Truy xuất hình ảnh dựa trên nội dung vào cuối của thời kỳ những năm đầu

======================================================================

Arnold W.M. Smeulders, \_Senior Member, IEEE\_, Marcel Worring, \_Member, IEEE\_,

Simone Santini, \_Member, IEEE\_, Amarnath Gupta, \_Member, IEEE\_, and Ramesh Jain, \_Fellow, IEEE\_

Tóm tắt - Bài viết trình bày đánh giá về 200 tài liệu tham khảo trong lĩnh vực truy xuất hình ảnh dựa trên nội dung. Bài viết bắt đầu bằng việc thảo luận về các điều kiện hoạt động của truy xuất dựa trên nội dung: các kiểu sử dụng, loại hình ảnh, vai trò của ngữ nghĩa và chênh lệch do cảm biến. Các phần tiếp theo thảo luận về các bước để tính toán cho các hệ thống truy xuất hình ảnh. Bước một trong bài đánh giá này là về việc xử lý hình ảnh cho việc truy xuất được sắp xếp theo màu sắc, kết cấu và hình học cục bộ. Các đặc điểm để truy xuất sẽ được thảo luận tiếp theo, được sắp xếp theo: các đặc điểm tích lũy và toàn cục, các điểm nổi bật, các đặc điểm về đối tượng và hình dạng, các dấu hiệu và kết hợp cấu trúc của chúng. Sự giống nhau của hình ảnh và đối tượng trong hình ảnh được xem xét theo từng loại đặc điểm, liên quan chặt chẽ với các loại và phương tiện phản hồi mà người dùng hệ thống có khả năng cung cấp thông qua tương tác. Chúng tôi thảo luận ngắn gọn về các khía cạnh của kỹ thuật hệ thống: cơ sở dữ liệu, kiến ​​trúc hệ thống và đánh giá. Trong phần kết luận, chúng tôi trình bày quan điểm của chúng tôi về: động lực của lĩnh vực, những tiếp nhận từ thị giác máy tính, những ảnh hưởng đến thị giác máy tính, vai trò của sự giống nhau và sự tương tác, nhu cầu về cơ sở dữ liệu, vấn đề về mặt đánh giá và vai trò của chênh lệch do ngữ nghĩa.

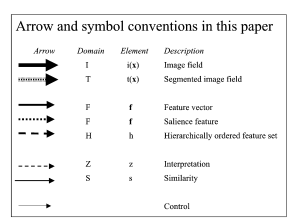
1. Giới thiệu

Có những thứ trong tác phẩm "The Scream" của Munch hay "Wivenoe Park" của Constable mà không từ ngữ nào có thể truyền tải. Chúng phải được quan sát tận mắt. Cũng tương tự đối với bức ảnh về sa mạc Kalahari, một tế bào đang phân chia hoặc biểu cảm trên khuôn mặt của một diễn viên đóng vai Vua Lear. Những điều này ngoài diễn đạt của lời nói. Hãy thử tưởng tượng một biên tập viên chụp ảnh mà không nhìn thấy chúng, một bác sĩ X quang quyết định mô tả bằng lời nói. Hình ảnh phải được xem và tìm kiếm dưới dạng hình ảnh: theo đối tượng, theo phong cách, theo mục đích.

Nghiên cứu về truy xuất hình ảnh dựa trên nội dung ngày nay là một chuyên ngành sôi nổi, đang dần mở ra theo chiều rộng. Như đã xảy ra trong quá trình trưởng thành của nhiều ngành, sau khi thành công sớm trong một vài ứng dụng, nghiên cứu hiện đang tập trung vào các vấn đề sâu hơn, thách thức các vấn đề khó khăn ở ngã rẽ của ngành học sinh ra nó: thị giác máy tính, cơ sở dữ liệu và truy xuất thông tin.

Ở giai đoạn hiện tại của nghiên cứu truy xuất hình ảnh dựa trên nội dung, thật thú vị khi nhìn lại từ đầu và xem những ý tưởng ban đầu nào đã nở rộ, những ý tưởng nào không nở rộ, và những ý tưởng đã bị lỗi thời bởi bối cảnh luôn thay đổi của điện toán. Vào tháng 2 năm 1992, Quỹ Khoa học Quốc gia Hoa Kỳ (USNSF) đã tổ chức một hội thảo tại Redwood, California, để "xác định các lĩnh vực nghiên cứu chính cần được các nhà nghiên cứu giải quyết cho các hệ thống quản lý thông tin trực quan sẽ hữu ích trong khoa học, công nghiệp, y tế, môi trường, giáo dục, giải trí và các ứng dụng khác "[81]. Nhìn nhận lại, hội thảo đã làm một công việc tuyệt vời để xác định các vấn đề chưa được giải quyết mà các nhà nghiên cứu nên thực hiện. Cụ thể, hội thảo đã tuyên bố chính xác rằng "Hệ thống quản lý thông tin thị giác không nên được coi là một ứng dụng của kỹ thuật tiên tiến (trong thị giác máy tính và cơ sở dữ liệu) để quản lý và xử lý hình ảnh" và "các nhà nghiên cứu thị giác máy tính nên xác định các đặc điểm cần thiết cho \_sự hiểu biết về hình ảnh tương tác\_, thay vì nhấn mạnh của ngành đối với các kỹ thuật tự động trong thời điểm hiện tại" (đã thêm đoạn in nghiêng nhấn mạnh). Về các lĩnh vực ứng dụng có thể có, hội thảo đã xem xét chủ yếu các vấn đề Grand Challenge, như dự báo thời tiết, mô hình sinh học, hình ảnh y tế, hình ảnh vệ tinh, v.v. Không còn nghi ngờ gì nữa, những người tham gia đã thấy đủ để khẳng định cho việc sử dụng dung lượng tính toán và lưu trữ lớn cần thiết cho cơ sở dữ liệu trực quan. Điều này diễn ra vào năm 1992. Trước hội thảo này nhiều năm, Hội nghị về các ứng dụng cơ sở dữ liệu của các ứng dụng hình ảnh được tổ chức tại Florence vào năm 1979, có lẽ là một trong những hội nghị đầu tiên thuộc loại đó [13]. Trong phần giới thiệu, người ta nói rằng: "Một mặt, điều này đã tạo điều kiện cho sự tiến bộ của cơ sở dữ liệu tích hợp [...], mặt khác là sự tiến bộ của các ứng dụng xử lý đồ họa và hình ảnh (nói ngắn gọn: hình ảnh)." Sau đó, tác giả tiếp tục phàn nàn rằng: "Sự phát triển trong hai lĩnh vực này về mặt truyền thống không liên quan đến nhau", một quan điểm vẫn còn rất chính xác ngày nay.

Mũi tên và quy ước biểu tượng trong bài viết này



Hình 1. Luồng dữ liệu và quy ước ký hiệu như được sử dụng trong bài viết này. Các kiểu mũi tên khác nhau thể hiện các cấu trúc dữ liệu khác nhau.

Ngay sau hội thảo USNSF, trình duyệt Internet Mosaic đã được phát hành, tạo ra cuộc cách mạng Web nhanh chóng thay đổi cục diện. Trong cùng một thời đại, một loạt các cảm biến thị giác kỹ thuật số mới đã xuất hiện. Số lượng hình ảnh mà người dùng trung bình có thể đạt được tăng lên đáng kể chỉ trong một vài năm. Ngay lập tức, các công cụ lập chỉ mục của Web hoặc các kho lưu trữ kỹ thuật số trở nên cấp bách.

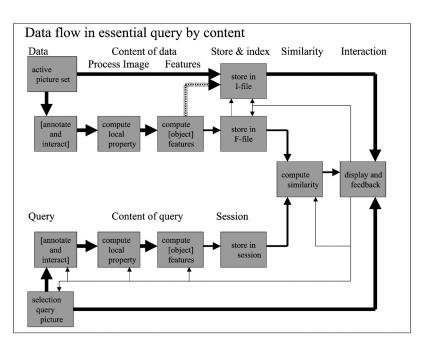
Trong bài viết này, chúng tôi trình bày một quan điểm về những gì chúng tôi muốn gọi là "thời kỳ những năm đầu" của việc truy xuất hình ảnh dựa trên nội dung. Mặc dù các bài báo truy xuất hình ảnh dựa trên nội dung được xuất bản trước năm 1990 rất hiếm, gần như chắc chắn đã lỗi thời và ít ảnh hưởng trực tiếp ngày nay, số lượng bài báo được xuất bản từ năm 1997 thì rất ngoạn mục. Nhiều tới mức, trên thực tế, việc biên soạn một đánh giá toàn diện về kỹ thuật tiên tiến thời đó đã vượt quá khả năng của một bài nghiên cứu như thế này. Việc chọn lọc là cần thiết và cùng với đó là sự cần thiết phải thiết lập một số tiêu chí chọn lọc. Bên cạnh các tiêu chí hiển nhiên (tính hoàn thiện của bài viết, tầm quan trọng tới lĩnh vực), chúng tôi cũng đã xem xét tới tiêu chí khả năng tiếp cận người đọc. Điều đó có nghĩa là, chúng tôi đã ưu tiên lựa chọn các bài viết tạp chí hơn là các bài viết hội nghị, bất cứ khi nào có thể. Chúng tôi cũng cảm thấy rằng lĩnh vực này còn quá non trẻ và biến đổi nhanh để tạo một danh mục lịch sử chính xác và chúng tôi đã không nỗ lực theo hướng đó.

Chúng tôi lựa chọn các kiểu sử dụng và các kiểu tính toán làm nguyên tắc hàng đầu trong bài đánh giá của chúng tôi. Chúng tôi theo dõi dữ liệu đi qua quy trình tính toán và xem xét các quy trình thay thế có cùng vị trí trong luồng (Hình 2). Trong các sơ đồ luồng dữ liệu, chúng tôi sử dụng các quy ước được chỉ ra trong Hình 1. Chúng tôi tập trung vào các phương pháp tính toán để hướng tới tổng quan dựa trên công cụ thay vì tổng quan dựa trên hệ thống. Sự lựa chọn này ngụ ý rằng các tham chiếu mô tả các hệ thống hoàn chỉnh được phân chia, trong đó các phần của phương thức sẽ xuất hiện trong một số phần của bài viết. Về tổng quan dựa trên hệ thống, xem [141].

Chúng tôi bó buộc chủ đề trong giới hạn hình ảnh tĩnh và coi cơ sở dữ liệu video như một chủ đề riêng biệt. Truy xuất video có thể được coi là một chủ đề rộng hơn so với truy xuất hình ảnh vì video được xây dựng từ các hình ảnh đơn lẻ. Từ góc độ khác, việc truy xuất video có thể được coi là đơn giản hơn so với truy xuất hình ảnh do video tiết lộ các đối tượng dễ dàng hơn khi các điểm tương ứng với một đối tượng di chuyển cùng nhau. Trong hình ảnh tĩnh, ý định truyền tải câu chuyện của tác giả nằm ở việc chọn khung, chiếu sáng và bố cục. Ngoài ra, video có dòng thời gian tuyến tính, đặc điểm quan trọng đối với cấu trúc câu chuyện của video, tương tự như trong văn bản. Chúng tôi sẽ bàn về truy xuất video ở vị trí khác, ví dụ, [1], [16].

Bài viết được kết cấu như trong Hình 2. Đầu tiên chúng ta thảo luận về phạm vi của truy xuất dựa trên nội dung trong Phần 2. Trong phần đó, các đặc điểm của miền và nguồn kiến ​​thức đang được thảo luận. Sau đó, mô tả nội dung được phân tích theo hai bước. Đầu tiên, trong Phần 3, các phương pháp xử lý ảnh theo màu sắc, kết cấu và hình dạng cục bộ sẽ được thảo luận. Chúng đóng vai trò là bước tiền xử lý để phân vùng mảng dữ liệu và tính toán các đặc điểm, như được thảo luận trong Phần 4, Trong Phần 5, chúng tôi thảo luận về việc diễn giải một hình ảnh đơn lẻ và sự giống nhau giữa một cặp hình ảnh. Định nghĩa truy vấn, hiển thị và tương tác là chủ đề của Phần 6. Bài viết kết luận ở cấp độ hệ thống: lập chỉ mục, kiến ​​trúc hệ thống và đánh giá hiệu suất. Mỗi chương được kết thúc bằng thảo luận về kỹ thuật tiên tiến hiện tại.

Luồng dữ liệu trong truy vấn thiết yếu theo nội dung



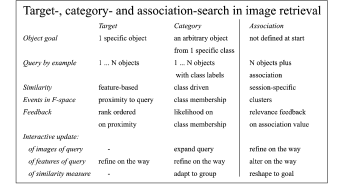
Hình 2. Các thành phần thuật toán cơ bản của truy vấn bằng ví dụ hình ảnh được chụp trong sơ đồ luồng dữ liệu, sử dụng các quy ước của Hình 1.

2. Phạm vi

Trong các tài liệu nghiên cứu, có thể thấy rất nhiều phương pháp và hệ thống truy xuất dựa trên nội dung. Trong phần này, chúng tôi thảo luận về các kiểu ứng dụng, danh mục hình ảnh, ảnh hưởng của khung cảnh và vai trò của kiến ​​thức miền, và chênh lệch do ngữ nghĩa giữa các đặc điểm hình ảnh và người dùng.

2.1. Các ứng dụng của truy xuất dựa trên nội dung

Tìm kiếm theo mục tiêu, phân loại và mối liên hệ trong truy xuất hình ảnh



Hình 3. Ba kiểu mục đích của các hệ thống truy xuất dựa trên nội dung.

Trong [31], chúng tôi thấy ba loại mục đích người dùng rộng lớn khi sử dụng hệ thống, xem Hình 3.

\* Có một lớp rộng lớn các phương pháp và hệ thống nhằm mục đích duyệt qua một tập hợp lớn các hình ảnh từ các nguồn không xác định. Người dùng \_tìm kiếm theo mối liên hệ (association)\_ ngày từ đầu thường không có mục đích cụ thể nào ngoài việc tìm kiếm những điều thú vị. Tìm kiếm theo mối liên hệ thường bao hàm sự tinh chỉnh lặp đi lặp lại của tìm kiếm, sự giống nhau hoặc các ví dụ mà tìm kiếm được bắt đầu. Các hệ thống trong thể loại này thường có tính tương tác cao, trong đó đặc tả có thể được phác họa [30] hoặc đưa ra bằng hình ảnh ví dụ. Ví dụ thực tế lâu đời nhất của một hệ thống như vậy có lẽ là [88]. Kết quả tìm kiếm có thể được thao tác tương tác bằng phản hồi liên quan [68], [51]. Để hỗ trợ tìm kiếm các kết quả có liên quan, các nguồn khác ngoài hình ảnh cũng được sử dụng, ví dụ [168], [21].

\* Một lớp người dùng khác \_tìm kiếm theo mục tiêu\_ vào một hình ảnh cụ thể. Kết quả tìm kiếm có thể là một bản sao chính xác của hình ảnh trong tâm trí, như khi tìm kiếm trong danh mục tác phẩm nghệ thuật, ví dụ: [48]. Tìm kiếm theo mục tiêu cũng có thể áp dụng cho một hình ảnh khác của cùng một đối tượng mà người dùng có hình ảnh. Đây là tìm kiếm theo mục tiêu bằng ví dụ. Tìm kiếm theo mục tiêu cũng có thể được áp dụng khi người dùng có một hình ảnh cụ thể và mục tiêu được chỉ định tương tác tương tự như một nhóm các ví dụ đã cho, ví dụ [31]. Các hệ thống này phù hợp để tìm kiếm tem, nghệ thuật, linh kiện công nghiệp và danh mục nói chung.

\* Lớp ứng dụng thứ ba, \_tìm kiếm theo phân loại\_, nhắm vào việc truy xuất một đại diện hình ảnh tùy ý của một lớp cụ thể. Nó có thể là trường hợp người dùng có một ví dụ và kết quả tìm kiếm là các yếu tố khác của cùng một lớp. Phân loại có thể được truy xuất lấy từ nhãn hoặc xuất hiện từ cơ sở dữ liệu [170], [186]. Trong tìm kiếm theo phân loại, người dùng có thể có sẵn một nhóm hình ảnh và kết quả tìm kiếm là hình ảnh bổ sung của cùng một lớp [28]. Một ứng dụng điển hình của tìm kiếm theo phân loại là danh mục các biến thể. Trong [74], [79], các hệ thống được thiết kế để phân loại nhãn hiệu. Các hệ thống trong danh mục này thường tương tác với một định nghĩa riêng của miền về sự giống nhau.

Ba kiểu sử dụng trên không bao quát hết tất cả các trường hợp [42]. Nghiên cứu [121] trên tạp chí đã xác định năm kiểu sử dụng điển hình: tìm kiếm một hình ảnh cụ thể, duyệt tổng thể để đưa ra lựa chọn tương tác, tìm kiếm để minh họa một tài liệu và tìm kiếm các phần chỉ để bổ trợ cho giá trị thẩm mỹ của hình ảnh. Các nỗ lực để hình thành một phân loại chung các yêu cầu của người dùng đối với các hình ảnh tĩnh và chuyển động được tìm thấy trong [6]. Nghiên cứu này và các nghiên cứu tương tự cho thấy phạm vi truy vấn rộng hơn so với việc chỉ truy xuất hình ảnh dựa trên sự hiện diện hoặc vắng mặt của các đối tượng có đặc điểm hình ảnh đơn giản.

2.2. Miền hình ảnh và chênh lệch do cảm biến

Trong các danh mục của hình ảnh được xem xét -- miền hình ảnh `I` -- có sự phân biệt chuyển tiếp giữa các miền hẹp và rộng [160]. Ở một đầu của phổ phân biệt, chúng ta có miền hẹp:

> Một miền hẹp có độ biến thiên có giới hạn và có thể dự đoán được trong tất cả các khía cạnh liên quan về diện mạo của nó.

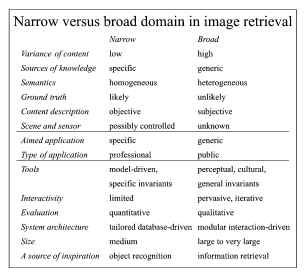
Trong một miền hẹp, người ta nhận thấy độ biến thiên có giới hạn của nội dung của hình ảnh. Thông thường, các hoàn cảnh ghi hình là tương tự nhau trên toàn bộ miền. Ví dụ, trong miền hẹp của thạch bản, việc ghi hình được thực hiện dưới ánh sáng trắng với góc nhìn từ phía trước và không bị cản trở. Ngoài ra, khi sự xuất hiện của đối tượng có độ biến thiên giới hạn, mô tả ngữ nghĩa của hình ảnh thường được xác định rõ và thường là duy nhất. Một ví dụ khác về miền hẹp là một tập hợp các góc nhìn phía trước của các khuôn mặt được ghi trên nền trống. Mặc dù mỗi khuôn mặt là duy nhất và có độ biến thiên lớn trong các chi tiết hình ảnh, có những ràng buộc rõ ràng về hình học, vật lý và màu sắc liên quan đến miền. Miền sẽ rộng hơn nếu các khuôn mặt được chụp từ đám đông hoặc từ khung cảnh ngoài trời. Trong trường hợp đó, các biến đổi về chiếu sáng, sự hỗn loạn trong khung cảnh, các vật cản và góc nhìn sẽ có tác động lớn đến việc phân tích.

Ở đầu kia của phổ, chúng ta có miền rộng:

Một miền rộng có sự biến thiên không giới hạn và không thể đoán trước về diện mạo của nó ngay cả với cùng một ý nghĩa ngữ nghĩa.

Trong các miền rộng, hình ảnh là đa hình và ngữ nghĩa của chúng chỉ được mô tả một phần. Nó có thể là trường hợp có các đối tượng dễ thấy trong khung cảnh mà lớp đối tượng không xác định hoặc thậm chí sự diễn giải cho khung cảnh không phải là duy nhất. Một lớp rộng của hình ảnh có thể được tìm thấy trong các kho ảnh lớn [168] hoặc các kho lưu trữ ảnh khác [42]. Lớp rộng nhất có sẵn cho đến nay là tập hợp các hình ảnh tồn tại trên Internet.

So sánh miền hẹp và miền rộng trong truy xuất hình ảnh



Hình 4. Tham chiếu nhanh về so sánh giữa miền hẹp và miền rộng.

Nhiều vấn đề được quan tâm trên thực tiễn lại có miền hình ảnh ở giữa hai đầu cực phổ này, xem Hình 4. Các khái niệm rộng và hẹp rất hữu ích trong việc mô tả các kiểu sử dụng, trong việc chọn lọc các đặc điểm và trong thiết kế hệ thống. Trong miền hình ảnh rộng, chênh lệch giữa mô tả đặc điểm và diễn giải ngữ nghĩa thường rộng. Đối với các miền hình ảnh hẹp, chuyên biệt, chênh lệch giữa các đặc điểm và diễn giải ngữ nghĩa của chúng thường nhỏ hơn, do đó các mô hình riêng của miền có thể hữu ích. Đối với khuôn mặt, nhiều mô hình hình học đã được đề xuất, cũng như các mô hình được học qua thống kê [127]. Các mô hình tính toán này không có sẵn cho các miền hình ảnh rộng vì số lượng mô hình tính toán cần thiết là rất lớn.

Đối với các miền hình ảnh rộng nói riêng, người ta phải sử dụng các nguyên tắc hợp lệ nói chung. Liệu sự chiếu sáng của miền có phải màu trắng để có ánh sáng màu? Liệu nó giả định các đối tượng được xác định và hiển thị đầy đủ, hay liệu khung cảnh có thể chứa các đối tượng lộn xộn và cản trở? Đây có phải là bản ghi 2D của khung cảnh 2D hay bản ghi 2D của khung cảnh 3D không? Các đặc điểm nhất định của chiếu sáng, hiện diện hoặc thiếu vắng vật cản, lộn xộn và sự khác biệt trong góc nhìn camera, tất cả quyết định các yêu cầu về phương pháp truy xuất.

> Chênh lệch do cảm biến là khoảng cách giữa vật thể trong thế giới thực và thông tin trong một mô tả (điện toán) có nguồn gốc từ một bản ghi hình của khung cảnh đó.

Chênh lệch do cảm biến khiến cho việc mô tả các đối tượng trở thành một bài toán được mô tả tồi: Nó mang lại sự không chắc chắn trong những gì được biết về trạng thái của đối tượng. Chênh lệch do cảm biến đặc biệt sâu sắc khi thiếu kiến thức chính xác về các điều kiện ghi hình. Các bản ghi 2D của các đối tượng 3D khác nhau có thể giống hệt nhau. Nếu không có thêm kiến ​​thức, người ta phải quyết định rằng chúng \_có khả năng\_ đại diện cho cùng một đối tượng. Ngoài ra, bản ghi 2D của cảnh 3D chứa thông tin tình cờ cho cảnh đó và cảm biến đó, nhưng người ta không biết phần nào của thông tin có liên quan đến cảnh. Sự không chắc chắn do chênh lệch do cảm biến không chỉ xảy ra theo góc nhìn, mà còn theo các vật cản (mà khiến các phần thiết yếu giúp phân biệt hai vật thể có thể nằm ngoài tầm nhìn), vật lộn xộn và chiếu sáng.

Việc so sánh các cách diễn giải khác nhau có thể làm giảm chênh lệch do cảm biến. Các hệ thống truy xuất hình ảnh dựa trên nội dung có thể cung cấp hỗ trợ cho việc phân biệt thông qua việc loại bỏ một số giải thích có khả năng, giống như trong xử lý ngôn ngữ tự nhiên.

2.3. Kiến thức miền

Các nguồn kiến ​​thức chung được sắp xếp theo các lớp đồng đẳng

cultural = văn hóa

categorical = phân loại

geometric = hình học

physical = vật lý

perceptual = tri giác

literal = nghĩa đen

```

Hình 5. Khi tìm kiếm một chiếc "ghế", chúng ta có thể hài lòng với bất kỳ đối tượng nào có tên đó, nghĩa là chúng ta tìm kiếm các vật đồng đẳng theo định nghĩa của con người. Khi chúng ta tìm kiếm tất cả các ghế một chân, chúng ta thêm một ràng buộc bổ sung cho lớp chung và giới hạn lớp đồng đẳng. Điều tương tự cũng đúng khi tìm kiếm một chiếc ghế màu đỏ, tức là thêm một điều kiện độc lập với các ràng buộc hình học. Khi chúng ta tìm kiếm một chiếc ghế đồng đẳng về mặt tri giác đối với một chiếc ghế nhất định, ít nhất phải có sự đồng đẳng về vật lý và hình học. Cuối cùng, khi chúng ta tìm kiếm chính xác chiếc ghế có cùng hình ảnh với chiếc ghế đó, sự đồng đẳng theo nghĩa đen được yêu cầu, tất nhiên vẫn bỏ qua các biến thể do nhiễu.

Trong tìm kiếm trực quan, việc trình bày rõ ràng kiến ​​thức về miền là rất quan trọng để giảm bớt chênh lệch do cảm biến. Trong số các nguồn kiến ​​thức chung, chúng tôi đề cập đến:

\* Các quy luật về sự đồng đẳng và giống nhau về cú pháp (nghĩa đen) xác định mối quan hệ giữa các pixel hình ảnh hoặc các đặc điểm hình ảnh bất kể nguyên nhân vật lý hoặc tri giác nào về nó. Chẳng hạn, việc xem xét hai hình ảnh tương tự nhau vì cả hai đều thể hiện một số sắc thái màu xanh lam có chọn lọc ở phần trên của hình có hiệu quả trong việc tách các cảnh ngoài trời khỏi các hình ảnh khác. Đó là sự giống nhau về mặt cú pháp bởi vì phương pháp này không tham chiếu đến các lý do mà sự giống nhau này tồn tại (trong trường hợp này là sự tán xạ trên bầu trời) hoặc lý do tri giác mà hai hình ảnh này sẽ xuất hiện tương tự đối với một người quan sát. Tương tự như vậy, không gian màu RGB có hiệu quả trong sự giống nhau theo nghĩa đen (vì nó hiệu quả trong nghệ thuật [65]) trong khi nó không đại diện cho quá trình hình thành màu vật lý hoặc quá trình nhận thức màu sắc.

\* Các quy luật mô tả tri giác của con người về sự đồng đẳng và tương đồng rất quan trọng bởi vì chúng xác định sự đồng đẳng trên cơ sở người dùng trải nghiệm sự đồng đẳng. Về màu sắc, CIE-Lab và không gian Munsell được thiết kế để phù hợp với tri giác của con người về một số vật phẩm dễ thấy nhất định trong hình ảnh [177], việc hướng tính toán các đặc điểm miền rộng đến các điểm và vùng này [157], [138] là điều tự nhiên. Tương tự, một kiến ​​trúc hợp lý về mặt sinh học [76] của các đơn vị xử lý bao quanh trung tâm có khả năng chọn các khu vực mà con người cũng sẽ tập trung vào đầu tiên.

\* Các quy luật vật lý mô tả sự đồng đẳng và khác biệt của hình ảnh dưới sự khác biệt về tính chất cảm biến và bề mặt đối tượng. Vật lý của sự chiếu sáng, phản xạ bề mặt và hình thành hình ảnh có ảnh hưởng chung đến hình ảnh. Các quy luật vật lý chung có thể được sử dụng cho các lớp đối tượng lớn. Một ví dụ phổ biến là quy luật phản xạ ánh sáng đồng đều từ các vật thể mờ. Các quy luật này được khai thác để thiết kế các đặc điểm màu sắc thể hiện sự đồng đẳng bất kể tư thế và góc nhìn.

\* Các quy tắc hình học và tôpô mô tả sự đồng đẳng và khác biệt của các kiểu mẫu trong không gian. Khi hai vật thể đồng đẳng nhau về mặt hình học, tính chất vật lý của bề mặt hoặc điều kiện vật lý của cảm biến có thể khác nhau. Một ví dụ về các định luật hình học được sử dụng trong truy xuất là đối với tất cả các hình ảnh có chiều sâu, các chi tiết cục bộ gần đường chân trời sẽ xuất hiện nhỏ hơn. Ngoài ra, đường chân trời được định nghĩa hình học là một đường ảo chứa các tiêu điểm. Một ví dụ khác về các định luật hình học là sự biểu hiện của các mối quan hệ không gian [22] hoặc tôpô [172] giữa các đối tượng.

\* Các quy tắc dựa trên phân loại mã hóa các đặc điểm chung cho lớp `z` của không gian của tất cả các khái niệm `Z`. Nếu `z` là lớp chứa tất cả các ấm trà, các đặc điểm sẽ bao gồm sự hiện diện của vòi ấm. Các phân loại hầu như chỉ được sử dụng trong miền hẹp. Kiến thức miền có thể đóng vai trò ràng buộc bổ sung đối với các phẩm chất hình ảnh theo nghĩa đen, các định luật vật lý hoặc hình học bổ sung hoặc các tập quán nhân tạo riêng của miền. Khi miền là bản vẽ kỹ thuật, kiến ​​thức hình học chi tiết sẽ chỉ đạo việc phát hiện các ký hiệu. Trong nghệ thuật thời trung cổ, màu sắc và vị trí tương đối của các vật thể có ý nghĩa tượng trưng [30], tạo ra một tập hợp các ràng buộc hữu ích trong tìm kiếm. Mỗi miền ứng dụng có một tập các ràng buộc riêng.

\* Cuối cùng, các tập quán nhân tạo hoặc các kiểu mẫu liên quan đến con người đưa ra các quy tắc về sự đồng đẳng và khác biệt dựa trên văn hóa. Khi nói về văn hóa, chúng tôi cũng giả định tới ngôn ngữ. Trong quá trình tìm kiếm hình ảnh trong nhà, người ta có thể kiểm tra nhiều đường thẳng và góc vuông góc như là một tiêu chí lựa chọn đầu tiên. Đồ dùng có kích thước xác định để cho phép cầm. Thời trang quyết định màu sắc [95].

Các quy luật này được sắp xếp như trong Hình 5.

2.4. Công dụng và người dùng, chênh lệch do ngữ nghĩa

Chúng tôi cho rằng hầu hết sự thất vọng với các hệ thống truy xuất sớm xuất phát từ việc không nhận ra sự tồn tại của chênh lệch do ngữ nghĩa và hậu quả của nó đối với việc thiết lập hệ thống.

> Chênh lệch do ngữ nghĩa là việc thiếu sự trùng lắp giữa thông tin mà người ta có thể trích xuất từ ​​dữ liệu trực quan và sự diễn giải cùng dữ liệu đó của người dùng trong một tình huống nhất định.

Mô tả bằng ngôn ngữ hầu như luôn luôn theo ngữ cảnh, trong khi một hình ảnh có thể đứng tách biệt. Mô tả ngôn ngữ của một hình ảnh là một nhiệm vụ khó khăn, và có lẽ là không thể [146]. Người dùng thường tìm kiếm hình ảnh có chứa các đối tượng nhất định hoặc truyền tải một thông điệp nhất định. Mô tả hình ảnh, mặt khác, phụ thuộc vào các đặc điểm dựa trên dữ liệu và cả hai có thể không liên quan nhau. Sự liên kết giữa một hệ thống ngữ nghĩa hoàn chỉnh với dữ liệu hình ảnh sẽ đòi hỏi ít nhất là giải quyết vấn đề nhận dạng đối tượng chung từ một hình ảnh duy nhất. Vì vấn đề này chưa được giải quyết, nghiên cứu tập trung vào các phương pháp khác nhau để liên kết ngữ nghĩa cấp cao hơn với các quan sát dựa trên dữ liệu.

Như được chỉ ra trong hình 2, các phương tiện đặc trưng ngữ nghĩa trực tiếp nhất đòi hỏi chú thích bằng từ khóa hoặc chú thích. Điều này làm sự tiếp cận dựa trên nội dung để truy xuất thông tin [135]. Những phản đối phổ biến đối với việc dán nhãn là chi phí và độ bao quát. Về mặt chi phí, việc dán nhãn hàng ngàn hình ảnh là một công việc cồng kềnh và tốn kém đến mức việc triển khai tính cân đối kinh tế của cơ sở dữ liệu có thể sẽ giảm. Để giải quyết vấn đề, các hệ thống trong [21], [142] sử dụng chương trình khám phá Internet, thu thập hình ảnh và chèn chúng vào hệ phân loại được xác định trước dựa trên cơ sở văn bản xung quanh chúng. Một cách tiếp cận tương tự cho các thư viện kỹ thuật số được thực hiện bởi [24]. Về mặt độ bao quát, việc ghi nhãn hiếm khi hoàn tất, phụ thuộc nhiều vào ngữ cảnh và trong mọi trường hợp, có một phần đáng kể các yêu cầu mà ngữ nghĩa không thể được ghi lại chỉ bằng cách dán nhãn [6], [64]. Cả hai phương pháp sẽ chỉ nối lại chênh lệch do ngữ nghĩa trong các trường hợp đơn lẻ.

2.5. Thảo luận về Phạm vi

Điểm mấu chốt trong truy xuất dựa trên nội dung là người dùng tìm kiếm sự giống nhau về mặt ngữ nghĩa, nhưng cơ sở dữ liệu chỉ có thể cung cấp sự giống nhau bằng cách xử lý dữ liệu. Đây là những gì chúng tôi gọi là chênh lệch do ngữ nghĩa. Đồng thời, chênh lệch do cảm biến giữa các thuộc tính trong một hình ảnh và các thuộc tính của đối tượng đóng vai trò làm hạn chế việc truy xuất nội dung của hình ảnh.

Chúng tôi đã thảo luận về các ứng dụng truy xuất dựa trên nội dung theo ba loại rộng: tìm kiếm theo mục tiêu, tìm kiếm theo phân loại và tìm kiếm theo mối liên hệ. Tìm kiếm theo mục tiêu liên quan tới truyền thống khớp mẫu trong thị giác máy tính. Những thách thức mới trong truy xuất dựa trên nội dung là số lượng lớn các đối tượng để tìm kiếm, các đặc tả truy vấn không đầy đủ, các mô tả hình ảnh không đầy đủ và sự biến thiên của điều kiện cảm biến và trạng thái đối tượng. Tìm kiếm theo phân loại được xây dựng trên các phương pháp nhận dạng đối tượng và nhận dạng mẫu thống kê trong thị giác máy tính. Những thách thức mới trong truy xuất dựa trên nội dung nếu so với thành tựu của nhận dạng đối tượng là các thao tác tương tác trên kết quả, số lượng lớp đối tượng thường rất lớn và việc thiếu giai đoạn đào tạo rõ ràng để tinh chỉnh đặc điểm và phân loại.

Tìm kiếm theo mối liên hệ càng ngày bị loại bỏ khỏi hầu hết các truyền thống thị giác máy tính. Nó bị cản trở nhiều nhất bởi chênh lệch do ngữ nghĩa. Chừng nào còn chênh lệch này, việc sử dụng truy xuất dựa trên nội dung để duyệt sẽ không nằm trong tầm ngắm của đại chúng vì con người đã quen với việc dựa vào ngữ nghĩa ngay lập tức khi họ nhìn thấy hình ảnh. 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 chênh lệch do ngữ nghĩa giữa sự đơn giản của các đặc điểm hình ảnh có sẵn và sự phong phú của ngữ nghĩa người dùng.

Chúng tôi phân tích các đặc điểm của miền hình ảnh, kiến ​​thức miền và các loại sử dụng làm yếu tố chính quyết định chức năng của hệ thống. Một sự khác biệt quan trọng là giữa các miền rộng và hẹp. Miền càng rộng, việc duyệt hoặc tìm kiếm theo mối liên hệ càng có thể là giải pháp phù hợp. Miền càng hẹp, việc ứng dụng kiến ​​thức miền sẽ càng thành công. \_Thách thức đối với các công cụ tìm kiếm hình ảnh trên một miền rộng là điều chỉnh công cụ theo miền hẹp mà người dùng có trong tâm trí thông qua đặc tả, ví dụ và tương tác.\_