Raport Badawczy

Research Report

RB/43/2016

Content-based image retrieval tools and techniques

T. Jaworska

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Prof. dr hab. inż. Janusz Kacprzyk

Warszawa 2016

Tatiana Jaworska

Content-Based Image Retrieval Tools and Techniques

In the beginning was an image.

To my mother
who inspired me
to develop intellectually

Contents

3 Image Representations

3.1 Introduction. Forms of Image Representation

With regard to computers, an image can be represented in different forms. The two most frequently used types use at present are vector and raster images. Vector graphics, generated as geometric shapes by mathematical equations, can be scaled, rotated, moved, or otherwise manipulated to any degree without any loss of quality, and displayed or printed at whatever resolution is available on a monitor or printer. An example of a vector image is shown in Fig. 3.1 a) as a line image and in Fig. 3.1 b) as a colour one. In turn, raster graphics is made up of pixels, each of a different colour, arranged to display an image. A major difference is that raster image pixels do not retain their appearance as size increases, when you blow a photograph up, it becomes blurry for this reason. These two representations can be compared in Fig. 3.1 c) vector zoom and d) raster zoom.

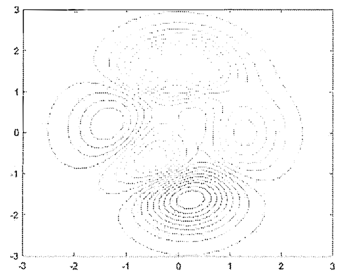
Here, we analyse visual information so, first of all, we have to explain how we understand the notion of image.

Definition 3.1 (image)

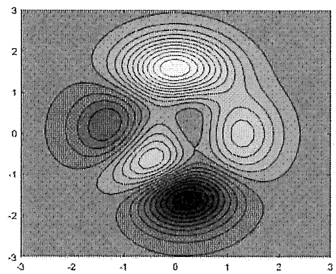
An image is a two dimensional function f: $\mathbf{R}^2 \to \mathbf{R}$ such as $f(x,y)$ describing the intensity at the position (x,y). For digital images I, we have a pixel array I: $[a,b] \times [c,d] \to [0,1]$. Colour images consist of three colour components:

$$I: (x,y) = \begin{bmatrix} r(x,y) \\ g(x,y) \\ b(x,y) \end{bmatrix} \tag{3.1}$$

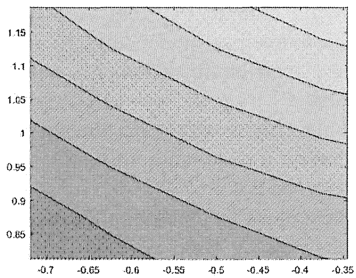
Generally, we analyse and store in the DB raster images, but in some cases we use vector images, for example, as a prompt for the user in the GUI (see Fig. 3.2), where we exploit the easy zooming (details in Chapter 8).



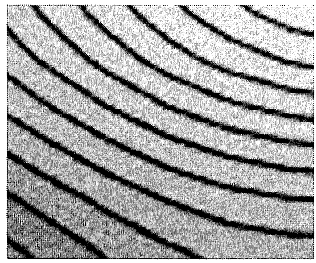
a)



b)



c)



d)

Fig. 3.1 Two most often used image representations: raster and vector; a) vector representation, b) vector with colour filling, c) close-up of the vector representation, d) close-up of the raster representation.

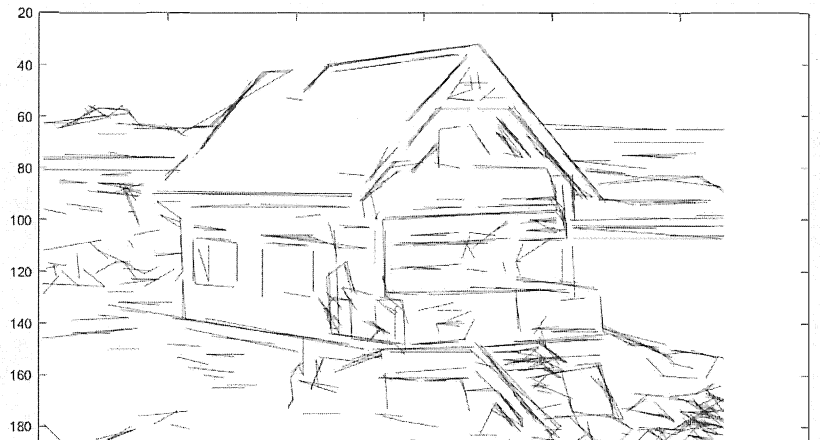


Fig. 3.2 Example of a vector image - used as a prompt in the GUI.

3.2 Visual Feature Descriptors

Feature extraction is a process of selecting a map of the form $X=f(Y)$, by which a sample $\mathbf{y}=[y_1,y_2,\ldots,y_q]$ in a q-dimensional measurement space Ω_Y is transformed into a point $\mathbf{x}=[x_1, x_2, \ldots, x_{q'}]$ in a q'-dimensional feature space Ω_X, where $q'<q$. This task is realized to generate the optimal characteristics necessary for the process of recognition and reduce the dimensionality of space Ω_Y in order to apply effective computable algorithms of classification. We understand the space Ω_Y as a model which we construct based on a subset of selected variables.

The exact definition of feature often depends on the problem or the type of application. Feature detection is low-level image processing because it operates on the level of pixels and is the easiest to extract by the computer system. As a matter of fact, all CBIR systems use the feature detection level, some concentrate on one feature, for instance, colour or shape, others use their combination, but only the more advanced move further to semantic retrieval.

Below we present the most commonly used algorithms which are also useful in search engine construction. We would like to present a wide range of methods as a background for our content-based image retrieval system.

To begin with, we describe our approach to colour segmentation as the first point in constructing our CBIR system which, step by step, will be presented throughout this book.

3.3 Colour Information

Colour is a commonly used feature because its layout in the image is the key information, whereas the simpler systems extract only global features from the colour image. The more advanced ones use colour information about regions or separate segments [34].

Each pixel of the image can be represented as a point in a 3D colour space. Many colour spaces for image retrieval, including *RGB, Munsell, CIE L*a*b*, CIE L*u*v*, HSV* are used depending on the aims and the method of image acquisition.

Colour information gives the opportunity to construct such descriptors as:

- **colour moments** (mean, variance and skewness) [56], [17] help to describe colour distribution in the whole image, which is the basis for many CBIR retrieval processes. Nevertheless, they do not give spatial information about pixels;
- **the colour histogram** is easy to compute and invariant in terms of scaling and rotation, however, it also fails to provide spatial information about pixels, so many images have similar histograms;
- **the colour coherence vector** (CCV) [57] is constructed based on the colour histogram. In this case, each histogram bin, a separate one for each colour, is

partitioned into two parts: coherent if it belongs to a large uniformly-coloured region, and incoherent in the opposite case. It means that two pixels a and a' are coherent if they belong to region C, such that $a, a' \in C$ and there exists a path in C between a and a'. For the image, the CCV is defined as the vector $[(\alpha_1, \beta_1), (\alpha_2, \beta_2), \ldots, (\alpha_N, \beta_N)]$, where α_i denotes the number of coherent pixels of the i^{th} colour bin, whereas β_N denotes the number of coherent pixels. The additional spatial information included in the CCV improves the results of retrieval in comparison to the simple colour histogram.
- **the colour correlogram**, also called a second-order histogram, describes the spatial correlation of pairs of colours. A colour correlogram is a table indexed by colour pairs, where the k^{th} entry for (i,j) specifies the probability of finding a pixel of colour j at a distance r from a pixel of colour i in the image. The colour correlograms for all sets of colours are rather large, therefore, a simplified version is an auto-correlogram which is a spatial relationship only between points of identical colour.

In our approach, we omitted the colour description globally. We only used colour information, specifically RGB colour space, to segment separate objects from an image (in details see sec. 4.2).

3.3 Texture Information

Texture is one of the most important visual cues to identify homogeneous regions [58]. The goal of texture classification is to identify each uniform texture region, whereas the goal of texture segmentation is to obtain the boundary map and further separate regions characterized by different textures.

For our purpose, the key operation is segmentation, in a nutshell, we can assume that image texture is an attribute representing the spatial arrangement of grey or colour levels of the pixels in a region [59]. Hence, the intensity variations in an image, characterizing texture, generally reflect physical variations in the real scene. To model these variations the following issues need to be addressed:

- pixel colour value in a spatial neighbourhood;
- spatial distributions of these values;
- their resolution or scale;
- the unrecognizability of separate primitive objects in a texture region.

Basically, texture representation methods can be classified into four categories: structural, statistical, fractal [60] and transformational [34] as it can be seen in Fig. 3.. The first category of methods can be divided into morphological operators and adjacency graphs presenting texture as structural primitives and their placement rules. The primitive can be as simple as a single pixel that can take a grey value, but it is usually a collection of pixels. The placement rule is defined by a tree grammar. A texture is then viewed as a string in the language defined by the grammar whose terminal symbols are the texture primitives. An advantage of this method is that it can be used for texture generation, as well as texture analysis. The patterns generated by the tree grammars can also be regarded as ideal textures in Zucker's model [61]. They are more effective when we have a regular texture.

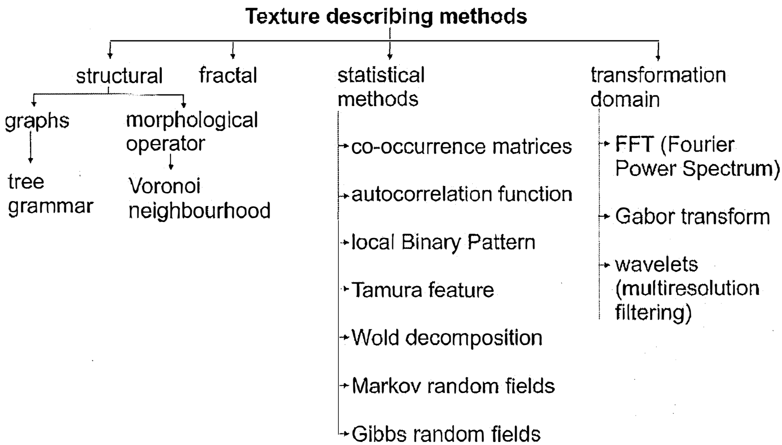


Fig. 3.3 The categories of texture describing methods.

Another example is Voronoi features [62], which were proposed because the local spatial distributions of tokens are reflected in the shapes of Voronoi polygons. Many of the perceptually significant characteristics of a token's environment are manifest in the geometric properties of Voronoi neighbourhoods. In order to apply geometrical methods to grey level images, we only need to first extract tokens from images.

The statistical methods describe texture by statistical distribution of image intensity. There are numerous statistical texture representations:

– co-occurrence matrices. Spatial grey level co-occurrence estimates image properties related to second-order statistics. Haralick [63] suggested the use of the $G \times G$ grey level co-occurrence matrix $P_{\mathbf{d}}$ for a displacement vector $\mathbf{d} = (dx, dy)$, defined as follows: the entry (i,j) of $P_{\mathbf{d}}$ is the number of occurrences of the pair of grey levels i and j which are a distance $\mathbf{d}$ apart. Formally, it is given as:

$$P_{\mathbf{d}}(i,j) = |\{((x,y),(t,v)) : I(x,y) = i,\ I(t,v) = j\}| \tag{3.2}$$

where $(x,y),(t,v) \in N \times N$, $(t,v) = (x+dx,\ y+dy)$, and $|.|$ is the cardinality of a set. Based on this matrix some useful texture features can be described, such as: energy, entropy, contrast, homogeneity or correlation.

– Autocorrelation function – can be used to assess the amount of regularity as well as the fineness – coarseness of the texture in the image. Formally, the autocorrelation function of an image $I(x,y)$ is defined as follows:

$$\sigma(x,y) = \frac{\sum_{u=0}^{N} \sum_{v=0}^{N} I(u,v) I(u+x, v+y)}{\sum_{u=0}^{N} \sum_{v=0}^{N} I^2(u,v)} \tag{3.3}$$

– Local Binary Pattern (LBP) operator [64] is, originally, based on a 3×3 pixel neighbourhood (see the *example* sub-table in Table 3.1). Image pixels in each neighbourhood of a pixel (*i,j*) are exchanged into a binary threshold map where 1 is for a pixel larger than the central pixel and 0 where the values are less than the central one (see the *threshold* sub-table in Table 3.1). The values of the pixels in the threshold map are multiplied by the weights given to the corresponding pixels. The weights are the power of 2 where the number of neighbourhood is the exponent (see the *weights* sub-table in Table 3.1). Finally, the values of the eight weighted pixels are summed to obtain one factor (the *result* can be seen in Table 3.1). The LBP histogram computed over a region is used as a texture description. Because of the LBP design, it is invariant under any monotonic grey scale transformation and provides information about the spatial structure of the local image texture. Due to its 3×3 window operation, however, feature distributions may be sensitive to geometric distortion. This operator was extended later [65] for neighbourhoods of different sizes, for instance, circular neighbourhood and bilinear interpolation of non-integer pixel values.

Table 3.1 Computation of Local Binary Pattern (LBP).

Example			Threshold			Weights			Result		
8	5	2	1	0	0	1	2	4	1	0	0
7	6	1	1	x	0	8	x	16	8	x	0
9	13	7	1	1	1	32	64	128	32	64	128

LBP = 1+8+32+64+128=233

The original LBP operator was defined to only deal with the spatial information. Later, it was extended to a spatiotemporal representation for dynamic texture analysis. For this purpose, the so-called Volume Local Binary Pattern (VLBP) operator was proposed [66].

– Tamura feature; Tamura et al. [67] considered six basic textural features:

- coarseness – relates to distances of notable spatial variations of grey levels, that is, implicitly, to the size of the primitive elements forming the texture.
- contrast – measures how grey levels vary in the image and to what extent their distribution is biased to black or white. The second-order and normalised fourth-order central moments of the grey level histogram are used to define the contrast.
- directionality – measured the distribution of oriented local edges against their directional angles using the Sobel edge detector (for details see sec. 3.4.1).
- line-likeness - is defined as an average coincidence of the edge directions.

- regularity – is defined as the normalised sum of the standard deviations of the corresponding above-mentioned feature.
- roughness – feature is given by simply summing the coarseness and contrast measures.

- Wold decomposition, [68], [59] provides three different components to describe texture: *harmonic, evanescent,* and *non-deterministic,* corresponding to *periodicity, directionality,* and *randomness* introduced by his predecessors. Periodic textures have a strong harmonic component, highly directional textures have a strong evanescent component, and less structured textures tend to have a stronger non-deterministic component. The deterministic periodicity of the image is analysed using the autocorrelation function. The corresponding Wold feature set consists of the frequencies and the magnitudes of harmonic spectral peaks (e.g. the largest peaks). The nondeterministic (random) components of the image are modelled with the multiresolution simultaneous autoregressive (MR-SAR) process. The retrieval uses matching of the harmonic peaks and the distances between the MR-SAR parameters. The similarity measure involves a weighted ordering based on the confidence level in the query pattern regularity.
- Markov random fields [69]. Random field models consider an image as a 2D array of random scalars (grey values) or vectors (colours). In other words, the signal at each pixel location is a random variable. Each type of texture is characterised by a joint probability distribution of signals that accounts for spatial inter-dependence, or interaction among the signals. The interacting pixel pairs are usually called neighbours, and a random field texture model is characterized by the geometric structure and quantitative strength of interactions among the neighbours. If pixel interactions are assumed translation invariant, the interaction structure is given by a set N of characteristic neighbours of each pixel. This results in the Markov random field model where the conditional probability of signals in each pixel (i,j) depends only on the signals in the neighbourhood $\{(i+m, j+n)$: (m,n) from the set $N\}$.
- Gibbs random fields (GRF). This theory is borrowed from Gibbs principal ensembles of statistical thermodynamics. We move from particles to pixels and still analyse potential and energy functions. Hence, GRF assigns a probability mass function to the entire lattice:

$$P(\mathrm{X} = \mathrm{x}) = \frac{1}{Z}\exp\left[-\sum_{c_i \in C} E(c_i)\right], \quad \forall \mathbf{x} \in \Omega \tag{3.4}$$

where Z is a normalizing constant known as the partition function and $E(c_i)$ is the energy function.

For texture analysis, general generic Gibbs random field models with multiple pairwise pixel interactions allow to relate the desired neighbourhood to a set of most ‘energetic’ pairs of the neighbours. A Gibbs distribution is usually defined with respect to cliques, i.e. proper subgraphs of a neighbourhood graph on the lattice. A clique is a particular spatial configuration of pixels, in which all its members are statistically dependent on each other. Then the interaction structure itself and relative frequency distributions of signal co-occurrences in the selected pixel pairs can serve as texture features.

Many natural surfaces have a statistical quality of roughness and self-similarity at different scales. Fractals are very useful and have become popular in modelling these properties in image processing but scale variations can have a great impact on the imaged appearance of a texture. Self-similarity across scales in fractal geometry is a crucial concept. The fractal dimension gives a measure of the roughness of a surface.

Fractal-based texture analysis was introduced by Pentland in 1984 [70]. To apply the fractal model to an image surface, we need to assume that: the intensity random function $I(x)$ is a fractal Brownian function and the fractal dimension of a fractal Brownian function is invariant over transformations of scale[1]. In order to obtain the fractalness of an image, Pentland introduced the description of the image change $\Delta I = I(x + \Delta x) - I(x)$ with scale as follows (eq. (5) [70]):

$$E(|\Delta I_{\Delta x}|)\ \|\Delta x\|^{-H} = E(|\Delta I_{\Delta x=1}|) \tag{3.5}$$

where: $E(|\Delta I_{\Delta x}|)$ is the expected value of the change in intensity ΔI over distance Δx, H is the Hurst exponent [71], [72]. Equation (3.5) is the mutual relation of the image intensities expressed in a statistical way.

We can assume that $\kappa = E(|\Delta I_{\Delta x=1}|)$, hence we obtain in the above equation $E(|\Delta I|) = \kappa\ \|\Delta x\|^{H}$. By applying log to both sides we have

$$\log E(|\Delta I|) = \log \kappa + H \log \|\Delta x\|. \tag{3.6}$$

The Hurst exponent H can be obtained by using the linear least-squares regression to estimate the slope of the grey-level difference $GD(k)$ versus k in a log-log scale over the interval $k = [1, s]$, s - the maximum value, where:

---

[1] Following Mandelbrot [276] the increments of a random function $\{X(t,\omega); -\infty < t < \infty\}$ are said to be self-similar with parameter $H \geq 0$ if for any $h > 0$ and any moment t_0

$$\{X(t_0+\tau,\omega) - X(t_0,\omega)\} \triangleq \{h^{-H}[X(t_0+h\tau,\omega) - X(t_0,\omega)]\}.$$

If $X(t_0,\omega)$ has self-similarity and stationary increments and is mean square continuous, then $0 \leq H < 1$ there is a constant V such that

$$E\,[\,X(t_0+\tau,\omega) - X(t_0,\omega)]^2 = V\tau^{2H}.$$

For images, following Pentland [70], instead of time t we speak about spatial dimension x, so we have $E[I(x+\Delta x) - I(x)]^2 = V\,\Delta x^{2H}$.

$$GD(k) = \frac{\sum_{x=1}^{N}\sum_{y=1}^{N-k-1}|I(x,y)-I(x,y+k)|}{2N(N-k-1)} + \\ + \frac{\sum_{x=1}^{N-k-1}\sum_{y=1}^{N}|I(x,y)-I(x+k,y)|}{2N(N-k-1)} \tag{3.7}$$

The fractal dimension *FD* can be derived from the relation *FD*=3-*H*. The approximation error of the regression line fit should be determined to prove that the analyzed texture is fractal, and thus be efficiently described using fractal measures. A small value of the fractal dimension *FD* implies that a large value of the Hurst exponent *H* represents fine texture, while a large *FD*, implying a smaller *H* value, corresponds to coarse texture [73].

Recently, texture descriptors have been based on transformational models. Let us recall the basic notions of the unitary transform. A general linear operation on the input image $I(x,y)$ results in an $M\times N$ output image $U(m,n)$ which is defined by:

$$U(m,n) = \sum_{x=0}^{K-1}\sum_{y=0}^{J-1} I(x,y)O(x,y;m,n) \tag{3.8}$$

where: $O(x,y;m,n)$ is the operator kernel.

Based on this universal rule we can chronologically describe the most useful and therefore the most common transformational methods:

- Fourier power spectra and Fast Fourier Transform (FFT) [74]. For image function $I(x,y)$ we compute its Fourier transform as:

$$F(u,v) = \frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} I(x,y)\exp\left\{\frac{-2\pi i}{N}(xu+yv)\right\} \tag{3.9}$$

  where (u,v) are the spatial frequencies and the quantity $|F(u,v)|^2$ is defined as the power spectrum which, in fact, is the modulus of a complex number. In the image terms the energy distribution of the power spectrum reflects the periodical structure of a texture, whereas the directional nature of the texture is reflected in the direction distribution of energy in the power spectrum. Frankly speaking, the limitation at high frequencies is the image resolution.

- The Gabor transform [75], [76]. The Fourier transformation is an analysis of the global frequency content in the signal. Many applications require the analysis to be localized in the spatial domain. This is usually handled by introducing spatial dependency into the Fourier analysis. The classical way is using the windowed Fourier transform. Considering one dimension, the windowed Fourier transformation of a sinusoidal wave $f_{u_0}(x) = e^{iu_0x}$ (illustrates in Fig. 3.4) is defined as: $F(u) = 2\pi\delta(u - u_0)$, where δ is the Dirac function. Then its energy is spread over the frequency interval in the neighbourhood of a u_0:

$$\left[u_0 - \frac{\sigma_u}{2}, u_0 + \frac{\sigma_u}{2}\right].$$

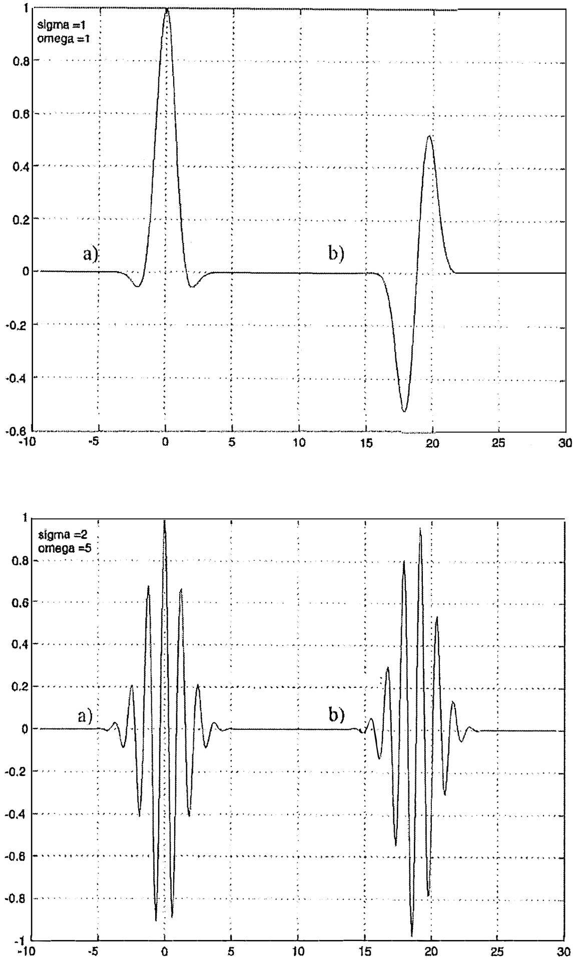


Fig. 3.4 Gabor function, where a) the real part of the function and b) the imaginary part of the function [79].

When the 2D window function $w(r)$ is Gaussian,

$$w(r) = \frac{1}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}; \qquad r^2 = x^2 + y^2 \tag{3.10}$$

the transform becomes a 2D Gabor transform [77], [78]:

$$G(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\pi\left[\frac{(x-x_0)^2}{\sigma_x^2} + \frac{(y-y_0)^2}{\sigma_y^2}\right]\right\} \cdot e^{i(u_0x+v_0y)} \tag{3.11}$$

where (x_0,y_0) is the centre of the receptive window in the spatial domain and (u_0,v_0) is the optimal spatial frequency of the filter in the frequency domain. σ_x and σ_y are the standard deviations of the elliptical Gaussian along x and y. The 2D Gabor function is thus a product of an elliptical Gaussian and a complex plane wave.

The 2D Gabor function consists of a sinusoidal plane wave of a certain frequency and orientation modulated by a Gaussian envelope given by:

$$g(x,y) = \exp\left\{-\frac{1}{2}\left[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right]\right\} \cos(2\pi\omega_0(x\cos\theta + y\sin\theta)) \tag{3.12}$$

where ω_0 and θ are the frequency and phase of the sinusoidal wave, respectively. Then a set of Gabor filters can be obtained by appropriate dilations and rotations of $g(x,y)$ for angles $\theta = \frac{n\pi}{K}$, $n = 0,1,\ldots, K\text{-}1$ (see Fig. 3.5). In this case, the Gabor transform of an image $I(x,y)$ is defined as:

$$W_n(x,y) = \int I(x,y)\, \overline{g_n}(x-x_0, y-y_0)\, dx_0 dy_0 \tag{3.13}$$

for which:

$$\mu_n = \int |W_n(x,y)|\, dxdy \tag{3.14}$$

$$\sigma_n = \sqrt{\int (|W_n(x,y)| - \mu_n)^2 dxdy} \tag{3.15}$$

where μ_n is the mean and σ_n is the standard deviation of the magnitude of $W_n(x,y)$ for a particular orientation, where n denotes the specific subband.
The texture analyzers based on 2D Gabor functions offer a strong correlation with the actual human segmentation and respective visual field profiles.

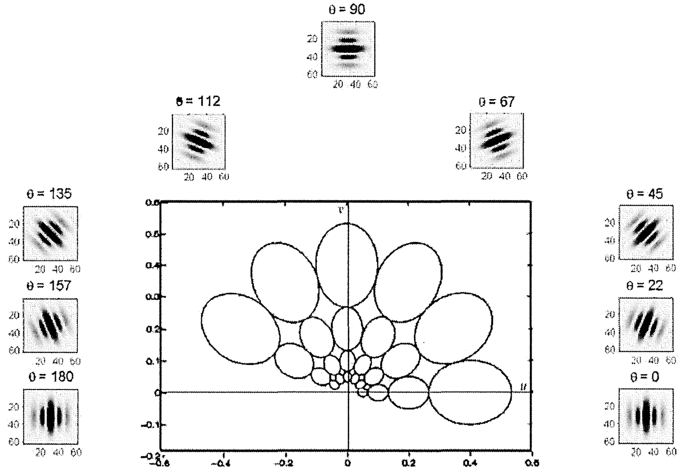


Fig. 3.5 Examples of 2D Gabor functions for particular angles $\theta = \frac{n\pi}{K}$, where K is the number of orientations. The outside windows present 2D Gabor filters, where $K = 9$. The central contours correspond to the half-peak magnitude of the filter responses in the set of Gabor filters with the upper and lower centre frequency of interest: $\omega_h = 0.4$ and $\omega_l = 0.05$, respectively, six orientations ($K = 6$), and four scales ($S = 4$), followed by **[80]**.

- The wavelet transformations are a big group of methods focused on the multiresolution analysis concept. Generally, the structures to be recognized differ significantly in size. Hence, it is impossible to define *a priori* an optimal resolution for image analysis. Burt [81] and Crowley [82] have each introduced pyramidal implementation to compute image details in different resolutions. A multiresolution analysis (MRA) yields a scale-invariant interpretation of the image. A multiresolution representation provides a simple hierarchical framework for interpreting the image information. In different resolutions, details of an image generally characterize different physical structures of the scene; in a coarse resolution, these details correspond to larger structures represented by 'big' image components.

  The idea of wavelets is based on a basic function called a wavelet (3.16) with two parameters: one - s, characterizing the scale, the other one - u, indicating the position of the function, introduced instead of the sinusoidal basic function with one parameter ω in the Fourier transform.

$$\psi_{su}(x) = \frac{1}{\sqrt{s}}\psi\left(\frac{x-u}{s}\right) \tag{3.16}$$

Hence, the 1D continuous wavelet transform is the projection of an $f(x)$ signal, in the $L^2(R)$, onto the function family $\{\psi_{su}, s > 0, u \in R\}$ generated from the single function ψ by translation and dilation:

$$[W_\psi f](s,u) = <\psi_{su}, f> = s^{-\frac{1}{2}} \int_{-\infty}^{+\infty} \overline{\psi\left(\frac{x-u}{s}\right)}\, f(x)dx \qquad (3.17)$$

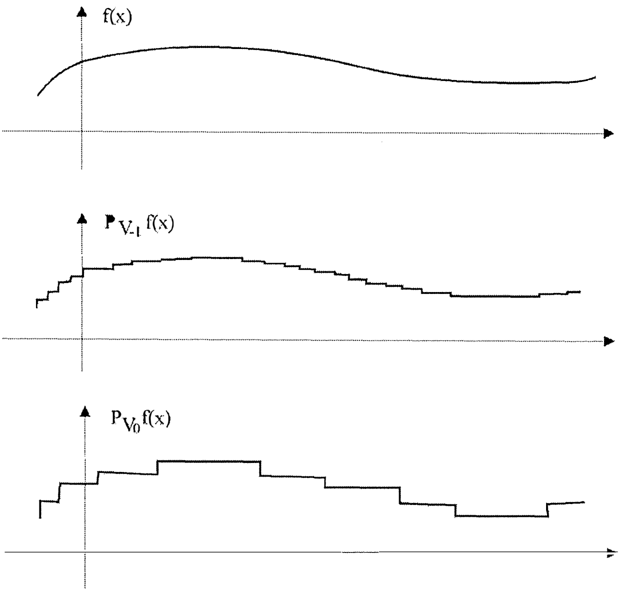


Fig. 3.6 A function $f(x)$ and its projection onto two consecutive levels V_{-1} and V_0 of the multiresolution analysis [83].

The idea of this method is presented in Fig. 3.6.
The redundancy of the continuous wavelet transform (3.17) can be cleared by discretizing both the scale factor s and the translation u. Then, we need a dyadic scale space, $s=2^j$ and $u=k$ with $j,k \in Z$ where Z is an integer. The fragment of the orthogonal basis with levels from V_{-3} to V_{-7} for Symmlet wavelets can be seen in Fig. 3.7.
The theory of multiresolution signal decomposition was developed by Mallat [84], [85] and Meyer [86] and thus the paradigm for constructing wavelets was established. Polish mathematicians were also involved in this analysis [87].

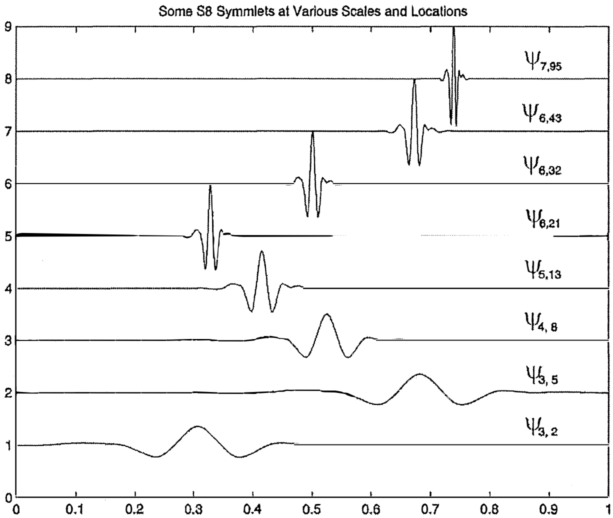


Fig. 3.7 An example of the dyadic Symmlet wavelets. A scale j and location k are presented for each wavelet $\psi_{j,k}$ on the right side [79].

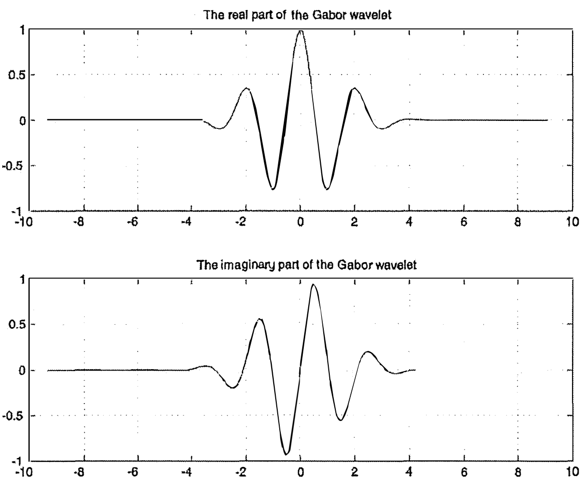


Fig. 3.8 The real and imaginary parts of the Gabor wavelet for σ=2 and ω=3 which are 'larger' than the example of the subset shown in Fig. 3.5 [79].

Many different wavelets have been introduced over the years, some of them are real and some others are complex. The Gabor wavelet (see Fig. 3.8) is an example of a function in the complex domain. Each ellipse from Fig. 3.5 represents the frequency support of a dyadic wavelet $\hat{\psi}^{K}_{2^j}$. This support size is proportional to 2^{-j} and its position rotates when K is modified.

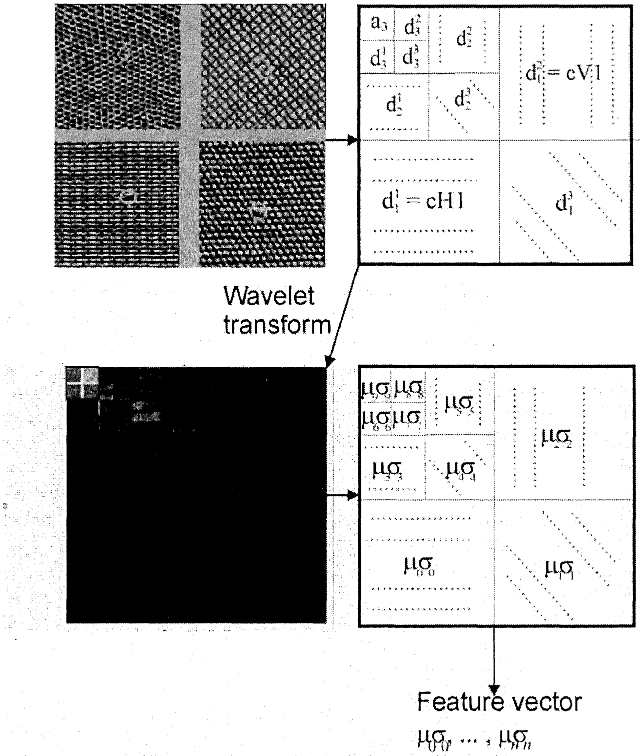


Fig. 3.9 A texture classifier flow chart based on the Gabor wavelet transformation (follows [80], [88]).

The discrete Gabor wavelet transform (GWT), for example, is applied in texture recognition and segmentation. Sebe and Lew [80] prepared an efficient method based on GWT parameters, such as μ and σ to be used as the texture feature vector. Fig. 3.9 presents four different grey texture representatives (top left) and the organization of wavelet image coefficients $d^{p}_{\mathrm{j,k,l}} = \langle I, \psi^{p}_{\mathrm{j,k,l}} \rangle$ [88]. The dotted lines show the direction of details (top right). In the bottom left square we can see a wavelet transform for texture images, whereas in the bottom right square we can see the organization of GWT coefficients (cf. (3.14) and (3.15)) that constitute feature vector $f = \{\mu_n, \sigma_n\}$.

At the same time, Faizal and Fausi [89] also used the DWT decomposition scheme and they noticed that the DWT of image $M{\times}M$ gave as a result $(3K{+}1) \times M^2$ coefficients. Based on this fact, they transformed wavelet coefficients to the 3D domain where they looked for clusters whose centroids characterized a texture.

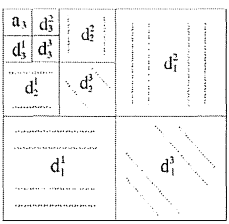
3.3.1 The Texture Approach to the Hybrid Semantic System

One of the reasons why so many kinds of wavelets were introduced in the previous section was to apply them to shape description. Our aim was to describe precisely the surface shape of an image function and we found out that the 'shape' of different wavelets corresponds more or less with the surface shape of the analysed image function. We should mention that from the shape matching point of view, one of the most important features of 2D wavelets is their detail directionality which is shown in Fig. 3.10 a) (called by Mallat as coefficients [88]). Thanks to this fact the convolution of wavelet and image element results in maximum values for concurrent element shapes (compare Fig. 3.9).

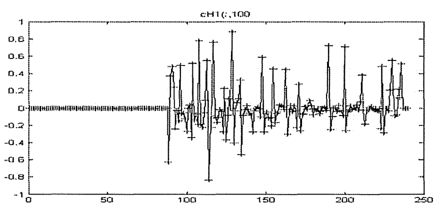
This wavelet property was applied by Jaworska [22], in order to describe the shape and size of the geometrical architectural texture. These parameters are different ones than those defined by Sebe and Lew or Faizal and Fausi in the previous subsection. We used the simplest wavelets, namely, the Haar wavelets which best fit a geometrical texture.

If we compute the convolution of an image consisting of regular tiles or bricks and vertical and horizontal details, we obtain a 2D transform whose maximum values are placed in the connection spots among these tiles or bricks. One dimension example of wavelet coefficients obtained from the convolution horizontal details and the 100th line of image segment is shown in Fig. 3.10 b).

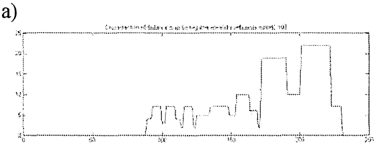
Having computed horizontal details, we have measured separately distances between maxima (shown in Fig. 3.10 c)) and between minima for each column of this matrix (shown in Fig. 3.10 d)). In this way, we created a 2D distance map in one direction for the analysed object. Repeating this procedure for vertical details, we obtained the second 2D distance map for the other direction. Based on 2D FWT maps of the object, texture is parametrised calculating the change distributions in the horizontal and vertical direction, respectively (see Fig. 3.10 g) and h)).



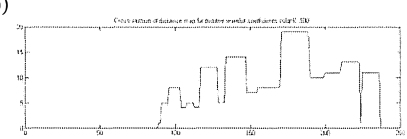
a)



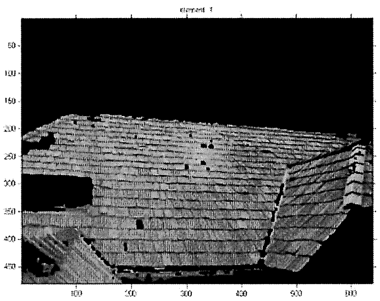
b)



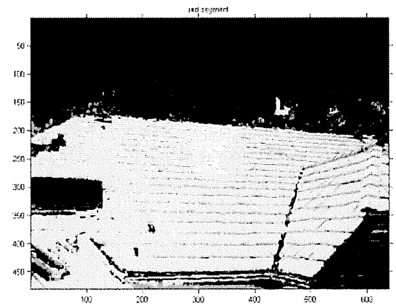
c)



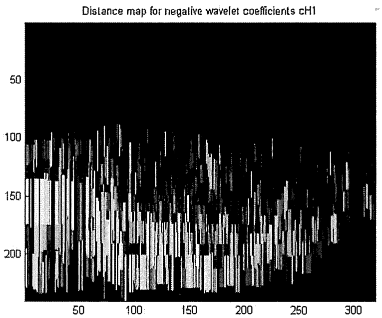
d)



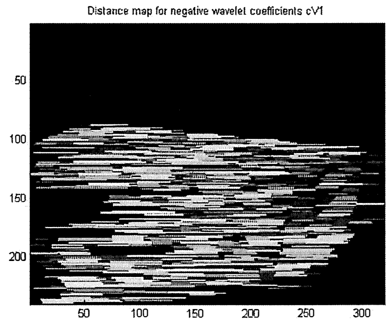
e)



f)



g)



h)

Fig. 3.10 Distance maps of texture calculated based on the 2D FWT with Haar wavelets. a) The disposition of wavelet image coefficients d_j^p where j is a multiresolution level, and a_j is an approximation at j level. b) Horizontal wavelet coefficients presented along the 100th column of the image transform (for the Haar wavelet, where j = 1). c) Cross-section through the 100th column of the distance map for positive horizontal wavelet coefficients. d) Cross-section through the 100th column of the distance map for negative horizontal wavelet coefficients. e) Original image of a roof segment (the segment was separated from the whole image based on our colour algorithm (cf. subsect. 4.2.3). f) The red component of the original image. g) Distance map for negative horizontal wavelet coefficients cH1. h) Distance map for negative vertical wavelet coefficients cV1 [49].

As a result of the above described algorithm, we obtain two ranges for the horizontal texture object components h and two others for the vertical one v.

$$T_p = \frac{[h_{\min 1,2}; h_{\max 1,2}]}{[v_{\min 1,2}; v_{\max 1,2}]} \tag{3.18}$$

3.4 Edge Detection

We understand an edge as a discontinuity in the image brightness which helps us to identify curved lines – edges separated segments. Therefore, in an ideal case, the result of applying an edge detector to an image may lead to a set of connected curves that indicate the boundaries of objects, the boundaries of surface markings, as well as curves that correspond to discontinuities in surface orientation.

Generally, to detect edges we have to assume an image model in which discontinuities of image brightness are likely to correspond to:

- discontinuities in depth;
- discontinuities in surface orientation;
- changes in material properties;
- variations in scene illumination.

The edge or contour can be defined as a parametric curve, polygon, or B-spline, but this can cause problems with the description of non-uniformed topological objects.

The method presented below covers grey images, because edge detection for colour images is more complicated. If a pixel falls on the boundary of an object in an image, then its neighbourhood is a zone of grey-level transition. Edge detection operators examine each pixel neighbourhood, and quantify the slope, as well as the direction, of the grey level transition.

There are several methods to do this, for example:

- thresholding [90]
- watershed algorithm [91];
- gradient methods;
- active contours;
- Hough transform;
- fuzzy thresholding [92];

3.4.1 Gradient Methods

In gradient methods we treat the slope and direction of a potential edge as the magnitude and direction of the gradient vector, respectively, we apply the second derivative of the intensity of a 2D image function $I(x,y)$, namely, the Laplacian:

$$\nabla^2 I(x,y) = \frac{\partial^2}{\partial x^2} I(x,y) + \frac{\partial^2}{\partial y^2} I(x,y) \tag{3.19}$$

The Laplacian is a linear, shift-invariant operator and its transfer function is equal to zero at the origin of a frequency space. Fig. 3.11 presents an example of edges and both derivatives.

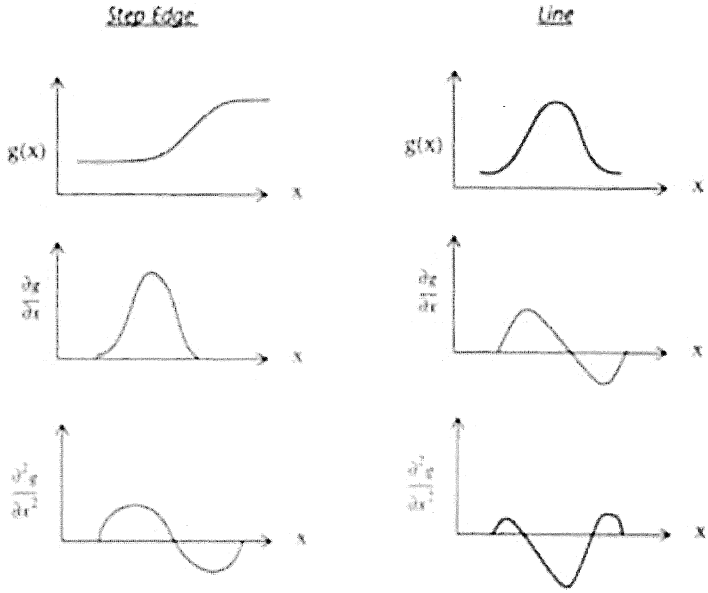


Fig. 3.11. The kind of edges (at the top), the first derivative of the edges (in the middle), the second derivative of the edges (at the bottom).

At present, for the discrete image version, most methods are based on convolution with a set of directional derivative masks **M** – filters:

$$\mathrm{H} = \mathrm{I} * \mathrm{M} \tag{3.20}$$

$$H(m,n) = \sum_i \sum_j I(i,j)\, M(m-i, n-j) \tag{3.21}$$

where the exemplary masks $\mathbf{M}_{[3\times 3]}$, for a discrete Laplacian $H(m,n)$ are shown in Table 3.2.

Table 3.2 Laplacian convolution kernels.

0	-1	0
-1	4	-1
0	-1	0

-1	-1	-1
-1	8	-1
-1	-1	-1

The other well-known edge operators are suggested by: Sobel [93], Prewitt [94] and Kirsch [90]. In all of them each pixel in the image is convolved with both kernels. One kernel responds maximally to a generally vertical edge and the other to a horizontal edge. The maximum value of the two convolutions is taken as the output value for that pixel. The kernels for the Sobel edge operator are shown in Table 3.3, whereas their results are presented in Fig. 3.12 c).

Table 3.3 The Sobel convolution kernels.

-1	0	1
-2	0	2
-1	0	1

1	2	1
0	0	0
-1	-2	-1

The kernels for the Prewitt edge operator are shown in Table 3.4.

Table 3.4 Prewitt convolution kernels

-1	0	1
-1	0	1
-1	0	1

1	1	1
0	0	0
-1	-1	-1

The Canny edge detector [95] is known to many as the optimal one for a number of reasons; the first and most obvious being its low error rate. It is important that edges presented in images should not be missing and that there are no responses to non-edges. The second reason is that the edge points are well localized, while the third that there is only one response to a single edge.

In the Canny algorithm the Gaussian filter, based on the 5×5 mask, is used to smooth the image because the larger the Gaussian mask, the lower the detector's rate of sensitivity to noise. Next, the Sobel operator is applied to estimate the gradient Gx in the x-direction and Gy in the y-direction. The magnitude, the so-called *edge strength* of the gradient is thus approximated, using the formula:

$$|G| = |Gx| + |Gy| \tag{3.22}$$

Then we find the edge direction

$$\theta = \arctan\left(\frac{Gy}{Gx}\right) \tag{3.23}$$

See the result of the Canny edge operator applying to one segmented colour layer in Fig. 3.12 d).

We also use this algorithm to create contour map for the whole image as it is seen in Fig. 3.2.

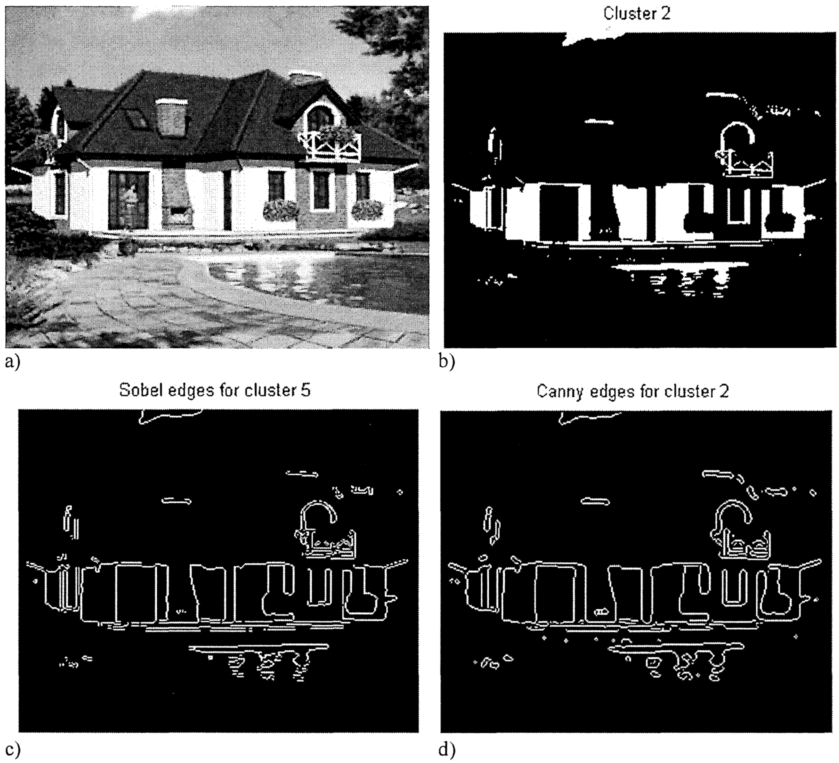


Fig. 3.12 An example of edge detection. a) the original image, b) a layer segmented by clustering, c) an example of the Sobel method for the layer from b), d) an example of the Canny method for the layer from b).

3.4.2 Boundary Tracking by Active Contours

In contrast to gradient-based representation, where the boundaries are detected based on the pixel intensity, the parametric model of active contours leads to the energy minimization problem.

Definition 3.2. (active contours)

Formally, let ρ be a metric (for instance the Euclidean metric) in $\mathbf{R}^2$ and $K(\mathbf{x_0}, \varepsilon) = \{\mathbf{x} \in \mathbf{R}^2 : \rho(\mathbf{x}_0, \mathbf{x}) < \varepsilon\}$ be a sphere with the centre $\mathbf{x}_0 \in \mathbf{R}^2$ and radius $\varepsilon < 0$. A set $c \subseteq \mathbf{R}^2$ is a contour if and only if there exists a function f: $\mathbf{R}^2 \to \mathbf{R}$, such as:
$$c = \{x \in R^2 : \bigvee_{\varepsilon > 0} \bigwedge_{x_1, x_2 \in K(x_0, \varepsilon)} f(x_1) \geq 0 \wedge f(x_2) < 0\}.$$

Active contours, or snakes, are computer-generated curves that move within images to find object boundaries [96]. A traditional snake is a parametric curve $C(p)=\begin{bmatrix} x(p) \\ y(p) \end{bmatrix}$, where $p\in[0,1]$, that moves through the spatial domain Ω of an image $I(x,y)$. A snake, which we call the gradient vector flow (GVF) snake, begins with the calculation of a field of forces, called the GVF forces, over the image domain.

$$J(C)=E_{int}(C)+E_{ext}(C) \tag{3.24}$$

The external energy function E_{ext} is derived from the image so it moves towards the image contour:

$$E_{ext}=\int_0^1 P(C(p))\,dp=-\nabla P(C(p)) \tag{3.25}$$

where $P(x,y)$ is a convolution of image $I(x,y)$ (seen as a line) with a 2D Gaussian function $G_\sigma(x,y)$ with a standard deviation σ, as follows:

$$P(x,y)=-|\nabla G_\sigma(x,y)*I(x,y)|^2 \tag{3.26}$$

The internal energy E_{int} controls the snake like a physical object resistant to both stretching and bending, towards the image boundaries:

$$E_{int}=\frac{1}{2}\int_0^1 \alpha|C'(p)|^2+\beta|C''(p)|^2\,dp \tag{3.27}$$

where the first derivative $C'(p)$ models stretching and elasticity, whereas the second derivative $C''(p)$ models bending and rigidity, where α and β are weight parameters.

A snake that minimizes $J(C)$ must satisfy the Euler equation:

$$\alpha C''(p)-\beta C''''(p)-\nabla P(C(p))=0 \tag{3.28}$$

that can be viewed as a force balance equation

$$F_{int}+F_{ext}^{(p)}=0 \tag{3.29}$$

where: $F_{int}=\alpha C''(p)-\beta C''''(p)$ and $F_{ext}^{(p)}=-\nabla E_{ext}$.

The GVF forces create the gradient of an image edge map (see Fig. 3.13).

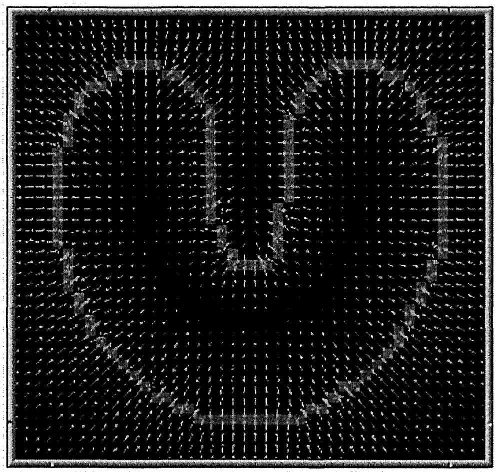


Fig. 3.13 A gradient vector flow (GVF) field for a U-shaped object. These vectors will pull an active contour towards the object boundary. (Follows: *Active Contours, Deformable Models, and Gradient Vector Flow* Chenyang Xu and Jerry L. Prince web page: http://www.iacl.ece.jhu.edu/static/gvf/)

In comparison with the classical edge detection techniques, snakes have multiple advantages:

- They produce closed and smooth object boundaries.
- They autonomously and adaptively search for the minimum state.
- External image forces act upon the snake in an intuitive manner.
- Incorporating Gaussian smoothing in the image energy function introduces scale sensitivity.

But they also have some key drawbacks:

- They are sensitive to local minima states.
- Minute features are often ignored during energy minimization over the entire contour.
- Their accuracy depends on the convergence policy.

3.4.3 Hough Transform

The classical Hough transform was concerned with the identification of lines in the image, but later the Hough transform was extended to identify the positions of arbitrary shapes, most commonly circles or ellipses.

The simplest variant of the Hough transform is used to detect of straight lines. In general, the straight line $y = mx + b$ can be represented as a point (b,m) in the parameter space. However, vertical lines pose a problem. Instead, Duda and Hart [97] propose the polar coordinate representation of a line:

$$\rho = x\cos\theta + y\sin\theta, \tag{3.30}$$

where ρ is the distance from the origin to the closest point on the straight line, and θ is the angle between the x axis and the line connecting the origin with that closest point (see Fig. 3.14).

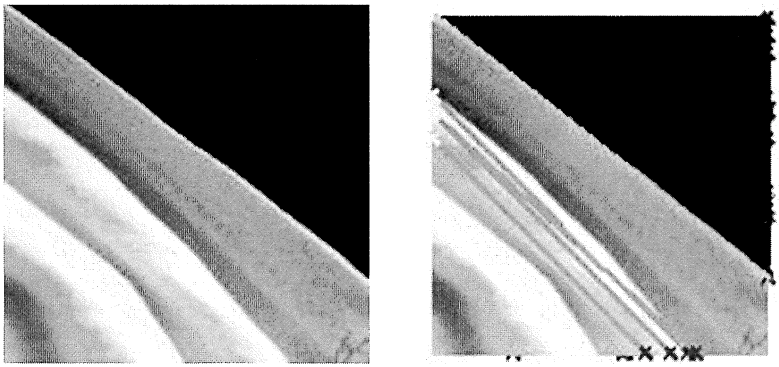


Fig. 3.14 Left: The original image. Right: Lines (green) found by the Hough transform.

Each point in image generates the sinusoid in Hough space (Fig. 3.15), and each point along this sinusoid corresponds to the ρ-θ values for a single line passing through the original point.

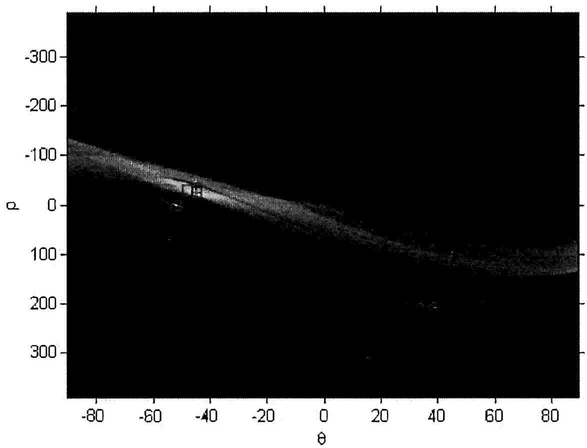


Fig. 3.15 The Hough transform space. White sinusoids represents lines visible in Fig. 3.14.

The Hough space is an accumulator space which means that it sums up the votes of many pixels in the image, and points in Hough space that have a large total vote are then interpreted as indicating the corresponding alignment on the real space image. To construct the Hough transform, every point present in the real-space image casts its votes into the Hough space for each of the lines that can possibly pass through it.

3.5 Shape Information

Shape extraction is a non-trivial operation, but shape-based methods are particularly challenging due to the intrinsic difficulties in dealing with shape location and recognition. Nevertheless, there is no doubt that shape is one of the basic features describing image content, hence, we define the key properties of shape:

- identifiability: shapes which are found perceptually similar by humans have the same/analogous features that are different from the others;
- translation, rotation, scale and affine invariance;
- the location, the rotation and the scaling changes of the shape must not affect the extracted features;
- occultation invariance: when some parts of a shape are occulted by other objects, the feature of the remaining part must not change compared to the original shape.
- statistically independent of two features. This represents compactness of the representation.
- reliability: as long as one deals with the same pattern, the extracted features must remain the same.

Shape description can be generally divided into two kinds of methods: contour-based and region-based. Under each kind, the methods are further divided into a structural and global approach based on whether the shape is represented as a whole or by segments (primitives). The whole breakdown is shown in Fig. 3.16.

Contour shape techniques only exploit shape boundary information. Zhu et al. [98] use salient contours, extracted from bottom-up contour grouping, as tokens for image-model shape matching. Shape matching with contours instead of isolated edges has several advantages. Long salient contours are more distinctive, which leads to efficiency of the search as well as the accuracy of shape matching. Furthermore, an accidental alignment causing false positive detections is removed by requiring the entire contour to match whole objects.

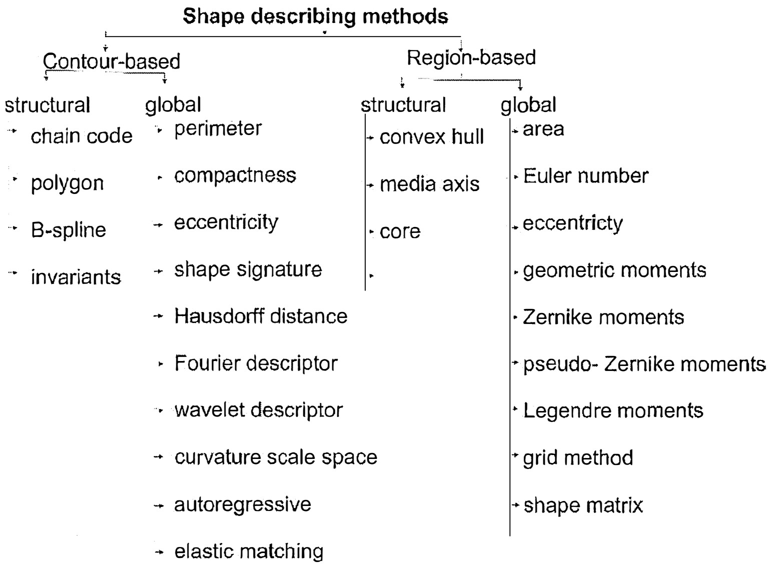


Fig. 3.16 Shape describing methods [99].

Boundary-based methods such as [100] represent shapes by the locations of the maxima of its curvature scale space (CSS) image. Shapes are smoothed by selecting the appropriate scale and then matched by shifting the CSS contours so that the major maxima of one image overlap that of the other [101]. The shape boundaries are approximated using planar curves and are progressively simplified through discrete curve evolution based on a novel relevance measure [100]. The weakness of the boundary-based approach is that it does not represent the interior of the shape [102] and is, therefore, very sensitive to spatial reconfigurations of parts and local boundary perturbations.

In region-based techniques, all the pixels within a shape region are taken into account to obtain the shape representation. Common region based methods use moment descriptors to describe shapes [103]. Other region based methods include the grid method, shape matrix, convex hull and media axis.

3.5.1 The Shape Approach to the Hybrid Semantic System

According to the shape descriptors presented above our approach belongs to the region-based group of methods (see Fig. 3.16). Generally, we define shape descriptors for separated segments. The method used to find these segments is introduced in section 4.2.3.

Assuming that we have segments described as a set of pixels, we can apply two types of moments: moments of inertia and Zernike moments. The former are very efficient as shape descriptors and can be calculated as:

$$\mu_{pq} = \sum_{x} \sum_{y} (x - \bar{x})^p \, (y - \bar{y})^q I(x, y), \qquad p, q = 0,1,2 \tag{3.31}$$

where p and q are the number of moments, $\bar{x}$ and $\bar{y}$ are segment centroids.
The second most efficient kind of descriptors is Zernike moments [104].

Definition 3.3 (Zernike moments)

Zernike moments are a set of complex polynomials $\{V_{pq}(x,y)\}$ which form a complete orthogonal set over the unit disk of $x^2 + y^2 \leq 1$. Hence, the definition of 2D Zernike moments with p^{th} order with repetition q for intensity image $I(x,y)$ of the image is described as:

$$Z_{pq} = \frac{p+1}{\pi} \iint\limits_{x^2+y^2 \leq 1} V^*_{pq}(x,y)\, I(x,y)\, dx dy \tag{3.32}$$

where: $V^*_{pq}(x, y) = V_{p,-q}(x, y)$. (3 33)

Generally, the first 10 Zernike moments, i.e. those from Z_{00} to Z_{33}, are sufficient as a shape feature (see Fig. 3.17). The scale invariance is obtained by/thanks to normalization of Z_{00} by the total number of image pixels.

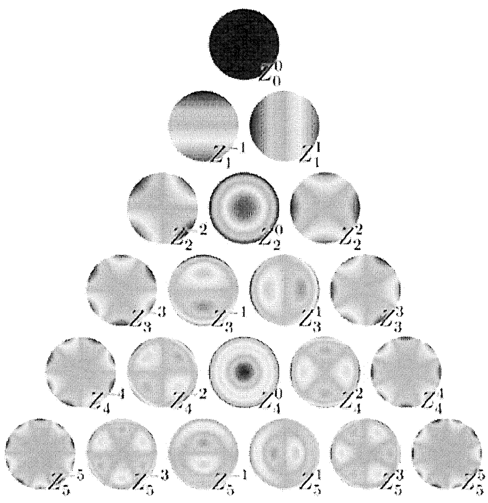


Fig. 3.17 The first Zernike base functions (followed Wikipedia).

Characteristic features of Zernike moments are: (i) invariance to rotation only; (ii) the translation invariance is achieved by the location of the original image centroid in the centre of the coordinates, (iii) the scale invariance is obtained by normalizing Z_{00} by the total number of image pixels.

Although the Zernike moment descriptor has a robust performance, it has several shortcomings. First, the kernel of Zernike moments is complex to compute, and the shape has to be normalized into a unit disk before deriving the moment features. Second, the radial features and circular features captured by Zernike moments are not consistent, one is in the spatial domain and the other is in the spectral domain. It does not allow multi-resolution analysis of a shape in a radial direction. Third, the circular spectral features are not captured evenly at each order, which can result in a loss of significant features which are useful for shape description.

To overcome these shortcomings, a generic Fourier descriptor (GFD) has been proposed by Zhang and Lu [99]. The GFD is acquired by applying a 2D Fourier transform on a polar-raster *PF:*

$$PF = \sum_{r}\sum_{i} I(r, \theta_i) \exp\left[2\pi j\left(\frac{r}{R}\rho + \frac{2\pi i}{T}\varphi\right)\right] \tag{3.34}$$

where: $0 \leq r < R$ and $\theta_i = i(2\pi/T)$; $0 \leq i < T$; $0 \leq \rho < R$, $0 \leq \varphi < T$. R and T are the radial frequency resolution and angular frequency resolution, respectively. The normalized coefficients are the GFD. The similarity between two shapes is measured by the city block distance between their GFDs.

It has been found that methods operating within the spatial domain suffer from two main drawbacks: noise sensitivity and a high dimension of the feature vector. The problems can be solved in four ways: histogram, moments, scale space and spectral transforms.

3.6 Local Feature Descriptors

Local feature descriptors represent a group of methods which allow the user to find local image structures in a repeatable way and to encode them in a representation that is invariant to a range of image transformations, such as translation, rotation, scaling, and affine deformation. The purpose of introducing local feature descriptors is to provide a representation that enables the user to efficiently match local structures of images. For this objective, the feature extractors must fulfil three important criteria:

1. The feature extraction process should be repeatable and precise, so that the same features are extracted on two images showing the same object.
2. At the same time, the features should be distinctive, so that different image structures can be held apart from each other.

3. Proper feature descriptors should be resistant to accidental variance of features and invariant to scaling and rotation of image.

Herein, we present a group of seven local feature descriptors developed recently. Their general advantage is high precision in matching the required object, but their drawback is long running time and the necessity to provide a query-by-example.

3.6.1 Scale-Invariant Feature Transform (SIFT)

The scale-invariant feature transform (SIFT) was introduced by Lowe [24], [23] to identify objects in two images, even if these objects were cluttered or under partial occlusion, because the SIFT feature descriptor is invariant to uniform scaling and orientation. In like manner, it is partially invariant to affine distortion and illumination changes.

The algorithm starts from key-points detection in order to identify locations and scales that can be repeatedly assigned under differing views of the same object. Key-point locations are defined as maxima and minima of the difference of the Gaussians $G(x,y,\sigma)$ applied in a scale-space to a series of smoothed and resampled images.

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \tag{3.35}$$

where $L(x,y,\sigma)$ is the product of a convolution. This difference of Gaussians is calculated for two nearby scales with k factor:

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \tag{3.36}$$

The scale-space extrema detection produces too many key-point candidates, so at first, for each candidate key-point, interpolation of nearby data is used to accurately determine its position. The interpolation is done using the quadratic Taylor expansion of the difference-of-Gaussian scale-space function, D with the candidate key-point as the origin. This Taylor expansion is given by:

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T\frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x} \tag{3.37}$$

for $\mathbf{x} = (x,y,\sigma)^{\mathrm{T}}$.

Low contrast candidate points and edge response points along an edge are discarded. Dominant orientations are assigned to localized key-points. These steps ensure that the key-points are more stable for matching and recognition. The SIFT descriptors robust to local affine distortion are then obtained by considering pixels around a radius of the key location, blurring and resampling of local image orientation planes.

The next step is the orientation assignment when each key-point is assigned one or more orientations based on local image gradient directions. First, the Gaussian-smoothed image $L(x,y,\sigma)$ at the key-point's scale σ is taken so that all computations are performed in a scale-invariant manner. For an image sample $L(x,y)$ at scale σ, the gradient magnitude $m(x,y)$ and orientation $\theta(x,y)$ are precomputed using pixel differences:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \quad (3.38)$$

$$\theta(x,y) = \arctan\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right) \quad (3.39)$$

assuming that $\forall(x,y)$ in the neighbourhood (x_0,y_0).

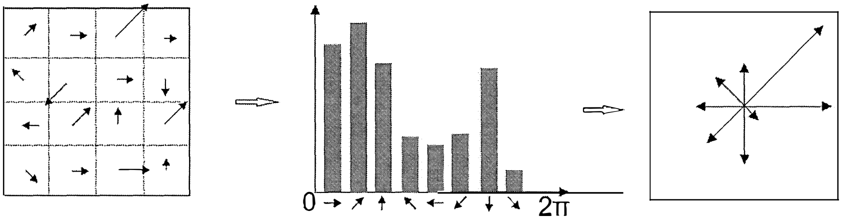


Fig. 3.18 The gradient magnitude and orientation at each point of a 4x4 set of samples (on the left) which are accumulated into orientation histograms with 8 bins each (in the middle). The key-points descriptor summarizes the contents over 4x4 subregions, as shown on the right, with the length of each arrow corresponding to the sum of the gradient magnitudes near that direction within the region. Peaks in the orientation histogram correspond to dominant directions of local gradients.

The SIFT key samples generated at a larger scale are given twice the weight of those at a smaller scale. This means that the larger scale is in effect able to filter the most likely neighbours for the smaller scale. This also improves recognition performance by giving more weight to the least-noisy scale. To avoid the problem of boundary effects in bin assignment, each key-point match votes for the 2 closest bins in each dimension, giving a total of 16 entries for each hypothesis and further broadening the pose range.

Hough transform (as it has been described in subsect. 3.4.3, cf. (3.30)) is used to cluster reliable model hypotheses to search for keys that agree upon a particular model pose. When feature clusters are found to vote for the same pose of an object, the probability of the interpretation being correct is much higher than for any single feature. An entry in a hash table is created predicting the model location, orientation, and scale from the match hypothesis. The hash table is searched to identify all clusters of at least 3 entries in a bin, and the bins are sorted into decreasing order of size.

Each identified cluster is then subject to a verification procedure in which a linear least squares solution is performed for the parameters of the affine transformation relating the model to the image. The affine transformation of a model point $[x\ y]^T$ to an image point $[u\ v]^T$ can be written as below:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{3.40}$$

where the model translation is $[t_x\ t_y]^T$ and the affine rotation, scale, and stretch are represented by the parameters m_1, m_2, m_3 and m_4. In order to find the transformation parameters the equation (3.40) can be reformulated to group the unknowns into a column vector.

$$\begin{bmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \\ & & \cdots & & & \\ & & \cdots & & & \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} u \\ v \\ \vdots \end{bmatrix} \tag{3.41}$$

This equation presents a single match, but any number of further matches can be added, with each match contributing two more rows to the first and last matrix. At least 3 matches are needed to provide a solution. We can write this linear system as

$$A\mathbf{x} \approx \mathbf{b} \tag{3.42}$$

where A is a known n-by-p matrix ($n > p$), $\mathbf{x}$ is an unknown p-dimensional parameter vector, and $\mathbf{b}$ is a known n-dimensional measurement vector. Therefore, the minimizing vector $\mathbf{x}$ is a solution of the normal equation.

$$\mathrm{A}^T\mathrm{A}\mathbf{x} = \mathrm{A}^T\mathbf{b} \tag{3.43}$$

hence, we obtain:

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b} \tag{3.44}$$

which minimizes the sum of the squares of the distances from the projected model locations to the corresponding image locations.

3.6.2 RootSIFT

SIFT was originally designed, by Lowe [23], to be used with the Euclidean distance, but since there is a histogram comparison, Arandjelović and Zisserman

[105] introduced alternative histogram distance measures, namely the Hellinger kernel.

The Hellinger kernel for two L_1 normalized histograms, x and y (i.e. $\sum_i^n x_i = 1$ and $x_i \geq 0$), is defined as follows:

$$H(x,y) = \sum_{i=1}^{n} \sqrt{x_i y_i} \tag{3.45}$$

where n is a number of vector with unit Euclidean norm such as: $||\mathbf{x}||_2 = 1$.

The RootSIFT application slightly increases the average precision of retrieval.

3.6.3 Rotation-Invariant Generalization of SIFT (RIFT)

RIFT is a rotation-invariant variant of the SIFT method dedicated to texture images where the notion of orientation is difficult to define. The RIFT descriptor is constructed using circular normalized patches divided into concentric rings of equal width and within each ring a gradient orientation histogram is computed [106], [55]. To maintain rotation invariance, the orientation is measured at each point relative to the direction pointing outward from the center (see Fig. 3.19).

When the size of the Laplacian-of-Gaussian (LoG) kernel matches with the size of a blob-like structure, the response attains an extremum:

$$|\mathrm{LoG}(\mathbf{x}, \sigma_n)| = \sigma_n^2 |L_{xx}(\mathbf{x}, \sigma_n) + L_{yy}(\mathbf{x}, \sigma_n)|. \tag{3.46}$$

The LoG kernel can therefore be interpreted as a matching filter.

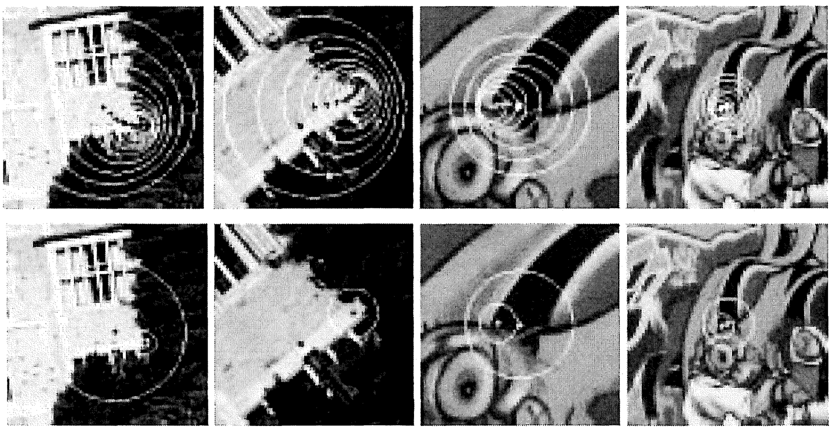


Fig. 3.19 Scale invariant interest point detection: (Top) Initial multi-scale Harris points (selected manually) corresponding to one local structure [106].

3.6.4 Fisher Vector (FV)

The Fisher kernel has been proposed, in the context of measuring the amount of information that an observable random variable X carries about an unknown parameter λ of a distribution that models X. Formally, it is the variance of the score, or the expected value of the observed information. From the theory of information we know that the entropy of a random value V is:

$$H(V) = - \sum p(v) \log(p(v)) \tag{3.47}$$

In the context of image retrieval the FV are usually ℓ^2-normalized since, as proved it Perronnin et al. [107] this is a way to eliminate the fact that distinct images contain different amounts of image-specific information. The Fisher Vector is applied to non-binary local features, using the Gaussian Mixture Model to represent the average distribution $p(X\,|\,\lambda)$.

For sample $X = \{x_t,\ t = 1,\dots,T\}$ of observations, which is a set of T local descriptors extracted from an image and $p(X\,|\,\lambda)$ represents a probability density function with λ parameters, the gradient ∇_λ of the log-likelihood describes the contribution of the parameters to the generation process [107]:

$$G(X|\lambda) = \frac{1}{T}\nabla_\lambda \log p(X|\lambda) \tag{3.48}$$

The dimensionality of this vector depends only on the number of parameters λ, not on the number of patches T.

Further, this approach has been moved to an image classification, where the Fisher kernel K method is to derive a function that measures the similarity between two sets of data X and Y, such as the sets of local descriptors extracted from two images. The idea is to characterize a signal with a gradient vector derived from a probability density function (pdf) which models the generation process of the signal [108].

$$K(X,Y) = \mathrm{G}'(X|\lambda)\ \mathrm{F}^{-1}(\lambda)\ \mathrm{G}(Y|\lambda) \tag{3.49}$$

where Fisher information matrix: $\mathrm{F}^{-1}(\lambda) = L'(\lambda)L(\lambda)$.

This representation can then be used as input to a discriminative classier. For the problem of image categorization the input signals are images. Perronnin and Dance proposed to use, as a generative model the Gaussian Mixture Models (GMM) which approximates the distribution of low-level features in images, i.e. a visual vocabulary.

Krapac and Šegvić proposed in [109] to use large FVs to object location on video images. They detect multiple objects on complex backgrounds, for instance, road signs.

3.6.5 Vectors of Locally Aggregated Descriptors (VLAD)

The VLAD method analyzes the local descriptors contained in an image to create statistical summaries that still preserve the effectiveness of local descriptors and allow treating them as global descriptors [110]. These image descriptors were designed to be very low dimensional (e.g. 16 bytes per image).

This method encodes a set of local feature descriptors $F = (x_1, \ldots, x_k)$, extracted from an image treated as a codebook with k visual words, using a dictionary based on a clustering method, such as GMM or k-means clustering. Each local descriptor x_j is then associated with its nearest centroid $\mathrm{NN}(x_j) = \mu_i$.

$$v_{i,j} = \sum_{x_j:\mathrm{NN}(x_j)=\mu_i} (x_j - \mu_{i,j}) \tag{3.50}$$

where $i = 1, \ldots, k$ is the index of visual word (k – number of centroids) and $j = 1, \ldots, d$ is the local descriptor component. Hence, the whole image representation dimension is $D = k \times d$.

For each cluster, the residual vectors (i.e. the difference between the centroid and the associated descriptors) are accumulated and the sum of the residual is concatenated into a single vector $V = [v_1^T \ldots v_k^T]$. Next, vector v is normalised by $v{:=}v/||v||_2$ and the Euclidean distance is sufficient to compare two VLADs.

3.6.6. Features from accelerated segment test (FAST)

FAST is a corner detection method, introduced by Rosten and Drummond in 2006 [111], which could be used to extract feature points and later used to track and map objects [112]. A FAST corner detector uses a circle of 16 pixels (a Bresenham circle of $r = 3$) to classify whether a candidate point p is actually a corner. Each pixel in the circle is labelled from integer number 1 to 16 clockwise. For a set of N contiguous pixels, if the pixels in the circle are all brighter than the intensity of candidate pixel p (denoted by I_p), plus a threshold value t, or all darker than the intensity of candidate pixel p, minus threshold value t, then p is classified as a corner. The conditions can be written as:

1. A set of N contiguous pixels S, $\forall\, x \in S$, the intensity of x denoted by (I_x) can be $I_x > I_p + t$;
2. A set of N contiguous pixels S, $\forall\, x \in S$, $I_x < I_p - t$;

So when either of the two conditions is met, candidate p can be classified as a corner. There is a tradeoff between selecting N, the number of contiguous pixels and the threshold value t. Then, N is usually selected as 12. A high-speed test method could be applied to exclude non-corner points.

Generally, the FAST detector is employed to find objects in video frames because of its effectiveness.

3.6.7 Oriented FAST and Rotated BRIEF (ORB)

ORB is basically a fusion of the FAST key-point detector and a BRIEF descriptor with many modifications to enhance the performance introduced by Rublee et al. [113] in 2011. First, it uses FAST to find key-points, then apply the Harris [114] corner measure to find top N points among them. It also uses a pyramid to produce multiscale-features.

In order to compute orientation, they found moments of order p and q, such as:

$$m_{pq} = \sum_x \sum_y x^p y^q I(x,y), \quad p,q = 0,1,2 \tag{3.51}$$

and intensity centroids: $C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right)$.

Then, a vector $\overrightarrow{OC}$ from the corner's center, O, to the centroid, can be constructed. The orientation of the patch then simply is:

$$\theta = \operatorname{atan2}(m_{01}, m_{10}), \tag{3.52}$$

where atan2 is the arctangent function with two arguments. In order to improve the rotation invariance of this measure authors made sure that moments are computed with x and y remaining within a circular region of radius r. They empirically selected r to be the patch size, so that that x and y run from $[-r, r]$.

Next, the Binary Robust Independent Elementary Features (BRIEF) descriptor is used for a simple binary test τ between pixels in a smoothed image patch **p**, as follows:

$$\tau(\mathrm{p}; \mathrm{x}, \mathrm{y}) := \begin{cases} 1: & \mathrm{p}(x) < \mathrm{p}(y) \\ 0: & \mathrm{p}(x) \geq \mathrm{p}(y) \end{cases} \tag{3.53}$$

where **p**(x) is the intensity of **p** at a point x. The feature is defined as a vector of n binary tests:

$$f_n(\mathrm{p}) := \sum_{1 \leq i \leq n} 2^{i-1} \tau(\mathrm{p}; \mathrm{x}_i, \mathrm{y}_i) \tag{3.54}$$

The test pairs of **x** and **y** are selected by Gaussian distribution around the centre of the patch or PCA for good discrimination. As tests for typical frames of size 640×480 proved, the ORB descriptor gives significant time decreasing.

3.7 Standardization Efforts - MPEG-7

The Moving Picture Experts Group (MPEG) [115] is a working group of authorities that was formed by ISO and IEC to set standards for audio and video compression and transmission. The MPEG-7 is a multimedia content description interface that uses XML to store metadata, and can be attached to the timecode in order to tag particular events. The description of the standard can be found in [116].

The main elements of the MPEG-7 standard are:

- Descriptors (D) that define the syntax and the semantics of each feature (metadata element);
- Description Schemes (DS) that specify the structure and semantics of the relationships between their components which may be both descriptors and description schemes;
- A Description Definition Language (DDL) to define the syntax of the MPEG-7 Description Tools and to allow the creation of new description schemes and, possibly, descriptors and to allow the extension and modification of existing description schemes;

The above-mentioned tools deal with binarization, synchronization, transport and storage of descriptors (see Fig. 3.20) [117].

MPEG-7 Visual Description Tools consist of basic structures and descriptors that cover the following basic visual features: color, texture, shape, motion, localization, and face recognition. Each category consists of elementary and sophisticated descriptors.

MPEG-7 Multimedia Description Schemes (DSs) are metadata structures for describing and annotating audio-visual (AV) content. The DSs provide a standardized way of describing in XML the important concepts related to AV content description and content management in order to facilitate searching, indexing, filtering, and access. The DSs are defined using the MPEG-7 Description Definition Language (DDL), which is based on the XML Schema language, and are instantiated as documents or streams. The resulting descriptions can be expressed in a textual form (i.e., human readable XML for editing, searching, filtering) or compressed binary form (i.e., for storage or transmission).

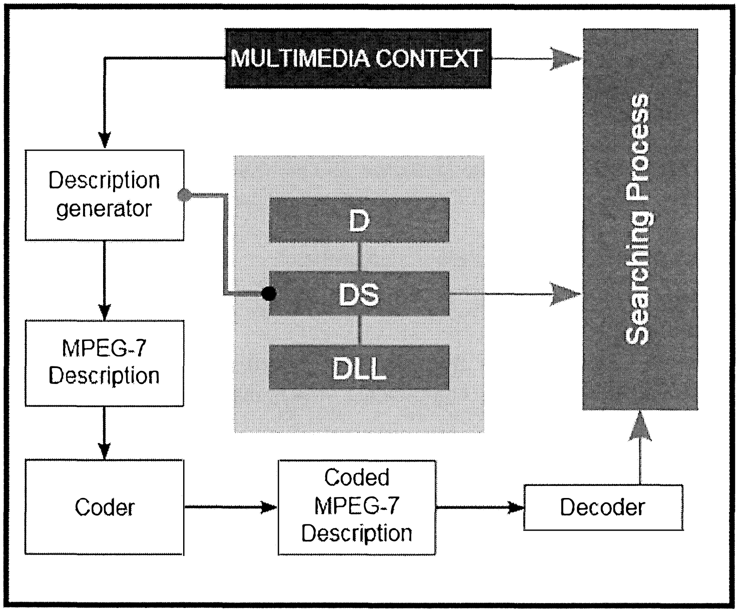


Fig. 3.20 Relations between different tools and the elaboration process of MPEG 7 [117].

3.8 Global Versus Local Comparison of Features

A feature is defined to capture a certain visual property of an image, either globally for the entire image or locally for a small group of pixels. So far, we have described local features i,e. related to the most basic structure of images, such as pixels. To reduce computation, an image may be divided into small, non-overlapping structures like lines, pixel neighbourhoods, subsets of pixels or patterns, and later the features are computed individually for every structure. The features are still local because of the small block size, but the amount of computation is radically reduced in comparison with that needed for obtaining features around every pixel.

In turn, global features capture the overall characteristics of an image. The advantage of global extraction is its high speed for both extracting features and computing similarity.

Both global and local features can be represented as a feature vector. There is one global feature for an image as well as many feature vectors **x**, **y** with local features, also for this image, additionally, to each feature its weight $\xi(i)$ can be attributed. Then the Euclidean distance, measuring the distance between $\mathbf{x}$ and $\mathbf{y}$ vector, can be calculated as:

$$d_E = \sqrt{(\mathbf{x}-\mathbf{y})^T \mathrm{diag}(\xi_i^2)(\mathbf{x}-\mathbf{y})} \tag{3.55}$$

Generally, a histogram was widely used as a global feature because it is very fast to compute and easy to compare with another histogram. A histogram can then be treated as a k-dimensional vector ($f_1, f_2, \ldots, f_k$), where f_i is the frequency of the l^{th} bin. We assume that two distributions (two images) have two histograms $H_0(i)$ and $H_1(i)$ with $0<i<n$ bins each.

We can, for example, compare histograms based on the Minkowski distance (with p=3):

$$d_{L_p}(H_0, H_1) = \left[\sum_{i=1}^{n} |H_0(i) - H_1(i)|^p\right]^{\frac{1}{p}} \tag{3.56}$$

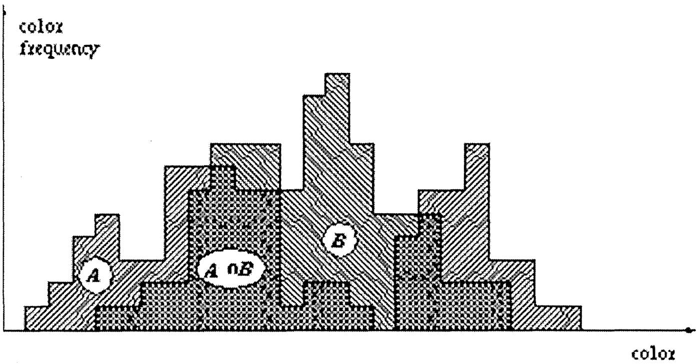


Fig. 3.21 Histogram intersection.

This kind of distance is very sensitive to even small shifts because the same bars must be compared with each other.

The histogram intersection [118] is widely used because of its ability to handle partial matches when the areas of the two histograms (the sum over all the bins (3.57)) are different (see Fig. 3.21). It is shown by Swain and Ballard [118] that when the areas of the two histograms are equal, the histogram intersection is equivalent to the (normalized) L_1 distance.

$$d_\cap(H_0, H_1) = 1 - \frac{\sum_i \min(H_0(i) - H_1(i))}{\sum_i H_1(i)} \tag{3.57}$$

When the feature vector represents relative frequency distribution (e.g., a normalised grey level co-occurrence histogram), for example, texture features, the dissimilarity can also be measured by the relative entropy, or Kullback-Leibler (K-L) divergence. Let D_J denote the divergence between two distributions, $H_0(i)$ and $H_1(i)$, which is based on vector quantization. Then:

$$D_J(H_0, H_1) = \sum_{i=1}^{n} H_0(i) \log \frac{H_0(i)}{H_1(i)} \tag{3.58}$$

This dissimilarity measure is asymmetric and does not represent a distance because the triangle inequality is not satisfied. The symmetric distance is obtained by averaging H_0 and H_1 [119]. In histogram notation we can describe this distance as:

$$D_J(H_0, H_1) = \sum_{i=1}^{n} \left[H_0(i) \log \frac{H_0(i)}{m_i} + H_1(i) \log \frac{H_1(i)}{m_i} \right] \tag{3.59}$$

where: $m_i = \frac{H_0(i)+H_1(i)}{2}$. Additionally, the K-L divergence is sensitive to histogram binning.

In 1998 Rubner et al. [12] introduced the Earth Mover's Distance (EMD) which was understood as the minimal cost that must be paid to transform one distribution (histogram) into the other, where there is a *distance* $d_{i,j}$ (between bin i and j) and is meant as the distance between the basic features that are aggregated into the histogram. Computing the EMD is based on a solution to the well-known *transportation problem.*

Given two histograms H_i and H_j, the EMD is:

$$\text{EMD}(H_i, H_j) = \frac{\min\limits_{f_{i,j}} \sum_{i,j} (f_{i,j} d_{i,j})}{\sum_{i,j} (f_{i,j})} \tag{3.60}$$

for the following constraints: $f_{i,j} \geq 0$, $\sum_j f_{i,j} \leq H_i$; $\sum_i f_{i,j} \leq H_j$;

$$\sum_{i,j} (f_{i,j}) = \min \left\{ \sum_{i} (H_i) , \sum_{j} (H_j) \right\}$$

where: $f_{i,j}$ denotes the flow which allows us to move some amount of 'mass' from H_i to H_j and vice versa; the flow cannot be higher than neither H_i nor H_j and the last constraint describes the maximum mass possible to move.

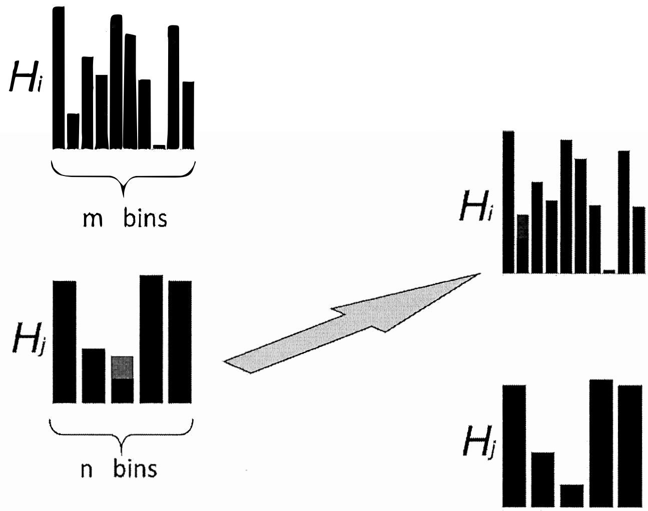


Fig. 3.22 Transport of 'mass' from H_j to H_i.

The EMD naturally extends the notion of a distance between individual elements to that of a distance between sets, or distributions, of elements. The advantages of the EMD over previous definitions of distribution distances are significant.

If the ground distance is a metric and the total weights of two signatures are equal, the EMD is a true metric, which allows endowing image spaces with a metric structure.

3.8 From Features to Signature

Generally, a signature describes the image information (also known as a global image descriptor) and it can also be seen as a mathematical function. The principal purpose of this function is to extract from a large image data structure.

The objective is that computed signatures enable us to determine similarities (i.e have matching features) between images that represent, for instance, the same scene but captured from different points of view. It means that most of the visual applications required that two images with high perceptual similarities have resembling signatures.

Signature generation functions can be roughly classified in three main groups, depending on the input data used to generate the global descriptor:

1. Appearance-based: the signature is calculated from texture, colour information, transformations in the frequency space or matrix factorizations .

2. Feature-based: the signature is calculated from the image key-points and their descriptors like those used in the SIFT, RIFT, etc.
3. Region-based: the signature is defined on the distance between sets of vectors, which is not as obvious as defining distance between single vectors.

A signature $\{\mathbf{s}_j = (\mathbf{m}\, j;\ w\mathbf{m}\, j)\}$ can represent a set of feature clusters, where a cluster is represented by its mean (or mode) m_j, and by the fraction $w\mathbf{m}\, j$ of pixels that belong to that cluster, $1 \leq j \leq n$, where n depicts the complexity of the particular image and the representative m_j is a d-dimensional vector. In general, vector quantization algorithms can be used for the clustering, as long as they are applied on every image independently, and they adjust the number of clusters to the complexities of individual images. On this assumption, a histogram $\{h_j\}$ can be viewed as a signature $\{\mathbf{s}_j = (\mathbf{m}\, j;\ w\mathbf{m}\, j)\}$ in which j^{th} cluster maps the point m_j with the central value in bin j of the histogram, and then w_j is equal to h_j, which better represents the image content.

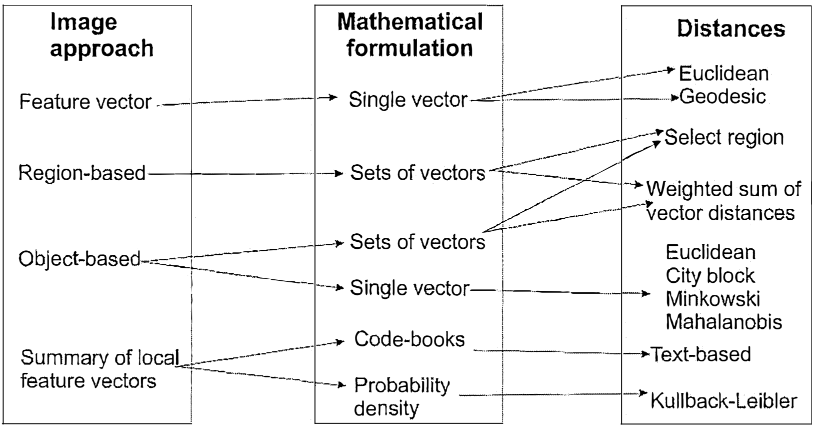


Fig. 3.23 Different types of image similarity measure and their mathematical formulations.

Additionally, the global signature based on the image histogram can be seen as: $\{(z_1', f_1'), \ldots, (z_k', f_k')\}$, where f_l' is the percentage of $x_{i,j}$'s grouped into cluster l, and z_l' is a bin's location. The collection of pixels (i, j) for which $x_{i,j}$'s are in the same cluster forms a relatively homogeneous region because the common cluster forces closeness between the visual features in $x_{i,j}$'s.

As shown in Fig. 3.23, similarity computation can be performed with feature vectors, region-based signature, object-based signature, or summarized local features.

References

[1] Y. Yao, Y. Zeng, N. Zhong and X. Huang, "Knowledge Retrieval," in *Proceedings of the 2007 IEEE/WIC/ACM International Conference on Web Intelligence*, Silicon Valley, USA, 2007.

[2] "http archive," 2016. [Online]. Available: http://httparchive.org/trends.php?s=Top1000&minlabel=Jan+20+2011&maxlabel=Oct+15+2014#bytesImg&reqImg.

[3] S. Nandagopalan, B. S. Adiga and N. Deepak, "A Universal Model for Content-Based Image Retrieval," *World Academy of Science, Engineering and Technology,* vol. 46, pp. 644-647, 2008.

[4] M. Yasmin, S. Mohsin, I. Irum and M. Sharif, "Content Based Image Retrieval by Shape, Color and Relevance Feedback," *Life Science Journal,* vol. 10, no. 4s, pp. 593-598, 2013.

[5] M. Rehman, M. Iqbal, M. Sharif and M. Raza, "Content Based Image Retrieval: Survey," *World Applied Sciences Journal,* vol. 19, no. 3, pp. 404-412, 2012.

[6] Y. J. Lee, . L. C. Zitnick and M. F. Cohen, "ShadowDraw: Real-time User Guidance for Freehand Drawing.," *ACM Transactions on Graphics (TOG),,* vol. 30, no. 4, pp. 1-27, July 2011.

[7] T. M. Lehmann, M. O. Güld, C. Thies, B. Fischer, D. Keysers, K. Spitzer, H. Ney, M. Kohnen, H. Schubert and B. B. Wein, "Content-Based Image Retrieval in Medical Applications," *Methods on Imformatic in Medicine,* vol. 43, pp. 354-361, 2004.

[8] S. Antani, J. Cheng, J. Long, R. L. Long and G. R. Thoma, "Medical Validation and CBIR of Spine X-ray Images over the Internet," in *Proceedings of IS&T/SPIE Electronic Imaging. Internet Imaging VII*, San Jose, C, 2006.

[9] R. K. Srihari, "Automatic Indexing and Content-Based Retrieval of Captioned Images," *IEEE Computer,* vol. 28, no. 9, pp. 49-56, September 1995.

[10] V. Khanaa, M. Rajani, K. Ashok and A. Raj, "Efficient Use of Semantic Annotation in Content Based Image Retrieval (CBIR)," *International Journal of Computer Science Issues,* vol. 9, no. 2, pp. 273-279, March 2012.

[11] C. Carson, S. Belongie, H. Greenspan and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," *IEEE Transaction on Pattern Analysis and Machine Intellignece,* vol. 24, no. 8, pp. 1026-1038, Aug. 2002.

[12] Y. Rubner, C. Tomasi and L. J. Guibas , "The Earth Mover's Distance as a Metric for Image Retrieval," *International Journal of Computer Vision,* vol. 40, no. 2, pp. 99-121, 2000.

[13] B. Xiao , X. Gao, D. Tao i X. Li, „Recognition of Sketches in Photos," w *Multimedia Analysis, Processing and Communications*, tom 346, W. Lin, D. Tao, J. Kacprzyk , Z. Li, E. Izquierdo i H. Wang , Redaktorzy, Berlin, Springer-Verlag, 2011, pp. 239-262.

[14] T. Kato, "Database architecture for content-based image retrieval," in *Proceedings of SPIE Image Storage and Retrieval System*, San Jose, CA, USA, 1992, April,.

[15] V. N. Gudivada and V. V. Raghavan , "Content-Based Image Retrieval Systems," *IEEE Computer,* vol. 28, no. 9, pp. 18-22, Sep. 1995.

[16] M. Flickner, H. Sawhney, W. Niblack , J. Ashley, Q. Huang, B. Dom, M. Gorkani , J. Hafner, D. Lee, D. Petkovic, D. Steele and P. Yanker , "Query by Image and Video Content: The QBIC System," *IEEE Computer,* vol. 28, no. 9, pp. 23-32, September 1995.

[17] V. E. Ogle and M. Stonebraker, "CHABOT: Retrieval from a Relational Database of Images," *IEEE Computer,* vol. 28, no. 9, pp. 40-48, September 1995.

[18] R. Mehrotra and J. E. Gary, "Similar-Shape Retrieval in Shape Data Management," *IEEE Computer,* vol. 28, no. 9, pp. 57-62, Sep. 1995.

[19] M. Nakazato i T. S. Huang, „3D MARS: Immersive Virtual Reality for Content-Based Image Retrieval," w *IEEE International Conference on Multimedia and Expo*, Tokyo, August 22-25, 2001.

[20] S. Saurin, „Saurin Shah Portfolio," 2014. [Online]. Available: http://www.shahsaurin.com/projects_demo/threejs-webgl/ .

[21] G. Chang, M. J. Healey , J. A. M. McHugh i J. T. L. Wang, Mining the World Wide Web: An Information Search Approach., Norwell: Kluwer Academic, 2001.

[22] T. Jaworska, „Object extraction as a basic process for content-based image retrieval (CBIR) system.," *Opto-Electronics Review,* tom 15, nr 4, pp. 184-195, Dec. 2007.

[23] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Internationa Journal of Computer Vision,* vol. 60, no. 2, pp. 91-110, 2004.

[24] D. G. Lowe, "Object Recognition from local scale-invariant features," in *International Conferences on Computer Vision*, Corfu, Greece, 1999.

[25] . C. Leininger , „Fusion d'images : des outils au service des neurochirurgiens," June 2006. [Online]. Available: https://interstices.info/jcms/c_16870/fusion-d-images-des-outils-au-service-des-neurochirurgiens.

[26] M. R. Azimi-Sadjadi, J. Salazar and S. Srinivasan, "An Adaptable Image Retrieval System With Relevance Feedback Using Kernel Machines and Selective Sampling," *IEEE Transactions on Image Processing,* vol. 18, no. 7, p. 1645 1659, 2009.

[27] J. Urban, J. M. Jose and C. J. van Rijsbergen, "An adaptive technique for content-based image retrieval," *Multimedial Tools Applied,* no. 31, pp. 1-28, July 2006.

[28] X. S. Zhou and T. S. Huang, "Relevance Feedback in Image Retrieval: A Comprehensive Review," *ACM Multimedia Systems,* vol. 8, no. 6, pp. 536-544, 2003.

[29] L. Zhang, L. Wang and W. Lin, "Conjunctive patches subspace learning with side information for collaborative image retrieval," *IEEE Transactions on Image Processing,* vol. 21, no. 8, pp. 3707-3720, 2012.

[30] M. M. Rahman, S. K. Antani and G. R. Thoma, "A query expansion framework in image retrieval domain based on local and global analysis," *Information Processing and Management,* vol. 47, pp. 676-691, 2011.

[31] L. Zhang, L. Wang and W. Lin, "Generalized biased discriminant analysis for content-based image retrieval," *IEEE Transactions on System, Man, Cybernetics, Part B - Cybernetics,* vol. 42, no. 1, pp. 282-290, 2012.

[32] L. Zhang, L. Wang and W. Lin, "Semi-supervised biased maximum margin analysis for interactive image retrieval," *IEEE Transactions on Image Processing,* vol. 21, no. 4, pp. 2294-2308, 2012.

[33] L. Wang, W. Lin and L. Zhang, "Geometric Optimum Experimental Design for Collaborative Image Retrieval," *IEEE Transactions on Circuits and System for Video Technology,* vol. 24, pp. 346-359, 2014.

[34] F. Long, H. Zhang and D. D. Feng, "Fundamentals of content-based image retrieval," in *Multimedia Information Retrieval and Management Technological Fundamentals and Applications.*, New York, Sprainger-Verlag, 2003, pp. 1-26.

[35] S. Gould and X. He, "Scene Understanding by labellilng Pixels," *Communications of the ACM,* vol. 57, no. 11, pp. 68-77, November 2014.

[36] J. Yao, S. Fidler and R. Urtasun, "Describing the Scene as a Whole: Joint Object Detection, Scene Classification and Semantic Segmentation," in *The 26th IEEE Conference on Computer Vision and Pattern Recognition*, Providence, Rhode Island, 2012.

[37] L.-J. Li, H. Su, . E. P. Xing and L. Fei-Fei, "Object Bank: A High-Level Image Representation for Scene Classification and Semantic Feature Sparsification," in *24th Annual Conference on Neural Information Processing Systems* , Vancouver, Canada, 2010.

[38] D. M. Wells, A. P. French, A. Naeem, O. Ishaq and R. Traini, "Recovering the dynamics of root growth and development using novel image acquisition and analysis methods," *Phisiological Transactions of The Royal Society B,* no. 367, p. 1517–1524, 2012.

[39] C. Steger, M. Ulrich and C. Wiedemann, Machine Vision Algorithms and Applications, Weinheim: Wiley-VCH, 2008.

[40] J. Wan, D. Wang, S. C. Hoi, P. Wu, J. Zhu, Y. Zhang and J. Li, "Deep Learning for Content-Based Image Retrieval: A Comprehensive Study," in *Proceedings of the ACM International Conference on Multimedia*, Orlando, Florida, 3-7 Nov. 2014.

[41] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, no. 12, pp. 1349-1380, Dec 2000.

[42] T. Jaworska, „A Search-Engine Concept Based on Multi-Feature Vectors and Spatial Relationship," w *Flexible Query Answering Systems*, tom 7022, H. Christiansen, G. De Tré, A. Yazici, S. Zadrożny i H. L. Larsen, Redaktorzy, Ghent, Springer, 2011, pp. 137-148.

[43] C.-R. Su, J.-J. Chen and K.-L. Chang, "Content-Based Image Retrieval on Reconfigurable Peer-to-Peer Networks," in *International Symposium on Biometrics and Security Technologies*, 2013.

[44] "List of CBIR engines," 2015. [Online]. Available: http://en.wikipedia.org/wiki/List_of_CBIR_engines.

[45] L.-J. Li, C. Wang, Y. Lim, D. M. Blei and L. Fei-Fei, "Building and Using a Semantivisual Image Hierarchy," in *IEEE Conference on Computer Vision and Pattern Recognition* , June, 2010.

[46] F. Wu, Advances in Visual Data Compression and Communication: Meeting the Requirements of New Applications, CRC Press, 2014, p. 513.

[47] J. G. Kolo, K. P. Seng, L.-M. Ang and S. R. S. Prabaharan, "Data Compression Algorithms for Visual Information," in *Informatics Engineering and Information Science*, vol. 253, A. A. Manaf, . S. Sahibuddin, . R. Ahmad , . S. M. Daud and . E. El-Qawasmeh , Eds., Berlin, Springer-Verlag, 2011, pp. 484-497.

[48] N. Sharda, "Multimedia Transmission ober Wireless Sensor Networks," in *Visual Information Processing in Wireless Sensor Networks: Technology, Trends and Applications*, L. Ang, Ed., 2011.

[49] T. Jaworska, „Object extraction as a basic process for content-based image retrieval (CBIR) system.," *Opto-Electronics Review,* tom 15, nr 4, pp. 184-195, December 2007.

[50] T. Jaworska, "Database as a Crucial Element for CBIR Systems," in *Proceedings of the 2nd International Symposium on Test Automation and Instrumentation*, Beijing, China, 16-20 Nov., 2008.

[51] T. Jaworska, "Application of Fuzzy Rule-Based Classifier to CBIR in comparison with other classifiers," in *11th International Conference on Fuzzy Systems and Knowledge Discovery*, Xiamen, China, 19-21.08.2014.

[52] T. Jaworska, "Spatial representation of object location for image matching in CBIR," in *New Research in Multimedia and Internet Systems*, vol. 314, A. Zgrzywa, K. Choroś and A. Siemiński, Eds., Wrocław, Springer, 2014, pp. 25-34.

[53] T. Jaworska, "Query techniques for CBIR," in *Flexible Query Answering Systems*, vol. 400, T. Andreasen, H. Christiansen, J. Kacprzyk, H. Larsen, G. Pasi, O. Pivert, G. De Tre, M. A. Vila, A. Yazici and S. Zadrożny, Eds., Cracow, Springer, 2015, pp. 403-416.

[54] Y.-J. Zhang, Y. Gao and Y. Luo, "Object-Based Techniques for Image Retrieval," in *Multimedia Systems and Content-Based Image Retrieval*, S. Deb, Ed., Hershey, London, IDEA Group Publishing, 2004, pp. 156-181.

[55] T. Tuytelaars and K. Mikolajczyk, "Local Invariant Feature Detectors: A Survey," *Computer Graphics and Vision*, vol. 3, no. 3, p. 177–280, 2007.

[56] W. Niblack, M. Flickner, D. Petkovic, P. Yanker, R. Barber, W. Equitz, E. Glasman, C. Faloutsos and G. Taubin, "The QBIC Project: Querying Images by Content Using Colour, Texture and Shape," *SPIE*, vol. 1908, pp. 173-187, 1993.

[57] G. Pass and R. Zabith, "Histogram refinement for content-based image retrieval," *IEEE Workshop on Applications of Computer Vision*, pp. 96-102, 1996.

[58] M. Pietikäinen, Ed., Texture Analysis in Machine Vision, vol. 40, World Scientific, 2000.

[59] N. Sebe and M. S. Lew, "Texture Features for Content-Based Retrieval," in *Principles of Visual Information Retrieval*, M. S. Lew, Ed., London , Springer Science & Business Media, 2013, pp. 50-81.

[60] M. Tuceryan and A. K. Jain, "Texture Analysis," in *The Handbook of Pattern Recognition and Computer Vision*, 2 ed., C. H. Chen, L. F. Pau and P. S. P. Wang, Eds., World Scientific Publishing Co., 1998, pp. 207-248.

[61] S. W. Zucker, "Toward a Model of Texture," *Computer Graphics and Image Processing*, vol. 5, pp. 190-202, 1976.

[62] N. Ahuja, "Dot Pattern Processing Using Voronoi Neighborhoods," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, no. 4, pp. 336-343, May 1982.

[63] R. M. Haralick, "Statistical and Structural Approaches to Texture," *Proceedings of the IEEE*, vol. 67, pp. 786-804, 1979.

[64] . M. Pietikäinen, T. Ojala and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distributions.," *Pattern Recognition*, vol. 29, no. 1, pp. 51-59, January 1996.

[65] T. Ojala, M. Pietikäinen and T. Mäenpää, "Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns.," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.

[66] M. Pietikäinen, A. Hadid, G. Zhao and T. Ahonen, Computer Vision Using Local Binary Patterns, vol. 40 in Computational Imaging and Vision, Springer Science & Business Media, 2007.

[67] H. Tamura, S. Mori i T. Yamawaki, „Texture features corresponding to visual perception," *IEEE Transactions On Systems, Man and Cybernetics*, tom 8, pp. 460-473, 1978.

[68] R. Sriram , J. M. Francos and W. A. Pearlman, "Texture coding using a Wold decomposition model.," *IEEE Transactions of Image Processing* , vol. 5, no. 9, pp. 1382-1386, 1996.

[69] G. L. Gimel'farb and A. K. Jain, "On retrieving textured images from an image data base.," *Pattern Recognition*, vol. 29, no. 9, pp. 1461-1483, 1996.

[70] A. P. Pentland, "Fractal-based description of natural scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 661-674., June 1984.

[71] B. B. Mandelbrot, Fractal Geometry of Nature, New York: Freeman, 1982.

[72] H. E. Hurst, "Long-term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers,* vol. 116, no. 1, pp. 770-799, 1951.

[73] S. Ezekiel and J. A. Cross, "Fractal-based Texture Analysis," in *APCC/OECC'99, Joint Conference of 5th Asia-Pacific Conference on Communications (APCC) and 4th Opto-Electronics and Communications Conference (OECC),* 1999.

[74] J. Millard, P. Augat, T. M. Link, M. Kothari, D. C. Newitt, H. K. Genant, and S. Majumdar, "Power Spectral Analysis of Vertebral Trabecular Bone Structure from Radiographs: Orientation Dependence and Correlation with Bone Mineral Density and Mechanical Properties," *Calcified Tissue International,* vol. 63, pp. 482-489, 1998.

[75] S. Selvarajah and S. R. Kodituwakku, "Analysis and Comparison of Texture Features for Content Based Image Retrieval," *International Journal of Latest Trends in Computing,* vol. 2, no. 1, pp. 108-113, March 2011.

[76] G. M. Haley and B. S. Manjunath, "Rotation-Invariant Texture Classification Using a Complete Space-Frequency Model," *IEEE Transactions on Image Processing,* vol. 8, no. 2, Feb. 1999.

[77] D. Gabor, "Theory of communication," *Journal of the Institution of Electrical Engineers ,* pp. 445 - 457, 1946.

[78] T. S. Lee, "Image Representation Using 2D Gabor Wavelets," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,* vol. 18, no. 10, October 1996.

[79] T. Jaworska, "Point-to-point correspondence into stereo pair of images," Silesian University of Technology, Gliwice, Poland, 2001.

[80] N. Sebe and M. S. Lew, "Wavelet Based Texture Classification," in *Proceedings. 15th International Conference on Pattern Recognition,* 2000.

[81] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE TRANSACTIONS ON COMMUNICATIONS,* Vols. COM-3l, no. 4, pp. 532-540, April 1983.

[82] J. L. Crowley, "A representation for visual information," 1987.

[83] I. Daubechies, Ten lectures on wavelets, Philadephia: Society for Industrial and Applied Mathematics, 1992.

[84] S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 11, no. 7, pp. 674-693, 1989.

[85] S. Mallat, "Multiresolution Approximation and Wavelet Orthonormal Bases of L2(R)," *Transactions American Mathematical Society,* vol. 315, no. 1, pp. 69-87, 1989.

[86] Y. Meyer, Les ondelettes. Algorithmes et applications, Paris: Armand Colin, 1992.

[87] P. Wojtaszczyk, Wavelet Theory (in Polish), Warsaw: PWN, 2000.

[88] S. Mallat, A wavelet tour of signal processing, Academic Press, 1998.

[89] M. Faizal, A. Fauzi and P. H. Lewis, "Automatic texture segmentation for content-based image retrieval application," *Pattern Analysis and Applications,* vol. 9, p. 307–323, 2006.

[90] R. A. Kirsch , "Computer determination of the constituent structure of biological images," *Computers and Biomedical Research,* vol. 4, no. 3, p. 315–328, July 1971.

[91] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 13, no. 6, p. 583–598, 1991.

[92] O. Basir, H. Zhu and F. Karray, "Fuzzy Based Image Segmentation," in *Fuzzy Filters foe Image processing,* vol. 122, Berlin, Springer, 2003, pp. 101-128.

[93] H. M. Sobel, Multivariate Observations, Wiley, 1984.

[94] J. M. S. Prewitt, "Object Enhancement and Extraction," in *Picture Processing and Psychopictorics,*, B. S. B. S. Lipkin and A. Rosenfeld, Eds., NY, Academic Press, 1970.

[95] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vols. PAMI-8, no. 6, pp. 679-698, 1986.

[96] C. Xu and J. L. Prince, "Snakes, Shapes, and Gradient Vector Flow," *IEEE TRANSACTIONS ON IMAGE PROCESSING,* vol. 7, no. 3, pp. 359-369, March 1998.

[97] R. O. Duda and P. E. Hart, "Use of the HOUGH Transformation to Detect Lines and Curves in Pictires," 1971.

[98] Q. Zhu, . L. Wang, Y. Wu and J. Shi,, "Contour Context Selection for Object Detection: A Set-to-Set Contour Matching Approach,," in *The 10th European Conference on Computer Vision (ECCV)*, Marseille, France, 2008.

[99] D. Zhang and G. Lu, "Review of shape representation andd escription techniques," *Pattern Recognition,* vol. 37, p. 1 – 19, 2004.

[100] S. Abbasi, F. Mokhtarian and J. Kittler, "Curvature scale space image in shape similarity retrieval," *Multimedia Systems,* no. 7, p. 467–476, 1999.

[101] C.-J. Sze, H.-R. Tyan, H.-Y. M. Liao, C.-S. Lu and S.-K. Huang, "Shape-based Retrieval on a Fish Database of Taiwan," *Tamkang Journal of Science and Engineering,* vol. 2, no. 3, pp. 63-173 , 1999.

[102] T. B. Sebasian and B. B. Kimia, "Curves vs Skeltons in Object Recognition," in *Proceedings of International Conference on Image Processing,* Thessaloniki, 7-10 Oct. 2001 .

[103] L. Kotoulas i I. Andreadis, „Image analysis using moments," w *Proceedings of 5th International Conference on Technology and Automation*, Thessaloniki, Greece, 2005.

[104] M. R. Teague, "Image analysis via the general theory of moments," *Journal of the Optical Society of America,* vol. 70, no. 8, pp. 920-930, 1980.

[105] R. Arandjelović and A. Zisserman, "Three things everyone should know to improve object retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, 2012.

[106] K. Mikolajczyk and C. Schmid, "Scale & Affine Invariant Interest Point Detectors," *International Journal of Computer Vision,* pp. 63-86, 2004.

[107] F. Perronnin, J. Sanchez and T. Mensink, "Improving the Fisher Kernel for Large-Scale Image Classification," in *European Conference on Computer Vision, Lecture Notes in Computer Science*, Heraclion, Greece, Sep, 2010.

[108] F. Perronnin and C. Dance, "Fisher Kernels on Visual Vocabularies for Image Categorization," in *Proceeding Computer Vision and Pattern Recognition*, 2007.

[109] J. Krapac and S. Śegvić, "Weakly Supervised Object Localization with Large Fisher Vectors," in *Proceedings of the 10th International Conference on Computer Vision Theory and Applications*, Berlin, 11-14 Mar, 2015.

[110] H. Jegou, M. Douze, C. Schmid and P. Perez, "Aggregating local descriptors into a compact image representation," in *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, 13-18 June, 2010.

[111] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *IEEE International Conference on Computer Vision*, 2005.

[112] E. Rosten i T. Drummond, „Machine learning for high-speed corner detection," w *European Conference on Computer Vision*, 2006.

[113] E. Rublee, V. Rabaud , K. Konolige and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *IEEE International Conference on Computer Vision (ICCV)*, Barcelona, Spane, 6-12, Nov, 2011.

[114] M. Brown, R. Szeliski i S. Winder, „Multi-image matching using Multi-Scale Oriented Patches," *Computer Vision and Pattern Recognition,* nr 2, pp. 510-517, 2005.

[115] The Moving Picture Experts Group, "MPEG," [Online]. Available: http://mpeg.chiariglione.org/. [Accessed 2015].

[116] MPEG, "MPEG standards - Full list of standards developed or under development," 20 April 2010. . [Online]. Available: http://mpeg.chiariglione.org/standards.htm.

[117] I. JTC1/SC29/WG11, "CODING OF MOVING PICTURES AND AUDIO MPEG-7". Palma de Mallorca, Spain Patent N6828, Oct. 2004.

[118] M. J. Swain and D. H. Ballard, "Color Indexing," *International Journal of Computer Vision,* vol. 7, no. 1, pp. 11-32, 1991.

[119] V. Castelli i L. D. Bergman, Redaktorzy, Image Databases: Search and Retrieval of Digital Imagery, New York: Wiley, 2002.

[120] J.-J. Chen, C.-R. Su, W. E. L. Grimson, J.-L. Liu and D.-H. Shiue, "Object Segmentation of Database Images by Dual Multiscale Morphological Reconstructions and Retrieval Applications," *IEEE Transactions on Image Processing,* vol. 21, no. 2, pp. 828-843, Feb. 2012.

[121] P. Melin and O. Castillo, Hybrid Intelligent Systems for Pattern Recognition Using Soft Computing. An Evolutionary Approach for Neural Networks and Fuzzy Systems., Berlin: Springer, 2005, p. 272.

[122] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms., New York: Plenum Press, 1981, p. 272.

[123] Y. Cheng , "Mean Shift Mode Seeking, and Clustering," *IEEE TRANSACTIONS on PATTERN ANALYSIS and Machine Intelligence,* vol. 17, no. 8, Aug, 1995.

[124] G. Seber, Multivariate Observations, New York: Wiley, 1984, p. 686.

[125] H. Späth, Cluster analysis algorithms for data reduction and classification of objects, vol. 4, Pensilvania University: E. Horwood, 1980, p. 226.

[126] M. Acharyya and M. K. Kundu, "An adaptive approach to unsupervised texture segmentation using M-Band wavelet transform," *Signal Processing,* no. 81, pp. 1337-1356, 2001.

[127] L. J. Latecki and R. Lakamper, "Application of planar shape comparison to object retrieval in image databases," *Pattern Recognition,* no. 35, pp. 15-29, 2002.

[128] W.-B. Goh and K.-Y. Chan, "A Shape Descriptor for Shapes with Boundary Noise and Texture," in *British Machine Vision Conference,* Norwich, 24 June, 2003.

[129] C. Xu and J. Liu, "2D Shape Matching by Contour Flexibility," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,* vol. 31, no. 1, Jan. 2009.

[130] J. Mutch and D. G. Lowe, "Object class recognition and localization using sparse features with limited receptive fields," *International Journal of Computer Vision (IJCV),* vol. 80, no. 1, pp. 45-57, Oct 2008.

[131] T. Serre, L. Wolf and T. Poggio, "Object Recognition with Features Inspired by Visual Cortex," in *Proceedings on Computer Vision and Pattern Recognition,* Los Alamos, 2005.

[132] Y. Li and L. G. Shapiro, "Object Recognition for Content-Based Image Retrieval," Dagstuhl Seminar, Leibniz, Austria, 2002.

[133] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener and C. Roux, "Fast Wavelet-Based Image Characterization for Highly Adaptive Image Retrieval," *IEEE Transactions on Image Processing,* vol. 21, no. 4, pp. 1613-1623, April 2012.

[134] B. V. Dasarathy, Ed., Nearest neighbor (NN) norms : NN pattern classification techniques, 6th ed., Los Alamitos, Callifornia: IEEE Computer Society Press, 1991.

[135] C. Cortes and V. Vapnik , "Support-Vector Networks," *Machine Learning,* vol. 20, p. 273–297, 1995.

[136] I. Rish, "An empirical study of the Naïve Bayes classifier," in *Proceedings of the IJCAI-2001 Workshop on Empirical Methods in AI,* Brussels, 2001.

[137] G. P. Zhang, "Neural Networks for Classification: A Survey," *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and reviews,* vol. 30, no. 4, pp. 451-462, Nov 2000.

[138] J. M. Ali, "Content-Based Image Classification and Retrieval: A Rule-Based System Using Rough Sets Framework," in *Artificial Intelligence for Maximizing Content Based Image Retrieval,* Z. Ma, Ed., NY, Springer, 2009, pp. 68-82.

[139] T. Jaworska, "Towards Fuzzy Classificaton in CBIR," in *Information Systems Architecture and Technology.* , Vols. Knowledge Based Approach to the Design, Control and Decision Support, J. Świątek, L. Borzemski, A. Grzech and Z. Wilimowska, Eds., Wrocław, Oficyna Wydawnicza Politechniki Wrocławskiej, 2013, pp. 53-62.

[140] U. M. Fayyad and K. B. Irani, "The attribute selection problem in decision tree generation," in *the 10th National Conference on Artificial Intelligence, AAAI,* 1992.

[141] L. Breiman , J. Friedman , C. J. Stone and R. A. Olshen, Classification and Regression Trees, New York: Chapman and Hall, 1984, p. 368.

[142] J. R. Quinlan, "Induction of Decision Trees," *Machine Learning,* vol. 1, pp. 81-106, 1986.

[143] J. R. Quinlan, C4.5: Programs for Machine Learning, San Mateo: Morgan Kaufmann Publishers, 1993.

[144] H. Schulz, B. Waldvogel, R. Sheikh and S. Behnke, "CURFIL: Random Forests for Image Labeling on GPU," in *Proceedings of the 10th International Conference on Computer Vision Theory and Applications,* Berlin, 11-14 Mar, 2015.

[145] J. Ylioinas, J. Kannala, A. Hadid and . M. Pietikainen, "Learning Local Image Descriptors Using Binary Decision Trees," in *Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV 2014),*, Steamboat Springs, CO, USA,, 2014.

[146] B. Bouchon-Meunier and C. Marsala, "Fuzzy decision tree and databases," in *Flexible Query Answering Systems,* T. Andreasen, H. Christiansen and H. L. Larsen, Eds., Kluwer Academic Publisher, 1997, pp. 277-288.

[147] J. D. M. Rennie, L. Shih, J. Teevan and D. R. Karge, "Tackling the Poor Assumptions of Naive Bayes Text Classifiers," in *Proceedings of the 20th International Conference on Machine Learning,* Washington, DC, USA, 2003.

[148] N. M. Murty and S. V. Devi, Pattern Recognition: An Algorithmic Approach, vol. z serii Undergraduate Topics in Computer Science, Springer Science & Business Media, 2011, p. 263.

[149] L. Wang, Ed., Support Vector Machines: Theory and Applications, Berlin: Springer, 2005, p. 450.

[150] H. Ishibuchi and Y. Nojima, "Toward Quantitative Definition of Explanation Ability of Fuzzy Rule-Based Classifiers," in *IEEE International Conference on Fuzzy Systems,* Taipei, Taiwan, June 27-39, 2011.

[151] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems,* vol. 13, no. 4, pp. 428-435, 2005.

[152] K. Nozaki, H. Ishibuchi and H. Tanaka , "Adaptive fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems,* vol. 13, no. 4, pp. 238-250, 1996.

[153] H. Ishibuchi and Y. Nojima, "Toward Quantitative Definition of Explanation Ability of Fuzzy Rule-Based Classifiers," in *IEEE International Conference on Fuzzy Systems,* Taipei, Taiwan, June 27-39, 2011.

[154] T. Jaworska, "Application of Fuzzy Rule-Based Classifier to CBIR in comparison with other classifiers," in *11th International Conference on Fuzzy Systems and Knowledge Discovery*, Xiamen, China, 2014.

[155] S. K. Candan and W.-S. Li, "On Similarity Measures for Multimedia Database Applications," *Knowledge and Information Systems,* vol. 3, pp. 30-51, 2001.

[156] A. Hamilton-Wright and D. W. Stashuk, "Constructing a Fuzzy Rule Based Classification System Using Pattern Discovery," in *Annual Meeting of the North American Fuzzy Information Processing Society*, 2005.

[157] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *Nature,* vol. 521, pp. 436-444, 28 May 2015.

[158] C. Olah, „Conv Nets: A Modular Perspective," blog, July 2014. [Online]. Available: http://colah.github.io/posts/2014-07-Conv-Nets-Modular/.

[159] A. Krizhevsky, I. Sutskeve and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, 2012.

[160] MathWorks Inc., "Deep learning with MATLAB," 2016. [Online]. Available: https://www.mathworks.com/discovery/deep-learning.html.

[161] C.-C. Chang and T.-C. Wu, "An exact match retrieval scheme based upon principal component analysis," *Pattern Recognition Letters,* vol. 16, pp. 465-470, 1995.

[162] D. S. Guru and P. Punitha, "An invariant scheme for exact match retrieval of symbolic images based upon principal component analysis," *Pattern Recognition Letters* , vol. 25, p. 73–86, 2004.

[163] S. Rolewicz, Functional Analysis and Control Theory: Linear Systems, vol. Series: Mathematics and its applications, Warsaw: PWN-Polish Scientific Publishers , 1987.

[164] J. Z. Wang, J. Li and G. Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture LIbraries," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,,* vol. 23, no. 9, pp. 947-963, Sep. 2001.

[165] C. Mallows, "A Note on Asymptotic Joint Normality," *The Annals of Mathematical Statistics,* vol. 43, no. 2, pp. 508-515., 1972.

[166] D. Zhou, J. Li and H. Zha, "A new Mallows distance based metric for comparing clusterings," in *Proceedings of the 22nd International Conference on Machine Learning,*, Bonn,m Germany, Aug. 2005.

[167] E. Pękalska and R. P. Duin, The Dissimilarity Representation for Pattern Recognition. Foundations and Applications., 1 ed., Vols. Series in Machine Perception and Artificial Intelligence - Vol. 64, New Jersey, London: World Scientific, 2005, p. 607.

[168] B. Ko and H. Byun, "Integrated Region-Based Image Retrieval Using Region's Spatial Relationships," in *Proceedings of 16th International Conference on Pattern Recognition*, 11-15 Aug. 2002.

[169] C. Beecks, M. S. Uysal and T. Seidl, "A Comparative Study of Similarity Measures for Content-Based Multimedia Retrieval," in *Multimedia and Expo (ICME)*, Suntec City, 19-23 July, 2010.

[170] T. Jaworska, "A Search-Engine Concept Based on Multi-Feature Vectors and Spatial Relationship," in *Flexible Query Answering Systems*, vol. 7022, H. Christiansen, G. De Tré, A. Yazici, S. Zadrożny and H. L. Larsen, Eds., Ghent, Springer, 2011, pp. 137-148.

[171] T. Jaworska, "An Asymmetric Approach to Signature Matching," in *Multimedia and Network Information Systems*, vol. 506, A. Zgrzywa, K. Choraś and A. Siemiński, Eds., Wrocław, Springer, 2016, pp. 27-37.

[172] G. Wu, E. Y. Chang and N. Panda, "Formulating context-dependent similarity functions," in *The 13th annual ACM international conference on Multimedia*, Singapore , Nov., 2005.

[173] A. Natsev and J. R. Smith, "A study of image retrieval by anchoring," in *IEEE International Conference on Multimedia and Expo*, Lausanne, Switzerland , Aug. 2002.

[174] C.-T. Nguyen, X. Wang, J. Liu and Z.-H. Zhou, "Labeling Complicated Objects: Multi-View Multi-Instance Multi-Label Learning," in *28th AAAI Conference on Artificial Intelligence*, Hilton Québec Canada, June, 2014.

[175] H. Mueller, W. Mueller, S. Marchand-Maillet and T. Pun, "A Framework for Benchmarking in CBIR," *Multimedia Tools and Applications,* no. 21, pp. 55-73, 2003.

[176] D. A. Narasimhalu , M. S. Kankanhalli and J. Wu, "Benchmarking Multimedia Databases," *Multimedia Tools and Applications,* vol. 4, no. 3, p. 333–356, May 1997.

[177] J. R. Smith, "Image retrieval evaluation," in *IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL '98)*, Santa Barbara, 1998.

[178] A. Dimai , "Assessment of effectiveness of content-based image retrieval systems," in *3rd International Conference on Visual Information Systems (VISUAL '99)*, Amsterdam, The Netherlands, 1999.

[179] E. L. van den Broek, T. Kok, T. E. Schouten and L. G. Vuurpijl, "Human-Centered Content-Based Image Retrieval," in *Proceedings of XIII Conference on Human Vision and Electronic Imaging*, Feb. 14, 2008.

[180] M. Everingham, A. S. Eslami, L. Van Gool, C. K. I. Williams, J. Winn and A. Zisserman, "The PASCAL Visual Object Classes Challenge: A Retrospective," *International Journal of Computer Vision,* no. 111, p. 98–136, 2015.

[181] Corel comp., "The COREL Database for Content based Image Retrieval".

[182] Z. Yang and C.-C. Jay Kuo, "Learning image similarities and categories from content analysys and relebance feedback," in *Proceedings of the ACM Multimedia Workshops. Multimedia00'*, Los Angeles, CA, USA , Oct 30 - Nov 03, 2000.

[183] the Eastman Kodak Company, [Online]. Available: http://r0k.us/graphics/kodak/.

[184] D.-C. He and A. Safia, "Multiband Texture Database," 2015. [Online]. Available: http://multibandtexture.recherche.usherbrooke.ca/.

[185] D.-C. He and A. Safia, "New Brodatz-based Image Databases for Grayscale Color and Multiband Texture Analysis," *ISRN Machine Vision,* vol. Article ID 876386, pp. 1-14, 2013.

[186] N. Rasiwasia, P. J. Moreno and N. Vasconcelos, "Bridging the Gap: Query by Semantic Example," *IEEE TRANSACTIONS ON MULTIMEDIA,* vol. 9, no. 5, pp. 923-938, Aug 2007.

[187] X. Wang, S. Qiu, K. Liu i X. Tang, „Web Image Re-Ranking Using Query-Specific Semantic Signatures," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,* tom 36, nr 4, pp. 810-823, April 2014.

[188] M. Everingham, L. Van Gool, C. K. I. Williams, A. Zisserman, J. Winn, A. S. Eslami and Y. Aytar , "The PASCAL Visual Object Classes Homepage," 2015. [Online]. Available: http://host.robots.ox.ac.uk/pascal/VOC/index.html.

[189] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, USA, June, 2009.

[190] L. Fei-Fei, K. Li, O. Russakovsky, J. Krause, J. Deng and A. Berg, "ImageNet," Stanford Vision Lab, Stanford University, Princeton University, 2014. [Online]. Available: http://www.image-net.org/.

[191] G. Griffin, A. D. Holub and P. Perona, "The Caltech 256," California Institute of Technology, Los Angeles, 2006.

[192] G. Griffin, "Caltech256," 2006. [Online]. Available: http://www.vision.caltech.edu/Image_Datasets/Caltech256/.

[193] J. Philbin, O. Chum and M. a. S. J. a. Z. A. Isard, "Object Retrieval with Large Vocabularies and Fast Spatial Matching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[194] J. Philbin, R. Arandjelović and A. Zisserman, "The Oxford Buildings Dataset," Department of Engineering Science, University of Oxford, Nov 2012. [Online]. Available: http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/.

[195] J. Philbin, O. Chum and M. a. S. J. a. Z. A. Isard, "Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases," in *IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, USA, 23-28 June, 2008.

[196] J. Philbin i A. Zisserman, „The Paris Dataset," Visual Geometry Group, Department of Engineering Science, University of Oxford , 2008. [Online]. Available: http://www.robots.ox.ac.uk/~vgg/data/parisbuildings/.

[197] B. C. Becker, "PubFig83 + LFW Dataset," 2015. [Online]. Available: http://www.briancbecker.com/blog/research/pubfig83-lfw-dataset/.

[198] B. C. Becker and E. G. Ortiz, "Evaluating Open-Universe Face Identification on the Web," in *CVPR 2013, Analysis and Modeling of Faces and Gestures Workshop.*, Portland, Oregon, USA, 23-28 June, 2013.

[199] P.-S. P. Chen, "Entity-relationships model – Toward a Unified View of Data," *ACM Transactions on Database Systems,* vol. 1, no. 1, pp. 9-36, 1976.

[200] R. Barker, Entity-Relationship Modelling. Case MethodSM, London, : Addison-Wesley, 1995.

[201] R. Barker and C. Longman , Function and Process Modelling. Case MethodSM, London: Addison-Wesley Pub. Co., 1993.

[202] K. Rodden and K. R. Wood, "How Do People Manage Their Digital Photographs?," in *SIGCHI Conference on Human Factors in Computing Systems*, Ft. Lauderdale, Florida, USA., April 5–10, 2003.

[203] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGEN,* vol. 22, no. 12, pp. 1349 - 1380, Dec 2000.

[204] X. Wang, K. Liu and X. Tang, "Query-Specific Visual Semantic Spaces forWeb Image Re-ranking.," in *Computer Vision and Patern Recognition Paper*, 2011.

[205] W. Niblack, M. Flickner, D. Petkovic, P. Yanker, R. Barber, W. Equitz, E. Glasman, C. Faloutsos and G. Taubin, "The QBIC Project: Querying Images by Content Using Colour, Texture and Shape," *SPIE,* vol. 1908, pp. 173-187, 1993.

[206] B. Xiao , X. Gao, D. Tao and X. Li, "Recognition of Sketches in Photos," in *Multimedia Analysis, Processing and Communications*, vol. 346, W. Lin, D. Tao, J. Kacprzyk , Z. Li, E. Izquierdo and H. Wang , Eds., Berlin, Springer-Verlag, 2011, pp. 239-262.

[207] J.-H. Lim and J. S. Jin, "A structured learning framework for content-based image indexing and visual query," *Multimedia Systems,* vol. 10, p. 317–331, 2005.

[208] J. Assfalg, A. Del Bimbo and P. Pala, "Three-Dimensional Interfaces for Querying by Example in Content-Based Image Retrieval," *IEEE Transactions on Visualization and Computer Graphics* , vol. 8, no. 4, pp. 305-318, Oct-Dec 2002.

[209] J. Fauqueur and N. Boujemaa, "Mental image search by boolean composition of region categories," *Multimed Tools and Applications,* vol. 31, p. 95–117, 2006.

[210] T. Jaworska, "Multi-criteria object indexing and graphical user query as an aspect of content-based image retrieval system.," in *Information Systems Architecture and Technology,* L. Borzemski, A. Grzech, J. Świątek and Z. Wilimowska, Eds., Wrocław, Wrocław Technical University Publisher, 2009, pp. 103-112.

[211] . B. Moghaddam, H. Biermann and D. Marg, "Regions-of-Interest and Spatial Layout for Content-Based Image Retrieval," *Multimedia Tools and Applications,* vol. 14, no. 2, pp. 201-210, June 2001.

[212] M. M. Rahman, S. K. Antani and G. R. Thoma, "A query expansion framework in image retrieval domain based on local and global analysis," *Information Processing and Management,* vol. 47, pp. 676-691, 2011.

[213] J. Fauqueur, "Instantaneous mental image search with range queries on multiple region descriptors," Cambridge, UK, Jan, 2005.

[214] Y. Liu, D. Zhang, G. Lu and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition,* vol. 40, pp. 262-282, 2007.

[215] J. C. Cubero, N. Marín, J. M. Medina, E. Pons and A. M. Vila, "Fuzzy Object Management in an Object-Relational Framework," in *Proceedings of the 10th International Conference IPMU*, Perugia, Italy, 4-9 July, 2004.

[216] F. Berzal, J. C. Cubero, J. Kacprzyk, N. Marín, A. M. Vila and S. Zadrożny, "A General Framework for Computing with Words in Object-Oriented Programming.," in *International Journal of Uncertainty. Fuzziness and Knowledge-Based Systems.*, vol. 15 (Suppl), Singapore, World Scientific Publishing Company, 2007, pp. 111 131,.

[217] W. Plant and G. Schaefer, "Visualization and Browsing of Image Databases," in *Multimedia Analysis, Processing and Communications*, vol. 346, W. Lin, D. Tao, J. Kacprzyk , Z. Li, E. Izquierdo and H. Wang, Eds., Berlin, Springer, 2011, pp. 3-57.

[218] K. Rodden, „Evaluating similarity-based visualisations as interfaces for image browsing," University of Cambridge, Cambridge, 2002.

[219] K. Rodden, K. R. Wood, W. Basalaj and D. Sinclair, "Evaluating a Visualisation of Image Similarity as a Tool for Image Browsing," in *IEEE Symposium on Information Visualisation*, 1999.

[220] W. Basalaj, "Proximity visualisation of abstract data," University of Cambridge, Cambridge, 2001.

[221] C. Faloutsos and K. Lin, "Fast Map: A Fast Algorithms for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets," in *ACM SIGMOD international conference on Management of data* , New York, USA, May, 1995.

[222] L. F. D. Santos, R. L. Dias and M. X. Ribeiro, "Combining Diversity Queries and Visual Mining to Improve Content-Based Image Retrieval Systems: The DiVI Method," in *IEEE International Symposium on Multimedia*, Miami, Dec. 2015.

[223] A. Bursuc and T. Zaharia, "ARTEMIS@ MediaEval 2013: A Content-Based Image Clustering Method for Public Image Repositories," *ACM Multimedia,* pp. 18-19, Oct. 2013.

[224] C. Chen, G. Gagaudakis and P. Rosin, "Similarity-Based Image Browsing," in *Proceedings of the 16th IFIP World Computer Congress, International Conference on Intelligent Information Processing*, Beijing, China, 2000.

[225] T. Kohonen, "The Self_Organizing Map," *Proceedings of IEEE,* vol. 78, no. 9, pp. 1464-1480, Sep. 1990.

[226] A. Csillaghy , H. Hinterberger and A. B. Benz, "Content-Based Image Retrieval in Astronomy," *Information Retrieval Journal,* vol. 3, no. 3, pp. 229-241, 2000.

[227] Y. Rui and T. S. Huang, "Relevance Feedback Techniques in Image Retrieval," in *Principal of Visual Information Retrieval,* M. S. Lew, Ed., London, Springer, 2001, pp. 219-258.

[228] V. Mezaris, I. Kompatsiaris and M. G. Strintzis, "An ontology approach to object-based image retrieval," in *Proceedings of International Conference on Image Processing ICIP 2003.*, 2003.

[229] A. D. Gudewar and L. R. Ragha, "Ontology to Improve CBIR System," *International Journal of Computer Applications,* vol. 52, no. 21, pp. 23-30, 2012.

[230] C. Doulaverakis, E. Nidelkou, A. Gounaris and Y. Kompatsiaris, "A Hybrid Ontology and Content-Based Search Engine For Multimedia Retrieval," in *Workshop Proceedings in Advances in Databases and Information Systems ADBIS '2006,* Thessaloniki, 2006.

[231] O. Allani, N. Mellouli, H. B. Zghal, H. Akdag and H. B. Ghzala, "A Relevant Visual Feature Selection Approach for Image Retrieval," in *Proceedings of the 10th International Conference on Computer Vision Theory and Applications,* Berline, 11-14 Mar, 2015.

[232] O. Russakovsky and L. Fei-Fei, "Attribute Learning in Large-scale Datasets," in *Proceedings of the 12th European Conference of Computer Vision (ECCV), 1st International Workshop on Parts and Attributes.,* Crete, Greece, 2010.

[233] T. Hofmann, "Probabilistic latent semantic analysis," in *Proceedings of the15th Conference on Uncertainty in Artificial Intelligence,* Stockholm, 1999.

[234] D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research,* vol. 3, pp. 993-1022, 2003.

[235] L. Fei-Fei and P. Perona, "A Bayesian Heirarcical Model for Learning Natural Scene Categories," in *Computer Vision & Pattern Recognition CVPR,* 2005.

[236] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman and W. T. Freeman, "Discovering objects and their location in images," in *Proceedings of Internationa Conference of Computer Vision,* Beijing, 2005.

[237] J. Bautista-Ballester, J. Verges-Llahi and D. Puig, "Using Action Objects Contextual Information for a Multichannel SVM in an Action Recognition Approach based on Bag of VisualWords," in *Proceedings of the 10th International Conference on Computer Vision Theory and Applications,* Berlin, 11-14 Mar, 2015.

[238] T. Kinnunen, J.-K. Kamarainen, L. Lensu and H. Kälviäinen, "Unsupervised object discovery via self-organisation," *Pattern Recognition Letters,* no. 33, p. 2102–2112, Aug 2012.

[239] J. Urban, J. M. Jose and C. J. van Rijsbergen, "An adaptive technique for content-based image retrieval," *Multimedial Tools Applied,* no. 31, pp. 1-28, July 2006.

[240] L. Zhang, L. Wang and W. Lin, "Generalized biased discriminant analysis for content-based image retrieval," *IEEE Transactions on System, Man, Cybernetics, Part B - Cybernetics,* vol. 42, no. 1, pp. 282-290, 2012.

[241] L. Zhang, L. Wang and W. Lin, "Semi-supervised biased maximum margin analysis for interactive image retrieval," *IEEE Transactions on Image Processing,* vol. 21, no. 4, pp. 2294-2308, 2012.

[242] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science,* vol. 290, no. 5500, pp. 2323-2326, Dec. 2000.

[243] S.-F. Chang, W. Chen and H. Sundaram, "Semantic Visual Templates: Linking Visual Features to Semantics," in *International Conference on Image Processing, 1998. ICIP 98.,* Chicago, 1998.

[244] Y. Zhuang, X. Liu and Y. Pan, "Apply Semantic Template to Support Content-based Image Retrieval," in *the Proceeding of IS&T and SPIE Storage and Retrieval for Media Databases 2000,* San Jose, California, USA, Jan, 2000.

[245] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross and K. Miller, "Introduction to WordNet: An On-line Lexical Database," *Communications of the ACM,* vol. 38, no. 11, pp. 39-41 , Nov. 1995.

[246] M. Mucha and P. Sankowski, "Maximum Matchings via Gaussian Elimination," in *Proceedings of the 45th Annual Symposium on Foundations of Computer Science (FOCS'04),* 2004.

[247] Z. Wang , A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image Qualifty Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing,* vol. 13, no. 4, p. 600–612, April 2004.

[248] E. Candes, L. Demanet, D. Donoho and L. Ying, "Fast Discrete Curvelet Transforms," 2006.

[249] I. Aizenberg, N. N. Aizenberg and J. P. Vandewalle, Multi-Valued and Universal Binary Neurons, Springer US, Springer Science+Business Media Dordrecht, 2000, p. 276.

[250] T. Yamashita, T. Watasue, Y. Yamauchi and H. Fujiyoshi, "Improving Quality of Training Samples Through Exhaustless Generation and Effective Selection for Deep Convolutional Neural Networks," in *Proceedings of the 10th International Conference on Computer Vision Theory and Applications*, Berlin, 11-14 Mar, 2015.

[251] F. Jurišić, I. Filković and Z. Kalafatić, "Evaluating the Effects of Convolutional Neural Network Committees," in *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016)*, Rome, Italy, 27-29 Feb, 2016.

[252] H. H. Aghdam, E. J. Heravi and D. Puig, "Analyzing the Stability of Convolutional Neural Networks against Image Degradation," in *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016)*, Rome, Italy, 27-29 Feb, 2016.

[253] S. Srinivasulu and P. Sakthivel , "Extracting Spatial Semantics in Association Rules for Weather Forecasting Image," in *Trendz in Information Sciences & Computing(TISC2010)*, Chennai, 17-19 Dec. 2010.

[254] A. Moumtzidou, V. Epitropou, S. Vrochidis, K. Karatzas, S. Voth, A. Bassoukos, J. Moßgraber, A. Karppinen, J. Kukkone and I. Kompatsiaris, "A model for environmental data extraction from multimedia and its evaluation against various chemical weather forecasting datasets.," *Ecological Informatics,* no. 23, p. 69–82, Sep. 2014.

[255] K. Choroś, "False and Miss Detections in Temporal Segmentation of TV Sports News Videos - Causes and Remedies," in *New Research in Multimedia and Internet Systems*, Advances in Intellignet Systems and Computing ed., vol. 314, A. Zgrzywa, . K. Choroś and A. Siemiński, Eds., Wrocław, Springer, 2015, pp. 35-46.

[256] J. Li, „The application of CBIR-based system for the product in electronic retailing," w *2010 IEEE 11th International Conference on Computer-Aided Industrial Design & Conceptual Design (CAIDCD)*, Yiwy, China, 17-19 Nov. 2010.

[257] G. De Tre, D. Vandermeulen, J. Hermans, P. Claes, J. Nielandt and A. Bronselaer, "Bipolar Comparison of 3D Ear Models," in *Information Processing and Management of Uncertainty in Knowledge-Based Systems - 15th International Conference - IPMU*, Montpellier, France, 2014.

[258] A. E. Carpenter, "Extracting Rich Information from Images," in *Cell-Based Assays for High-Throughput Screening*, P. A. Clemons, N. J. Tolliday and B. K. Wagner , Eds., Springer, 2009, pp. 193-211.

[259] M. Mansourvar and M. A. Ismail, "Content-Based Image Retrieval in Medical Systems," *International Journal of Information Technology,* vol. 20, no. 2, pp. 1-9, 2014.

[260] A. Obero and M. Singh, "Content Based Image Retrieval System for Medical Databases (CBIR-MD) - Lucratively tested on Endoscopy, Dental and Skull Images," *IJCSI International Journal of Computer Science Issues,* vol. 9, no. Issue 3, No 1, May 2012.

[261] M. S. Chaibou and K. Kalti, "A New Labeled Quadtree-based Distance for Medical Image Retrieval," in *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016)*, Rome, Italy, 27-29 Feb., 2016.

[262] H.-s. Kim, H.-W. Chang, H. Liu, J. Lee and D. Lee, "BIM: IMAGE MATCHING USING BIOLOGICAL GENE SEQUENCE ALIGNMENT," 2010. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5414214.

[263] A. T. Inc., "image pattern recognition using vector quantization - uszczegółowić". the United States Patent and Trademark Office Patent 7,502,519, 2009.

[264] J. Mallik, A. Samal and S. L. Gardnerb, "A content based image retrieval system for a biological specimen collection," *Computer Vision and Image Understanding,* vol. 114, no. 7, p. 745–757, July 2010.

[265] G. Csurka, J. Ah-Pine and S. Clinchant, "Unsupervised Visual and Textual Information Fusion in CBMIR Using Graph-Based Methods," *ACM Transactions on Information Systems,* vol. 33, no. 2, pp. 9:1--9:31, Feb, 2015.

[266] L. Anselin and S. J. Rey, Eds., Perspectives on Spatial Data Analysis, Berlin: Springer, 2010, p. 290.

[267] C. Hahne, A. Aggoun, S. Haxha, V. Velisavljevic and J. C. J. Fernández, "Light field geometry of a standard plenoptic camera," *Optics Express,* vol. 22, no. 22, pp. 26659-26673, Nov. 2014.

[268] S. Cloix, T. Pun and D. Hasler, "Real-time Scale-invariant Object Recognition from Light Field Imaging," in *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016)*, Rome, Italy, 27-29 Feb., 2016.

[269] *IEEE Transactions on Image Processing,* vol. 13, no. 3, p. all, March 1994.

[270] S. Lyu, D. Rockmore i H. Farid, „A digital technique for art authentication," *Proceedings of the National Academy of Sciences of the United States of America,* tom 101, nr 49, p. 17006–17010, 7 Dec. 2004.

[271] M. Aubry, B. C. Russell and J. Sivic, "Painting-to-3D Model Alignment Via Discimanative Visual Elements," *ACM Transactions on Graphics,* vol. 28, no. 4, pp. 1-14, Article No. 106 , Aug. 2009.

[272] J. K. Gilbert, Ed., Visualization in Science Education, Springer Science & Business Media, 2006, p. 346.

[273] E. Alepis and M. Virvou, Object-Orianted User Interfaces fro Personalized Mobile Learning, vol. 64, J. Kacprzyk and J. C. Lakhimi, Eds., Heidelberg: Springer, 2014, p. 129.

[274] G. Ghiani, M. Manca and F. Paternò, "Authoring Context-dependent Cross-device User Interfaces based on Trigger/Action Rules," in *The 14th International Conference on Mobile and Ubiquitous Multimedia*, Linz, Austria, 30 Nov. - 2nd Dec. 2015.

[275] Z. Raisi, F. Mohanna and M. Rezaei, "Applying Content-Based Image Retrieval Techniques to Provide New Services for Tourism Indusry," *International Journal of Advanced Networking and Applications,* vol. 6, no. 2, pp. 2222-2232, Oct. 2014.

[276] W. Premchaiswadi, "An Image Search for Tourist Information Using a Mobile Phone," *WSEAS Transactions on Information Science and Applications,* vol. 4, no. 7, pp. 532-541, Apr 2010.

[277] M. Markkula and E. Sormunen, "Searching for Photos - Journalists' Practices in Pictorial IR," in *Electronic Workshops in Computing – Challenge of Image Retrieval*, Newcastle, UK,, Feb. 1998.

[278] D. Gurari, S. D. Jain, M. Betke and K. Grauman, "Pull the Plug? Predicting If Computers or Humans Should Segment Images," in *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, June, 2016.

[279] R. Datta, T. Joshi, J. Li and J. Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," *ACM Computing Surveys,* vol. 40, no. 2, pp. 5:1-5:60, Apr. 2008.

[280] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review,* vol. 10, no. 4, pp. 422-437, October 1968.

[281] A. Kundu and J.-L. Chen, "Texture classification using QMF bank-based subband decomposition," *CVGIP: Graphical Models and Image Processing,* vol. 54, no. 5, p. 369–384, 1992.

[282] C. Xu and J. L. Prince, "Snakes, Shapes, and Gradient Vector Flow," *IEEE TRANSACTIONS ON IMAGE PROCESSING,* vol. 7, no. 3, pp. 359-369, March 1998.

[283] "Fast Wavelet-Based Image Characterization for Highly Adaptive Image Retrieval," *IEEE Transactions on Image Processing,* 2012.

[284] D. Eads, D. Helmbold and E. Rosten, "Boosting in Location Space," Santa Cruz, 2013.

[285] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack and D. Petkovic, "Efficient and Effective Querying by Image Content.," *Journal of Intelligent Information Systems,* vol. 3, pp. 231-262, 1994.

[286] M. Koyuncu and B. Cetinkaya, "A Component-Based Object Detection Method Extended with a Fuzzy Inference Engine," in *Proceedings of the International Conference on Fuzzy Systems Fuzz-IEEE2015,* Istambul, 2015.

[287] J. Philbin, O. Chum and M. a. S. J. a. Z. A. Isard, "Object Retrieval with Large Vocabularies and Fast Spatial Matching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2007.

[288] K. Chen, "Deep and Modular Neural Networks," in *Handbook Computational Intelligence,* 1 ed., J. Kacprzyk and W. Pedrycz, Eds., Berlin, Springer, 2015, pp. 473-494.

[289] A. Huneiti and M. Daoud, "Content-Based Image Retrieval Using SOM and DWT," *Journal of Software Engineering and Applications,* no. 8, pp. 51-61, Feb 2015.

[290] L. Deng and D. Yu, "Deep Learning Methods and Applications," in *Foundations and Trends in Signal Processing,* Vols. 7, nos. 3–4, Now the essence of knowledge, 2014, p. 197–387.

[291] J. Bautista-Ballester, J. Verges-Llahi and D. Puig, "Using Action Objects Contextual Information for a Multichannel SVM in an Action Recognition Approach based on Bag of VisualWords," in *Proceedings of the 10th International Conference on Computer Vision Theory and Applications,* Berlin, 11-14 Mar, 2015.

[292] O. Allani, N. Mellouli, H. B. Zghal, H. Akdag and H. B. Ghzala, "A Relevant Visual Feature Selection Approach for Image Retrieval," in *VISAPP 2015 - International Conference on Computer Vision Theory and Applications,* Berlin, 2015.

[293] R. K. Srihari , "Automatic indexing and content-based retrieval of captioned images," *IEEE Computer,* pp. 49 - 56, Sep. 1995.

[294] Y. Liu, D. Zhang, G. Lu and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition,* vol. 40, pp. 262-282, 2007.

[295] S. K. Pal and P. Mitra, Pattern Recognition Algorithms for Data Mining. scalability, Knowledge Discovery and Soft Granular Computing., London, New York: Chapman and Hall CRC Press Company, 2004, p. 244.

[296] C. Beecks, M. S. Uysal and T. Seidl, "Signature Quadratic Form Distances fer Content-Based Similarity," in *ACM Multimedia,* Beijing, China, Oct. 19-24, 2009.

[297] H. E. Hurst, „Long-term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers,* pp. 770-808, 1951.

[298] N. Sebe and M. S. Lew, "Texture Features for Content-Based Retrieval," in *Principles of Visual Information Retrieval,* M. S. Lew, Ed., Springer Science & Business Media, 2013, pp. 50-81.

[299] I. Rish, "An empirical study of the naive Bayes classifier," in *IJCAI-2001 workshop on Empirical Methods in AI,* 2001.

[300] R. Datta, J. Li and J. Z. Wang, "Content-Based Image Retrieval - Approaches and Trends of the New Age," in *Multimedia Information Retrieval (MIR '05)*, Singapour, 2005.

[301] T. Jaworska, "The Concept of a Multi-Step Search-Engine for the Content-Based Image Retrieval Systems," in *Information Systems Architecture and Technology. Web Information Systems Engineering, Knowledge Discovery and Hybrid Computing*, Wrocław, 2011.

[302] Z. Wang , A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image Qualifty Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing,* vol. 13, no. 4, p. 600–612, April 2004.

Index

List of Figures