Color for CBIR

Wei-Ta Chu

J.R. Smith, Chapter 11 of Image Database: Search and Retrieval of Digital Imagery, edited by V. Castelli and L.D. Bergman, John Wiley & Sons, 2002.

Color Descriptor Metrics

Table 11.1. Summary of the Eight Color Histogram Descriptor Metrics ($\mathcal{D}1 - \mathcal{D}8$)

Metric	Description	Category
$\mathcal{D}1$	Histogram L_1 distance	Minkowski-form ($r = 1$)
$\mathcal{D}2$	Histogram L_2 distance	Minkowski-form ($r = 2$)
$\mathcal{D}3$	Binary set Hamming distance	Binary Minkowski-form ($r = 1$)
$\mathcal{D}4$	Histogram quadratic distance	Quadratic-form
$\mathcal{D}5$	Binary set quadratic distance	Binary quadratic-form
$\mathcal{D}6$	Histogram Mahalanobis distance	Binary quadratic-form
$\mathcal{D}7$	Histogram mean distance	First moment
$\mathcal{D}8$	Histogram moment distance	Higher moments

Metric Properties

- A metric is a dissimilarity (distance) measure that satisfies the following properties
  - Nonnegativity: $D(\boldsymbol{a}, \boldsymbol{b}) \geq 0$
  - Reflexivity: $D(\boldsymbol{a}, \boldsymbol{b}) = 0$ if and only if $\boldsymbol{a} = \boldsymbol{b}$
  - Symmetry: $D(\boldsymbol{a}, \boldsymbol{b}) = D(\boldsymbol{b}, \boldsymbol{a})$
  - Triangle inequality: $D(\boldsymbol{a}, \boldsymbol{b}) + D(\boldsymbol{b}, \boldsymbol{c}) \geq D(\boldsymbol{a}, \boldsymbol{c})$
  - E.g. Euclidean distance in *d* dimensions

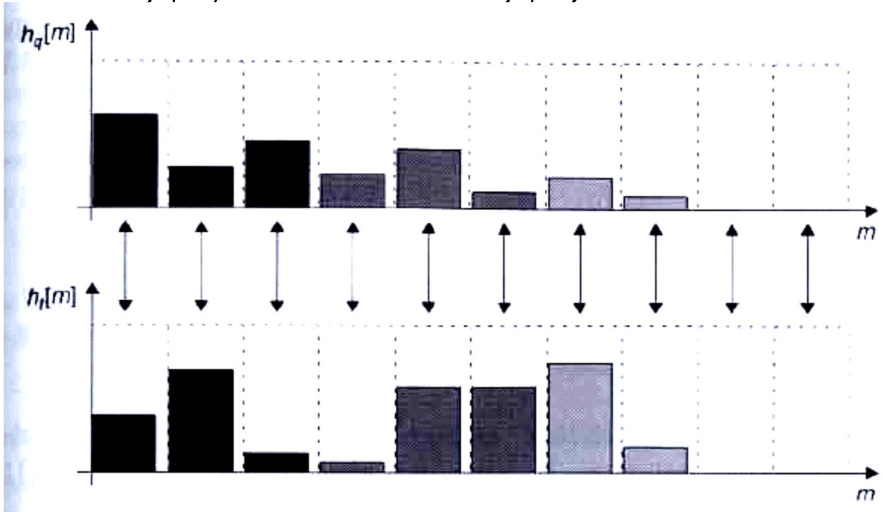
$$D(\boldsymbol{a}, \boldsymbol{b}) = \left(\sum_{k=1}^{d} (a_k - b_k)^2\right)^{\frac{1}{2}}$$

Minkowski-Form Metrics

- The first category of metrics for color histogram descriptors is based the Minkowski-form metric.

$$d^r_{q,t} = \sum_{m=0}^{M-1} |h_q(m) - h_t(m)|^r$$

- Only account for a specific color to the proportion of the same color, but not to the proportions of other similar colors.
- d(a dark red image, a lighter red image) maybe the same as d(a dark red image, a blue image)



Histogram Intersection (D1)

- If histograms are not normalized

$$d'_{q,t} = 1 - \frac{\sum_{m=0}^{M-1} \min[h_q(m), h_t(m)]}{\min(|\mathbf{h}_q|, |\mathbf{h}_t|)}$$

- If histograms are normalized, $|\mathbf{h}_q|=|\mathbf{h}_t|$

$$\mathcal{D}1(q,t) = \sum_{m=0}^{M-1} |h_q(m) - h_t(m)|$$

  - It's recognized as the Minkowski-form metric with $r = 1$ and is commonly known as the "walk" or "city block" distance.

Histogram Euclidean Distance (D2)

- The Minkowski-form metric with $r = 2$

$$\mathcal{D}2(q,t) = D2^2 = (\mathbf{h}_q - \mathbf{h}_t)^T(\mathbf{h}_q - \mathbf{h}_t) = \sum_{m=0}^{M-1}[h_q(m) - h_t(m)]^2$$

- The Euclidean distance can be decomposed as

$$\mathcal{D}2(q,t) = \mathbf{h}_q^T\mathbf{h}_q + \mathbf{h}_t^T\mathbf{h}_t - 2\mathbf{h}_q^T\mathbf{h}_t. \tag{11.14}$$

Given the following normalization of the histograms, $\|\mathbf{h}_q\| = \mathbf{h}_q^T\mathbf{h}_q = 1$ and $\|\mathbf{h}_t\| = \mathbf{h}_t^T\mathbf{h}_t = 1$, the Euclidean distance is given by

$$\mathcal{D}2(q,t) = 2 - 2\mathbf{h}_q^T\mathbf{h}_t. \tag{11.15}$$

- D2 can be derived from the inner product of the query $\mathbf{h}_q$ and target color histogram $\mathbf{h}_t$.

Binary Set Hamming Distance (D3)

- Binary sets count the number of colors with a frequency of occurrence within the image exceeding a predefined threshold.
- A binary set **s** is an M-dimensional binary vector with an *i*-th entry equal to 1 if the *i*-th entry of the color histogram **h** exceeds T and equal to zero otherwise.
- Binary sets indicate the presence of each color but do not indicate an accurate degree of presence.

$$\mathcal{D}3(q,t) = \frac{|\mathbf{s}_q - \mathbf{s}_t|}{|\mathbf{s}_q||\mathbf{s}_t|}$$

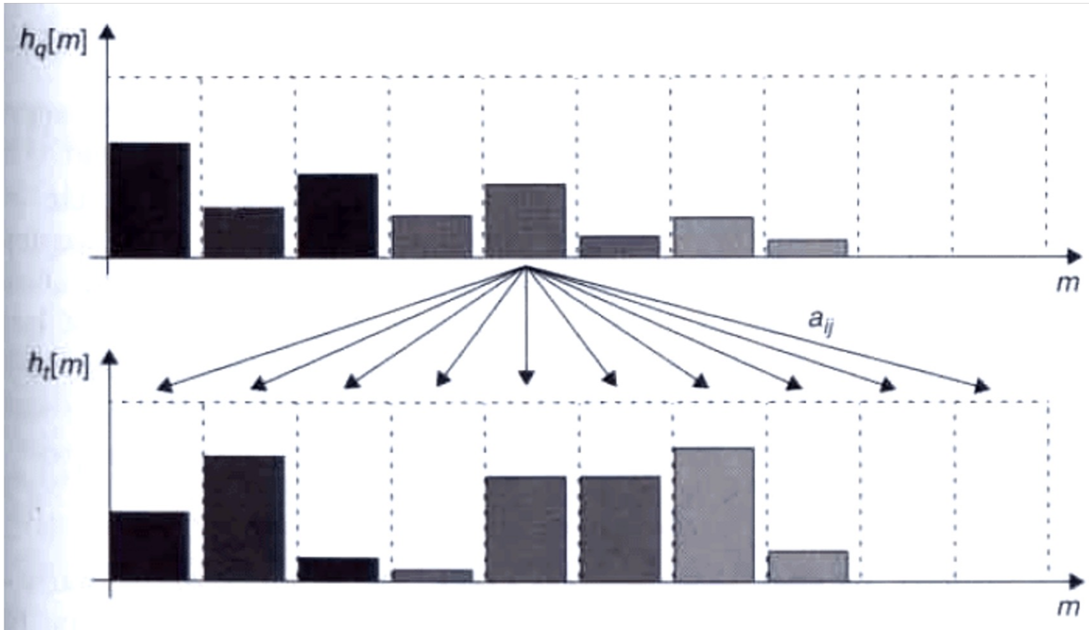
|.| denotes the sum of the elements of the vector.

$$\mathcal{D}3(q,t)|\mathbf{s}_q||\mathbf{s}_t| = \mathbf{s}_q \odot \mathbf{s}_t$$

Exclusive OR

Quadratic-Form Metrics

- Quadratic-form metrics consider the cross-relation of the bins.
- The quadratic-form metrics compare all bins and weight the inter-element distance by pairwise weighting factors.



Histogram Quadratic Distance Measure (D4)

- The quadratic-form distance between color histograms $\mathbf{h}_q$ and $\mathbf{h}_t$ is given by

$$\mathcal{D}4(q, t) = D4^2 = (\mathbf{h}_q - \mathbf{h}_t)^T \mathbf{A} (\mathbf{h}_q - \mathbf{h}_t)$$

- $A=[a_{ij}]$, and a_{ij} denotes the similarity between histogram bins with indices i and j.
- In a naïve implementation, the color histogram quadratic distance is computationally expensive.

Histogram Quadratic Distance Measure (D4)

$$\mathcal{D}4(\mathrm{q},\mathrm{t}) = D4^2 = (\mathbf{h}_\mathrm{q} - \mathbf{h}_\mathrm{t})^T \mathbf{A} (\mathbf{h}_\mathrm{q} - \mathbf{h}_\mathrm{t})$$

1 by N N by N N by 1

$$\begin{bmatrix} d_1 & d_2 & \cdots & d_N \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & & & \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} = \begin{bmatrix} d'_1 & d'_2 & \cdots & d'_N \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix}$$

$$d'_1 = d_1 a_{11} + d_2 a_{21} + d_3 a_{31} + \cdots + d_N a_{N1}$$

Example of Quadratic Distance

- Histogram Quadratic Distance $d^{hist}_{q,t} = (\mathbf{h}_q - \mathbf{h}_t)^t \mathbf{A}(\mathbf{h}_q - \mathbf{h}_t)$

$$\mathbf{A}_{\text{red, orange, blue}} = \begin{bmatrix} 1.0 & 0.9 & 0.0 \\ 0.9 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$$

- Consider a pure red image **x**=[1.0, 0.0, 0.0]^T, and a pure orange image **y**=[0.0, 1.0, 0.0]^T.
- The quadratic distance between **x** and **y** is 0.2.
- The Euclidean distance between **x** and **y** is $\sqrt{2}$.

Hafner, et al. , "Efficient Color Histogram Indexing for Quadratic Form Distance Functions," IEEE Trans. on PAMI, vol. 17, no. 7, pp. 729-736, Jul., 1995

Binary Set Quadratic Distance (D5)

- The quadratic form between two binary descriptor sets $\mathbf{s}_q$ and $\mathbf{s}_t$ is given by

$$\mathcal{D}5(q,t) = D5^2 = (\mathbf{s}_q - \mathbf{s}_t)^T \mathbf{A} (\mathbf{s}_q - \mathbf{s}_t).$$

Histogram Mahalanobis Distance (D6)

- A special case of the quadratic-form metric in which the transform matrix A is given by the covariance matrix, that is $A = \sum^{-1}$.
- Idea: colors that are widely prevalent across all images are not likely to help in discriminating among different images.

Histogram Mahalanobis Distance (D6)

the Mahalanobis distance, the color histogram descriptor vectors are treated as random variables $\mathbf{X} = [x_0, x_1, \cdots, x_{M-1}]$. Then, the *correlation* matrix is given by $R = [r_{ij}]$, where $r_{ij} = \mathbf{E}\{x_i x_j\}$. In this notation, $\mathbf{E}\{Y\}$ denotes the expected value of the random variable Y. The *covariance* matrix is given by $\mathbf{\Sigma} = [\sigma_{ij}^2]$, where $\sigma_{ij}^2 = r_{ij} - \mathbf{E}\{x_i\}\mathbf{E}\{x_j\}$.

The Mahalanobis distance between color histograms is obtained by letting $\mathbf{X}_q = \mathbf{h}_q$ and $\mathbf{X}_t = \mathbf{h}_t$, which gives

$$D6(q,t) = D6^2 = (\mathbf{X}_q - \mathbf{X}_t)^T \mathbf{\Sigma}^{-1} (\mathbf{X}_q - \mathbf{X}_t). \tag{11.25}$$

In probability theory and statistics, covariance is a measure of how much two variables change together.

Covariance Matrix

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \qquad \Sigma_{ij} = \mathrm{E}\big[(X_i - \mu_i)(X_j - \mu_j)\big] \qquad \mu_i = \mathrm{E}(X_i)$$

Covariance matrix

$$\Sigma = \begin{bmatrix} \mathrm{E}[(X_1 - \mu_1)(X_1 - \mu_1)] & \mathrm{E}[(X_1 - \mu_1)(X_2 - \mu_2)] & \cdots & \mathrm{E}[(X_1 - \mu_1)(X_n - \mu_n)] \\ \mathrm{E}[(X_2 - \mu_2)(X_1 - \mu_1)] & \mathrm{E}[(X_2 - \mu_2)(X_2 - \mu_2)] & \cdots & \mathrm{E}[(X_2 - \mu_2)(X_n - \mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n - \mu_n)(X_1 - \mu_1)] & \mathrm{E}[(X_n - \mu_n)(X_2 - \mu_2)] & \cdots & \mathrm{E}[(X_n - \mu_n)(X_n - \mu_n)] \end{bmatrix}.$$

Histogram Mahalanobis Distance (D6)

In the special case when the bins of the color histogram, x_i, are uncorrelated, that is, when all the covariances $r_{ij} = 0$ when $i \neq j$, $\mathbf{\Sigma}$ is a diagonal matrix [35]:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_0^2 & & & 0 \\ & \sigma_1^2 & & \\ & & \ddots & \\ 0 & & & \sigma_{M-1}^2 \end{bmatrix} \tag{11.26}$$

In this case, the Mahalanobis distance reduces to

$$\mathcal{D}6(q,t) = \sum_{m=0}^{M-1} \left[\frac{x_q(m) - x_t(m)}{\sigma_m} \right]^2, \tag{11.27}$$

which is a weighted Euclidean distance. When the x_i are not uncorrelated, it is

Mean Color Distance (D7)

- Compute from the mean of the color histogram of each of the color channels.
- Mean color descriptor: $\boldsymbol{v} = (\overline{r}, \overline{g}, \overline{b})$

$$\mathcal{D}7 = (\mathbf{v}_q - \mathbf{v}_t)^T (\mathbf{v}_q - \mathbf{v}_t)$$ (inner product, Euclidean distance)

Color Moment Distance (D8)

- The color moment descriptor based on variance can be represented by $\boldsymbol{\sigma}^2 = (\sigma_r^2, \sigma_g^2, \sigma_b^2)$
- The color moment distance is given by

$$\mathcal{D}8 = (\boldsymbol{\sigma}_q^2 - \boldsymbol{\sigma}_t^2)^T (\boldsymbol{\sigma}_q^2 - \boldsymbol{\sigma}_t^2)$$

Evaluation

- Testbed
  - N=3100 color photographs, including animals, sports, scenery, people, and so forth.
  - Four benchmark queries, including sunsets, flowers, nature, and lions
  - Relevance score:
    - 1 if it belonged to the same class as the query image
    - 0.5 if partially relevant to the query image
    - 0 otherwise

Evaluation

- A benchmark query is issued to the system, the system retrieves the image in rank order, then for each cutoff value *k*, the following values are computed.

True positive
- $A_k = \sum_{n=1}^{k} V_n$, is the number of relevant results returned among the top k,

False positive
- $B_k = \sum_{n=1}^{k}(1 - V_n)$, is the number of irrelevant results returned among the top k,

False negative
- $C_k = \sum_{n=k+1}^{N} V_n$, is the number of relevant results not returned among the top k,

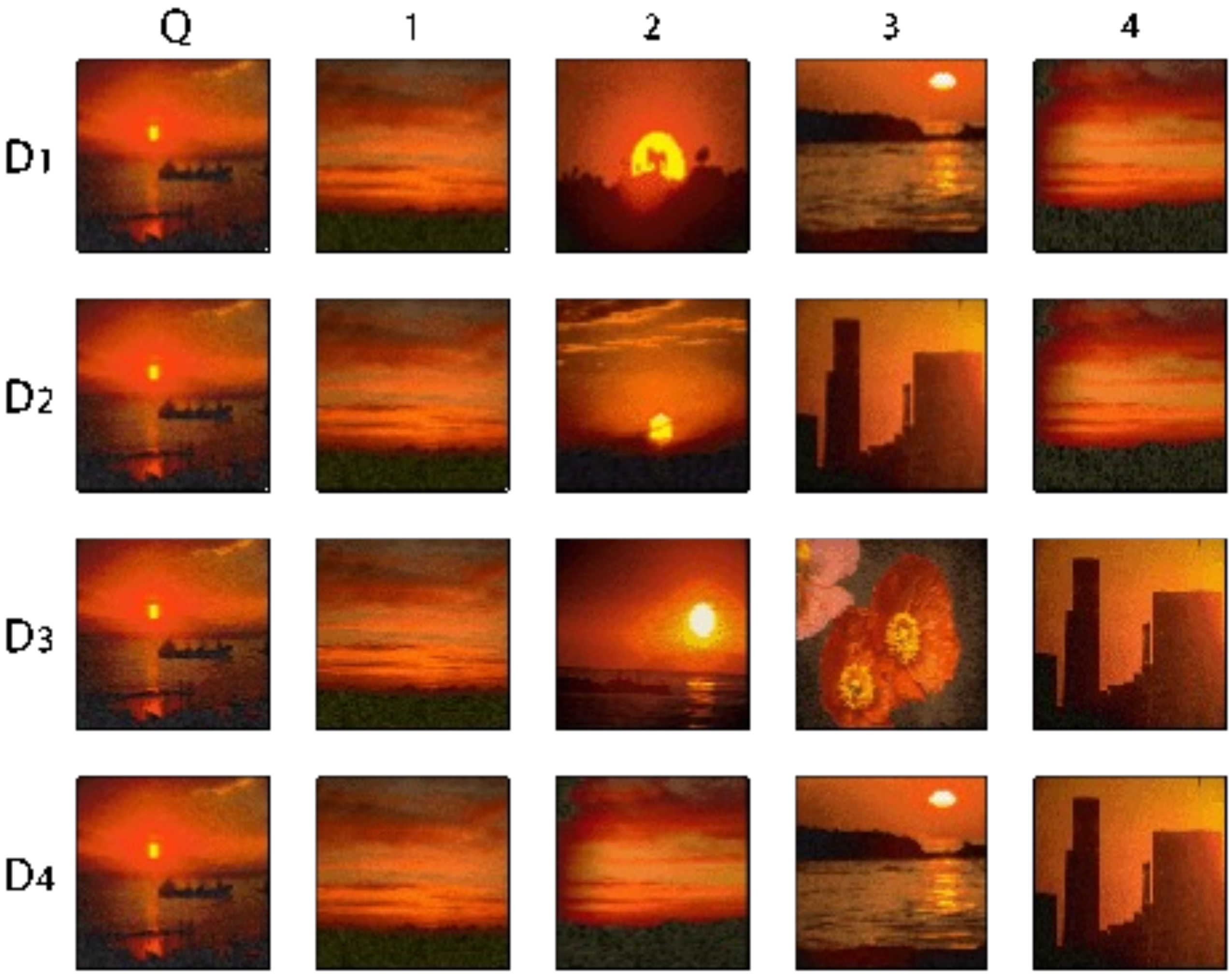
True negative
- $D_k = \sum_{n=k+1}^{N}(1 - V_n)$, is the number of irrelevant results not returned among the top k.

Evaluation

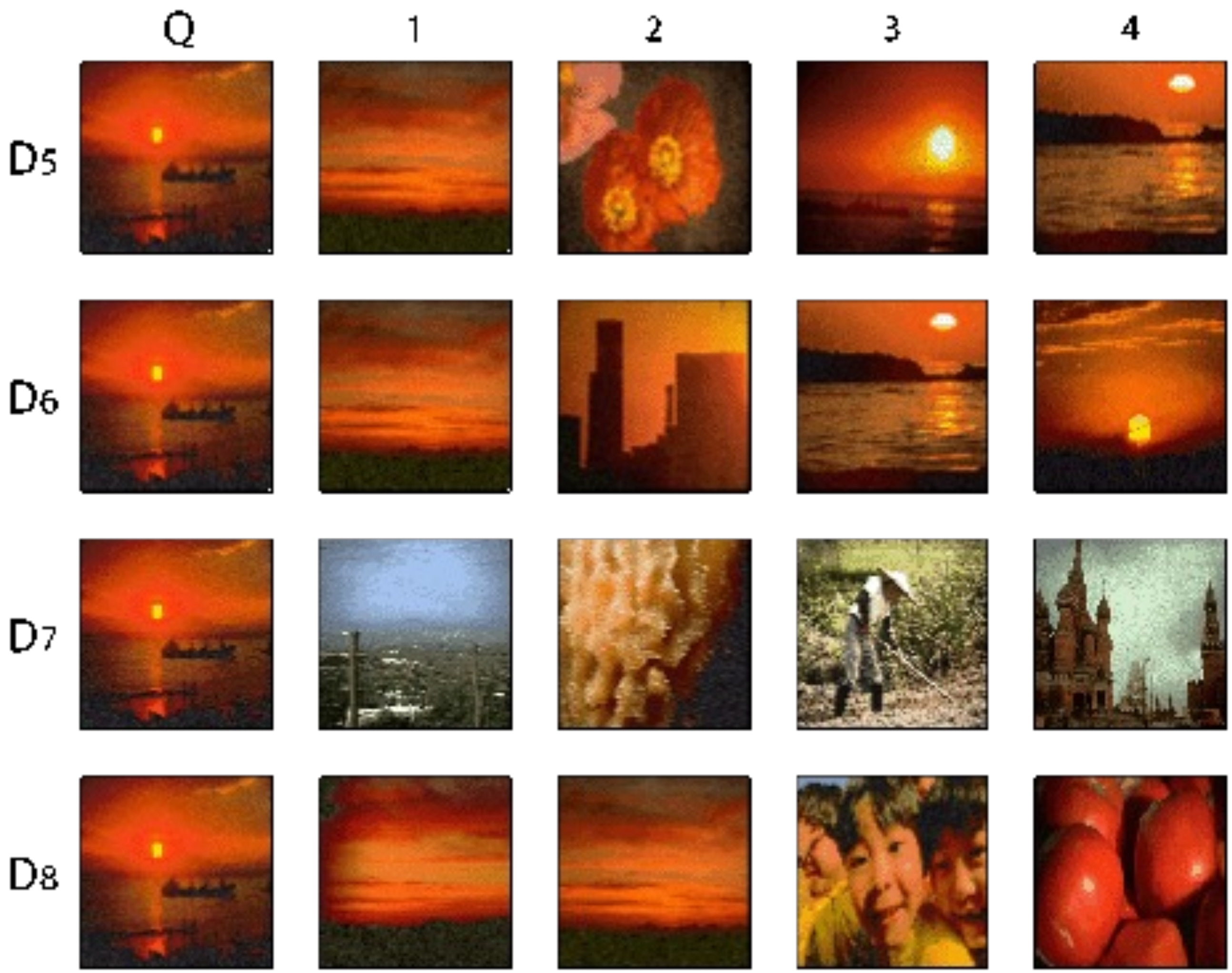
- Recall: $R_k = \dfrac{A_k}{A_k + C_k}$ indicates the proportion of desired results that are returned among the k best matches;
- Precision: $P_k = \dfrac{A_k}{A_k + B_k}$ measures the efficiency with which the relevant items are returned among the best k matches;
- Fallout: $F_k = \dfrac{B_k}{B_k + D_k}$ measures the efficiency of rejecting nonrelevant items.

(False positive rate)

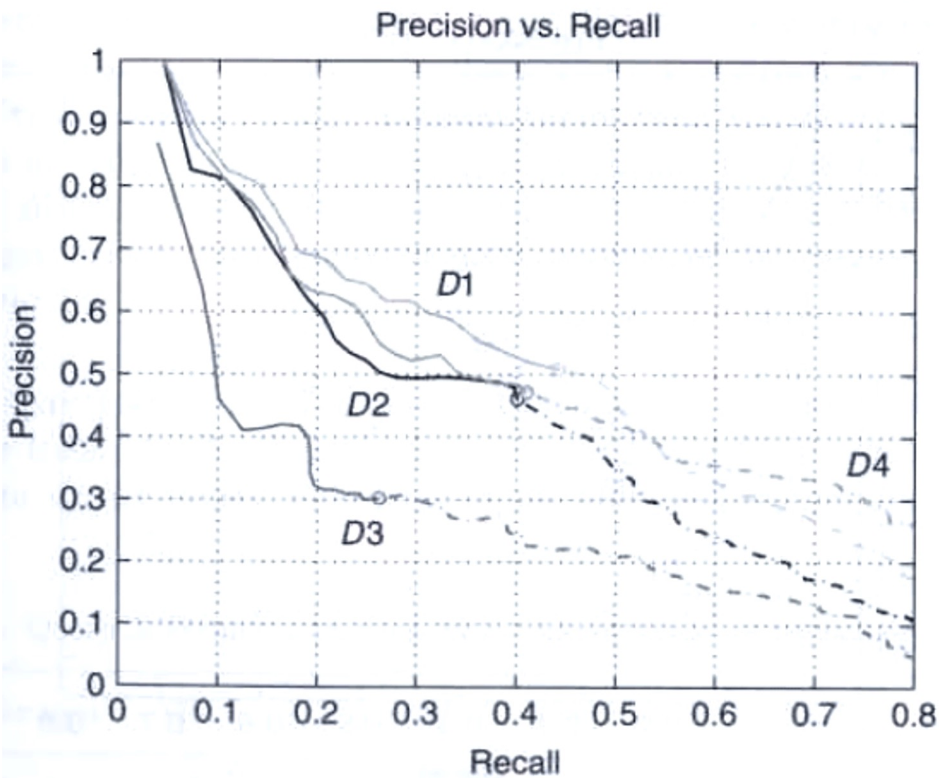
Query 1: Sunsets

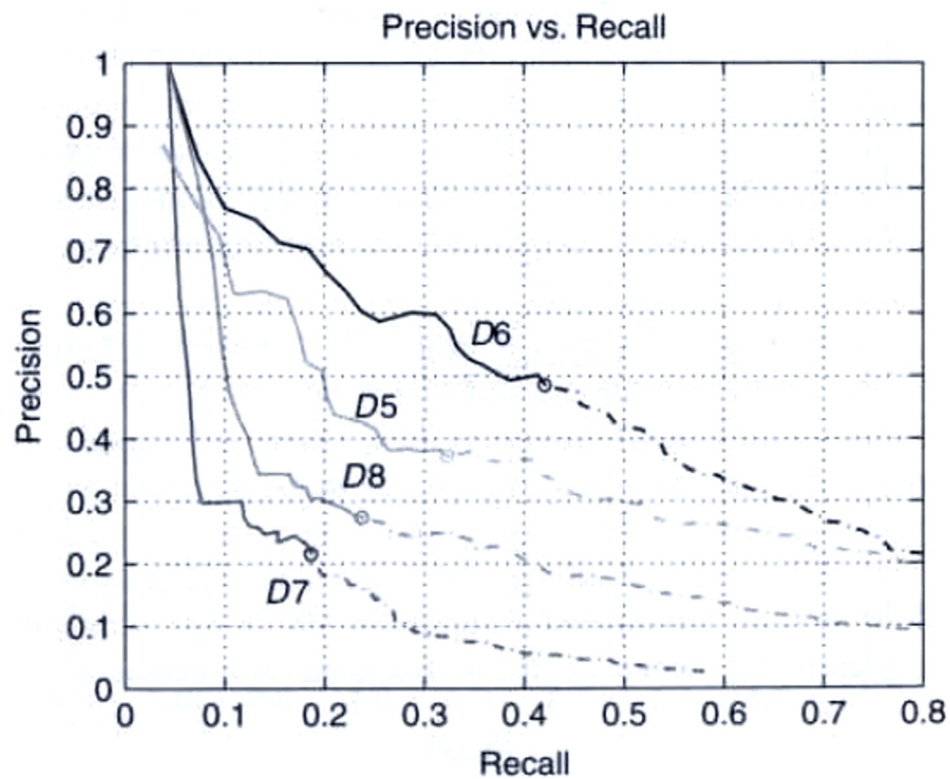


Query 1: Sunsets



Performance -- Sunsets





D7 has the worst performance. Using the mean of each color channel did not sufficiently capture color information in the images.

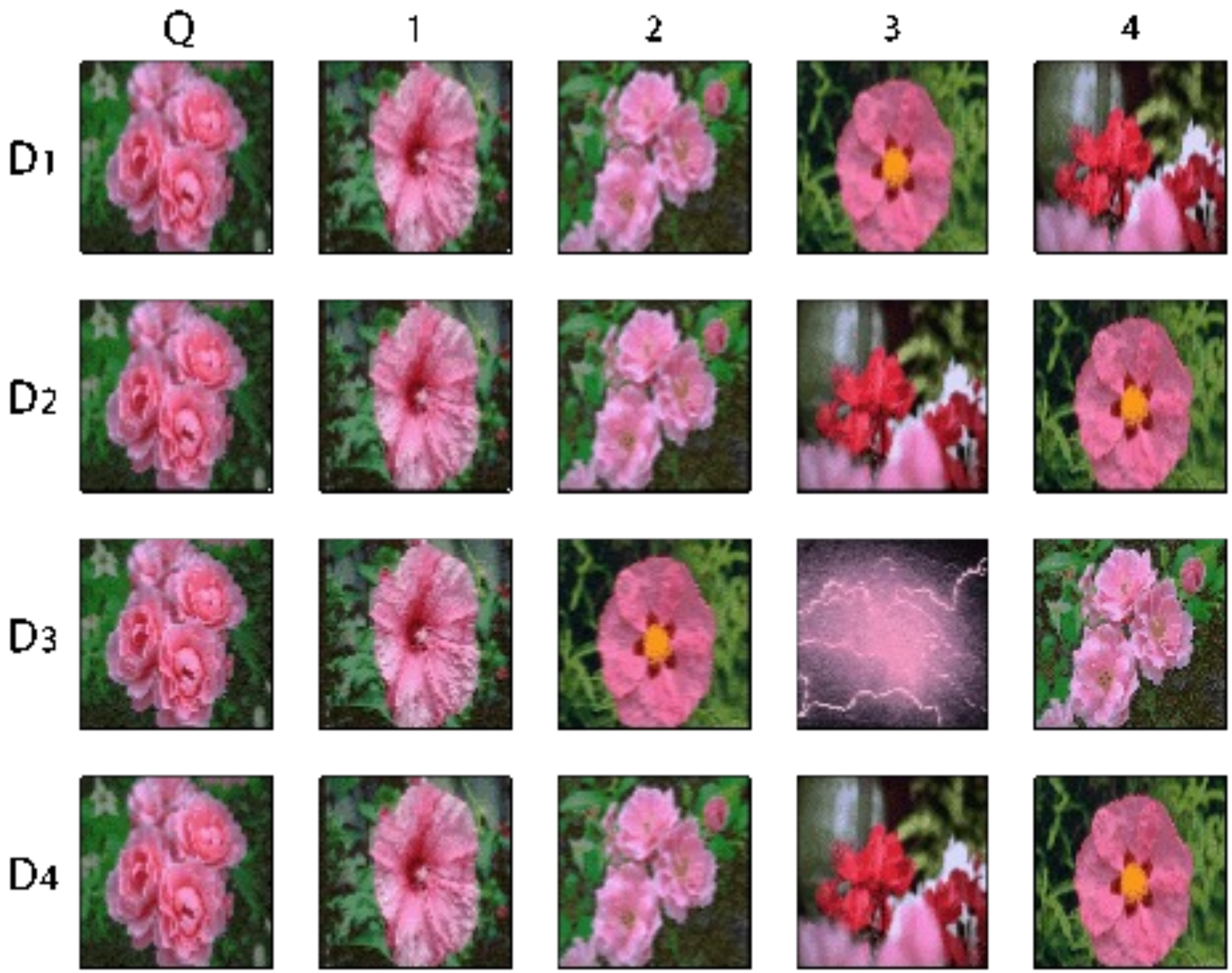
Performance -- Sunsets

Table 11.2. Query 1: Sunsets—Comparison of Eight Distance Metrics

21 Pink Flower Images	*D1*	*D2*	*D3*	*D4*	*D5*	*D6*	*D7*	*D8*
# Relevant in Top 10 (20)	6(10)	5(9)	4(6)	5(9)	4(7)	5(9)	2(4)	3(5)
# Retrieved to Obtain 5 (10) Relevant Ones	8(20)	9(24)	16(46)	9(22)	12(32)	8(21)	29(186)	18(56)

The quadratic-form metrics D4 and D6 performed well in Query 1.
The retrieval effectiveness for D1 was slightly better.
This query does not show that the quadratic-form color histogram metrics improve performance substantially over the Minkowski-form metrics.

Query 2: Flowers



Query 2: Flowers

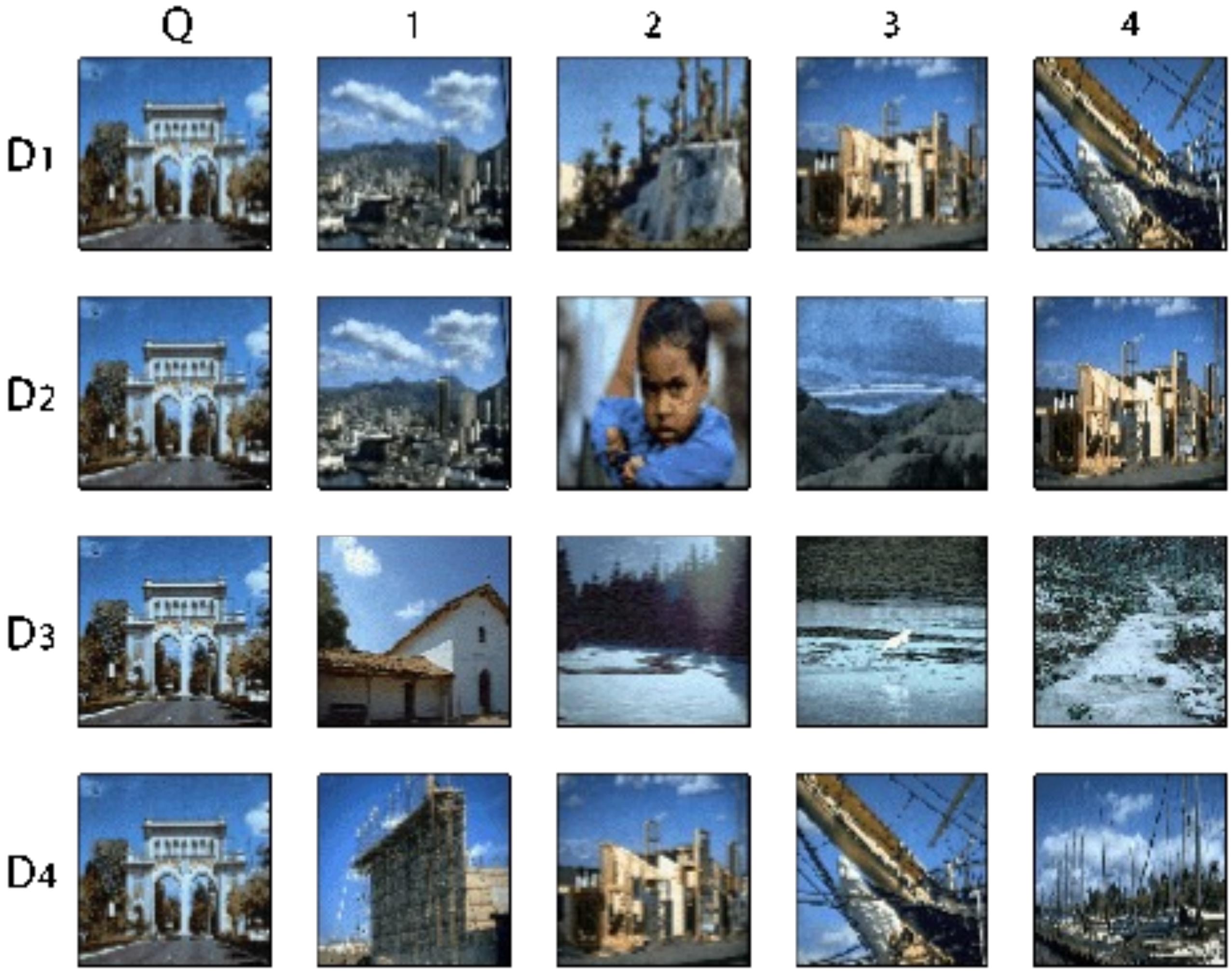
Performance -- Flowers

Table 11.3. Query 2: Flowers—Comparison of Eight Distance Metrics

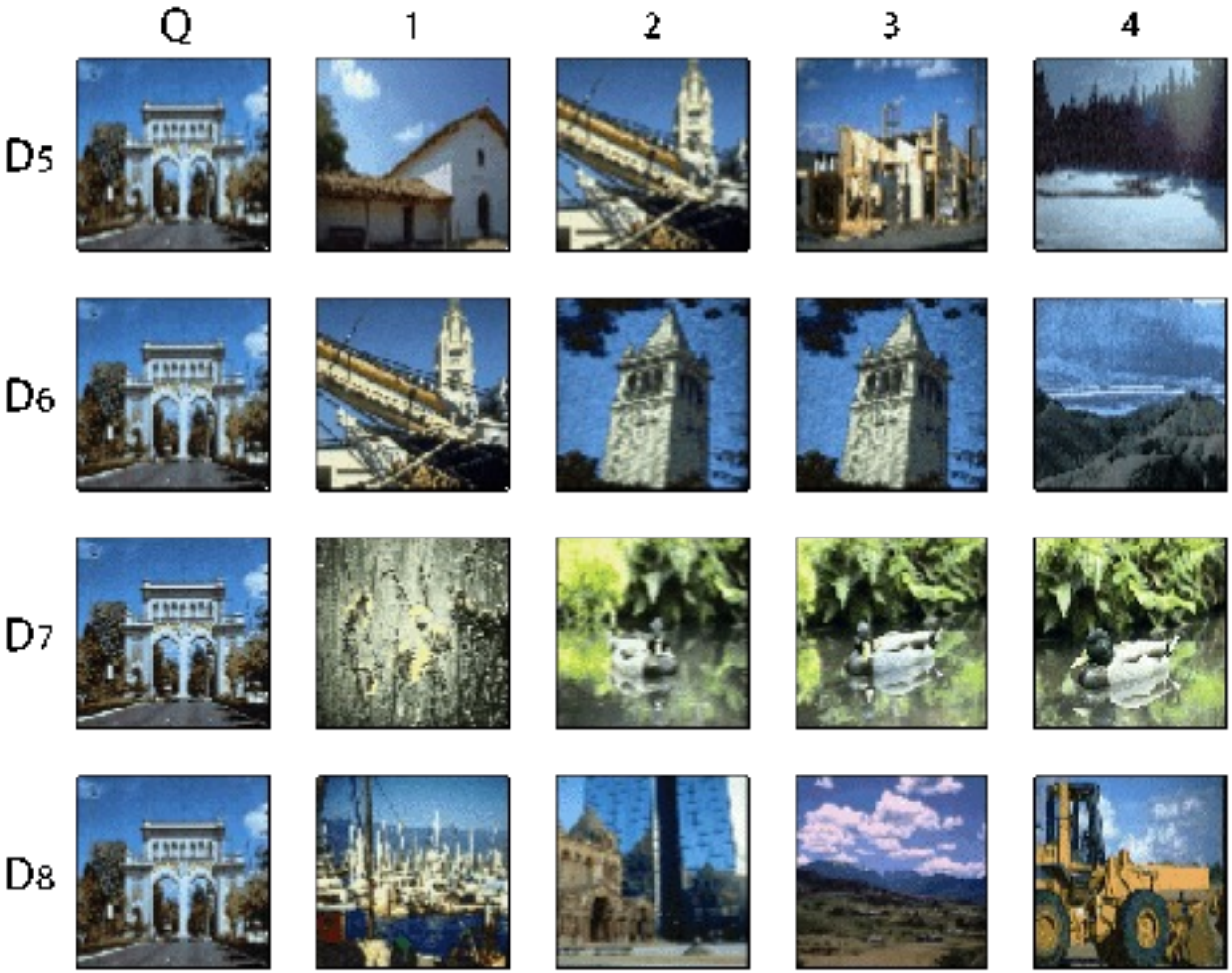
23 Sunset Images	D1	D2	D3	D4	D5	D6	D7	D8
# Relevant in Top 10 (20)	6(9)	6(9)	5(6)	6(9)	5(8)	7(10)	2(3)	3(3)
# Retrieved to Obtain 5 (10) Relevant Ones	7(21)	7(24)	10(62)	7(21)	8(34)	7(18)	48(291)	43(185)

D1, D2, D4, and D6 are similar in providing good retrieval results.
D3 produced a substantial drop in retrieval effectiveness.
The binary set quadratic-form metric D5 improved the retrieval effectiveness over D3.

Query 3: Nature



Query 3: Nature



Performance -- Nature

Table 11.4. Query 3: Nature with Blue Sky—Comparison of Eight Distance Metrics

42 Nature Images	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$	$D7$	$D8$
# Relevant in Top 10 (20)	5(8)	5(8)	3(4)	5(8)	3(5)	5(9)	1(1)	2(3)
# Retrieved to Obtain 5 (10) Relevant Ones	10(25)	10(26)	27(76)	10(26)	20(62)	9(23)	148(462)	33(108)

The most difficult one. The test set contains many images with blue colors that don't come from blue skies.

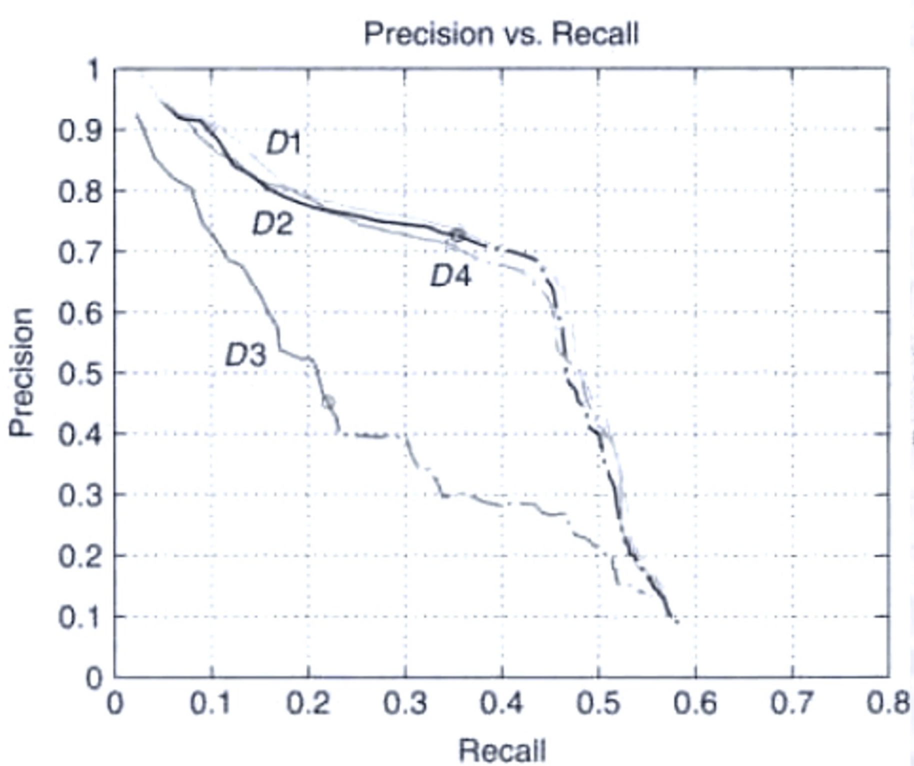
Typical queries for unconstrained color photos. The color information provides only a partial filter of the semantic content of the actual image.

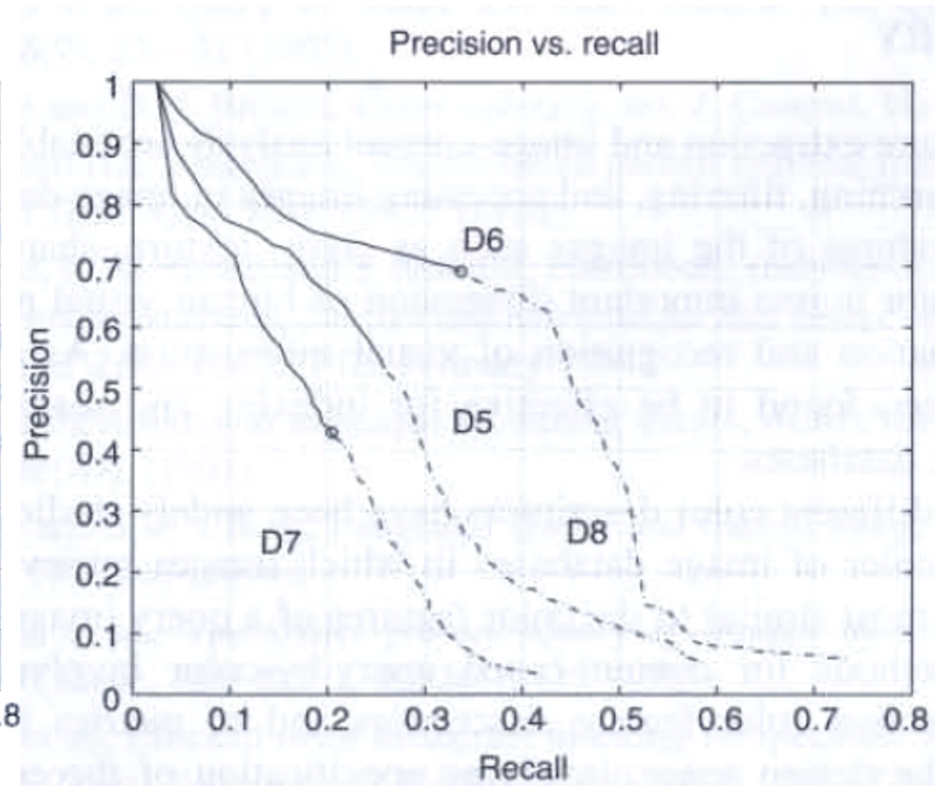
D1, D2, D4, and D6 are similar.
D3 produced a substantial drop in retrieval effectiveness, while D5 improves the retrieval effectiveness over D3.

Query 4: Lions

Query 4: Lions

Performance -- Lions





Performance -- Lions

Table 11.5. Query 4: Lions — Comparison of Eight Distance Metrics

41 Lion Images	$\mathcal{D}1$	$\mathcal{D}2$	$\mathcal{D}3$	$\mathcal{D}4$	$\mathcal{D}5$	$\mathcal{D}6$	$\mathcal{D}7$	$\mathcal{D}8$
# Relevant in Top 10 (20)	7(14)	7(14)	6(9)	7(14)	6(9)	7(13)	5(8)	7(10)
# Retrieved to Obtain 5 (10) Relevant Ones	6(13)	6(14)	8(26)	6(14)	7(22)	6(14)	8(32)	7(18)

D1, D2, D4, and D6 are found to be excellent.

Summary

- D1, D2, D4, and D6 are found to be more effective than that of D3, D5, D7, and D8.
- Simple color descriptors D7 and D8 are not effective.
- Performance is poor for binary color sets (D3 and D5)
- D1, D2, D4, and D6 are approximately the same.