

min-Hash

min-Hash

min-Hash is an efficient representation of a set A_i

$$m(\mathcal{A}_i, f) = \arg \min_{X \in \mathcal{A}_i} f(X)$$

set of visual words

hash function

visual word

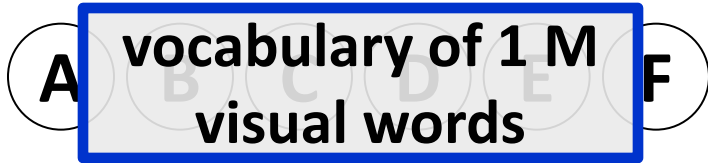
min-Hash is a locality sensitive hashing (LSH) function m that selects an element (visual word) $m(A_i)$ from each set A_i of visual words detected in image i so that

$$P\{m(\mathcal{I}_1) == m(\mathcal{I}_2)\} = \frac{|\mathcal{I}_1 \cap \mathcal{I}_2|}{|\mathcal{I}_1 \cup \mathcal{I}_2|}$$

Image similarity $\text{sim}(\mathcal{I}_1, \mathcal{I}_2) = \frac{|\mathcal{I}_1 \cap \mathcal{I}_2|}{|\mathcal{I}_1 \cup \mathcal{I}_2|}$

min-Hash

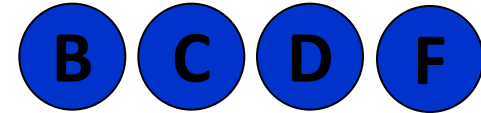
Vocabulary



Set I_1



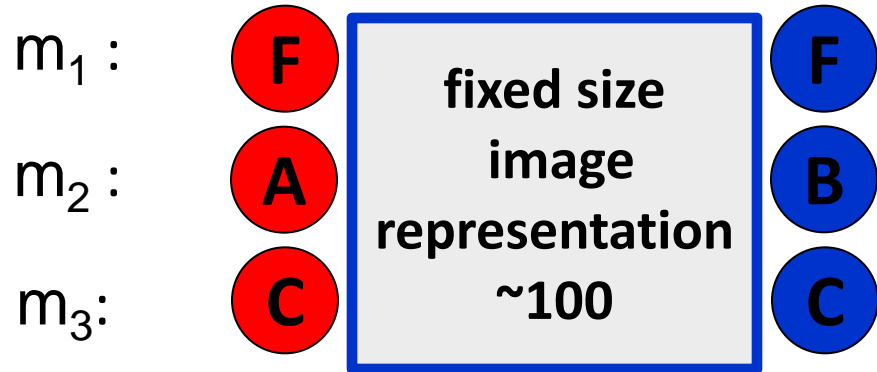
Set I_2



Random orderings

3 6 2 5 4 1
1 2 6 3 5 4
3 2 1 6 4 5

min-Hash




$$\text{sim}(I_1, I_2) = 1/2$$

Estimated similarity of I_1 and I_2 from 3 min-Hashes = $2/3$

Set Overlap and min-Hash

I_1 :  I_2 : 

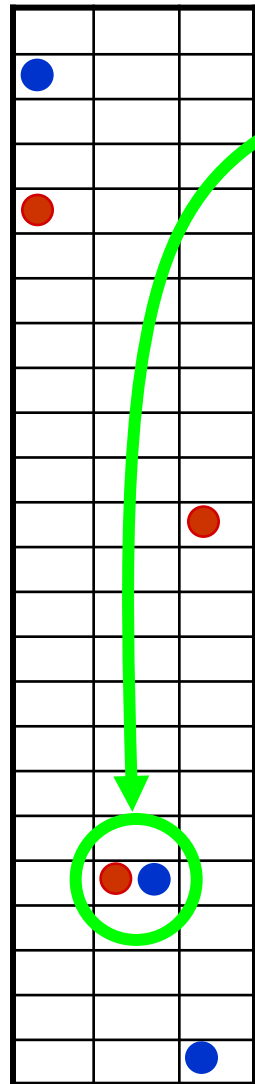
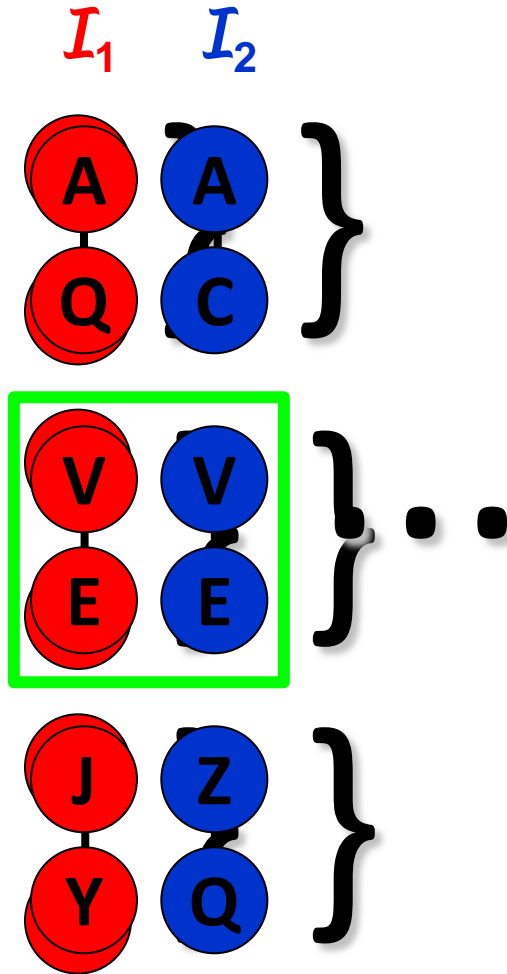
$I_1 \cup I_2$: 

$m(I_1) = m(I_2)$ 

$$P(m(I_1) = m(I_2)) = \frac{\text{Smiley}}{\text{Smiley} + X} = \frac{|I_1 \cap I_2|}{|I_1 \cup I_2|}$$

min-Hash

a sketch = s -tuple of min-Hashes



k hash tables

Sketch collision

collision:

all s min-Hashes must agree

$$P\{\text{collision}\} = \text{sim}(I_1, I_2)^s$$

retrieval:

1. generate k sketches
2. at least one of k sketches must collide

$P\{\text{retrieval}\} =$

$$1 - (1 - \text{sim}(I_1, I_2)^s)^k$$

Probability of Retrieving an Image Pair

Images of the same object and unrelated images

Near duplicate Images

$s = 3, k = 512$



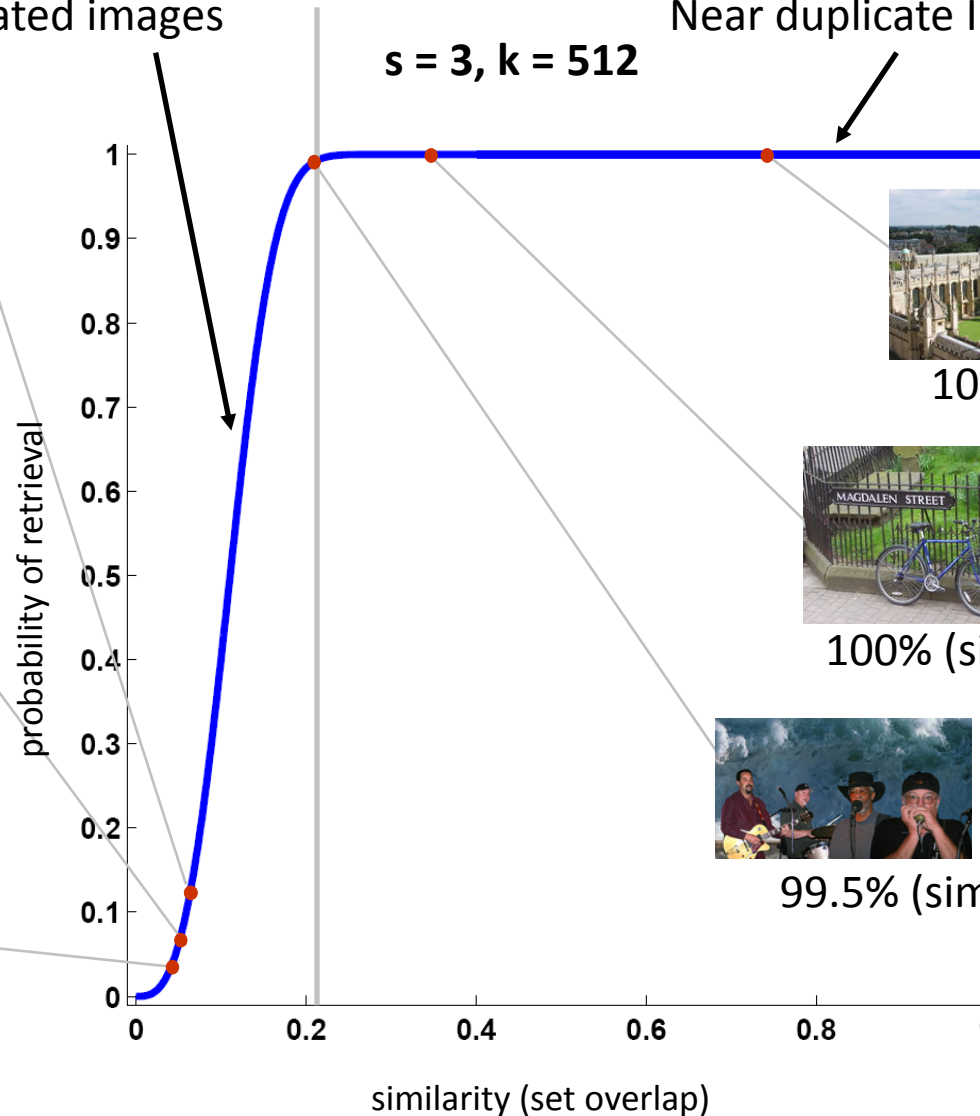
13.9 % (sim = 0,066)



8.9 % (sim = 0.057)



5.1% (sim = 0.047)



100% (sim = 0.746)



100% (sim = 0.322)



99.5% (sim = 0,217)