Rasterized Image Databases for Image Compression

Zoya Bylinskii*, Maria Shugrina*, Andrew Spielberg*, Wei Zhao*

CSAIL

bigdata@CSAIL
MIT BIG DATA INITIATIVE

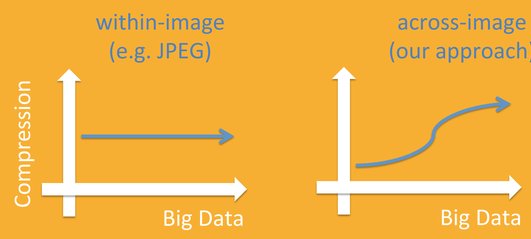
6.830
Database Systems
Final Project

MOTIVATION

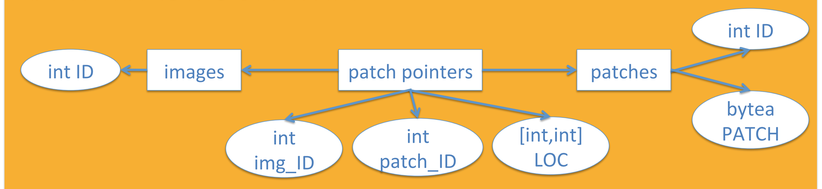
- More than 1.8 billion images uploaded to the internet every day
- Redundancy in large image collections can lead to more efficient storage

PATCH-BASED IMAGE CODING

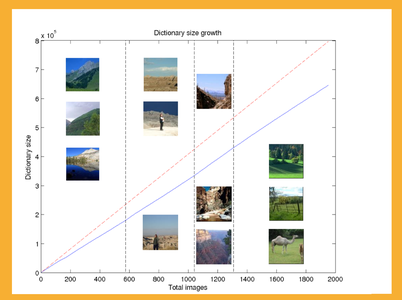
- Lossy compression
- Compression savings INCREASE as the database size INCREASES
- Rasterization is a simple operation – easily deconstruct and reconstruct images using patches

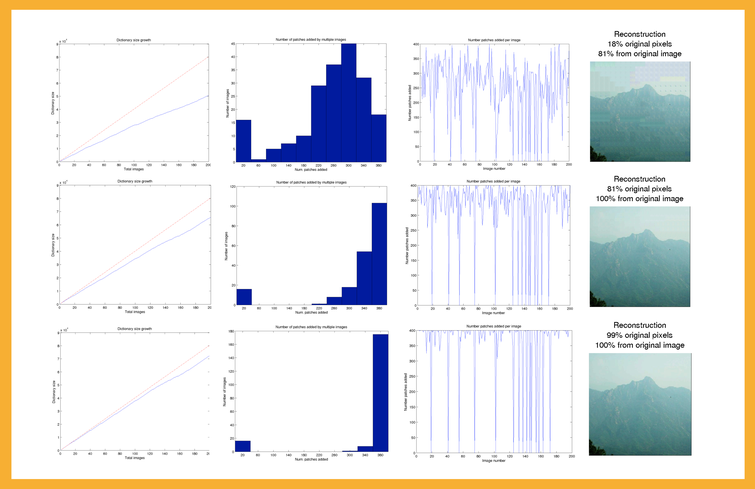


DATABASE SCHEMA



PARAMETER SELECTION





RESULTS

APPLICATIONS

APPROACH

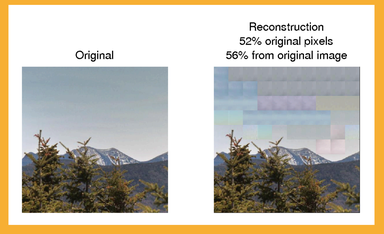
Algorithm 1 Insert Image I into database

1: $Patches \leftarrow$ `CutIntoPatches`(I, patch_size=n)
2: **for** P in $Patches$ **do**
3: $SimPat \leftarrow$`FindLikelySimilarPatches`(P)
4: $P_{closest} \leftarrow argmin\{S(P, P_i)\}$
5: **if then** $S(P, P_i) > T$
6: `insert` P into `patches`

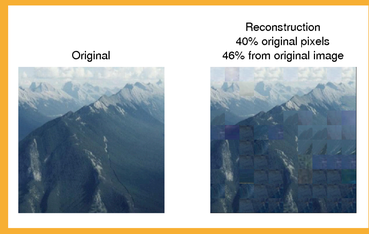
Computing similarity per color channel: $S(P_1, P_2, i) = \frac{||P_1(i) - P_2(i)||^2}{n^2}$

- CIE (LUV) color space
- Threshold per color channel: $S(P_1, P_2, i) < T_i$

The importance of the right color space and a thresholding each color channel:



The importance of the right similarity function and threshold:



The importance of the right patch size: