Performance Improvement of Hybrid Content Based Image Retrieval
Devashish Purandare, Amol Rokade, and Prashant Kumar (Group #4)
Sinhgad College of Engineering, Pune, Savitribai Phule Pune University
Under the guidance of: Prof. Girija G. Chiddarwar

Abstract
A Hybrid approach in Content Based Image Retrieval (CBIR) provides higher accuracy as compared to a global approach but at the same time it comes with the high computational complexity of the local approach. Our implementation of the hybrid CBIR system has reduced complexity and better performance than the original approach, using various methods such as caching, parallelism, map-filter, approximation and reducing comparisons. Our results indicate that the performance can be improved by a huge margin without sacrificing accuracy.

Feature extraction & Similarity

- **Global Similarity**
  \[ \text{sim}_{\text{Global}}(P, Q) = \min \left( \text{sim}_{\text{XOR}}(P, Q), \text{sim}_{\text{XOR}}(P, G), \text{sim}_{\text{XOR}}(Q, G) \right) \]

- **Local Similarity**
  \[ \text{sim}_{\text{Local}}(P, Q) = \frac{\text{sim}_{\text{XOR}}(P, Q) - \text{sim}_{\text{XOR}}(P, G) - \text{sim}_{\text{XOR}}(Q, G)}{\text{sim}_{\text{XOR}}(P, G) + \text{sim}_{\text{XOR}}(Q, G) - 100} \]

- **Total Distance**
  \[ \text{dist}(P, Q) = \text{dist}_{\text{XOR}}(P, Q) + \text{dist}_{\text{XOR}}(P, G) \]

Optimizations
- 1. **Caching** - MongoDB
- 2. **Reducing Loops**.
- 3. **Parallelization and multicore processing**.
- 4. **Optimizing loops**.
- 5. **Reducing data size**.
- 6. **Junking lower values**.
- 7. **Reducing unnecessary computations**.
- 8. **Using GPU to speed up the system**.
- 9. **Map - Filter - Reduce to optimize search**.

Pre-processing
- **Original**
- **Threshold**
- **Cropped**
- **Border Extraction**
- **Rotated**
- **Resized**

The System

- **Local Feature Extraction**:
  1. Area division - 16 equal parts.
  2. 2D - Discrete Wavelet Transform.

- **Similarity Measurement**:
  1. XOR for global features.
  2. Taxicab Norm for local features.
  3. Normalization with highest value.
  4. Divide each distance by the maximum distance to find the %age similarity.

Performance Evaluation

1. Image retrieval with very high accuracy.
2. Database and caching improve the performance several times over the original implementation without affecting accuracy.
3. Multi core processing, and other optimizations improve processing significantly over basic implementation.
4. Map-filter can improve performance significantly while slightly reducing accuracy.

Graphical User Interface

Linear Scaling

Unnecessary Calculations

Map - Filter