ColorWorld – a Color-based Image Indexing and Retrieval System

Group No. 3

Ke Wang 903058889

Qichao Chu 903060535

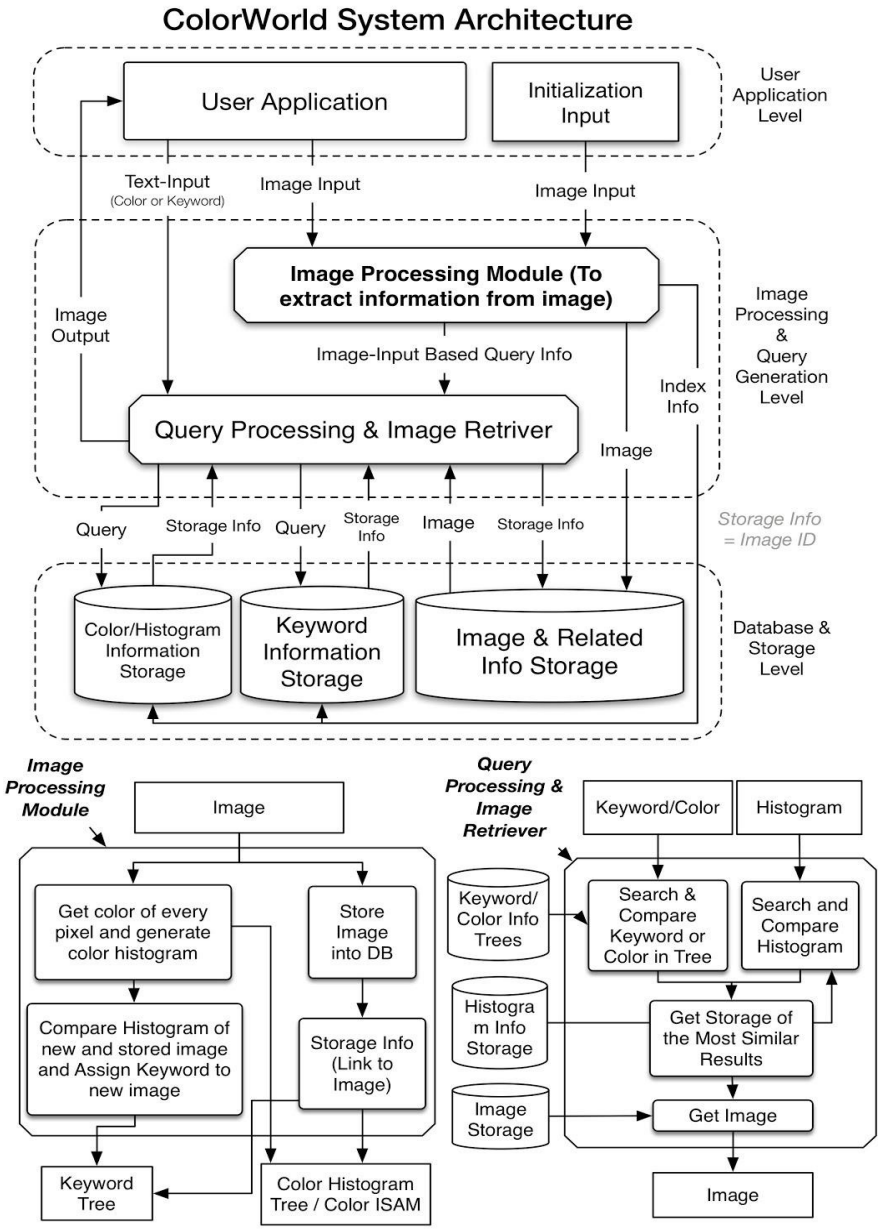
Goal and Idea

- Use color attributes and keywords to retrieve images
  - Use color histogram and color coherence vector
  - Compare the distance to get related images
- Idea
  - Different color space
  - Different distance measurements
  - B+ Tree and Keyword-Tree
- Dataset from Washington University
  - Images with different topics
  - Images with description files
  - Need to combine different format for our use

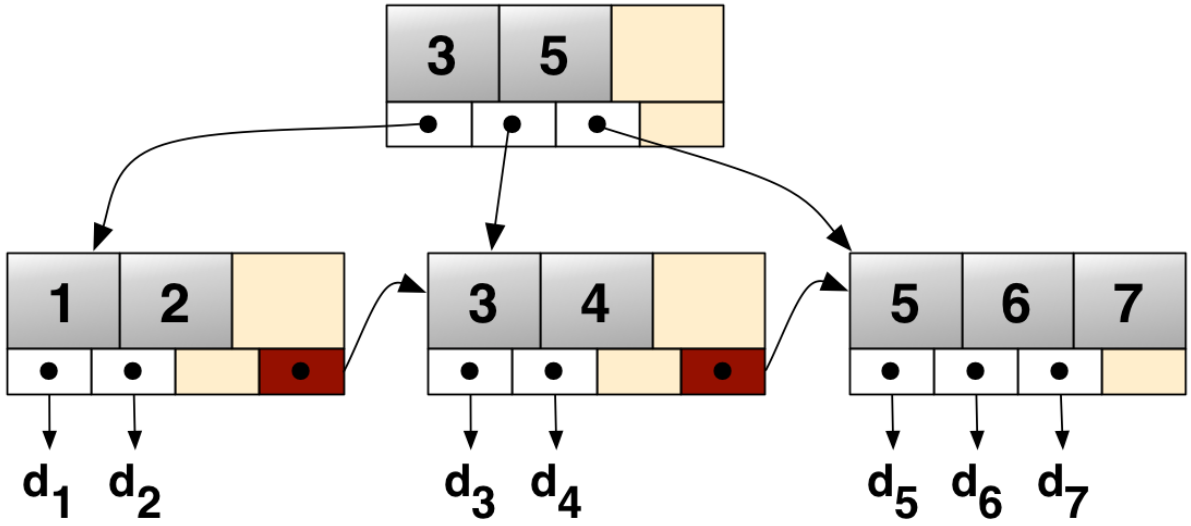
Architecture

Main components:

- Hash Table
  - *Color -> Image ID*
- B+ Tree
  - *Image iD -> Image Desc*
- Keyword Tree
  - *Keyword -> Image ID, Confidence*
- Color Histogram
- Color Coherence Vector

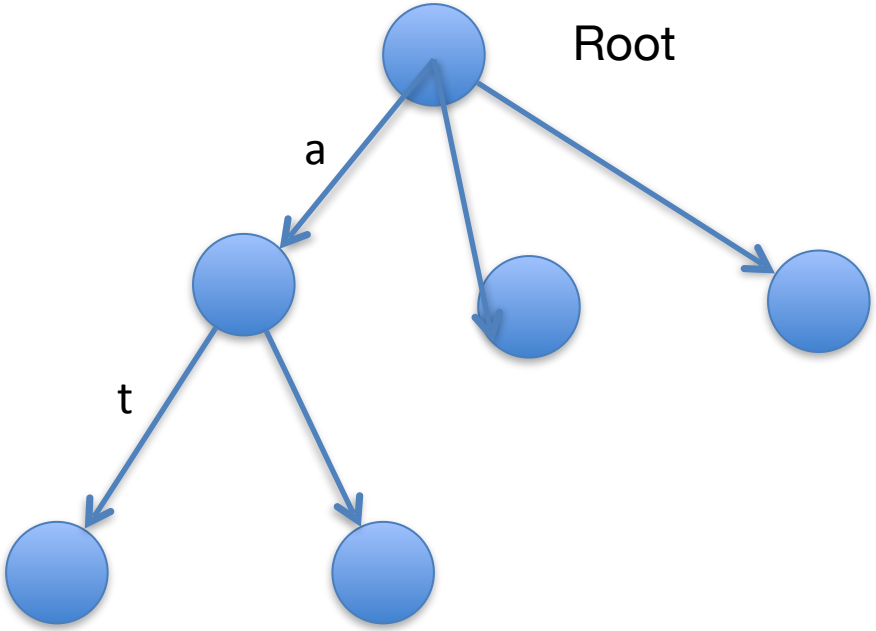


Data Structure - B+ Tree



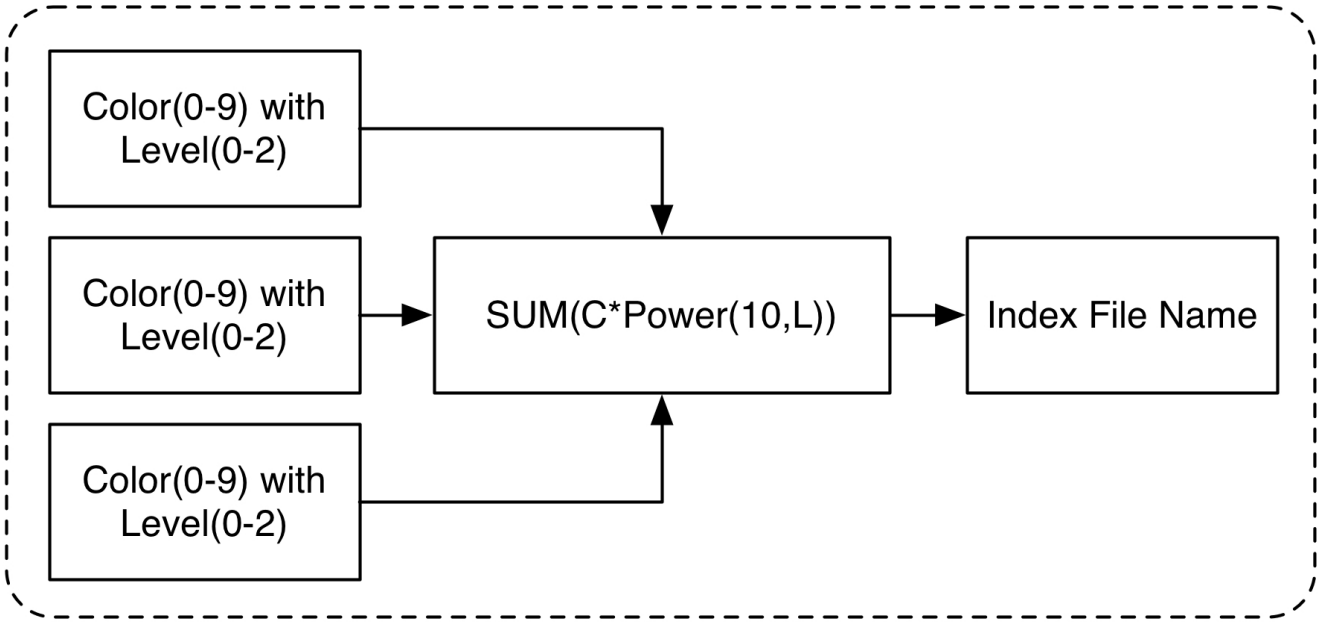
Leaves:
Image ID and its description

Data Structure - Keyword Tree



MAP:
Image ID -> Confidence

Data Structure - Color Hash Table



MAP: Color Info -> Image ID

Color Histogram and CCV

- Color Histogram
  - Choose color space and divide the space
    - RGB -> [8,8,8]
    - HSV -> [16,3,3]
  - int[ ][ ][ ] histogram
    - Denotes how many pixels are in this divided space
  - Compute the distance between two histograms to find relativity
    - Euclidean, Intersection, Quadratic
- CCV Color Coherence Vector
  - Take into consideration coherent pixels and incoherent ones
  - Each bucket -> (a,b)

Accomplishments and TODO

- Accomplishments:
  - Basic indexing (color attributes, keywords) and retrieval (related images, keywords)
- ColorWorld supports:
  - Search by keyword
  - Search by image
  - Vote up or down keyword for an image
- TODO
  - Serialization and deserialization
  - Add new images
  - Assign keyword to new images

Real-World Example