

Nhập môn Công nghệ Thông tin

Chapter 3 – Data Representation (Biểu diễn dữ liệu)

Khoa Công nghệ Thông tin
Đại học Ngoại ngữ - Tin học TP HCM
2019

1

Data (Dữ liệu) và Máy tính

Máy tính là thiết bị đa phương tiện, xử lý rất nhiều loại thông tin

Máy tính lưu trữ, trình bày và giúp chúng ta sửa đổi

- Số
- Chữ
- Âm thanh
- Hình ảnh và đồ họa
- Video

Tất cả được lưu trữ dưới dạng chữ số nhị phân(**bits**)

2

2

Thông tin tương tự và số (Analog and Digital Information)

Máy tính là hữu hạn

Làm thế nào để chúng ta biểu diễn một thế giới vô tận?

Chúng ta biểu diễn vừa đủ để thỏa mãn nhu cầu tính toán và các giác quan thị giác và âm thanh của chúng ta

3

3

Thông tin tương tự và số (Analog and Digital Information)

Thông tin có thể được biểu diễn bằng một trong hai cách: tương tự hoặc số

Analog data (Dữ liệu tương tự)

Một cách biểu diễn liên tục, tương tự như thông tin thực tế

Digital data (Dữ liệu số)

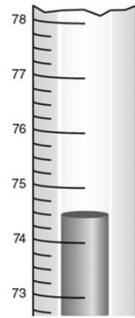
Một sự biểu diễn rời rạc, phá vỡ thông tin thành các phần tử riêng biệt

4

4

Thông tin tương tự và số (Analog and Digital Information)

Nhiệt kế thủy ngân
là một thiết bị tương tự



HÌNH 3.1 Nhiệt kế thủy ngân liên tục tăng tỷ lệ thuận với nhiệt độ

5

5

Thông tin tương tự và số (Analog and Digital Information)

Máy tính không thể làm việc tốt với dữ liệu tương tự, vì vậy chúng ta số hóa dữ liệu

Digitize (số hóa)

Chia dữ liệu thành từng phần và thể hiện các phần đó một cách riêng biệt

Tại sao chúng ta sử dụng nhị phân để thể hiện dữ liệu số hóa?

6

6

Tín hiệu điện tử

Thông tin quan trọng về tín hiệu điện tử

- Một tín hiệu tương tự liên tục làm dao động điện áp lên xuống
- Tín hiệu số chỉ có trạng thái cao hoặc thấp tương ứng với hai chữ số nhị phân
- Tất cả các tín hiệu điện tử (cả tương tự và số) đều làm suy giảm khi chúng di chuyển xuống một dòng
- Điện áp của tín hiệu biến động do hiệu ứng môi trường

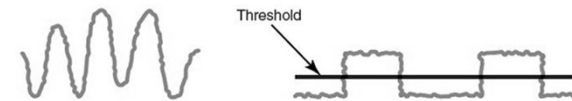
7

7

Tín hiệu điện tử (Cont'd)



HÌNH 3.2 Tín hiệu tương tự và tín hiệu số



HÌNH 3.3 Sự suy giảm tín hiệu tương tự và tín hiệu số

Định kỳ, tín hiệu số được **tái chuẩn hóa tần số** để lấy lại hình dạng ban đầu

8

8

Biểu diễn nhị phân

- Mỗi bit có thể là 0 hoặc 1, vì vậy nó có thể biểu diễn một sự lựa chọn giữa hai khả năng (hay "hai điều")
- Hai bits có thể biểu diễn bốn điều (Why? Hint: 00, 01, 10, 11.)

Có bao nhiêu thứ 3 bits có thể biểu diễn?

Có bao nhiêu thứ 4 bits có thể biểu diễn?

Có bao nhiêu thứ 8 bits có thể biểu diễn?

9

9

B

1 Bit	2 Bits	3 Bits	4 Bits	5 Bits
0	00	000	0000	00000
1	01	001	0001	00001
	10	010	0010	00010
	11	011	0011	00011
		100	0100	00100
		101	0101	00101
		110	0110	00110
		111	0111	00111
			1000	01000
			1001	01001
			1010	01010
			1011	01011
			1100	01100
			1101	01101
			1110	01110
			1111	01111
				10000
				10001
				10010
				10011
				10100
				10101
				10110
				10111
				11000
				11001
				11010

10

10

Biểu diễn các số tự nhiên

8-bit Binary Representation	Natural Number
01111111	127
01111110	126
...	...
00000011	3
00000010	2
00000001	1
00000000	0

- Dễ! Chỉ cần chuyển đổi sang nhị phân
- Máy tính lưu trữ dữ liệu theo từng khối có kích thước cố định, vì vậy chúng ta có các số 0 ở đầu

Các số nguyên bao gồm những gì mà số tự nhiên không có?

11

11

Biểu diễn các số âm

Signed-magnitude number representation (Biểu diễn số dạng độ lớn có dấu)

- Được con người sử dụng
- Máy tính thường sử dụng các hệ thống biểu diễn số nguyên (Integer) để đại diện cho cả số dương và số âm.
- Dấu thể hiện thứ tự (số âm đứng trước số dương theo thứ tự tăng dần)
- Các chữ số biểu thị độ lớn (khoảng cách từ 0)



12

12

Biểu diễn các số âm

Vấn đề: Hai số 0 (dương và âm) (-0 và +0)

Không có vấn đề gì đối với con người, nhưng sẽ gây ra sự phức tạp không cần thiết trong máy tính

Giải quyết: Biểu diễn số nguyên bằng cách liên kết chúng với số tự nhiên

Một nửa số tự nhiên sẽ đại diện cho chính chúng

Nửa còn lại sẽ đại diện cho số nguyên âm

13

13

Biểu diễn các số âm

Được gọi là biểu diễn phần bù của mười, bởi vì chúng ta có thể sử dụng công thức này để tính toán biểu diễn của một số âm

$$\text{Negative}(I) = 10^k - I, \text{ where } k \text{ is the number of digits}$$

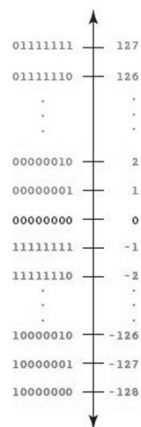
Ví dụ, -3 là Âm (3), do đó, sử dụng hai chữ số, biểu diễn của nó là Âm (3) = $100 - 3 = 97$

Chúng ta nhận được gì nếu thử ở dạng nhị phân?

14

14

Biểu diễn các số âm



Two's Complement (Phần bù của hai)

(Dòng số nhị phân dễ đọc hơn khi viết theo chiều dọc)
(Dùng biểu diễn trong máy tính)

Hãy nhớ bảng hướng dẫn cách biểu diễn số tự nhiên đã học?

Điều gì thú vị ở bit ngoài cùng bên trái?

15

15

Biểu diễn các số âm

Two's Complement (Phần bù của hai)

Để biểu diễn số âm trong dạng này, ta thực hiện phép đảo bit (chuyển 0 thành 1 và ngược lại) và sau đó cộng thêm 1 vào kết quả.

Số	Số dương	Số âm
-5	0000101	11111011 (bù 2)

16

16

Biểu diễn các số âm

Phép cộng và phép trừ giống như trong phép tính số bù mười

$$\begin{array}{r} -127 \quad 10000001 \\ + \quad 1 \quad 00000001 \\ \hline -126 \quad 10000010 \end{array}$$

Điều gì sẽ xảy ra nếu giá trị tính toán không khớp?

17

17

Tràn số (Number Overflow)

Nếu mỗi giá trị được lưu trữ bằng 8 bit, thì $127 + 3$ tràn:

$$\begin{array}{r} 01111111 \\ + 00000011 \\ \hline 10000010 \end{array}$$

Rõ ràng, $127 + 3$ là -126 . Hãy nhớ khi chúng ta nói rằng chúng ta sẽ luôn thất bại trong nỗ lực lập bản đồ một thế giới vô hạn vào một cỗ máy hữu hạn?

Hầu hết các máy tính sử dụng 32 hoặc 64 bit cho số nguyên, nhưng luôn có vô số số không được biểu diễn

18

18

Biểu diễn số thực

Số thực (Real numbers) là số có một phần nguyên và một phần phân số (một trong hai số đó có thể bằng 0)
Ví dụ số thực ở cơ số 10:

104.32
0.999999
357.0
3.14159

Ở bên trái dấu thập phân, trong cơ số 10, chúng ta có vị trí hàng đơn vị, hàng chục, hàng trăm, v.v.

10^{-1} , 10^{-2} , 10^{-3} ...

19

19

Biểu diễn số thực

Giá trị thực trong cơ số 10 có thể được xác định theo công thức sau đây trong đó phần định trị là một số nguyên:

$$\text{sign} * \text{mantissa} * 10^{\text{exp}}$$

(Dấu * Phần định trị * 10^{exponent})

Dấu (dương hoặc âm)
Phần định trị (mantissa) được tạo thành từ các chữ số trong giá trị với điểm cơ số ở bên phải
Số mũ (exponent) xác định điểm cơ số được dịch chuyển như thế nào so với phần định trị.

Biểu diễn này được gọi là dấu chấm động bởi vì điểm cơ số "floats"

20

20

Biểu diễn số thực

Giá trị thực Biểu diễn dấu chấm động	Giá trị thực Biểu diễn dấu chấm động
12001.00	$12001 * 10^0$
-120.01	$-12001 * 10^{-2}$
0.12000	$12000 * 10^{-5}$
-123.10	$-12310 * 10^{-2}$
155555000.00	$15555 * 10^4$

21

21

Biểu diễn số thực

Công thức để biểu diễn số thực ở cơ số 2 như sau:

$$\text{sign} * \text{mantissa} * 2^{\text{exp}}$$

(Dấu * Phần định trị * 2^{exponent})

22

22

Biểu diễn số thực

Biểu diễn dạng khoa học (Scientific notation)

Một dạng biểu diễn dấu phẩy động, trong đó dấu thập phân được giữ ở bên phải của chữ số ngoài cùng bên trái

12001.32708 là $1.200132708E+4$ trong ký hiệu khoa học (E + 4 là cách máy tính hiển thị $\times 10^4$)

Ký hiệu khoa học của 123.332 là gì?

Ký hiệu khoa học của 0.0034 là gì?

23

23

Biểu diễn văn bản

Biểu diễn văn bản như thế nào?

Số lượng ký tự để biểu diễn là hữu hạn (whew!), Vì vậy hãy liệt kê tất cả chúng và gán cho mỗi ký tự một chuỗi nhị phân

Bộ ký tự (Character set)

- Là một tập hợp các ký tự, biểu tượng và ký hiệu được sử dụng để biểu diễn ngôn ngữ, văn bản và các dạng thông tin khác trong máy tính và hệ thống.
- Mỗi bộ ký tự thường có một số lượng ký tự cố định, và các ký tự này có thể được mã hóa bằng các giá trị số hoặc dãy bit.
- Các bộ ký tự thường được sử dụng để biểu diễn các ngôn ngữ khác nhau trên thế giới

24

24

Bộ ký tự ASCII

ASCII là viết tắt của **American Standard Code for Information Interchange**

Bộ ký tự cơ bản ASCII gồm 128 ký tự bao gồm chữ cái, số và ký tự đặc biệt. Mỗi ký tự được mã hóa bằng 7 bit

- Sau đó ASCII mở rộng đã phát triển sử dụng 8 bit để biểu diễn cho một ký tự

Có bao nhiêu ký tự có thể được trình bày?

25

25

Ảnh xạ bộ ký tự ASCII

Left Digit(s)	Right Digit	ASCII									
		0	1	2	3	4	5	6	7	8	9
0		NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT
1		LF	VT	FF	CR	SO	SI	DLE	DC1	DC2	DC3
2		DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS
3		RS	US	□	!	"	#	\$	%	&	'
4		()	*	+	,	-	.	/	0	1
5		2	3	4	5	6	7	8	9	:	;
6		<	=	>	?	@	A	B	C	D	E
7		F	G	H	I	J	K	L	M	N	O
8		P	Q	R	S	T	U	V	W	X	Y
9		Z	[\]	^	_	`	a	b	c
10		d	e	f	g	h	i	j	k	l	m
11		n	o	p	q	r	s	t	u	v	w
12		x	y	z	{		}	~	DEL		

FIGURE 3.5 The ASCII character set

26

26

Extended ASCII

000	001	002	003	004	005	006	007	008	009	010	011	012	013	014	015
NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
016	017	018	019	020	021	022	023	024	025	026	027	028	029	030	031
DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US	
032	033	034	035	036	037	038	039	040	041	042	043	044	045	046	047
[]	^	_	`	a	b	c	d	e	f	g	h	i	j	k
048	049	050	051	052	053	054	055	056	057	058	059	060	061	062	063
l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{
064	065	066	067	068	069	070	071	072	073	074	075	076	077	078	079
{		}	~	DEL											
080	081	082	083	084	085	086	087	088	089	090	091	092	093	094	095
P	Q	R	S	T	U	V	W	X	Y	Z	[]	^	_	
096	097	098	099	100	101	102	103	104	105	106	107	108	109	110	111
`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
p	q	r	s	t	u	v	w	x	y	z	[]	^	_	DEL
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
€	£	¥	¢	¤	¦	§	¨	©	ª	«	¬	®	¯	°	±
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿		
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255
ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

27

27

Bộ ký tự Unicode

- Extended ASCII: Phiên bản mở rộng của bộ ký tự ASCII cung cấp 256 ký tự, đủ cho tiếng Anh những không đủ để sử dụng quốc tế
- Bộ ký tự Unicode sử dụng 16 bits cho mỗi ký tự

Ảnh xạ có thể biểu thị bao nhiêu ký tự?

28

28

Bộ ký tự Unicode

Code (Hex)	Character	Source
0041	A	English (Latin)
042F	Я	Russian (Cyrillic)
0E09	๙	Thai
13EA	Ꮚ	Cherokee
211E	℞	Letterlike symbols
21CC	⇌	Arrows
282F	⠋	Braille
345F	한	Chinese/Japanese/ Korean (common)

FIGURE 3.6 A few characters in the Unicode character set

29

29

Biểu diễn âm thanh

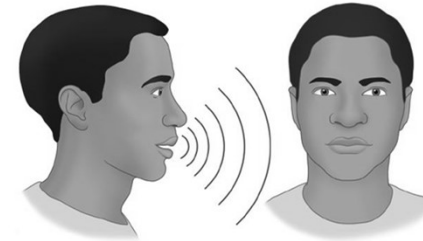


FIGURE 3.7 A sound wave vibrates our eardrums

Chúng ta cảm nhận được âm thanh khi một loạt sóng áp suất không khí làm rung chuyển màng trong tai, màng này sẽ gửi tín hiệu đến não

30

30

Biểu diễn thông tin âm thanh

Số hóa tín hiệu tương tự bằng cách

Lấy mẫu (Sampling): đo điện áp định kỳ và ghi lại dưới dạng giá trị số

— Lượng tử hóa (Quantization): biểu diễn điện áp dưới dạng số bằng cách sử dụng số bit hữu hạn

Bao lâu thì chúng ta nên lấy mẫu?

Tốc độ lấy mẫu khoảng 40.000 lần mỗi giây là đủ để tạo ra một sự tái tạo âm thanh hợp lý

31

31

Biểu diễn thông tin âm thanh

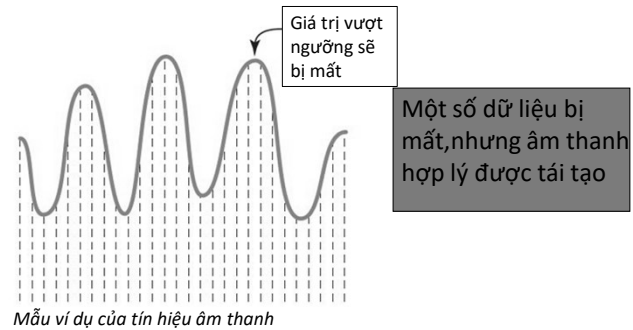
- Lưu trữ và xử lý: Chuỗi số biểu diễn tín hiệu âm thanh lưu trữ trong các tệp âm thanh như WAV, MP3, FLAC, vv.

- Khi cần phát lại âm thanh: máy tính sẽ sử dụng chuỗi số để tạo ra tín hiệu âm thanh liên tục mà tai người có thể nghe được.

32

32

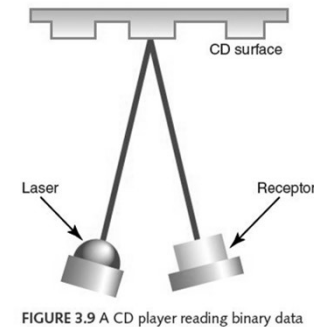
Biểu diễn thông tin âm thanh



33

33

Biểu diễn thông tin âm thanh



- CDs lưu trữ thông tin âm thanh (hoặc thông tin khác) bằng kỹ thuật số
 - Rỗ (Pits) (phản xạ kém)
 - Đất (Lands) (phản xạ tốt)
- Đọc bằng tia laser cường độ thấp
- Receptor chuyển đổi các phản xạ thành các chữ số nhị phân
- Chuỗi bit đại diện cho tín hiệu âm thanh

34

34

Định dạng âm thanh

Định dạng âm thanh: WAV, AU, AIFF, VQF, và MP3
– Sử dụng các kỹ thuật nén khác nhau

MP3 chiếm ưu thế:

- MPEG-2, tệp âm thanh lớp 3
- MPEG = Motion Picture Experts Group

Nén có tổn hao hay không tổn hao (hoặc cả hai)?

35

35

Biểu diễn Hình ảnh và Đồ họa

Màu sắc (Color)

- Chúng ta coi đó là điều hiển nhiên, nhưng thực sự nó là gì?

Võng mạc (Retinas) của mắt chúng ta có ba loại tế bào hình nón nhận ánh sáng

- Mỗi loại phản ứng với một tập hợp các tần số ánh sáng khác nhau
- Bộ não của chúng ta chuyển phản ứng đó thành nhận thức về màu đỏ, xanh lá cây hoặc xanh lam

36

36

Biển diễn Hình ảnh và Đồ họa

Màu sắc được biểu thị dưới dạng giá trị RGB (đỏ-lục-xanh)

Mô hình màu RGB (Red-Green-Blue): Trong mỗi màu được biểu diễn bằng một sự kết hợp của ba màu cơ bản: đỏ (Red), xanh lá cây (Green) và lam (Blue).

Giá trị RGB là (255, 255, 0) tối đa hóa sự đóng góp của màu đỏ và xanh lá cây và giảm thiểu sự đóng góp của màu xanh lam, dẫn đến màu vàng sáng

37

37

Biển diễn Hình ảnh và Đồ họa

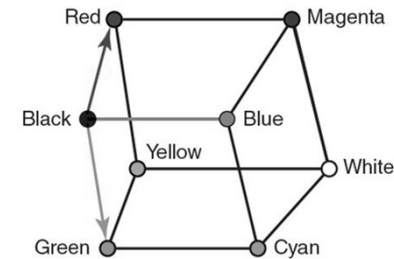


FIGURE 3.10 A three-dimensional color space

38

38

Biển diễn Hình ảnh và Đồ họa

Độ sâu màu (Color depth)

Lượng dữ liệu được sử dụng để biểu thị một màu HiColor

Độ sâu màu 16 bit: năm bit được sử dụng cho mỗi số trong một giá trị RGB với bit bổ sung đôi khi được sử dụng để biểu thị độ trong suốt

TrueColor

Độ sâu màu 24 bit: tám bit được sử dụng cho mỗi số trong giá trị RGB

39

39

Biển diễn Hình ảnh và Đồ họa

RGB VALUE			
Red	Green	Blue	Color
0	0	0	black
255	255	255	white
255	255	0	yellow
255	130	255	pink
146	81	0	brown
157	95	82	purple
140	0	0	maroon

Một vài giá trị RGB và màu sắc tương ứng

40

40

Số hóa hình ảnh và đồ họa

- Pixels (các thành phần ảnh)
 - Các điểm màu trong hình ảnh (hay thiết bị thể hiện)
- Resolution(Độ phân giải)
 - Số lượng pixels trong ảnh (hay thiết bị)
- Raster Graphics(Đồ họa raster)
 - Coi hình ảnh là tập hợp các pixel
 - Các định dạng phổ biến nhất: BMP, GIF, PNG, và JPEG
- Vector Graphics(Đồ họa vector)
 - Coi hình ảnh là bộ sưu tập các đối tượng hình học
 - Các định dạng phổ biến nhất : Flash và SVG, EPS

41

41

Số hóa hình ảnh và đồ họa

- BMP (bitmap)
 - Bitmap images, còn gọi là raster images, biểu diễn hình ảnh dưới dạng một lưới hình vuông gọi là "pixel".
 - Độ sâu màu TrueColor hoặc ít hơn để giảm kích thước file
 - Rất thích hợp cho việc nén bằng cách mã hóa độ dài chạy
- GIF (indexed color)
 - Đây là một định dạng hình ảnh bitmap chủ yếu được sử dụng cho hình ảnh động. GIF sử dụng một số lượng giới hạn màu (thường là 256) để lưu trữ hình ảnh, làm cho nó thích hợp cho biểu diễn biểu tượng và hình động
 - Do đó, mỗi pixel chỉ yêu cầu 8 bit hoặc ít hơn
 - GIF động là chuỗi hình ảnh ngắn
- PNG (Portable Network Graphics)
 - Dự định thay thế ảnh GIF
 - Nén lớn hơn với phạm vi độ sâu màu rộng hơn
 - Không có hoạt ảnh

42

42

Số hóa hình ảnh và đồ họa

- JPEG (Joint Photographic Experts Group)
 - JPEG là một định dạng hình ảnh bitmap được sử dụng chủ yếu cho hình ảnh màu
 - Sử dụng thuật toán nén mất mát để giảm kích thước tệp, nhưng đồng thời có thể dẫn đến mất mát chất lượng
 - Tại sao? Thị giác của con người có xu hướng làm mờ các màu cùng nhau trong các khu vực nhỏ (khoa học!)
 - Làm sao? Chuyển đổi từ miền không gian sang miền tần số, sau đó loại bỏ các thành phần tần số cao (toán học!)
 - Nghe có vẻ quen? Về cơ bản, cùng một ý tưởng được sử dụng trong MP3
 - Mức độ nén có thể điều chỉnh

Tóm tắt đồ họa Raster:: BMP, GIF, PNG, and JPEG

Cái nào sử dụng nén không mất dữ liệu? Mất mát?

Bạn sẽ sử dụng cái nào cho nghệ thuật vẽ đường? Để có một bức ảnh màu

43

43

Số hóa hình ảnh và đồ họa

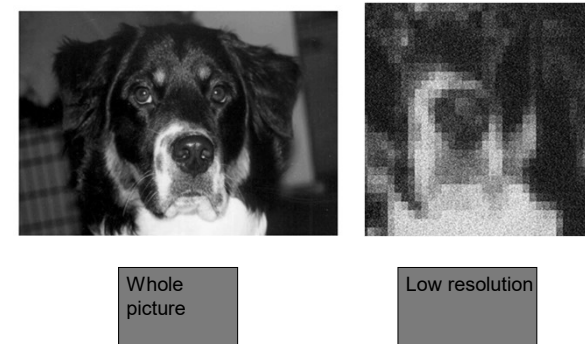


Figure 3.12 A digitized picture composed of many individual pixels

44

44

Đồ họa vector

Đồ họa Vector

- Định dạng mô tả hình ảnh dưới dạng đường thẳng và hình dạng hình học
- Đồ họa vectơ là một chuỗi các lệnh mô tả các hình dạng bằng cách sử dụng các thuộc tính toán học (ví dụ: hướng, chiều dài, độ dày, màu sắc)
- Đối với một số loại hình ảnh, kích thước file có thể nhỏ hơn với đồ họa raster vì không mọi pixel được mô tả.

45

45

Biểu diễn video

- **Serie of images : Loạt hình ảnh**
- **Play continuously, 24 frames/minute : Phát liên tục, 24 khung hình / phút**

Biểu diễn dữ liệu video chính là việc ghép nhiều hình ảnh tĩnh lại. Một số vấn đề về chất lượng của video phụ thuộc vào độ phức tạp của kỹ thuật nén video, được gọi là codec video.

46

46