Content-based image retrieval at the end of the early years  
(Kỹ thuật rút trích nội dung ảnh giai đoạn cuối của thời kỳ đầu)

Đại ý đây là kỹ thuật rút trích đặc trưng ảnh tốt nhất, cuối cùng của giai đoạn đầu tiền nghiên cứu

Mục tiêu của Paper này là đề cập đến việc kỹ thuật rút trích ảnh dựa vào các đặc trưng, cấu trúc, tính biểu diễn và sự tưởng đồng để từ đó rút ra các đặc trưng của một hình ảnh.

Contents

[1. Introduction (Giới thiệu) 3](#_Toc30445573)

[2. Scope (Mở rộng) 6](#_Toc30445574)

[2.1 Application of Content-based Retrieval (Ứng dụng của rút trích đặc trưng ảnh) 6](#_Toc30445575)

[2.2 The Image Domain and the Sensory Gap (Lĩnh vực ảnh và…) 6](#_Toc30445576)

[2.3 Domain Knowledge (Lĩnh vực tri thức) 7](#_Toc30445577)

[2.4 Use and User, the Semantic Gap () 7](#_Toc30445578)

[2.5 Discussion and Scope (Thảo luận và mở rộng) 7](#_Toc30445579)

[3. Description of Content : Image Processing (Mô tả nội dung : Xử lý ảnh) 8](#_Toc30445580)

[3.1 Color Image Processing (Xử lý màu sắc ảnh) 8](#_Toc30445581)

[3.2 Image Processing for Local Shape (Xử lý ảnh cho hình dạng ) 8](#_Toc30445582)

[3.3 Image Tecture Processing (Xử lý văn bản hình ảnh) 8](#_Toc30445583)

[3.4 Discussion on Image Processing (Thảo luận về xử lý ảnh) 8](#_Toc30445584)

[4. Description of Content : Features (Mô tả nội dung : Đặc trưng) 9](#_Toc30445585)

[4.1 Grouping Data (Nhóm dữ liệu) 9](#_Toc30445586)

[4.2 Global and Accumulating Features (Đặc trưng cục bộ và tích lũy) 9](#_Toc30445587)

[4.3 Salient Features (Đặc tính lộ rõ) 11](#_Toc30445588)

[4.4 Signs (Dấu hiệu) 11](#_Toc30445589)

[4.5 Shape and Object Features (Đặc tính hình dạng và đối tượng) 11](#_Toc30445590)

[4.6 Discription of Structure and Layout (Mô tả cấu trúc và bố cục của hình ảnh) 11](#_Toc30445591)

[4.7 Discussion on Features (Thảo luận về đặc trưng) 11](#_Toc30445592)

[5. Interpretation And Similarity (Sự biểu diễn và sự tương đồng) 12](#_Toc30445593)

[5.1 Sematic Interpretation (Biểu diễn ngữ nghĩa) 12](#_Toc30445594)

[5.2 Similarity between Features (Sự tưởng đồng giữa các đặc trưng) 12](#_Toc30445595)

[5.3 Similarity of Object Silhouettes (Sư tưởng đồng của bóng đối tượng) 12](#_Toc30445596)

[5.4 Similarity of Structural Features (Sự tưởng đồng của đặc tính cấu trúc) 12](#_Toc30445597)

[5.5 Similarity of Salient Features (Sự tưởng đồng của các đặc tính lộ rõ) 12](#_Toc30445598)

[5.6 Similarity of Sematic Level (Sự tưởng đồng của tầng ngữ nghĩa) 12](#_Toc30445599)

[5.7 Learning an Interpretation (Học cách biểu diễn) 12](#_Toc30445600)

[5.8 Discussion on Interpretation and Similarity (Thảo luận về sự biểu diễn và sự tưởng đồng) 12](#_Toc30445601)

[6. Interaction (Sự tương hỗ) 12](#_Toc30445602)

[6.1 Query Space : Definition and Initialization (Không gian truy vấn: Định nghĩa và khởi tạo) 12](#_Toc30445603)

[6.2 Query Specification (Đặc tả truy vấn) 13](#_Toc30445604)

[6.3 Query Space Display (Hiển thị không gian truy vấn) 13](#_Toc30445605)

[6.4 Interacting with Query Space (Tương tác với không gian truy vấn) 14](#_Toc30445606)

[6.5 Discussion on Interaction (Thảo luận về sự tương hỗ) 14](#_Toc30445607)

[7. System Aspects (Về mặt hệ thống) 14](#_Toc30445608)

[7.1 Storage and Indexing (Lưu trữ và lập chỉ mục) 14](#_Toc30445609)

[7.2 System Architectures (Kiến trúc hệ thống) 16](#_Toc30445610)

[7.3 System Evaluation (Đánh giá hệ thống) 17](#_Toc30445611)

[7.4 Discussion on System Aspects (Thảo luận về mặt hệ thống) 18](#_Toc30445612)

[8. Concluding Remarks (Nhận xét và kết luận) 19](#_Toc30445613)

[8.1 The driving force 19](#_Toc30445614)

[8.2 The heritage of computer vision 19](#_Toc30445615)

[8.3 The influence on computer vision 19](#_Toc30445616)

[8.4 Similarity and learning 20](#_Toc30445617)

[8.5 Interaction 20](#_Toc30445618)

[8.6 The need for databases 20](#_Toc30445619)

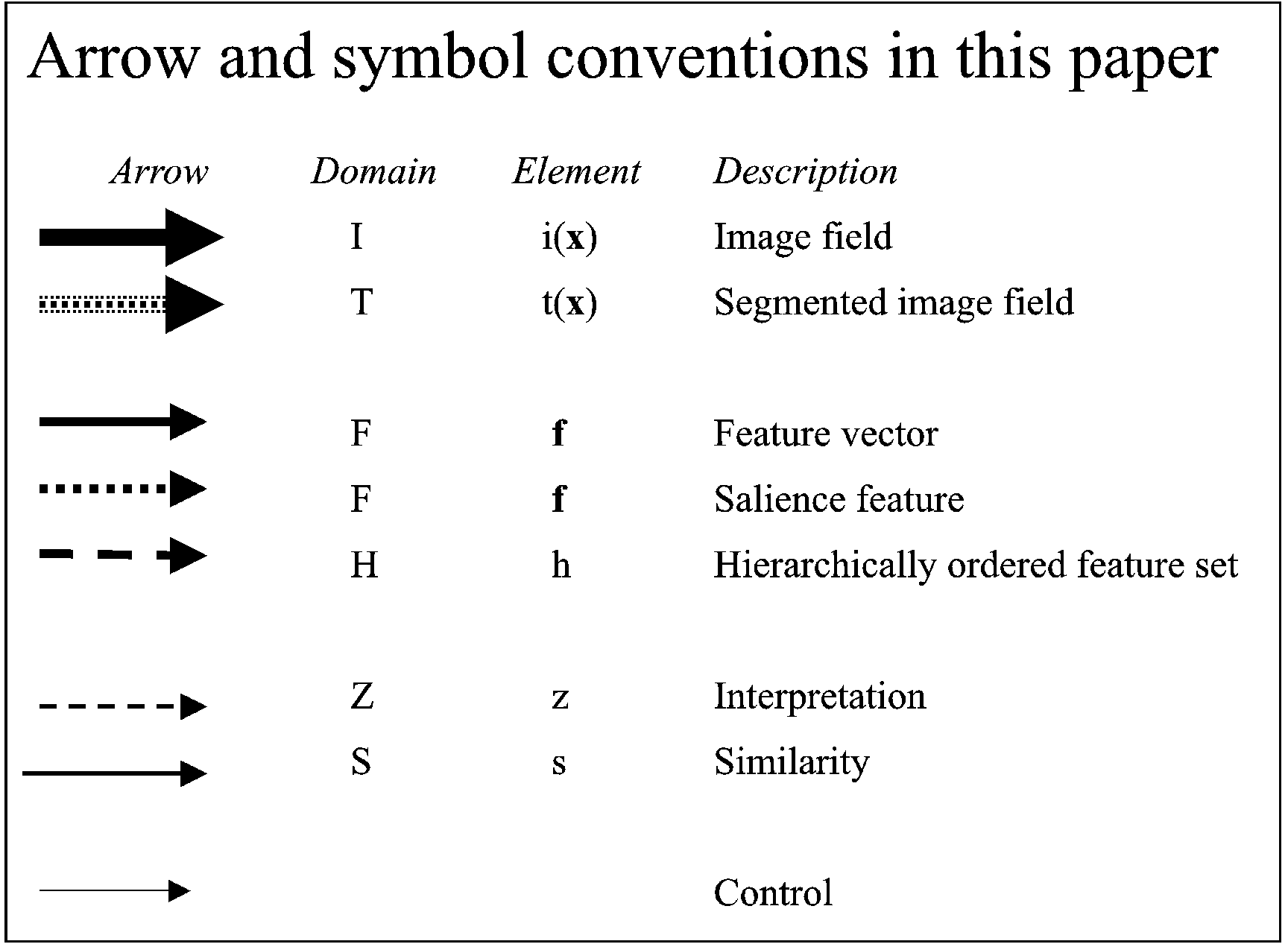
[8.7 The problem of evaluation 21](#_Toc30445620)

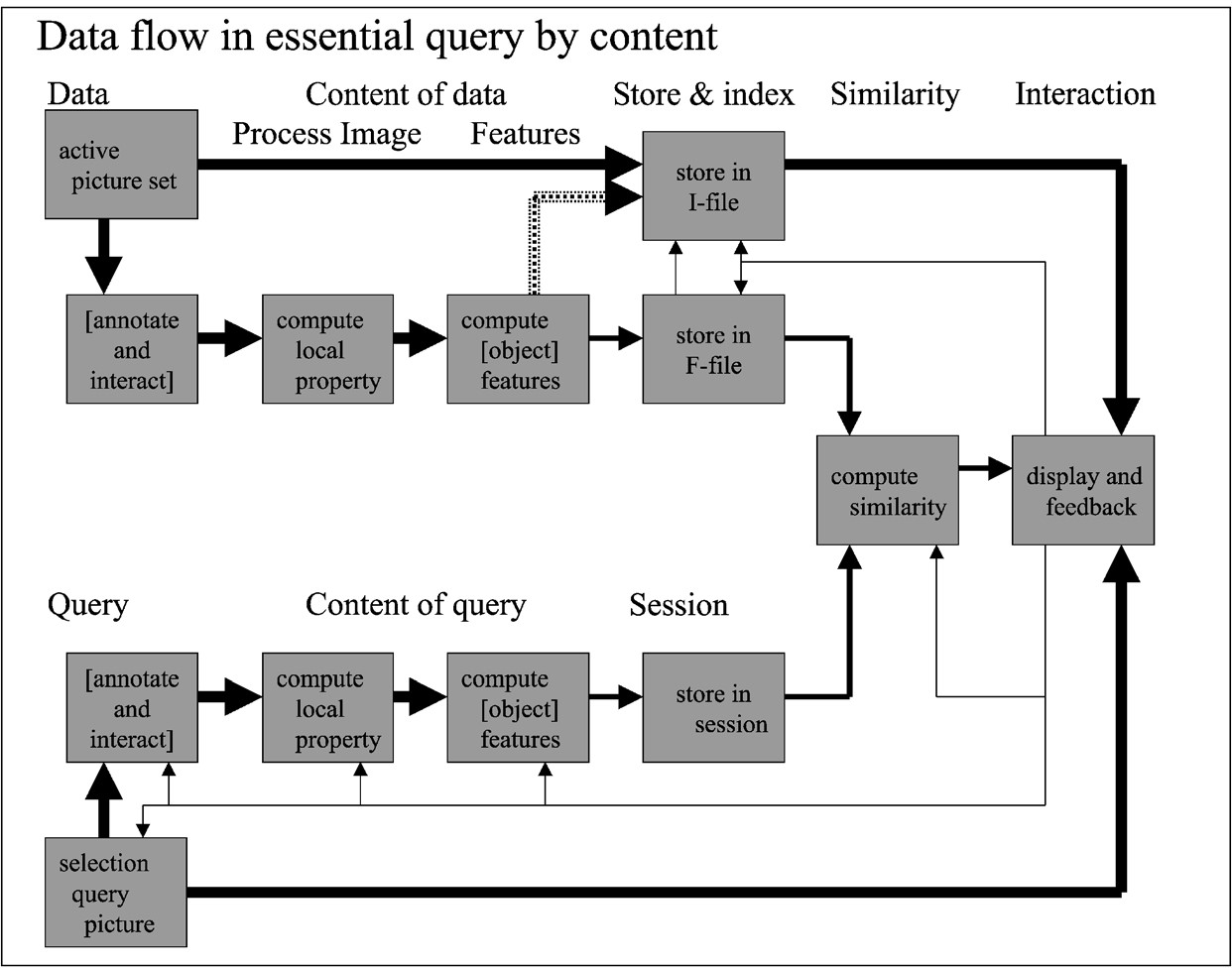
[8.8 The semantic gap and other sources 21](#_Toc30445621)

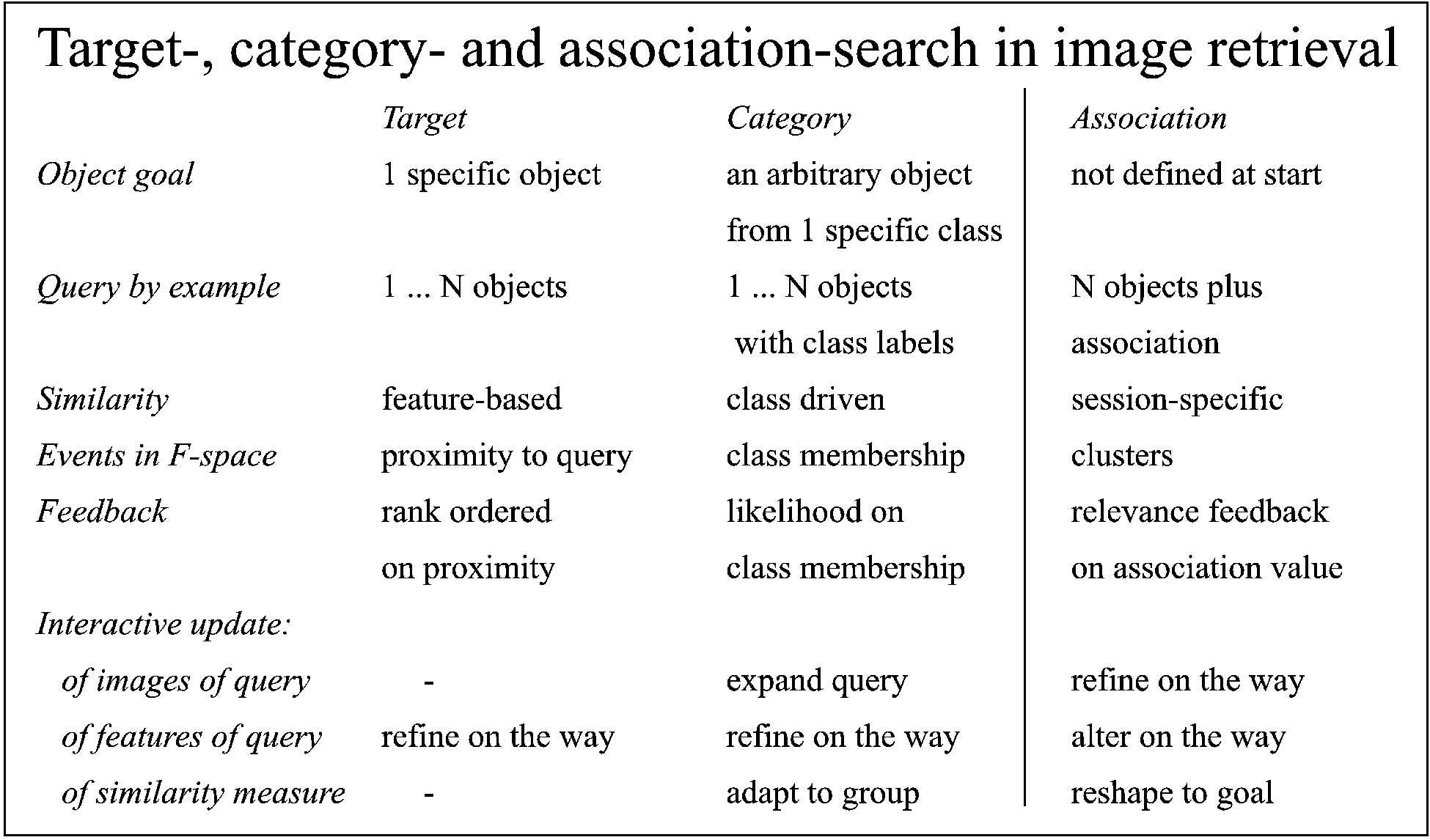
Phân công :

* Hoàng Minh Thanh : Chương 5
* Phan Minh Tâm : Chương 6
* Nguyễn Mạnh Tấn : Chương 1 & 2
* Nguyễn Anh Trúc : Chương 1 & 2
* Phan Huy Cường : Chương 3
* Trần Quang Cường : Chương 7 & 8
* Nguyễn Hoàng Sơn : Chương 7 & 8
* Trương Châu Hiền : Chương 4

# Introduction (Giới thiệu)



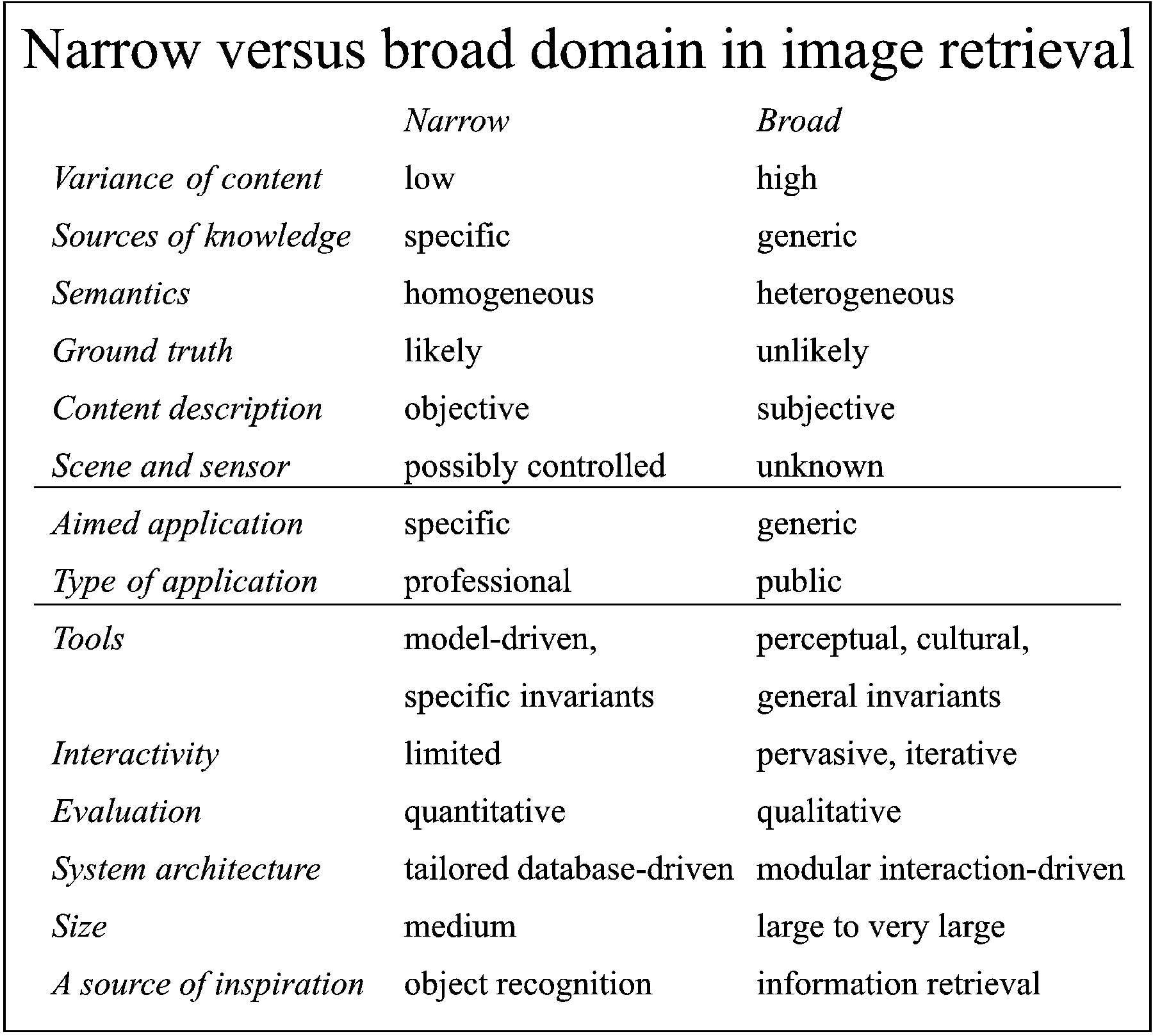




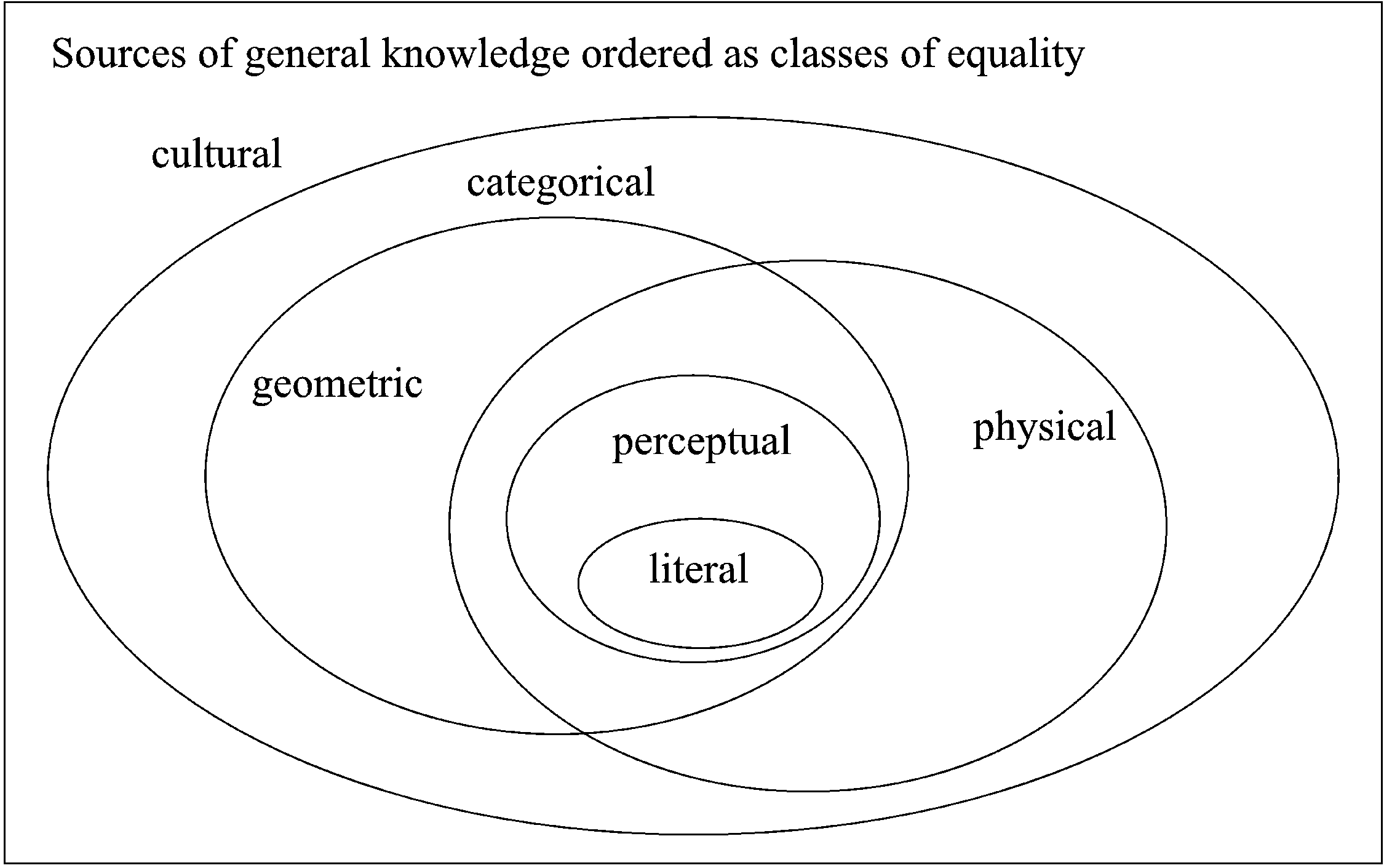
# Scope (Mở rộng)

## 2.1 Application of Content-based Retrieval (Ứng dụng của rút trích đặc trưng ảnh)

## 2.2 The Image Domain and the Sensory Gap (Lĩnh vực ảnh và…)

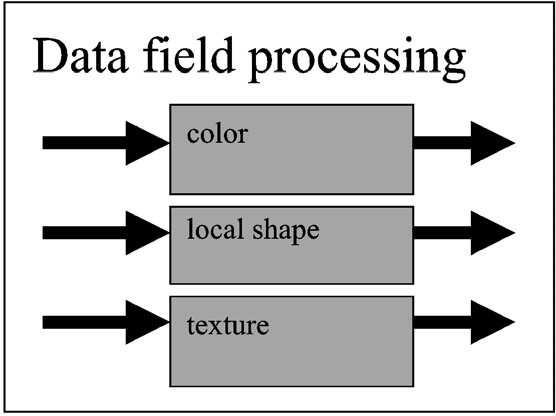


## 2.3 Domain Knowledge (Lĩnh vực tri thức)



## 2.4 Use and User, the Semantic Gap ()

## 2.5 Discussion and Scope (Thảo luận và mở rộng)



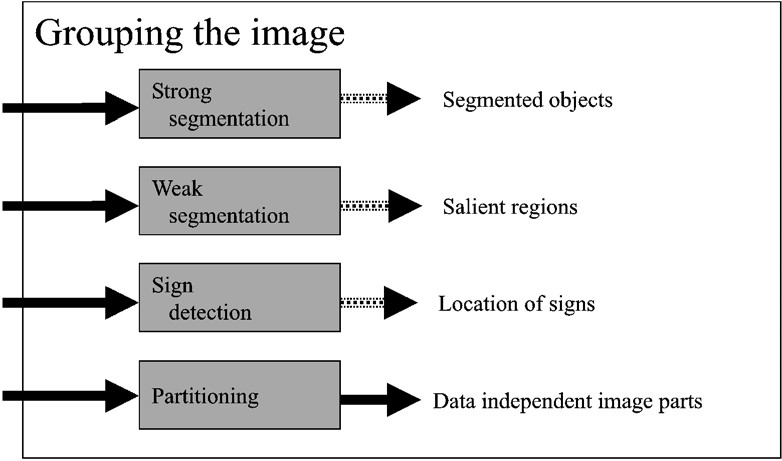
# Description of Content : Image Processing (Mô tả nội dung : Xử lý ảnh)

## 3.1 Color Image Processing (Xử lý màu sắc ảnh)

## 3.2 Image Processing for Local Shape (Xử lý ảnh cho hình dạng )

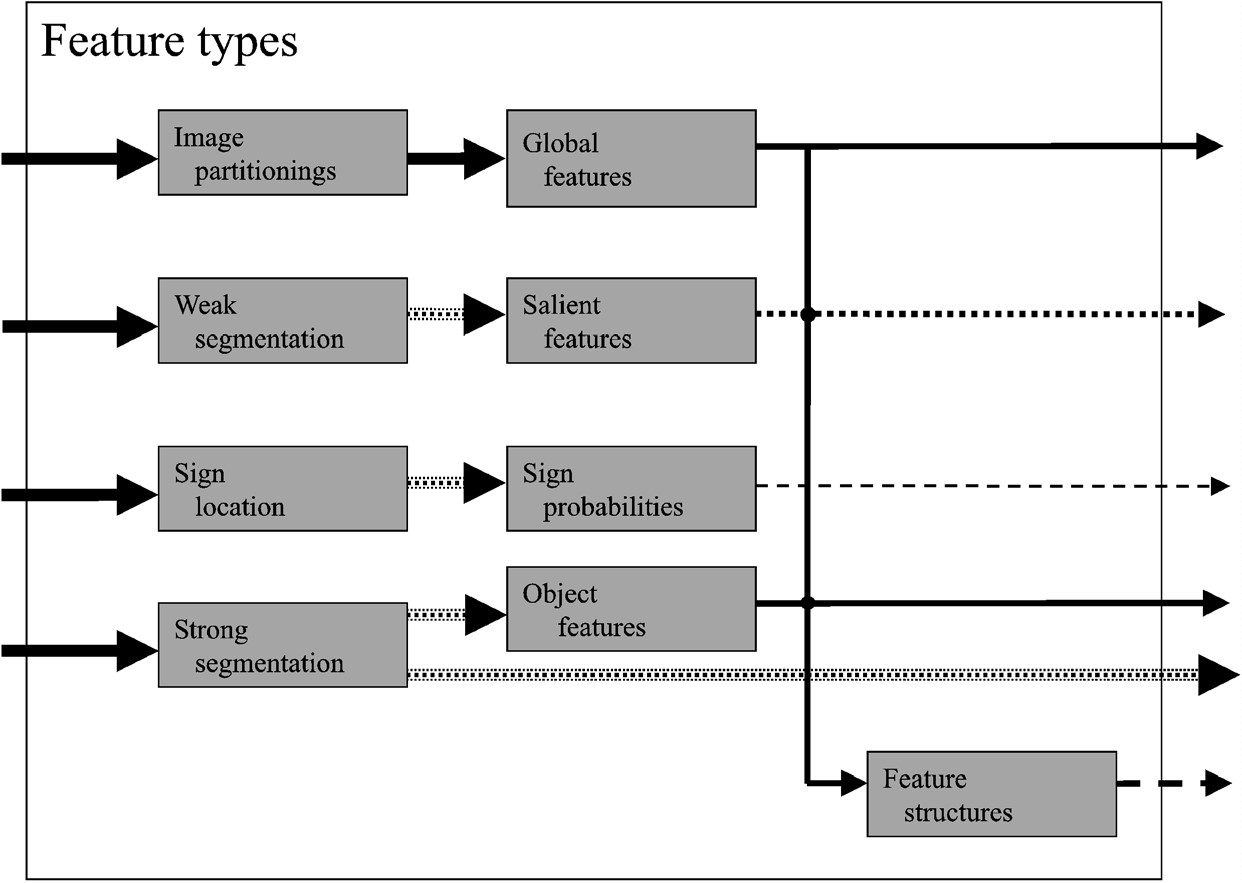
## 3.3 Image Tecture Processing (Xử lý văn bản hình ảnh)

## 3.4 Discussion on Image Processing (Thảo luận về xử lý ảnh)

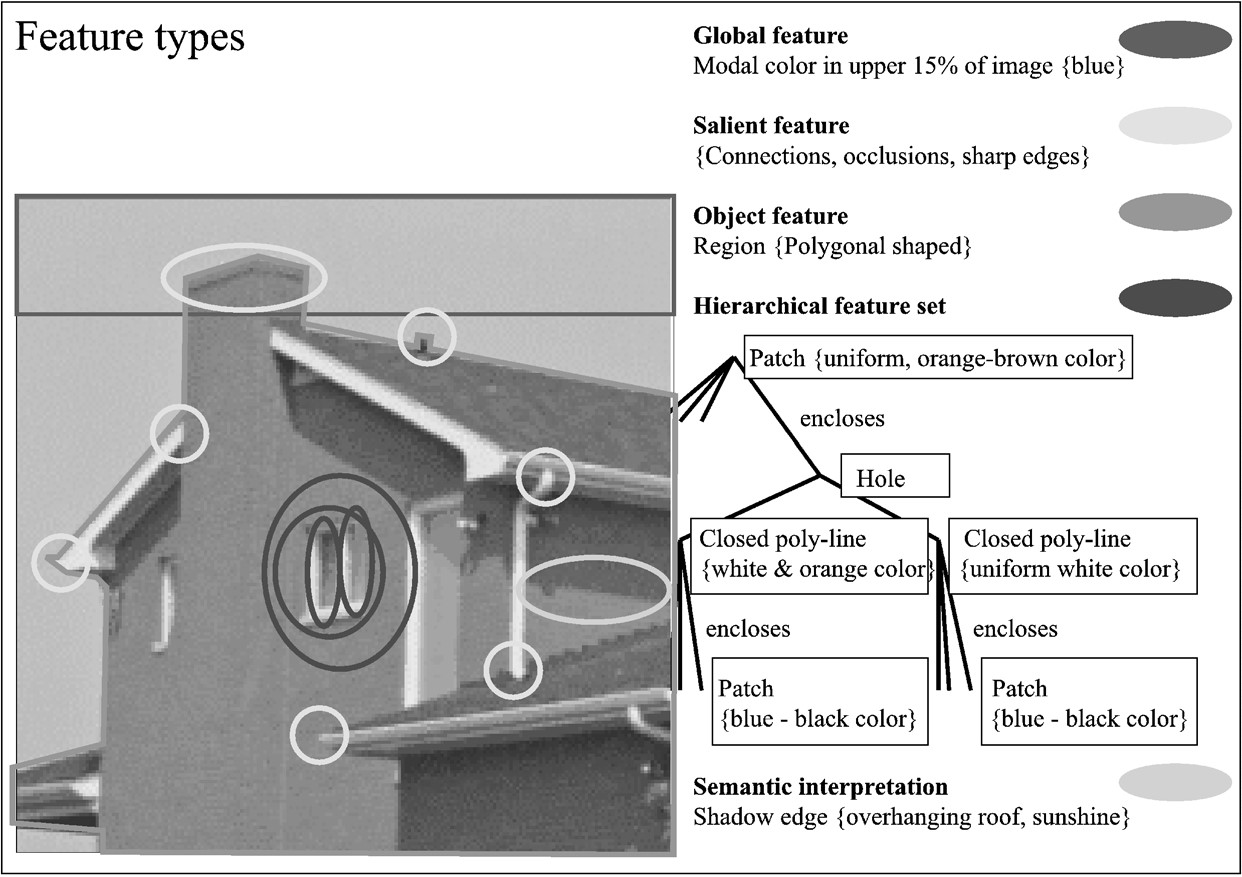


# Description of Content : Features (Mô tả nội dung : Đặc trưng)

## 4.1 Grouping Data (Nhóm dữ liệu)



## 4.2 Global and Accumulating Features (Đặc trưng cục bộ và tích lũy)



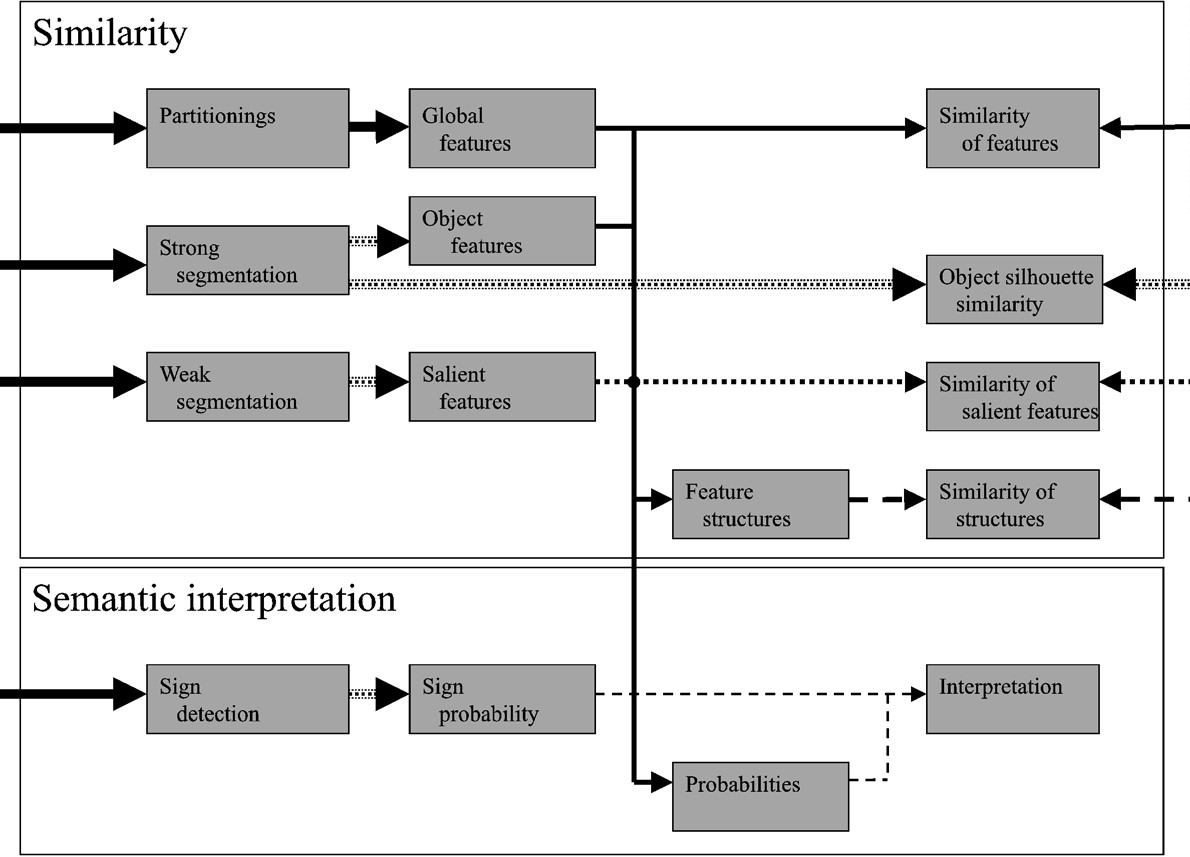
## 4.3 Salient Features (Đặc tính lộ rõ)

## 4.4 Signs (Dấu hiệu)

## 4.5 Shape and Object Features (Đặc tính hình dạng và đối tượng)

## 4.6 Discription of Structure and Layout (Mô tả cấu trúc và bố cục của hình ảnh)

## 4.7 Discussion on Features (Thảo luận về đặc trưng)



# Interpretation And Similarity (Sự biểu diễn và sự tương đồng)

## 5.1 Sematic Interpretation (Biểu diễn ngữ nghĩa)

## 5.2 Similarity between Features (Sự tưởng đồng giữa các đặc trưng)

## 5.3 Similarity of Object Silhouettes (Sư tưởng đồng của bóng đối tượng)

## 5.4 Similarity of Structural Features (Sự tưởng đồng của đặc tính cấu trúc)

## 5.5 Similarity of Salient Features (Sự tưởng đồng của các đặc tính lộ rõ)

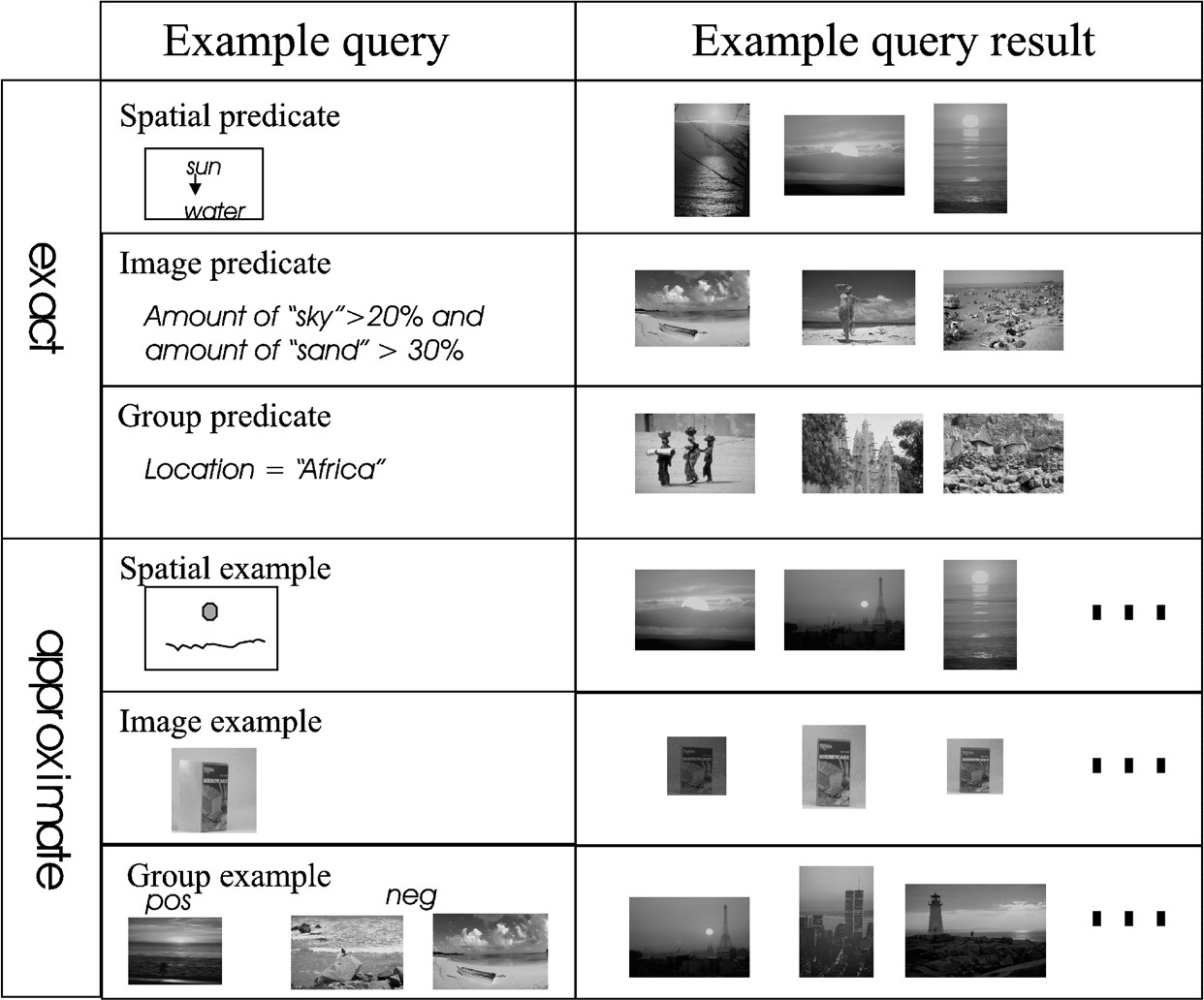
## 5.6 Similarity of Sematic Level (Sự tưởng đồng của tầng ngữ nghĩa)

## 5.7 Learning an Interpretation (Học cách biểu diễn)

## 5.8 Discussion on Interpretation and Similarity (Thảo luận về sự biểu diễn và sự tưởng đồng)

# 6. Interaction (Sự tương hỗ)

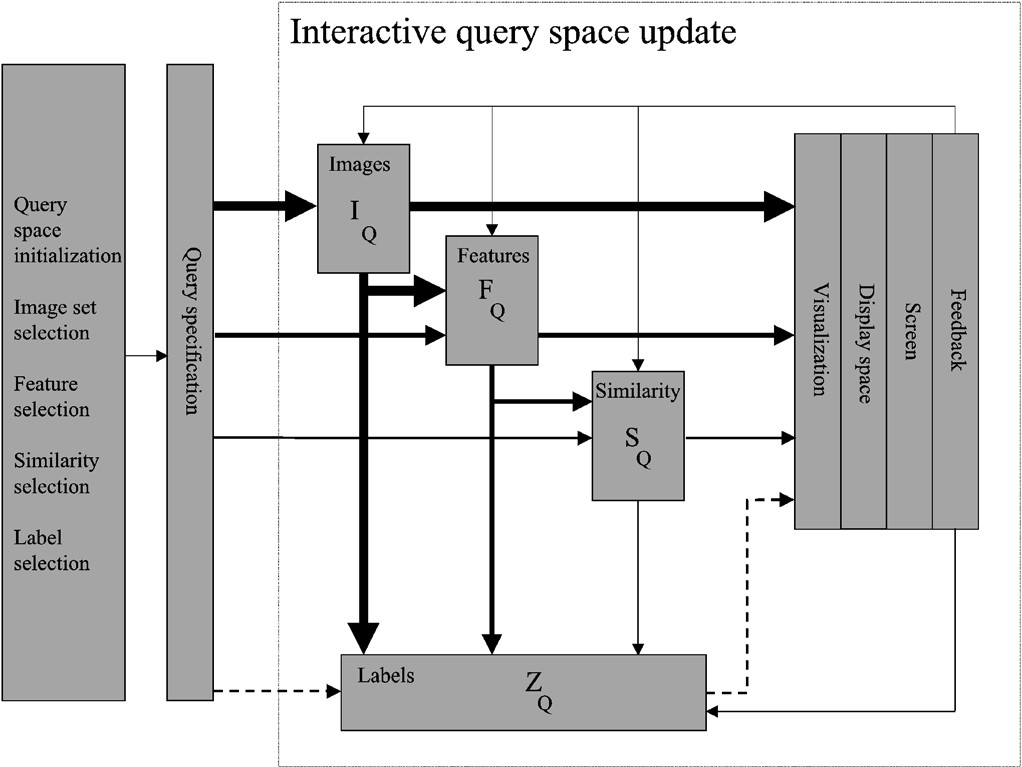
## 6.1 Query Space : Definition and Initialization (Không gian truy vấn: Định nghĩa và khởi tạo)



## 6.2 Query Specification (Đặc tả truy vấn)

## 6.3 Query Space Display (Hiển thị không gian truy vấn)

## 6.4 Interacting with Query Space (Tương tác với không gian truy vấn)



## 6.5 Discussion on Interaction (Thảo luận về sự tương hỗ)

# 7. System Aspects (Về mặt hệ thống)

## 7.1 Storage and Indexing (Lưu trữ và lập chỉ mục)

Chúng ta quan tâm tới nội dung cuối cùng của hình ảnh, sự lặp lại của tất cả hình ảnh trong tập dữ liệu của vector O ở phần trước, chúng ta đã thảo luận về việc translate từ vector image Fq’ so sánh với fd của tập dữ liệu trên cơ sở hàm tương tự.

Trong thực tế, điều thú vị nhất trong ứng dụng là xử lý trên các tập dữ liệu lớn, nơi mà thống kê ảnh hưởng tới hình ảnh và cách học ảnh hưởng tới tập dữ liệu. Đối với tập dữ liệu hình ảnh lớn, việc performance tính toán là quan trọng. Khi lưu các vector đặc trưng, với mọi bản ghi tương ứng với mỗi vector, chúng ta buộc phải xử lý qua tất cả các vector đó. Trong trường hợp đó chúng ta phải duyệt qua từng record tương ứng với các phép tính toán trên vector đó. Vì vậy việc xử lý là O(N), số lần xử lý của việc tính toán khi thêm một record vào DB và update record đó. Việc quét tuyến tính sẽ đem đến thời gian phản hồi ngoài mong đợi, nhất là do các tập dữ liệu trên 10000 hình ảnh trở lên.

Ngoài ra, số lượng hình ảnh, do phân giải cũng được xem xét. Ví dụ, biểu đồ histogram of texture-based retrieval, hình ảnh có một vector chín chiều cho mỗi pixel được nén thành biểu đồ 512 bin với tổng số 512^2 biểu đồ 512 bins mỗi hình ảnh.

Có một sự gần chính xác trong việc áp dụng lý thuyết, thời gian response đã giảm 15 lần khi sử dụng k-d tree cho 20 nearest neighbor truy vấn với N = 500000 hình ảnh với M = 78 chiều với e = 0,1. Một cải tiến nữa là cây B tree k-d là dạng tổng quát hóa đa chiều của cây B-tree tiêu chuẩn với khả năng phân tách của k-d tree. Đây là cấu trúc dữ liệu cân bằng với độ sâu bằng gốc, với hiệu xuất O(logN).

Kỹ thuật phân vùng data, với mỗi điểm trong không gian đặc trưng, một vùng đại diện cho vùng lân cận của vector đó. R-tree là một cấu trúc phân vùng dữ liệu như vậy để lập chỉ mục các vùng siêu hình chữ nhật trong không gian M chiều. Các nút kad của R-tree đại diện cho các hình chữ nhật giới hạn tối thiểu của các tập hợp các vector đặc trưng. Một nút bên trong là một hình chữ nhật bao gồm các hình chữ nhật của tất cả các phần tử con của nó. R-tree+ là một biến thể không cho phép các hình chữ nhật giới hạn tối thiểu trong một nút trùng nhau. Trong cây R-tree\*, hình chữ nhật giới hạn tối thiểu có thể chồng lên nhau. The VAM-split R-tree chia dọc theo chiều của phương sai tối đa, vì vậy có performance toort hơn R-tree\*. Do tiêu chí phân tách phụ thuộc vào mức độ lan truyền của dữ liệu thay vì số lượng dữ liệu trong một nhóm, nên nó chứng tỏ là rất hiệu quả, ngay cả đối với dữ liệu được phân cụm và cho các chiều tương quan. A VAM-split R-tree có thể được xây dựng nhanh hơn hai bậc so với R-tree, mất it hơn một giây chho truy vấn 20 lân cận gấn nhất trong cơ sở dữ liệu hình ảnh N = 100000 hình ảnh.

Cây SS-tree được cải tiến hơn nữa, SR-tree [87], sử dụng giao điểm của siêu cầu giới hạn tối thiểu và siêu chữ nhật giới hạn tối thiểu làm vùng giới hạn của một yếu tố dữ liệu. Khi kích thước tăng lên, nó kết hợp, đối với một vùng giới hạn, lợi thế của thể tích của hình chữ nhật với đường kính siêu nhỏ của siêu cầu. Nó đã được chứng minh rằng cây SR có hiệu quả trong các kích thước vector tính năng thấp và cao.

Distance-based index structures là các kỹ thuật phân vùng không gian dựa trên ví dụ và do đó rất phù hợp cho truy vấn khi dữ liệu không gian. Ý tưởng chính là chọn một điểm và chia phần còn lại của không gian tính năng thành các nhóm M0 theo các vòng tròn đồng tâm xunh quanh. Đây là điểm thuận lợi đầu tiên của distance-based index. VP-tree được đặc trưng cho các vector nhiều chiều. Cây MVP là một cấu trúc dữ liệu tĩnh sử dụng nhiều điểm ví dụ (vantage) tại mỗi nút. Điểm thuận lợi đầu tiên được sử dụng để tạo phân vùng M0. Và trong mỗi phân vùng, một điểm thuận lợi thứ hai tạo ra M0 nhiều phân chia hơn. Cây MVP, với M’ = 3, Mmaxpernode = 13, sử dụng số liệu L2 được tìm thấy để thực hiện các phép tính khoảng cách ít hơn trong khi so sánh vector cho tìm kiếm lân cận gần nhất so với kỹ thuật cạnh tranh. Cây M được đề xuất trong [27] là một chiến lược lập chỉ mục mạnh mẽ và có thể mở rộng hơn, sử dụng tam giác của không gian số liệu, nhưng đồng thời, vẫn giữ các thuộc tính phân vùng dữ liệu của các phương thức truy cập không gian như cây R và cây SS. M-tree tương tự như MVP-tree theo nghĩa là cả hai đều phân chia không gian dựa trên thể tích hình cầu xung quanh một số vectơ tham chiếu. Tuy nhiên, M-tree là local, chỉ số cân bằng. Do đó, chúng ít bị ảnh hưởng bởi sự phân bố vectơ trong không gian đặc trưng.

Mặc dù M-tree không thể đảm bảo hiệu suất trường hợp xấu nhất, nhưng nó đã được chứng minh là hoạt động tốt trong thực tế, với chi phí tính toán khoảng cách của O(logM) trong hầu hết các trường hợp. Nó cũng có lợi thế là xử lý trực tiếp các tính năng có thể được biểu diễn trong không gian số liệu nhưng không phải trong không gian vectơ, không giống như các kỹ thuật như FastMap [46] hoặc chia tỷ lệ đa chiều, xấp xỉ không gian tính năng với không gian vectơ duy trì xấp xỉ cùng số liệu.

Thật đáng tiếc rằng chưa có quá nhiều công việc phân chia giữa các ngành về tầm nhìn và cơ sở dữ liệu đã được thực hiện, với một vài ngoại lệ trong các hệ thống thương mại [48], [61] và nghiên cứu [46], [85], [171]

## 7.2 System Architectures (Kiến trúc hệ thống)

Nhiều hệ thống như Photobook [128], PictoSeek [57] và [116] bắt nguồn từ tầm nhìn máy tính. Trong các hệ thống như vậy, dữ liệu và tính năng thường được lưu trữ trong các tệp được xử lý theo tên. Từ quan điểm kiến trúc, phương pháp này có khả năng gặp phải các vấn đề về tính toàn vẹn và hiệu suất của dữ liệu khi cố gắng mở rộng quy mô cơ sở dữ liệu lớn và số lượng lớn người dùng.Số lượng lớn là một vấn đề và bất kỳ công cụ tìm kiếm Web nào khác, trong đó nhấn mạnh vào việc web sử dụng cơ sở dữ liệu làm nơi lưu trữ. Trong các kiến trúc như vậy, một sự khác biệt rõ ràng giữa việc index và không index. Việc đơn giản hóa DB rất tốt. Do đó, với các cấu trúc chỉ mục tĩnh, đơn giản, người ta có thể có được hiệu suất tốt trong quá trình truy xuất.

Một bước vượt xa cơ sở dữ liệu chỉ đọc là việc sử dụng hệ thống quản lý cơ sở dữ liệu tiêu chuẩn được mở rộng để truy xuất hình ảnh. Giảm truy xuất hình ảnh dưới dạng mô-đun trình cắm trong cơ sở dữ liệu hiện có giải quyết vấn đề toàn vẹn cho nội dung hình ảnh và cho phép cập nhật động. Cơ sở dữ liệu tiêu chuẩn duy trì một kênh trao đổi dữ liệu hẹp giữa công cụ tìm kiếm và dữ liệu và do đó, hiệu suất khá kém.

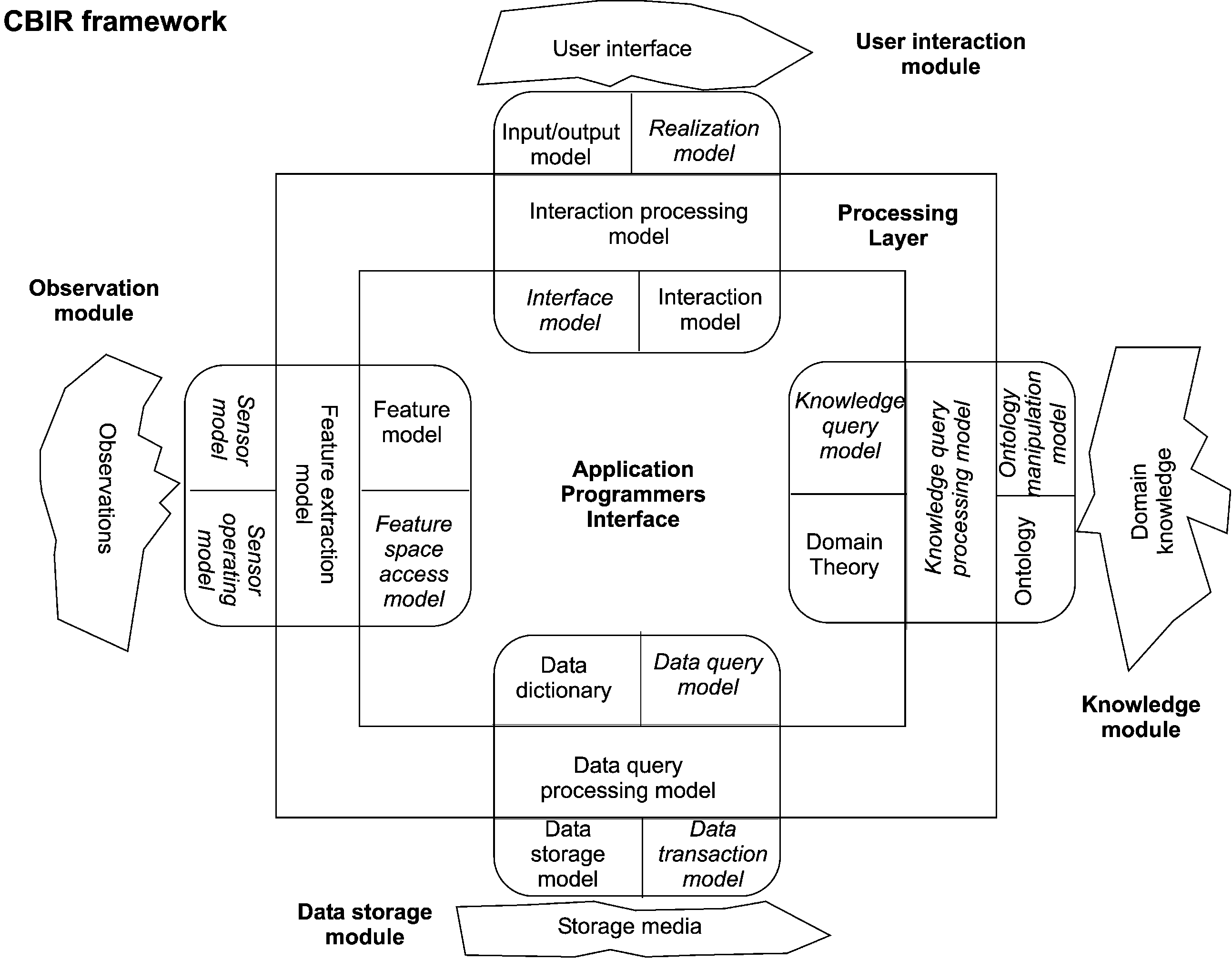
Đối với các hệ thống hoàn chỉnh hơn, các kiến trúc chi tiết sẽ được xem xét. Một trong những đóng góp ban đầu để làm như vậy có thể được tìm thấy trong hệ thống CORE [193], là cơ sở cho nhiều ứng dụng khác nhau. Kiến trúc được hệ thống được chia làm nhiều loại:  quản lý cơ sở dữ liệu chung, trên đó các mô-đun để phân tích, lập chỉ mục, đào tạo và truy xuất. Theo cách tiếp cận như vậy, cấu trúc cơ sở dữ liệu chiếm ưu thế trong quản lý kiến thức, tính toán tính năng và các công cụ trực quan. Kiến trúc chung được mô tả trong [60] dựa trên mô hình chi tiết của các loại thông tin khác nhau. Mỗi loại thông tin lưu giữ dữ liệu của nó trong một kho lưu trữ riêng thay vì cơ sở dữ liệu hợp nhất, đặc biệt tách biệt các data-driven và ngữ nghĩa. Ưu điểm là các đơn vị xử lý, một đơn vị cho mỗi loại dữ liệu mới, có thể được kết hợp nhanh chóng. Hạn chế là sự khác biệt giữa các tính năng mang thông tin và tính năng dựa trên dữ liệu thay đổi theo mức độ kiến thức trong hệ thống, sẽ gây cản trở trong việc phân tích kiến thức.

Kiến trúc Infoscopes [82] tuân theo sự phân biệt tương tự các tính năng. Nó bổ sung một chiều hướng mới cho kiến trúc của các hệ thống dựa trên nội dung bằng cách đưa ra kiến thức rõ ràng cho các phần khác nhau của hệ thống.

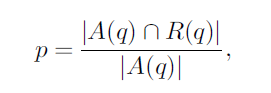
Hệ thống Piction [155] sử dụng thông tin hình ảnh và văn bản để làm cho kiến thức rõ ràng. Các hệ thống cho đến nay tập trung chính vào dữ liệu và sau đó về xử lý kiến thức. Ngược lại, các hệ thống được xem xét trong [141] và đặc biệt là hệ thống MARS- [142], dựa trên mô hình truy xuất thông tin. Sự tương tác với người dùng được coi là rất quan trọng cho một hệ thống thành công. Họ đề xuất một kiến trúc cho các hệ thống trong tương lai có cấu trúc xử lý tuần tự từ các tính năng đến tương tác người dùng.

Tích hợp và trực quan hóa cơ sở dữ liệu [96] đã mang lại các kỹ thuật trực quan hóa cấu trúc và nội dung của cơ sở dữ liệu hình ảnh. Sự tương đồng dựa trên nội dung giữa các hình ảnh không được khai thác. Các hệ thống trong [68], [184] đã tích hợp sự tương tác và trực quan hóa dựa trên nội dung, cũng như các kỹ thuật cơ sở dữ liệu để truy xuất hình ảnh có liên quan. Hệ thống cơ sở dữ liệu El NinÄo [146] đề xuất một kiến trúc cho việc tích hợp từ xa thông qua một bộ trung gian.

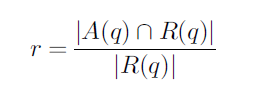
## 7.3 System Evaluation (Đánh giá hệ thống)



Việc đánh giá truy xuất hình ảnh là một chủ đề khó khăn nhưng rất cần thiết để triển khai thành công các hệ thống và tính hữu dụng của chúng trong các ứng dụng thực tế. Việc đầu tiên để đánh giá cơ sở dữ liệu hình ảnh là sự khôi phục thông tin, trong đó các kỹ thuật đánh giá dựa trên người dùng đã đạt đến một mức độ đáng kể [143]. Giả sử tập dữ liệu D và truy vấn q. Thông qua việc sử dụng các đối tượng của con người, bộ dữ liệu có thể được chia thành hai bộ: tập hợp các hình ảnh liên quan đến truy vấn q, R(q) và phần bổ sung của nó, tập hợp các hình ảnh không liên quan R(q). Giả sử rằng truy vấn q được cung cấp cho một tập dữ liệu và nó trả về một tập hợp các hình ảnh A(q) làm câu trả lời. Độ chính xác của câu trả lời là tỷ lệ của các hình ảnh được trả về phù hợp với truy vấn:



trong khi recall là tỉ lệ của những ảnh có liên quan được được mô tả bởi công thức:



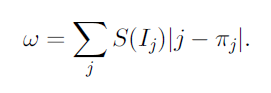
Mặc dù độ chính xác và thu hồi là một việc quan trọng trong truy xuất thông tin, trong bối cảnh cơ sở dữ liệu hình ảnh không đủ vì hai lý do:

* Việc lựa chọn một bộ có liên quan trong cơ sở dữ liệu hình ảnh có nhiều vấn đề hơn so với cơ sở dữ liệu văn bản. Trong trường hợp tài liệu, phán đoán của con người về một tài liệu nhất định có liên quan đến một hoàn cảnh nhất định là tương đối dễ và có mối liên hệ chặt chẽ giữa phán đoán này. Trong trường hợp của một hình ảnh, mức độ liên quan kém ổn định hơn nhiều do tính chất của một hình ảnh được tách ra liên quan đến bối cảnh ngôn ngữ.
* Lý do thứ hai là cơ sở dữ liệu hình ảnh thường không trả về một tập hợp kết quả tương ứng.

Mặc dù về nguyên tắc, một kết quả truy vấn của toàn bộ cơ sở dữ liệu, kích thước của bộ câu trả lời A(q) thường được giảm xuống còn k lần hình ảnh phù hợp nhất. Khi số lượng hình ảnh liên quan lớn hơn k, việc thu hồi là vô nghĩa như một thước đo chất lượng của cơ sở dữ liệu.

Mặc dù có những hạn chế, độ chính xác và recall là các phép đo hữu ích trong các trường hợp đặc biệt. Đặc biệt, khi cơ sở dữ liệu hình ảnh dựa trên ngữ nghĩa được cung cấp bởi label hoặc mô tả văn bản, độ chính xác và recall được sử dụng một cách hữu ích [164].

Trong [119], vấn đề được xem xét là đo hiệu năng của cơ sở dữ liệu mà không sử dụng khái niệm tập hợp có liên quan. Họ giả định rằng một cơ sở dữ liệu lý tưởng sẽ lấy một truy vấn q và cung cấp một thứ tự lý tưởng của cơ sở dữ liệu: Z= [z1, ..., zk]. Cơ sở dữ liệu lý tưởng cũng sẽ cung cấp thước đo mức độ phù hợp cho từng hình ảnh, S(I) [0,1]Cơ sở dữ liệu được thử nghiệm, ở đầu bên kia, sẽ sắp xếp hình ảnh, được đưa ra cùng một truy vấn q, như ZQ = [z1, ..., zk], trong đó 1, k là hoán vị của 1, … k.  Sự dịch chuyển của hình ảnh Ij giữa hai thứ tự được đưa ra bởi jj 􏰆 j j, và tất cả các chuyển vị có thể được thêm và cân bằng sự liên quan của các hình ảnh tương ứng có được sự dịch chuyển có trọng số



## 7.4 Discussion on System Aspects (Thảo luận về mặt hệ thống)

Về kiến trúc hệ thống, chúng ta duy trì hệ thống truy xuất dựa trên nội dung được phát triển và kết quả của việc tính toán, kiến thức, sự tương tác người dùng và lưu trữ những dữ liệu đó lại. Đối với các kiến trúc hệ thống được thảo luận ở trên, chúng ta kết luận rằng hầu hết các hệ thống có một sự sáng tạo. Chúng ta cần có 1 framework cho việc truy xuất hình ảnh dự trên nội dung. Framework này dựa trên các giao thức giữa các module (xem hình 13). Framework này tuân theo các rule: hướng đối tưỡng, nhiệm vụ, lớp trừu tượng, ẩn dữ liệu và giao thức như CORBA.

# 8. Concluding Remarks (Nhận xét và kết luận)

## 8.1 The driving force

Content-base image retrieval phát triển ngày càng nhanh, nó phát triển giống như 1 nền kinh tế, chúng ta biết rằng những gì chúng ta nói ở đây sẽ sớm lỗi thời và chúng ta hy vọng rằng những mô hình tính toán sẽ tồn tại lâu hơn. Lợi thế cho content-base image retrieval là sự phổ biến của các ứng dụng cảm ứng, internet, giá thành rẻ cho các thiết bị lưu trữ. Content-base image retrieval sẽ phát triển theo nhiều hướng khác nhau: đối tượng mới, mục đích mới, cách sử dụng, kiểu tương tác, độ lớn của tập data set, nhiều cách để giải quyết vấn đề. Content-base image cần 1 nền tảng chính xác nhất. 1 số bài báo đã viết về 1 số cách giải quyết vấn đề nhưng không rõ ràng lắm. Một bài toán classifications có nhiều mục địch rất hữu ích ở đây, bao gồm những tiêu chí rất là khó để phân biệt được các lĩnh vực. Mặc dù vài năm gần đây, nó đã được làm rõ rằng content-base image retrieval không phải là “bình cũ rượu mới” nữa. Nó sẽ phụ thuộc về mọi thứ vì chúng tôi tin rằng content-base image retrieval cuối cùng sẽ không hoàn toàn được quyết định trong lĩnh vực computer vision.

## 8.2 The heritage of computer vision

Một khó khăn quan trọng cần khắc phục trước khi content-base image retrieval dựa trên nội dung là nhận ra rằng không cần thiết phải giải quyết toàn bộ vấn đề về hình ảnh đầu vào. Một hệ thống phải đủ để trình bày hình ảnh tương tự dưa trên định nghĩa của người dùng. Tất nhiên, càng đi sâu vào ngữ nghĩa của các bức tranh, sự hiểu biết về bức tranh sẽ càng sâu sắc hơn, nhưng điều đó rất có thể dựa trên việc phân loại hình ảnh thay vì dựa trên sự hiểu biết chính xác. Chúng tôi đã thảo luận về các loại của content-base image retrieval dựa trên nội dung theo ba loại: tìm kiếm mục tiêu, tìm kiếm danh mục và tìm kiếm theo liên kết.

Target search: dựa trên pattern và nhận dạng đối tượng. Những khó khăn là số lượng các đối tượng cần tìm kiếm, tính không hoàn chỉnh, sự thay đổi của môi trường và trạng thái đối tượng.

Category search: xây dựng trên các vấn đề nhận dạng đối tượng và nhận dạng mẫu thống kê. Những khó khăn trong content-base image retrieval so với nhận dạng đối tượng là kết quả thao tác có được, số lượng lớn và quá trình training không được minh bạch.

Search by association: mục đích không được xác định khi bắt đầu. Giới hạn của computer-vision là ở các tính năng và chức năng. Do đó, search by association bị giới hạn bởi ngữ nghĩa.

## 8.3 The influence on computer vision

Content-base image retrieval cung cấp một cái nhìn khác về những vấn đề computer vision truyền thống.

Đầu tiên: content-base image retrieval mang đến một tập dữ liệu lớn trong khi những bài báo cáo gần đây cho thấy thử nghiệm là dưới 100 hình, một bài báo hiện đại trong truy xuất dựa tren hàng nghìn hình ảnh. Tất nhiên, mục đích khác nhau cho computer vision và content-base retrieval. Việc tập hợp dữ liệu của hình ảnh dễ hơn so với tập hợp dữ liệu trong computer-vision. Để xử lý một ngàn hình ảnh đòi hỏi phần mềm và phương pháp phải mạnh.

Thứ 2: content-base retrieval gặp phải tình trạng không có phương pháp cho những segment mạnh. Các hệ thống content-base retrieval phải xử lý dựa trên segment gián đoạn. Đầu tiên, weak of segment đã được giới thiệu trong content-base retrieval. Trong weak of segment, kết quả thu được là những region giống nhất theo tiêu chí.

Thứ 3: content-base retrieval đã phục hồi sự thủ vị trong xử lý ảnh màu. Vì nội dung là hướng đền người dùng nên màu sắc không được bỏ qua. Mục đích của xử lý màu sắc ở đây là giảm ảnh hưởng bởi điều kiện sense.

Cuối cùng: chú ý tính bất biến. Tính bất biến là 1 khía cạnh.

## 8.4 Similarity and learning

Similarity là một cách giải thích của hình ảnh dựa trên sự khác biệt giữa hai yếu tố hoặc 1 nhóm yếu tố. Đối với mỗi loại cần có những biện pháp khác nhau. Đối với sự tương đồng (similiarity) chú ý đến biểu đồ historgram và hiệu quả tính toán. Biểu đồ histogram đang nhận được chú ý rất lớn từ cộng đồng DB cho các tập dữ liệu lớn. Điều này thuận lợi cho việc áp dụng truy xuất trên nhiều lĩnh vực.

Chúng ta thiêng về tầm quan trọng của human-base hơn sự tương đồng chung chung. Ngoài ra, kết nối giữa ngữ nghĩa hình ảnh, dữ liệu hình ảnh và bối cảnh truy vấn sẽ phải được làm rõ hơn trong tương lai.

Learning nhanh chóng thu hút sự chú ý như một phương tiện để xây dựng các mô hình rõ ràng. Việc học ngày nay được thực hiện nhờ sự sẵn có của các tập dữ liệu lớn và các máy móc mạnh, cho phép người dùng tạo thành các label, từ các label được gắn nhãn đến các label không được gắn nhãn. Learning dường như thành công với các tập dữ liệu lớn, được gắn label trên các miền hẹp. Learning từ các tập data được gán label sẵn có khả năng thành công cao hơn so với học không giám sát. Theo quan điểm của chúng tôi, để mang lại ngữ nghĩa cho user, learning là bắt buộc.

## 8.5 Interaction

Chúng tôi coi sự tương tác hình trong image retrieval là 1 trong những thay đổi lớn với computer vision truyền thống (trong hội thảo 1992). Đưa người dùng vào kiểm soát và trực quan hoá content luôn là nguyên tắc hàng đầu trong nghiên cứu thông tin. Ngày càng có nhiều kỹ thuật được sử dụng hoặc phát mình trong content-base image retrieval. Text retrieval và image retrieval chia sẻ nhu cầu trực quan hoá nội dung thông tin theo cách có ý nghĩa. Cũng như chấp nhận phản hồi từ người dùng hơn là 1 nguồn dữ liệu thô.

User tương tác trong image retrieval có đặc điểm khác nhau so với text retrieval. Không cần dịch từ từ ngữ sang yếu tố hình ảnh. Ngoài các loại truy vấn tiêu chuẩn có 6 loại truy vấn tiêu chuẩn khác nhau đã được đề cập trong bài này. Mỗi loại yêu cầu công cụ giao diện và interaction patterns. Trực quan hoá trong không gian image retrieval là cực kỳ quan trong với user. Trong khi đó, 2 hoặc 3 chiều được sử dụng trong truy vấn liên kết, target search and category search theo sau đó. Trong mọi trường hợp, một dòng đồ họa máy tính và thực tế ảo được dự đoán trong tương lai gần. Lập chỉ mục một bộ dữ liệu để sử dụng tương tác là một thách thức lớn vì hệ thống không thể lường trước hoàn toàn hành động của người dùng.

## 8.6 The need for databases

Khi các tập dữ liệu tăng và khi các tập dữ liệu lớn hơn có nhiều vấn đề thú vị hơn, các khía cạnh tính toán có thể bị bỏ qua. Sự kết nối giữa image retrieval dựa trên nội dung và database có thể sẽ tăng lên trong tương lai. Đã có những nỗ lực, nhưng cho đến nay, các vấn đề như định nghĩa ngôn ngữ truy vấn phù hợp, tìm kiếm hiệu quả trong nhiều chiều phần lớn chưa được giải quyết.

Khi các dữ liệu thực sự lớn xuất hiện trong hàng trăm ngàn hình ảnh, cơ sở dữ liệu không nên bị bỏ qua, như một thành phần thiết yếu của hệ thống truy xuất dựa trên nội dung. Ngoài ra, khi hiệu suất tương tác là cần thiết, việc lưu trữ và lập chỉ mục phải được tổ chức trước. Các tập dữ liệu lớn như vậy sẽ có ảnh hưởng đến việc lựa chọn các tính năng như chi phí tính toán và khả năng quyết định hiệu quả của chúng. Đối với các tập dữ liệu rất lớn, không thể bỏ qua việc tích hợp tính toán và lập chỉ mục. Khi nói về index trong computer vision, sự quan trọng vẫn là về những gì cần lập chỉ mục. Sự khác biệt đã trở nên nhỏ hơn gần đây, nhưng chúng tôi tin rằng hầu hết các công việc vẫn phải được thực hiện. Hơn nữa, trong việc xử lý các kích thước vectơ lớn, việc mở rộng các truy vấn và mở rộng truy vấn theo cách hữu ích cho nhiều người dùng vẫn chưa được giải đáp. Để có hiệu quả, cần thực hiện tính năng nén hình ảnh.

## 8.7 The problem of evaluation

Rõ ràng là việc đánh giá hiệu năng hệ thống là điều cần thiết để đánh giá các phương pháp tốt và không tốt. Điều này là do sự phát triển dựa trên nội dung, nhưng cũng do những khó khăn khách quan. Trong đó tương tác là một thành phần quan trọng trong hầu hết các hệ thống, rất khó để đánh giá performance của tập dữ liệu. Sự thành công của hệ thống phụ thuộc rất nhiều vào bộ công cụ của hệ thống so với truy vấn. Ngoài ra, điều hợp lý là một tập dữ liệu lớn bao gồm một số tập dữ liệu nhỏ hơn để có kích thước lớn. Một tiêu chuẩn tham khảo dựa vào đó các thuật toán mới có thể được đánh giá đã giúp lĩnh vực nhận dạng văn bản rất lớn, xem http://trec.nist.gov

## 8.8 The semantic gap and other sources

Một điểm quan trọng trong content-base retrieval là khoảng cách ngữ nghĩa, trong đó ý nghĩa của hình ảnh hiếm khi rõ ràng. Việc sử dụng truy xuất dựa trên nội dung sẽ không nằm trong mong đợi của con người vì con người đã quen với việc dựa vào ngữ nghĩa ngay lập tức khi họ nhìn thấy một hình ảnh và họ mong muốn một máy tính cũng làm như vậy. Mục đích của các hệ thống truy xuất dựa trên nội dung phải là để cung cấp hỗ trợ tối đa trong việc thu hẹp khoảng cách ngữ nghĩa giữa sự đơn giản của các tính năng hình ảnh có sẵn và sự phong phú của ngữ nghĩa người dùng.

Một cách để giải quyết khoảng cách ngữ nghĩa từ các nguồn bên ngoài hình ảnh bằng cách tích hợp các nguồn thông tin khác về hình ảnh trong truy vấn. Thông tin về một hình ảnh có thể đến từ một số nguồn khác nhau: nội dung hình ảnh, nhãn gắn liền với hình ảnh, hình ảnh được nhúng trong văn bản, v.v. Chúng tôi vẫn có những cách rất cơ bản để tích hợp thông tin này để tối ưu hóa. Trong số này, việc tích hợp xử lý ngôn ngữ tự nhiên và thị giác máy tính đáng được chú ý.