**Nội dung**

[**4. Description of Content: Features** 2](#_Toc30451034)

[**4.1. Grouping Data** 2](#_Toc30451035)

[**4.2. Global and Accumulating Features** 5](#_Toc30451036)

[**4.3. Salient Features** 8](#_Toc30451037)

[**4.4. Signs** 10](#_Toc30451038)

[**4.5. Shape and Object Features** 10](#_Toc30451039)

[**4.6. Description of Structure and Lay-Out** 12](#_Toc30451040)

[**4.7. Discussion on Features** 14](#_Toc30451041)

Segmentation: Khoanh vùng | Phân đoạn | ...

|  |
| --- |
| Arrow and symbol conventions in i  Arrow  Domain  1  z  Element  i(x)  t(x)  z  Description  Image field  Segmented ima±  Feature vector  Salience feature  Hierarchically o  Interoretation |
| Fig. 1. Data flow and symbol conventions as used in this paper. Different styles of arrows indicate different data structures. |
| Hình 1. Quy ước kí hiệu và "data flow" sử dụng trong bài báo này. Kiểu mũi tên khác nhau chỉ thị cấu trúc dữ liệu khác nhau. |

|  |
| --- |
| Grouping the image  Strong  segmentation  Weak  segmentation  Sign  detection  Segmented obje  Salient regions  Location of si |
| Fig. 7. Symbolic representation of different ways to group image data. |
| Hình 7. Biểu diễn tượng trưng cho các cách khác nhau để nhóm dữ liệu ảnh. |

**4. Description of Content: Features**

*Mô tả về nội dung (bằng): Các đặc trưng)*

In the first section, we discuss the ultimate form of spatial data by grouping the data into object silhouettes, clusters of points or point-sets. In subsequent sections, we leave the spatial domain to condense the pictorial information into feature values.

Phần đầu bàn luận về hình dạng cuối của dữ-liệu-không-gian bằng cách gom nhóm dữ liệu thành những đối tượng ngược sáng (đối tượng mang màu tối, xung quanh sáng hơn), những cụm các điểm hoặc những tập hợp điểm. Trong phần này, sử dụng miền không gian để cô đọng thông tin hình ảnh thành các giá trị đặc trưng.

**4.1. Grouping Data**

*(Nhóm dữ diệu)*

In content-based image retrieval, the image is often divided in parts before features are computed from each part, see Fig. 7. Partitionings of the image aim at obtaining more selective features by selecting pixels in a trade-off against having more information in features when no subdivision of the image is used at all. We distinguish the following partitionings:

Trong tra cứu ảnh dựa vào nội dung, hình ảnh thường được chia thành nhiều phần trước khi các đặc trưng được rút trích (tính toán) từ mỗi phần trên, xem hình 7. Các phân vùng của ảnh có mục đích lấy được nhiều hơn các đặc trưng chọn lọc bằng cách chọn các điểm ảnh, nhưng đánh đổi bằng việc có ít thông tin trong các đặc trưng hơn so với khi không có chia nhỏ hình ảnh. Phân biệt các kiểu phân vùng sau đây:

* When searching for an object, it would be most advantageous to do a complete object segmentation

|  |
| --- |
| Feature types  Image  partitionings  Weak  segmentation  Sign  location  Strong  Global  features  Salient  features  Sign  probabilities  Object  features |
| Fig. 8. The different types of features using the data flow conventions of Fig. 1. |
| Hình 8. Các kiểu đặc trưng khác nhau, sử dụng quy ước "data flow" của Hình 1. |

first: "Strong segmentation is a division of the image data into regions in such a way that region contains the pixels of the silhouette of object in the real world and nothing else, specified by: ."

Khi thực hiện tìm kiếm một đối tượng, sẽ thuận tiện nhất nếu lấy được nguyên-mảnh-hoàn-chỉnh-của-đối-tượng trước (hay tách được đối tượng ra khỏi ảnh trước). Một mảnh tốt là một vùng của ảnh (T) chứa các điểm ảnh của đối tượng (O) mà không lẫn điểm ảnh không liên quan, kí hiệu .

It should be noted immediately that object segementation for broad domains of general images is not likely to successed, with a possible exception for sophisticated techniques in very narrow domains.

Cũng nên chú ý rằng, tách vật thể (tách đối tượng) từ những hình ảnh có nội dung chung chung, nội dung dàn trải, phạm vi rộng, dường như khó thành công, trừ khi sử dụng những kĩ thuật tinh vi đối trong những phạm vi hẹp.

* The difficulty of achieving strong segmentation may be circumvented by weak segmentation where grouping is based on data-driven properties: "Weak segmentation is a grouping of the image data in conspicuous regions T internally homogenous according to some criterion, hopefully with ."

Để có thể phá vỡ khó khăn trong việc đạt được strong segmentation, sử dụng weak segmentation, việc gom nhóm dựa vào các tính chất data-driven: "Weak segmentation là một nhóm dữ liệu của hình ảnh thể hiện một vùng rõ ràng T đồng nhất dựa trên một vài tiêu chí, với hi vọng rằng ."

The criterion is satisfied if region is within the bounds of object , but there is no guarantee that the region covers all of the object's area. When the image contains two nearly identical objects close to each other, the weak segmentation algorithm may falsely observe just one patch. Fortunately, in content-based retrieval, this type of error is rarely obstructive for the goal. In [125], the homogenity criterion is implemented by requesting that colors be spatially coherent vectors in region. Color is the criterion in [50], [126]. In [20], [112], the homogeneity criterion is based on color and texture. The limit case of weak segmentation is a set of isolated points [148], [57]. No homogeneity criterion is needed then, but the effectiveness of the isolated points rest on the quality of their selection. When occlusion is present in the image, weak segmentation is the best one can hope for. Weak segmentation is used in many retrival systems, either as purpose of its own or as a preprocessing stage for data-driven model-based object segmentation.

Tiêu chí được thõa mãn nếu vùng nằm trong biên của object , nhưng không có gì đảm bảo rằng vùng đó bao phủ toàn bộ đối tượng. Khi một ảnh chứa hai đối tượng gần giống nhau và sát cạnh nhau, thuật toán weak segmentation có thể quan sát nhầm thành một. Tuy nhiên, trong tra cứu ảnh dựa vào nội dung, kiểu lỗi này hiếm khi cản trở kết quả đạt được. Trong tài liệu [125], tiêu chí đồng nhất được cài đặt dựa trên yêu cầu về màu sắc là các vector không gian liên tục trong vùng. Màu sắc là tiêu chí trong tài liệu [50], [126]. Trong tài liệu [20], [112], tiêu chí đồng nhất dựa trên màu sắc và kết cấu. Trường hợp hạn chế của weak segmentation là tập các điểm cô lập, trong tài liệu [148], [57]. Tiêu chí đồng nhất không còn cần thiết sau đó, nhưng tính hiệu quả của các điểm cô lập lại nằm trên chất lượng sự lựa chọn của chúng. Khi xảy ra occlusion (bị che, đối tượng này che đối tượng khác) trong hình ảnh, weak segmentation là thứ tốt nhất có thể nghĩ đến. Weak segmentation được sử dụng trong một vài hệ thống tra cứu, với mục đích riêng hoặc trong giai đoạn tiền xử lí cho việc tách vật thể dựa trên model hướng dữ liệu (data-driven model-based object segmentation).

* When the object has a (nearly) fixed shape, like a traffic light or an eye, we call it a sign: "Localizing signs is finding an object with a fixed shape and semantic meaning, with ."

Khi vật thể (gần như) có hình dạng cố định, chẳng hạn đèn giao thông hoặc con mắt. Những thứ như vậy được gọi là sign (dấu hiệu): "Cục bộ hóa / khoang vùng / định vị signs là tìm đối tượng với hình dạng cố định và có ý nghĩa về mặt ngữ nghĩa, ."

Signs are hepful in content-based retrieval as they deliver an immediate and unique semantic interpretation.

Signs giúp ích trong tra cứu ảnh dựa vào nội dung, khi chuyển tải diễn giải ngữ nghĩa ngay lập tức và duy nhất.

* The weakest form of grouping is partitioning: "A partitioning is a division of the data array regardless of the data, symbolized by: ."

Dạng yếu nhất của gom nhóm là phân vùng: "Một phân vùng là sự phân chia của mảng dữ liệu chứa dữ liệu bất kì. Kí hiệu: ."

The area may be the entire image or a conventional partitioning as the central part of the image against the upper, right, left, and lower parts [67]. The feasibility of fixed partitioning comes from the fact that images are created in accordance with certain canons or normative rules, such as placing the horizon about 2/3 up in the pictures or keeping the main subject in the central area. This rule is often violated, but this violation in itself has semnatic significance. Another possibility of partitioning is to divide the image in tiles of equal size and summarize the dominant feature values in each tile [130].

Vùng có thể là toàn bộ ảnh hoặc phân vùng quy ước là phần trung tâm của ảnh đối với phần trên, phải, trái và dưới [67]. Tính khả thi của phân vùng cố định đến từ thưc tế là hình ảnh được tạo tùy theo luật / quy tắc cụ thể, chẳng hạn đặt đường chân trời lên 2/3 trong hình hoặc giữ cho chủ đề chính (đối tượng chính) nằm trong vùng trung tâm. Luật này thường hay bị vi phạm, nhưng sự vi phạm mang trong nó ngữ nghĩa quan trọng. Một khả năng khác của phân vùng là chia ảnh thành từng ổ có kích thước như nhau, và tóm tắt các giá trị đặc trưng chiếm ưu thế trong mỗi ô [130].

Each of these four approaches to partioning leads to a preferred type of features, as summarized in Fig. 8 and illustrated in Fig. 9, where feature hierarchies are used to make a combination on all types.

Mỗi cách trong 4 cách tiếp cận phân vùng dẫn đến một kiểu được ưu thích của đặc trưng, như đã tóm tắt trong Fig. 8 và được mô tả trong Fig. 9, nơi mà hệ thống thứ bậc đặc trưng được sử dụng để tạo một kết hợp của tất cả các kiểu.

**4.2. Global and Accumulating Features**

*(Đặc trưng toàn cục và tăng cường)*

In the computational process, features are calculated next. The general class of accumulating features aggregate the spatial information of a partitioning irrespective of the image data. A special type of accumulative features are the global features which are calculated from the enitire image. Accumulating features are symbolized by:

where represent an aggregations operation (the sum in this case, but it may be an more complex operator). is the set of accumulative features or a set of accumulative features ranked in a histogram. is part of feature space . is the partitioning over which the value of is computed, see Fig. 9 for an illustration. In the case of global features, represents the image, otherwise, represents a fixed tiling of the image. The operator may hold relative weights, for example, to compute transform coefficients.

Trong quy trình tính toán, đặc trưng được tính toán tiếp theo. Phân lớp chung của các đặc trưng chồng chất, tổng hợp thông tin không gian của phân vùng không phân biệt thuộc dữ liệu hình ảnh. Một kiểu đặc biệt của đặc trưng chồng chất là đặc trưng toàn cục, được tính toán từ toàn bộ hình ảnh. Đặc trưng chồng chất được kí hiệu bởi:

với biểu diễn cho thao tác tổng hợp (trong trường hợp này là phép tính tổng, nhưng có thể là phép toán phức tạp hơn). là tập các đặc trưng tích lũy hoặc tập các đặc trưng tích lũy xếp hạng trong một biểu đồ tần suất (histogram). là một phần của không gian đặc trưng . là phân vùng mà trên đó giá trị của được tính, xem Fig. 9 để minh họa. Trong trường hợp của đặc trưng toàn cục, biểu diễn hình ảnh, còn không, biểu diễn cho một ô cố định của ảnh. Toán tử có thể giữ các trọng số tương đối, chẳng hạn, để tính toán hệ số biến đổi.

|  |
| --- |
| Feature types  _ct•r  Global feature  Modal color in upper 1'  Salient feature  {Connections, occlusi01  Object feature  Region {Polygonal shal  Hierarchical feature st  Patch {uniform, o  encloses  Hole  Closed poly-line  {white & orange color  encloses |
| Fig. 9. Minh họa cho các kiểu đặc trưng khác nhau, được đề cập trong paper. |

A simple but very effective approach to accumulating features is to use the histogram, that is, the set of features ordered by histogram index . The original idea to use histograms for retrieval comes fron Swain and Ballard [169], who realized that the power to identify an object using color is much larger than that of a gray-valued image. As a histogram loses all information about the location of an object in the image, [169], [41] project the histogram back into the image to locate it by searching for best matches. A histogram may be effective for retrieval as long as there is a uniquenss in the color pattern held against the pattern in the rest of the entire data set. In addition, the histogram shows an obvious robustness to translation of the object and rotation about the viewing axis. Swain and Ballard also argue that color histograms change slowly with change in viewpoint and scale and with occlusion.

Một cách tiếp cận đơn giản để tích lũy đặc trưng là sử dụng histogram, nghĩa là, tập đặc trưng sắp xếp bởi histogram chỉ mục . Ý tưởng ban đầu sử dụng histogram cho tra cứu bắt đầu từ Swain và Ballard [169], hai người này nhận ra sức mạnh để xác định đối tượng sử dụng màu sắc lớn hơn nhiều so với ảnh xám. Trong histogram không còn chứa thông tin về vị trí của đối tượng trong ảnh, [169], [41] chiếu histogram trở lại ảnh để định vị đối tượng bằng cách tìm kiếm để tìm vị trí khớp nhất. Histogram có thể hiệu quả cho truy vấn chừng nào có sự khác biệt lớn trong mẫu chống lại mẫu trong phần còn lại của tập dữ liệu. Thêm nữa, histogram thể hiện rõ ràng sự mạnh mẽ để tịnh tiến đối tượng và xoay quanh trục xem (viewing axis).

All of this is in favor of the use of histograms. When very large data sets are at stake, plain histogram comparison will saturate the discrimination. For a 64-bin histogram, experiments show that, for reasonable conditions, the discriminating power among images is limited to 25,000 images [167]. To keep up performance, in [125], a joint histogram is used, providing discrimination among 2250,000 images in their database, rendering 80 percent recall among the best 10 for two shots from the same scene using simple features. Other joint histograms add local texture or local shape [61], directed edges [78], and local higher order structures [48].

Tất cả điều này ủng hộ việc sử dụng histogram. Khi tập dữ liệu rất lớn bị đe dọa, việc so sánh histogram thuần túy sẽ bão hòa sự phân biệt. Với histogram 64-bin (giỏ), các thí nghiệm chỉ ra rằng, với các điều kiện hợp lí, sức mạnh phân biệt giữa các ảnh bị giới hạn ở 25.000 ảnh [167]. Để giữ hiệu suất, trong [125], histogram kết hợp (joint hisrogram) được sử dụng, cung cấp sự phân biệt giửa 250.000 ảnh trong cơ sở dữ liệu của họ, kết xuất 80% recall (độ đo recall) giữa the best 10 for two shots từ cùng cảnh sử dụng đặc trưng đơn giản. Các histogram kết khác thêm kết cấu cục bộ hoặc hình dạng cục bộ [61], các cạnh trực tiếp [78], và các cấu trúc cục bộ cao bậc hơn [48].

Another alternative is to add a dimension representing the local distance. This is the correlogram [73], defined as a three-dimensional histogram where the colors of any pair are along the first and second dimension and the spatial distance betwen them along the thid. The autocorrelogram defining the distances betwen pixels of identical colors is found on the diagonal of the correlogram. A more general version is the geometric histogram [134], with the normal histogram, the correlogram, and several alternatives as special cases. This also includes the histogram of the triangular pixel values reported to outperform all of the above as it contains more information. To avoid an explosion of dimensions of the histogram, one could also perfer to reconsider the quality of the information along each of the dimensions. In Section 3, we have considered invariant representations suited to enrich the information on the axes of the histogram as it rules out the accidental influence of sensing and scene conditions.

Một cách thay thế là thêm một chiều để biểu diễn khoảng cách cục bộ. Đó là correlogram [73], định nghĩa như là một histogram 3 chiều, với màu sắc của mỗi cặp trải theo chiều thứ nhất và thứ hai, và khoảng cách không gian giữa chúng trải theo chiều thứ ba. Correlogram tự động (Autocorrelogram) định nghĩa khoảng cách giữa các pixel của các các màu giống nhau được tìm thấy trên đường chéo của correlogram. Một phiên bản tổng quát hơn là histogram hình học (geometric histogram) [134], với histogram thông thường, correlogram, và một vài thay thế trong trường hợp đặc biệt. Điều này cũng bao gồm histogram của triangular pixel values reported để làm tốt hơn tất cả vì mang nhiều thông hơn. Để tránh sự bùng nổ số chiều của histogram, người ta có thể sẽ thích xem xét lại chất lượng của thông tin trải trên mỗi chiều. Trong phần 3 (section 3), chúng ta đã xem xét các biểu diễn không thay đổi phù hợp để làm giàu thông tin trên các trục của histogram, loại trừ ảnh hưởng ngẫu nhiên của cảm nhận và các điều kiện cảnh.

A different view on accumulative features is to demand that all information (or all relevant information) in the image is preserved in the feature values. When the bitcontent of the features is less than the original image, this boils down to compression transforms. Many compression transforms are known, but the quest is for transforms simultaneously suited as retrieval features. As proper querying for similarity is based on a suitable distance function between images, the transform has to be applied on a metric space. The components of the transform have to correspond to semantically meaningful characteristics of the image. Finally, the transform should admit indexing in compressed form yielding a big computaional advantage over having the image be untransformed first. Schneier and Abdel-Mottaleb [149] is just one of many where the cosine-based JPEG-coding scheme is used for image retrieval. The JPEG-transform fulfills the first and third requirement, but fails on a lack of semantics. In the MPEG-standard, the possibility of including semantic descriptors in the compression transform is introduced [29]. For an overview of feature indexes in the compressed domain, see [103]. In [92], a wavelet packet was applied to texture images and, for each packet, entropy and energy measures were determined and collected in a feature vector. In [75], vector quantization was applied in the sapce of coefficients to reduce its dimensionality. This approach was extended to incorporate the metric of the color space in [146]. In [77], a wavelet transform was applied independently to the three channels of a color image and only the sign of the most significant coefficients is retained. In a recent paper [3], a scheme is offered for a broad spectrum of invariant descriptors suitable for application on Fourier, wavelets, and splines and for geometry and color alike.

Một góc nhìn khác về đặc trưng tích lũy là nhu cầu về tất thông tin (hoặc tất cả thông tin thích hợp) trong ảnh được giữ trong các giá trị đặc trưng. Khi nội dung nhỏ (bitcontent) của đặc trưng nhỏ hơn ảnh gốc, this boils down to biến đổi nén. Một vài biến đổi nén được biết, nhưng nhiệm vụ là cho biến đổi đồng thời phù hợp như truy vấn đặc trưng. Như truy vấn phù hợp cho tính tương tự là dựa trên một hàm khoảng cách thích hợp giữa các ảnh, biến đổi phải áp dụng trên không gian metric. Các thành phần của biến đổi phải tương ứng với các tính chất có ý nghĩa về ngữ nghĩa của ảnh. Cuối cùng, biến đổi nên cho phép lập chỉ mục (indexing) trong dạng nén cho ra lợi thế tính toán lớn so với có ảnh không được chuyển đổi từ đầu. Schneier và Abdel-Mottaleb [149] chỉ là một trong số những tài liệu mà biến đổi cosine của nén JPEG được sử dụng cho truy xuất hình ảnh. Biến đổi JPEG đáp ứng yêu cầu thứ nhất và thứ ba, nhưng thất bại ở việc thiết mất ngữ nghĩa. Trong chuẩn MPEG, khả năng bao gồm mô tả ngữ nghĩa trong biến đổi nén được giới thiệu [29]. Để có tổng quan về chỉ mục đặc trưng trong lĩnh vực nén, xem [103]. Trong [92], một wavelet packet được áp dụng vào kết cấu của ảnh và, với mỗi packet, entropy và độ đo năng lượng được xác định và cho vào vector đặc trưng. Trong [75], lượng hóa vector được áp dụng vào không gian hệ số để giảm chiều. Cách tiếp cận này được mở rộng để kết hợp metric của không gian màu trong [146]. Trong [77], một biến đổi wavelet được áp dụng một cách độc lập vào ba kênh màu của ảnh và chỉ mỗi dấu của hệ số quan trọng nhất được giữ lại. Trong một bài báo gần đây [3], một lược đồ được yêu cầu cho phổ rộng của các mô tả không liên quan phù hợp cho các ứng dụng trên Fourier, wavelets, và splines và cho hình học và cho màu sắc như nhau.

Another type of complete feature sets capturing all information in the image is to use moments. Their invariant combination of moments [72] and [89] have been successfully employed in retrieval of objects in [48], especially when the image contains just the object.

Một kiểu khác của tập đặc trưng toàn diện bắt lấy tất cả thông tin trong ảnh là sử dụng moment. Kết hợp bất biến của các moments [72] và 89] đã sử dụng thành công trong truy vấn của vật thể trong [48], đặc biệt khi ảnh chỉ chứa mỗi đối tượng.

**4.3. Salient Features**

*(Đặc trưng quan trọng | Đặc trưng nổi bật)*

Another way to avoid the brittleness of strong segmentation is to opt for weak segmentation. This leads to a grouping of the data into homogenous regions. From the merged regions, a selection is made on their salience. The most conspicuous regions are stored. The limit case of a weak segmentation is the dectection of conspicuous point, see Fig. 9. Salient features may be covered by the generic equation:

where stands for a local selection operation and operator maximizes the saliency of the processed image field . The area over which the value of is searched for is usually the whole image, although there would be no objection to concentrating on the center or top part of the image in search for specific events.

Một cách khác để tránh giòn (brittleness) cho khoanh vùng mạnh là lựa chọn khoanh vùng yếu. Điều này dẫn đến một nhóm dữ liệu ở trong một khu vực đồng nhất. Từ các khu vực được sáp nhập, sự lựa chọn được tạo trên tính nổi bật của chúng. Những khu vực dễ thấy nhất sẽ được lưu trữ. Trường hợp hạn chế của khoang vùng yếu là sự phát hiện những điểm nổi bật, xem Fig. 9. Đặc trưng nổi bật có thể được bao quát bởi phương trình chung sau:

với là toán tử lựa chọn cục bộ và toán tử cực đại sliency của trường ảnh đã xử lí . Khu vực trên giá trị của được tìm kiếm thông thường là toàn bộ ảnh, mặc dù sẽ không có objection để tập trung tại trung tâm hoặc phần trên của ảnh trong tìm kiếm sự kiện riêng.

As the information of the image is condensed into just a limited number of feature values, the information should be selected with precision for greatest saliency and proven robustness. That is why saliency in [100] is defined as the special points which survive longest when gradually blurring the image in scale space. Also, in [138], lifetime is an important selection criterion for salient points in addition to wiggliness, spatial width, and phase congruency. To enhance the quality of salient descriptions, in [178], invariant and slient features of local patches have been considered. In each case, the image is summarized in a list of conspicuous points. In [148], salient and invariant transitions in gray value images are recorded. Similarly, in [57], [55], photometric invariance is the leading principle in summarizing the image in salient transition in the image. Salient feature calculations lead to sets of regions or points with known location and feature values capturing their salience.

Thông tin của ảnh được cô đọng lại thành số lượng có giới hạn các giá trị đặc trưng, thông tin nên được lựa chọn với độ chính xác để có greatest saliency và chứng minh tính vững chắc. Điều đó giải thích tại sao sliency trong [100] được định nghĩa là những điểm đặc biệt, mà tồn tại lâu nhất khi dần dần làm nhòe ảnh (blurring) trong không gian tỉ lệ (scale space). Ngoài ra, trong [138], vòng đời là một tiêu chí lựa chọn quan trọng cho các điểm nổi bật ngoài wiggliness (lắc), chiều dài không gian, pha đồng dư. Để tăng cường chất lượng của mô tả nổi bật, trong [178], đặc trưng bất biến và nổi bật của của vệt cục bộ đã được xem xét. Trong mỗi trường hợp, ảnh là tổng hợp trong một danh sách của những điểm dễ thấy. Trong [148], chuyển tiếp nổi bật và bất biến trong ảnh xám được ghi lại. Tương tự, trong [57], [55], trắc quan bất biến là nguyên tắc đầu để tóm tắt ảnh trong chuyển đổi nổi bật của ảnh. Đặc trưng nổi bật dẫn tới các tập của khu vực hoặc điểm biết vị trí và giá trị đặc trưng nắm bắt salience của chúng.

In [20], first, the most conspicuous homogenous regions in the image are derived and mapped into feature space. Then, expectation-maximization [37] is used to determine the parameters of a mixture of Gaussians to model the distribution of points into the feature space. The means and covariance matrices of these Gaussians, projected on the image plane, are represented as ellipsoids, and direction. The average values of the color and texture descriptions inside the ellipse are also stored.

Trong [20], đầu tiên, những khu vực đồng nhất dễ thấy nhất trong ảnh được lấy đạo hàm và ánh xạ vào không gian đặc trưng. Sau đó, expectation-maximization (cực đại ước lượng trung bình) được sử dụng để phát hiện các tham số của các Gaussian thành mô hình phân phối của các điểm vào không gian đặc trưng. Ma trận trung bình và hiệp phương sai của các Gaussians, chiếu trên mặt phẳng ảnh, được biểu diễn dạng eellip, và hướng. Các giá trị trung bình của mô tả màu sắc và kết cấu bên trong ellipse cũng được lưu trữ.

**4.4. Signs**

*(dấu hiệu)*

When one of the possible interpretations of an image is so preponderant that it can be considered the meaning of the image, the image holds a sign, characterized by the probability on interpretation :

with symbols as in (5). The anaysis leads to a localization of a sign with its probability. Typical signs are an icon, a character, a traffic light, or a trademark. In the case of maps, the interpretation of map symbols and their spatial relationships provides access to the content of the map [144]. Other systems based on signes are designed with specific application domains in mind, like OCR from an image [200], faces to detect from the image [197], medical image [90], [17], textile [95], art [65], or detecting the constituent components of silhouettes of plants based on a visual lexicon in [180].

Khi một trong các nội suy có thể của một ảnh là nặng hơn (có ưu thế hơn), điều có có thể được coi là ý nghĩa của ảnh, ảnh giữ một dấu hiệu, đặc trưng bởi xác suất trên nội suy :

với kí hiệu như trong (5). Quá trình phân tích dẫn tới cục bộ hóa của một dấu hiệu và xác suất của nó. Thông thường, các dấu hiệu là một biểu tượng, một kí tự, một cột đèn tín hiệu, hoặc nhãn hiệu. Trong trường hợp của bản đồ, nội suy của kĩ hiệu bản đồ và quan hệ không gian cung cấp tiếp cận đến nội trung của bản đồ [144]. Các hệ thống dựa trên kí hiệu được thiết kế với những lĩnh vực ứng dụng cụ thể trong suy nghĩ, giống như OCR từ một ảnh [200], khuông mặt để phát hiện từ ảnh [197], ảnh y khoa [90]. [17], dệt may [95], hội họa [65], hoặc phát hiện các thành phần cấu thành của hình ngược sáng của thực vật dựa trên từ điển trực quan trong [180].

For signs, a strong semantic interpreation is within graps and the undisputed semantic interpretation brings clarity in interpreting the image. That is attractiveness of using signs, in spite of the fact that the analysis tends to become application-oriented.

Với các dấu hiệu, một nội suy ngữ nghĩa mạnh là với sự hiểu biết và nội suy ngữ nghĩa không thể tranh cãi mang lại sự rõ ràng trong giải thích ảnh. Đó là sức hấp dẫn của việc sử dụng dấu hiệu, mặc dù trong thực thế, quá trình phân tích có xu hướng trở thành hướng ứng dụng.

**4.5. Shape and Object Features**

*(Đặc trưng hình dạng và vật thể)*

The theoretically best way to enhance object-specific information contained in images is by segmenting the object in the image. But, as discussed above, the brittleness of segmentation in broad domains. In fact, in many cases, it is not necessary to know exactly where an object is in the image as long as one can identify the presence of the object by its unique characteristics. When the domain is narrow, a tailored segmentation algorithm may be needed more and, fortunately, also be better feasible. When segmentation is applied, we have:

where is the data field resulting from the processing above (equal to the image when is the identity operator), is the segmentation operator for object , and indicates the object are . For shape, is a (possibly ordered) set of features from for :

where represents an aggregation operation and is the functional computing shape in this case. Object internal features are computed similar to (4).

Về mặt lý thuyết, cách tốt nhất để nâng cao thông tin đối tượng cụ thể được chứa trong ảnh, bằng cách khoang vùng đối tượng trong ảnh. Nhưng, như đã thảo luận ở trên, tính dễ gãy của khoanh vùng trong miền (domain) rộng. Thực tế, trong nhiều trường hợp, không cần thiết phải biết chính xác vật thể ở đâu trong ảnh chừng nào người ta có thể xác định sự hiện diện của vật thể bằng những tính chất riêng biệt của nó. Khi miền hẹp hơn, một thuật toán khoang vùng vừa vặn có thể cần thiết nhiều hơn và, may mắn, cũng khả thi hơn. Khi khoanh vùng được áp dụng, chúng ta có:

trong đó là trường dữ liệu kết quả từ quá trình xử lí trên trên (tương đương với ảnh khi toán tử xác định), là toán tử khoanh vùng cho vật thể , và chỉ thị vật thể là . Với hình dạng, là một tập (có thể có thứ tự) của các đặc trưng từ với :

trong đó biểu diễn phép tổng hợp và là chức năng tính toán hình dạng trong trường hợp này. Các đặc trưng bên trong vật thể được tính toán tương tự (4).

The object internal features are largely identical to the accumulative features, now computed over the object area. They need no further discussion here.

Đặc trưng bên trong đối tượng phần lớn giống nhau để trở thành đặc trưng tăng cường, bây giờ được tính toán trên khu vực vật thể. Chúng không cần phải thảo luận thêm nữa.

An abundant comparison of shape for retrieval can be found in [109], evaluating many features on a 500-element trademark data set. Straightforward features of general applicability include Fourier features and moment invariants of the object this time, sets of consecutive boundary segments, or encoding of contour shapes [43].

Một so sánh phong phú về hình dạng để tra cứu có thể được tìm thấy trong [109], đánh giá nhiều đặc trưng trên tập dữ liệu nhãn hiệu 500-thành phần. Đặc trưng chân thật của khả năng ứng dụng chung bao gồm đặc trưng Fourier và moment bất biến của vật thể, tập các đoạn đường biên liên tục, hoặc encoding của hình dạng contour [43].

For retrieval, we need a shape representation that allows a robust measurement of distances in the presence of considerable deformations. Many sophisticated models widely used in computer vision often prove too brittle for image retrieval. On the other hand, the (interactive) use of retrieval makes some mismatch acceptable and, therefore, precision can be traded for robustness and computational efficiency.

Để tra cứu, cần một biểu diễn hình dạng cho phép đo đạc vững khoảng cách khi có sự hiện diện của những biến dạng đáng kể. Nhiều mô hình tinh vi được dùng rộng rãi trong thị giác máy tính thường chứng minh quá dễ đổ vỡ cho tra cứu hình ảnh. Mặc khác, sử dụng (tương tác) tra cứu khiến một số sai lệch có thể chấp nhận và, vì vậy, độ chính xác có thể được đánh đổi cho tính vững chắc và tính toán hiệu quả.

More sophisticated methods include elastic matching and multiresolution representation of shapes. In alastic deformation of image portions [36], [122] or modal matching techniques [150], image patches are deformed to minimize a cost functional that depends on a weighed sum of the mismatch of the two patches and on the deformation energy. The complexity of the optimization problem depends on the number of points on the contour. Hence, the optimizations is computationally expensive and this, in spite of the greater precision of these methods, has limited their diffusion in image databases.

Nhiều phương pháp tinh vi khác, bao gồm elastic matching (so khớp đàn hồi) và biểu diễn đa-phân-giải (multiresolution representation) của hình dạng. Trong biến dạng đàn hồi của những phần ảnh [36]. [122] hoặc kĩ thuật mô hình so khớp [150], những mảnh ảnh được biến dạng để cực tiểu chức năng tính chi phí mà ở đó dựa trên tổng có tính trọng số của hai mảnh không khớp và trên năng lượng biến dạng. Độ phức tạp của vấn đề tối ưu phụ thuộc vào số lượng điểm trên đường đồng mức (contour). Vì vậy, tối ưu là chức năng tính toán nặng và điều này, mặc dù tốt hơn trong tính chính xác của những phương pháp đó, nhưng có giới hạn sự khuếch tán của chúng trong cơ sở dữ liệu hình ảnh.

Multiscale models of contours have been studied as a representation for image databases in [116]. Contours were extracted from images and progressively smoothed by dividing them into regions of constant sign of the second derivative and progressively reducing the number of such regions. At the final step, every contour is reduced to an ellipsoid which could be characterized by some of the features in [48]. A different view on multiresolution shape is offered in [94], where the contour is sampled by a polygon and then simplified by removing points from the contour until a polygon survives selecting them on perceptual grounds. When computational efficiency is at stake, an approach for the description of the object boundaries is found in [201], where an ordered set of critical points on the boundary are found from curvature extremes. Such sets of selected and ordered contour points are stored in [108] relative to the basis spanned by an arbitrary pair of the points. All point pairs are used as a basis to make the redundant representation geometrically to make the redundant representation geometrically invariant, a technique similar to [192] for unordered points sets.

Mô hình multiscale (đa tỉ lệ) của các đường viền được nghiên cứu như là một biểu diễn cho cơ sở dữ liệu hình ảnh trong [116]. Đường viền được tách khỏi ảnh và dần dần được làm trơn bằng cách chia chúng vào các khu vực không đổi dấu của đạo hàm bậc hai và dần dần giảm số lượng của những khu vực trên. Tại bước cuối, mỗi đường viền được giảm thành mặt ellipse mà có thể được đặc trưng bởi một số đặc trưng trong [48]. Một cách nhìn khác trên hình dạng đa-phân-giải được cung cấp trong [94], với đường viên được lấy mẫu bằng đa giác và sau đó đơn giản hóa bằng cách loại bỏ những điểm từ đường viền cho đến khi đa giác vẫn tồn tại chọn chúng on perceptual grounds. Khi tính toán hiệu quả bị đe dọa, một cách tiếp cận cho mô tả đường biên của vật thể được tìm thấy trong [201], với tập có thứ tự các điểm quan trọng trên đường biên được tìm thấy từ cực trị của đường cong. Những tập như vậy từ những điểm trên đường viền được chọ và sắp xếp được chứa trong [108] liên quan đến kéo dài cơ sở (basis spanned) bởi một cặp bất kì của các điểm. Tất cả các cặp điểm được sử dụng như là cơ sở để tạo biểu diễn dự phòng hình học để tạo biểu diễn dự phòng hình học bất biến, một kĩ thuật tương tự với [192] cho tập điểm không thứ tự.

For retrieval of objects in 2D-images of the 3D-worlds, a viewpoint invariant description of the contour is important. A good review of global shape invariants is given in [140].

Để tra cứu vật thể trong ảnh hai chiều trong thế giới ba chiều, một mô tả điểm nhìn bất biến của đường viên là quan trọng. Một xem xét tốt của hình dạng toàn cục bất biến hình dạng toàn cục được đưa ra trong [140].

**4.6. Description of Structure and Lay-Out**

*(Mô tả của cấu trúc và bố trí)*

When feature calculations are available for different entities in the image, they may be stored with a relationship between them, see Fig. 9 for an illustration. Such as structural feature set may contain feature values plus spatial relationships, a hierarchically ordered set of feature values, or relationships between point sets or object sets. The process is symbolized by:

where indicates the th part of the th object and is an (ordered) spatial relationship describing object in elements. Structural and layout feature descriptions are captured in a graph, hierarchy, or any other ordered set of feature values and their relationships.

Khi tính toán đặc trưng có sẵn cho các thực thể khác nhau trong ảnh, chúng có thể được chứa với quan hệ giữa chúng, xem mô tả Fig. 9. Tập đặc trưng cấu trúc như vậy có thể chứa giá trị đặc trưng công với quan hệ không gian, một tập thứ tự phân cấp của giá trị đặc trưng, hoặc quan hệ giữa các tập điểm hoặc tập vật thể. Quá trình có thể kí hiệu như sau:

trong đó biểu thị cho phần thứ của vật thể thứ và là quan hệ không quan (có thứ tự) mô tả vật thể trong thành phần. Mô tả của đặc trưng cấu trúc và bố trí được đem vào đồ thị, cấu trúc phân cấp, hoặc bất kì tập hợp có thứ tự của giá trị đặc trưng và quan hệ giữa chúng.

To that end, in [107], [50], lay-out descriptions of an object are discussed in the form of a graph of relations between blobs. A similar lay-out description of an image in terms of graph representing the spatial relations between the objects of interest was used in [129] for the description of medical images. In [53], a graph is formed of topological relationships of homogenous RGB-regions. When selected features and the topological relationships are viewpoint invariant, the description is viewpoint invariant, but the selection of RGB-representation as used in the paper will only suit that purpose to a limited degree. The systems in [70], [163] study spatial relationships between regions, each characterized by locations, size, and features. In the later system, matching is based on the 2D-string representation founded by Chang and Hau [22]. For a narrow domain, in [129], [132], the relevant elements of a medical X-ray image are characterized separately and joined together in a graphj that encodes their spatial relations.

Đến cuối cùng, trong [107], [50], mô tả vế bố trí của một vật thể được thảo luận trong dạng của đồ thị và quan hệ giữa đốm màu. Mô tả cấu trúc tương tự của một ảnh về mặt biểu diễn đồ thị cho quan hệ không gian giữa các vật thể quan tâm được dùng trong [129] cho mô tả của ảnh y khoa. Trong [53], một đồ thị được tạo từ đồ hình quan hệ của các vùng RGB đồng nhất. Khi đặc trưng được lựa chọn và đồ hình quan hệ là điểm nhìn bất biến, thì mô tả là điểm nhìn bất biến, nhưng lựa chọn của biểu diễn RGB như được dùng trong bài báo chỉ sẽ phù hợp cho mục đích đó tới một mức độ giới hạn. Hệ thống trong [70], [163] nghiên cứu quan hệ không gian giữa các vùng, mỗi vùng có đặc tính về vị trí, kích thước, đặc trưng. Trong hệ thống về sau, so khớp dựa trên biểu diễn chuỗi hai chiều, tìm thấy bởi Chang và Hau [22]. Đối với miền hẹp, trong [129], [132], các thành phần thích hợp của ảnh X-quang y khoa được đặc trưng tách biệt và kết với nhau thành một đồ thị mà trong đó encode quan hệ không gian giữa chúng.

Starting from a shape description, the authors in [94] decompose an object into its main components, making the matching between images of the same object easier. Automatic identification of salient regions in the image, based on nonparametric clustering followed by decompostion of the shapes found into limbs, is explored in [52].

Bắt đầu từ một mô tả hình dạng, các tác giả trong [94] phân tách một vật thể thành các thành phần chính của nó, khiến việc so khớp giữa các ảnh cùng chung vật thể dễ dàng hơn. Tự động nhận biết các khu vật nổi bật trong hình ảnh, dựa trên gom cụm phi tham số theo sau là phân tách của các hình dạng tìm thấy vào trong các nhánh, được khám phá trong [52].

**4.7. Discussion on Features**

*(Thảo luận về đặc trưng)*

Also in the description of the image by features, it should be kept in mind that for retrieval a total understanding of the image is rarely needed. Strong segmentation of the scene and complete feature descriptions may not be necessary at all to achieve a similarity ranking. Of course, the deeper one goes into the semantics of the pictures, the deeper the understanding of the picture will also have to be, but the critical point in the advancement of content-based retrieval is the semantic meaning of the image that rarely self-evident.

Cũng trong mô tả của hình ảnh bằng các đặc trưng, nên nhớ rằng, tra cứu "hiểu biết tổng thể" của ảnh là hiếm khi cần. Khoanh vùng mạng của cảnh và mô tả đặc trưng toàn diện có thể không cần thiết để đạt được thứ hạng tương tự. Tất nhiên, càng đi sâu vào ngữ nghĩa của hình ảnh, sự hiểu biết về hình ảnh càng sâu hơn, nhưng điểm quan trọng trong sự tiến bộ của tra cứu dựa vào nội dung là ý nghĩa ngữ nghĩa của ảnh hiếm khi rõ ràng.

The theoretically best approach to a semantic interpretation of an image remains the use of a strong segmentation of the scene. Automatic strong segmnetation is, however, hard to achieve, if not impossible for general domains. The brittleness of strong segmentation is mostly unsurpassable obstacle when describing the content of images by describing the content of its objects. Especially for broad domains and for sensory conditions where clutter and occlusion are to be expected, automatic strong segmentation is hard, if not impossible. In that case, segmentation is to be done by hand when retrieval relies on it.

Cách tiếp cận tốt nhất theo lí thuyết cho diễn dịch ngữ nghĩa của ảnh vẫn còn việc sử dụng khoanh vùng mạnh của cảnh. Tự động khoanh vùng mạng khó để đạt được, nếu không muốn nói là bất khả thi đối với các miền chung. Tính gãy của khoanh vùng mạnh là trở ngại chủ yếu không thể vượt qua khi mô tả nội dung của ảnh bằng mô tả nội dung của các vật thể của ảnh. Đặc biệt cho miền rộng và điều kiện cảm giác, nơi lộn xộn và tắc sẽ được dự kiến, tự động khoanh vùng mạnh là khó, nếu không phải là không thể. Trong trường hợp này, khoanh vùng được thực hiện bằng tay khi tra cứu dựa trên nó.

Narrow domains such as trademark validation, the identification of textiles, and the recognition of fish depend on the shape of the object, assessing similarity of fish depend on the shape of the object, assessing similarity on the basis of the silhouettes. The fine-to-coarse decompostions are attractive in their discriminating power and computational efficiency. Again, a major bottleneck is the highly accurate segmentation of the object (as well as a frontal viewpoint of the object). In selected narrow domains, this may be achieved by recording the object against a clear background.

Lĩnh vực hẹp, chẳng hạn thẩm định nhãn hiệu, nhận biết vải dệt, nhận diện cá dựa trên hình dạng của đối tượng, đánh giá giống nhau của cá dự trên hình dạng của đối tượng, đánh giá sự tương tự trên nền tảng của hình ngược sáng. Phân tách từ mịn đến thô hấp dẫn với sức mạnh phân biệt và hiệu suất tính toán. Một lần nữa, nút cổ chai lớn là khoanh vùng chính xác cao của vật thể (cũng như là điểm nhìn phía trước của vật thể). Trong những miền hẹp được chọn, điều này có thể đạt được bằng ghi lại vật thể đối với với nền rõ ràng.

General content-based retrieval system have dealt with segmentation brittleness in a few ways. First, a weaker version of segmentation has been introduced in content-based retrieval. In weak segmentation, the result is a homogenous region by some criterion, but not necessarily convering the complete object silhouette. It results in a fuzzy, blobbly description of objects, rather than a precise segmentation. Salient features of the weak segments capture the essential information of the object in a nutshell. The extreme form of the weak segmentation is the selection of a salient point set as the ultimately efficient data reduction in the representation of an object, very much like the focus-of-attention algorithms for an earlier age. Only points on the interior of the object can be used for identifying the object and conspicuous points at the borders of objects have to be ignored. Little work has been done on how to make the selection. Weak segmentation and salient features are a typical innovation of content-based retrieval. It is exptected that salience will receive much attention in the further expansion of the field, especially when computational considerations gain in importance.

Hệ thống tra cứu dựa vào nội dung thông thường có thể xử lí với những khoanh vùng gãy theo một số cách. Thứ nhất, phiên bản yếu hơn của khoanh vùng đã được giới thiệu trong tra cứu dựa vào nội dung. Trong khoanh vùng yếu, kết quả là vùng đồng nhất bởi một vài tiêu chí, nhưng cần thiết phải bao phủ hoàn bộ bóng đen ngược sáng của vật thể. Kết quả là mô tả mờ, "blobby" về vật thể, thay vì một khoanh vùng chính xác. Đặc trưng nổi bật của khoanh vùng yếu chiếm lấy thông tin chủ yếu của vật thể một cách ngắn gọn. Hình thức cực đoạn của khoanh vùng yếu là sự lựa chọn của điểm nổi bật như là sự giảm dữ liệu cuối cùng hiệu quả tring biểu diễn của vật thể, rất giống như thuật toán tập trung của chú ý (focus-of-attention) trong thời kì đầu. Chỉ những điểm thuộc phần trong của vật thể có thể sử dụng để xác định vật thể và những điểm rõ ràng tại biên của vật thể phải bị bỏ qua. Công việc nhỏ đã được thực hiện về cách tạo lựa chọn. Khoanh vùng yếu và đặc trưng nổi bật là đổi mới điển hình của tra cứu dựa vào nội dung. Người ta hi vọng rằng, chỗ nổi bật có thể nhận được nhiều sự chú ý trong những mở rộng về xa hơn của lĩnh vực, đặc biệt khi cân nhắc tính toán đạt được quan trọng.

The alternative is to do no segmentation at all. Content-based retrieval has gained from the use of accumulative features, computed on the global image or partitioning thereof, disregarding the content, the most notable being the histogram. Where most attention has gone to color histograms, histograms of local geometric properties and texture are following. To compensate for the complete loss of spatial information, the geometric histogram was recently defined with an additional dimension for the spatial layout of pixel properties. As it is superset of the histogram, an improved discriminability for large data sets is anticipated. When accumlative features are calculated from the central part of a photograph may be very effective in telling them apart by topic, but the center does not always reveals the purpose. Likewise, features calculated from the top part of a picture may be effective in telling indoor scenes from outdoor scenes, but again this holds to a limited degree. A danger of accumulative features is their inability to discriminate among different entities and semantic groupings will increase the power of accumulative descriptors to capture the content of the image.

Thay thế là không khoanh vùng nữa. Tra cứu dựa vào nội dung đã thu được từ việc dùng các đặc trưng tích lũy, tính toán trên ảnh toàn cục hoặc phân vùng của chúng, không quan tâm nội dung, đáng chú ý nhất là histogram. Khi những gì gây chú ý nhất đều vào trong lược đồ màu (color histograms), histogram của thuộc tính cục hình học cục bộ và kết cấu là sau. Để đền bù mất mát cho mất mát toàn diện của thông tin không gian, lược đồ hình hình học được định nghĩa gần đây với chiều bổ sung cho bố trí không gian của thuộc tính điểm ảnh. Với ý nghĩa là tập cha (superset) của histogram, một cải thiện phân biệt cho tập dữ liệu lớn hơn được dự đoán. Khi đặc trưng tích lũy được tính toán từ phần trung tâm của ảnh chụp có thể sẽ rất hiệu quả trong việc phân biệt chúng theo chủ đề, nhưng trung tâm không phải lúc nào cũng tiết lộ mục đích. Tương tự, đặc trưng được tính toán từ phần trên của hình ảnh có thể hiệu quả trong việc phân biệt cảnh trong nhà và cảnh ngoài trời, nhưng một lần nữa là điều này giữ ở mức độ giới hạn. Điều nguy hiểm của đặc trưng tích lũy là chúng không có khả năng phân biệt giữa các thực thể khác nhau và nhóm ngữ nghĩa sẽ tăng sức mạnh của mô tả tích lũy để bắt lấy nội dung của ảnh.

Structual descriptions match well with weak segmentation, salient regions, and weak semantics. One has to be certain that the structure is within one object and not an acciental combination of patches which have no meaning in the object world. The same brittleness of strong segmentation lurks here. We expect a sharp increase in the research of local, partial, or fuzzy structural descriptors for the purpose of content-based retrieval, especially of broad domains.

Mô tả cấu trúc khớp tốt với khoanh vùng yếu, khu vực nổi bật, và ngữ nghĩa yếu. Người ta phải chắc chắn rằng cấu trúc bên trong một vật thể và không phải là kết hợp tình cờ của các mảnh mà không có ý nghĩa gì trong thế giới vật thể. Tương tự cho tính giòn gãy của khoanh vùng mạnh ẩn nấp ở đây. Chúng ta mong đợi một sự tăng mạnh mẽ trong tìm kiếm cục bộ, từng phần, hoặc mô tả cấu trúc trúc mờ cho mục đích của tra cứu dựa trên nội dung, đặc biệt là miền rộng.