Downloaded from UvA-DARE, the institutional repository of the University of Amsterdam (UvA)
http://dare.uva.nl/document/196834

File ID	196834
Filename	Chapter 3 Features and Content Extraction

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type	Dissertation
Title	Giving Eyes to ICT! Or How does a computer recognize a cow?
Author	B.A.M. Schouten
Faculty	Faculty of Science
Year	2001

FULL BIBLIOGRAPHIC DETAILS:
http://dare.uva.nl/record/96150

Copyright

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other then for strictly personal, individual use.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)

Chapter 3

Features and Content Extraction

Interaction with visual content is essential to visual information retrieval. To describe the visual content in multimedia databases low level aspects of an image like color, together with high level concepts like objects are characterized by *features*. Low level features are directly related to perceptual aspects of image content. Several aspects of color, texture or object-shape can be modeled. In order to retrieve similar images within VIR systems these features are compared to the features of the other images present in the database.

A *feature space* consists of all features of all images within a database. If a feature space is endowed with a (similarity) metric, it will be called a *query space* [28]. In most systems features are represented through a numerical vector (n-tuple of numerical values). The query space is therefore modeled as a suitable n-dimensional feature space. Standard mathematical distances like the *Euclidean distance* or the *Minkovsky distance* are used to measure the distance between two points in this feature space. The *similarity measure* can be adapted by giving weights to the importance of certain features in the image (Figure 3.1).

3.1 Examples of Features

Content of multimedia-objects like video, images, sounds are described by several features, directly related to the content of the multimedia object. Features of images can be subdivided into:

- **Perceptual features.**
  Directly observable features like color distribution, textural properties and shape properties.

- **Semantic features.**
  Features related to the concept pictured in the image. Suppose a car has an appearance in an image, seeing it as an object is a quality of the observer related to the concept.
- **Psychological features.**
  Like emotions, appreciation (nice and ugly) and dislike.

A perceptual feature can be modeled more or less independently of the user but often loses its meaning without context. Semantic and psychological features depend heavily on the user. These features cannot be modeled without any knowledge of the context (domain knowledge) wherein the system is used. We confine ourselves to a short description of the most commonly used features like color and texture. For an extensive overview we refer the reader to Bimbo [3] or Huang et al. [11].

3.1.1 Color

Color is one of the most powerful features to describe an image with. The presence and distribution of colors induce sensations and convey meanings to the observer [3]. In the Bauhaus period artist and designers like Itten developed color schemes from a perceptual point of view (Figure 3.2). In his book *Visual Perception*, Tom Cornsweet [6] has made an extensive study.

Aspects of color can be modeled by several color attributes. Usually color stimuli are represented as points in three-dimensional color spaces (channels). There are several ways of expressing color channels, amongst them RGB Space (Figure 3.3). Each color in RGB is expressed as the combination of the primary colors: Red, Green and Blue. This model is used in most hardware (Televisions, Computers and Screens).

To describe low-level color properties of an image, color histograms are used. A color histogram denotes the joint probability of the intensities of the three color channels. A similarity measure is achieved by histogram matching and/or color moments. An aspect important to all features is the robustness to a change in lighting conditions or variations of the image like rotating the canvas or rescaling. Gevers [9] studied the invariances of color spaces to camera viewpoint, orientation and position of the object as well as changes in the color and intensity of the illumination.

Click A.5

As an example we refer the reader to the Blobworld system of the University of California, Berkeley:

`http://elib.cs.berkeley.edu/photos/blobworld.`

In this example, color queries can be processed by region. It will be clear that combining color information to the spatial relationships within an image (regions, objects) will enhance the results.

3.1.2 Texture

Texture is almost as effective in describing an image as color. [2, 5, 13, 16, 19, 20, 25, 22, 11]. Texture can be analyzed from both a mathematical and a psychological point of view. The advantage of a profound mathematical theory and *human perception* join together in using fractal geometry for feature extraction.

In most texture one finds a structural element which is repeated in the image by a placement rule. From a mathematical point of view one can distinguish *statistical* and *structural* texture.
A primitive (building block) is arranged according to a certain placement rule [26, 35]

$$f = R(e) \tag{3.1}$$

R represents the relation or *placement rule* and e denotes an element. Structural texture is characterized by a precise definition of R and e. Statistical texture by a more macroscopic view, with R and e exhibiting variance.

Tamura amongst others [32] was one of the first to recognize the importance of modeling in accordance with human perception. In psychological experiments by testers, several aspects of both structural and statistical texture were distinguished. She proposed:

Click A.6

contrast , *coarseness* , *directionality* , *line-likeness* , *regularity* and *roughness*

as the main aspects to distinguish the several aspects of texture. These principles have been used in e.g. QBIC [22] and FracFeat, a fractal feature extractor we builded which will be discussed in Section 6.1.

Fractal image compression is concerned with finding similarities between parts of an image and records the spatial relationships between them, see Figure 3.4. These spatial relationships relate to the placement rules of Tamura in a very simple way. Fractal features are well equipped to classify images into natural scenes and human environment (e.g. cities) [29]. Fractal features can be made invariant to a wide range of image variations, like contrast scaling, rotation of the image canvas and even folds. This makes fractal feature extraction applicable within the domain of textiles and fashion [30, 31]. The above aspects are subject of three papers from Chapter 7, 8 and 9. In Chapter 5 we provide a mathematical background on the use of *fractal geometry* for feature extraction.

In the early 1990's when wavelets were introduced, many researchers started to study the wavelet transform for texture representations. Among the different transforms, the Gabor [16] transform shows very good results in modeling both mathematical and perceptual aspects.

3.2 Questions, Questions, and more Questions.

Since there are no *all encompassing truths* in the perception of visual aspects, extracting meaningful features is a challenging research topic.

- **Never enough !**
  A person maybe recognized best by a scar. Faloutsos describes in his 'Gemini' approach a way of looking at the problem of finding the most useful feature:

  Question: If we are allowed to use only one numerical feature to describe each data object, what should this feature be ?[8]

  From this question immediately new questions arise like, which feature to choose, why, and how ?

- **Absolutely arbitrary !**
  Because of subjectivity in perception there does not exist a single best representation for a given feature; there are multiple representations which characterize the given feature from different aspects.

  The amount of features to be extracted from the image is endless. The choice of the features in the system depends heavily on the domain. Moreover, the same image can be retrieved for different reasons over time and the system should be able to come up with the right features and similarity measures every time. Because features are present in the system and provided by the "owner" of the system it is almost impossible to provide the right features. Features are only *partially relevant.*[1]. Another aspect is the invariance of a feature to perturbations of the image. An object can be seen from different angles, images can be rotated, lighting may vary.

- **Never right !**
  Similarity is provided by mathematically defined measures and features. The meaning of the measure to the user are ambiguous. Similarity is subjective and strongly dependent on culture, personal taste and again on different meanings the user can have at different times. In Figure 3.5 we show an example of the images returned by the Virage system [1], as input the hand of a human being was presented.

Some improvements are widely recognized: features should always be an integral part of the database in such a way that the user is able to relate to the system in an intuitive and transparent way. The user should be in the loop, adding knowledge about the similarity and the features wanted, and communicating with the system using *relevance feedback* and learning. Simple similarity metrics are not enough [20]. Click A.7

[1] In content based image retrieval images are returned to the user according to a search process. We distinguish the *partial relevance* of the feature extracted from an image and the *partial relevance* of the returned image to the user: *partial relevance of images*. The latter is subject of Chapter 4, [4]

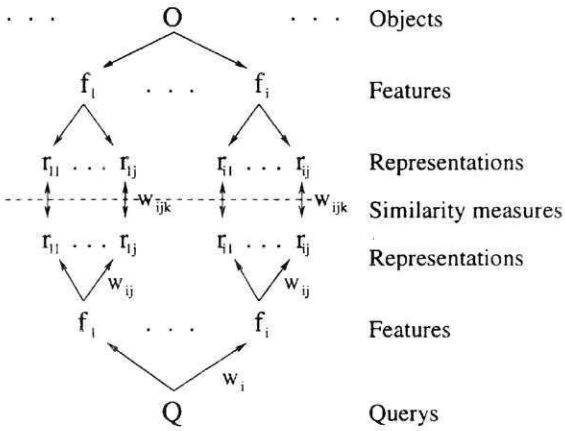

Figure 3.1: Schematic overview [27] of a query process by feature extraction and similarity search. A query image Q is represented by features $f_1, ..., f_i$. Each feature f_i is "composed" of several aspects $r_{i,1}, ..., r_{i,j}$. Weights $w_{i,j}$ are used to represent the importance of the different aspects. A *similarity measure* (dotted line) "compares" the aspects of the query image Q with the aspects of the other images O in the database. Weights $w_{i,j,k}$ can be given to "tune" the similarity measure.

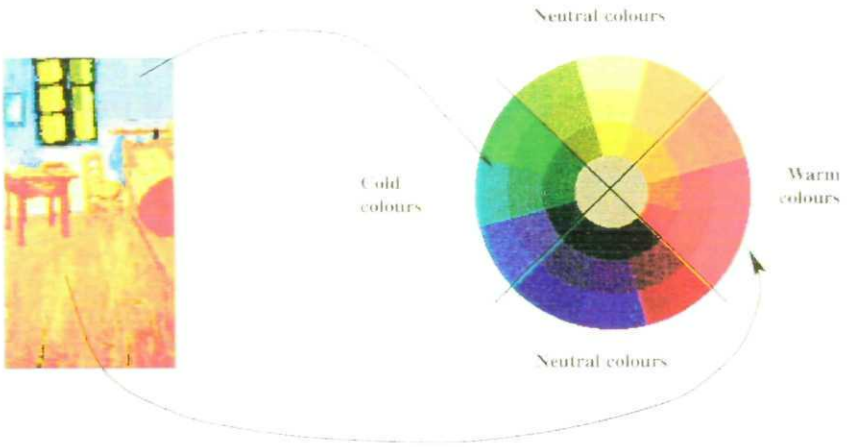

Figure 3.2: Ittens color circle, representing warm and cold colors located in opposite positions.

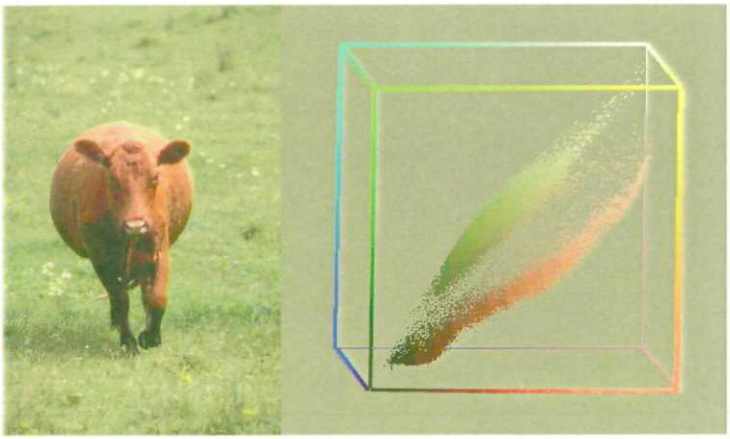

Figure 3.3: RGB space is used to plot color values for each pixel in the image (left). In this plot, the object is clearly seperated from the background (right).

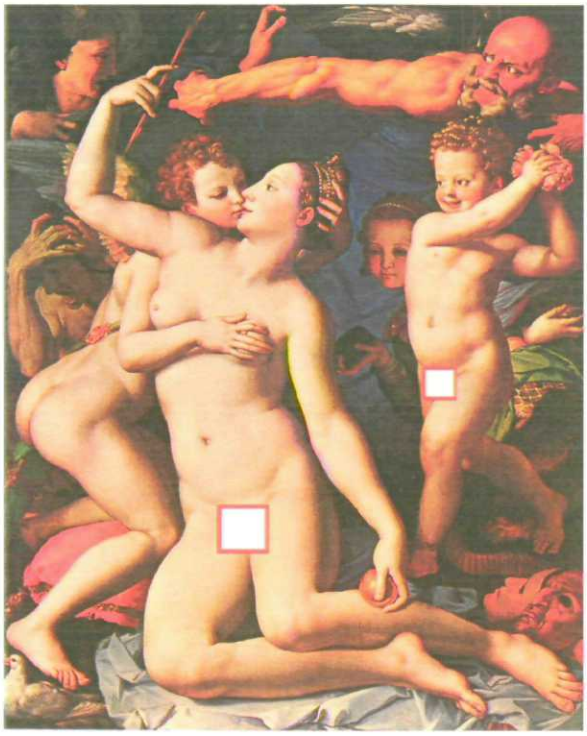

Figure 3.4: A fractal encoder searching for "similar" image blocks. The coder keeps track of the spatial relationships between the similar blocks, used to characterize the image with.

Figure 3.5: *Virage* system [1] at work. *top:* User is presented random images (query by example). Weights can be given to the importance of the features. *bottom:* Retrieved images, including the image of a pig.