Noname manuscript No.
(will be inserted by the editor)

Retrieval of High–Dimensional Visual Data: current state, trends and challenges ahead

Antonio Foncubierta–Rodríguez · Henning Müller · Adrien Depeursinge

Received: date / Accepted: date

Abstract Information retrieval algorithms have changed the way we manage and use various data sources, such as images, music or multimedia collections. First, free text information of documents from varying sources became accessible in addition to structured data in databases, initially for exact search and then for more probabilistic models. Novel approaches enable content–based visual search of images using computerized image analysis making visual image content searchable without requiring high quality manual annotations. Other multimedia data followed such as video and music retrieval, sometimes based on techniques such as extracting objects and classifying genre. 3D (surface) objects and solid textures have also been produced in quickly increasing quantities, for example in medical tomographic imaging. For these two types of 3D information sources, systems have become available to characterize the objects or textures and search for similar visual content in large databases. With 3D moving sequences (i.e., 4D), in particular medical imaging, even higher–dimensional data have become available for analysis and retrieval and currently present many multimedia retrieval challenges.

This article systematically reviews current techniques in various fields of 3D and 4D visual information retrieval and analyses the currently dominating application areas. The employed techniques are analysed and regrouped to highlight similarities and complementarities among them in order to guide the choice of optimal approaches for new 3D and 4D retrieval problems. Opportunities for future applications conclude the article. 3D or higher–dimensional visual information retrieval is expected to grow quickly in the coming years and in this respect this article can serve as a basis for designing new applications.

Keywords 3–dimensional objects · Visual information retrieval · 3D retrieval · 4D retrieval · high–dimensional objects

Antonio Foncubierta-Rodríguez
University of Applied Sciences Western Switzerland (HES–SO)
TechnoArk 3, 3960 Sierre, Switzerland E-mail: antonio.foncubierta@hevs.ch

1 Introduction

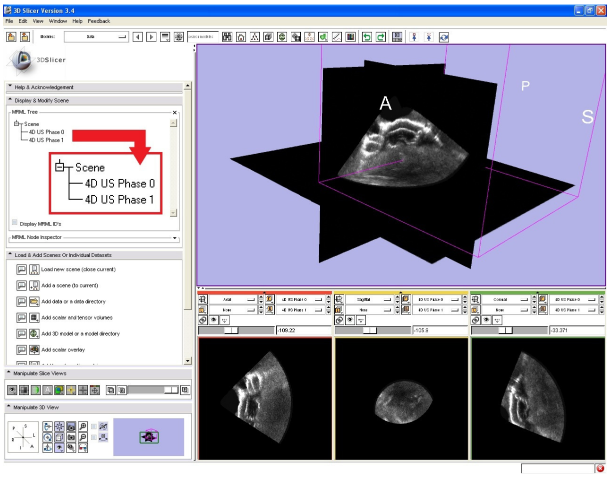
Multidimensional visual information encompasses a wide set of data containers ranging from images (2D), videos (2D plus time), to 3D surface models of objects, 3D solid models such as tomographic medical images or 4D temporal series of volume data. Images, volumes and videos are all part of multidimensional multimedia data. However, a distinction is needed in order to separate the mature, well–established 2D image retrieval domain from the developing higher dimensional (3D, 4D, 5D) retrieval domains. When a distinction between both data types is needed, the terms low–dimensional visual information and high–dimensional visual information will be used. In this work we use the term multidimensional information referred to n–D visual data or objects with n equal or greater than two, including images, videos, 3D models or 4D visual objects.

The amount of multidimensional data available has enormously increased in the past years: e.g. the video hosting website YouTube[1], founded in 2005, receives more than 60 hours of new video every minute (in early 2012) [1]. Other domains, such as medical imaging, produce an enormous amount of multidimensional information every day [8]. Such large quantities of data are difficult to manually categorize for further access or reuse. Whereas some tasks may be suitable for text–based retrieval, either with structured or free–text queries (e.g., retrieval of press events or images of particular geographical regions), other domains require specific retrieval paradigms to perform an efficient search in large databases, where adding textual annotations is not feasible or subjective and error–prone (e.g., feelings that are invoked by visual data). This is the case of high–dimensional visual information, where understanding and interpreting is time–consuming and not so intuitive: e.g., a 2D image can be understood immediately without interaction, whereas a 3D volume or video requires either sliding through slices or browsing a sequence through time. Figure 1 shows examples of interfaces for viewing high–dimensional visual data. This also motivates the use of computer–based approaches for analyzing high–dimensional data, due to the limitations of displaying dimensions larger than three for human inspection. The use of additional data together with visual–only information has proven to be valuable for retrieval and classification purposes [35]. This extra information is often included in the same container or file format: e.g. the DICOM[2] standard enables the storage of metadata together with images, providing *context to the visual content* [125,96]. However, not all domains can deal with metadata to the same extent, and its usefulness is strongly related to the application. E.g., in medical information retrieval, age can be a very selective criteria for specific conditions and diseases, but not for others.

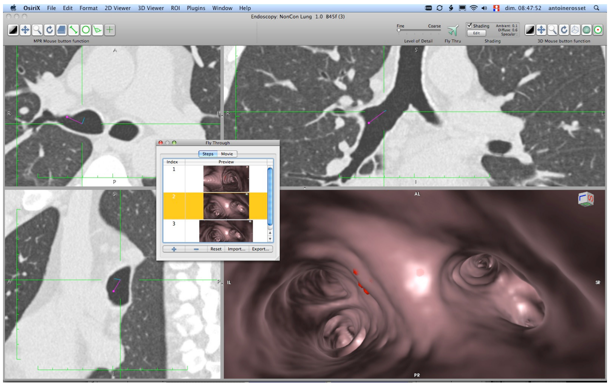
The aforementioned challenges, namely the complexity of the content as well as the enormous size of the data collections, show an urgent need for visual content–based retrieval systems. In the past decade multidimensional information retrieval beyond 2D image retrieval has been attracting an increasing interest from the research community [117,123]. Visual 2D image retrieval was extended to higher dimensions. The number of publications in these fields has grown from dozens of papers in the year 2000 to hundreds by the end of 2010. A query with the keywords *3D retrieval*, *video retrieval* and *image retrieval* in the publication search system Sco-

[1] `http://www.youtube.com/`, as of 3 May 2012.

[2] Digital Imaging and Communications in Medicine.



(a) 3D Slicer showing Multi–Planar Rendering (MPR) and a slicing view of ultrasound imaging (`http://www.slicer.org/`, as of 3 May 2012).



(b) OsiriX showing MPR and surface rendering of CT (Computer Tomography) imaging for virtual endoscopy (`http://www.osirix-viewer.com/`, as of 3 May 2012).

Fig. 1 Interfaces for viewing high–dimensional medical data showing the possibility of having views through slices of the volume or render surface–based views of the 3D data.

pus[3] clearly shows this trend for topics covering the "multidimensional" category (see Figure 2). The highest growth period for multidimensional visual information retrieval research occurred around the year 2005 when important contributions were published: the Princeton benchmark initiative for 3D objects [116], the first Shape Retrieval Contest (SHREC) [127] and comprehensive reviews of the literature on 3D object retrieval [17, 123]. This analysis can be limited by the maturity of the field: i.e., once a domain is well–stablished, researchers may tend to use less often terms that are redundant within this community.

[3] http://www.scopus.com/, as of 3 May 2012.

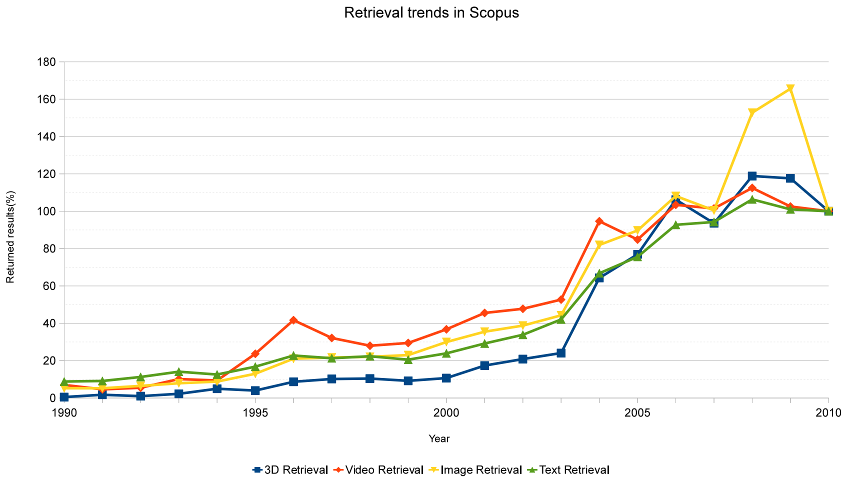


Fig. 2 Evolution of the number of articles found in Scopus for various queries containing the keyword *retrieval* in the title, keywords or abstract. 100% corresponds to the number of articles found in 2010 for each category.

In this article, a review of the high–dimensional visual information retrieval domain is presented, describing the most important applications and techniques found in the literature. The aim of this article is to find similarities among techniques across domains to foster cross–domain synergies between applications and techniques. The article provides a brief description of the most common methods available to researchers that face a high–dimensional retrieval task classified by data dimensionality rather than content type. In this sense, it is complementary to previously published reviews of content and concept–based retrieval systems for images [118, 113, 97, 30, 4], videos [90, 119] and 3D objects [123, 17].

The rest of this paper is organized as follows: Section 2 describes the review methodology used for the paper, Section 3 lists the main applications for high–dimensional visual information retrieval, and Section 4 summarizes the most widely employed techniques and how they differ from the ones used for 2D image retrieval. The specific challenges for the high–dimensional case and conclusions are explained in Section 5.

2 Methods

A systematic analysis of the research literature was executed to retrieve the research trends in the field and the most important papers being published in the last more than ten years. The research–oriented search engine Scopus was chosen because of the large amount of publications that it indexes, including but not limited to those published by Elsevier, Springer, ACM (Association for Computing Machinery), IEEE (Institute of Electronic and Electrical Engineers) and SPIE. Scopus might include fewer publications than Google Scholar but in general the publications listed are of high quality and the references are complete. Most important journals and conferences dealing with multidimensional visual information retrieval are covered. A set of queries were performed to find a total of 5564 relevant publications (see Table 1). Abstracts were analyzed using an online keyword extraction tool[4] that provides stop–word lists for the English language. Results were divided based on time periods for which the growth pace of the number of publications is approximately stable according to Figure 2: publications before the year 2000, publications from the year 2000 to 2005 and publications after 2005. This allows obtaining a more detailed picture of what are the most important trends in the field. Similar methods have previously been used to analyze the impact of publications in [124].

Table 1 Number of papers retrieved by the Scopus search engine for various queries and time periods.

(a) Query: 3D retrieval refined with shape, model or surface.

	before 2001	2001-2005	after 2005	Total
Papers	125	514	1627	2266

(b) Query: visual information retrieval refined with 3D, 4D, 5D, multidimensional, image, volume or volumetric data and not video.

	before 2001	2001-2005	after 2005	Total
Papers	15	40	84	139

(c) Query: video retrieval refined with visual or content–based.

	before 2001	2001-2005	after 2005	Total
Papers	534	959	1666	3159

3 Applications

In this section, the main applications domain of multidimensional retrieval are presented. Applications are regrouped based on the nature of their data as follows: Section 3.1 deals with surface–based model retrieval, including watertight models and polygon soup models. Section 3.2 takes into account full–support data, i.e., multidimensional data that can be defined as a solid volume in 3D or a hyper–volume of higher dimensionality, also treating the case when images of two or more dimensions are sampled in time, such as in general–purpose video or 3D+t medical imaging.

[4] `http://www.tagcrowd.com/`, as of 3 May 2012.

3.1 Surface–based model retrieval

Model–based retrieval includes a set of applications requiring the ability to recognize and retrieve 3D surfaces with similar shapes.

Definition 1 *Let $\mathcal{A}, \mathcal{B}$ be two subsets of a Euclidean space (see Eq. 1). The subsets are said to have the same shape if there is a rotation matrix $\mathbf{R}$, a not null scaling factor s and a displacement vector $\mathbf{d}$ that transform every point $\mathbf{y} \in \mathcal{B}$ into one point $\mathbf{x} \in \mathcal{A}$ satisfying Eq. 2.*

$$\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^n, \tag{1}$$

$$\mathbf{x} = s\mathbf{R}\mathbf{y} + \mathbf{d} \tag{2}$$

This definition of shape is often too rigid, and more flexible definitions are used for practical applications. Some research communities define shape from a topological point of view [34,43] whereas other applications stress the importance of partial matching in shape analysis [93].

Results from the online text analysis tool in Figure 3 show that research moves from technology–centered studies [101,131] based on general–purpose polygonal retrieval [14,122] to application–focused research [19,68,99,146]. Another remarkable trend is that face recognition is a novel yet active topic in multidimensional research, with a high number of publications in the past ten years.

By far the most frequent application for model–based retrieval is general–purpose object retrieval without a clear real–life application described by the authors. Existing model–based datasets are particularly well suited for general–purpose applications where the ground truth consists of widely accepted categories (e.g., people, animals, buildings, etc.) [116]. On the other hand, it is often difficult to find publicly available datasets specific to a certain topic, where most research groups evaluate only their own datasets [43], as Bustos et al. describe in [17]. Some of these topic–specific, real–life applications for model–based retrieval include, but are not limited to:

- face recognition [85,111,134,145];
- retrieval of pieces for industry processes [25,28,43];
- retrieval of artistic and architectural objects [68,115].

Illustrations of the above mentioned applications are depicted in Figure 4.

3.2 Full–support retrieval

Surface–based model retrieval deals with external aspect of objects, specifically with concepts like shape, structure or topology. In contrast, some applications require knowledge of the internal aspects of visual data, dealing with concepts like texture or density. These applications are covered by full–support data, which describe objects across all possible dimensions.

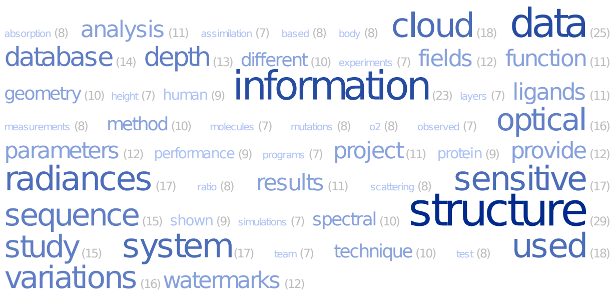
The concept of full–support data can be described using signal processing concepts such as the intrinsic dimension of a multiple variable signal [15]. The intrinsic dimension of an N–variable signal is the minimum number M of variables needed to represent the signal.

Definition 2 *The intrinsic dimension M of the signal f (see Eq. 3) is the smallest number for which the relation in Eq. 4 is true for all $\mathbf{x}$, for some M–variable function g and $M \times N$ not null matrix $\mathbf{A}$.*

$$f(\mathbf{x}) = f(x_1, x_2, \ldots, x_N), \tag{3}$$
$$f(\mathbf{x}) = g(\mathbf{A}\mathbf{x}) \tag{4}$$

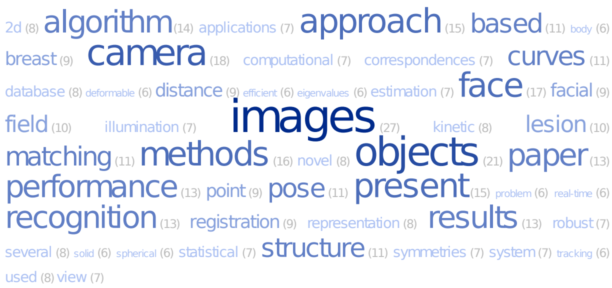
In this section we consider the full–support case, i.e., when $rank(\mathbf{A}) = N$, with $N \geq 3$, meaning signals that, requiring at least 3 variables to be indexed, are described by the smallest possible number of variables.

Results from queries shown in Tables 1(b) and 1(c) were analyzed in order to extract the most frequent applications. A further distinction can be made based on the nature of the variables. The subset of applications where all variables are referring to spatial dimensions is described in Section 3.2.1 whereas the applications



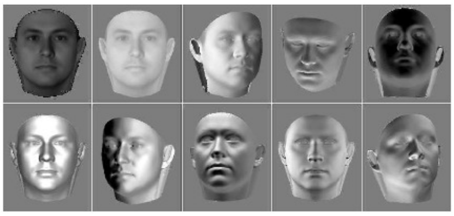
(a) Publications before 2001.

(b) Publications between 2001 and 2005.

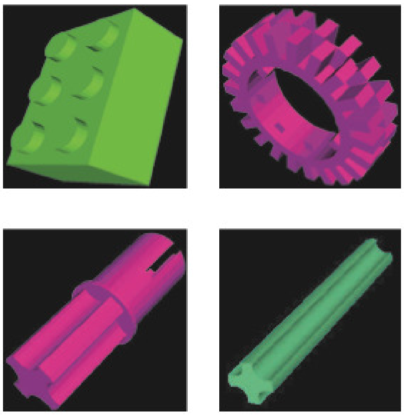


(c) Publications after 2005.

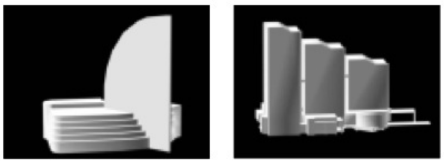
Fig. 3 Keywords found in 2266 abstracts from publications on surface–based model retrieval regrouped by publication period.



(a) Face recognition [145].



(b) Retrieval of mechanical pieces for industrial processes [28].



(c) Retrieval of architectural objects [115].

Fig. 4 Examples of surface–based retrieval applications.

with intrinsic dimension equal or greater than 3 containing at least one variable referring to time are considered in Section 3.2.2.

3.2.1 Spatial–only full–support data

Although the extension from 2D images to 3D might appear intuitive, acquisition methods and applications have strongly limited the spread of retrieval techniques for this type of data as shows the number of publications on the topic (see Table 1(b)). Due to the opacity of matter, optic acquisition is often not possible for these applications, so most of the techniques used for extracting the matter properties from within a volume are those capable of showing an insight into matter, such as X–ray, magnetic resonance and ultrasound imaging or 3D confocal microscopy. This list of techniques is enough to justify why the most frequent application for

3D full–support retrieval is medical imaging, as it can be seen in the tag cloud from Figure 5, where the keyword *medical* is among the most frequent terms found in the texts.

Applications where full–support information is used for retrieval are the following:

– Medical image retrieval for computer–assisted diagnosis with a specific clinical application [37].
– General purpose medical image retrieval for PACS (Picture Archival and Communication System) browsing [13, 59, 75].

Retrieval and classification techniques are closely related, since both often have identical feature–extraction steps; sometimes classification is achieved after a retrieval process. Retrieval has been defined as a classification task between relevant and not relevant (usually without training data), for instance in the *Binary Independence Retrieval* model [98]. For this reason a growth of the use of full–support texture would make it possible to find retrieval systems based on existing classification–based applications. E.g., in the geology field, several classification applications have been proposed [55, 57, 71] and retrieval applications may evolve from these as the techniques related to visual description of geological and other three–dimensional data spread within the related community.

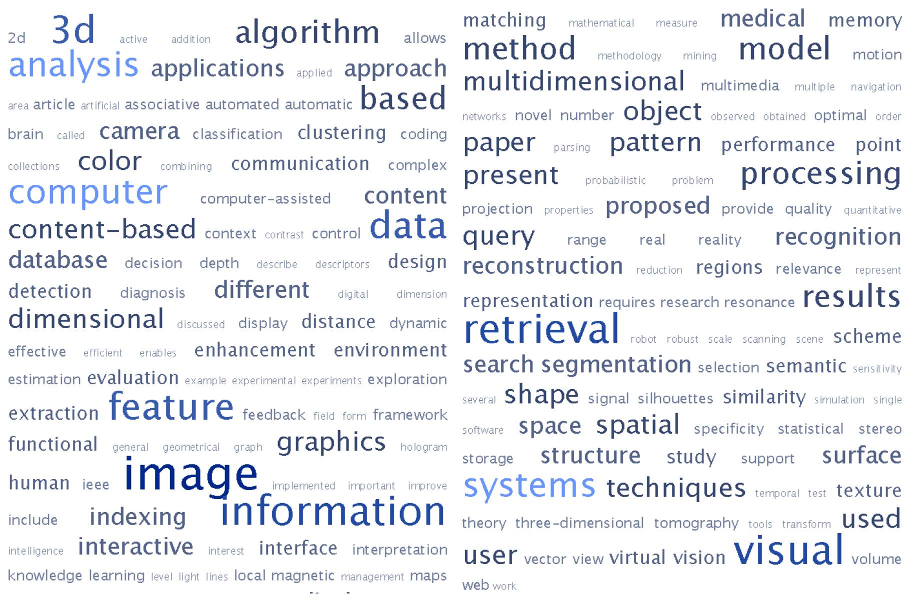


Fig. 5 Keywords found in 139 abstracts from full–support retrieval publications.

3.2.2 Space and time volumetric data

In concordance with the explosion of user–generated video content mentioned in Section 1, there has been enormous efforts for video retrieval research in the past years. Video retrieval, as shown in Table 1(c), is by far the subtype of multidimensional retrieval that received the highest attention also thanks to the availability of large test collections created in the TRECVID benchmark.

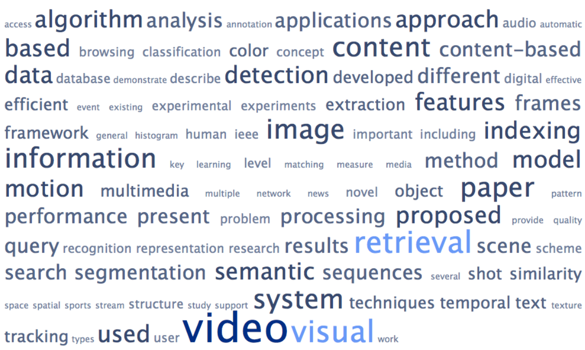


Fig. 6 Keywords found in 3159 abstracts from video retrieval publications.

As can be seen in Figure 6, video retrieval often focuses on the understanding of the *semantics* and *syntactics* of visual information to provide a way of indexing videos [6]. This includes scene classification and shot boundary detection [86], an area where big efforts where made in the 1990's [49,62]. With spoken text, videos also have a possibility to extract semantic information from the sound. The most common application for video retrieval is large–scale audiovisual collection management [94,137]. Evaluation of video retrieval is also very active and standardized, with important contributions from TRECVID[5], videoCLEF [83,84], and MultimediaEval[6].

4 Techniques for visual information retrieval

Efficient visual information retrieval requires facing two challenges: on the one hand the problem of accurately describing the information encompassed in a *visual container* is tackled by using computer vision and image processing, also known as feature extraction. On the other hand the problem of dealing with large amounts of complex information for achieving fast and accurate results that are relevant to the query is approached by using machine learning and information retrieval techniques. Figure 7 contains an overview of a generic visual information retrieval system, distinguishing the visual description phase and the information retrieval step.

Visual information can be retrieved in different ways. In some domains, it is possible to define categorical elements that enable description and retrieval: e.g., a film can be described in terms of the genre (comedy, drama, science–fiction, etc.). Some domains require retrieving documents without attending to categories, but

[5] http://trecvid.nist.gov/, as of 3 May 2012.
[6] http://www.multimediaeval.org/, as of 3 May 3012.

to similarities. E.g., a film can be described in terms of its length in minutes or aspect ratio, and therefore similar films would have a similar length and aspect ratio. This idea is further extended using the concept of feature vectors.

Definition 3 *Let $f_1, f_2, \ldots, f_n \in \mathbb{R}$ be n numerical values representing n features or characteristics that apply to visual information elements or documents. Then, a feature space $\mathcal{F} \subseteq \mathbb{R}^n$ can be constructed for all the valid values of $f_1, f_2, \ldots, f_n$ where each dimension is related to one of the features. A visual information element or document X can then be mapped to a point in the feature space, the point represented by the values of the features $f_1 = x_1, f_2 = x_2, \ldots, f_n = x_n$ for the document. The vector $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{F} \subseteq \mathbb{R}^n$ is called feature vector of the document X.*

Two documents X and Y with feature vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be similar if $d(\mathbf{x}, \mathbf{y}) < T$ is true for some distance measure d and a given threshold T.

In general, not only distances are used as similarity measures, other metrics and (dis–)similarity measures can be used attending to the type of features used and the desired properties of the retrieval system.

Techniques for defining feature vectors out of visual content in high dimensional data are further explored in Section 4.1, the description of similarity, distance measures and other information retrieval techniques are outlined in Section 4.2 and methods for fusing several retrieval techniques and feature vectors as well as metadata is explained in Section 4.3. Finally, Section 4.4 deals with the representation challenges for high–dimensional visual information.

4.1 Visual information description

There are various approaches for describing visual information in multidimensional data. The choice of one or another is often related to the application of the retrieval system. For instance, for machine parts retrieval shape is more important than texture, and therefore information extraction methods are focused on shape and surface quantification. However, the main distinction among methods is whether they are 3D–native or they use a *divide and conquer* approach to multidimensionality, working on lower dimension spaces and aggregating this information later on, e.g., analyzing 3D–images slice by slice.

4.1.1 High dimensional approaches

In this section we consider methods that obtain information from all dimensions simultaneously: for instance, methods based on mapping properties of a 3D model onto a 3D sphere but not those that map data onto a planar surface; similarly, we consider high dimensional approaches that analyze images computing features in 3D neighborhoods as opposed to 2D neighborhoods. A distinction is made between the techniques that involve shape or surface information and those that also include volumetric features such as 3D texture.

Shape description From very simple statistics to complex topological graphs, shape is widely used for 3D retrieval, since object matching is also one of the clearest applications. Table 2 shows a description and classification of popular methods.

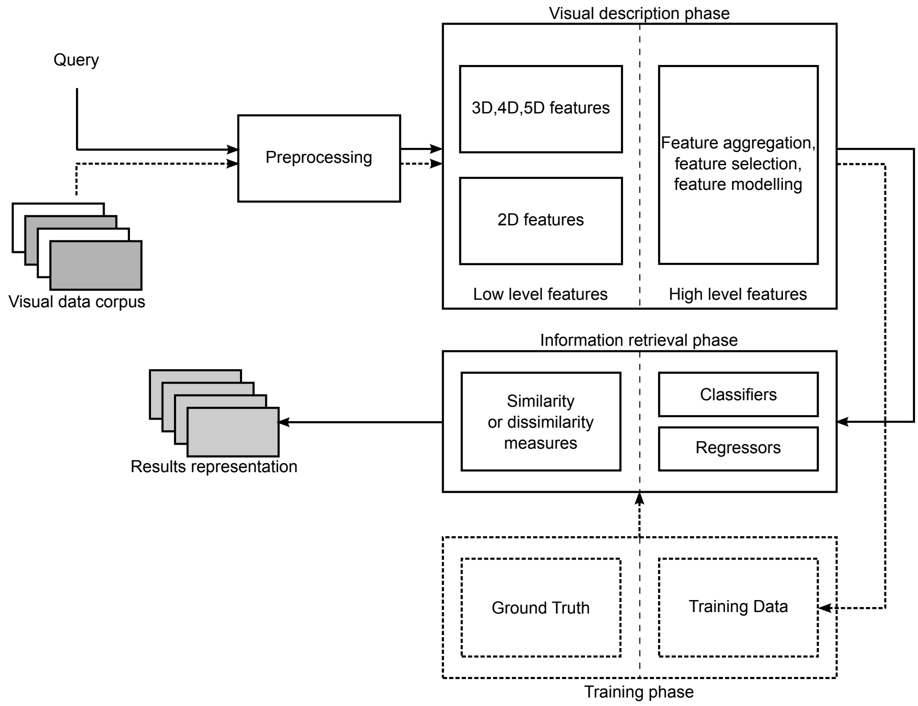


Fig. 7 Overview of a generic visual information retrieval system. The high–dimensional visual data from the retrieval corpus (dashed line) is processed and used as training data for supervised or unsupervised machine learning methods. The high–dimensional visual data from the query (full line) is processed in a similar way but is not involved in the learning process.

Full–support data description. Both volumetric images and videos contain information as a series of images, sampled in space and in the case of videos, also in time. Despite the similar nature of information, different approaches are often used. For instance, some techniques are tightly related to video, where there has been a big effort by the Motion Picture Expert Group (MPEG) in finding a multimedia information description model with the MPEG–7 standard; whereas visual pattern description in the field of spatial–only information, often known as solid or full–support texture [106,32], has been approached in other ways. A summary of common full–support description techniques is shown in Table 3.

4.1.2 Low dimensional approaches

Due to the complexity of the multidimensional visual information, the high dimensional description task is often reduced to multiple 2D feature extractions. For instance, a 3D model can be described by *view–based techniques*, i.e., a set of 2D images are computed based on views of the object from various perspectives. By reducing the dimensionality, common 2D–descriptors can be used, often at the cost of missing a complete characterization of the object unless the number of views grows sufficiently. Table 4 lists some low–dimensional techniques.

Table 2 Shape description methods in 3D.

(a) Point–based methods.

Methods	Explanation	Examples
Distance distributions	Probability distribution of distances between points sampled on the surface of an object.	[101, 67, 100]
Shape histograms	The volume that the object fills is divided in bins (radial divisions, angular divisions, both, or other divisions), the object is described by the histogram of occurrences according to these bins.	[9, 122]
Geometric moments	The object is considered a random process of 3 variables, described in terms of statistical moments.	[112, 44]
Spherical harmonics, raycast descriptors	The object is described by evaluating the intersection points with a predefined set of rays casted from the surface of a sphere containing the object.	[112, 130, 64]

(b) Surface–based methods.

Methods	Explanation	Examples
Point signatures	The object is sampled on its surface and to each point a signature describing the local curvature of the surface is assigned.	[27, 26]
Extended Gaussian image	The object is placed inside a Gaussian sphere, and a histogram is computed from the intersection with the sphere of the normal vectors on the surface of the object.	[53, 136]

(c) Topology and volume–based methods.

Methods	Explanation	Examples
Topological and skeleton based descriptors	The object is described in topological terms according to the relationships of its subparts. A skeleton of a volumetric model might be generated as a descriptor of the object.	[65, 121]

4.2 Information retrieval

A retrieval system needs to be able to provide relevant documents to a query based on the concept of (visual) similarity. Although being a critical step, visual description (or visual features) is not enough for achieving a relevant versus non–relevant classification or to rank documents according to visual similarity. The visual description step in a retrieval system consists of finding a set of features or descriptors that are meaningful for the retrieval purpose: i.e., that can code the differences and similarities among the items to be retrieved. Once these features have been obtained, the final step involves a decision–making process to find a mapping that aggregates the information of the visual descriptors to a class or a ranking. To achieve this, two strategies can be used: defining of (dis–)similarity measures and/or using machine learning methods on training data.

When using similarity or dissimilarity measures, training data is not always required for the system to work. It can perform retrieval directly on the data set by sorting the items according to the chosen (dis–)similarity measure with respect to the query item. One of the the simplest and still most frequently used techniques is the k–nearest neighbor (kNN) search, where the retrieved items consist of the k documents closest to the query item in the feature space. kNN works well if several local groupings or clusters of documents/objects exist in the feature

Table 3 Full–support description methods in 3D.

(a) Geometry–based methods.

Methods	Explanation	Examples
Textons, texels	full–support information is described relying on the assumption that the observed pattern is constituted by elementary units, called *textons* or *texels*, repeated with varying spatial distributions, sizes and orientations.	[72,132, 133]
Measures from binarized images	By binarizing the images, higher–level geometric measures can be extracted from the volumes, such as uniformity, granularity, volume, surface and others.	[102–104, 74]

(b) Spectral–based methods.

Methods	Explanation	Examples
Fourier analysis	The information is approximated by a linear combination of basis functions in a given direction. In order to have local information of the data, Fourier analysis requires that the transformation is applied in a window around the interest point.	[56,78,82]
Filter–based methods	Instead of using the windowed Fourier transform, local spectral properties are obtained by convolving the information signal with a given template. The template or filter is a function of limited support with given direction, scale and phase properties. These functions can be tailored to detect specific features: such as edges or corners.	[2,11,16, 58,81,91, 109]
Multiscale analysis	Multiscale analysis can be achieved by a a filterbank of templates at different scales organized in a pyramid. One of the most common multiscale approaches is the Wavelet Transform (WT), but other filters or templates can be used to describe multidimensional patterns.	[3,31,51, 58,70,89, 103,107, 114,142, 144]

(c) Statistical and stochastic methods.

Methods	Explanation	Examples
Co–occurrence methods	Statistical measures based on the co–occurrence between the gray or color values of pairs of pixels at predefined relative positions.	[11,18,21, 22,52,66, 76,79,81, 126]
Run–Length methods	Run–length is an encoding method that describes data by computing the number of consecutive repetitions of the same value. In multidimensional data, a run–direction is first defined, and the number of consecutive voxels with the same value is computed. With this description, higher–level statistical measures are computed.	[76,77, 138–140]
Local Binary Patterns	Local Binary Patterns (LBP) compute the statistics of the spatial organization of voxels on the surface of (hyper–) spherical neighborhoods of the voxels. They are gray–scale invariants, and since they characterize spherical frequencies they are related to spherical harmonics.	[46,47,58, 105,107]
Markov Random Fields	3D Gaussian Markov random fields encode the relationships between values of voxels in volumetric spherical neighborhoods.	[42,108]

(d) Video–specific methods.

Methods	Explanation	Examples
Compressed domain descriptors	Exploiting the compression features to compose a feature vector for video comparison. For instance: the Discrete Cosine Transform (DCT) coefficients or motion vectors derived from coding standards such as MPEG–2 or H.264.	[92,128]
MPEG–7 descriptors	MPEG–7 Visual description tools include the visual basic structures (such as description tools for grid layout, time series, and spatial coordinates) and visual description tools that describe color, texture, shape, motion, localization and faces.	[117,92, 128,45,60]

Table 4 Low dimensional methods.

Methods	Explanation	Examples
Spin Images	By defining a set of normal vectors to sampled points on the surface of the object, a 2–dimensional histogram is defined by projecting the object points in a neighborhood of the sample point onto a plane defined by the vector.	[5,12,39]
Silhouettes and depth images	The objects are described by several 2D images corresponding to the views from a fixed number of points. If the distance information is kept then the image is called depth–image, whereas if the distance information is discarded, the resulting image is a binary silhouette.	[10,20,24, 88]
Slice or frame based	The volume is described by individually processing each of the slices or frames, or a selection of them. For instance, a compressed video can be described by the features that describe each of the so–called *keyframes*.	[50,37]

space without very clear class boundaries. The definition of closest strongly depends on the distance metric used. Most (dis–)similarity measures are based on the computing the Euclidean distance between two elements in the feature space. For example, let the query item Q be represented by the N–dimensional feature vector $\mathbf{f^Q} = (f_1^Q, f_2^Q, \ldots, f_N^Q)$ and an item i in the dataset be represented by the feature vector $\mathbf{f^i} = (f_1^i, f_2^i, \ldots, f_N^i)$, then a dissimilarity measure based on the Euclidean distance can be defined as $d_{i,Q} = \sqrt{(f_1^Q - f_1^i)^2 + (f_2^Q - f_2^i)^2 + \cdots + (f_N^Q - f_N^i)^2}$. Other distance measures are often used instead of the Euclidean distance, according to the desired properties of the measure or the specific characteristics of the feature vector,e.g, the Mahalanobis distance, the earth mover's distance or histogram intersection. Therefore, there has been much interest in comparing distance metrics for this purpose [40,110].

Machine learning methods are also very popular in the information retrieval step as shown in Figure 7. A machine learning method requires training data as a previous experience in order to accurately predict the relevance of the items for the query. Machine learning methods can be classified as supervised or unsupervised, depending on whether ground truth was available during the training.

From a classification point of view, supervised methods try to find the best boundaries between classes by making decisions knowing the labels assigned to a given training set [80]. One of the most frequently found methods in supervised learning are Support Vector Machines (SVM) [23] that also lead to best results in many visual information retrieval benchmarks [95]. Another trend in supervised learning are relevance feedback methods, where the retrieval system evolves by using the manual feedback from the user [61,137].

Representation of complex concepts with low–level features as presented in Section 4.1 remains difficult due to the so–called *semantic gap* between computer–understandable low–level features and human–understandable high–level semantic concepts. Various techniques try to reduce this gap, either using machine learning methods A relatively recent trend among machine learning methods is the *bag–of–words* approach, which extends a concept from text retrieval to the visual information retrieval field. *Bag–of–words* or *bag–of–visual–words* attempt to learn concepts from the features, clustering the feature space into densely populated regions that might represent visual concepts on the images. The histogram of visual words is subsequently used as a descriptor of a volume or part of it [50, 141]. *Bag–of–words* can be considered unsupervised during the clustering phase, and supervised if the features were obtained using a supervised machine learning method.

4.3 Fusion of descriptors and retrieved elements

As seen in Section 4.1, a visual information element can be described by different types of features. Moreover, some domains use valuable metadata that can significantly improve retrieval efficiency. In Section 4.2, some approaches to retrieval have been introduced. It is therefore clear that on the one hand, some features might be better suited for some retrieval applications than others; and on the other hand, some information retrieval techniques might provide better, faster or more accurate results than others. However, some applications might benefit from a combination of techniques. E.g., results can significantly improve when integrating clinical data into content–based image retrieval, [35, 147]; in the video analysis domain, multimodal approaches[7] have proven to be more effective than unimodal approaches [120, 69, 7]. These situations are dealt by using *fusion techniques*.

Fusion techniques are often classified into early and late fusion. Based on the definitions given by Snoek et al. [120], early and late fusion can be defined as follows:

Definition 4 (Early Fusion) *Fusion scheme that integrates unimodal features before making decisions such as classification, concept–learning, retrieval.*

Definition 5 (Late Fusion) *Fusion scheme that first reduces features to separately made decisions (classes, scores, rankings, etc.), then these are integrated.*

In general, the term *early fusion* refers to the combination of various types of features into a single descriptor and *late fusion* refers to the combination of various lists of retrieved documents (runs) into a single, ranked list of elements.

Fusion of various sources of information can be triggered within the retrieval system by using query expansion techniques, which modify the original query based on available documents in the database or given rules.

Data fusion techniques, together with query expansion, have been widely used in benchmarking events like ImageCLEF [33] and TRECVID [135, 41, 29].

4.3.1 Early fusion approaches

Early fusion techniques combine descriptors in order to construct a higher dimensionality feature space, where all relevant features are present. The major disadvantage of this approach is the curse of dimensionality: as the the dimensionality of the feature space increases the density of elements in the space is reduced,

[7] In video analysis, multimodality refers to the use of multiple information sources for the same document: audio, text and visual information. This concept is easily generalized for other domains, for instance in medical imaging, visual information and metadata included in the DICOM headers.

scattering meaningful clusters of instances. To solve this problem, various feature selection, feature normalization [48] and feature weighting [143, 38] schemes have been used.

4.3.2 Late fusion approaches

Diversity among late fusion techniques is much broader than among early fusion approaches. Late fusion includes every technique that combines outputs of various systems into a single, sorted list of documents. Fusion techniques can be regrouped in three subcategories:

Rank–based: items are combined attending to their position in each of the previous lists of documents, either by intersection, union or another combination rule. These techniques often require reordering rules.

Score–based: items are combined attending to their relevance score, similarity or distance to the query item. These techniques require normalization of relevance scores among all systems.

Probability–based: items are assigned a score based on the probability of relevance, according to a trained fusion system [87]. These techniques require training queries with corresponding ground truth (relevance judgements).

A specific review on rank, score and probability–based fusion techniques by Donald and Smeaton [41] compares the performance of various techniques on TRECVID collections.

4.4 Data representation

Human intuition is often limited to three dimensions. Representation and understanding of higher dimensional data requires further knowledge and training. This limitation increases the difficulties faced by visual information retrieval systems at the result representation stage. Different strategies have been proposed to overcome this challenge, which can be grouped into the following categories:

Projection into lower dimensional space(s). Similar to the view–based techniques (see Section 4.1.2), visual information is projected into one or more lower dimensional spaces, often with samples at one of the discarded dimensions. These techniques are well known in the audiovisual domain [129], where audio information is often discarded for presentation and time is used as a sampled dimension: e.g. representation of a video by a series of thumbnails.

Interaction and virtual reality. Discarding one of the dimensions is often not easily possible, or there is no clear dimensionality that can be discarded a priori. In these cases, interactive techniques have been proposed to *enable* or *browse* dimensions according to users' needs. These methods are widely used in the medical domain, with virtual reality systems [54] or slice–browsing [36].

Addition of false visual attribute(s). When information about non–visual characteristics of high dimensional elements are needed, false visual attributes can be used. E.g.: transparency or false color have been widely used in volume rendering to represent concepts such as density or heat. Medical imaging makes often use of volume rendering [36] and false color to represent various anatomical structures and regions.

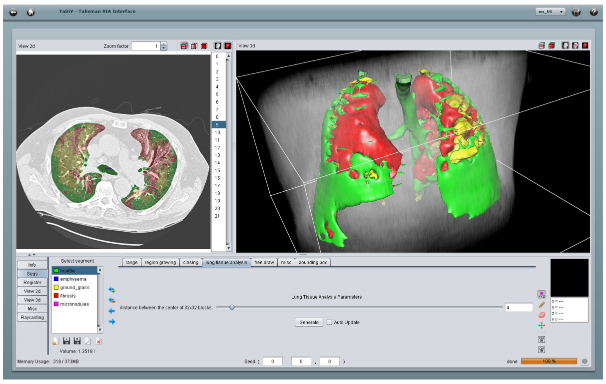


Fig. 8 Combination of data representation techniques in visual retrieval systems. Interactive slice–browsing and false color on the left pane and false color and transparency on the right pane. Source: [36].

Real–life systems often implement several methods separately or combined, in order to adapt to the users' workflow. For instance, the system shown in Figure 8 uses false visual attributes on the right pane and interactive slice–browsing on the left pane.

5 Conclusions and Challenges Ahead

In this paper a comprehensive review of the state of the art in high–dimensional visual information retrieval is presented. By systematically selecting and analyzing the publications of the past more than ten years in this field using SCOPUS, four major areas of interest were found: video retrieval as the most popular among all high–dimensional visual information retrieval applications; face recognition that is quickly gaining interest for its applications in the security industry and where 3D information has a clear added value over 2D; surface–based retrieval applications that include machinery retrieval of objects and related applications; and finally medical image retrieval that is by far the most popular application in spatial–only volumetric (often 3D texture) retrieval.

High–dimensional visual information retrieval has started solving some of the challenges regarding descriptors and machine learning in the domain. However, it still faces many challenges in terms of usability and scalability. High–dimensional visual information is a very large and complex data source. The main challenges are related to the difficulty of dealing with large datasets of very dense data. Feature extraction is time consuming and often produces a large number of visual descriptors.

A major challenge in visual information retrieval is related to the complexity of the data, which makes it difficult to find a small set of features that can accurately describe the documents. However, having a too large set of features will cause most basic machine learning methods such as k–NN to fail, due to the well–known curse of dimensionality [63]. This is one of the reasons for the bag–of–words approach attracting much interest, since it creates clusters of features that are relevant to the dataset defined by lower–level features. This lower dimensional set of features is based on the visual descriptors actually occurring in the data and allows for better distance measures and machine learning to be employed.

Research in high–dimensional visual information retrieval can profit from a closer collaboration among researchers. One of the most–common problems found in this field is the lack of publicly available datasets with annotated ground truth that can be shared by various research groups and therefore serve as baseline comparison for retrieval techniques. Benchmarking initiatives such as SHREC [127] in the field of shape–based retrieval or ImageCLEF [73] in the field of 2D image retrieval can become a powerful tool to create synergies among research groups to compare the various approaches and select best techniques for future applications.

Challenges in the medical field and on 3D solid textures are also multiple. Whereas 3D objects have the entire object information being relevant for retrieval in the case of 3D tissue types, in biomedicine, detection rather than full retrieval seems important as the volumes of interest relevant for retrieval are often very small and contain less than 1% of the volume to be analyzed. Detecting these regions of interest requires training data annotated by experts, a difficult task and often expensive as well. Based on a first detection step, then retrieval of similar volumes or cases could be performed. Whereas 3D surface models can be visualized easily, 3D texture is already hard to display and most often several views are required, as shown in Figure 1. Higher dimensional data will get even harder and new visualization methods need to be developed, for example to highlight abnormalities in ten energy levels of a 3D dual energy CT (Computer Tomography) of one patient, where visualization is far from trivial.

In general, retrieval from data in more than three dimensions can be regarded as one major challenge for the future. 3D cinema has already started and in medicine a large variety of imaging techniques produce more than 3D data such as PET/CT (Positron Emission Tomography / Computer Tomography) images, PET/MRI (Positron Emission Tomography / Magnetic Resonance Imaging) images or dual energy scanners. This will again increase the volume of data and will require data reduction before any retrieval can be attempted. Using approaches similar to visual words can help but also the basic descriptors will need to be adapted to multiple dimensions. Simple descriptors such as co–occurrence matrices are easy to adapt apart from the fact that an extremely large amount of data is being produced but for other descriptors e.g., wavelets) the formulation and usefulness beyond 3D might not be as trivial.

This article reviewed the literature on high–dimensional visual retrieval techniques. It can be shown that although video retrieval has been most popular over the past ten years, there are now many other developments, ranging from surface–based retrieval methods to solid 3D texture. Even higher dimensional data now becomes increasingly common, such as 3D cinema (3D plus time equals 4D) and also in the medical field where 4D image series become standard and where several volumes of the same patient can be produced combining CT and MRI or creating multiple energy images of CTs of the same patient. There are many challenges that retrieval applications will need to deal with in the future such as combining detection of regions of interest, dealing with computationally expensive analyses, and extremely large feature spaces. Visual user interfaces also need to be adapted as already 3D solid texture is hard to visualize and as dimensionality increases this will become hard. The techniques described in this article give an idea on what was done for past problems and how this can be employed to future challenges as well. This should allow to select techniques well for a problem at hand and compare new approaches to strong baselines of existing techniques.

6 Acknowledgments

This work was partially supported by the Swiss National Science Foundation (FNS) in the MANY project (grant 205321–130046), the EU 7^{th} Framework Program in the context of the Khresmoi project (FP7–257528), and the Center for Biomedical Imaging (CIBM).

References

1. YouTube. `http://www.youtube.com/t/press_statistics` (2012). [Online; accessed 14-March-2012]
2. Ahmed, M.N., Farag, A.A.: 3D segmentation and labeling using self-organizing Kohonen network for volumetric measurements on brain CT imaging to quantify TBI recovery. In: Proceedings of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine., *EMBS 1996*, vol. 2, pp. 738–739 (1996)
3. Akbari, H., Yang, X., Halig, L.V., Fei, B.: 3D segmentation of prostate ultrasound images using wavelet transform. In: Medical Imaging 2011: Image Processing, vol. 7962, p. 79622K. SPIE (2011)
4. Akgül, C., Rubin, D., Napel, S., Beaulieu, C., Greenspan, H., Acar, B.: Content–based image retrieval in radiology: Current status and future directions. Journal of Digital Imaging **24**(2), 208–222 (2011)
5. de Alarcón, P., Pascual-Montano, A., Carazo, J.: Spin images and neural networks for efficient content-based retrieval in 3D object databases. In: M. Lew, N. Sebe, J. Eakins (eds.) Image and Video Retrieval, *Lecture Notes in Computer Science*, vol. 2383, pp. 225–234. Springer Berlin / Heidelberg (2002)
6. Amir, A., Basu, S., Iyengar, G., Lin, C.Y., Naphade, M., Smith, J.R., Srinivasan, S., Tseng, B.: A multi-modal system for the retrieval of semantic video events. Computer Vision and Image Understanding **96**(2), 216 – 236 (2004)
7. Amir, A., Berg, M., Chang, S.F., Hsu, W., Iyengar, G., Lin, C.Y., Naphade, M., Natsev, A., Neti, C., Nock, H.J., Smith, J.R., Tseng, B., Wu, Y., Zhang, D.: IBM Research TRECVID-2003 video retrieval system. In: Proceedings of the TRECVID 2003 conference (2003)
8. Andriole, K.P., Wolfe, J.M., Khorasani, R.: Optimizing analysis, visualization and navigation of large image data sets: One 5000–section CT scan can ruin your whole day. Radiology **259**(2), 346–362 (2011)
9. Ankerst, M., Kastenmüller, G., Kriegel, H.P., Seidl, T.: 3D shape histograms for similarity search and classification in spatial databases. In: R. Güting, D. Papadias, F. Lochovsky (eds.) Advances in Spatial Databases, *Lecture Notes in Computer Science*, vol. 1651, pp. 207–226. Springer Berlin / Heidelberg (1999)
10. Ansary, T., Vandeborre, J.P., Mahmoudi, S., Daoudi, M.: A bayesian framework for 3D models retrieval based on characteristic views. In: 3DPVT 2004. Proceedings. 2nd International Symposium on 3D Data Processing, Visualization and Transmission, 2004., pp. 139 – 146 (2004)
11. Antel, S.B., Collins, D.L., Bernasconi, N., Andermann, F., Shinghal, R., Kearney, R.E., Arnold, D.L., Bernasconi, A.: Automated detection of focal cortical dysplasia lesions using computational models of their MRI characteristics and texture analysis. NeuroImage **19**(4), 1748–1759 (2003)
12. Assfalg, J., Bertini, M., Bimbo, A., Pala, P.: Content-based retrieval of 3D objects using spin image signatures. IEEE Transactions on Multimedia **9**(3), 589 –599 (2007)
13. Baum, K.G., Helguera, M., Krol, A.: Fusion viewer: A new tool for fusion and visualization of multimodal medical data sets. Journal of Digital Imaging **21**(SUPPL. 1), S59–S68 (2008)

14. Benedens, O., Busch, C.: Towards blind detection of robust watermarks in polygonal models. Computer Graphics Forum **19**(3), C–199–C–208 (2000)
15. Bennett, W.R., Davey, J.R.: Data transmission. McGraw–Hill (1965)
16. Bhalerao, A., Reyes-Aldasoro, C.: Volumetric texture description and discriminant feature selection for MRI. In: R. Moreno-Díaz, F. Pichler (eds.) Computer Aided Systems Theory - EUROCAST 2003, *Lecture Notes in Computer Science (LNCS)*, vol. 2809, pp. 573–584. Springer Berlin / Heidelberg (2003)
17. Bustos, B., Keim, D.A., Saupe, D., Schreck, T., Vranic, D.V.: Feature-based similarity search in 3D object databases. ACM Computing Surveys **37**(4), 345–387 (2005)
18. Cai, W., Liu, S., Wen, L., Eberl, S., Fulham, M.J., Feng, D.: 3D neurological image retrieval with localized pathology–centric CMRGlc patterns. In: 17th IEEE International Conference on Image Processing, ICIP 2010, pp. 3201–3204 (2010)
19. Chang, K.I., Bowyer, K.W., Flynn, P.J.: Multiple nose region matching for 3D face recognition under varying facial expression. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(10), 1695–1700 (2006)
20. Chen, D.Y., Tian, X.P., Shen, Y.T., Ouhyoung, M.: On visual similarity based 3D model retrieval. Computer Graphics Forum **22**(3), 223–232 (2003)
21. Chen, W., Giger, M.L., Li, H., Bick, U., Newstead, G.M.: Volumetric texture analysis of breast lesions on contrast–enhanced magnetic resonance images. Magnetic Resonance in Medicine **58**(3), 562–571 (2007)
22. Chen, X., Murphy, R.F.: Robust classification of subcellular location patterns in high resolution 3D fluorescence microscope images. In: 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, *EMBC 2004*, vol. 1, pp. 1632–1635 (2004)
23. Chen, Y., Zhou, X.S., Huang, T.: One-class svm for learning in image retrieval. In: Image Processing, 2001. Proceedings. 2001 International Conference on, vol. 1, pp. 34 –37 vol.1 (2001)
24. Cheng, P.C., Yeh, J.Y., Ke, H.R., Chien, B.C., Yang, W.P.: NCTU–ISU's evaluation for the user–centered search task at ImageCLEF 2004. In: Working Notes of the 2004 CLEF Workshop. Bath, England (2004)
25. Cheung, C.P., Godil, A.: A shape-based searching system for industrial components. In: Proceedings of the 15th International Conference on Web 3D Technology, Web3D '10, pp. 151–156. ACM (2010)
26. Chua, C.S., Han, F., Ho, Y.K.: 3D human face recognition using point signature. In: Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, pp. 233 –238 (2000)
27. Chua, C.S., Jarvis, R.: Point signatures: A new representation for 3D object recognition. International Journal of Computer Vision **25**, 63–85 (1997)
28. Cicirello, V., Regli, W.: Machining feature-based comparisons of mechanical parts. In: Shape Modeling and Applications, SMI 2001 International Conference on., pp. 176 –185 (2001)
29. Cooke, E., Ferguson, P., Gaughan, G., Gurrin, C., Jones, G.J.F., Le, H., Lee, H., Marlow, S., Donald, K.M., Mchugh, M., Murphy, N., E, N., Rothwell, R., Smeaton, A.F., Wilkins, P.: Trecvid 2004 experiments in dublin city university. In: Proceedings of the TRECVID 2004 conference (2004)
30. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys **40**(2), 1–60 (2008)
31. Depeursinge, A., Foncubierta-Rodríguez, A., Van De Ville, D., Müller, H.: Lung texture classification using locally—oriented riesz components. In: G. Fichtinger, A. Martel, T. Peters (eds.) Medical Image Computing and Computer Assisted Intervention —- MICCAI 2011, *Lecture Notes in Computer Science*, vol. 6893, pp. 231–238. Springer Berlin / Heidelberg (2011)
32. Depeursinge, A., Foncubierta-Rodríguez, A., Van De Ville, D., Müller, H.: Three–dimensional solid texture analysis and retrieval in biomedical imaging: review and opportunities. Medical Image Analysis (submitted)
33. Depeursinge, A., Müller, H.: Fusion techniques for combining textual and visual information retrieval. In: H. Müller, P. Clough, T. Deselaers, B. Caputo (eds.) ImageCLEF, *The Springer International Series On Information Retrieval*, vol. 32, pp. 95–114. Springer Berlin Heidelberg (2010)
34. Depeursinge, A., Müller, H.: Sensors, medical images and signal processing: Comprehensive multi–modal diagnosis aid frameworks. IMIA Yearbook of Medical Informatics **5**(1), 43–46 (2010)

35. Depeursinge, A., Racoceanu, D., Iavindrasana, J., Cohen, G., Platon, A., Poletti, P.A., Müller, H.: Fusing visual and clinical information for lung tissue classification in high–resolution computed tomography. Artificial Intelligence in Medicine **50**(1), 13–21 (2010)
36. Depeursinge, A., Vargas, A., Gaillard, F., Platon, A., Geissbuhler, A., Poletti, P.A., Müller, H.: Case-based lung image categorization and retrieval for interstitial lung diseases: clinical workflows. International Journal of Computer Assisted Radiology and Surgery **7**(1), 97–110 (2012)
37. Depeursinge, A., Zrimec, T., Busayarat, S., Müller, H.: 3D lung image retrieval using localized features. In: Medical Imaging 2011: Computer–Aided Diagnosis, vol. 7963, p. 79632E. SPIE (2011)
38. Deselaers, T., Weyand, T., Ney, H.: Image retrieval and annotation using maximum entropy. In: Working Notes of the 2006 CLEF Workshop. Alicante, Spain (2006)
39. Dinh, H., Kropac, S.: Multi-resolution spin-images. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 1, pp. 863 – 870 (2006)
40. Do, M.N., Vetterli, M.: Rotation invariant texture characterization and retrieval using steerable wavelet–domain hidden markov models. IEEE Transactions on Multimedia **4**(4), 517–527 (2002)
41. Donald, K., Smeaton, A.: A comparison of score, rank and probability-based fusion methods for video shot retrieval. In: W.K. Leow, M. Lew, T.S. Chua, W.Y. Ma, L. Chaisorn, E. Bakker (eds.) Image and Video Retrieval, *Lecture Notes in Computer Science*, vol. 3568, pp. 592–592. Springer Berlin / Heidelberg (2005)
42. El-Baz, A., Casanova, M., Gimel'farb, G., Mott, M., Switala, A., Vanbogaert, E., McCracken, R.: Dyslexia diagnostics by 3D texture analysis of cerebral white matter gyrifications. In: 19th International Conference on Pattern Recognition, ICPR 2008, pp. 1–4 (2008)
43. El-Mehalawi, M., Miller, R.A.: A database system of mechanical components based on geometric and topological similarity. part ii: indexing, retrieval, matching, and similarity assessment. Computer-Aided Design **35**(1), 95 – 105 (2003)
44. Elad, M., Tal, A., Ar, S.: Content based retrieval of vrml objects: an iterative and interactive approach. In: Proceedings of the sixth Eurographics workshop on Multimedia 2001, pp. 107–118. Springer-Verlag New York, Inc., New York, NY, USA (2002)
45. Fatemi, N., Lalmas, M., Rölleke, T.: How to retrieve multimedia documents described by MPEG-7. In: C. van Rijsbergen, I. Ounis, J. Jose, Y. Ding (eds.) Semantic Web and Information Retrieval (2004)
46. Fehr, J.: Rotational invariant uniform local binary patterns for full 3D volume texture analysis. In: Finnish Signal Processing Symposium (FINSIG), 2007. Oulu, Finland (2007)
47. Fehr, J., Burkhardt, H.: 3D rotation invariant local binary patterns. In: 19th International Conference on Pattern Recognition, 2008, ICPR 2008, pp. 1–4 (2008)
48. Ferecatu, M., Sahbi, H.: TELECOM ParisTech at ImageClefphoto 2008: Bi–modal text and image retrieval with diversity enhancement. In: Working Notes of the 2008 CLEF Workshop. Aarhus, Denmark (2008)
49. Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., Yanker, P.: Query by Image and Video Content: The QBIC system. IC **28**(9), 23–32 (1995)
50. Foncubierta-Rodríguez, A., Depeursinge, A., Müller, H.: Using multiscale visual words for lung texture classification and retrieval. In: H. Greenspan, H. Müller, T. Syeda Mahmood (eds.) Medical Content–based Retrieval for Clinical Decision Support, *MCBR–CDS 2011*, vol. 7075, pp. 69–79. Lecture Notes in Computer Sciences (LNCS) (2012)
51. François, R., Fablet, R., Barillot, C.: Robust statistical registration of 3D ultrasound images using texture information. In: Proceedings of the International Conference on Image Processing, 2003. ICIP 2003., vol. 1, pp. 581–584 (2003)
52. Friedrich, J.M.: Quantitative methods for three–dimensional comparison and petrographic description of chondrites. Computers & Geosciences **34**(12), 1926–1935 (2008)
53. Funkhouser, T., Min, P., Kazhdan, M., Chen, J., Halderman, A., Dobkin, D., Jacobs, D.: A search engine for 3d models. ACM Transactions on Graphics **22**(1), 83–105 (2003)
54. Gallo, L., Pietro, G.D., Coronato, A., Marra, I.: Toward a natural interface to virtual medical imaging environments. In: AVI '08: Proceedings of the working conference on Advanced visual interfaces, pp. 429–432. New York, NY, USA (2008)
55. Gao, D.: Volume texture extraction for 3D seismic visualization and interpretation. Geophysics **68**(4), 1294–1302 (2003)

56. Gao, D.: Texture model regression for effective feature discrimination: Application to seismic facies visualization and interpretation. Geophysics **69**(4), 958–967 (2004)
57. Gao, D.: Latest developments in seismic texture analysis for subsurface structure, facies, and reservoir characterization: A review. Geophysics **76**(2), 1–13 (2011)
58. Gao, X., Qian, Y., Hui, R., Loomes, M., Comley, R., Barn, B., Chapman, A., Rix, J.: Texture–based 3D image retrieval for medical applications. In: IADIS Multi Conference on Computer Science and Information System (MCCSIS) (2010)
59. Greenspan, H., Pinhas, A.T.: Medical image categorization and retrieval for pacs using the gmm-kl framework. IEEE Transactions on Information Technology in Biomedicine **11**(2), 190–202 (2007)
60. Gruhne, M.: Mp7qf: An mpeg-7 query format. In: Automated Production of Cross Media Content for Multi-Channel Distribution, 2007. AXMEDIS '07. Third International Conference on, pp. 15 –18 (2007)
61. Haas, M., Rijsdam, J., Thomee, B., Lew, M.S.: Relevance feedback: perceptual learning and retrieval in bio-computing, photos, and video. In: Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval, MIR '04, pp. 151–156. ACM, New York, NY, USA (2004)
62. Hanjalic, A., Lagendijk, R.L., Biemond, J.: A new method for key frame based video content representation, pp. 97–107. Eds. World Scientific (1997)
63. Hanka, R., Harte, T.P.: Curse of dimensionality: Classifying large multi–dimensional images with neural networks. In: Proceedings of the European Workshop on Computer–Intensive Methods in Control and Signal Processing (CIMCSP1996). Prague, Czech Republic (1996)
64. Healy, D.M., Rockmore, D.N., Kostelec, P.J., Moore, S.S.B.: FFTs for the 2-Sphere - Improvements and Variations. Tech. Rep. TR2002-419, Dartmouth College, Computer Science, Hanover, NH (2002)
65. Hilaga, M., Shinagawa, Y., Kohmura, T., Kunii, T.L.: Topology matching for fully automatic similarity estimation of 3D shapes. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, SIGGRAPH '01, pp. 203–212. ACM, New York, NY, USA (2001)
66. Huisman, A., Ploeger, L.S., Dullens, H.F.J., Jonges, T.N., Belien, J.A.M., Meijer, G.A., Poulin, N., Grizzle, W.E., van Diest, P.J.: Discrimination between benign and malignant prostate tissue using chromatin texture analysis in 3-D by confocal laser scanning microscopy. The Prostate **67**(3), 248–254 (2007)
67. Ip, C.Y., Lapadat, D., Sieger, L., Regli, W.C.: Using shape distributions to compare solid models. In: Proceedings of the seventh ACM symposium on Solid modeling and applications, SMA '02, pp. 273–280. ACM, New York, NY, USA (2002)
68. Isler, V., Wilson, B., Bajcsy, R.: Building a 3D virtual museum of native american baskets. In: Proceedings - Third International Symposium on 3D Data Processing, Visualization, and Transmission, 3DPVT 2006, pp. 954–961 (2007)
69. Iyengar, G., Nock, H.J.: Discriminative model fusion for semantic concept detection and annotation in video. In: Proceedings of the eleventh ACM international conference on Multimedia, MULTIMEDIA '03, pp. 255–258. ACM, New York, NY, USA (2003)
70. Jafari-Khouzani, K., Soltanian-Zadeh, H., Elisevich, K., Patel, S.: Comparison of 2D and 3D wavelet features for TLE lateralization. In: A.A. Amini, A. Manduca (eds.) Medical Imaging 2004: Physiology, Function, and Structure from Medical Images, vol. 5369, pp. 593–601. SPIE (2004)
71. Jerram, D.A., Higgins, M.D.: 3D analysis of rock textures: Quantifying igneous microstructures. ELEMENTS **3**(4), 239–245 (2007)
72. Julesz, B.: Textons, the elements of texture perception, and their interactions. Nature **290**(5802), 91–97 (1981)
73. Kalpathy-Cramer, J., Müller, H., Bedrick, S., Eggel, I., García Seco de Herrera, A., Tsikrika, T.: The CLEF 2011 medical image retrieval and classification tasks. In: Working Notes of CLEF 2011 (Cross Language Evaluation Forum) (2011)
74. Ketcham, R.A.: Computational methods for quantitative analysis of three–dimensional features in geological specimens. Geosphere **1**(1), 32–41 (2005)
75. Kim, J., Cai, W., Feng, D., Wu, H.: A new way for multidimensional medical data management: Volume of interest (voi)-based retrieval of medical images with visual and functional features. IEEE Transactions on Information Technology in Biomedicine **10**(3), 598–607 (2006)

76. Kim, T.Y., Choi, H.J., Hwang, H., Choi, H.K.: Three–dimensional texture analysis of renal cell carcinoma cell nuclei for computerized automatic grading. Journal of Medical Systems **34**(4), 709–716 (2010)
77. Kim, T.Y., Choi, H.K.: Computerized renal cell carcinoma nuclear grading using 3D textural features. In: IEEE International Conference on Communications Workshops, 2009. ICC Workshops 2009., pp. 1–5 (2009)
78. Kontos, D., Bakic, P.R., Carton, A.K., Troxel, A.B., Conant, E.F., Maidment, A.D.A.: Parenchymal texture analysis in digital breast tomosynthesis for breast cancer risk estimation: A preliminary study. Academic Radiology **16**(3), 283–298 (2009)
79. Korfiatis, P.D., Kalogeropoulou, C., Karahaliou, A.N., Kazantzi, A.D., Costaridou, L.I.: Vessel tree segmentation in presence of interstitial lung disease in MDCT. IEEE Transactions on Information Technology in Biomedicine **15**(2), 214–220 (2011)
80. Kotsiantis, S.B.: Supervised machine learning: A review of classification techniques. In: Proceeding of the 2007 conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies, pp. 3–24. IOS Press, Amsterdam, The Netherlands, The Netherlands (2007)
81. Kovalev, V.A., Kruggel, F.: Texture anisotropy of the brain's white matter as revealed by anatomical MRI. IEEE Transactions on Medical Imaging **26**(5), 678–685 (2007)
82. Krefting, D., Bart, J., Beronov, K., Dzhimova, O., J., F., Hartung, M., A., H., Knoch, T.A., Lingner, T., Mohammed, Y., Peter, K., Rahm, E., Sax, U., Sommerfeld, D., Steinke, T., Tolsdorff, T., Vossberg, M., Viezens, F., Weisbecker, A.: Medigrid: Towards a user friendly secured grid infrastructure. Future Generation Computer Systems **25**, 326–336 (2009)
83. Larson, M., Newman, E., Jones, G.: Overview of VideoCLEF 2008: Automatic generation of topic-based feeds for dual language audio-visual content. In: C. Peters, T. Deselaers, N. Ferro, J. Gonzalo, G. Jones, M. Kurimo, T. Mandl, A. Peñas, V. Petras (eds.) Evaluating Systems for Multilingual and Multimodal Information Access, *Lecture Notes in Computer Science*, vol. 5706, pp. 906–917. Springer Berlin / Heidelberg (2009)
84. Larson, M., Newman, E., Jones, G.: Overview of VideoCLEF 2009: New perspectives on speech-based multimedia content enrichment. In: C. Peters, B. Caputo, J. Gonzalo, G. Jones, J. Kalpathy-Cramer, H. Müller, T. Tsikrika (eds.) Multilingual Information Access Evaluation II. Multimedia Experiments, *Lecture Notes in Computer Science*, vol. 6242, pp. 354–368. Springer Berlin / Heidelberg (2010)
85. Lee, K.., Ho, J., Kriegman, D.J.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(5), 684–698 (2005)
86. Lienhart, R.: Reliable transition detection in videos: A survey and practitioner's guide. International Journal of Image and Graphics **1**, 469–486 (2001)
87. Lillis, D., Toolan, F., Mur, A., Peng, L., Collier, R., Dunnion, J.: Probability-based fusion of information retrieval result sets. Artificial Intelligence Review **25**, 179–191 (2006)
88. Loffler, J.: Content-based retrieval of 3D models in distributed web databases by visual shape information. In: Information Visualization, 2000. Proceedings. IEEE International Conference on, pp. 82 –87 (2000)
89. Lopes, R., Ayache, A., Makni, N., Puech, P., Villers, A., Mordon, S., Betrouni, N.: Prostate cancer characterization on MR images using fractal features. Medical Physics **38**(1), 83–95 (2011)
90. Marchand-Maillet, S.: Content–based video retrieval: An overview. Tech. Rep. 00.06, CUI - University of Geneva, Geneva, Switzerland (2000)
91. Mariolis, I., Korfiatis, P.D., Costaridou, L.I., Kalogeropoulou, C., Daoussis, D., Petsas, T.: Investigation of 3D textural features' discriminating ability in diffuse lung disease quantification in MDCT. In: IEEE International Conference on Imaging Systems and Techniques, IST 2010., pp. 135–138 (2010)
92. Mezaris, V., Kompatsiaris, I., Boulgouris, N., Strintzis, M.: Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval. Circuits and Systems for Video Technology, IEEE Transactions on **14**(5), 606 – 621 (2004)
93. Mitra, N.J., Guibas, L.J., Pauly, M.: Partial and approximate symmetry detection for 3D geometry. In: ACM Transactions on Graphics, vol. 25, pp. 560–568 (2006)
94. Moënne-Loccoz, N., Janvier, B., Marchand-Maillet, S., Bruno, E.: Handling temporal heterogeneous data for content-based management of large video collections. Multimedia Tools and Applications **31**, 309–325 (2006)

95. Müller, H., Clough, P., Deselaers, T., Caputo, B. (eds.): ImageCLEF – Experimental Evaluation in Visual Information Retrieval, *The Springer International Series On Information Retrieval*, vol. 32. Springer, Berlin Heidelberg (2010)
96. Müller, H., Kalpathy-Cramer, J.: Analyzing the content out of context — features and gaps in medical image retrieval. International Journal on Healthcare Information Systems and Informatics **4**(1), 88–98 (2009)
97. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content–based image retrieval systems in medicine–clinical benefits and future directions. International Journal of Medical Informatics **73**(1), 1–23 (2004)
98. Nallapati, R.: Discriminative models for information retrieval. In: ACM-SIGIR (2004)
99. Nguyen, D.., Kuhnert, L., Jiang, T., Thamke, S., Kuhnert, K..: Vegetation detection for outdoor automobile guidance. In: Proceedings of the IEEE International Conference on Industrial Technology, pp. 358–364 (2011)
100. Ohbuchi, R., Otagiri, T., Ibato, M., Takei, T.: Shape-similarity search of three-dimensional models using parameterized statistics. In: Computer Graphics and Applications, 2002. Proceedings. 10th Pacific Conference on, pp. 265 – 274 (2002)
101. Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D.: Shape distributions. ACM Transactions on Graphics **21**(4), 807–832 (2002)
102. Paulhac, L.: Outils et méthodes d'analyse d'images 3D texturées : application à la segmentation des images échographiques. Ph.D. thesis, Université François Rabelais – Tours, France (2009)
103. Paulhac, L., Makris, P., Gregoire, J.M., Ramel, J.Y.: Approche multirésolution pour la segmentation de textures dans les images ultrasonores 3D. In: XXIIe colloque GRETSI (traitement du signal et des images). Dijon, France (2009)
104. Paulhac, L., Makris, P., Gregoire, J.M., Ramel, J.Y.: Descripteurs de textures pour la segmentation d'images echographiques 3D. In: ORASIS'09 — Congrès des jeunes chercheurs en vision par ordinateur. Trégastel, France (2009)
105. Paulhac, L., Makris, P., Ramel, J.Y.: Comparison between 2D and 3D local binary pattern methods for characterisation of three–dimensional textures. In: Proceedings of the 5th international conference on Image Analysis and Recognition, ICIAR '08, pp. 670–679. Springer-Verlag, Berlin, Heidelberg (2008)
106. Pietroni, N., Cignoni, P., Otaduy, M.A., Scopigno, R.: Solid–texture synthesis: A survey. IEEE Computer Graphics and Applications **30**(4), 74–89 (2010)
107. Qian, Y., Gao, X., Loomes, M., Comley, R., Barn, B., Hui, R., Tian, Z.: Content–based retrieval of 3D medical images. In: The Third International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED 2011), pp. 7–12. IARIA (2011)
108. Ranguelova, E., Quinn, A.: Analysis and synthesis of three–dimensional Gaussian Markov random fields. In: Proceedings of the IEEE International Conference on Image Processing, *ICIP 99*, vol. 3, pp. 430–434 (1999)
109. Reyes-Aldasoro, C.C., Bhalerao, A.: Volumetric texture segmentation by discriminant feature selection and multiresolution classification. IEEE Transactions on Medical Imaging **26**(1), 1–14 (2007)
110. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision **40**(2), 99–121 (2000)
111. Samir, C., Srivastava, A., Daoudi, M.: Three-dimensional face recognition using shapes of facial curves. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(11), 1858–1863 (2006)
112. Saupe, D., Vranić, D.: 3D model retrieval with spherical harmonics and moments. In: B. Radig, S. Florczyk (eds.) Pattern Recognition, *Lecture Notes in Computer Science*, vol. 2191, pp. 392–397. Springer Berlin / Heidelberg (2001)
113. Sebe, N., Lew, M.S.: Texture features for content–based retrieval, pp. 51–85. Springer–Verlag, London, UK (2001)
114. Shen, L., Bai, L.: 3D Gabor wavelets for evaluating SPM normalization algorithm. Medical Image Analysis **12**(3), 375–383 (2008)
115. Shibata, T., Suzuki, M., Kato, T.: 3D retrieval system based on cognitive level - human interface for 3D building database. In: Proceedings - 2004 International Conference on Cyberworlds, CW 2004, pp. 107–112 (2004)
116. Shilane, P., Min, P., Kazhdan, M., Funkhouser, T.: The princeton shape benchmark. In: Shape Modeling Applications, pp. 167–178. Genova, Italy (2004)
117. Sikora, T.: The mpeg-7 visual standard for content description-an overview. Circuits and Systems for Video Technology, IEEE Transactions on **11**(6), 696 –702 (2001)

118. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content–based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(12), 1349–1380 (2000)
119. Snoek, C.G., Worring, M.: Concept-based video retrieval. Foundations and Trends in Information Retrieval **2**(4), 215–322 (2008)
120. Snoek, C.G.M., Worring, M., Smeulders, A.W.M.: Early versus late fusion in semantic video analysis. In: MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia, pp. 399–402. ACM, New York, NY, USA (2005)
121. Sundar, H., Silver, D., Gagvani, N., Dickinson, S.: Skeleton based shape matching and retrieval. In: Shape Modeling International, 2003, pp. 130 – 139 (2003)
122. Suzuki, M.T., Kato, T., Otsu, N.: Similarity retrieval of 3D polygonal models using rotation invariant shape descriptors. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, vol. 4, pp. 2946–2952 (2000)
123. Tangelder, J.W.H., Veltkamp, R.C.: A survey of content based 3D shape retrieval methods. In: Proceedings - Shape Modeling International SMI 2004, pp. 145–156 (2004)
124. Thornley, C.V., Johnson, A.C., Smeaton, A.F., Lee, H.: The scholarly impact of TRECVid (2003-2009). Journal of the American Society for Information Science and Technology **62**(4), 613–627 (2011)
125. Toussaint, G.T.: The use of context in pattern recognition **10**(3), 189–204 (1978)
126. Tsai, F., Chang, C.K., Rau, J.Y., Lin, T.H., Liu, G.R.: 3D computation of gray level co–occurrence in hyperspectral image cubes. In: A. Yuille, S.C. Zhu, D. Cremers, Y. Wang (eds.) Energy Minimization Methods in Computer Vision and Pattern Recognition, *Lecture Notes in Computer Science (LNCS)*, vol. 4679, pp. 429–440. Springer Berlin / Heidelberg (2007)
127. Veltkamp, R.C., Ruijsenaars, R., Spagnuolo, M., van Zwol, R., ter Haar, F.: SHREC2006 3D shape retrieval contest. Tech. rep., Department of Information and Computing Sciences, Utrecht University (2006)
128. Venkatesh Babu, R., Ramakrishnan, K.: Content-based video retrieval using motion descriptors extracted from compressed domain. In: Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on, vol. 4, pp. IV–141 – IV–144 vol.4 (2002). DOI 10.1109/ISCAS.2002.1010409
129. Viaud, M.l., Buisson, O., Saulnier, A., Guenais, C.: Video exploration: from multimedia content analysis to interactive visualization. In: Proceedings of the international conference on Multimedia, MM '10, pp. 1311–1314. ACM, New York, NY, USA (2010)
130. Vranic, D., Saupe, D.: Description of 3D-shape using a complex function on the sphere. In: ICME '02. Proceedings. 2002 IEEE International Conference on Multimedia and Expo, 2002., vol. 1, pp. 177 – 180 vol.1 (2002)
131. Vranic, D.V., Saupe, D., Richter, J.: Tools for 3D-object retrieval: Karhunen-loeve transform and spherical harmonics. In: 2001 IEEE Fourth Workshop on Multimedia Signal Processing, pp. 293–298 (2001)
132. Waksman, A., Rosenfeld, A.: Sparse, opaque three–dimensional texture, 2A: Visibility. Graphical Models and Image Processing **58**(2), 155–163 (1996)
133. Waksman, A., Rosenfeld, A.: Sparse, opaque three–dimensional texture, 2B: Photometry. Pattern Recognition **29**(2), 297–313 (1996)
134. Wang, X., Tang, X.: A unified framework for subspace face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(9), 1222–1228 (2004)
135. Westerveld, T., Ianeva, T., Boldareva, L., de Vries, A.P., Hiemstra, D.: Combining information sources for video retrieval – the lowlands team at trecvid 2003. In: Proceedings of the TRECVID 2003 conference (2003)
136. Wong, H.S., Cheung, K.K., Ip, H.H.: 3D head model classification by evolutionary optimization of the extended gaussian image representation. Pattern Recognition **37**(12), 2307 – 2322 (2004)
137. von Wyl, M., Mohamed, H., Bruno, E., Marchand-Maillet, S.: A parallel cross-modal search engine over large-scale multimedia collections with interactive relevance feedback. In: Demo at ACM International Conference on Multimedia Retrieval (ACM-ICMR'11). Trento, Italy (2011)
138. Xu, D.H., Kurani, A.S., Furst, J., Raicu, D.S.: Run–length encoding for volumetric texture. In: The 4th IASTED International Conference on Visualization, Imaging, and Image Processing – VIIP 2004. Marbella, Spain (2004)
139. Xu, Y., Sonka, M., McLennan, G., Guo, J., Hoffman, E.A.: Sensitivity and specificity of 3–D texture analysis of lung parenchyma is better than 2–D for discrimination of lung

pathology in stage 0 COPD. In: A.A. Amini, A. Manduca (eds.) SPIE Medical Imaging, vol. 5746, pp. 474–485. SPIE (2005)
140. Xu, Y., Sonka, M., McLennan, G., Guo, J., Hoffman, E.A.: MDCT–based 3D texture classification of emphysema and early smoking related lung pathologies **25**(4), 464–475 (2006)
141. Yang, J., Jiang, Y.G., Hauptmann, A.G., Ngo, C.W.: Evaluating bag-of-visual-words representations in scene classification. In: Proceedings of the international workshop on Workshop on multimedia information retrieval, MIR '07, pp. 197–206. ACM, New York, NY, USA (2007)
142. Yang, X., Schuster, D., Master, V., Nieh, P., Fenster, A., Fei, B.: Automatic 3D segmentation of ultrasound images using atlas registration and statistical texture prior. In: Medical Imaging 2011: Visualization, Image–Guided Procedures, and Modeling, vol. 7964, p. 796432. SPIE (2011)
143. van Zaanen, M., de Croon, G.: FINT: Find Images aNd Text. In: Working Notes of the 2004 CLEF Workshop. Bath, England (2004)
144. Zhan, Y., Shen, D.: Deformable segmentation of 3–D ultrasound prostate images using statistical texture matching method. IEEE Transactions on Medical Imaging **25**(3), 256–272 (2006)
145. Zhang, L., Samaras, D.: Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(3), 351–363 (2006)
146. Zhao, T., Nevatia, R.: Tracking multiple humans in complex situations. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(9), 1208–1221 (2004)
147. Zhou, X., Depeursinge, A., Müller, H.: Information fusion for combining visual and textual image retrieval. In: 20th IEEE International Conference on Pattern Recognition (ICPR), pp. 1590–1593 (2010)