Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study

Dengsheng Zhang and Guojun Lu

Gippsland School of Computing and Information Technology
Monash University
Churchill, Victoria 3842
Australia
dengsheng.zhang, guojun.lu@infotech.monash.edu.au
http://www.gscit.monash.edu.au/~dengs/
http://www.gscit.monash.edu.au/~guojunl/

Outline

- **Motivations**
- **Comparison of Contour Descriptors**
    - **Fourier Descriptors (FD)**
    - **Curvature Scale Space Descriptors (CSSD)**
    - **Retrieval Experiments**
    - **Discussions**
- **Comparison of Region Descriptors**
    - **Zernike Moments Descriptors (ZMD)**
    - **Grid Descriptors (GD)**
    - **Retrieval Experiments**
    - **Discussions**
- **Conclusions**

Motivations

- Content-Based Image Retrieval
- Investigate Techniques for Shape Representation
- Compare Other Techniques With MPEG-7 Shape Descriptors

FD-I

FD is derived by applying Fourier transform on shape signature, such as centroid distance function r(t).

- **Signature Derivation**

$$r(t) = ([x(t) - x_c]^2 + [y(t) - y_c]^2)^{1/2}, \quad t = 0, 1, \ldots, N-1$$

where $$x_c = \frac{1}{N}\sum_{t=0}^{N-1} x(t) \quad y_c = \frac{1}{N}\sum_{t=0}^{N-1} y(t).$$

and (x(t), y(t)) are shape boundary coordinates

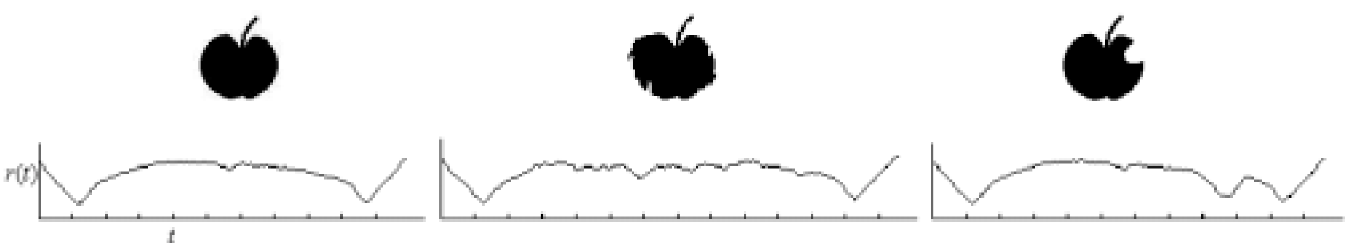


Figure 1. Three apple shapes on the top and their centroid distance signatures on the bottom.

FD-II

- **FD derivation**
  - **The discrete Fourier transform of *r*(*t*) is then given by**

$$a_n = \frac{1}{N}\sum_{t=0}^{N-1} r(t)\exp\left(\frac{-j2\pi nt}{N}\right),\ n = 0, 1, \ldots, N-1$$

    a_n are the Fourier transformed coefficients of r(t).

  - **Translation Normalization.** *a_n is translation normalized due to the translation invariance of r(t)*
  - **Rotation Normalization.** *Ignore the phase information of a_n and only retain the magnitude of a_n*
  - **Scaling Normalization.** *All the other coefficient magnitudes are normalized by $|a_0|$*
  - **Number of FDs Used.** 5, 10, 15, 30, 60 and 90 FDs are tested for retrieval, results show 10 FDs are sufficient for shape representation

- **Similarity Measurement.** *Euclidean distance*

$$d = \left(\sum_{i=1}^{N} | FD_i^Q - FD_i^T |^2\right)^{\frac{1}{2}}$$

CSSD-I

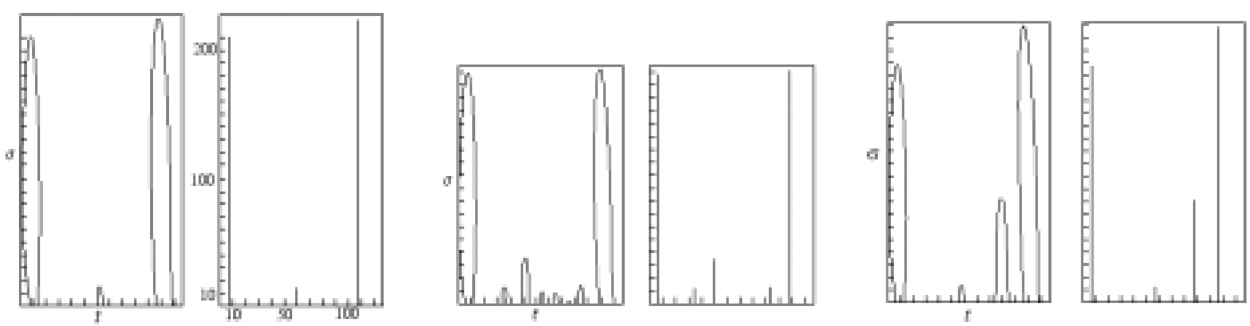
- **Curvature Derivation**

$$k(t) = (\dot{x}(t)\ddot{y}(t) - \ddot{x}(t)\dot{y}(t)) / (\dot{x}^2(t) + \dot{y}^2(t))^{3/2}$$

- **Gaussian Smooth of Curvature**

$$x'(t) = x(t) \otimes g(t,\sigma), \quad y'(t) = y(t) \otimes g(t,\sigma)$$

- **CSS Map Computation**
- **CSS Peak Extraction**



(a) (b) (c)

Figure 2. The CSS contour map and the CSS peaks map of (a) apple 1; (b) apple 2; (c) apple 3.

CSSD-II

- ## Normalization

  - **Translation Normalization.** ***CSSD is translation invariant due to curvature is translation invariant***
  - **Scale Normalization.** ***Normalizing all the shapes into fixed number of boundary points (e.g. 128).***
  - **Rotation Normalization.** ***Circular shifting the highest peak to the origin of the CSS map.***

- ## Similarity measurement

  - ***The similarity between two shapes A and B is measured by the summation of the peak differences between all the matched peaks and the peak values of all the unmatched peaks***
  - ***In order to increase robustness, four schemes of circular shifting matching are applied in order to tolerate variations of peak heights of potential matching peaks: shifting primary peak (the highest peak) of A to match the primary peak of B (other peaks of A are shifted accordingly); shifting primary peak of A to match secondary peak (second highest CSS peak) of B; shifting secondary peak of A to match the primary peak of B; shifting secondary peak of A to match the secondary peak of B.***
  - ***Mirror shape matching***

Retrieval Effectiveness

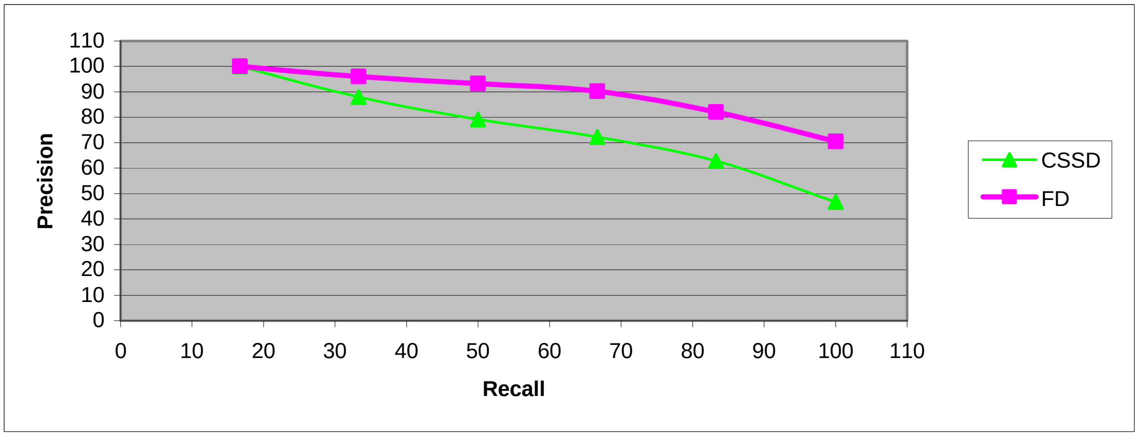
- **Framework. Java-based indexing and retrieval framework**

- **Platform. Widows platform on Pentium III-866 PC**

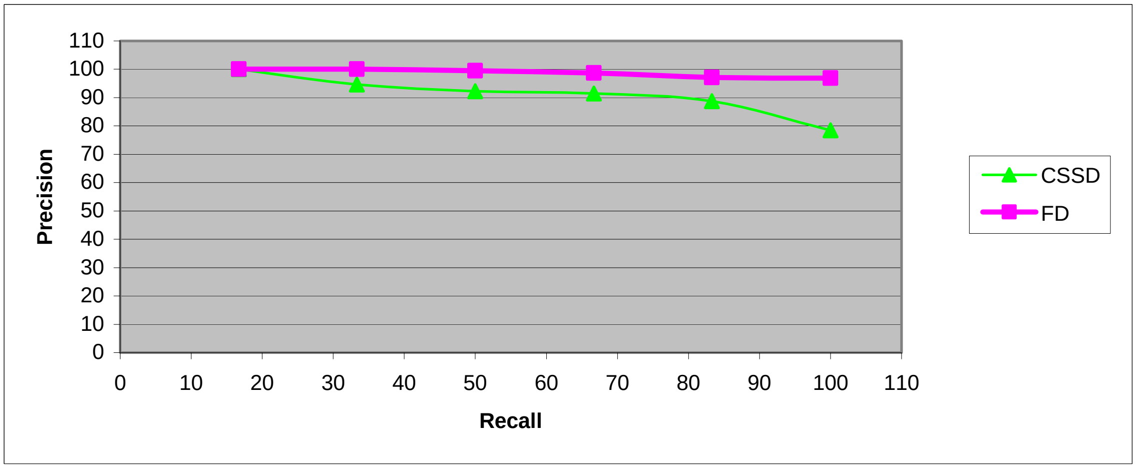
- **Database. MPEG-7 contour shape database—CE-1 Set A1, Set A2 and Set B.**
  Set A1 and Set A2, each consisting of 420 shapes of 70 classes, are for test of scale invariance and rotation invariance respectively. Set B consists of 1400 shapes for test of overall robustness

- **Query Method.** ***All the shapes in Set A1, A2 and B are used as queries***

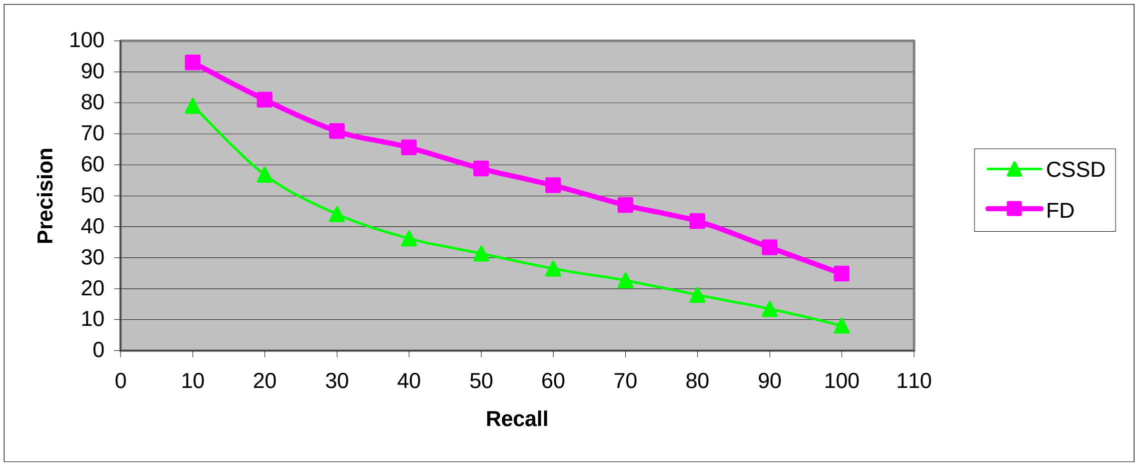
- **Evaluation Method. Precision and recall.**
  Precision: the ratio of the number of retrieved relevant shapes to the total number of retrieved shapes; Recall: the ratio of the number of retrieved relevant images to the total number of relevant shapes in the whole database. For each query, the precision of the retrieval at each level of the recall is obtained. The result precision of retrieval is the average precision of all the query retrievals.



(a)



(b)



(c)

Figure 3. Average precision and recall of retrieval using FD and CSSD on (a) Set A1; (b) Set A2; (c) Set B.

Computation Efficiency

- **Platform.** Windows platform on Pentium III-866 PC.
- **Database.** Set B of MPEG-7 contour shape database

Table 1. The elapsed time of feature extraction and retrieval for 1400 shapes

Time / Shape descriptors	Total time of feature extraction of 1400 shapes	Average time of feature extraction of each shape	Total time of retrieval of 1400 queries	Average time of retrieval of each query
FD	80960 ms	57.8 ms	54150 ms	38.6 ms
CSSD	120629 ms	86.1 ms	317178 ms	226.5 ms

Discussions—Contour-based Shape Descriptors

- **Feature Domains.** ***FD is obtained from spectral domain while CSSD is obtained from spatial domain.***

- **Dimensions.** ***Dimension of FD feature is constant, while dimension of CSSD feature varies for each shape.***

- **Computation Complexity.** ***The computation process of CSSD is more complex than that of FD. Computation of CSSD has an extra process of scaling normalization before CSSD extraction, and the extraction of the CSSD feature takes two processes, i.e., CSS map computation and CSS peaks extraction.***

- **Online Matching Computation.** ***The online matching of two sets of FDs is simply the Euclidean distance between two feature vectors of 10 dimensions. The online matching of two sets of CSSD involves at least 8 schemes of circular shift matching, and for each scheme of circular shift matching, it involves 6 shifts and the Euclidean distance calculation between two feature vectors of 6 dimensions.***

- **Type of Features Captured.** ***FD captures both global features and local features while CSSD does not capture global features.***

- **Parameters or Thresholds Influence.** ***For FD, the parameter is the number of FDs, which is predictable. For CSSD, the parameters are the number of sampling points, the threshold to eliminate short peaks and the tolerance value for peak position matching, which are determined empirically. In MPEG-7 [ISO00], four more empirical factors are used to implement this technique: the scale factor and the exponential factor for the peak transformation, the threshold to eliminate all the peaks in a shape and the database dependent value used for peak normalization.***

- **Hierarchical Representation.** ***FD supports hierarchical coarse to fine representation while CSSD does not. In order to support hierarchical representation, CSSD has to incorporate shape global features such as eccentricity and circularity which are unreliable.***

- **Suitability for Efficient Indexing.** ***FD is suitable to be organized into efficient data structure, while CSSD is not, due to its variable dimensions and complex distance calculation.***

ZMD-I

- **Zernike Polynomials:**

$$V_{nm}(x, y) = V_{nm}(\rho\cos\theta, \rho\sin\theta) = R_{nm}(\rho)\exp(jm\theta)$$

where

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!}\rho^{n-2s}$$

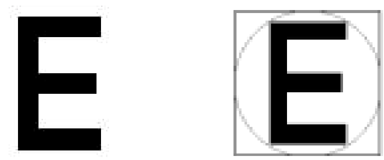
and ρ is the radius from (x, y) to the shape centroid, θ is the angle between ρ and x axis, n and m are integers and subject to n-|m| = even, |m|≤n. Zernike polynomials are a complete set of complex-valued function orthogonal over the unit disk, i.e., $x^2 + y^2 = 1$.

- **Complex Zernike Moments of Order *n* with Repetition *m***

$$A_{nm} = \frac{n+1}{\pi}\sum_x\sum_y f(x, y)V_{nm}^*(x, y), \qquad x^2 + y^2 \le 1$$

ZMD-II

- **Normalisation**
  - **Scale Normalization.** ***Scale shape into unit disk***

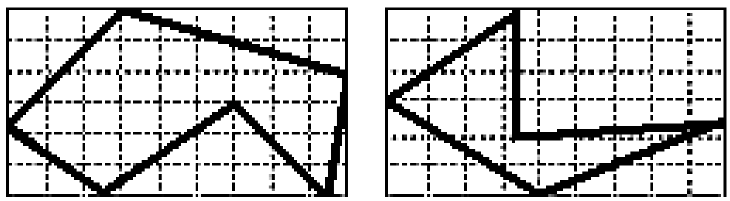


  - **Translation Normalization.** ***Shift shape centroid to the center of the unit disk***
  - **Rotation Normalization.** ***Ignore the phase information of A_{mn}, and only retain the magnitude of A_{mn}***
  - **Magnitude Normalization.** ***Magnitudes are normalized into* (0, 1) *with the shape mass***
  - **Number of ZMDs Used. 36 *ZMDs are used in accordance with MPEG-7 documents***
- **Similarity Measurement.** ***Euclidean distance***

$$d = \left(\sum_{i=1}^{N} | ZMD_i^Q - ZMD_i^T |^2\right)^{\frac{1}{2}}$$

GD-I

- **Grid Overlay.** ***Project shape onto a grid of fixed size (e.g.16×16) grid cells***



- **Cell Value Assignation.** ***Assign value of 1 to a cell if is covered by the shape and 0 if not***

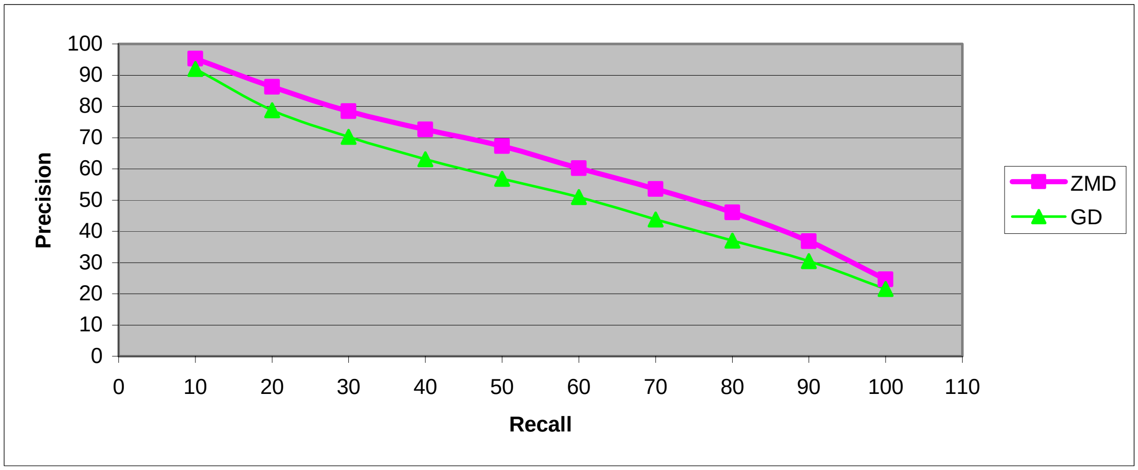
- **Acquire Shape Number.** ***A binary sequence is obtained by scanning the grid in left-right and top-bottom order. The shape numbers of the above two shapes are 001111000 011111111 111111111 111111111 111110011 001100011 and 001100000 011100000 111100000 111100000 011111100 000111000 respectively.***

GD-II

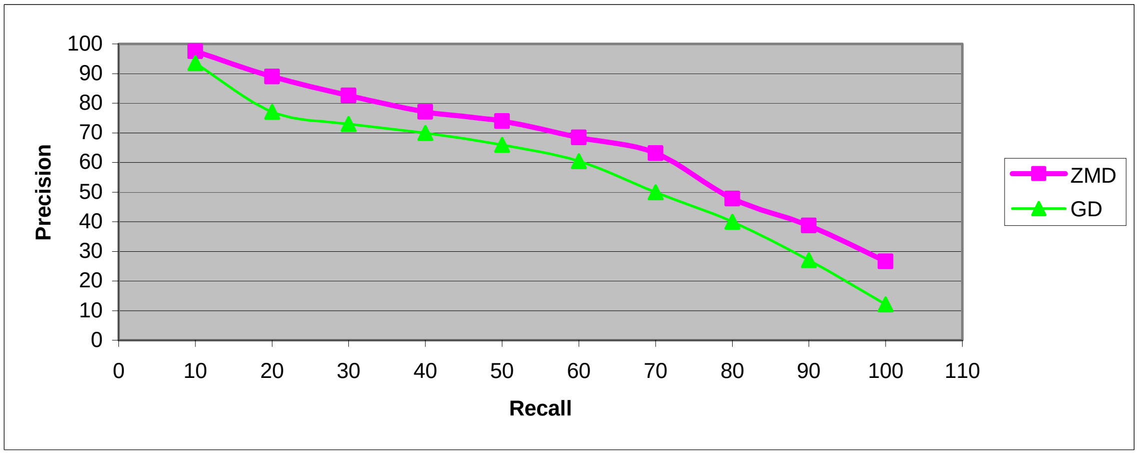
- **Normalization**
    - **Scaling Normalization.** ***Scale shape into fixed size bounding rectangle***
    - **Rotation Normalization.** ***Rotate shape major axis to horizontal direction***
    - **Translation Normalization.** ***Translate the rotated shape into the uppermost part of the bounding rectangle***
- **Similarity measurement**
    - **City Block Distance.** ***XOR operation on the two sets of shape numbers***

Retrieval Effectiveness

- **Framework.** ***Java-based indexing and retrieval framework***
- **Platform.** ***Widows platform on Pentium III-866 PC***
- **Databases.** ***Set B of MPEG-7 contour shape database (CE-1B) and MPEG-7 region shape database (CE-2). MPEG-7 region shape database consists of 3621 shapes of mainly trademarks***
- **Query Method.** ***All the 1400 shapes in CE-1B are used as queries; 31 classes (each has 21 members) of shapes from CE-2 are used as queries***
- **Evaluation Method.** ***Precision and recall***



(a)



(b)

Figure 4. (a) Retrieval performance of ZMD and GD on contour-based shapes; (b) Retrieval performance of ZMD and GD on region-based shapes.

Computation Efficiency

- **Platform.** ***Windows platform on Pentium III-866 PC***
- **Database.** ***MPEG-7 region-based shape database CE-2***

Table 2. Time info of feature extraction and retrieval of CE-2 shapes using region shape descriptors.

Time / Shape descriptors	Total time of feature extraction of 3621 shapes	Average time of feature extraction of each shape	Total time of retrieval of 651 queries	Average time of retrieval of each query
ZMD	4325010 ms	1194.4 ms	63854 ms	98 ms
GD	2628034 ms	725.7 ms	729909 ms	1121.2ms

Discussions—Region-based Shape Descriptors

- **Applications.** ***Both ZMD and GD are application independent.***
- **Feature domains.** ***ZMD is extracted from spectral domain while GD is extracted from spatial domain.***
- **Compactness.** ***The dimension of ZMD is low while that of GD is high.***
- **Robustness.** ***ZMD is the more robust to shape variations than GD.***
- **Computation complexity.** ***Both the offline and online computation of GD are more expensive than ZMD.***
- **Accuracy.** ***At the same level of recall, the retrieval precision of ZMD is higher than that of GD.***
- **Hierarchical representation.** ***Both ZMD and GD support hierarchical representation. The number of ZMDs can be adjusted to meet hierarchical requirement. For GD, hierarchical representation can be achieved by adjusting the cell size or combined with eccentricity and circularity. GMD does not support hierarchical representation because higher geometric moment invariants are difficult to obtain.***
- **Agree with human intuition.** ***GD is a more intuitive shape representation than ZMD.***

Conclusions—Contour-based Shape Descriptors

- **Drawbacks of CSSD compared with FD**
  - ***CSSD is only robust to local boundary variations, it's not robust in global sense***
  - ***The low dimension advantage is offset by its complex matching***
  - ***The retrieval performance of CSSD is low, the overall precision of similarity-based retrieval (on Set B of MPEG-7 contour shape database) only achieves 33.6***
  - ***CSSD does not support hierarchical representation. In order to support hierarchical representation, it has to incorporate other global shape features***
  - ***The representation and retrieval performance depend on empirical factors such as the number of sample points on the boundary, the threshold to eliminate short peaks and the tolerance value for peak position matching, etc.***
  - ***CSSD is not suitable for efficient indexing due to the expensive matching and variation of feature dimensions***
  - ***Based on these facts, we recommend that FD be included as one of the contour shape descriptors in MPEG-7***

Conclusions—Region-based shape descriptors

- ***In terms of low computation complexity, compact representation, robustness and retrieval performance, Zernike moments descriptors (ZMD) is more suitable for region-based shape retrieval than GD***

- ***GD agrees more with human intuition***