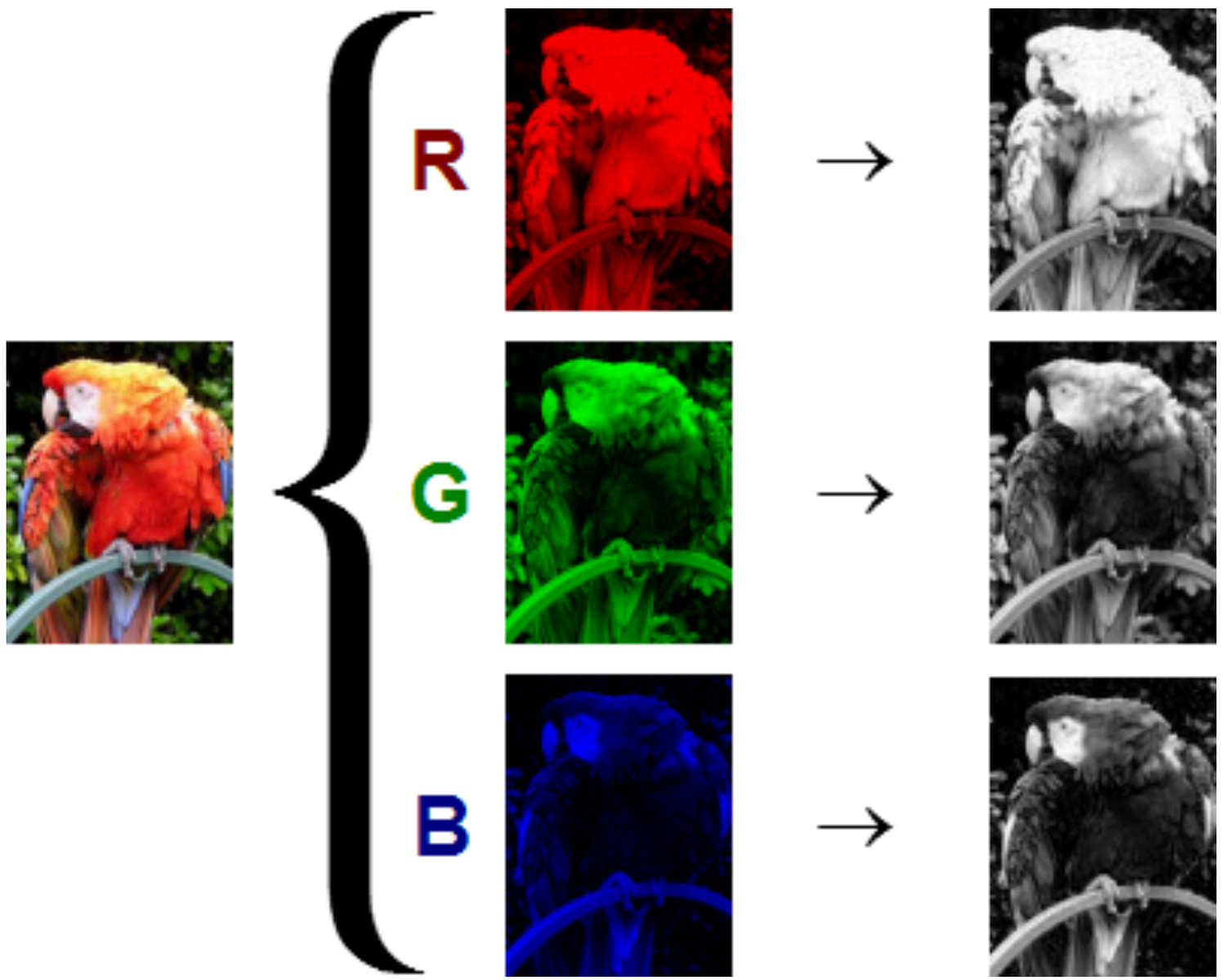


Computationally: HxWx3 array

What is a color image?

3 grayscale images, each representing one color channel

Often red, green and blue (RGB)



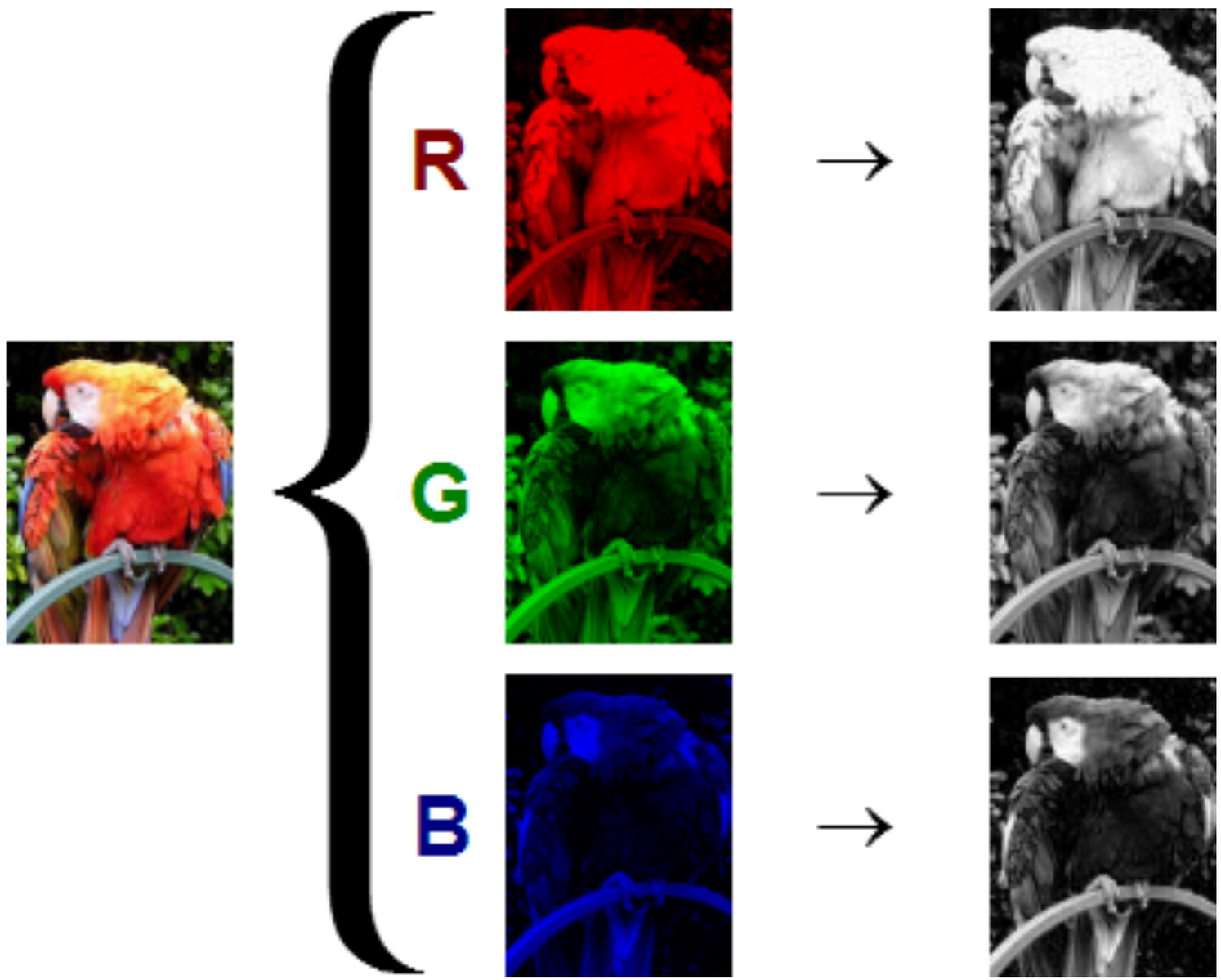
Computationally: HxWx3 array

Mathematically: $f : \mathbb{R}^2 \to \mathbb{R}^3$

What is a color image?

3 grayscale images, each representing one color channel

Often red, green and blue (RGB)



Computationally: HxWx3 array

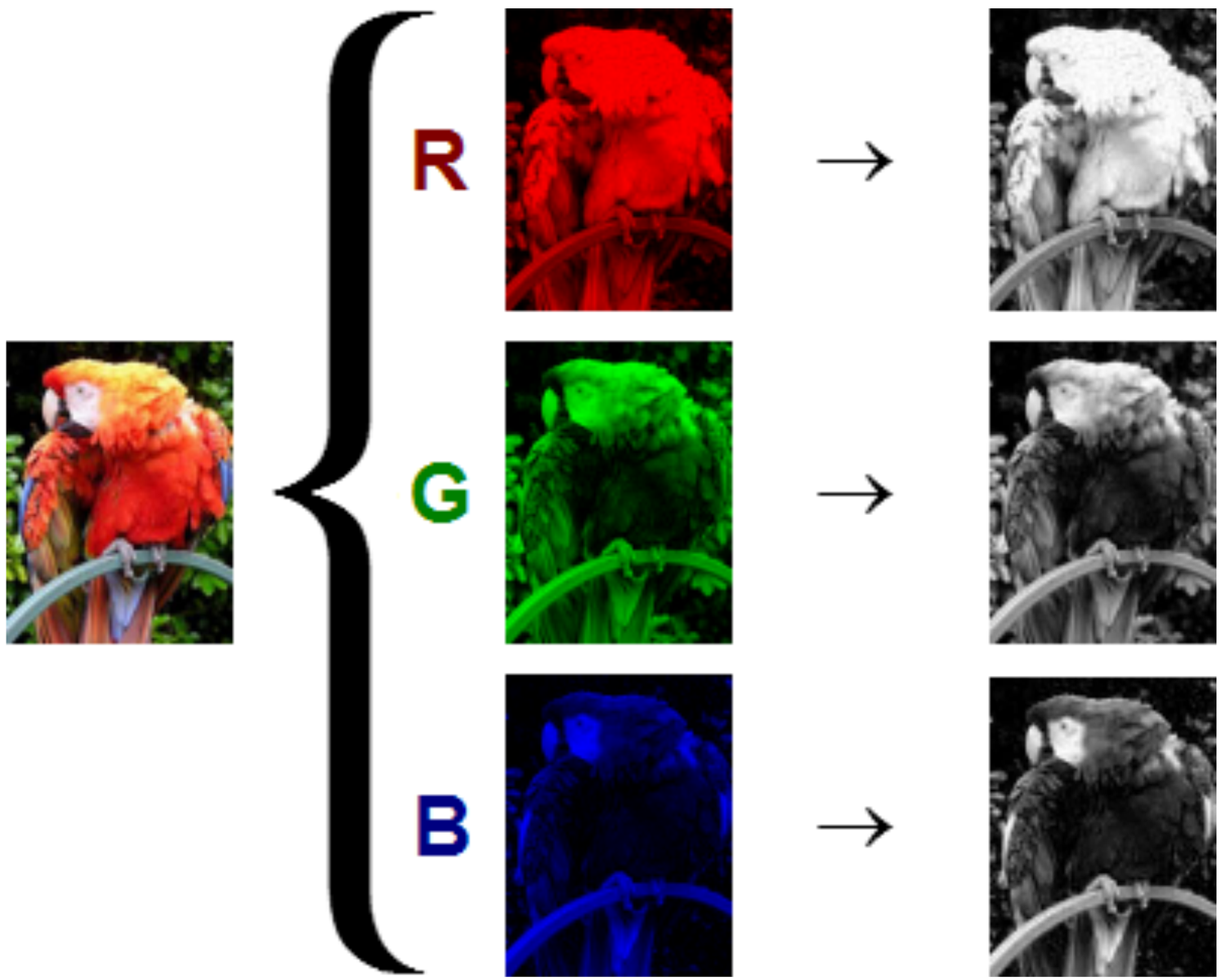
Mathematically: $f : \mathbb{R}^2 \to \mathbb{R}^3$

position

What is a color image?

3 grayscale images, each representing one color channel

Often red, green and blue (RGB)



Computationally: HxWx3 array

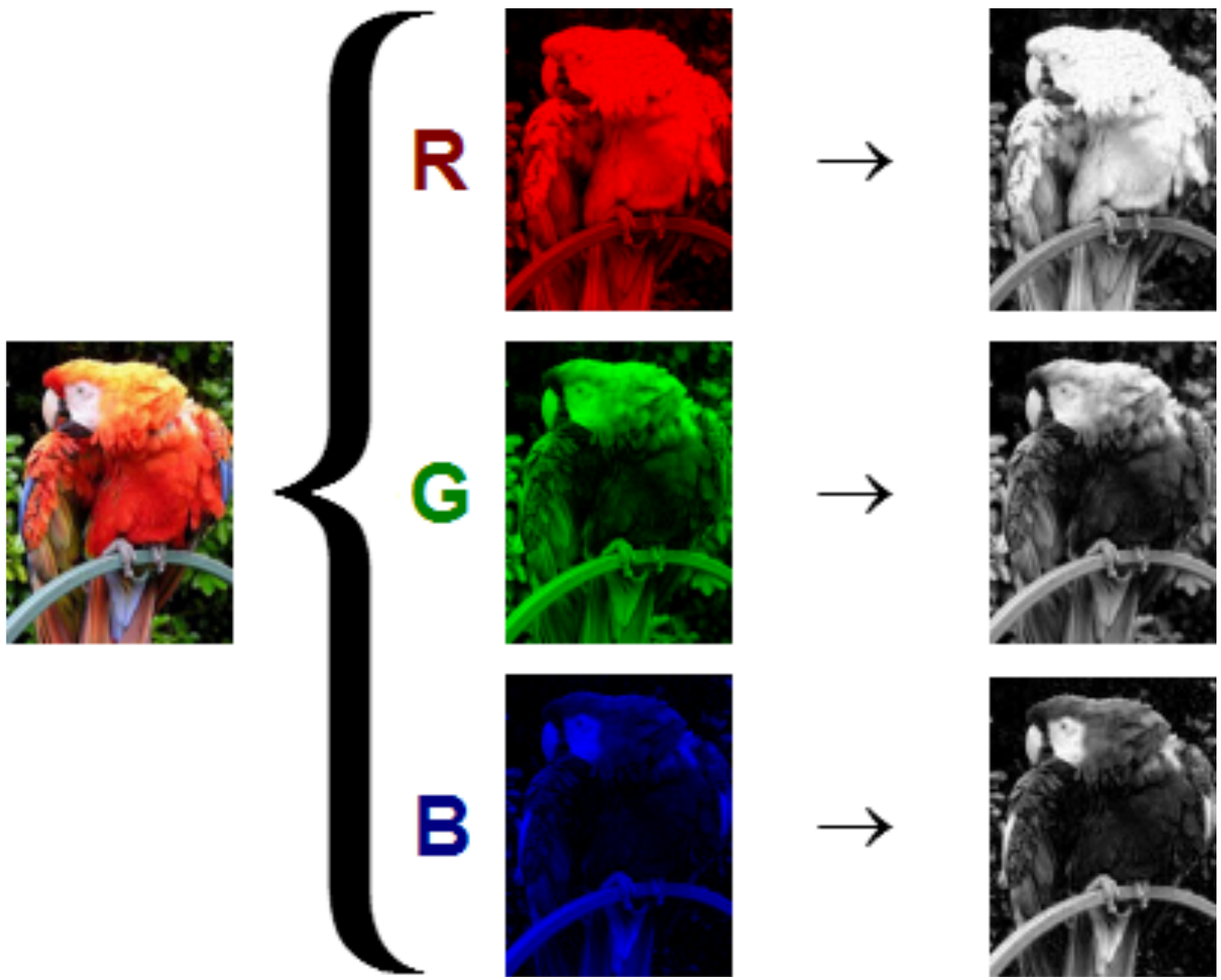
Mathematically: $f : \mathbb{R}^2 \to \mathbb{R}^3$

position color

What is a color image?

3 grayscale images, each representing one color channel

Often red, green and blue (RGB)



Computationally: HxWx3 array

Mathematically: $f : \mathbb{R}^2 \to \mathbb{R}^3$

position color

Not always RGB (HSL, Lab, ...)

Not always only 3 channels

Images are big

Efficiency matters - and not just big-O:

n is large, so even O(**n**) can be slow

This is a pretty small image.
It's 227 x 336 pixels, grayscale.
That's 95,842 bytes, or 93 kb

Images are big

Efficiency matters - and not just big-O:

n is large, so even O(**n**) can be slow

This is a pretty small image.
It's 227 x 336 pixels, grayscale.
That's 95,842 bytes, or 93 kb

A grayscale 1080p frame (1920x1080) is about 2MB.

Images are big

Efficiency matters - and not just big-O:

n is large, so even O(**n**) can be slow

This is a pretty small image.
It's 227 x 336 pixels, grayscale.
That's 95,842 bytes, or 93 kb

A grayscale 1080p frame (1920x1080) is about 2MB.
Want color? 6MB.

Images are big

Efficiency matters - and not just big-O:

n is large, so even O(**n**) can be slow

This is a pretty small image.
It's 227 x 336 pixels, grayscale.
That's 95,842 bytes, or 93 kb

A grayscale 1080p frame (1920x1080) is about 2MB.
Want color? 6MB.
Want video? ~30 frames per second; 900MB for a 5s clip

Images are big

Efficiency matters - and not just big-O:

n is large, so even O(**n**) can be slow

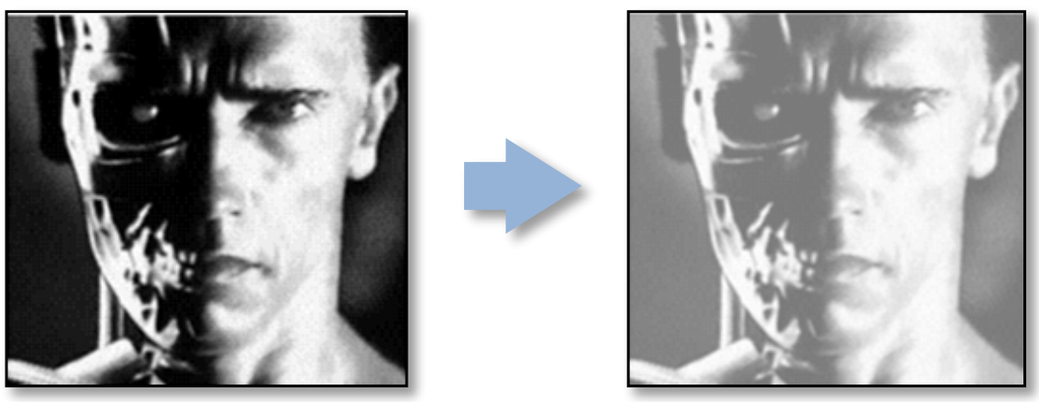
This is a pretty small image.
It's 227 x 336 pixels, grayscale.
That's 95,842 bytes, or 93 kb

A grayscale 1080p frame (1920x1080) is about 2MB.
Want color? 6MB.
Want video? ~30 frames per second; 900MB for a 5s clip
Want a movie? 1296 gigabytes for a 2-hour movie.

Image Processing: What can we do to an image?

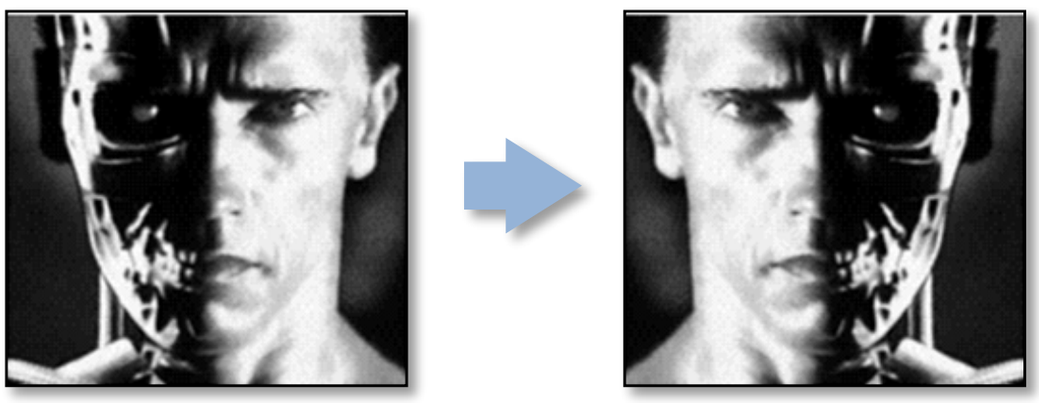
Written as a function, we can *transform* the image function to create altered functions (images):

photometric transformations



$$g\,(x,y) = f\,(x,y) + 20$$

geometric transformations

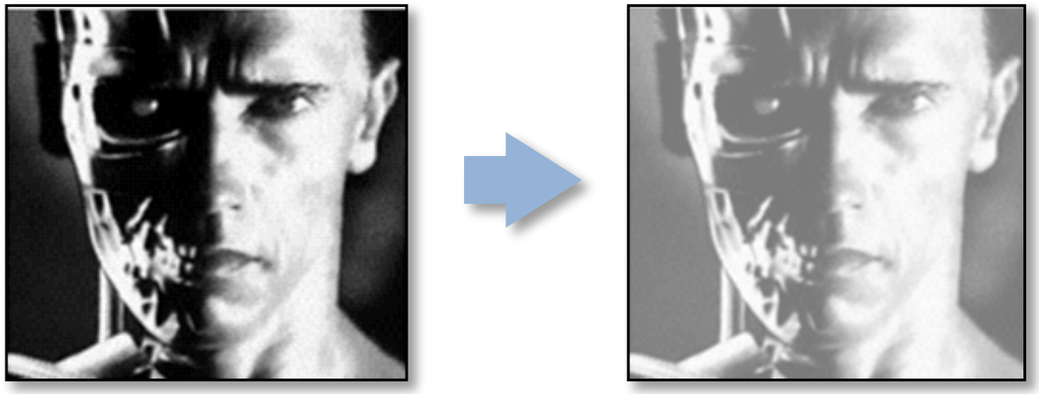


$$g\,(x,y) = f\,(-x,y)$$

Image Processing: What can we do to an image?

Written as a function, we can *transform* the image function to create altered functions (images):

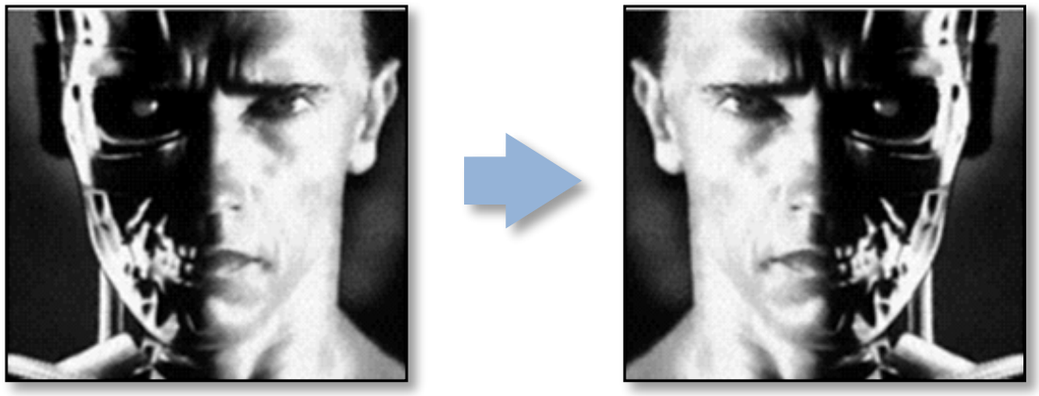
photometric transformations



(increase brightness)

$$g(x,y) = f(x,y) + 20$$

geometric transformations

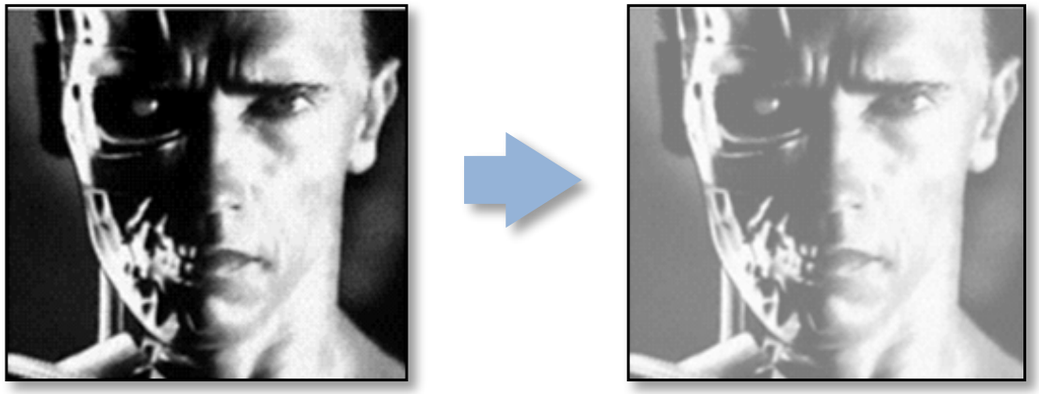


$$g(x,y) = f(-x,y)$$

Image Processing: What can we do to an image?

Written as a function, we can *transform* the image function to create altered functions (images):

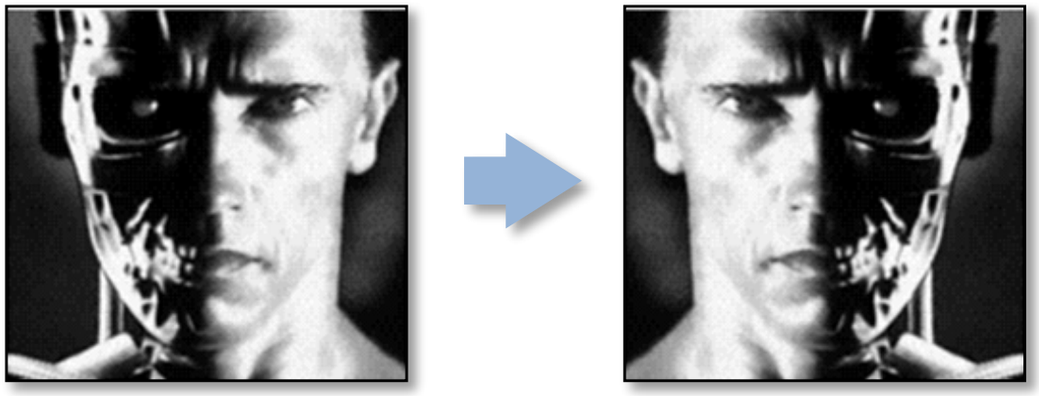
photometric transformations



(increase brightness)

$$g\,(x,y) = f\,(x,y) + 20$$

geometric transformations



(flip horizontally)

$$g\,(x,y) = f\,(-x,y)$$

Lecture 1 Problems

- In your breakout groups:
  - Find the "Problems" linked from today's lecture on the course webpage.
  - Open the Google Doc pinned in your group's text channel
  - Collaboratively write up answers to the problems in the google doc.

Real images aren't perfect

Real images are not only sampled, but they often have noise: unwanted variations in measured intensity value.

f(x, y)

Causes of noise (incomplete list):

- electronic variations in sensor chip
- analog-to-digital quantization
- film grain
- cosmic rays

Real images aren't perfect

Real images are not only sampled, but they often have noise: unwanted variations in measured intensity value.

f(x, y)

Causes of noise (incomplete list):

- electronic variations in sensor chip
- analog-to-digital quantization
- film grain
- cosmic rays

often, we can assume that noise is *random*

Real images aren't perfect

Real images are not only sampled, but they often have noise: unwanted variations in measured intensity value.

f(x, y)

Causes of noise (incomplete list):

- electronic variations in sensor chip
- analog-to-digital quantization
- film grain
- cosmic rays

often, we can assume that noise is *random*

(other times, we can't but we do anyway)

Real images aren't perfect

- Real images are not only sampled, but they often contain **noise**.

f(x, y)

???

How could we *denoise f*?