Pattern Analysis and Applications (2019) 22:1243–1292
https://doi.org/10.1007/s10044-019-00804-4

SURVEY

3D object recognition and classification: a systematic literature review

L. E. Carvalho[1,2] · **A. von Wangenheim**[1,2]

Received: 3 October 2017 / Accepted: 14 February 2019 / Published online: 27 February 2019

Abstract
In this paper, we present a systematic literature review concerning 3D object recognition and classification. We cover articles published between 2006 and 2016 available in three scientific databases (ScienceDirect, IEEE Xplore and ACM), using the methodology for systematic review proposed by Kitchenham. Based on this methodology, we used tags and exclusion criteria to select papers about the topic under study. After the works selection, we applied a categorization process aiming to group similar object representation types, analyzing the steps applied for object recognition, the tests and evaluation performed and the databases used. Lastly, we compressed all the obtained information in a general overview and presented future prospects for the area.

Keywords 3D object recognition · 3D object classification · 3D object representation · Systematic literature review

1 Introduction

The 3D object classification and recognition area, in the last few years, experienced a growing boosted by the popularization of 3D sensors and the increased availability of 3D object databases [41]. The methods developed for this purpose are applied in many domains like robotics, focusing on assisting the robot movement in an environment and performing object manipulation, security, where the techniques are employed to detect dangerous objects, and 3D general object detection, recognizing, for example, objects, faces, ears and so on.

The seminal works interested in solving 3D/multi-view object recognition as well as pose estimation started in the 1980s and early 1990s [22, 85, 173, 188, 203, 251, 293], which, for some authors [48], were considered the foundation of modern object recognition. A detailed history of 3D object recognition can be found in the book Computer Vision Detection, Recognition and Reconstruction [48]. This book contains not only a history about works interest in solving 3D object recognition problems, but also a selection of articles covering some of the talks and tutorials held during the first two editions of ICVSS (The International Computer Vision Summer School) on topics such as Recognition, Registration and Reconstruction. Each chapter provides an overview of these challenging topics with key references to the existing literature until 2009.

In order to provide a panorama of the area, identifying 3D object classification and recognition methods and the representation types or descriptions employed by those methods, we performed a systematic literature review. For this purpose, we employed Kitchenham's methodology [140] for systematic literature review, which is a well-established methodology.

The objectives of this paper are to: (1) present the methodology applied for the present systematic literature review and (2) perform an overall analysis aiming to identify the 3D object representation types, the general structure used in the analyzed works and how the evaluation and validation were performed.

The novelty in this review is twofold: First, to the author's knowledge, this is the first time that Kitchenham's methodology is applied for 3D object recognition. Second, differently from previous reviews for 3D object recognition, where the search was restricted to a specific type of method, we defined a 10-year window as our only search restriction

✉ L. E. Carvalho
lcarvalho@incod.ufsc.br

A. von Wangenheim
aldo.vw@ufsc.br

[1] Graduate Program in Computer Science, Federal University of Santa Catarina, Florianópolis, Brazil

[2] Image Processing and Computer Graphics Lab, National Brazilian Institute for Digital Convergence, Federal University of Santa Catarina, Florianópolis, Brazil

(2006–2016), analyzing all the works related to 3D object recognition in this period.

This review is organized as follows: Section 2 presents all the steps performed for the systematic literature review and details all the search parameters as well as the work selection criteria used. Section 3 shows an analysis of all the selected works, categorizing each work by the 3D object representation employed. Section 4 merges all the gathered information in a general overview. Lastly, the conclusion and future prospects for the area are given in Sect. 5.

2 Methodology and research description

This review is based on the methodology described by Kitchenham [140] for systematic literature review. Kitchenham's methodology has its objective "evaluate and interpret all the relevant researches available to a particular research question, topic or phenomenon of interest". Three guides, used for research in the healthcare field, were the base for Kitchenham's methodology [49, 99, 100, 244], and its principal feature is to keep the search reproducible when the same steps, keywords and tags are employed. The aforementioned methodology starts with the research question or topic definition. Then, keywords and search tags are defined to search works on scientific databases. Lastly, the exclusion criteria are applied for the selection of works that will be analyzed. Based on this methodology, we initially defined a research topic: **3D object recognition and classification**. After that, we defined keywords and tags, applying them on three scientific databases (ScienceDirect, ACM and IEEE) as follows:

ScienceDirect: *2017 < pub-date and pub-date > 2005 and TITLE-ABSTR-KEY("3D object classification") or TITLE-ABSTR-KEY("3D object recognition")*

ACM: *"3D object recognition" "3D object classification" published between 2006–2016*

IEEE: *("3D object recognition") OR ("3D object classification") and refined by Year: 2006–2016*

We obtained a total of 446 papers from a search with these keywords and tags. We analyzed the abstract of these papers, employing the following exclusion criteria:

- Papers written in other languages different from English;
- Repeated papers or papers that do not describe, in their abstracts, techniques for 3D object classification and/or recognition.
- Papers that use software for the recognition and/or classification part and do not focus on the employed techniques, citing only their use.

After the application of these exclusion criteria, a total of 277 papers were obtained, which were further analyzed. From these 277 works, 12 were related to books, 3D object database representation and methods evaluation and were, therefore, separated from the categorization process according to the object representation employed. The remaining 265 works were grouped into object representations categories, based on the works presented by Mhamdi [195] and Atmosukarto [20], extending the presented categories as new object representation methods appeared.

3 Analyzed Works

All the selected works were read, analyzed and categorized according to its main 3D object representation type. The performed analysis focused on how the objects are represented, which data type was used and how the experiments were conducted, taking into account the use of publicly available databases and the procedures for evaluation and results demonstration. Due to the extensive quantity of information, we built a technical report [40] composed of images and overviews for each analyzed work. In this review, we compressed the information obtained from such analysis, giving an overview from the 3D object recognition and classification area ranging from 2006 to 2016. The following sections demonstrate some object representations employed by the analyzed works.

3.1 Feature-based representations

Representations based on features are by far the most frequent type of representation form. Each 3D object has several features that along the last years were explored, pursuing a way of representing objects, which aggregates some very important characteristics such as high description and discrimination capacity, fast computation and low memory consumption, the last two almost mandatory for real-time applications in the robotic area. This search for features that better discriminate a 3D object resulted in several descriptors and keypoint detectors, feature vector constructions and pattern learning approaches to represent the object appearances, its geometric features and shapes. In order to better separate the feature-based representation, we categorized this representation type in the following sub-categories, as described in [20]: local features, global features and spatial maps. The global features distribution, defined by Atmosukarto, was grouped into the global features due to the categories similarity. We created another sub-category, global and local features, because some works employ both feature types to represent the objects. Each sub-category will be further analyzed in the following sections.

3.1.1 Local features

Local features are frequently described by points that are salient or encode information from the 3D object. This feature type, inside the feature-based representation, is the most frequently used, in the last few years, due to its "stronger discriminating power when differentiating objects that are similar in overall shape" [20]. Several ways of computing this feature type were presented along the years, and we will present some of them based on the analyzed works overview. Figure 1 shows some examples of works that employ local features for the object representation.

The scale-invariant feature transform (SIFT), presented by Lowe [174], used for describing salient points (keypoints) and representing the objects, was employed in several analyzed works as a form to extract keypoints. In some works, the extracted keypoints were tracked over an image sequence obtaining time-invariant features, which were used to train a classifier, e.g., support vector machine (SVM), for further object recognition [16, 54]. Some works used SIFT for detection and description associated with other methods, without the time-invariant construction [39, 55, 108, 124, 143, 212, 271].

Other works used SIFT associated with other feature descriptors or extended the SIFT descriptor: Sukhan Lee [154] merged the SIFT with 3D lines (geometric feature); Matas extracted the features using maximally stable extremal regions (MSER) [189], followed by a region's appearance characterization, by means of SIFT features, and the region outline, through the modified Fourier descriptor (MFD) [252]; Kim [135] employed a SIFT generalization extending the feature detector with the Harris corners detector; Salgian [257] analyzed the SIFT performance against two region descriptors (PCA-SIFT [123] and keyed context patches [264]), in the context of 3D object recognition. Later, he showed that by combining local image descriptors at feature level, the recognition performance can be significantly improved [258], where SIFT [175] and keyed context patches [264] were the features descriptors employed; Kim [133] combined the features such as SIFT, line and color, in an evidence selection and collection framework based on the Bayes' theorem; Naikal [205] also combined image features, such as SIFT features, to construct a vocabulary tree, which is employed in a distributed object recognition system for 3D object detection; Usui [297] used a SIFT extension, called affine SIFT [200], in a point-based 3D object recognition system, where the features are extracted from an input image and then matched with a database model descriptor; Rodner [247] presented an approach for generic object recognition, with depth and colored image information, using a SIFT variation, Opponent-SIFT features [261], for color feature encoding; Tangruamsub [281] combined two types of information, spatial and appearance information (SIFT and SURF, respectively), for 3D object recognition in real environments; Jeong [114] presents a feature selection method, with statistical modeling in real environments, for 3D object recognition and pose estimation, where some features were tested (SIFT, line [176] and color); Nakashika

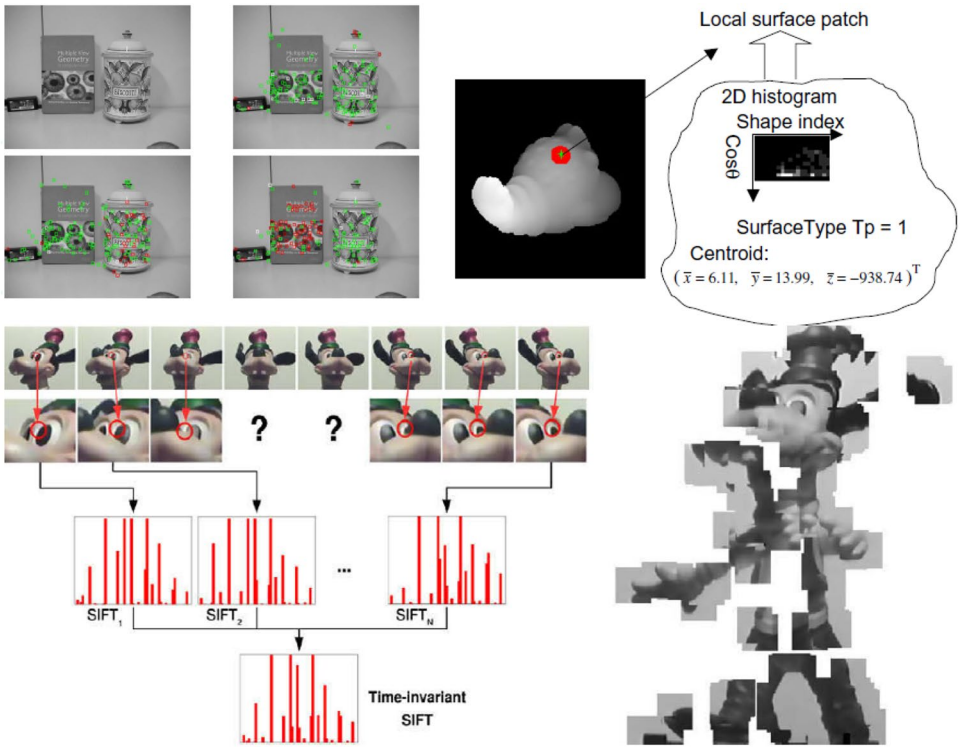

Fig. 1 Representation of the analyzed works that employ local features. Composition of images extracted from [16, 42, 212]

[208] combined the histograms of oriented normal vectors (HONVs), extracted from the depth image, and the SIFT, extracted from the RGB images, encoding the features using the locality-constrained linear coding (LLC), for 3D object recognition in RBG-D images; Soysal [276] presented an approach that can be seen as the integration between the invariant based on a detailed geometric approach presented in [306] with the modern methods based on local appearance, such as SIFT; Flitton [71] evaluated the object classification model, Bag of Words (BoW), performance as an approach for automatic dangerous objects detection in computed tomography 3D images. In order to do so, the combination of four 3D feature descriptors (density histogram (DH), density gradient histogram (DGH), SIFT and rotation-invariant feature transform (RIFT)), three codebook allocation methodologies (hard, kernel and uncertainty) and seven codebook sizes within a supervised machine learning framework, based on the support vector machine (SVM) classifier, were explored;

The Unique Signatures of Local Area Description (SHOT) descriptor [288], presented by Federico Tombari, was also employed in several analyzed works. The SHOT is a 3D descriptor that encodes histograms of the normals of the points within the support, which are the most representative surface local structures compared to plain 3D coordinates. Some analyzed works used the SHOT descriptor as originally proposed: Aldoma [10] employed the SHOT descriptor in his work, proposing a method to check hypotheses for 3D object recognition. In his method, after the keypoints were extracted at positions sampled uniformly on scene and model surfaces, and the SHOT descriptor computed, the calculated descriptors are matched to achieve a point-to-point match. This proposed method was also used to present a global hypothesis checking framework in cluttered scenes; Rodola [248] presented a framework for the recognition of known 3D objects in incomplete and cluttered scans, which employs the SHOT descriptor to construct the initial candidates that, in turns, are fed into a matching game. In general, a matching game [7] can be constructed by defining only four basic entities: a model points set, a data points set, a set of matching candidates and a pairwise compatibility function between them. The game's purpose is to operate a selection between the initial matches set elements, based on the entities, strategies and rewards linked to the strategies in order to achieve a equilibrium. This approach was also used in an earlier work, where the authors used a noncooperative game for the 3D object detection in cluttered scenes [6]; Palossi [223] presented a SHOT optimization through GPU-oriented programming, aiming to achieve real-time processing in typical 3D data sets; Gomes [82] employed the SHOT descriptor to propose the use of a mobile fovea approach, which reduces the 3D data sampling and the system object recovery processing of a point cloud; Rangel [239] proposed the use of a growing neural gas (GNG) [72] in order to reduce 3D point cloud noise and improve the 3D object recognition process. To do so, the recognition pipeline uses a uniform sampling method [255] as keypoint detector and the SHOT as object descriptor. On the other hand, a SHOT descriptor variation was proposed by Shaiek [268], which is inspired by the descriptors CSHOT and SHOT and named IndSHOT. The IndSHOT is composed by the model ID, index shape juxtaposed with the cosines histogram, surface type, keypoints 3D coordinates, mean shape index and standard deviation values.

Apart from those two forms of keypoint description, other descriptors and methods were also analyzed and are described below: Aiming to solve time consumption problem in car assembly lines, Pichler [231, 232] proposed a recognition scheme based on 3D models, in which the correspondence between model and scene spin Images is searched. More specifically, a technique was used to recognize keypoints in the spin images in order to find the most representative 3D model points and the scene objects.

Kushal [148] proposes a method to automatically construct 3D object models consisting of a dense set of small surface fragments and texture pattern descriptors from some stereo pairs. For the low-level image features description, the detector affine region, proposed by Mikolajczyk and Schmid [199], was used.

Lin [167] proposes a framework for summation invariants, along with four important summation invariant classes. These invariants are used to define a format representation for several applications, one of which is the 3D object recognition. In this application, surfaces summation invariants under both transformations, Euclidean and affine, are derived, and an algorithm for 3D face recognition, based on these invariants, is proposed.

Chen [46] proposes a geometric hashing method extension for 3D object recognition under perspective transformation. In this extension, 3D object aspects and constrained geometric structures are used to construct a hash table. The procedure for constructing the 3D object models hash table is fully described in Chen Zhe's work. For the recognition part, the following steps are performed: 2D image object line features extraction, constrained geometric structures search from the obtained lines and perspective invariants calculation from the structure; use of those invariants to index the hash table and candidate models selection according to the number of votes for each model; projection perspective matrix computation and use, transforming the 3D aspects into 2D image spaces; line features correspondence degree calculation and object identification according to the correspondence consistency. If the features do not match, the last two steps are repeated until all candidate models are verified.

Arana-Daniel [15] presents a detailed theory of the Clifford support vector machines (CSVMs) and its application

in classifying 3D objects, derived from point clouds, in multiple classes. The CSVM is introduced as a support vector machine generalization using Clifford's geometric algebra. The author explains the CSVM full foundation and theory and shows an experiment to classify objects in multiple classes with artificial and real training data.

Okada [215] describes the design and implementation of a knowledge-based 3D object recognition system and multitrack integration using a particle filter technique. To recognize the objects, the following knowledge of visual cues is defined: 3D object format information, object surface color histogram and straight edges visible on the object surface. These computed visual cues are integrated in a particle filter algorithm, which is widely used in tracking objects due to their robust features. The 3D object information is generated by two steps: First, 2D feature points are generated using the KLT feature extraction method. Then, a correlation-based stereo match is applied to calculate the feature points' disparity and to obtain a 3D distance from the points.

Li [164] proposes a method for recognizing depth images using supervised learning, which measures the similarity between depth images using their feature sets. The proposed method works as follows: For each depth image, the first image points are selected, which are characteristic and have salient geometric information. Each salient point is combined with a surface descriptor defined in the local surface patch near the point. After detecting the salient points and computing their local signatures, the depth images are represented as a set of non-ordered surface descriptors. A pyramid match kernel function is then used to measure the similarity between non-ordered feature sets. Finally, given a set of n labeled images classes, n classifiers are learned, using a similarity in pairs between these images, where each classifier separates one image class from the others. The input image can be submitted to these trained SVM classifiers, recognizing the object based on the most similar class.

Chen [42] presents a local descriptor for surface representation and 3D object recognition. The method presented has two stages: model construction and recognition. Initially, the feature points are defined in areas of wide format variation and measured through the format index. Then, for each feature point, the local surface patches (LSP), the surface type and a 2D histogram are calculated. This process is repeated for each model object to construct the models database. For the recognition part, the processes of extracting the feature points and calculating the LSPs are repeated. Then, one vote is placed in the hash table if the dissimilarity between the model and test LSP histograms falls within a threshold and if the surface type is the same. The model with the highest number of votes is indicated as the test object type. The author also used the same descriptor associated with a SVM classifier [43], where the hypotheses are ranked using the SVM learning algorithm ranking to generate a short list of model candidates for verification. The verification process is performed through the iterative closest point (ICP) application between the respective surfaces (Fig. 2).

Lee [159] presents a two-stage method for recognizing 3D objects based on 2D images. Initially, the image is described in terms of curvature scale space (CSS). The CSS is based on a 2D binary image to represent a parameterized closed curve shape at multiple scales. Using this representation, a phase correlation method is applied to form a new CSS image type. Based on this new representation, a feature vector is assembled by adding a separate sum of each arc size and scale, concatenating on a new vector called marginal-sum vector. The two-stage recognition part begins with the attempt to identify the test object category. To do so, the test object feature vector is projected in each category eigenspace in turns. The category eigenspace that provides the closest reconstruction to the test object feature vector is defined as the test object category. Once the category has been identified, the test object that best matches with objects in the category can be determined by calculating the Euclidean distance between the objects' points in the eigenspace.

Kietzmann [129] presents a new generalized relevance learning vector quantization (GRLVQ) used for object recognition. The proposed variation is an incremental GRLVQ version, the iGRLVQ, which allows an automatic selection of prototypes (codebook size) for each class. This automatic selection is performed through the initialization with only one prototype and the subsequent addition of new

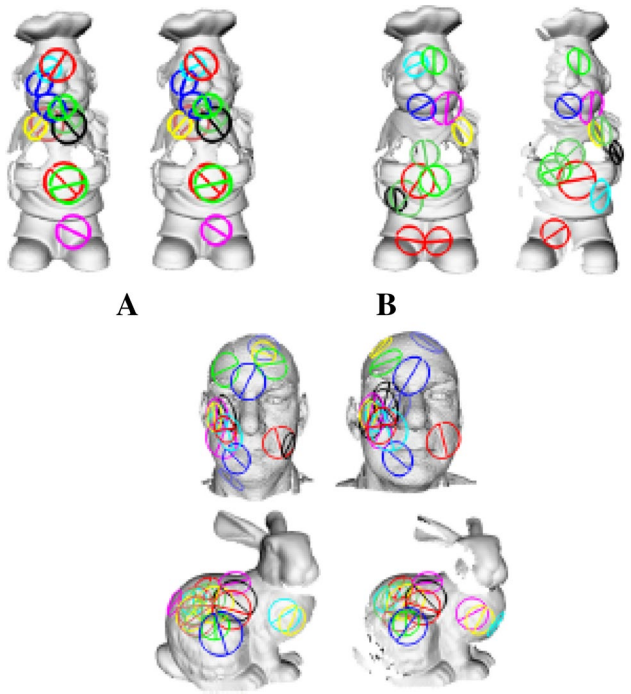

Fig. 2 Representation of the method proposed by Xinju Li. Figure extracted from [164]

representations for each object instance. Due to this on demand recruitment, comparable sparse models are created. Integrating the prototype-based learning mechanism with a generic set of visual features, embedded with the ability to learn real-time features, an effective architecture for object recognition is built. Based on the architecture, the recognition system called Feature and Incremental Learning of Objects Utility (FILOU) is presented, which classifies the target objects based on the performed training.

Zhong [336] presents an approach for 3D object recognition using the format representation method called intrinsic shape signatures (ISSs). The ISS consists of an intrinsic reference frame, which allows the view-invariant features extraction, a quick object pose registration and a discriminating feature vector encoding the 3D format features. The procedure of matching two ISSs is performed by comparing the format feature vectors. The pose estimation is performed directly by using the intrinsic reference frame.

The work presented by Kim [134, 136] proposed a method of scalable 3D object representation and a learning method for the recognition of everyday objects. The proposed method is based on the common frame constellation model (CFCM) and provides advantages in terms of computational efficiency and redundancy removal by sharing the object's view parameters. The local feature used is composed of an appearance vector and a local pose information. The appearance vector is generated by clustering the local features set extracted from the training images. Each training image is represented by a CFCM, where one part has an index for the appearance library and the other part contains information related to the reference frame. The local feature detection part is performed by the generalized robust-invariant feature (G-RIF), and the appearance clustering is performed through the combination of bottom-up and top-down methods. With the appearance library and CFCMs built, the next step is the object recognition. In this step, the scene is also represented by a set of CFCMs, and the corresponding pairwise hypotheses are constructed based on Hough transform. The generated hypotheses are accepted or rejected based on the bin size with an optimal threshold value.

Ho [104] presents a method for extracting local salient features from 3D models using their surface curvature. The surface curvature is calculated using the proposed multi-scalar algorithm, which is based on a local curvature measure, invariant to rotation and translation, known as curvedness. Different point curvedness values are calculated at multiple scales by fitting a surface of different sizes to its neighborhood. A set of reliable salient feature points is formed by the input surface space-scale representation set search. A method for assessing the confidence at each keypoint, based on the neighborhood curvedness values deviation, is also proposed. Tests with the proposed method were performed in a variety of 3D models with different noise levels in order to demonstrate the method effectiveness and robustness for salient feature extraction.

Elizabeth Gonzales [84] describes a method for 3D object recognition based on Fourier description clustering. The approach can be divided into two parts: The first part consists of Fourier description clusters calculation and database storing. The second part focuses on the object recognition, where the main steps for this stage are the following: Fourier descriptors calculation, classification process (discrimination), candidate selection, pose calculation and next best vision (NBV) algorithm utilization.

Gibbins [81] proposes and evaluates the local metrics use (such as Zernike moments, curvature, color representations and spin images, generally employed in 3D object recognition) for terrain structures classification. Initially, the author presents several local feature types to be used and evaluated in the land classification process. Then, real data samples (collected using a low-cost LADAR scanner and optical load) were used in order to evaluate the proposed feature types. These data have been converted into a point cloud with associated color information and were manually classified into four terrain types based on field observations. For each selected feature type, feature estimations were computed under spherical neighborhoods of 1,2 and 4 m. To minimize any correlation between test and training data, the terrain data were separated into two independent sets for training and testing. In the experiments performed, a classifier by vector quantization (VQ) was used in conjunction with linear discriminant analysis (LDA) to reduce the features.

Himmelsbach [102] describes a perception system based on LIDAR for robot mobility. The perception system is divided into three main steps: segmentation, classification and tracking. The segmentation is performed on an occupation grid, providing connected components of grid cells not belonging to the ground surface, i.e., objects. In the classification step, features of an one object identified in the previous step are extracted, capturing the local distribution of spatial properties and the reflectivity extracted on a support volume with fixed size around each point. To obtain a compact object description for classification, histograms are constructed based on features computed for each object point. Then, in a supervised learning framework, a SVM is trained to discriminate interest classes, based on manually labeled examples.

Kao [119] proposes a system for ship recognition with the purpose of researching and developing an effective ship contour capture at sea. The proposed system architecture starts with the contour features extraction using the gradient vector flow (GVF) method. The obtained contour, using the GVF, is not always closed; therefore, a closing process must be performed, while the edge contour is obtained. In the closing process, the Bresenham's line method is used to

connect two separate points. Then, the geometric eigenvalues are calculated after obtaining the image detailed contour. The object's complexity and the ratio between the longest and shortest object axes are employed as criteria for image recognition. All these geometric eigenvalues are utilized for rough matching, and the Fourier descriptor is used to perform a finer matching against database images.

The work presented by Mian [196] extends a work previously presented by the author and proposes a multiple scale keypoint detection algorithm for invariant local feature extraction. In relation to the work previously presented, the following differences can be highlighted: While the previous work uses a binary decision to select or reject a keypoint, the current work presents a quality measure to rank the keypoints. In the previous work, the keypoints and features were extracted on a fixed scale. However, in the current work, a technique to automatically select an appropriate scale at each keypoint and extract features invariant to scale is proposed. For the matching part, the calculated surface depth values are used to form a feature vector and a matching algorithm is employed, which clusters possible transformations between the queried object and the database models, for 3D object recovery from cluttered scenes.

Tombari [287] proposes a method to provide a unique local reference frame (LRF) for the purpose of enhancing 3D format descriptors. More specifically, the author proposes a descriptor that is based on the 3D shape context (3DSC) formulation. To generate a unique and unambiguous LRF, the approach proposed in [288] is used. To generate the proposed descriptor, unique shape context (USC), first the LRF is computed under a region around a feature point. Then, the spherical volume around the feature point is subdivided uniquely by means of a spherical lattice oriented with three repeatable directions provided by the LRF. Lastly, each network's bin accumulates a weighted sum of the surface points.

Also, Tombari [286] presents a method to detect free forms in 3D space in order to solve objects recognition task in 3D scenes with significant occlusion and clutter degrees. The presented method uses the 3D features detection and description to compute correspondences set between the 3D model and the current scene. The complete system can be divided into two stages: The first stage is an offline stage, in which the models detection and description occur as well as the Hough accumulator initialization. The second stage is the online stage, where the scene is analyzed, detecting and describing the scene features for the subsequent feature match between scene and model, followed by the voting stage, in 3D space Hough, in order to detected the object and determine its pose.

A rotation-invariant method based on the local feature configuration for detection, recognition and classification of 3D objects is proposed by Knoop [141]. The proposed method is an implicit shape model (ISM) generalization for the 3D case, with a refined voting scheme based on the Hough transform. The classification process using Hough's transform has three main parts: In the first part, a class model is learned from a set of training formats. In the second part, the learned class models are used to generate hypotheses of the probable researched sample class. In the last part, the hypothesis most likely to be the correct class is searched. The votes for each hypothesis are accumulated in a specific 4D space. The class with the highest probability, its location and scale are obtained by searching for the maximum value among all classes (Fig. 3).

Seatovic [61] presents a system for automatic plant treatment. The system combines an infrared laser triangulation sensor with a high-resolution camera to generate 3D weed images in a plantation. In the segmentation process, continuous surfaces patches are separated from each other. These 3D surface patches are compared to different criteria in the plant database, which contains surface parameters such as shape and surface state. If the object is recognized as a weed, its coordinates are computed and the leafs are sprayed with a herbicide. The complete system is fully described in [263].

Zhou [337] introduces a set of format-based features called Histograms of Categorized Shapes (HCS) for 3D ear recognition. Initially, to locate the ear in a depth image, the image is scanned from the upper left to the lower right corner with a fixed size detection window. At each position, the HCS feature vector is extracted and used to train a binary classifier. The classifier trained in this case was the SVM, which after training was employed for recognition.

A method for recognizing 3D objects based on their 2D curve projection invariants comparison is presented by Unel [295]. The proposed method starts with a depth image or a tessellated object representation. In both cases, the analyzed object orientation is computed, and when the input image is the depth image, the approach adjusts an algebraic surface to the object. Then, the surface data second-order moment

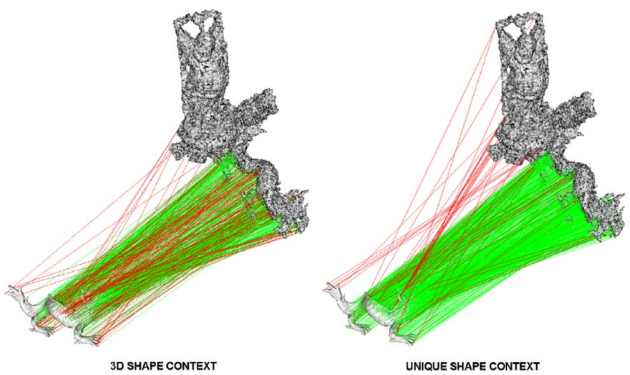

Fig. 3 Representation of the method proposed by Tombari. Figure extracted from [287]

matrix eigenvectors are computed. These eigenvectors imply three orthogonal directions in the space along which the surface data spread. The algebraic surface crossings with planes orthogonal to its main axes produce projection curves in the coordinate plane. When the input is the tessellated image, the analyzed object orientation is computed employing an almost object boundary convex hull. The quasi-convex hull induces an input object orientation in terms of the inertia axis, as defined by the moment inertia tensor. The crossings are obtained through the perpendicular planes intersection to the inertia axis with the tessellated model. Since crossing is planar curves, they are treated as projection curves. These projection curves are shown as affine equivalent, and two methods are proposed for the projection curves algebraic and geometric invariants construction. For the 3D object recognition, a mean similarity measure of the projection curves invariant vectors is employed.

An algorithm for 3D object recognition in sparse, non-segmented and noisy data is presented by Papazov [227]. The method consists of two phases, model preprocessing and recognition. In the model preprocessing stage, executed offline, each object model goes through an oriented point pairs sampling process that respect a tolerance value. For each point pair, a descriptor is computed, which is stored in a hash table. In the recognition stage, the following steps are performed: scene octree computation and calculation of the iterations required to reach a recognition probability greater than a previously defined value. In each iteration, the scene points sampling is performed, the points normal is estimated and the descriptors are computed. Then, the calculated descriptors are used as key to retrieve, from the models hash table, the model-oriented point pairs similar to the scene point pairs. After that, the transformations to align model and scene are computed and stored if the model-transformation pair is accepted by an acceptance function. Lastly, the conflicting hypotheses are filtered out from the list of solutions.

Akagunduz [2] proposes a 3D object recognition method invariant to transformations. First, the 3D surfaces are represented by 3D surface structures called multiple scale features. These surface structures are extracted from the depth images invariant to their size and sampling metrics. Then, multiple scale features are extracted with their scales using the space-scale 3D surfaces curvature. Triples of these multiple scale features are considered to represent the surfaces topology invariant to transformations, using them in a geometric hashing framework for object recognition. To observe the benefits of using multiple scale features, as opposed to feature extraction on a single scale, the 3D features were extracted with multiple scale search (MSFE) and without multiple scale search (SSFE).

A pipeline for learning-based 3D object recognition is described by Owechko [222]. The first step is a spatial suggestion process consisting of segmenting a point cloud into potential objects or suggestions. Then, related suggestions are merged and the resulting segmentation is used to select a 3D classifier based on the spatial properties suggestions. Next, the segment-centered regions of interest are used to generate a set of 3D features that capture geometric and topological properties of the point groups, within a region of interest. Lastly, statistical classifiers based on decision trees are trained with these features. The trained classifiers result is a object regions set which are segmented and semantically labeled.

An approach using a two-layer particle filter is proposed by Lee [155] for the 3D object pose recognition and estimation. In the upper layer, a set of candidate object's poses is identified and preserved in the search space as a set of super-particles. For each super-particle, a real pose probability is attributed, which is developed over time with the future evidence accumulation. In order to define the candidates for the object's pose, first, weak evidences are initially acquired, interpreting them in terms of possible object poses. These interpretations serve as regions of interest for a detailed investigation whereby the poses probabilities are computed for individual interpretations based on the probability and improbability of several features available in the corresponding regions of interest. During the probabilities computation, the object pose candidates are selected to be used as super-particles in the upper layer. In the lower layer, the pose uncertainties associated with the individual candidates are represented as particles which are subject to propagation over time. In summary, the two-layer particle filter algorithm can be described by the following steps: initialization, propagation, new super-particles generation, super-particles merging and resampling.

Zhou [338] presents a complete 3D object recognition system combining local and holistic features. This system was evaluated in an ear classification task and consists of four primary components: 3D ear segmentation; local feature extraction and matching; holistic feature extraction and matching; and a merging framework combining the local and holistic features at the matching score level. For the segmentation component, the method presented in [337] for 3D ear segmentation was employed. For the local feature extraction and matching component, the Histogram of Indexed Shapes (HIS) feature descriptor was introduced and extended to an object-centered 3D format descriptor called the Surface Patch Histogram of Indexed Shapes (SPHIS) for surface segment representation and correspondence. For the holistic feature extraction and matching component, the ear surface "voxelization" was proposed, to generate a representation from which an efficient voxel-wise comparison of gallery and test models can be performed. The correspondence scores obtained from the holistic and local features matching, between gallery and test models, are fused to generate

the final matching score through a weighted summation technique.

Decker [52] presents two approaches for automatic 3D object classification in 2D images. The first method is based on statistical modeling of wavelet features and uses the estimation by maximum likeness to determine the scene object class. This approach can be summarized in training and classification. In the training part, the object image is acquired from different viewpoints, preprocessed in one of the investigated color spaces, has its features extracted, the object area defined and the probability density function estimated. In the classification part, a feature vector set is determined from a test image and evaluated against the density functions of all considered objects classes. The second approach is based on robust local point descriptors. Initially, for training, SURF features are extracted from training images. Then, for the recognition part, the image features are matched geometrically by corresponding training and test images descriptors, in order to find the object with the highest correspondence to the queried image.

Hanai [97] presents a database of electronic parts and, initially, compares some features employed for the object classification task. Based on a survey presented by Akgul [4], two feature extraction methods, density-based framework (DBF) and CRSP, were chosen. The author also added one more method, the Surflet Pair Relation Histograms (SPRH), in order to analyze which method is most appropriate for the presented database classification. The measures to evaluate the methods were the Discounted Cumulative Gain (DCG) and nearest neighbor (NN). In NN, the percentage of the first closest matches belonging to the investigated class is used. A high NN score indicates the algorithm's potential for classification tasks. The DCG evaluates an entire sequence of ordered objects through a similarity measure and provides a greatest weight for the best ranked in the correct and incorrect ratio. The three methods were applied to the entire database, and the DCG and NN were used to evaluate the results.

Zarpalas [333] presents a descriptor for object recognition in 2.5D scenes in the presence of occlusion and clutter. The proposed compact regional format descriptor, called projection images, was designed to be robust against noise, partial occlusion and clutter. The projection images are formed by points projections on the plane centered on the base point, which is perpendicular to the visualization axis. For the recognition process, the projection images of known objects are extracted and stored in a database. Then, given a scene scan, the projection images are extracted and compared to those stored in the database through a matching between their points.

Petricek [229] presents a feature-based method for recognizing 3D objects in cluttered scenes. The proposed method is applied to polygonal meshes to establish a single and unambiguous local reference frame (LRF) and to create feature descriptors, similar to the MeshHOG. The recognition method based on this descriptor consists of two phases: learning and recognition. In the learning phase, the features are computed densely for the model. In the recognition phase, features on the same scale are computed for the scene. These scene features are then matched with model features to form a set of matching attempts, which provides preliminary object pose estimation. The object final hypothesis is generated by a consensus-based procedure (Fig. 4).

Yabushita [327] proposes a framework for 3D object recognition that requires few reference images. The proposed framework first estimates a 3D object model from a video sequence and generates a single target image through a spherical projection of the 3D model and the texture. After that, the object is recognized by matching the target image against the reference image stored in the database. The matching process initially detects keypoints in the target and database images and describes them using the SURF feature vector descriptor. Then, the distance between the target and database feature vectors is calculated. For each keypoint in the target image, the method searches for the k-nearest neighbor keypoints in the database images and computes the votes score from the distances for the n images. The database image receiving the highest number of votes becomes the recognition result.

A method for extracting geometric features based on gaze modeling, for 3D object recognition, is proposed by Maeda [184]. In the modeling process, local model surfaces are independently estimated for depth data parts constrained by several gaze domains. Then, since the features are extracted independently of each gaze domain, inconsistent or incoherent features can be obtained. Therefore, a stochastic method is introduced that allows to integrate such features by the reliability evaluation of each model gaze. In order to avoid generating multiple descriptors for an image, it is attempted to generate a format descriptor for a limited number of

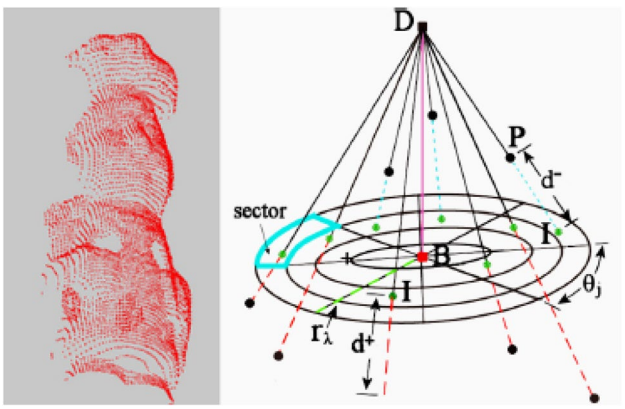

Fig. 4 Representation of the method proposed by Zarpalas. Figure extracted from [333]

feature points. The proposed descriptor, curvature distribution image (CDI), is generated based on the rates between surface curvatures to efficiently encode the curvature distribution near a reference point. In order to discuss the 3D format description performance by the type of curvature employed, five types of curvatures were used. Consequently, since an object is represented by a set of descriptive CDIs, the recognition is performed by finding the object at the database with the CDIs similar to the target object.

Bariya [24] presents a framework to explore the 3D geometric scale variability in depth images, which provides rich information to characterize the general geometry. To do so, a scale-space representation is derived through the normal field convolution of a depth image with Gaussian kernels, which results in a hierarchical feature set of different scales. Then, a local format descriptor is generated, which represents the surface structures that originate those features by sculpting and encoding the local surface falling within the feature support region. For object recognition, the following main steps are performed: Initially, a 3D object models library is constructed from the objects for recognition. Then, an interpretation tree is used to make all possible matches between scene and model features. Among all the hypotheses represented by the interpretation leaf nodes, only the strongest hypotheses are chosen for verification. Lastly, the hypothesis that produces the maximum value of overlap area is chosen and refined using the ICP algorithm. A similar framework was presented by the author for the recognition of 3D objects in cluttered images [23].

Socher [274] presents a model based on the combination of convolutional and recursive neural networks for the feature learning and classification in RGB-D images. The model starts with RGB and depth images and extracts features from the images, separately. Each image modality is inserted into a single-layer convolutional neural network (CNN), which provides translation-invariant low-level features and allows parts of an object to be deformable up to a certain extension. This layer outputs are then inserted into recursive neural networks (RNNs), which can learn compositional features and interactions between the parts. The RNNs hierarchically project the entries into smaller spaces through multiple layers with weights and nonlinearities attached. Finally, the concatenation of all the resulting vectors generates the final feature vector used in a softmax classifier.

A local feature descriptor for 3D object recognition is proposed by Jang and Woo [113]. The descriptor is a combination of local angle patterns (LAPs), gradient orientation and color information. The complete recognition system can be described in the following steps: First, feature points are extracted from the image. Then, for objects with low texture, the algorithm extracts feature points through a random sampling of positions in the object edges. After that, based on the extracted feature points, feature descriptors are constructed by combining LAP patterns, gradient orientation and color histograms of the fragments along the R, G and B spaces. Next, the descriptors are labeled to learn a codebook through random forest, and with the learned codebook, the proposed algorithm constructs a Bag of Words (BoWs). For the test, the trained classifier evaluates the queried BoWs ensuring that the highest probability is selected as the recognized object.

Liu [170] proposes a 3D object recognition method based on line drawing. Initially, a circular visual feature representation, using excitatory and inhibitory components, is presented to extract distinct information from the line drawings. This procedure is based on the process of forming a 3D object image in the retina. To simulate this process, the 3D object model is placed so as to match the center of gravity with the sphere center which delimits the 3D object. A dense viewpoint sampling is applied to the sphere surface, and for each viewpoint, a lighting model is applied to generate a object shadow image, which is converted to a line drawing using the CLD algorithm. Based on the representation by drawing lines, the Halton's quasi-random point sequence method is applied to uniformly sample a number of points within a region. At each sampling point, a circular histogram is established, in which each circular bin has the same radius difference and the maximum circle has the radius equal to one-fifth the size of the delimited area diagonal. The histogram of each sampling point has a number of bins, and if the number of points that fell in the bin is greater than a threshold, then that bin is activated and used in the recognition process. Based on this representation by feature histogram, a codebook approach is used to organize and process the correspondence based on the similarity metric presented.

Oleari [218] presents a low-cost stereo vision system designed for object recognition using fast point feature histogram (FPFH) [253]. The low-cost stereo vision system provides a precise and dense disparity image, which is transformed into a point cloud to perform the object recognition task. Since scene segmentation is not the proposed work focus, it is assumed that the object to be recognized is on a flat surface and within a delimited region. The recognition itself is a cluster-based recognition, which aims to compute the match between a selected cluster with an entry in the known models data set.

Bennamoun [29] presents a free-form 3D object recognition system based on surface feature descriptor. The system can be described by the following steps: first, for a randomly selected feature point, a local reference frame (LRF) is defined. Then, a feature descriptor, called Rotational Projection Statistics (RoPS), is constructed by computing the point distribution statistics in the 2D plane defined by the LRF. Finally, the recognition algorithm based on the RoPS features is presented, where the candidate model and the transformation hypotheses are generated by matching the scene

and library model features. This is performed by means of distance calculation between model-scene features using a kd tree and a voting system. These hypotheses are tested and verified by aligning the model to the scene.

Yabushita [326] proposes a 3D object recognition technique that allows the user to perform mobile visual searches of 3D objects. This technique, based on the framework previously proposed by the author [327], starts by extracting keypoints from all the captured video frames. These keypoints are tracked between two contiguous images by comparing their descriptors. The keypoint pairs, that have the most similar descriptors across a series of two or more images, are considered drawn from the same points in the object. Using this computation procedure, two steps are performed: multiple-view keypoints are gathered and the 3D coordinates of each grouped keypoint are estimated. The grouping and estimation processes results are used in the matching process [327], which focus on keypoints matching between query and database images through the comparison of their feature descriptors.

Kim [130] proposes a framework that explores the compatibility between object segmentation hypotheses in the image and the corresponding 3D map to determine the 3D object location. The framework is based on detection methods presented in [68, 228, 315], which identifies objects in images by means of bounding boxes. Based on these bounding boxes, the compatibility between the objects hypotheses within the box and the 3D map associated with the pixels within the box is explored. These object hypotheses, called Hypotheses object Foreground Masks (HFMs), are generated from the background and foreground object segmentation hypothesis within the bounding box. The object models are learned using a latent maximum-margin formulation. The features are extracted from appearances clues within the HFM, and the 3D descriptors are computed in the associated point cloud. The deformation costs for the relative distance between the object parts and the object root positions are calculated in 3D space, where a matching strategy is used. This 3D matching procedure involves the following steps: filters response maps projection in the 3D point cloud, scoring function definition and part responses sum according to their deformation costs. Once the object root location is obtained, the object parts location can be found by looking at the optimal displacement, similar to the 2D case [68].

An approach to 3D face recognition based on the frontal contours of heat propagation under the face surface is proposed by Abdelrahman [1]. The frontal contours are extracted automatically as the heat propagates from a set of automatically detected reference points, obtained through the method proposed in [67] for reference points detection. The approach encodes the local face features as well as the diffusion distance on the surface around these reference points. After calculating the heat kernels at each point, the 3D contours are drawn at the face surface 3D point that has the same heat value. A predetermined number of contours are used around each point, and each contour is sampled with a fixed number of points. This representation provides an ordered and finite set of 3D points per face. For the correspondence between two faces, the ICP is used to estimate rigid transformation parameters between the set of points, which correspond to the contours found for the two faces. The distance L2-norm between the acquired face and the gallery faces contour points, after registration, is used as distance measure, ranking the gallery faces based on this distance value (Fig. 5).

Yu [330] presents a robotic vision-based system, which can not only recognize different objects, but also estimate their pose through the use of a deep learning model. The deep learning model used is the Max-pooling Convolutional Neural Network (MPCNN). The author states that the deep learning model does not work well for object detection. Therefore, to work around this problem, he proposes the use of an object detection method to segment the object background. The data flow to the robotic system begins with the camera acquisition followed by the object detection method, with a learning method based on dictionary. Then, the MPCNN is used to recognize the objects and estimate their poses. Finally, the robot controller moves the robotic arm, picks up the object and moves it to a pre-intended position.

Lam [151] presents a system for recognizing 3D objects based on segment registration. This system has two stages, offline and online. In the offline stage, the training is performed, in which the models are preprocessed in segments of interest and in quantized point pairs. In the online stage, three phases are performed: segment of interest extraction, segment registration and verification by reprojection. In the segment of interest extraction, the algorithm employed can be divided into three parts: First, points of interest are extracted through operators like difference of normal operator. Then, the RANSAC line is used to estimate the segments boundaries as linear curves per patch and finally, the entire

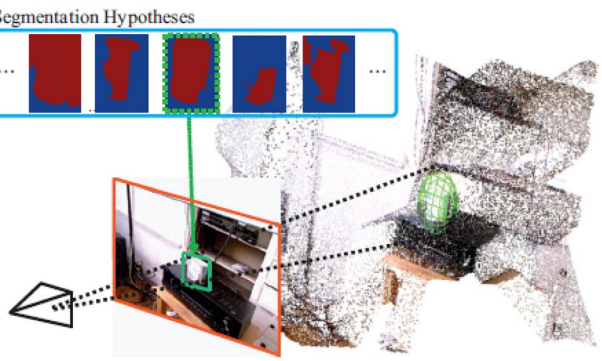

Fig. 5 Representation of the method proposed by Byung-soo Kim. Figure extracted from [130]

segment of interest is extracted using a region growing algorithm. In the segment registration phase, a pairing method such as 4PCS is used, which gives the model and scene segments, and performs a pairwise comparison, where each comparison will result in a candidate pose. Finally, in the reprojection verification part, each candidate pose is verified by reprojecting the model points in the scene. This process is performed in such a way that if the model instances, in the scene, are not strongly occluded, then the correspondence that provides the highest overlapping score will naturally be selected as the best pose estimation.

Ekekrantz [65] proposes a method called adaptive iterative closest keypoint (AICK) for registration of RGB-D data, based on the iterative closest point (ICP) algorithm principle. The AICK input is a set of invariant keypoints detected in each RGB-D frame, where each keypoint is associated with a 3D position in a local reference frame (LRF) and a feature descriptor. The AICK algorithm makes no assumption about the keypoint detection and description algorithms employed; however, it is desirable that the keypoint detector be invariant to rotation and scale. Thus, the SURF and ORB keypoint detectors were evaluated. In the original ICP, the Euclidean distance associated with the pairs of points between two clouds is used. In AICK, the Euclidean distance is replaced by the weighted Euclidean distance sum. By the author's choice, the Euclidean distance factor is completely neglected for the initial match. As a result, the algorithm does not require an initial guess for the transformation between the point clouds. Starting from the second iteration, correspondences which are geometrically close, but not so close in appearance, are allowed. This process continues until the appearance is no longer considered. At this point, the original ICP is performed for a finer registration.

Guo [92] presents a 3D object recognition algorithms that use not only format, but also color information. First, the descriptor previously presented by the author, Shape only Rotational Projection Statistics (S-RoPS), is extended to obtain the RoPS feature descriptor with color only (C-RoPS). To generate the C-RoPS, the same S-RoPS structure is used, replacing the spatial information with the color information. Then, given a local surface, a surface local reference frame (LRF) is constructed using the same approach used in S-RoPS. After that, the color parameters are employed to replace the coordinates and an approach similar to the S-RoPS is followed to generate the C-RoPS. The proposed recognition algorithm includes four modules: model representation, candidate model generation, transformation hypotheses generation, verification and segmentation. Each module is fully explained at [92].

Ma [182] presents the development of a mobile and customized manipulator for pose estimation and twist-lock grasping. For the perception part, an approach for 3D object recognition, using kernel principal component analysis (KPCA), based only on the depth information, is proposed. The perception process is divided into two parts: offline analysis, composed of data sampling, feature extraction and training based on the extracted features, and real-time processing, where, based on the feature training, the twist-lock is detected and the type and pose are identified. In the object detection and feature extraction parts, a technique combination is used to remove the object from the background and extract object features. Initially, the background subtraction method is applied to remove the object from the background. Then, a median filter is used to remove the noise followed by a Sobel operator for edge detection. The object position can be roughly estimated after edge detection. After locating the object, a set of kernel principal features of the depth images is used for 3D object description and recognition.

A method for classifying 3D objects based on local keywords and hidden Markov model (HMM) is presented by Guo Jing [115]. In the proposed approach, a vector of geometric features, based on the surface points Relative-Angle Context Distribution (RAC), is extracted. The local keywords are generated from clusters of RAC histograms. For the RAC histograms (HRAC) clustering part, the k-means algorithm is used. Then, each object is separated by the combined model, and a local keyword can be acquired. The local keyword is characterized by the cluster center. The clusters can also be associated with some kind of semantic meaning, such as the tiger's head or the bee's wing, and each cluster represents a unique local keyword. In the classification process, a first-order HMM was trained for each object class and employed for object classification.

A paradigm for real-time location and mapping, which uses 3D object recognition to jump over low-level geometry processing and produce incrementally constructed maps, directly at an object-oriented level, is proposed by Moreno [256]. This paradigm called SLAM++ can be described by the following steps: initially, an object database is created with a scan using the KinectFusion in a controlled environment, extracting a point mesh through the Marching Cubes algorithm. Then, a world representation with graph is used, where each node stores either the estimated object pose or the camera pose history in a given time frame. Each object node is annotated with a database object type and each object pose measure, from a camera, is stored in the graph as a factor which links a camera and an object poses. For the object's pose recognition, an approach similar to that employed in Drost [59] is used (Fig. 6).

Pang [225] describes a 3D object recognition method that combines machine learning procedures with 3D local features. The proposed method is divided into two modules, training and detection. In the training module, a detector is trained for each object class, using the Adaboost training procedure, with training samples generated from a prelabeled objects library. The object detectors consist of N

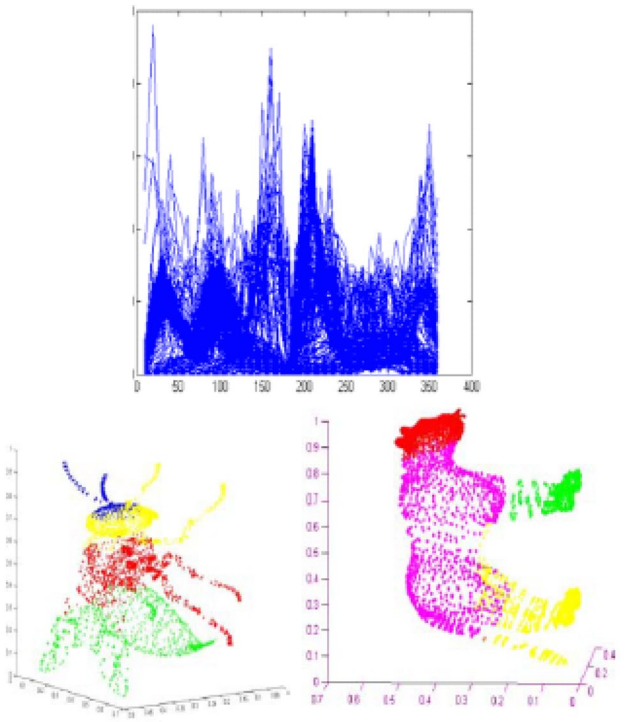

Fig. 6 Representation of the method proposed by Guo Jing. Figure extracted from [115]

weak classifiers, trained based on 3D image features (3D Haar feature), each with a weight. Each weak classifier evaluates a region candidate subset and returns a binary decision. The object detector, or strong classifier, is a combination of all the weak classifiers weights, which is compared to a predetermined threshold to decide whether the candidate region is a possible match. The detection module input is a 3D point cloud region. A 3D detection window is moved to search through the 3D image, evaluating the match between each point cloud cluster, within the detection window, and the target object. After exhaustive input point cloud scanning, all detected positive instances are processed through non-maximum suppression to identify the target object with the best match and confidence above a threshold.

Takei [278] proposes a method for locating and recognizing 3D object pose in cluttered scenes. The proposed approach extracts scene features using 3D computer graphics and employs the extracted features to achieve improved discriminatory performance. In an effort to improve discriminative performance, the diverse density method evaluates the features by using the distance in a feature space. The feature efficacy is evaluated using communion of positive samples and the negative samples separability. The proposed method uses vector pairs with high discriminatory performance for the matching process, evaluating the discriminatory performance by the concentration and separability degrees combination. The proposed method evaluates the discriminatory performance of all the vector pairs in the model object and selects vector pairs that have the highest evaluation values.

An approach to test the codebook feasibility for automatic threat classification in pre-segmented computed tomography images of luggage is investigated by Mounton [201]. To do so, the classification framework Bag of (Visual) Words (BoW) was employed. Based on this framework, the performance of five codebook models was compared using a variety of sampling strategies combinations (sparse via the SIFT 3D point of interest detector, and dense as recommended in [214]), feature encoding techniques (k-means clustering and extremely randomized clustering (ERC) forests) and classifiers (support vector machine and random forests) with an approach using visual cortex. The proposed techniques combination was evaluated in the context of classification of two target objects in computed tomographic images of luggage.

Xiangfei Qian [237] proposes a 3D object recognition method, which segments a 3D points set into a number of planar segments and extracts the inter-plane relationships (IPRs) for all segments. Based on the IPRs, the method determines the high-level feature (HLF) for each segment. Then, a plane classifier based on Gaussian mixture model (GMM) is used to classify each segment into a plane belonging to a certain model object. Finally, a recursive plane clustering procedure is performed to cluster the object model classified planes. The recognition method consists of five main procedures: 3D depth data acquisition; depth data plane extraction; feature extraction; GMM plane classifier design and training; and plane clustering.

Shang-Hung Lee [158] proposes moment-based 3D format registration method. The proposed method is one of the stages from the proposed 3D human body action recognition system, which consists of two stages, training and test. The training stage consists of six steps: 3D formats clustering through registration; local feature extraction; feature codebook generation; key poses detection; and in motion poses representation. From this application, a dynamic 3D format sequence is represented as a key poses sequence by matching each input 3D action format with the codebook poses to find the closest pose template. Finally, the SVM training takes place based on these key poses in motion representations. For the test part, the local features are extracted and the 3D actions are classified using the previously trained SVM classifier.

Guo [90] proposes an algorithm for registering multiple-view depth images. In this algorithm, a Rotational Projection Statistics (RoPS) features set is extracted from depth images pair, performing the correspondence between them. The two depth images are then registered using a transformation estimation method (CCV) and a Iterative closest Point (ICP) algorithm, for a more refined registration. Based on the pairwise registering algorithm, a multiple-view registration algorithm based on format

growth is proposed. Then, the format seed is initialized with a selected depth image, which is sequentially updated by performing a registration in pairs between the selected depth image and an input depth image. All input depth images are iteratively registered during the format growth process. Once the meshes corresponding to a particular format have been registered coarsely, the obtained results are refined with a multi-view register algorithm. Finally, a continuous 3D model is reconstructed for each format through a surface reconstruction and integration algorithm.

Ejima [64] proposes a 3D object recognition method based on reference point ensemble, which is a generalized Hough transform natural extension. The reference point ensemble consists of several landmarks color coded with green or red, where red landmarks are used to verify the hypothesis and the green landmarks are used for voting. In the proposed method, a set of reference point ensembles is generated by the local features of a given 3D scene. These local features are described by the Labeled-Surflet-Pair (L-Surflet-Pair), which is derived from Surflet-Pair, generating the reference point ensembles set from this description. Each generated reference point is a hypothetical 3D pose of a given object in the scene. The hypotheses going through the verification procedure by the red landmarks are used in the voting. The Hough voting is performed independently in each green dots space, which reduces the voting space to three dimensions. The effective object recognition is achieved by the change between two different modes: the individual mode, in which the independently hypotheses voting, in each Hough space of the green reference points, and the hypothesis verification with the red reference points are performed; the ensemble mode, in which the register verification occurs in a list of promising hypotheses and the total votes aggregation, is performed.

Geetha [180] proposes an algorithm which recognizes 3D object using 3D surface format features, 2D format border features and color features. For the 3D object format extraction, the first step is the reference points detection. To identify such points, two methods are used by the proposed approach: The first method uses the distance function first and second-order derivatives, for each point on the 3D surface, for a projection plane. The second method searches for some uniform point distances in the RGB-D images. After detecting the reference points, feature vectors, corresponding to the reference points, are computed using the principal curvature concept, calculated by the principal component analysis (PCA) algorithm. For the 2D format boundary extraction, the HOG descriptors are used, and for the color feature extraction, normalized histograms are employed. The feature vector of each object was stored in the database and later was used to obtain the closest correspondences, through the Euclidean distance measure.

The work presented by Mhamdi [195] uses a set of measurement functions for 3D object recognition. The size functions principle is to describe the 3D object by encoding the topological changes provided by its critical points and the link between them. A critical point can be a local maximum or minimum or a measurement function saddle point defined in the 3D object, and the idea is to describe the 3D object feature by a function. In the case under study, each 3D object is described by 18 measurement functions referring to 18 3D object portions in order to take advantage of different information present in each axis. The comparison between two 3D objects, described by the measurement functions, is performed through the similarity measure between the 18 functions associated with each object, which could be expressed by the minimum distance between the corresponding 18 functions distances associated with the object. Before performing the whole segment partitioning process and description through the 18 measurement functions, a preprocessing step occurs, which is divided into low-level processing and data normalization with the purpose of giving invariance to transformations, twists and articulations to the features to be extracted.

Guo [91] presents a local surface feature descriptor called Tri-Spin-Image (TriSI) used for 3D object recognition under occlusion and clutter. The assembled scheme for 3D object recognition consists of four parts: preprocessing, feature generation, feature comparison and hypothesis checking. In preprocessing, features points are uniformly selected from each model and model features are obtained by calculating the TriSI features at each feature point. These features of all models are used to construct a subspace derived from the PCA method application. Then, each model feature is projected into the generated subspace, obtaining compressed features, which are indexed and stored. The second part is the feature generation, which, given a scene, performs the same process of generating features projected in the subspace generated by the PCA. The third part uses the generated features and calculates, for each feature, the first and second closest distances between scene and stored model features. In the latter part, the models are ordered according to the received votes and are checked one by one based on the estimated transformations between the candidate and the model. In this verification step, a confidence score is assigned and the highest score transformation is used to align the model with the candidate.

Figueredo [69] proposes an algorithm for 3D object recognition starting from a point cloud of 3D rotated symmetric objects. The algorithm can be described in two steps: object description in terms of surflets and object recognition. The surflets are the basic units used to describe the surface shape and are represented by surface sampling points and the associated surface normals. The recognition process consists of a matching process between model and scene surflets

pairs. Based on a surflets pair, the point pair feature (PPF) can be calculated, which is defined by four tuples referring to the two associated surflets pairs. Searching for speed in the matching process between the model and the scene, the data structure used to represent the model description was a hash table, in which the key value is given by a discrete PPF, while the mapped value is the respective surflet pair. Based on the description through surflet, the entire model and scene matching process, transformations calculation and subsequent position estimation are performed (Fig. 7).

Yan Zhuang [341] presents a 3D object recognition framework, for a service robot, to eliminate false detections in cluttered environments. Initially, the laser point clouds are converted to bearing angle images and a Gentleboost-based approach is employed to detect multi-class objects. In order to solve the object variable scales problem, in the object detection, a scale coordination technique is adopted in each segmented sub-scene according to the 3D laser points spatial distribution. In addition, a semantic information, extracted from 3D laser points, is used to eliminate false object detection results. Finally, k-means clustering and Mahalanobis distance are employed to perform object segmentation on the laser point cloud.

Sanguino [262] presents an approach to detect and classify 3D objects using the generalized Hough method and a Kinect sensor. The algorithm considers feature points and the color spectrum as two interlaced processes to cooperatively recognize the object in a 2.5D scene. With this strategy, the algorithm automates preprocessing operations independently of scene and reduces the processing load on the object's point cloud for 3D object classification. The process of integrating color and format information is accomplished by simplifying the generalized Hough transform (GHT) and using the color spectrum as decision criterion. The following sequence of steps is used for segmenting the scene object: First, the RGB image is captured by the Kinect camera and converted to the HSI space. Then, the depth information is collected by the Kinect sensor and shown in gray scale. The distance information is applied to the input image, and the objects in the scene are discriminated. The Canny filter is applied to detect the contours around the objects, and the object identification result is obtained after the object removal from the scene. After the segmentation, the Hough spectrum and color spectrum combination is used for object recognition and classification based on a fitness value.

A feature description method called SHORT (Shell Histograms and Occupancy from Radial Transform) is presented by Takei [279]. The SHORT consists of a keypoint detector and a feature descriptor that uses a small amount of points in the restricted local regions. The keypoints detector evaluates the local format through the occupation use. In the object local region, the occupation differs from format to format because the local point cloud spatial extensibility differs from one format to another. In this way, the keypoints are detected by using the estimated occupation value as the format evaluation value at each point. The feature descriptor describes shell features from multiple scales. It also uses the points in the outer shell regions, in the sphere of multiple scales, and the estimated occupation with the keypoints detection. Then, it sets in advance the spherical shell regions that differ in scale at a keypoint and estimates a dominant direction vector by using the occupancy value and the point cloud in the configuration regions for the keypoint. Next, the descriptor computes the inner products histogram of the dominant direction vector and the directional vectors for the points in the shell regions, for the keypoints in each scale. Finally, an internal product histogram is integrated, in each scale, as being the shell features of multiple scales. The 3D Hough voting [286] method was employed as recognition algorithm.

The work proposed by Zou [344] seeks to explore a set of classifiers for the purpose of constructing a feature extraction

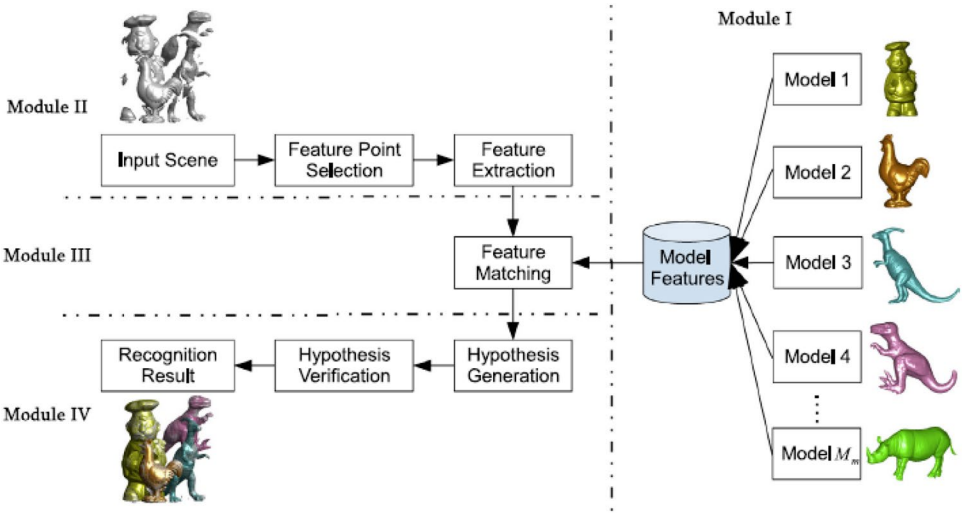

Fig. 7 Representation of the method proposed by Yulan Guo. Figure extracted from [91]

method for 3D object classification. The proposed method is based on the following idea: initially, a classifier is trained for each class and the outputs of all classifiers are combined as object feature. To construct the extraction method, the L2-norm regularized logistic regression was used due to the ease of developing the L2-norm update rule in a stochastic gradient rise form, which makes the proposed method scalable for training on a large data volume. The proposed method was compared with other three feature extraction methods (SIFT-based BOF, sparse coding and deep belief networks). In order to evaluate the proposed feature extraction method, the feature vectors were extracted for the selected databases and the average precision was calculated using the k-nearest neighbor search.

Xu [322] presents a method for 3D object recognition, which includes normal estimation, feature point selection, feature descriptor extraction, matching of scene and model features, hypotheses generation and verification. However, the author focused only on the first three developed parts (normal estimation, keypoint extraction and the local feature descriptor computation). For the normal estimation, the author analyzed the covariance matrix eigenvalues and eigenvectors created from the neighbors closest to the point in question. For the feature points selection, an extraction method called NARF (Normal Aligned Radial Feature) was employed. For the feature descriptor, the Fast Point Feature Histograms (FPFH) were used. The next steps to be implemented are: the KNN use for the feature matching and the Hough's voting for hypothesis generation and verification of global hypotheses.

Filipe [70] presents a method for detecting keypoints in 3D point clouds and performs a comparative evaluation between each 3D point detectors and 3D descriptors pair to evaluate their performance in recognizing objects and categories. The proposed 3D keypoint detection method, called Biologically Inspired 3D keypoint based on Bottom-Up Saliency (BIK-BUS), is a keypoint detector based on salience maps, which are determined by the computation of feature intensity conspicuity map. These conspicuity maps are fused into a salience map, and the focus of attention is sequentially directed to the map most salient points. Using this theory and the steps presented in [112] and [98] the 3D keypoint detector, BIK-BUS is generated. The pipeline used in the BIK-BUS performance evaluation was detailedly explained by the author.

Wang [302] uses a convolutional neural network (CNN) model to learn a RGB-D data feature set, which are delivered to a linear SVM classifier to classify objects. In the proposed work, the open-source framework Caffe was used with the SVM to classify the RGB-D data set. To perform the classification task, two Caffe networks are separately tuned using all RGB and depth images. To make the network trained in RGB images applicable to the depth images, first the missing depth values are filled out in each depth image and the filled depth images are converted into three channels using the method proposed in [93]. Then, the adjusted network is used to extract features of two imaging modalities, concatenating the features for the linear SVM training and testing against ground truth labels. The proposed approach was tested in object category recognition, classifying unseen objects under the training data categories.

The work presented by Hong [106] proposes a 3D object recognition method based on multiple-view data fusion. The method, named Multi-view Ensemble Manifold Regularization (MEMR), can be divided into two parts. The first part is training, which can be divided into the following steps: Initially, the image features, distribution and locality information are described and represented by Locality-constrained Linear Coding, extracted from different object views and represented with a feature vector. Next, a PCA method is applied in order to reduce the feature vector dimensionality. Then, the SVM kernel and the penalty matrix are computed for each view. Finally, an alternating optimization process is performed to obtain a refined combination of kernels and penalty matrices, thus obtaining the classifier based on multiple views. For the classification part, the same feature extraction process is used, submitting the extracted features to the trained classifier for classification.

Shah [265] presents a local surface description technique for 3D object recognition. The proposed technique begins with the keypoints detection phase. Then, once a keypoint has been detected, the predominant information from the surface next to it can be extracted and encoded in a local feature descriptor. The descriptor used is the local reference frame (LRF), which is represented by the keypoints themselves. As a next step, the normalized field vector is aligned with the LRF vectors to construct a rotational invariant local surface descriptor. This descriptor, called 3D-Vor descriptor, is derived from the calculated information associated with local surface vorticity. A similar approach was used by the author to introduce the 3D-Div [267] descriptor for 3D object recognition in low-resolution scenes. The object recognition part can be described by the following steps: scene representation through 3D-Div descriptor; feature correspondence between scene and models descriptors; hypotheses transformations generation; hypotheses verification; and segmentation (Fig. 8).

Kechagias–Stamatis [125] proposes a 3D descriptor, which removes the need for a local reference frame/axis (LRF/A), reducing the processing time required. The proposed descriptor, called Histogram of Distances (HoD), is based on multiple L2-norm metrics of local patches. The descriptor computation is inspired by the shape distributions. The main difference between the proposed descriptor and the shape distributions is the D1 function extension to a local base and the substitution of the involved reference

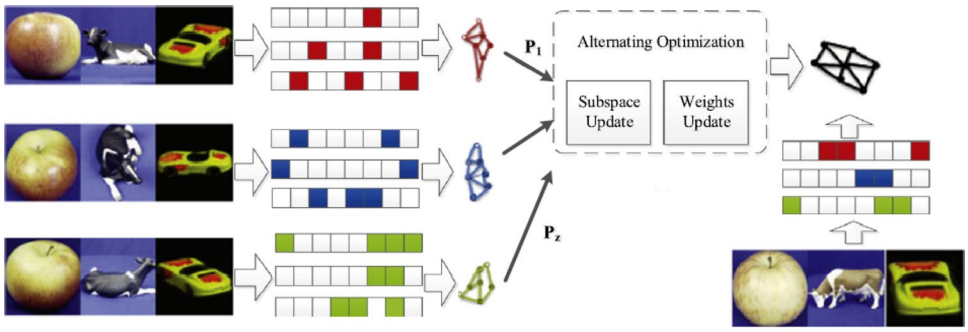

Fig. 8 Representation of the method proposed by Chaoqun Hong. Figure extracted from [106]

point centroid for the edge. Based on this new reference point, the L2-norm is calculated for all vertices in each local area, which was properly normalized and discretized to a predefined number of bins. Then, the normalized distances were encoded in a histogram called HoD. To increase the HoD descriptive power, the coarse and normalized distances were concatenated by selecting well-sized bins. The point cloud resolution invariance was performed by normalizing the HoD [288]. The mesh resolution invariance was extended by replacing the support radius metric with a multiple of each scene mesh resolution. The scene features were matched with all the model features based on its Euclidean and nearest neighbor distances criterion.

Bedkowski [25] presents an intelligent mobile application to support spatial mapping for the security area. In this context, the Complex Shape Histogram (CSH) is presented, which is a central framework component of artificial intelligence engine application's used to classify 3D point clouds with a SVM. Initially, robots acquire 3D data and record it using an enhanced version of the 6D SLAM algorithm. The 6D SLAM has been enhanced through the use of semantic classification, loop closure with CSH and parallel implementations. Then, the 3D object recognition is performed by the Objects of Potential Interest (OPI) detection and identification. To detect OPI, a 3D point nearest neighborhood search procedure is employed for each point in the current 3D measurement. Next, the current 3D scan alignment with the global reference model is performed using the ICP, minimizing the false detections. In order to identify the 3D objects, a knowledge base is constructed, where the training data set, composed of objects with assigned semantic labels, is prepared. Lastly, the 3D object recognition is accomplished through the observed objects classification, in semantic labels, based on the knowledge base. In order to classify the identified objects, the SVM is used based on the acquired point clouds and the training set composed of positive and negative objects examples.

Logoglu [171] proposes two 3D local descriptors for the object recognition task, Histograms of Spatial Concentric Surflet-Pairs (SPAIR) and Colored SPAIR (CoSPAIR). In order to calculate the SPAIR, initially, a 3D grid of N concentric spherical regions of same size is constructed. Then, for each spherical shell called level, relationships between surflet point pairs within a level and source point are calculated. After that, the generated histograms are normalized using a distinct points number in each level. The SPAIR descriptor is defined by three generated histograms concatenation in an order based on their distance from the center. In CoSPAIR, color/texture and format information are coded for each SPAIR descriptor level and the same concatenation process is performed to generate the descriptor. To test the proposed descriptors performance, initially, the database is divided into the query and reference sets. In both sets, the keypoints extraction, through the intrinsic shape signatures 3D (ISS3D), and the descriptors calculation occur, adding to the descriptors database the reference descriptors set and using them in the correspondence and voting processes against the query descriptors set.

The work presented by Shah [266] proposes a representation based on keypoints called Keypoints-based Surface Representation (KSR), used for 3D object recognition. The complete algorithm for 3D object recognition goes through two stages, training and recognition. In the training stage, the keypoints are first detected, and the KSR between the keypoints is computed for all 3D models, storing them in the object database. During the recognition stage, the KSRs, calculated for a given scene, are compared with the model KSRs using a linear correlation coefficient for matching KSRs. The match results between KSRs are used to vote on candidate models and generate the hypotheses to transform the model into the scene. The candidate models are checked in turns by aligning them with the scene using the hypothetical transformations. If the model candidate is precisely aligned with the scene portion, the candidate and the hypothesis are accepted. As a result, the scene points corresponding to the model are recognized and segmented, otherwise the hypothesis is rejected and the next one is verified.

3.1.2 Global features

Another feature type are the Global features, which are efficiently computed to represent 3D objects, as they

reduce its computation space, using fewer dimensions to describe the object. Mostly used in the early studies on 3D object representation, nowadays, the employment of global features has been gradually replaced by the local features. This replacement is motivated due to the fact that the global features are not discriminative enough when the objects have small differences, such as in intra-class retrieval cases or classification of very similar objects [20]. Despite this reduction on its use for object representation, still a good amount of analyzed works employed this feature type and will be further analyzed in this section. Figure 9 shows some examples of works that employ global features for the object representation.

The spin Image (SI) is one of the global feature descriptors employed by several analyzed works. This shape descriptor, presented by Johnson [116], is a data-level shape descriptor used to match surfaces represented as surface meshes. Yi Tan, for example, employs the SI to represent models and scene objects, exploring the similarity between models represented by it. Other works employing the SI for object matching and representation are: Stasse [277] presents a multi-resolution SI approach for object representation and recognition; Assfalg [17] shows a SI variation called Spin Image Signatures (SIS), which is developed under the SI approach with adaptations to support effective retrieval by content; Li [163] demonstrates a framework to identify partial 3D format in 3D CAD parts using the SI as descriptor; Ping [233] proposes the Tsallis entropy use to generate a concise SI representation, called Tsallis Entropy vector of Spin Image (TESI); Choi [47] proposed an improved SI version, which enhances the format discrimination performance, called Angular-Partitioned Spin Images (APSIs). This enhanced version improves the discriminative power and accuracy in detection by generating sub-spin Images for azimuthally partitioned cylindrical spaces; lastly, Iyappan [77] proposes a methodology for 3D ear recognition using SI and removing erroneously mapped features using 3D geometric surface properties.

Other two global feature descriptors that frequently appeared in the analyzed works are the Viewpoint Feature Histogram (VFH) and the moment-based descriptors. The VFH descriptor, presented by Rusu [254], is composed of two components: a viewpoint direction component and a surface shape component comprised an extended Fast Point Feature Histogram (FPFH) descriptor, presented previously by Rusu [253]. The analyzed works that use this descriptor are: In the seminal work, Rusu [254] presents the VFH in a system for recognizing the object and its pose; in a work presented by Bongale [32], where the VFH was used in a system for object recognition and tracking, which uses depth information of a low-cost sensor; and, lastly, in a work presenting a compressed VFH version for object classification based on format recognition, proposed by Salih [259].

A moment is a specific quantitative measure of the shape of a set of points [51]. Several moment-based descriptors were proposed along the years, and some of them are present in this review: Ramalingam [238] presents a fuzzy surface classification paradigm, which is an extension of conventional techniques based on sign of mean and Gaussian curvatures. In his work, a fuzzy moment-based recognition technique described and tested in [275] was employed; Ong [219] presents a theoretical framework for deriving scale and translation invariants for 3-D Legendre moments through the use of direct and indirect methods, employing the obtained invariants on 3D object recognition; Xu [323] proposes a 3D object recognition method, which uses some features, color moments, texture features, Hu's moment invariants and the affine moment invariants, extracted from each 2D image of 3D objects; Mavrinac [191] presents an approach for recognition of 3D objects in arbitrary poses, providing only a limited set of training view samples. This approach involves computing a disparity map and extract, from the map, a set of disparity map features (compactness, first Hu moments and the image general distribution intensity histogram); the method presented by Wan [300] shows a classification method, based on fuzzy KNN and Bayesian Rules, to determine whether a 3D object belongs to the human class,

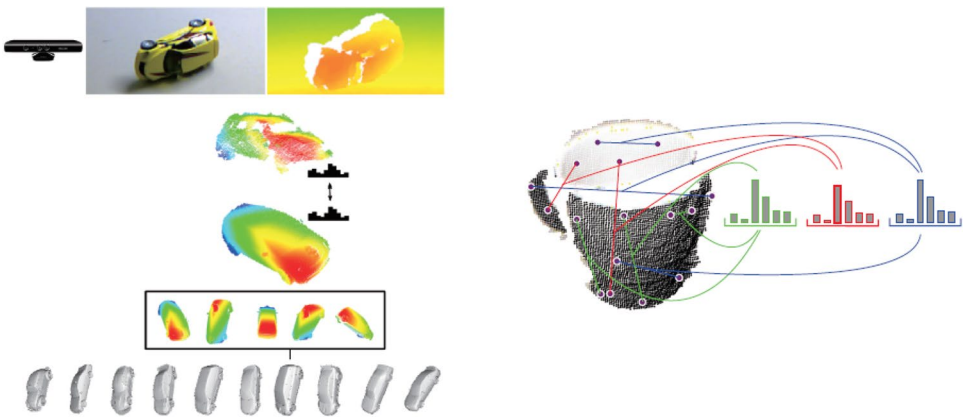

Fig. 9 Representation of analyzed works that employ global features. Composition of images extracted from [310, 311]

using the Zernike moments descriptor as visual features representation; Osman [221] presents a performance analysis of two moments, named Hu's and Zernike's moments, for object recognition; Yangye Wang [305] presents a method based on sub-areas edge moment for 3D object recognition in wireframe; Akbar [3] discusses the use of Clonal Selection Algorithm (CLONALG) and Particle Swarm Optimization (PSO) for 3D object recognition, extracting and using the Hu moments invariants as feature set; Ding [57] proposes a new generalized affine moment invariants called illumination invariant MSA moments. This method combines the traditional affine moment invariant (AMI), the multiscale autoconvolution (MSA) and MSA moment with the basic ideas used to construct lighting invariants, using the obtained invariants in format retrieval and object recognition tasks; Sheta [269] presents an image processing pipeline to recognize 3D objects based on their 2D image. In this pipeline, in the feature extraction stage, the segmented objects moments features are extracted. These features include seven Hu's moments, eleven Zernike's moments and six affine moments, which are used in a mathematical fuzzy model for object recognition; Bencharef [28] proposed a hybrid approach based on neural network and the combination of Hu & Zernike moments with geodesic descriptors for object recognition.

Other examples of descriptors employed are: In Shivaswamy study [272], a SVM extension, in order to make the classifier invariant to the sub-elements permutation of each entry, is proposed, demonstrating the method applicability in character recognition, 3D object recognition and in several UCI data. This extension, called permutation-invariant SVM, can be described by the following steps: Given a training set and a maximum number of iterations, the method calculates the centroid, the database radius and the hyperplane. Then, the Kuhn–Munkres algorithm is used to find the permutation for each example pattern that brings it near to the sphere centroid, ensuring that its decision limit margin only increases. Next, centroid, hyperplane, new radius and margin are recomputed in the exchanged data and the previously described process is repeated until it reaches the maximum iteration number. The author demonstrates the methods efficacy for 3D object classification with three experiments.

Raptis [240] proposes a system for objects/faces caricatures recognition. The innovation introduced is the 2D object/face caricatures, which are obtained in 3D and fused in terms of their contours. Additionally, these features are directly connected to all objects and faces stored in the database. A face/object is thus considered as the output of a detailed probabilistic Bayesian analysis of views contours. The features used are the object/face edge pixels that are extracted from the edges of low-level information. The faces were separated from the background using the C-means for two clusters (background and main) for the first view and the classifier employed used a nearest neighbor approach. Both pattern types are modeled as distributions, since they are vague due to imperfect lighting conditions and different facial postures.

An approach to transform 3D objects into string features, which represents the voxels distribution under a voxel grid, is presented by Assfalg [18]. In this approach, initially, 3D objects are mapped into a voxel grid. Then, the strings features computation is performed, which is obtained by iterating through the grid once for each dimension (x, y and z), creating a feature string for each dimension. After repeating this procedure for all dimensions, a three-dimensional object is described by a set of three strings. A basic measure of similarity in feature strings, called spectrum kernel, is used in order to determine the similarity between two input strings. Therefore, given two objects, the spectrum kernel between the objects representing strings is computed, thus obtaining three similarity values referring to the two object axes. These similarity values can be unified into a single value representing the similarity value between objects.

A technique for classifying 3D objects using a Global Geodesic Function (GGF), to intrinsically describe the object surface, is proposed by Aouada [12]. In order to compare objects efficiently, each object class is characterized by two parameters in the learning stage, class resolution feature and a threshold value. All classes are sorted in ascending order based on the resolution value, and a superclass is constructed by merging classes that share the same class resolution feature parameter. The process begins at the lowest resolution and initializes a control variable L with value 1. Then, the object GGF is computed, in the feature class resolution, in the position L and its resolution parameters are obtained. Lastly, a object resolution parameters comparison is performed against the class parameters in the L position, and if the similarity is established the search finish. Otherwise it moves to the next position $(L + 1)$, continuing the comparison process.

Xing [319, 320] proposes a system for 3D object reconstruction and recognition. This system consists of three subsystems: structural description extraction based on superquadric, objects reconstruction with multiple parts and 3D object recognition. The superquadric, used by the author previously in the 3D object classification by merging parts [318, 321], is a family of parametric formats that can describe a wide variety of 3D primitives formats with compact parameters. The superquadric-based structural description of 3D objects is implemented in two feature levels: geometrical and topological. The multi-part reconstruction system reconstructs the 3D models with different numbers of parts and different formats visually showing the 3D model. For the recognition between the 3D models library and a set of unknown objects a method formed by two stages, tree search

and similarity measurement are presented to obtain the classification results.

Drost [59] proposes a method that creates a model global description, based on oriented point pair features, matching this model locally through a fast-voting scheme. The model global description consists of all the model point pairs features and represents a point pair mapping in the feature space for the model, where similar model features are grouped. In the method offline phase, the model global description is created. In the online phase, a set of reference points in the scene is selected and all the other scene points are paired with the reference points to create point pair features. These features are matched with the model features, contained in the global description model, and a set of possible matches is retrieved. Each potential correspondence votes for an object pose through an efficient voting scheme.

Westell [307] presents a system for recognizing and locating objects inside closed and unknown environments. In the system, the user can describe the target object through format, size or color descriptions which uses these descriptions to automatically select the target object from an object database. When entering an environment and capturing multiple scene images from different viewpoints, an object recognition algorithm is used to locate possible matches and a stereo imaging device is employed to obtain the 3D coordinates of those matches. The recognition algorithm has two stages: data extraction and object identification. In the data extraction stage, the database color composition images are used to construct a unique representation of each object. In the object identification stage, a trust map is produced, based on the database match information for the scene, to identify an object within a scene image (Fig. 10).

Wohlkinger presents three approaches for object representation and matching. In the first approach, Wohlkinger [309] proposes to explore contextual knowledge through the use of sensor and hardware constraints, in the robotics and home domains. The proposed work starts with the requirement to invent a framework that can easily be extended to a new class of objects. To this end, the internet is used as the source to obtain the models for new objects. These data are used to calculate object classifiers by extending the 3D descriptor harmonics to the robotic domain constraints, matching them against the database to find the class closest to the object. For the models acquisition part based on the internet, the same approach presented by the author in [310] is adopted. In the second approach, a format descriptor based on format functions, for partial point clouds, is presented [310]. This descriptor, called Ensemble of Shape Functions (ESF), has the capability of training in synthetic data and classify the object provided from a depth sensor in a single partial view. The classification task is defined as a 3D retrieval task, locating the nearest neighbors of synthetically generated CAD models views, to the point cloud generated with a Kinect like depth sensor. The models ESF correspondence with the point cloud ESF is performed by means of the L1-distance. In the third approach, the author presents an 3D descriptor adaptation for the D2 shape distribution, originally defined solely for 3D model matching [311]. The presented 3D format descriptor adaptation for 2.5D data allows the calculation of real-time features directly from the 3D points. In addition, it is shown how such a descriptor can be used in a framework, which uses a semiautomatic approach to acquire from the internet the data required for training.

Chen [44] presents a scheme for automatic object detection. In the proposed approach, the object detection scheme can identify target objects automatically in depth images using an initial object segmentation process, to subdivide all possible objects in the scene, and then apply a classification process based on objects geometric constraints and angle-of-view histogram. The method of segmenting objects in the scene is detailed in [44]. The classification framework uses a combination of several object features and its objective is achieved by finding the relevant dimensions correlation, the point density and the vision histogram of angles between the segmented point cloud and the object models stored in a database.

Liang [165] presents a 3D object recognition and pose estimation method using a deep learning model. Initially, two deep belief networks (DBN) are trained separately

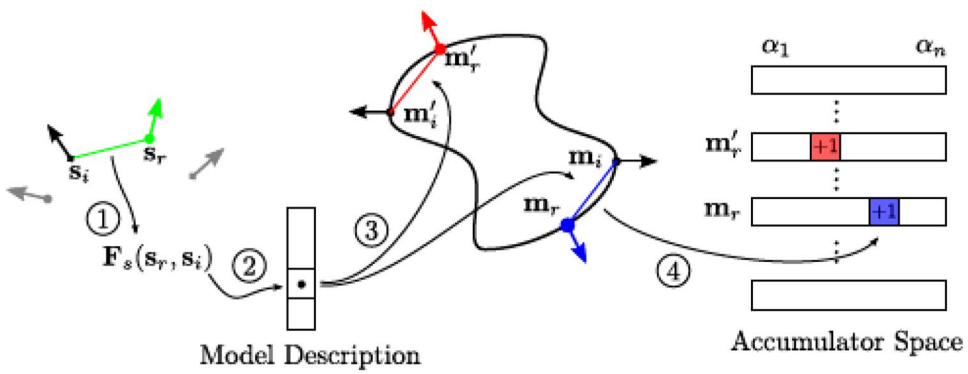

Fig. 10 Representation of the method proposed by Drost. Figure extracted from [59]

before connecting the last layers to train a classifier. To overcome failure to detect objects in the deep learning model, an object detection method based on K-means clustering is used, thus extracting the object from the background before recognition. The methods based on deep learning have the ability to recognize or predict a large patterns set by learning sparse features from a small set of patterns. For this reason, the deep learning model can be trained with a small pose number and then predict a large set of poses with the trained model. In the proposed system, different object poses imply in different classes, which means that a class represents a object pose in the DNB model. Therefore, a pose estimation problem is simplified to a classification problem.

Ribeiro [245] presents an effective global localization technique using soft 3D object recognition to estimate the pose according to a given map reference points. A depth sensor acquires partial view for each observed object, from which the proposed algorithm extracts the robot relative pose to the object, based on a library of Partial View Heat Kernel (PVHK) descriptors. In the proposed algorithm, the same approach used in [35], to compare the descriptors and to estimate the probabilities, is employed. Also, the distance function was used as proposed in [35], i.e., the descriptors are matched by comparing the curves format defined by the graph, where the temperature is plotted as a limiting length function. The localization algorithm was validated in a diverse set of experiments in a closed environment using everyday objects.

Rocha [246] presents a framework for 3D object recognition in an industrial context. The object recognition is realized in two different phases, learning and classification. In the learning phase, a data sampling and a segmentation process are performed. Then, following the segmentation procedure, the feature extraction is executed, where five different features were selected: dimensions (height, width and length), part surface area convex hull and the holes mean diameter. Based on these features organized into a transformation vector, a SVM was trained. In the classification phase, the same process chain used in the training phase is employed in the initial steps: sampling through the HRI table, HRI data segmentation and feature extraction. Lastly, the object classification based on the trained classifier is performed.

Beksi [26] presents a dictionary learning framework using RGB-D cloud-point covariance descriptors to perform object classification. The dictionary learning in combination with RGB-D covariance descriptors provides a point cloud data compact and flexible description. The covariance descriptors encapsulate features (position, color, normals and so on) on the object's point cloud by means of a single positive definite matrix, which characterizes the object. These covariance descriptors are used to create a dictionary representing the object, and a set of these dictionaries can be used to classify a new point cloud into a object class.

Luo [178, 179] introduces a model-based 3D object recognition method and search through a 7-DoF robot with online obstacle prevention for factory automation. The complete system is designed using a state machine divided into three stages: teaching, working and idle stages. Each stage is divided into some sub-states, and focusing on object classification and pose estimation, these particular sub-states will be analyzed. First, a recognition database is generated, offline, using the object CAD model. Following the recognition pipeline using global descriptors, for each obtained frame, a preprocessing step is performed to segment the object's cluster. The preprocessing includes: subsampling, outliers removal, region of interest segmentation, RANSAC planes segmentation and object clustering. The final result is an object's point cloud cluster from which the global descriptor is computed. After that, an approximate search for the nearest neighbor is performed in the database, selecting the k best candidates from which the most similar candidate will be elected.

A method for efficient 3D object recognition with occlusion is presented by Xia [314]. The proposal to use a method based on deep learning begins with the construction of a multi-view format model based on 3D objects. This model is constructed through the use of a coding/decoding deep learning network to represent the features. The network used to learn the features is a restricted Boltzmann machine (RBM) block composition associated with a deep belief network (DBN). The training objects feature set are learned by the DNB and used as input by the random forest algorithm to classify the objects according to the labels representing several object classes.

Zhu [340] presents a boosted cross-domain categorization (BCDC) framework that uses labeled data from other domains as auxiliary data to expand the original learning system intra-class diversity. The presented classification framework operates jointly with a cross-domain dictionary learning method [339] and shares the same basic principles of the training instances impacts sequential updating, but it attempts to sequentially update and reproduce the dissimilar data samples representation instead of assigning a lower weight to them. The BCDC has as input the target domain tagged data, the auxiliary domain data, the maximum number of iterations and the weak classifier. The algorithm initializes the data distribution as uniform. Then, the cross-domain dictionary learning method is applied to the auxiliary and target domain data, previously initialized with the uniform distribution. Next, for the provided number of iterations, the data distribution is set and the learned data dictionary is reproduced as an additional set of auxiliary data, based on the auxiliary data domain under the data distribution set previously. After that, the hypothesis is

computed, the hypothesis error is calculated, the factors are set and the new weight vector is updated. The output from this process are: a strong classifier and an updated auxiliary domain of instances representations.

Tateno [283] proposes a framework that is capable of conducting a real-time incremental 3D scene segmentation while being reconstructed via Simultaneous Localization and Mapping (SLAM). The proposed framework starts with the 3D reconstruction, which is based on a Kinect Fusion approach, a real-time method that estimates the camera pose of a mobile sensor and depends of the volumetric representation called truncated signed distance function (TSDF). Successively, the incremental segmentation algorithm, based on a recent approach proposed in [282], is applied. However, contrary to the approach in [282], the incremental 3D segments merge is performed within a specific representation based on voxel called label Volume. Finally, the recognition part, inspired by the global descriptor pipeline proposed in [9], is performed to provide the 3D matching between a scene depth map and a rendered views set of each 3D model (Fig. 11).

Kasaei [122] presents an object descriptor called Global Orthographic Object Descriptor (GOOD) built to be robust, descriptive and efficient to compute and use. The proposed descriptor is constructed with the following steps: Initially, using an object point cloud, the PCA method is applied and the object three principal axes are determined, applying a disambiguation method to define the three principal axes directions and to calculate a local reference frame (LRF). With the LRF calculated, the next step is to concatenate the object orthographic projections in the three orthogonal planes. Each projection is described by a distribution matrix with the same bin's number and size, to ensure the correct comparison between different object formats. To ensure invariance with the point cloud density, the matrix is normalized and then converted into a vector. The three projection vectors are concatenated, thus producing the descriptor. Several experiments were performed to evaluate the object descriptor performance in relation to the description power, scalability, robustness and efficiency.

Naji [207] presents an approach for 3D object recognition based on heat equations. These equations are used to calculate the geodesic distance between any point pairs, in the Riemannian manifold, using a heat kernel transformation. The adopted recognition system can be described by the following steps: heat equations calculation for the 3D objects in the database; basic features calculation derived from the heat equations; classifier training based on the previously calculated features; and test objects classification into classes using the trained classifier.

3.1.3 Global and local features

The analyzed works, presented in the last few years, demonstrated a possible tendency for the feature-based representation. This tendency is related to the jointly use of global and local features. Some analyzed works tried to combine the local features discriminative power with the global features representation, seeking a most discriminative and efficient representation type. Those works are briefly described in this section. Figure 12 shows some examples of works that employ global and local features for the object representation.

Ayoub [21] proposes a method for recognizing 3D objects captured by an active stereo vision sensor. The recognition and classification system proposed is divided into three parts: data acquisition, feature extraction and classification. The data acquisition part is performed through a stereo

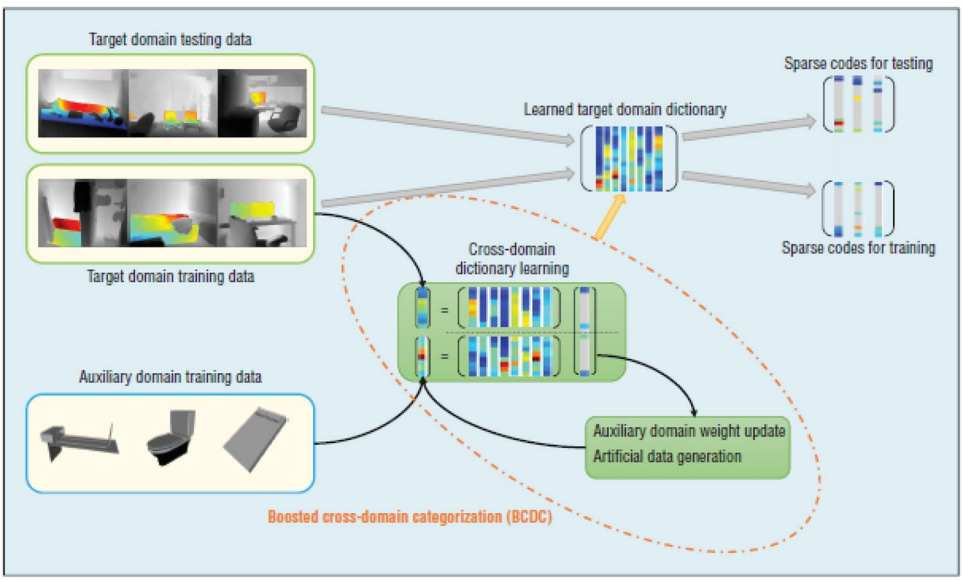

Fig. 11 Representation of the method proposed by Fan Zhu. Figure extracted from [340]

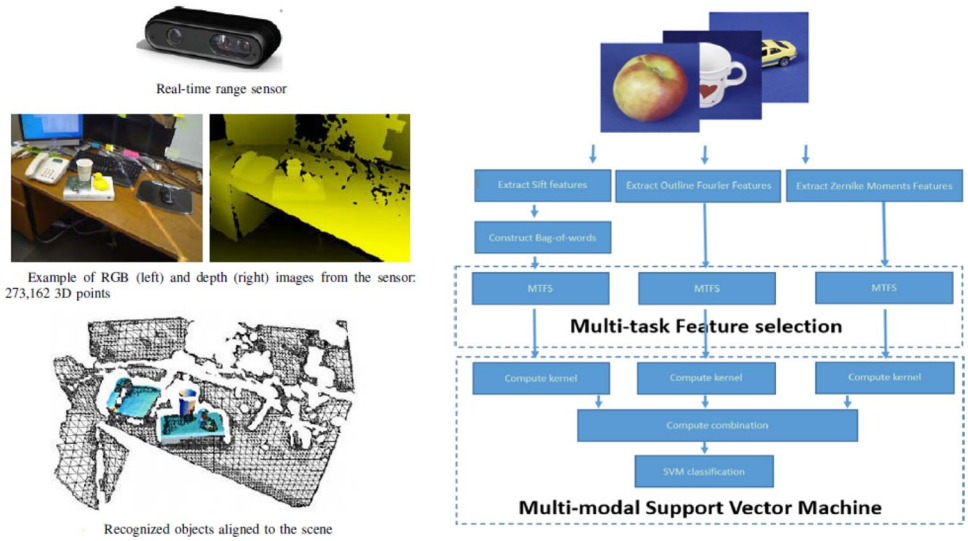

Fig. 12 Representation of analyzed works that employ local and global features. Composition of images extracted from [41, 131]

vision system and the preprocessing techniques application, to improve the acquired data quality. At the end of this part, a 3D point cloud is obtained. In the feature extraction part, given the acquired point cloud, features that will be used in a SVM classifier are extracted. In the classification part, the SVM classifier is trained and evaluated with two different features sets. The first set is the traditional feature set consisting of the 3D points from the respective point cloud surface. The second set is a suitable feature set based on depth histogram.

Eunyoung Kim [131] presents an approach for object recognition that boost the dissimilarity between queried and similarly shaped objects by maximizing the visibility context use. A point pair feature was designed, containing a discriminative description inferred from the visibility context. Also, a pose estimation method has been proposed, which locates the objects using the point pairs correspondence. The approach's initial step is to estimate surface normal and point sampling based on multiple point mesh resolutions, thus obtaining a simplified mesh. Then, three visibility types are computed based on the depth value of each image pixel and its corresponding 3D position in the world coordinate. After that, point pairs features are calculated based on two oriented 3D points, so that each coordinate value is mapped to an integer value, using these values as keys to find a set of corresponding point pairs extracted from the models. For each match, the potential corresponding model pose in the scene is computed and voted for final recognition. In order to reduce the computational complexity of calculating such features, a new point pair feature, that contains inferred descriptions of the previously described visibility contexts, is proposed. The visibility context boosts the discriminating power of each point pair feature, through the implicit imposition of consulted model global features, and reduces the number of spurious matches.

Sánches [249] introduces a descriptor designed for object class representation. The so-called SCurV descriptor explores 3D format information by computing and incorporating the surfaces curvature and the projected local surface points distributions in a 3D object-oriented and view-dependent descriptor. These different information sources are combined in a simple but effective way of combining different features to improve classification results. Therefore, the proposed descriptor is the result of computing the following quantities: an object-centered global representation based on surface curvature; a local representation centered on a viewpoint providing degrees of flatness, concavity and convexity; and the final descriptor, which is the result of the tensor product computation between the two previously calculated representations. To test the proposed 3D descriptor for object classification, the classifier margin-based regression, which is an SVM extension, was employed.

Garstka [79] proposes an adaptive approach for 3D object classification. In this approach, appropriate algorithms for 3D point cloud feature description are selected via reinforcement learning depending on the objects' properties to be classified. The proposed approach objective is the autonomous learning of a combined and optimized application of several 3D feature description algorithms for the purpose of increasing the 3D point clouds overall classification rate. The main steps for classification are: Given a 3D point cloud, a collection of global properties are extracted. These values are used by the reinforcement learning agent to select the first algorithm. In the second step, the Intrinsic Shape Signature algorithm is employed to determine the points of interest. During the third step, one of the local 3D feature description algorithms will be applied. As result, a local 3D feature description set is obtained. Each of the determined feature descriptions is quantized to be binned in a histogram. In the last step, the histogram values are used as input vector

for an SVM classifier to identify the object appropriate class. This classification pipeline is enhanced by a reinforcement learning agent.

Naguib [204] proposes a classifier based on Tree-Augmented Naïve (TAN) Bayesian Network. The employed feature space was separated into true/false regions, which allows to drive the Bayesian a priori conditional probabilities inference of a statistical database. The true/false regions were also used to estimate the expected posterior probabilities of each object under specific active conditions. These expectations are used to select a set of optimal features under this environment and, autonomously, to rebuild the Bayesian network. The complete system can be described by the following steps: system training; target object acquisition; octree segmentation; probability distributions update for all database object features; construction of a discrimination power table and discrimination probability calculation; optimal feature set selection (height, mean width and SIFT) and conditional probability table construction; computation of the reliability associated with these sufficient conditions; optimal feature set measurement; and feature probability calculation to correspond to the target object.

Kasaei [120, 121] presents an efficient approach capable of learning and recognizing object categories in an iterative way, without the need to know objects in advance. The first step is object detection, which involves distance filtering, subsampling and object's point cloud clustering. The object detector periodically requests a list of all objects currently at the table top. The object detection module creates a new perception pipeline for each detected new object. Each pipeline includes modules for object tracking, feature extraction and object recognition. The object tracker works based on a particle filter, which uses geometric information, as well as color and normal surface data, to predict the most likely next object pose. The object tracker sends the object point cloud to the feature extraction module, which computes the feature for the object view provided by means of a 3D-format descriptor (spin images associated with key points extraction). The object features are kept in memory, and the user can provide the category labels for these objects. The object labeling, manipulated by the user interface module, triggers the object conceptualization module. In such situation, the object conceptualize reads the memory with object current category, as well as the feature set describing it, and creates or enhances the object category. During recognition, a classification rule by nearest neighbor is used to estimate the detected object category label.

Lee [157] presents an approach to accurately recognize industrial objects and estimate their poses based on a Bayesian framework with optimal feature selection and model-based point cloud matching. The framework consists mainly of two parts: 3D object recognition with multiple evidences and model-based 3D object pose estimation. In the recognition process, candidate evidences include global and local features extracted from the RGB-D data, i.e., the 3D SIFT, CLB and shape descriptors extracted from the point cloud. After collecting the candidate features, the feature selection based on the t-test is applied to choose sufficient features for different target objects as support evidences. Finally, the measured features are compared with each object in the database using a Bayesian network. The most likely object is considered as the recognition result.

Chen [41] proposes the SVM use associated with three feature representation modalities for 3D object classification. The proposed framework begins with the feature extraction through SIFT, Outline Fourier and Zernike moments from a database. In order to make a discriminative representation, a relevant feature subset is selected from the features set shared by the representation modalities. Based on the features selected, the next step is to train the SVM based on a multi-kernel SVM approach, which maintains each model representation independence while using each modality features in the classificatory process. The final step is to use SVM to classify the new object being searched.

Li [162] presents a hierarchical semantic segmentation algorithm, which partitions a densely cluttered scene into different object regions. A typical convolutional architecture involves two main steps: local filters convolution over input signals and filter responses pooling within a predefined neighborhood. In Chi Li's work, two alternative pooling spaces were explored: SIFT (gradient) and FPFH (3D geometry). The traditional pooling in the space domain cannot be applied directly to high-dimensional pooling domains such as SIFT and FPFH due to the exponential pooling bins growth number. Thus, a generalized pooling approach based on K-means and on the nearest neighbor search, for arbitrary pooling domains, is presented. First, a proposed region hierarchy is presented, which avoids relying on the region fusion heuristics used in most scene segmentation techniques. This region hierarchy exploits a large set of partial object regions ranging from local to global patterns. Then, the multi-domain pooled features are efficiently propagated across this generic region hierarchy, and the semantic regions labels, on all scales, are combined for robust semantic segmentation. After the semantic analysis, the original point cloud scene is partitioned into regions with homogeneous semantic labels. Lastly, similarly to [161], the object postures are estimated for each segmented class with an objRecRANSAC.

3.1.4 Spatial maps

The last of feature-based representation subclasses are the works that describe the 3D object with the spatial maps representation. This representation type describes the object by capturing and preserving physical locations on the object [20]. The most suitable definition found for maps was: "a

symbolic depiction emphasizing relationships between elements of some space, such as objects, regions, or themes" [34]. This definition agrees with the map usage on object recognition, due to the fact that, in this area, a map representation is employed to represent an object in terms of distance or some other symbolic meaning. Figure 13 shows some examples of works that employ spatial maps for the object representation. The analyzed works that employ map as representation are shown by Sen Wang, Kordelas, Donghui Wang, Atmosukarto, Pintilie, Eunyoung Kim and Rodrigues.

Sen wang presents two works, where a framework [303] and its use on object recognition and 3D partial surfaces stitching [304] are proposed, employing a Least Squares Conformal Shape Images (LSCSIs) generated from least squares conformal maps (LSCMs). This representation simplifies the 3D shape correspondence problem to a 2D image matching problem. Also using a map representation, Kordelas employs a distance map representation for 3D object recognition in cluttered scenes [144, 145]. The proposed approach can be divided into two parts: distance map extraction and storage for each 3D object possibly on the scene and distance map extraction from the scene for further matching between scene and objects distance maps, performed through a similarity metric application.

Donghui Wang [301] presents a method for recognizing 3D objects from depth images under an arbitrary pose by means of fast sphere correlation. First, all the extended Gaussian image (EGI) views under different viewpoints are extracted and combined into a Gaussian sphere to form a feature description for each object. Then, the examined depth image, in an arbitrary pose, is represented as a phase-encoded Fourier transform (PFT) feature. This PFT feature is mapped on a Gaussian hemisphere by coordinates transformation and intensity scaling. Next, the spherical correlation algorithm, based on spherical harmonic functions, is used to match and measure the similarity between the mapped PFT and the combined EGIs. Both, the combined vision EGIs and the PFT vision EGIs, can be considered as two feature functions in the S^2 unit sphere. Thus, the 3D object recognition task from a depth image can be converted to a spherical correlation between two spherical functions.

Atmosukarto [20] shows a 3D format representation and a methodology for classification developed for craniofacial dysmorphology studies. This work is based on another work presented by the same author in 2008 [19]. The proposed methodology begins with the objects rescaling to fit them to a fixed-length bounding box. The next step consists of two phases: feature extraction at a lower level and feature aggregation in medium level. In the purpose of this study, the salience points, determined at the medium-level feature aggregation, that best serve for the craniofacial disorders classification are used. To find the 3D object salient points, a learning approach was selected. A salient point classifier has been trained with a set of training points provided by application experts. For classification, most methods require the use of a 3D descriptor or a signature to describe the object format and properties. The proposed method signature is based on 3D salient points mapped in 2D planes through a longitude–latitude transformation. The 3D object classification is then accomplished by training a classifier using the 2D object map, classifying into animal heads, human or specific objects depending on the database used.

Another example of spatial map use is presented by Pintilie, where an approach to calculate conformal map and use it for 3D object classification is shown [234]. The proposed approach for conformal map computation goes through an optimization problem and is explained in detail by the author. For the classification, the approach's basic idea is to classify the transformations that maps the 3D object into a 2D conformal map instead of classifying the raw data directly, i.e., the transformation that produces the conformal map will be used to classify the objects.

In Eunyoung Kim work [132], a framework to classify free-form objects in point clouds is presented. The proposed framework initially segments the scene object candidates and then identifies the class for each candidate. The framework can be briefly described in three main parts: the database hierarchical structure, the online learning process and the object classifier. The model used to construct the hierarchical structured database (HSD) was the TAX model, which constructs an object classes hierarchical model from unlabeled depth images, by mapping each image to a path in a tree composed by L nodes. To this structure, an object representation called visual word was proposed. In the online learning process, a learning procedure, which incrementally updates the existing HSD,

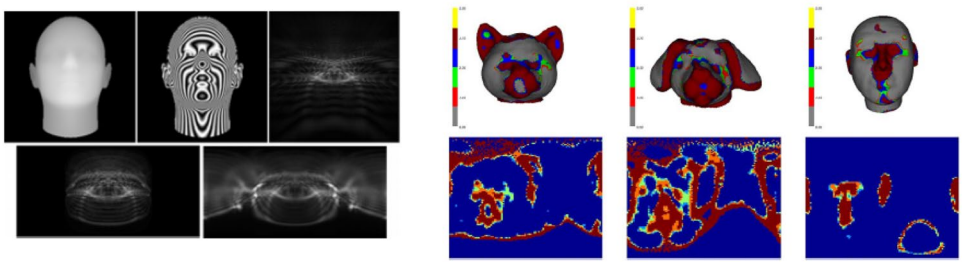

Fig. 13 Representation of analyzed works that employ spatial maps. Composition of images extracted from [20, 301]

is used to infer the HSD structure. The object classifier uses information from object patterns, the distribution, the structure assembled by HSD, the k-mean and the Bhattacharyya coefficient to infer the object labels.

The work proposed by Rodrigues presents a cortical model for 3D face recognition from their 2D projections [250]. In Rodrigues work, first the number of 2D feature templates required for the representation of all views was studied. Each face template is represented by symbolic and salience maps; more specifically, each template is represented by 400 maps (5 views × 4 type of events x 20 scales). The recognition scheme compares the input image representations with the templates previously computed through a similarity metric.

Lastly, we have the work proposed by Yu [331], where an approach for face recognition using 3D directional corner points (3D DCPs), which combines structural connectivity information with spatial information from 3D faces, is presented. In this approach, the 3D surfaces are represented by 3D DCPs derived from ridge and valley curves. After the valley and ridge curves detection, on the 3D surface, a corner point detection process, which is based on the Douglas–Peucker algorithm, is applied to generate the 3D DCPs. Then, after representing the surface with the 3D DCP, a point-to-point conversion process was developed to calculate the difference between two 3D DCPs. The dissimilarity between the two faces is then calculated through a global conversion process between two 3D DCP sets. Both the face to be matched and the face in the face gallery go through the same processing steps: normalization process, ridge and valley curves detection and description through 3D DCP, which are finally matched.

3.2 View-based representation

The view corresponds to an image representing how something, e.g., object, landscape or scene of interest, is visualized. Figure 14 shows some examples of works that employ views for object representation. There are few analyzed works that use view as the main representation either in the best view selection or in multiple-view analysis, searching for correspondences between them. For example, Deinzer presents an approach to solve the optimal viewpoint selection and fusion problem for optimal 3D object recognition [53]. His approach can be defined as a two-step approach, where in the first step, action-value function estimation occurs, and in the second step, if in any moment the viewpoint fusion returns a state S as classification result, the camera movement that maximizes the expected cumulative and the reward weight are selected. The presented approach can be described as an optimization approach, which can be solved by global adaptive random search algorithm application followed by a local simplex. In his experiments, the author showed the influence of optimal viewpoint selection on object recognition performance.

Different from Deinzer approach's, the work presented by Rui Nian shows a probabilistic scheme for 3D object recognition from 2D view sequence [211]. The proposed scheme can be described in three steps: preprocessing, model view-based learning and recognition. In the preprocessing part, collected database images, which consists in 2D video sequences of object's types, are processed in order to extract image features and suppress possible artifacts that may be difficult for the object recognition. In the model view-based learning part, views are clustered together and the model views are responsible for minimizing distances

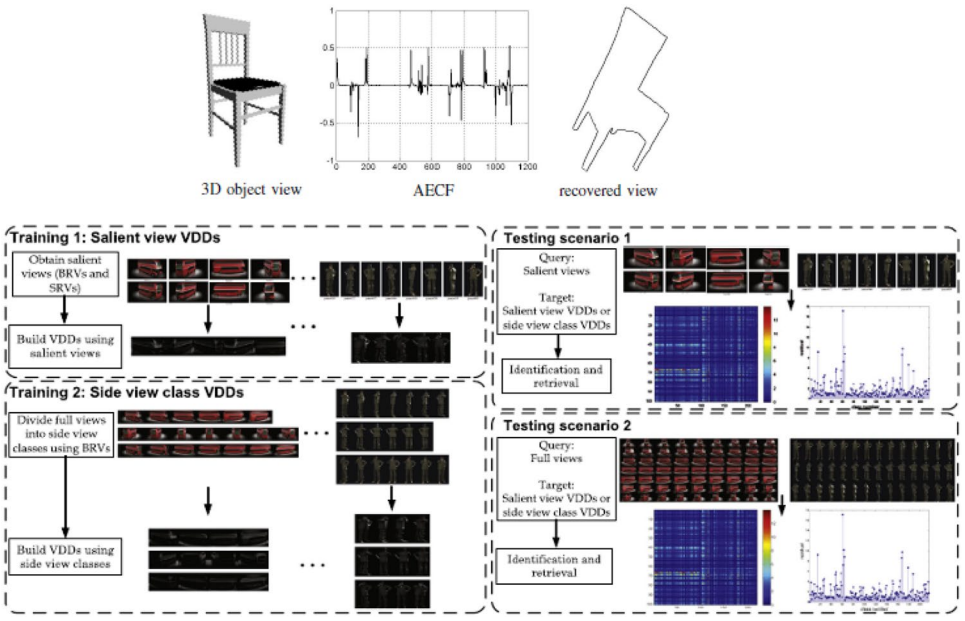

Fig. 14 Representation of analyzed works that employ view-based representation. Composition of images extracted from [45, 296]

for all other elements in the same cluster. The model views are composed by generalizations of members in each cluster, which corresponds to different views of each object. The recognition part is performed through probability density function estimation, i.e., for a given input image is decide whether the object is present or not in the scene based on which probability is higher.

Another example of 3D recognition method from a small number of 2D images taken from arbitrary viewpoints is presented by Zografos [343]. The proposed system starts with the selection of two base views, followed by the selection of a number of correspondent reference points, located in points of limits discontinuity, edges and other prominent features. When an appropriate number of reference points was selected, the Delaunay triangulation is used in order to produce consistent and correspondent triangular meshes for all the images. The recognition system itself involves choosing the appropriate coefficients from a linear combination of views (LCV), thus synthesizing an image which is compared to the target image using a similarity metric.

Other methods that employ view-based representations are presented by Jun-Hai Zhai, Vázquez, Luciw, Polat, Dimov, Urdiales, Bo Pang, Ulrich, Elons, Domingo Mery, Efremova, Guan pang, Faulhammer and Yi-Chen Chen and are further presented in this section.

Zhai [334] proposes a 3D object recognition method based on view, which consists of three steps. In the first step, a wavelet transformation is used to decompose object view images into different frequencies sub-images. In the second step, for each sub-image, features are extracted using a singular value decomposition (SVD) approach. Since an image can be viewed as a matrix, the SVD can be used to extract image features. These extracted features are combined to construct a feature vector from the original image. Finally, in the third step, the constructed vector is fed into a SVM to classify other objects.

Vázquez presents a method based on views and some biological aspects from the children vision in the early stages of life for 3D object recognition [298, 299]. The biological aspects used are related to response to low frequencies in the early stages and some conjectures about how a children detects subtle characteristics of an object. For the low-frequency response, the proposed method employs low-pass filter to remove image high-frequency components. Then, subtle image features are detected through a random selection of stimulating points. Lastly, as learning device, a dynamic associative memory (DAM) is used to learn features and perform object recognition.

Luciw proposes a Topographic Class Grouping (TCG) mechanism, which explains how top-down connections influence the feature detector type developed and their placement in the neuronal plane [177]. The top-down connections boost variations in the neuronal plane between class direction during the training step, where different views from the objects are used for training. The proposed mechanism demonstrates an increasing distance between input samples belonging to different classes, which results on a larger separation of neurons belonging to different classes. Therefore, neurons that answer to the same class stay relatively close. After the training step, the multilayer in-place learning network used for the mechanism development is employed on 3D object classification task.

The work demonstrated by Polat uses genetic algorithms (GAs) and general regression neural network (GRNN) [235] for pattern recognition based on 3D object poses/views [236]. His method applies the GA for GRNN optimization; thus, the first step is initial population generation, which is generated randomly from the database. Followed by fitness value computation, selection of individuals, mutation and crossover generating a new population and repeating this process until the stop condition is reached. The AG output is then used to train the GRNN for object classification without feature extraction.

A heuristic approach for 3D object recognition, through multiple 2D projections of the object of interest, is presented by Dimov [56]. In his approach, the object identification is interpreted as a conventional content-based image retrieval (CBIR) problem, where an arbitrary input image, from a given object, is treated as a search sample inside a big database with a set of object images (appearances) from several views from the object. A given object in front of a camera is considered a dynamic 3D object represented by a series of 2D appearances projections. If an appropriate number of those images are stored in an image database, then one can search, in that database, the image more similar to the input image. Furthermore, it is possible to localize a sequence of images in a descendant order of similarity with the input image.

Urdiales presents a method of view planning, in order to choose the best view sequence for 3D object recognition [296]. The proposed method works as follows: first an initial view is acquired and objects in the database, which the differences with the input vector are lower than a determined threshold are found. Then, the map of candidate clusters is compared aiming to reduce the number of existing candidates. After that, the difference between each two cluster maps is computed and accumulated to check in which points they all differ more. In order to avoid different view learning problems, when the second view is acquired all the candidates are realigned. In this point, the maps are correctly aligned and the recognition, based on hidden Markov model (HMM), can be applied to the views with maximum difference.

Bo Pang proposes a way of describing 3D models by a series of 2D projected images, applying this description on a 3D object recognition system [224]. The system can

be described through two main parts: training and recognition. In the training part, a set of 2D projected images are selected for effectively represent each 3D model. The selection process uses feature extraction, through Zernike and Fourier descriptor and trace features, and implements an effective view selection with feature merging and clustering by multiple learning, obtaining key views from the projected images. In the recognition process, the input image extracted signatures are compared with the data of each model in the database and their distance is treated as dissimilarity score. The object with the lowest dissimilarity score becomes the searching result.

Ulrich presents an approach for 3D object instance recognition and pose determination in one single camera image [294]. Initially, a hierarchical model is generated based only on geometric information from 3D CAD model from the object. During the hierarchical model generation, only object geometric information, important for the recognition process, is included into the hierarchical model. The hierarchical model generation main task is to derive a 2D hierarchy of object views, which can be used to find the object efficiently in an image. The different object views are automatically created through a virtual camera positioning around the 3D object and by its projection in the image plane of each virtual camera. During the recognition phase, the generated hierarchical model is used to recognize the 3D object in a single camera image and to determine the object's pose according to the camera coordinate system, by applying a 2D correspondence method through the complete hierarchical model.

The work introduced by Elons shows a technique to deal with pose variations in the 3D object recognition process [66]. The proposed technique uses pulse-coupled neural network (PCNN) to generate a unique signatures through images acquired on different angles. Two parts compose this technique: model construction and recognition. In the model construction part, in the applied case for hand signal recognition, each hand signal to be recognized must be placed in a stable position in the center of a circular turntable and, with two cameras, the hand signals images are acquired. The acquired images are used to produce a signature through the PCNN, where the two signatures, referent to each camera, are weighted and linearly combined to produce a 3D signature of the image. After the signature database construction, a neural network, Multilayer perceptron, is used to learn and classify the input images, which pass through the same 3D signature generation process.

Domingo Mery proposes an automatic method based on multiple X-ray images from different views for regular object recognition [194]. The method is composed from two steps: monocular analysis, where it is possible to obtain detections on each view in the sequence, and multiple-view analysis, where it is possible to recognize interest objects using the correspondences in all views. The method also can be divided into two stages: the offline stage consists of the geometric model estimation and learning from multiple views. The online stage is performed, using the geometric model, in order to recognize the interest object in a test image sequence.

Efremova presents visual ventral neural network model for recognition and classification of 3D objects [62, 63]. The model represents a module hierarchy, which resembles V1–V4 areas and the inferior temporal cortex, and its architecture is based on the neural network concept named Self-Organized Map (SOM). More specifically, the concept of SOM associated with a radial basis function (RBF) network is used. This architecture was trained with a training set of different views from the objects and subsequently used for 3D object classification.

An approach described by Guan pang shows a 3D recognition method applied in point cloud [226], which projects the 3D point cloud in several range images from several viewpoints, transforming the 3D recognition problem in a sequence of 2D detection problems. To assure that the original 3D information wasn't lost, the 3D for 2D projection is performed in multiple-view angles. After the input point cloud be projected in multiple 2D view images, each 2D view is used to localize the target object. Then, all the 2D detection results are reprojected to 3D space for a combined 3D object localization estimation. The final object detection is performed only if several 2D re-projected detections occur in a nearby 3D region, thus filtering the detections of multiple views.

An online method based on multiple views, which combines acquired environmental information, merging individual recognition outputs from single views, is presented by Faulhammer [76]. The proposed method uses RGB-D data and transfers, continuously, the hypotheses constructed at several points in a framework to gather the maximum amount of information for all the objects in the scene. Additionally, the proposed approach allows the online recognition improvement, i.e., the recognition is improved on each new observation. The single-view recognizer generates, for each point cloud, a set of candidate objects (hypotheses) potentially present in the scene. These hypotheses are obtained using a single-view recognition system proposed by Aldoma [9]. After positioning refinement with ICP, a final verification stage returns a hypothesis subset which better represent the scene according to a global criteria [9]. To explore information from multiple views, a graph is created with vertices representing the information from a single view and edges connecting those views, when the views share a common object hypothesis. Using the constructed graph, the single views are merged in a unique global representation from the scene through the camera pose estimation.

A method to determine 3D object salient views is proposed by Chen [45]. The proposed method has two stages: In

the first stage, views are extracted, cropped and resized from a video sequence. Then, border scores are computed using a scatter-based metric to estimate the Boundary Representative Views (BRVs) and to determine the side view classes. In the second stage, for each side, a set of side representative views (SRVs), which better represent a corresponding side, is selected through the representation error minimization. Based on these two types of views, BRVs and SRVs, dictionaries named view-dependent dictionaries (VDDs) are built, which encodes object geometry information through views. The VDDs are then used on the object retrieval and recognition.

3.3 Graphs

Graph is a structure amounting to a set of objects in which some pairs of the objects are in some sense related. The objects correspond to mathematical abstractions called vertices and the relation of pairs of vertices named edges [291]. The graph representation is often applied in several applications being 3D object recognition one of them. Examples of works that employ graph for 3D object recognition are presented by Marini, Shengping Xia, Aouada, Noma, Huimin Ma, Bonev, Kuk-Jin, Mengjie Hu and Madi, which are further analyzed in this section. Figure 15 shows some examples of works that employ graph for object representation.

Marini propose a method to construct creative 3D object class prototypes described by structural signatures encoded by an attribute graph, which summarizes topological and geometric shape aspects [186]. With each class prototype computed, the classification of a queried model can be performed through model comparison with each class prototype. In this way, the structural class descriptor, coded as graph, is compared with the consulted object structural descriptor.

Shengping Xia uses R-SIFT features to construct a class-specific hyper graph (CSHG), which encodes in a comprehensive way, SIFT and global geometry restrictions [313]. Furthermore, the CSHG captures efficiently multiple objects appearance instances, using this trained graph structure to classify other input objects.

Aouada presents a method for partitioning 3D complex shapes into simple parts, focused on matching and recognition of 3D objects [13]. The object partitioning counts with object topological extraction using Reeb graphs. The recognition employs a kernel-based technique for Reeb graphs registration, aiming further pairwise comparison between primitive shapes through a distance metric.

Noma presents a sparse shape representation using graph, tested on 3D object and numeric digits recognition [213]. The proposed graph representation encodes uniformly spaced sampling of object's contour points, representing each point by a graph node. After the representation

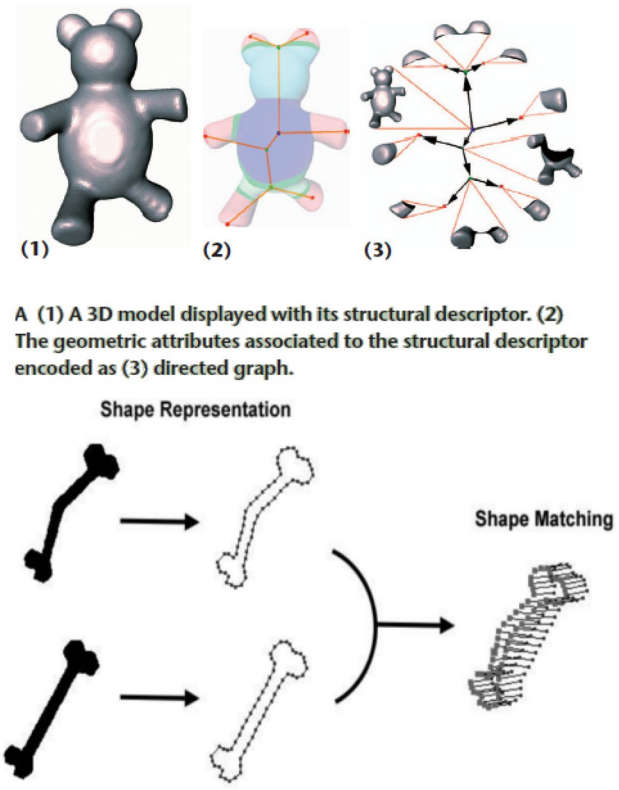

Fig. 15 Representation of analyzed works that employ graph-based representation. Composition of images extracted from [186, 213]

computation, the metric for object shape matching between the graph representations, based on sum of beliefs computed by belief propagation method, is defined. To test the representation and similarity metric, a KNN classifier was used.

Huimin Ma proposes a multiple resolution system for 3D object recognition [181] based on human visual model. The recognition system has two main parts: construction of an aspect graph library and queried 3D object recognition. After the aspect graph library construction, the queried object is compared, through Hausdorff distance, with the aspect graph representations stored, in order to perform the object recognition.

Bonev [30, 31] evaluates graph structure measures for 3D object classification. Initially, for each object, three Reeb graphs are extracted: one based on geodesic distance, one based on the object mass center distance and the last based on the distance from the center of the sphere that circumscribes the triangles mesh. For each graph, nine different measures are computed and transformed in histograms: Complexity Flow, Friedler, Adjacency Spectrum, Degrees, Perron-Frobenius, N.Laplacian Spectrum, Node Centrality, Commute Times 2 and Commute Times 1. All the histograms compose a bag of features, where a feature selection process is performed to select features that better represent a class for further object classification.

Yoon [328] presents a framework for 3D object recognition based on local feature invariant and their 3D information obtained from stereo pair. After local feature extraction, extracted with G-RIF [137] or SIFT, the features are matched between the left and right images and the 3D coordinates from the matched features are computed. To represent object and its 3D information, a directional Attributed Relational Graph (ARG) was used. The framework, basically, calculates the ARG representation for each object to be recognized and stores it in the database and then generates and verifies the recognition hypotheses by comparing the ARG representation from the target object with the ARG representations previously computed and stored.

Hu proposed a 3D object recognition method based on aspect graph aware [109]. The proposed method is composed of two stages: the offline stage, where the Bundler motion method structure is applied into the object data set and the background point are manually removed, to obtain the object point cloud model and generate, from this model, the aspect graph aware representation, and the online stage, where coarse 2D–3D correspondences are produced, by similarity computation between SIFT descriptors from input image and 3D model, and refined via a two-stage filter application, which removes false correspondences. Lastly, the object pose estimation is determined via RANSAC associated with EPnP and the localization is obtained through 3D model reprojection.

Lastly, Madi [183] proposes an algorithm for measuring the distance between 3D objects represented by triangular tessellations graphs. This distance is based on the triangular tessellations decomposition into triangular stars, which is a connected component composed by the union between a triangle and its neighborhood. The method, basically, describes the object by the triangular star representation and computes the distance, dissimilarity measure, between triangular star from two triangular tessellations, using this measure to better determine the correspondence between triangular tessellations pairs.

3.4 Models

The word model has a vast quantity of meanings depending on its application area; for example, there are mathematical model, 3D model, model theory, and so on. For the 3D object recognition area, one can also find several model types that are employed for object representation. Figure 16 shows some examples of works that employ models for object representation. Based on the analyzed works, we can provide some examples of models used:

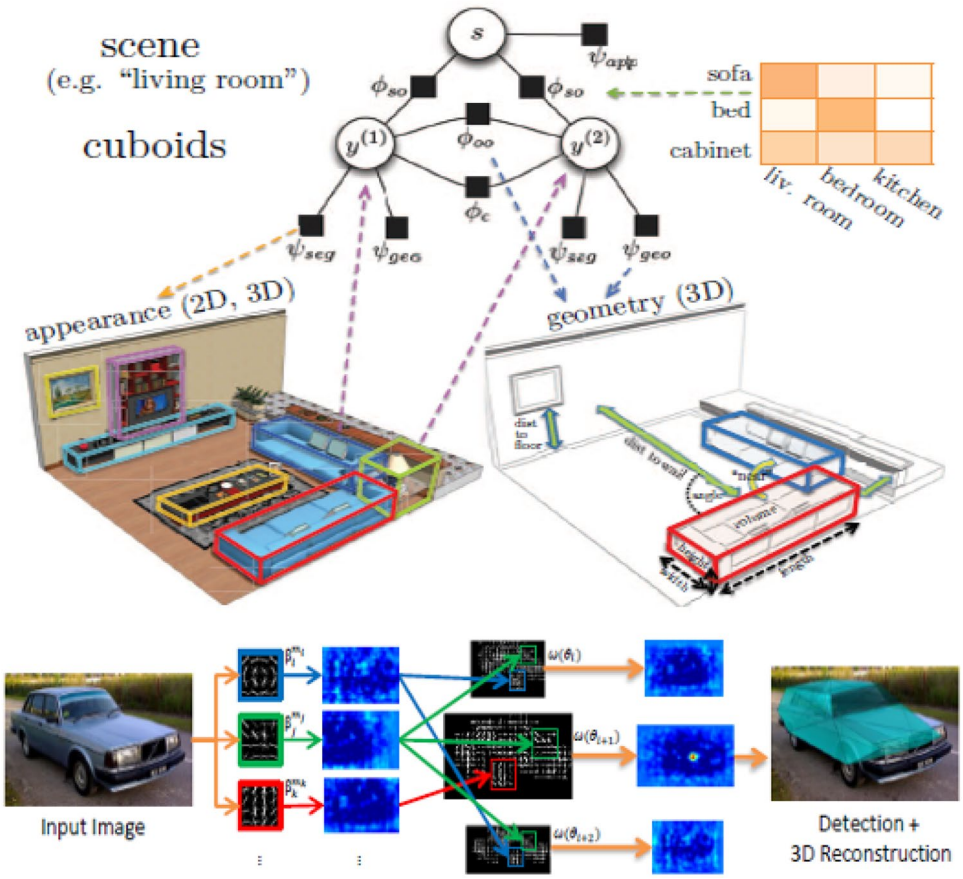

Fig. 16 Representation of analyzed works that employ model-based representation. Composition of images extracted from [101, 166]

- Geometric models: this category of model uses geometric features to represent and match 3D objects. Some examples of analyzed works that use this model type are presented by Aouat, Truong, Khatun, Sukhan Lee, Dahua Lin, Hejrati and Mabel M. Zhang. Aouat proposes a method based on quasi-invariant geometric features, divided into two stages: geometric model database construction and object model retrieval [14]. The object model retrieval is based on index calculation of different object viewpoints and matching between those indexes with the previously calculated geometric database model indexes. Truong uses pairs of perpendicular lines to represent the object visible face and to test the recognition of a parallelepiped model [292]. Then, using the coverage ratio estimation, the most accurate match between pairs of detected perpendicular lines and the database model is found. Khatun [127, 128] presents an approach for 3D shape representation using ellipsoids and, based on this ellipsoidal representation, extracts wavelet features which composes the 3D shape feature vector. This 3D shape feature vector, also classified as a descriptor, is tested on model representation description, calculating the approximation error according to the model. Deeply related to another work co-authored by him [176], Sukhan Lee [156] proposes a probabilistic approach, employing negative and positive evidences for 3D object recognition in cluttered scenes. The mains steps employed in his approach are: acquisition using a stereo pair; 3D line extraction; line grouping based on model constraints; polyhedral model representation described according to edges, vertex and surfaces; pose hypothesis generation and verification based on model features; and, probabilities computation and classification, using Bayesian rules. An CPMC framework [38] extension to 3D, aiming to solve jointly scene comprehension and 3D object recognition problems, is presented by Dahua Lin [166]. The proposed extension includes depth information, to the already use appearance information, into the physical and statistical iteration modeling between objects and scene, as well as inter-object iterations, in terms of hypothesis cuboids in cloud points regions. Then, for the jointly detection problem, a conditional random field is formulated to model contextual relation between 3D objects, based on the appearance, geometry and contextual cues integration. Hejrati proposes an approach for 3D object reconstruction and recognition [101]. This approach is based on analysis by synthesis strategy, where a forward synthesis model builds possible geometric interpretations of the world and the interpretation that better suits with visual evidence measured is selected. The method basically consists on the use of a non-rigid structure from motion to learn and estimate the base shape and the feature vector for each training example, followed by the optimal reconstruction inference through a brute force searching scheme. Slightly different from the aforementioned methods, the method proposed by Mabel M. Zhang presents a 3D triangle histogram for 3D object classification by tactile sensing. The proposed descriptor is built by a set of contact point sampling on the object surface and its information used to train a linear SVM for object test set classification.
- Covariance and Appearance Manifold-based models: This type of methods constructs models based on appearance manifold and convariance matrix to represent 3D objects for the task of 3D object recognition. As example of this model category are the works proposed by Lina [168, 169], where different forms of manifold construction are presented and compared, on 3D object recognition task. For the recognition with an input image, the Mahalanobis distance metric is employed;
- Surface models: surface-based models employ surfaces as object representation, performing recognition with the extracted surface features. The type of surface employed on the analyzed works varies. Kushal, for example, uses partial surface models, learned by feature pattern correspondence repeated through the training images of each class [149], as base representation for the correspondence between object model and test image. On the other hand, Ibrayev uses curved surfaces from tactile Data [110] to represent the object. For each surface model in a database, a look-up table is constructed to store the pre-computed principal curvatures. Then, for object recognition, a robotic arm with a touch sensor obtain data points in the object surface over three concurrent curves, which are used for comparison against the pre-computed principal curvature from the surface models.
- Distribution-based models: Distribution-based models represent the object through its features distributions. One example of analyzed work using this model representation is presented by Wentao fan, where a statistical framework for 3D object modeling and recognition is proposed. The complete process of model representation and recognition of test models is obtained by the application of a hierarchical Pitman-Yor (HPY) [285] process of Beta-Liouville [33] mixture distribution.
- Hybrid models: models belonging to this model category, combine different kinds or levels of feature information to compose a object model representation. These are the cases of the analyzed works presented by Anand, Raytchev e Kent. Anand shows an active vision-based system for 3D object recognition and pose estimation [11] which employs an autonomous robot team, data fusion of multiple sensors and a self-organization mechanism to complete the task. The combined model representation has as base a radial probability distribution function graph (PDF), a set of directional PDFs and a confi-

dence level tree. On the other hand, Raytchev proposed a model, named visibility map, which encodes a compact model from the 3D object through the use of different object views [243]. This model performs the different views encoding by using a binary vector, where each vector index refers to another vector with representative features. For the model comparison with test images, a suitable metric for the visibility subspace is used. At last, the work presented by Kent demonstrates a system for object model construction, recognition and manipulation enabled by web robotic advances [126]. The proposed system merges a point cloud pair, based on point cloud feature set and a set of correct and incorrect manual labeled registrations, to construct the model. The result is a object represented by a set of multiple object models, which is employed on the object recognition and manipulations tasks.

- Boltzmann machine models: this model is a type of network of symmetrically connected, neuron-like units that make stochastic decisions about whether to be on or off [103]. Inspired by this model, the work proposed by Nair presents a new high-level model for deep belief nets (DBNs), evaluated on 3D object recognition task [206]. The proposed model is a third-order Boltzmann machine trained using a hybrid algorithm which combines generative and discriminatory gradients. After training with the selected number of hidden layers and units per layer, the next step is testing the model on 3D object recognition task.
- Statistical models: this is a type of mathematical model that embeds a set of assumptions related to a sampled data, often representing, in considerably idealized form, the data-generating process [192]. A example using this type of model is given by Grzegorzek, where a probabilistic approach, for 3D object localization and classification in 2D images, is presented [87]. His approach can be divided into two stages: in the first stage, named the training stage, objects and context statistical models are learned separately. In the second stage, named recognition, one of the three recognition algorithms, embed on the proposed system, is used to classify and localize the object in a test image.
- CAD models: CAD stands for computer-aided design; therefore, CAD models are digital representations employed on creation, modification, analysis, or optimization of a design [86]. For 3D object recognition, 3D CAD models are frequently used as object model input, which are further processed into other type of simplified representation. However, some of the analyzed works use CAD models not only as a simple input, but also in the complete recognition process. For example, we have the work proposed by Muhammad zia, where CAD models are employed on scene interpretation problem, through their use on object's occurrence probability computation for further combination into a probability of a particular scene hypothesis [342]. Another work using CAD models is presented by Pengfei Han [96], which uses CAD models for 3D object recognition and pose estimation in monocular images. The CAD model is rendered in different azimuths over an assumed inclination, provided by mobile device inertial sensors, to hypothesize possible azimuths of the object and compare it with the input image through a contour match.
- Octree-based models: Octree is a tree data structure where each node has precisely eight-node children [150]. Therefore, a model based on this data structure encodes representative information taking advantage on its structure features. Yong-Hui Xu [325] used the octree model to build a segmentation model, which eliminates inherited noise by label transfer procedure and retrieves the observed object lost parts. The proposed method aims to recognize and segment 3D objects in an RGB-D image captured by Kinect with information of labeled images in a database.

3.5 Subspace

A subspace is a vector space that is contained within another vector space. A vector space is a collection of objects called vectors, which may be added together and multiplied by numbers, satisfying certain requirements, called axioms [95]. For 3D object recognition, the representation using subspace consists on its construction based on 3D object feature vectors for further use on subspaces similarity computation and matching. Example of analyzed works employing subspace representation are presented by Fukui, Jianing Wu, Hotta, Akihiro, Igarashi, Kise and Kobayashi. Figure 17 shows some examples of works that employ subspace for object representation.

Fukui employed subspace representation in three of his analyzed works. In the first two, he presented Kernel Orthogonal Mutual Subspace Method (KOMSM) [75] and Kernel Constrained Mutual Subspace Method (KCMSM) [74] which are derived directly from Kernel Mutual Subspace Method (KMSM), which is the Mutual Subspace Method (MSM) nonlinear extension. The MSM classifies a pattern set based on canonical angle between linear subspace classes. The proposed variations deal with the pattern distribution nonlinearity and orthogonalization, aiming to classify patterns based on the similarity measured between subspace classes. In his third work, Fukui [73] presents a framework to extract local shape differences between two distinct objects for posterior use on shape classification. The author geometrically defines the concept of difference subspace (DS), which represents the difference components between two subspaces. Then, the DS is generalized for

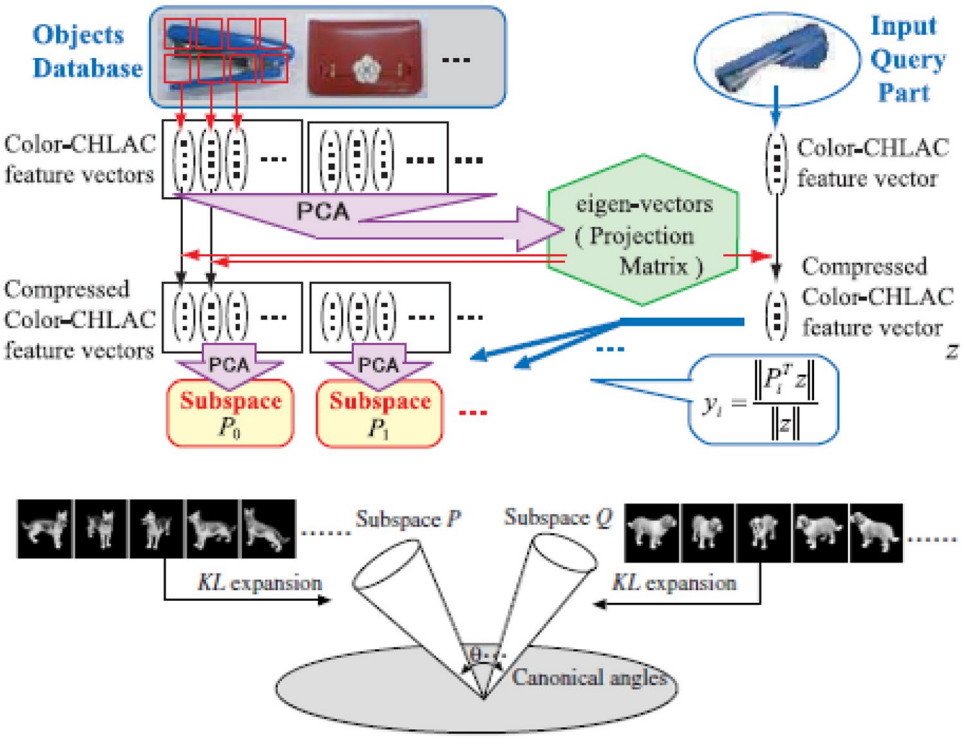

Fig. 17 Representation of analyzed works that employ subspace-based representation. Composition of images extracted from [75, 118]

multiple class subspaces, constructing the generalized difference subspace (GDS). The GDS use on subspace methods, such as subspace method (SM) and MSM, is shown in terms of recognition capacity.

Jianing Wu proposes a 3D object classification method, where the main idea is the feature vector distribution approximation with multiple local subspaces [312]. However, it is difficult to optimize the local subspace number and their dimension. Therefore, multiple local subspace sets are generated and combined, through ensemble learning algorithm, for posterior classification of input objects based on class subspace similarity measurement.

Hotta presents another object classification method based on subspaces [107]. The proposed method starts with Gabor feature extraction followed by the Gabor feature local parts extraction and uses for subspace construction, via Kernel principal component analysis. The constructed subspaces, for each object class, are used for classification of an input object, where the same procedure of feature extraction and subspace construction is performed for the input object. The similarity computation, between input and training object subspaces, is measured using the Class-Featuring Information Compression (CLAFIC).

Akihiro proposes a MSM theoretical extension named Compound Mutual Subspace Method (CPMSM), which can be applied for 3D object classification [5]. The CPMSM can be divided into two steps: in the learning step, a Karhunen–Loève (KL) expansion is applied on the training image set to obtain reference subspaces, and the difference subspace is computed from the obtained reference subspaces. In the test step, the KL expansion is applied on input image set to obtain the input subspace and measure the similarity between input and class subspaces.

Igarashi presents a method for measure the similarity between shapes [111], for 3D object recognition, using 3D shape subspaces constructed by a factorization method, such as Kanade–Lucas–Tomasi (KLT). Similar to the subspace methods previously mentioned, the proposed method uses canonical angle to measure the similarity between constructed shape subspaces.

A method for recognizing objects using a linear subspace method in a 3D feature space, based on Color Cubic Higher-order Local Auto-Correlation (Color-CHLAC) [117], is proposed by Kanezaki [118]. The Color-CHLAC features are calculated using 3D voxel data color and format information. The proposed recognition system based on the linear subspace method can be described by the following steps: As preprocessing step, the Color-CHLAC feature vectors, for the subdivided parts of each model in the database, are calculated. These feature vectors are used to calculate the subspace bases defined by each object. In the recognition step, a feature vector is extracted from a scene part for query. Then, this queried part is matched against the objects in the database, through feature vector projection, into each database object subspace, and similarity calculation, which is employed for ranking the database objects and electing the object candidate.

Three subspace-based recognition methods, on which the main feature is the use of a high subspace number generated from an equally high number of local features, are presented by Kise [138]. The local features are extracted via SIFT chain features, which are employed for subspace construction through PCA. Aiming to match the local features with a high number of constructed subspaces, a simple approximation using the nearest neighbor criteria was performed. The three proposed methods based on this approximation are: one method performing the match between one simple subspace with the local feature through feature vector projection into the subspace; a method with two-step matching, which on the first step the subspaces are used to select subspace candidates and on the second step subspace candidates are examined in a large dimension to select the best candidate; and one method with mutual subspaces, where query and stored object are represented by subspaces and the similarity is measured based on the subspaces canonical angles.

A MSM generalization, called generalized mutual subspace method (gMSM), is proposed by Kobayashi [142]. The gMSM has focused on minimizing the subspace dimensionality problem, applying it for object image set classification. The proposed method inserts a smooth weighting into the base vectors composing the subspace, without definitely selecting a number of principal bases, i.e., weights are added into calculation process, aiming to soft the binary weight (1 or 0) at selection of space base vectors. Therefore, the gMSM is a generalization that classifies a vector set (subspaces) through a similarity calculation between them based on their canonical angles, which are derived from smooth weighted combination of subspace bases.

3.6 Tensors

Tensors are mathematical objects which can be used to describe physical properties. They are a generalization of scalar and vectors. In fact, a scalar is a zero-rank tensor, and a vector is a first-rank tensor [37]. Its application for 3D object recognition lies on the description and matching of models and scene objects. For example, Mian [197] employs tensors associated with 3D object model, built from multiple object range images, to perform the matching between scene and library model objects, through a vote scheme. The methods proposed by Ben-Yaacov, Smeets, Yaniv Gur and Orts-Escolano also employ tensors as object representation. Figure 18 shows some examples of works that employ subspace for object representation.

Ben-Yaacov proposed a method using tensors of Implicit polynomials (IPs) [27] as representation, in order to derive a 3D rotation-invariant set, aiming the IP-based 3D object recognition. Then, this set is feed to a classifier for further object recognition. Smeets [273] presented an approach for object recognition of inelastic deformation invariant objects, employing diffusion distance tensors (DDT) as 3D object representation and performing the object recognition through the direct comparison between modal representations. Gur [94] proposed a method that uses spherical harmonics (SH) and contravariant tensors generation for the construction of rotation-invariant feature vectors applied on 3D object recognition. His method performs tensors generation and contraction, with SH coefficients, to generate the set of invariants used on the feature vector construction. Lastly, Orts-Escolano proposed a hardware implementation [220], with GPGPU, of the tensors representation presented by Mian [198], explaining in detail the construction process and the use for object recognition.

3.7 Other forms of representation

This section includes other forms of representations that have only a few works that make use of it. For example, the correlation filters set representation, presented by Loo [172], which is employed for automatic target recognition in real time, through a 3D object classification, based on fringe-adjusted joint transform correlator (FJTC). Figure 19 shows some examples of works that employ other forms of object representation.

We also analyzed the following representations:

- Content-adaptive pyramid: a representation based on content adjustable pyramid, presented in [146], used for classification of 3D images;
- Level curves: the representation using level curves, presented by Mahiddine [185], is employed for 3D partial objects retrieval task based on level curves matching. In the proposed approach, initially a set of level curves is generated from a 3D object stored in the database. Then, the level curves for each queried partial object are extracted and compared against the level curves representing the 3D object stored, searching for an object match.
- Depth aspect image: the method presented by Kitaaki [139], for 3D object recognition, is basically composed of two steps: pose estimation and positioning. For the coarse step, a depth aspect image (DAI) representation and matching, between model and scene, is employed. For the fine step, a hierarchical modified iterative closest point (HM-ICP) [217] is used.
- Volumetric descriptor: two analyzed works used a volumetric descriptor as object representation. The first one, presented by Gafar [78], represents the object through its partitioning in spherical shells, followed by the feature extraction, based on this area descriptor, and matching between reference and queried objects. The second work shows a 3D object volumetric description, named volumetric accelerator (VOLA), proposed by Xu [324]. The

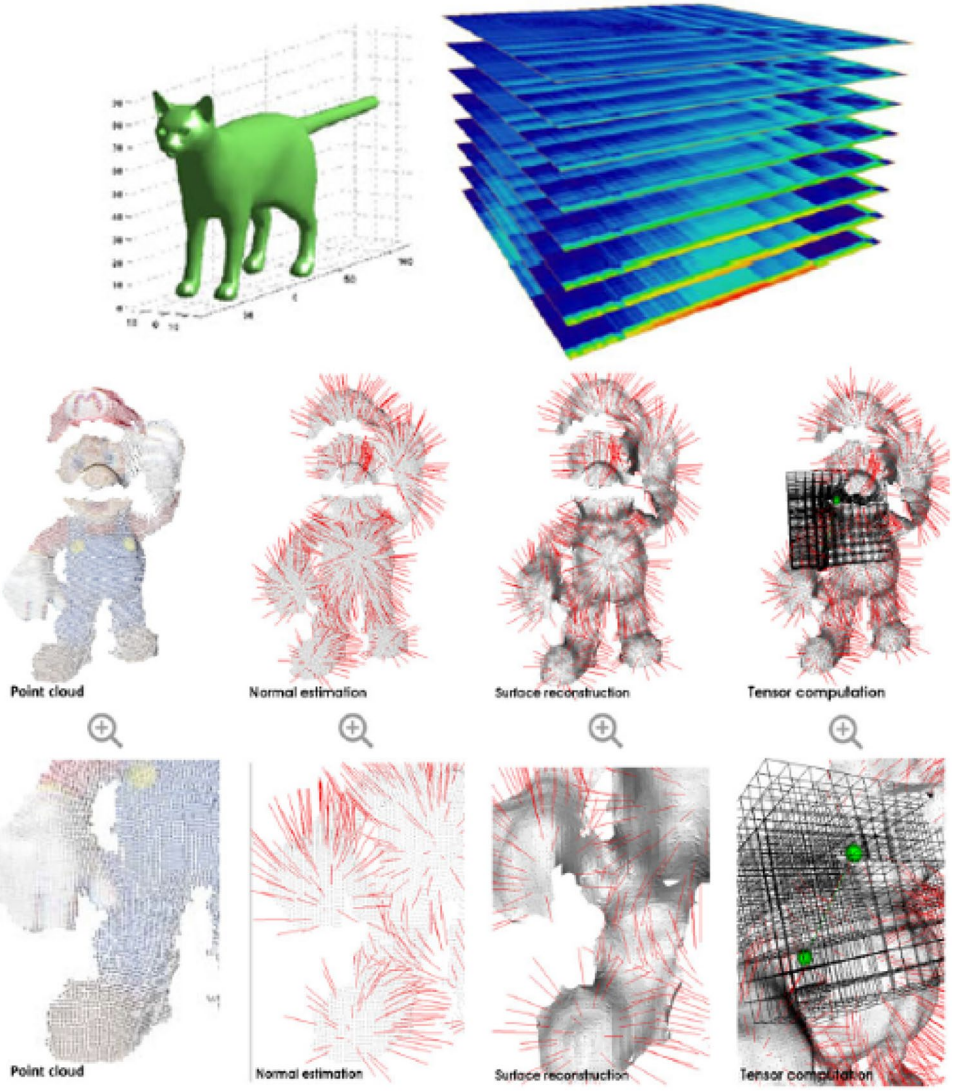

Fig. 18 Representation of analyzed works that employ tensor-based representation. Composition of images extracted from [220, 273]

VOLA representation deals with volumetric irregular grids, known as voxels, and stores only a single information bit to represent each voxel, employing a octree data structure to store the voxels. This simplified representation is used to represent the 3D objects, reducing the computational complexity from the convolutional neural networks (CNNs), when applied to 3D object recognition.

- Eigenimages: the representation with eigenimages, presented by Zhang [335], is used on the object pose estimation. The eigenimages are deducted from eigenspace constructed by a training set. Then, based on a given input sample, its eigenimages projections are computed and the closest training images are deducted with correlation between training and test eigenimages.
- Holography: Nelleri [209] and Kumar [147] present two methods using the digital holography representation for 3D object recognition. In Nelleri, one hologram of each 3D object to be classified is electronically stored using digital holography setup. Then, the electronic holograms are processed aiming to retrieve complex values from the 2D image, which corresponds to the 3D object points in the object plane. Lastly, applying the Mexican hat digital wavelet matched filtering, on the reconstructed image, higher correlation peaks are obtained, using them to discriminate and recognize objects regardless of their location in the scene. In Kumar, a digital holography from one perspective of two different 3D objects are simulated, using Fresnel digital holography, and the corresponding 3D images are numerically reconstructed using Fresnel–Kirchhoff integral. Then, the reconstructed 3D images are compared against the target 3D object by means of two proposed correlation strategies, joint fractional Fourier transform (JFRT) and nonlinear JFRTC (NJFRTC).
- Images photon-counting: the image photon-counting representation, presented by Do [58], is employed for

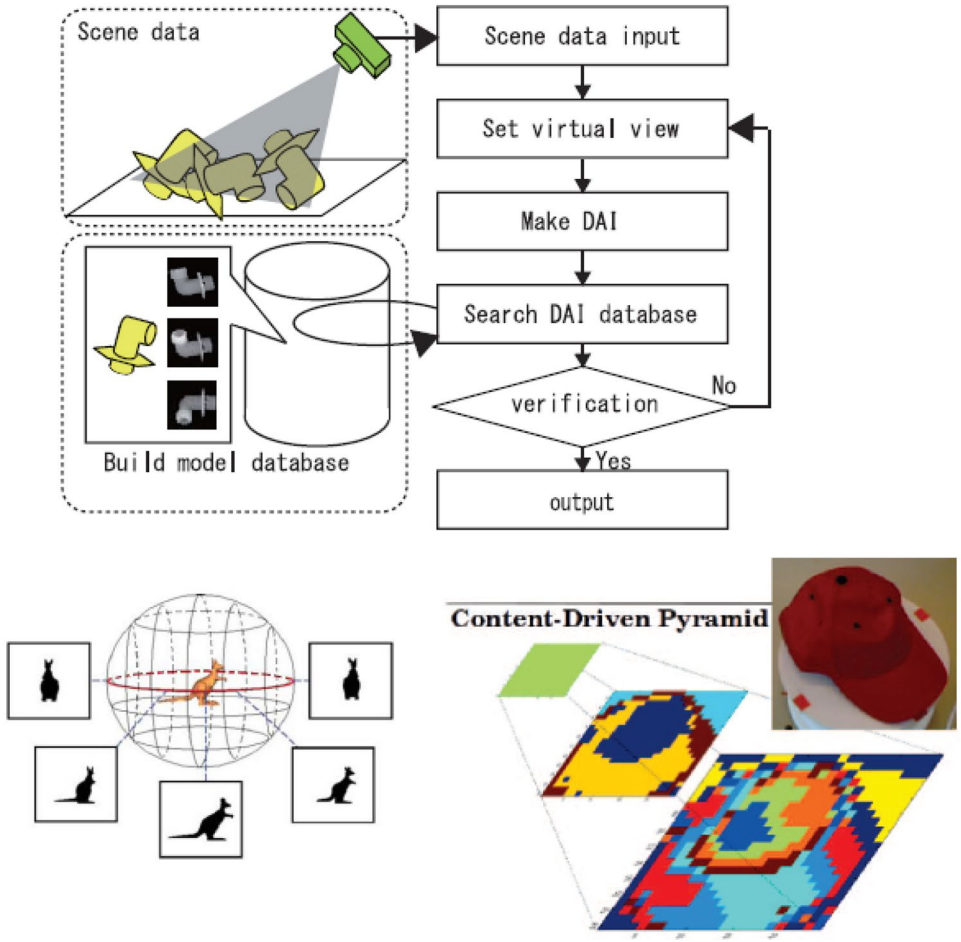

Fig. 19 Representation of analyzed works that employ other forms of representation (depth aspect image, holography and content-adaptive pyramid, respectively). Composition of images extracted from [139, 146, 335]

object recognition on 3D integral image. This representation from the acquired integral image is generated using a Poisson distribution. Then, a kurtosis-maximization-based algorithm is employed to extract independent features from the photon-counting training set. Lastly, a photon-counting image of an unknown input scene is computed and classified with k-nearest neighbor (KNN) and cosine angle metrics.

- Multiplex complex amplitude: the recognition of 3D objects using complex amplitude description, proposed by Yoshikawa [329], encodes directly the height object information in complex amplitude phase based on Fourier transform profilometry (FTP). The object recognition is performed by 2D correlation using the complex amplitude, denominating this 3D recognition method as Fourier transform profilometric correlator (FPC) and the complex amplitude as height transformed complex amplitude (HCA).
- Plane parts: the plane parts representation, introduced in [316], aims to describe the object as simple plane parts for posterior grouping in 3D to represent object portions. The work proposed by Xiang [317] uses each of these groups as 3D aspectlet, which are automatically generated. These 3D aspectlets are then used to provide evidences for the localization and object pose estimation.
- Pseudo-random binary sequences: the pseudo-random binary sequences (PRBS) [280] is a string sequence representing the Point Cloud Non-Uniform Rational Basis Spline (NURBS) model [36], used by Ravari for non-supervised object classification in range images [241] and for comparison between 3D objects [242], according to a relevant Kolmogorov complexity.
- Sliced curvature scale space: Okal [216] presents a representation via sliced curvature scale space (SCSS), which explores the scale-space theory via the space curvature, extending it to represent 3D objects. The SCSS is an extension of the curvature scale space (CSS), which is developed by repeatedly convolving a signal with a Gaussian kernel. To extend the CSS, a mechanism of slicing, through which the 3D object is seen as a set of infinitely close slices of thin plates packaged together, was adopted. Using this 3D object representation, the author trained a SVM aiming to classify objects into multiple class.

– Templates: the works proposed by Lee [153] and Zang [332] used templates to perform retrieval and recognition of 3D objects, respectively. In [153], the proposed system is composed of two steps: The first step is an offline stage, where reference templates from the 3D target object are obtained and its eigenvectors are computed from a set of training sub-images. The second step is an online stage, where occlusion removal, recursive compensation of the occlusion removal and object recognition, through cross-correlation between the reconstructed image, without occlusion and the reference templates, are performed. In [332], a template-based matching method is employed for the recognition and tracking of 3D objects. The matching method is performed through the comparison between queried image and template images, rendered using OpenGL with different viewpoints. For the object tracking, a edge-based method proposed in [60] is employed.

In addition to the representations presented, the work presented by Muja [202] demonstrates an infrastructure implementation, named REcognition INfrastructure (ReIn), capable of combining several 2D/3D object recognition and pose estimation techniques, in parallel, as dynamically loadable plugins.

3.8 Other works

Other works that were analyzed, that fit the established criteria, but were not directly related to the representation of 3D objects, are presented in this section. The first work is the method proposed by Marques [187], where 3D–2D correspondence solution is presented. The solution discussed by Marques is described over general principles, defined by the author, on which a unique and optimal solution, for the correspondence 2D–3D, should be obtained. Another work, which could not be categorized on the previously presented categories, is shown by Wohlkinger [308]. In his study, the methodology, the steps and all the features of the constructed database, named 3DNet, are presented.

Different from the two above-mentioned works, are the studies shown by Salti, Elizabeth González, Megherbi, Mateo, Sharman and Yulan Guo, which fit on a category of reviews and analysis of methods. Salti [260], for example, presents a study evaluating the performance of different keypoint detectors according to the task of 3D object recognition. Elizabeth González [83] presents a qualitative and quantitative analysis over the 2D shape recognition methods performance, when employed on the recognition of 3D objects. Megherbi [193] presents a comparison of classification approaches for automatic threat objects identification in computed tomography images. Mateo [190] describes a study analyzing 3D descriptors based on normal surface for 3D object recognition. Sharman shows a systematic review of segmentation and 3D object recognition algorithms, employing a total of 20 works for this analysis. Lastly, Yulan Guo provides a survey with 3D object recognition methods based on surface features [88] and an evaluation study over 3D feature descriptors [89].

3.9 Books

Our search also yielded three books in the results, which provided information from the object's recognition introductory part (*An Introduction to Object Recognition* [290]), until the study of more advanced methods and techniques for recognition, representation and tracking of 3D objects (*Representations and Techniques for 3D Object Recognition and Scene Interpretation* [105] and *Visual Perception and Robotic Manipulation: 3D Object Recognition, Tracking and Hand-Eye Coordination* [284]).

4 Discussion

Based on the analyzed works, we can visualize two general pipelines for object recognition/classification. Both first pipeline, illustrated by Fig. 20, and second pipeline, illustrated by Fig. 21, have similar and different steps. As similar steps, we have data acquisition, preprocessing (optional) and data representation. As different steps, we can refer to the way that object recognition/classification is performed: In the first pipeline, the data representation chosen is compared through a matching or similarity calculation methods, aiming to classify/recognize an object. In other words, the input data are compared with object database representations, previously calculated in an offline stage, assigning to the input data, a classification or object representation which best corresponds to it. In the second pipeline, the chosen representation is employed on training a classifier, performed offline

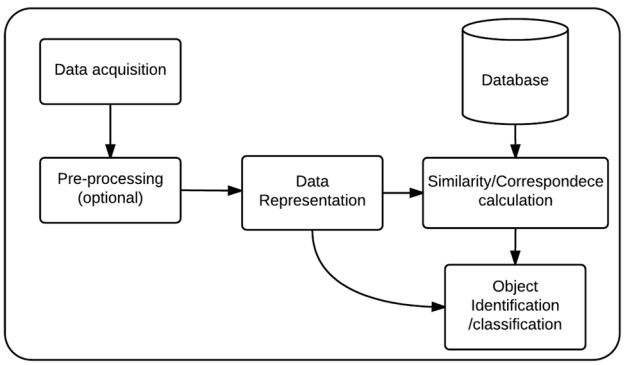

Fig. 20 First pipeline form

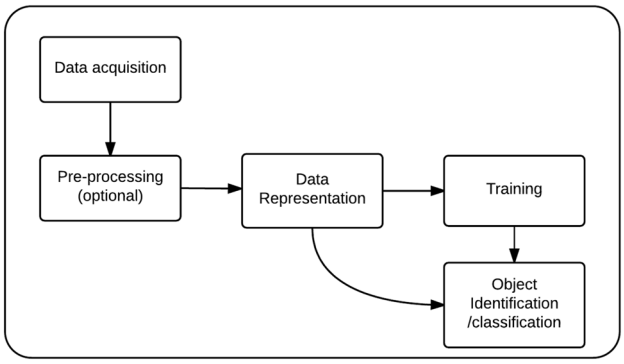

Fig. 21 Second pipeline form

or in some cases online [132], for posterior use on object classification/recognition. Each step from both pipelines will be further discussed in the next sections.

4.1 Data acquisition

The data acquisition techniques found on the analyzed works start with simple camera [330], stereo pair [154] and Kinect [325] acquisitions and go through acquisition with a tactile sensor [110]. However, not all the works performed their own data acquisitions, using on these cases public databases, e.g., in [129]. The next section lists some of these public databases used. Other works performed simulations based on the proposed method, for example the work proposed by Anand [11], or employed only synthetic data for the experiments, for example the work proposed by Kordelas [144].

4.2 Data set and databases available

Several public databases were found in our research. Some of them are composed with 3D object models, and others have images from RGB-D cameras or are composed by several 2D object images acquired from different camera positions. Some examples of public databases are: Grand Challenge v2.0 database (FRGC v2.0) [230], Stanford 3D Scanning Repository [50], Bologna Dataset VI [289], Amsterdam Library of Object Images [80], ETH-80 [160], Princeton Shape Benchmark [270], COIL [210], MNIST [152] and, additionally, in [88] Yulan Guo shows a table with 16 popular range images databases. More database examples can be found in our technical report [40].

4.3 Preprocessing: optional

This step, optionally employed in some works, involves the possible application of the following methods:

- Filtering: responsible for noise removal [220] or to reach a better sampling [218];
- Selection of region of interest (ROI): the ROI selection helps to focus on the object area only [202], reducing computational cost and improving recognition results.
- Segmentation: the segmentation process is employed in several works to simplify the image into segments [102], generating some indications of possible object localization and reducing the search cost.
- Normalization: employed in some works for reach some data invariance [57].

4.4 Representations

The representations presented in this work are divided in seven categories. Inside each specific category, we have several variations, e.g., in the feature-based representation there is a variety of feature descriptors (SHOT [288], SI[116], VFH [254], SIFT [175] and so on). Other examples of this variation inside a category are present on model category, on which there are geometric [14], statistical [87] and CAD [342] models and on graph category, where there are Reeb graph [31] and class-specific hyper graph [313] representations. The seven categories employed on the categorization process are: feature-based, views, graphs, models, subspace, tensors and the last category, called other forms of representation, which groups representations with minor frequency in the analyzed works. Figure 22 shows each category frequency, and Fig. 23 shows, for the feature-based representation, the feature-type frequencies.

4.5 Similarity and correspondence calculation

For the similarity and correspondence calculation, the methods used in the analyzed works employ mainly a matching computation between scene and object descriptors, defined by the authors [8, 223], or use a distance metric to measure the similarity between descriptors [311] or between objects [1]. These correspondence methods are generally associated with vote schemes [281] and/or alignment methods [8], aiming to obtain and verify hypotheses, culminating on the object recognition/classification.

4.6 Classifiers

As well as the similarity and correspondence calculation, several classifiers, employed on the object recognition/classification, were used on the analyzed works. Among them, we can highlight the neural network classifiers, in its great majority using the support vector machine classifier [21, 54] or its variations [15]. Some works employed classifiers based on deep learning models [274, 314, 330], others used the k-nearest neighbor associated with Euclidean distance

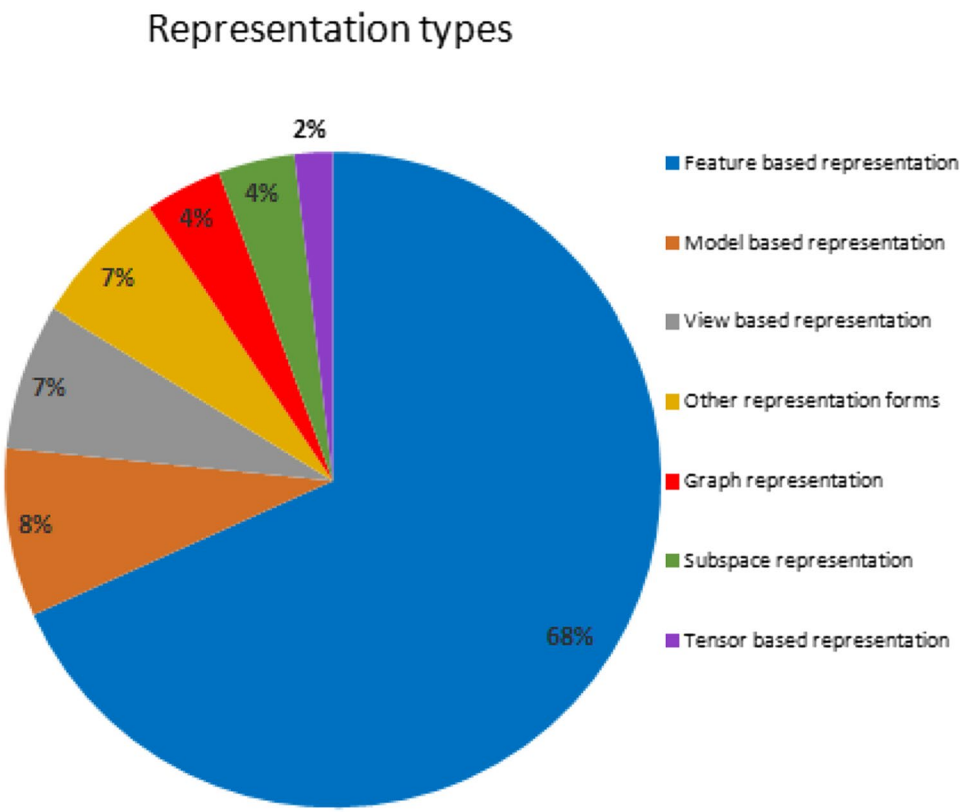

Fig. 22 Analyzed works representation types chart

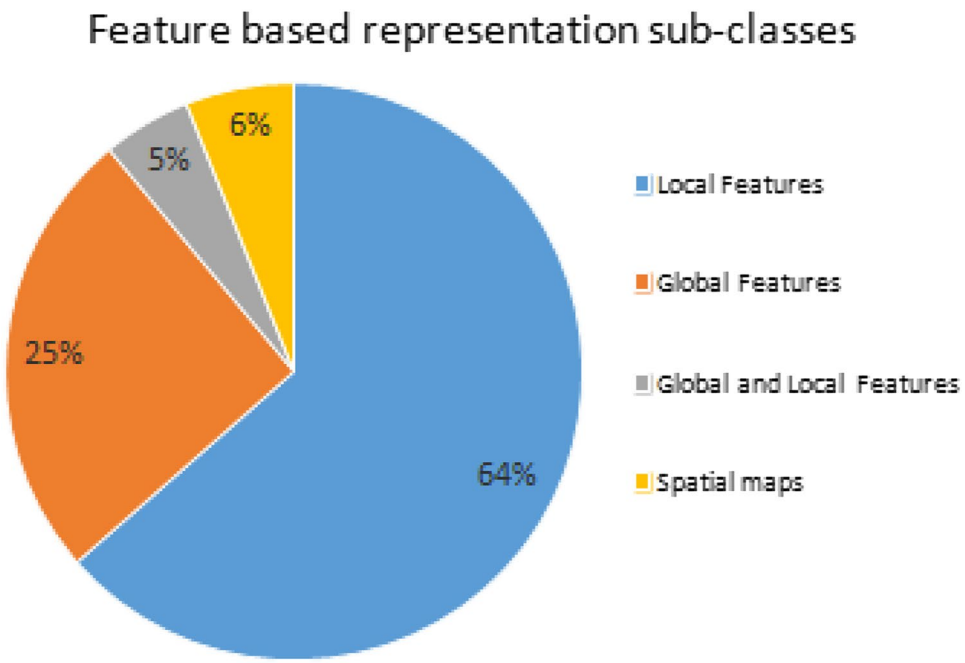

Fig. 23 Feature-based representation chart, grouped by the feature type

[122, 158, 309]. There are also a few works that utilized fuzzy [191, 269], fuzzy associated with Bayesian networks, fuzzy associated with neural networks [221] and probabilistic models [237].

4.7 Validation and results report

The great majority of the analyzed works used as result evaluation metrics the successful recognition/classification rates. Additionally, some works demonstrated their results

according to the receiver operating characteristics (ROC) and cumulative match characteristic (CMC) [273] curves. Others evaluated their results based on classification/recognition error and on the equal error rate (EER) [167]. Independently of the method used to demonstrate and evaluate the results, a large part of the analyzed works compared their results against other methods, considered state of the art on that moment, to demonstrate their work strengths and abilities.

Another important concept to evaluate the representation type and the methods used for classification is the noise sensitiveness. The great majority of analyzed works used noise reduction techniques to remove or reduce its effect in their methods and representations, avoiding to perform a specific analysis about noise sensitiveness [20, 42, 220]. Some works ignore the noise concept in their study, just by using synthetic data and simulations [11, 172]. Others used noise as evaluation metric, where a method/representation was considered robust and having good results when the final result had high classification rates in the presence of different noise levels [104]. Also, a few works do not even have the word noise in their study [129].

5 Conclusion

The proposed literature review analysis allows us to notice a growing study over 3D object recognition/classification and representation methods. Those studies, boosted by popularization of 3D data acquisition equipment, showed applications on several areas such as human health, security and industrial. The analyzed works also demonstrated studies comparing several classification/recognition and representation methods on 3D object recognition/classification tasks using a variety of public databases.

This review sought to condense the studies published between 2006 and 2016 available on three scientific databases, aiming to understand the 3D object recognition/classification area and to analyze the representations, experiments, evaluation and validation methods employed. Additionally, this literature review with methods analysis, general overview, performance evaluation analysis and future prospects is useful for researches working on the 3D object recognition/classification area.

5.1 Future prospects

From this literature review over the 3D object recognition/classification area, we can make some predictions of the direction that this area is taking:

- Due to the large number of methods employing feature-based representation, noticeable in Fig. 22, we can predict a trend of continuity with this representation type, which would also include the addition of more features into the future proposed feature-based descriptors;
- Some works present feature descriptors combinations, which provide indicatives that the descriptors combination would be a possible solution to increase the objects' representation discriminative power;
- Several publicly databases were presented in the last few years for the 3D object recognition area. Thus, due to the increasing popularization of 3D data acquisition equipment, a strong tendency that more 3D object databases will emerge in the next years;
- Based on those methods used in the robotic area and the proposal of a few GPU methods [220, 223], we can predict an increase in the research methods and techniques improving speed computation and reducing data dimensionality. These improvements will emerge aiming the application of 3D recognition methods on real-time applications with several input information;
- We can also infer an increase in the use of combined methods and learning models such as deep learning, due to the current increase in the use of those types of methods in several areas;
- Lastly, there is a high probability of proposal of variations with well-known descriptors, using different similarity metrics, combinations and measures.

Acknowledgements This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

1. Abdelrahman M, Farag AA, El-Melegy M (2013) Heat front propagation contours for 3d face recognition. In: 2013 IEEE sixth international conference on biometrics: theory, applications and systems (BTAS), pp 1–6
2. Akagunduz E, Ulusoy I (2010) 3D object recognition from range images using transform invariant object representation. Electron Lett 46(22):1499–1500
3. Akbar H, Suryana N, Sahib S (2011) Training neural networks using clonal selection algorithm and particle swarm optimization: a comparisons for 3D object recognition. In: 2011 11th International conference on hybrid intelligent systems (HIS), pp 692–697
4. Akgül CB, Sankur B, Yemez Y, Schmitt F (2009) 3D model retrieval using probability density-based shape descriptors. IEEE Trans Pattern Anal Mach Intell 31(6):1117–1133
5. Akihiro N, Fukui K (2011) Compound mutual subspace method for 3d object recognition: a theoretical extension of mutual subspace method. In: Proceedings of the 2010 international conference on computer vision—volume Part II, ACCV'10. Springer, Berlin, pp 374–383
6. Albarelli A, Rodolà E, Bergamasco F, Torsello A (2011) A non-cooperative game for 3d object recognition in cluttered scenes.

In: 2011 International conference on 3D imaging, modeling, processing, visualization and transmission, pp 252–259
7. Albarelli A, Rodolà E, Torsello A (2010) A game-theoretic approach to fine surface registration without initial motion estimation. In: 2010 IEEE computer society conference on computer vision and pattern recognition, pp 430–437
8. Aldoma A, Tombari F, Di Stefano L, Vincze M (2012) A global hypotheses verification method for 3d object recognition. In: Proceedings of the 12th European conference on computer vision—volume part III, ECCV'12. Springer, Berlin, pp 511–524
9. Aldoma A, Tombari F, Prankl J, Richtsfeld A, Stefano LD, Vincze M (2013) Multimodal cue integration through hypotheses verification for rgb-d object recognition and 6dof pose estimation. In: 2013 IEEE international conference on robotics and automation, pp 2104–2111
10. Aldoma A, Tombari F, Stefano LD, Vincze M (2016) A global hypothesis verification framework for 3d object recognition in clutter. IEEE Trans Pattern Anal Mach Intell 38(7):1383–1396
11. Anand S, Kirmani A, Shrivastava S, Chaudhury S, Bhaumik B (2006) A microscopic framework for distributed object-recognition pose-estimation. In: 2006 9th International conference on control, automation, robotics and vision, pp 1–6
12. Aouada D, Feng S, Krim H (2007) Statistical analysis of the global geodesic function for 3d object classification. In: 2007 IEEE international conference on acoustics, speech and signal processing—ICASSP '07, vol 1, pp I-645–I-648
13. Aouada D, Krim H (2009) Meaningful 3d shape partitioning using morse functions. In: 2009 16th IEEE international conference on image processing (ICIP), pp 417–420
14. Aouat S, Laiche N, Souami F, Larabi S (2008) 3D object indexing and recognition. Appl Math Comput 196(1):318–332
15. Arana-Daniel N, Bayro-Corrochano E (2006) Mimo svms for 3d object classification. In: The 2006 IEEE international joint conference on neural network proceedings, pp 1628–1635
16. Arnaud E, Odone F, Verri A (2006) Trains of keypoints for 3d object recognition. In: Proceedings of the 18th international conference on pattern recognition—volume 02, ICPR '06. IEEE Computer Society, Washington, DC, pp 1014–1017
17. Assfalg J, Bertini M, Bimbo AD, Pala P (2007) Content-based retrieval of 3-d objects using spin image signatures. IEEE Trans Multimed 9(3):589–599
18. Assfalg J, Borgwardt KM, Kriegel HP (2006) 3dstring: A feature string kernel for 3d object classification on voxelized data. In: Proceedings of the 15th ACM international conference on information and knowledge management, CIKM '06. ACM, New York, NY, pp 198–207
19. Atmosukarto I, Shapiro LG (2008) A learning approach to 3d object representation for classification. In: Proceedings of the 2008 joint IAPR international workshop on structural, syntactic, and statistical pattern recognition, SSPR & SPR '08. Springer, Berlin, pp 267–276
20. Atmosukarto I, Wilamowska K, Heike C, Shapiro LG (2010) 3D object classification using salient point patterns with application to craniofacial research. Pattern Recognit 43(4):1502–1517
21. Ayoub J, Granado B, Romain O, Mhanna Y (2010) 3-d object recognition based on svm and stereo-vision: application in endoscopic imaging. In: 2010 International conference of soft computing and pattern recognition, pp 198–201
22. Ballard D (1981) Generalizing the hough transform to detect arbitrary shapes. Pattern Recognit 13(2):111–122
23. Bariya P, Nishino K (2010) Scale-hierarchical 3d object recognition in cluttered scenes. In: 2010 IEEE computer society conference on computer vision and pattern recognition, pp 1657–1664
24. Bariya P, Novatnack J, Schwartz G, Nishino K (2012) 3D geometric scale variability in range images: features and descriptors. Int J Comput Vis 99(2):232–255
25. Bedkowski J, Majek K, Majek P, Musialik P, Pelka M, Nüchter A (2016) Intelligent mobile system for improving spatial design support and security inside buildings. Mob Netw Appl 21(2):313–326
26. Beksi WJ, Papanikolopoulos N (2015) Object classification using dictionary learning and rgb-d covariance descriptors. In: 2015 IEEE international conference on robotics and automation (ICRA), pp 1880–1885
27. Ben-Yaacov H, Malah D, Barzohar M (2010) Recognition of 3d objects based on implicit polynomials. IEEE Trans Pattern Anal Mach Intell 32(5):954–960
28. Bencharef O, Fakir M, Minaoui B, Hajraoui A, Oujaoura M (2012) Color objects recognition system based on artificial neural network with zernike, hu amp; geodesic descriptors. In: 2012 6th International conference on sciences of electronics, technologies of information and telecommunications (SETIT), pp 338–343
29. Bennamoun M, Sohel FA, Guo Y, Lu M, Wan J (2013) 3d free form object recognition using rotational projection statistics. In: Proceedings of the 2013 IEEE workshop on applications of computer vision (WACV), WACV '13, pp 1–8. IEEE Computer Society, Washington, DC, USA
30. Bonev B, Escolano F, Giorgi D, Biasotti S (2010) High-dimensional spectral feature selection for 3d object recognition based on reeb graphs. In: Proceedings of the 2010 Joint IAPR international conference on structural, syntactic, and statistical pattern recognition. Springer, Berlin, pp 119–128
31. Bonev B, Escolano F, Giorgi D, Biasotti S (2010) Information-theoretic feature selection from unattributed graphs. In: Proceedings of the 2010 20th international conference on pattern recognition, ICPR '10, pp 930–933. IEEE Computer Society, Washington, DC, USA
32. Bongale P, Ranjan A, Anand S (2012) Implementation of 3d object recognition and tracking. In: 2012 international conference on recent advances in computing and software systems, pp 77–79
33. Bouguila N (2012) Infinite liouville mixture models with application to text and texture categorization. Pattern Recognit Lett 33(2):103–110
34. Bovee C, Thill J (2015) Business communication essentials: a skills-based approach. Pearson Education, London
35. Brandão S, Veloso M, Costeira JP (2014) Multiple hypothesis for object class disambiguation from multiple observations. In: 2014 2nd International conference on 3D vision, vol 1, pp 91–98
36. Brujic D, Ainsworth I, Ristic M (2011) Fast and accurate nurbs fitting for reverse engineering. Int J Adv Manuf Technol 54(5):691–700
37. University of Cambridge, What is a tensor? URL https://www.doitpoms.ac.uk/tlplib/tensors/what_is_tensor.php. Dissemination of IT for the Promotion of Materials Science. Accessed 29 May 2017
38. Carreira J, Sminchisescu C (2012) Cpmc: automatic object segmentation using constrained parametric min-cuts. IEEE Trans Pattern Anal Mach Intell 34(7):1312–1328
39. Carrer L, Yarovoy AG (2014) Concealed weapon detection using uwb 3-d radar imaging and automatic target recognition. In: The 8th European conference on antennas and propagation (EuCAP 2014), pp 2786–2790
40. Carvalho L, von Wangenheim A (2017) Literature review for 3D object classification/recognition. Tech. rep., Technical Report INCoD/LAPIX.01.2017.E.v01
41. Chen F, Ji R, Cao L (2016) Multimodal learning for view-based 3d object classification. Neurocomputing 195:23–29 **(learning for medical imaging)**
42. Chen H, Bhanu B (2007) 3D free-form object recognition in range images using local surface patches. Pattern Recognit Lett 28(10):1252–1262

43. Chen H, Bhanu B (2009) Efficient recognition of highly similar 3d objects in range images. IEEE Trans Pattern Anal Mach Intell 31(1):172–179
44. Chen LC, Nguyen XL, Lin ST (2012) Automatic object detection employing viewing angle histogram for range images. In: 2012 IEEE/ASME international conference on advanced intelligent mechatronics (AIM), pp 196–201
45. Chen YC, Patel VM, Chellappa R, Phillips PJ (2015) Salient views and view-dependent dictionaries for object recognition. Pattern Recognit 48(10):3053–3066
46. Chen Z, Zhao R, Zhang Y (2006) Geometric hashing using 3d aspects and constrained structures. In: 2006 8th International conference on signal processing, vol. 2
47. Choi KS, Kim DH (2013) Angular-partitioned spin image descriptor for robust 3d facial landmark detection. Electron Lett 49(23):1454–1455
48. Cipolla R, Battiato S, Farinella GM (eds.) (2010) Computer vision: detection, recognition and reconstruction, studies in computational intelligence, vol. 285. Springer. URL http://dblp.uni-trier.de/db/series/sci/sci285.html
49. Collaboration C (2003) Cochrane Reviewers' Handbook. Version 4.2.1. National Health and Medical Research Council
50. Curless B, Levoy M (1996) A volumetric method for building complex models from range images. In: Proceedings of the 23rd annual conference on computer graphics and interactive techniques, SIGGRAPH '96. ACM, New York, NY, pp 303–312
51. Daley DJ, Vere-Jones D (2008) An introduction to the theory of point processes, probability and its applications (New York), vol II, 2nd edn. Springer, New York
52. Decker P, Thierfelder S, Paulus D, Grzegorzek M (2011) Dense statistic versus sparse feature-based approach for 3d object recognition. Pattern Recognit Image Anal 21(2):238–241
53. Deinzer F, Denzler J, Derichs C, Niemann H (2006) Aspects of optimal viewpoint selection and viewpoint fusion. In: Proceedings of the 7th Asian conference on computer vision—volume part II, ACCV'06, pp 902–912. Springer, Berlin
54. Delponte E, Arnaud E, Odone F, Verri A (2006) Analysis on a local approach to 3d object recognition. In: Proceedings of the 28th conference on pattern recognition, DAGM'06. Springer, Berlin, pp 253–262
55. Delponte E, Noceti N, Odone F, Verri A (2007) Appearance-based 3d object recognition with time-invariant features. In: 14th International conference on image analysis and processing (ICIAP 2007), pp 467–474
56. Dimov D, Zlateva N, Marinov A (2009) Cbir over multiple projections of 3d objects. In: Proceedings of the 2009 Joint COST 2101 and 2102 international conference on biometric ID management and multimodal communication. Springer, Berlin, pp 146–153
57. Ding H, Li X, Zhao H, Xiao W (2012) A new generalized affine moment invariants for shape retrieval and object recognition. In: 2012 8th IEEE international symposium on instrumentation and control technology (ISICT) proceedings, pp 137–142
58. Do CM, Javidi B (2009) Three-dimensional object recognition with multiview photon-counting sensing and imaging. IEEE Photonics J 1(1):9–20
59. Drost B, Ulrich M, Navab N, Ilic S (2010) Model globally, match locally: Efficient and robust 3d object recognition. In: 2010 IEEE computer society conference on computer vision and pattern recognition, pp 998–1005
60. Drummond T, Cipolla R (2002) Real-time visual tracking of complex structures. IEEE Trans Pattern Anal Mach Intell 24(7):932–946
61. Šeatović D, Kutterer H, Anken T (2010) Automatic weed detection and treatment in grasslands. In: Proceedings ELMAR-2010, pp 65–68
62. Efremova N, Asakura N, Inui T, Abdikeev N (2011) Inferotemporal network model for 3d object recognition. In: The 2011 IEEE/ICME international conference on complex medical engineering, pp 555–560
63. Efremova NA, Inui T (2014) An inferior temporal cortex model for object recognition and classification. Sci Tech Inf Process 41(6):362–369
64. Ejima T, Enokida S, Kouno T, Ideguchi H, Horiuchi T (2014) 3D object recognition based on the reference point ensemble. In: 2014 International conference on computer vision theory and applications (VISAPP), vol 3, pp 261–269
65. Ekekrantz J, Pronobis A, Folkesson J, Jensfelt P (2013) Adaptive iterative closest keypoint. In: 2013 European conference on mobile robots, pp 80–87
66. Elons AS, Abull-ela M, Tolba M (2013) A proposed { PCNN } features quality optimization technique for pose-invariant 3d arabic sign language recognition. Appl Soft Comput 13(4):1646–1660
67. Everingham M, Sivic J, Zisserman A (2006) Hello! my name is... buffy—automatic naming of characters in tv video. In: In BMVC
68. Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D (2010) Object detection with discriminatively trained part-based models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645
69. de Figueiredo RP, Moreno P, Bernardino A (2015) Efficient pose estimation of rotationally symmetric objects. Neurocomputing 150:126–135
70. Filipe S, Itti L, Alexandre LA (2015) Bik-bus: biologically motivated 3d keypoint based on bottom-up saliency. IEEE Trans Image Process 24(1):163–175
71. Flitton G, Mouton A, Breckon TP (2015) Object classification in 3d baggage security computed tomography imagery using visual codebooks. Pattern Recognit 48(8):2489–2499
72. Fritzke B (1994) A growing neural gas network learns topologies. In: Proceedings of the 7th international conference on neural information processing systems, NIPS'94, pp. 625–632. MIT Press, Cambridge, MA, USA
73. Fukui K, Maki A (2015) Difference subspace and its generalization for subspace-based methods. IEEE Trans Pattern Anal Mach Intell 37(11):2164–2177
74. Fukui K, Stenger B, Yamaguchi O (2006) A framework for 3d object recognition using the kernel constrained mutual subspace method. In: Proceedings of the 7th Asian conference on computer vision—volume part II, ACCV'06. Springer, Berlin, pp 315–324
75. Fukui K, Yamaguchi O (2007) The kernel orthogonal mutual subspace method and its application to 3d object recognition. In: Proceedings of the 8th Asian conference on computer vision—volume part II, ACCV'07. Springer, Berlin, pp 467–476
76. Fäulhammer T, Zillich M, Vincze M (2015) Multi-view hypotheses transfer for enhanced object recognition in clutter. In: 2015 14th IAPR international conference on machine vision applications (MVA), pp 10–13
77. G, I.I., Prakash S (2016) False mapped feature removal in spin images based 3d ear recognition. In: 2016 3rd international conference on signal processing and integrated networks (SPIN), pp 620–623
78. Gafar MF, Hemayed EE (2010) Surface area distribution descriptor for object matching. J Adv Res 1(3):233–241
79. Garstka J, Peters G (2015) Adaptive 3-d object classification with reinforcement learning. In: 2015 12th International conference on informatics in control, automation and robotics (ICINCO), vol 02, pp 381–385
80. Geusebroek JM, Burghouts GJ, Smeulders AW (2005) The amsterdam library of object images. Int J Comput Vis 61(1):103–112

81. Gibbins D, Swierkowski L (2009) A comparison of terrain classification using local feature measurements of 3-dimensional colour point-cloud data. In: 2009 24th International conference image and vision computing New Zealand, pp 293–298
82. Gomes RB, da Silva BMF, de Medeiros Rocha LK, Aroca RV, Velho LCPR, Gonçalves LMG (2013) Efficient 3d object recognition using foveated point clouds. Comput Gr 37(5):496–508
83. González E, Adán A, Feliú V (2012) 2D shape representation and similarity measurement for 3d recognition problems: an experimental analysis. Pattern Recognit Lett 33(2):199–217
84. González E, Adán A, Feliú V, Sánchez L (2008) Active object recognition based on fourier descriptors clustering. Pattern Recognit Lett 29(8):1060–1071
85. Grimson W, Lozano-Perez T (1985) Recognition and localization of overlapping parts from sparse data in two and three dimensions. In: Proceedings. 1985 IEEE international conference on robotics and automation, vol 2, pp 61–66. https://doi.org/10.1109/ROBOT.1985.1087320
86. Groover MP Jr, Zimmers EW Jr (1997) CAD/Cam: computer-aided design and manufacturing, 1st edn. Prentice Hall PTR, Upper Saddle River
87. Grzegorzek M, Izquierdo E (2007) Statistical 3d object classification and localization with context modeling. In: 2007 15th European signal processing conference, pp 1585–1589
88. Guo Y, Bennamoun M, Sohel F, Lu M, Wan J (2014) 3D object recognition in cluttered scenes with local surface features: a survey. IEEE Trans Pattern Anal Mach Intell 36(11):2270–2287
89. Guo Y, Bennamoun M, Sohel F, Lu M, Wan J, Kwok NM (2016) A comprehensive performance evaluation of 3d local feature descriptors. Int J Comput Vis 116(1):66–89
90. Guo Y, Sohel F, Bennamoun M, Wan J, Lu M (2014) An accurate and robust range image registration algorithm for 3d object modeling. IEEE Trans Multimed 16(5):1377–1390
91. Guo Y, Sohel F, Bennamoun M, Wan J, Lu M (2015) A novel local surface feature for 3d object recognition under clutter and occlusion. Inform Sci 293:196–213
92. Guo Y, Sohel FA, Bennamoun M, Wan J, Lu M (2013) Integrating shape and color cues for textured 3d object recognition. In: 2013 IEEE 8th conference on industrial electronics and applications (ICIEA), pp 1614–1619
93. Gupta S, Girshick R, Arbeláez P, Malik J (2014) Learning rich features from rgb-d images for object detection and segmentation. In: European conference on computer vision. Springer, pp 345–360
94. Gur Y, Johnson CR (2014) Generalized hardi invariants by method of tensor contraction. In: 2014 IEEE 11th international symposium on biomedical imaging (ISBI), pp 718–721
95. Halmos P (1948) Finite dimensional vector spaces. Annals of mathematics studies. Princeton University Press, Princeton
96. Han P, Zhao G (2015) Cad-based 3d objects recognition in monocular images for mobile augmented reality. Comput Gr 50:36–46
97. Hanai R, Yamazaki K, Yaguchi H, Okada K, Inaba M (2011) Electric appliance parts classification using a measure combining the whole shape and local shape distribution similarities. In: Proceedings of the 2011 international conference on 3D imaging, modeling, processing, visualization and transmission, 3DIMPVT '11. IEEE Computer Society, Washington, DC, pp 296–303
98. Harris C, Stephens M (1988) A combined corner and edge detector. In: Proceedings of the fourth Alvey vision conference, pp 147–151
99. Health N, (Australia) MRC, Staff N, (2000) How to review the evidence: systematic identification and review of the scientific literature. Handbook series on preparing clinical practice guidelines. National Health and Medical Research Council, Canberra
100. Health N, (Australia) MRC, Staff N (2000) How to Use the Evidence: Assessment and Application of Scientific Evidence Handbook series on preparing clinical practice guidelines. National Health and Medical Research Council, Canberra
101. Hejrati M, Ramanan D (2014) Analysis by synthesis: 3d object recognition by object reconstruction. In: 2014 IEEE conference on computer vision and pattern recognition, pp 2449–2456
102. Himmelsbach M, Luettel T, Wuensche HJ (2009) Real-time object classification in 3d point clouds using point feature histograms. In: 2009 IEEE/RSJ international conference on intelligent robots and systems, pp 994–1000
103. Hinton GE, Sejnowski TJ (1983) Optimal perceptual inference. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, New York, pp 448–453
104. Ho HT, Gibbins D (2008) Multi-scale feature extraction for 3d models using local surface curvature. In: Proceedings of the 2008 digital image computing: techniques and applications, DICTA '08. IEEE Computer Society, Washington, DC, pp 16–23
105. Hoiem D, Savarese S (2011) Representations and techniques for 3D object recognition and scene interpretation, 1st edn. Morgan & Claypool Publishers, San Rafael
106. Hong C, Yu J, You J, Chen X, Tao D (2015) Multi-view ensemble manifold regularization for 3d object recognition. Inform Sci 320:395–405
107. Hotta K (2009) Pose independent object classification from small number of training samples based on kernel principal component analysis of local parts. Image Vis Comput 27(9):1240–1251
108. Hsiao E, Collet A, Hebert M (2010) Making specific features less discriminative to improve point-based 3d object recognition. In: 2010 IEEE computer society conference on computer vision and pattern recognition, pp 2653–2660
109. Hu M, Wei Z, Shao M, Zhang G (2015) 3-D object recognition via aspect graph aware 3-d object representation. IEEE Signal Process Lett 22(12):2359–2363
110. Ibrayev R, Jia YB (2012) Recognition of curved surfaces from one-dimensional tactile data. IEEE Trans Autom Sci Eng 9(3):613–621
111. Igarashi Y, Fukui K (2011) 3d object recognition based on canonical angles between shape subspaces. In: Proceedings of the 10th Asian conference on computer vision—volume part IV, ACCV'10. Springer, Berlin, pp 580–591
112. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259
113. Jang Y, Woo W (2012) Local feature descriptors for 3d object recognition in ubiquitous virtual reality. In: Proceedings of the 2012 international symposium on ubiquitous virtual reality, ISUVR '12. IEEE Computer Society, Washington, DC, , pp 42–45
114. Jeong W, Lee S, Kim Y (2011) Statistical feature selection model for robust 3d object recognition. In: 2011 15th International conference on advanced robotics (ICAR), pp 402–408
115. Jing G, Mingquan Z, Chao L (2013) 3d object classification based on local keywords and hidden markov model. In: 2013 Fourth international conference on digital manufacturing automation, pp 1–4
116. Johnson AE, Hebert M (1999) Using spin images for efficient object recognition in cluttered 3d scenes. IEEE Trans Pattern Anal Mach Intell 21(5):433–449
117. Kanezaki A, Harada T, Kuniyoshi Y (2010) Partial matching of real textured 3d objects using color cubic higher-order local auto-correlation features. Vis Comput 26(10):1269–1281
118. Kanezaki A, Nakayama H, Harada T, Kuniyoshi Y (2010) High-speed 3d object recognition using additive features in a linear subspace. In: 2010 IEEE international conference on robotics and automation, pp 3128–3134

119. Kao CH, Hsieh SP, Peng CC (2010) Study of feature-based image capturing and recognition algorithm. In: ICCAS 2010, pp 1855–1861
120. Kasaei SH, Oliveira M, Lim GH, Lopes LS, Tomé AM (2014) An interactive open-ended learning approach for 3d object recognition. In: 2014 IEEE international conference on autonomous robot systems and competitions (ICARSC), pp 47–52
121. Kasaei SH, Oliveira M, Lim GH, Seabra Lopes L, Tomé AM (2015) Interactive open-ended learning for 3d object recognition: an approach and experiments. J Intell Robot Syst 80(3–4):537–553
122. Kasaei SH, Tomé AM, Lopes LS, Oliveira M (2016) Good: a global orthographic object descriptor for 3d object recognition and manipulation. Pattern Recognition Letters 83(part 3):312–320 **(efficient shape representation, matching, ranking, and its applications)**
123. Ke Y, Sukthankar R (2004) Pca-sift: a more distinctive representation for local image descriptors. In: Proceedings of the 2004 IEEE computer society conference on computer vision and pattern recognition, 2004. CVPR 2004, vol 2, pp II-506–II-513 vol 2
124. Keaikitse M, Govender N, Warrell J (2013) Comparison of active sift-based 3d object recognition algorithms. In: 2013 Africon, pp 1–5
125. Kechagias-Stamatis O, Aouf N (2016) Histogram of distances for local surface description. In: 2016 IEEE international conference on robotics and automation (ICRA), pp 2487–2493
126. Kent D, Behrooz M, Chernova S (2014) Crowdsourcing the construction of a 3d object recognition database for robotic grasping. In: 2014 IEEE international conference on robotics and automation (ICRA), pp 4526–4531
127. Khatun A, Chai WY, Iskandar DA, Islam MR (2011) The effectiveness of ellipsoidal shape representation technique for 3d object recognition system. In: 2011 7th international conference on information technology in Asia, pp 1–6
128. Khatun A, Wang YC, Islam MR, Iskandar DA (2010) 3d shape recognition using wavelet transform based on ellipsoid. In: 2010 International conference on intelligent and advanced systems, pp 1–6
129. Kietzmann TC, Lange S, Riedmiller M (2008) Incremental grlvq: learning relevant features for 3d object recognition. Neurocomputing 71(13–15):2868–2879 **(artificial neural networks (ICANN 2006)/engineering of intelligent systems (ICEIS 2006))**
130. Kim Bs, Xu S, Savarese S (2013) Accurate localization of 3d objects from rgb-d data using segmentation hypotheses. In: 2013 IEEE conference on computer vision and pattern recognition, pp 3182–3189
131. Kim E, Medioni G (2011) 3d object recognition in range images using visibility context. In: 2011 IEEE/RSJ international conference on intelligent robots and systems, pp 3800–3807
132. Kim E, Medioni G (2011) Scalable object classification using range images. In: Proceedings of the 2011 international conference on 3D imaging, modeling, processing, visualization and transmission, 3DIMPVT '11, pp 65–72. IEEE Computer Society, Washington, DC, USA
133. Kim H, Lee J, Lee S (2009) Environment adaptive 3d object recognition and pose estimation by cognitive perception engine. In: Proceedings of the 8th IEEE international conference on computational intelligence in robotics and automation, CIRA'09. IEEE Press, Piscataway, pp 532–539
134. Kim S, Kweon IS (2006) Scalable representation and learning for 3d object recognition using shared feature-based view clustering. In: Proceedings of the 7th Asian conference on computer vision—volume part II, ACCV'06, pp 561–570. Springer, Berlin
135. Kim S, Kweon IS (2007) Robust model-based scene interpretation by multilayered context information. Comput Vis Image Underst 105(3):167–187
136. Kim S, Kweon IS (2008) Scalable representation for 3d object recognition using feature sharing and view clustering. Pattern Recognit 41(2):754–773
137. Kim S, Yoon KJ, Kweon IS (2008) Object recognition using a generalized robust invariant feature and gestalt's law of proximity and similarity. Pattern Recognit 41(2):726–741
138. Kise K, Kashiwagi T (2011) 1.5 million subspaces of a local feature space for 3d object recognition. In: The first Asian conference on pattern recognition, pp 672–676
139. Kitaaki Y, Okuda H, Kage H, Sumi K (2008) High speed 3-d registration using gpu. In: 2008 SICE annual conference, pp 3055–3059
140. Kitchenham B (2004) Procedures for performing systematic reviews. Tech. rep., joint technical report TR/SE-0401
141. Knopp J, Prasad M, Van Gool L (2010) Orientation invariant 3d object classification using hough transform based methods. In: Proceedings of the ACM workshop on 3D object retrieval, 3DOR '10. ACM, New York, pp 15–20
142. Kobayashi T (2013) Generalized mutual subspace based methods for image set classification. In: Proceedings of the 11th Asian conference on computer vision—volume part I, ACCV'12, pp 578–592. Springer, Berlin
143. Kootstra G, Ypma J, de Boer B (2007) Exploring objects for recognition in the real word. In: 2007 IEEE international conference on robotics and biomimetics (ROBIO), pp 429–434
144. Kordelas G, Daras P (2007) Recognizing 3d objects using ray-triangle intersection distances. In: 2007 IEEE international conference on image processing, vol 6, pp VI-173–VI-176
145. Kordelas G, Daras P (2010) Viewpoint independent object recognition in cluttered scenes exploiting ray-triangle intersection and { SIFT } algorithms. Pattern Recognit 43(11):3833–3845
146. Kounalakis T, Boulgouris NV, Triantafyllidis GA (2016) Content-adaptive pyramid representation for 3d object classification. In: 2016 IEEE international conference on image processing (ICIP), pp 231–235
147. Kumar D, Nishchal NK (2015) Three-dimensional object recognition using joint fractional fourier transform correlators with the help of digital fresnel holography. Optik Int J Light Electron Opt 126(20):2690–2695
148. Kushal A, Ponce J (2006) Modeling 3d objects from stereo views and recognizing them in photographs. In: Proceedings of the 9th European conference on computer vision - volume part II, ECCV'06. Springer, Berlin, ,pp 563–574
149. Kushal A, Schmid C, Ponce J (2007) Flexible object models for category-level 3d object recognition. In: 2007 IEEE conference on computer vision and pattern recognition, pp 1–8
150. Laboratory RPIIP, Meagher D (1980) Octree encoding: a new technique for the representation, manipulation and display of arbitrary 3-d objects by computer. Stanford University, Stanford
151. Lam J, Greenspan M (2013) 3d object recognition by surface registration of interest segments. In: Proceedings of the 2013 international conference on 3D Vision, 3DV '13, pp 199–206. IEEE Computer Society, Washington, DC, USA
152. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
153. Lee BG, Liliana Shin DH (2010) Enhanced computational integral imaging system for partially occluded 3d objects using occlusion removal technique and recursive { PCA } reconstruction. Opt Commun 283(10):2084–2091
154. Lee S, Kim E, Park Y (2006) 3D object recognition using multiple features for robotic manipulation. In: Proceedings 2006 IEEE

international conference on robotics and automation, 2006. ICRA 2006, pp 3768–3774
155. Lee S, Lu Z (2011) Dependable 3d object recognition with two-layered particle filter. In: Proceedings of the 5th international conference on ubiquitous information management and communication, ICUIMC '11. ACM, New York, NY, pp 37:1–37:8
156. Lee S, Lu Z, Kim H (2011) Probabilistic 3d object recognition with both positive and negative evidences. In: Proceedings of the 2011 international conference on computer vision, ICCV '11. IEEE Computer Society, Washington, DC, pp 2360–2367
157. Lee S, Wei L, Naguib AM (2016) Adaptive bayesian recognition and pose estimation of 3d industrial objects with optimal feature selection. In: 2016 IEEE international symposium on assembly and manufacturing (ISAM), pp 50–55
158. Lee SH, Cheng SC, Chang CC (2014) Moment-preserving techniques for 3d shape registration and recognition. In: 2014 International symposium on computer, consumer and control, pp 516–519
159. Lee TK, Drew MS (2007) 3D object recognition by eigen-scale-space of contours. Springer, Berlin, pp 883–894
160. Leibe B, Schiele B (2003) Analyzing appearance and contour based methods for object categorization. In: Proceedings of 2003 IEEE computer society conference on computer vision and pattern recognition, 2003. vol 2, pp 409–415
161. Li C, Boheren J, Hager GD (2015) Bridging the robot perception gap with mid-level vision. In: International symposium on robotics research (ISRR)
162. Li C, Bohren J, Carlson E, Hager GD (2016) Hierarchical semantic parsing for object pose estimation in densely cluttered scenes. In: 2016 IEEE international conference on robotics and automation (ICRA), pp 5068–5075
163. Li X, Godil A, Wagan A (2008) 3d part identification based on local shape descriptors. In: Proceedings of the 8th workshop on performance metrics for intelligent systems, PerMIS '08. ACM, New York, pp 162–166
164. Li X, Guskov I (2007) 3d object recognition from range images using pyramid matching. In: 2007 IEEE 11th international conference on computer vision, pp 1–6
165. Liang D, Weng K, Wang C, Liang G, Chen H, Wu X (2014) A 3d object recognition and pose estimation system using deep learning method. In: 2014 4th IEEE international conference on information science and technology, pp 401–404
166. Lin D, Fidler S, Urtasun R (2013) Holistic scene understanding for 3d object detection with rgbd cameras. In: Proceedings of the 2013 IEEE international conference on computer vision, ICCV '13, pp 1417–1424. IEEE Computer Society, Washington, DC, USA
167. Lin WY (2006) Robust geometrically invariant features for two-dimensional shape matching and three-dimensional face recognition. Ph.D. thesis, University of Wisconsin at Madison, Madison, WI, USA. AAI3234848
168. Lina: Recognition of 3d objects in various capturing conditions using appearance manifolds. In: 2010 the 2nd international conference on computer and automation engineering (ICCAE), vol 2, pp 349–352 (2010)
169. Takahashi T, Ide I, Murase H (2008) Construction of appearance manifold with embedded view-dependent covariance matrix for 3d object recognition. IEICE Trans Inf Syst 91(4):1091–1100
170. Liu YJ, Fu QF, Liu Y, Fu XL (2012) 2d-line-drawing-based 3d object recognition. In: Proceedings of the first international conference on computational visual media, CVM'12. Springer, Berlin, pp 146–153
171. Logoglu KB, Kalkan S, Temizel A (2016) Cospair: colored histograms of spatial concentric surflet-pairs for 3d object recognition. Robot Auton Syst 75:558–570
172. Loo CH, Elsherbeni AZ (2008) Optoelectronic 3-d object classification from 2-d images. J Lightw Technol 26(18):3248–3255
173. Lowe DG (1987) Three-dimensional object recognition from single two-dimensional images. Artif Intell 31(3):355–395
174. Lowe DG (1999) Object recognition from local scale-invariant features. In: Proceedings of the international conference on computer vision, vol 2, ICCV '99. IEEE Computer Society, Washington, DC, p 1150
175. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
176. Lu Z, Lee S, Kim H (2011) Probabilistic 3d object recognition based on multiple interpretations generation. In: Proceedings of the 10th Asian conference on computer vision—volume part IV, ACCV'10. Springer, Berlin, pp 333–346
177. Luciw MD, Weng J (2008) Topographic class grouping with applications to 3d object recognition. In: 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), pp 3987–3994
178. Luo RC, Kuo CW (2015) A scalable modular architecture of 3d object acquisition for manufacturing automation. In: 2015 IEEE 13th international conference on industrial informatics (INDIN), pp 269–274
179. Luo RC, Kuo CW, Chung YT (2015) Model-based 3d object recognition and fetching by a 7-dof robot with online obstacle avoidance for factory automation. In: 2015 IEEE international conference on robotics and automation (ICRA), pp 2647–2652
180. Geetha M, Paul MP, Kaimal MR (2014) An improved content based image retrieval in rgbd images using point clouds. In: 2014 International conference on communication and signal processing, pp 828–832
181. Ma H, Huang T, Wang Y (2010) Multi-resolution recognition of 3d objects based on visual resolution limits. Pattern Recognit Lett 31(3):259–266
182. Ma S, Zhou C, Zhang L, Hong W, Tian Y (2013) 3d object recognition using kernel pca based on depth information for twist-lock grasping. In: 2013 ieee international conference on robotics and biomimetics (ROBIO), pp 2667–2672
183. Madi K, Paquet E, Seba H, Kheddouci H (2015) Graph edit distance based on triangle-stars decomposition for deformable 3d objects recognition. In: Proceedings of the 2015 international conference on 3D vision, 3DV '15, pp 55–63. IEEE Computer Society, Washington, DC, USA
184. Maeda M, Nakamae T, Inoue K (2012) Surface matching by curvature distribution images generated via gaze modeling. In: Proceedings of the 21st international conference on pattern recognition (ICPR2012), pp 2194–2197
185. Mahiddine A, Merad D, Drap P, m. Boï J (2014) Partial 3d-object retrieval using level curves. In: 2014 6th international conference of soft computing and pattern recognition (SoCPaR), pp 77–82
186. Marini S, Spagnuolo M, Falcidieno B (2007) Structural shape prototypes for the automatic classification of 3d objects. IEEE Comput Gr Appl 27(4):28–37
187. Marques M, Costeira J (2009) Lamp: linear approach for matching points. In: 2009 16th IEEE international conference on image processing (ICIP), pp 2113–2116
188. Marr D (1982) Vision: a computational investigation into the human representation and processing of visual information. Henry Holt and Co., Inc., New York
189. Matas J, Chum O, Urban M, Pajdla T (2004) Robust wide-baseline stereo from maximally stable extremal regions. Image Vis Comput 22(10):761–767 **(British machine vision computing 2002)**
190. Mateo CM, Gil P, Torres F (2014) A performance evaluation of surface normals-based descriptors for recognition of objects using cad-models. In: 2014 11th international conference on

informatics in control, automation and robotics (ICINCO), vol 02, pp 428–435
191. Mavrinac A, Shawky A, Chen X (2008) A fuzzy associative approach for recognition of 3d objects in arbitrary pose. In: 2008 IEEE international conference on fuzzy systems (IEEE world congress on computational intelligence), pp 710–715
192. McCullagh P (2002) What is a statistical model? Ann Stat 30(5):1225–1310. https://doi.org/10.1214/aos/1035844977
193. Megherbi N, Han J, Breckon TP, Flitton GT (2012) A comparison of classification approaches for threat detection in ct based baggage screening. In: 2012 19th IEEE international conference on image processing, pp 3109–3112
194. Mery D, Riffo V, Zuccar I, Pieringer C (2013) Automated x-ray object recognition using an efficient search algorithm in multiple views. In: Proceedings of the 2013 IEEE conference on computer vision and pattern recognition workshops, CVPRW '13. IEEE Computer Society, Washington, DC, pp 368–374
195. Mhamdi MAA, Ziou D (2014) A local approach for 3d object recognition through a set of size functions. Image Vis Comput 32(12):1030–1044
196. Mian A, Bennamoun M, Owens R (2010) On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes. Int J Comput Vis 89(2):348–361
197. Mian AS, Bennamoun M, Owens R (2006) Three-dimensional model-based object recognition and segmentation in cluttered scenes. IEEE Trans Pattern Anal Mach Intell 28(10):1584–1601
198. Mian AS, Bennamoun M, Owens RA (2006) A novel representation and feature matching algorithm for automatic pairwise registration of range images. Int J Comput Vis 66(1):19–40
199. Mikolajczyk K, Schmid C (2002) An affine invariant interest point detector. In: Proceedings of the 7th European conference on computer vision-part I, ECCV '02. Springer, London, pp 128–142
200. Morel JM, Yu G (2009) Asift: a new framework for fully affine invariant image comparison. SIAM J Imaging Sci 2(2):438–469
201. Mouton A, Breckon TP, Flitton GT, Megherbi N (2014) 3d object classification in baggage computed tomography imagery using randomised clustering forests. In: 2014 IEEE international conference on image processing (ICIP), pp 5202–5206
202. Muja M, Rusu RB, Bradski G, Lowe DG (2011) Rein—a fast, robust, scalable recognition infrastructure. In: 2011 IEEE international conference on robotics and automation, pp 2939–2946
203. Murase H, Nayar SK (1996) Learning by a generation approach to appearance-based object recognition. In: Proceedings of the 13th international conference on pattern recognition, 1996, vol 1. IEEE, pp 24–29
204. Naguib AM, Lee S (2015) An adaptive evidence structure for bayesian recognition of 3d objects. In: Proceedings of the 9th international conference on ubiquitous information management and communication, IMCOM '15. ACM, New York, pp 75:1–75:8
205. Naikal N, Yang AY, Sastry SS (2010) Towards an efficient distributed object recognition system in wireless smart camera networks. In: 2010 13th international conference on information fusion, pp 1–8
206. Nair V, Hinton GE (2009) 3d object recognition with deep belief nets. In: Proceedings of the 22nd international conference on neural information processing systems, NIPS'09. Curran Associates Inc., pp 1339–1347
207. Naji D, Fakir M, Bouikhalene B, Elayachi R (2016) Recognition of 3d objects using heat diffusion equations and random forests. In: 2016 13th international conference on computer graphics, imaging and visualization (CGiV), pp 161–166
208. Nakashika T, Hori T, Takiguchi T, Ariki Y (2014) 3d-object recognition based on llc using depth spatial pyramid. In: Proceedings of the 2014 22nd international conference on pattern recognition, ICPR '14. IEEE Computer Society, Washington, DCpp 4224–4228
209. Nelleri A, Gopinathan U, Joseph J, Singh K (2006) Three-dimensional object recognition from digital fresnel hologram by wavelet matched filtering. Opt Commun 259(2):499–506
210. Nene SA, Nayar SK, Murase H (1996) Columbia object image library (coil-100). Technical report, computer vision Laboratory, Department of Computer Science, Columbia University
211. Nian R, Ji G, Zhao W, Feng C (2007) Probabilistic 3d object recognition from 2d invariant view sequence based on similarity. Neurocomputing 70(4–6):785–793. Advanced Neurocomputing Theory and Methodology selected papers from the international conference on intelligent computing 2005 (ICIC 2005) international conference on intelligent computing 2005
212. Noceti N, Delponte E, Odone F (2009) Spatio-temporal constraints for on-line 3d object recognition in videos. Computer vision and image understanding 113(12):1198–1209. Special issue on 3D Representation for Object and Scene Recognition
213. Noma A, Cesar Jr, RM (2010) Sparse representations for efficient shape matching. In: Proceedings of the 2010 23rd SIBGRAPI conference on graphics, patterns and images, SIBGRAPI '10. IEEE Computer Society, Washington, DC, pp 186–192
214. Nowak E, Jurie F, Triggs B (2006) Sampling strategies for bag-of-features image classification. Springer, Berlin, pp 490–503
215. Okada K, Kojima M, Tokutsu S, Maki T, Mori Y, Inaba M (2007) Multi-cue 3d object recognition in knowledge-based vision-guided humanoid robot system. In: 2007 IEEE/RSJ international conference on intelligent robots and systems, pp 3217–3222
216. Okal B, Nüchter A (2013) Sliced curvature scale space for representing and recognizing 3d objects. In: 2013 16th international conference on advanced robotics (ICAR), pp 1–7
217. Okuda H, Kitaaki Y, Hashimoto M, Kaneko S (2006) HM-ICP: fast 3-d registration algorithm with hierarchical and region selection approach of M-ICP. JRM 18(6):765–771
218. Oleari F, Rizzini DL, Caselli S (2013) A low-cost stereo system for 3d object recognition. In: 2013 IEEE 9th international conference on intelligent computer communication and processing (ICCP), pp 127–132
219. Ong LY, Chong CW, Besar R (2007) An approach to 3-d object recognition using legendre moment invariants. In: 2007 International conference on intelligent and advanced systems, pp 671–674
220. Orts-Escolano S, Morell V, Garcia-Rodriguez J, Cazorla M, Fisher RB (2015) Real-time 3d semi-local surface patch extraction using gpgpu. J Real Time Image Process 10(4):647–666
221. Osman MK, Mashor MY, Arshad MR, Saad Z (2009) 3d object recognition using manfis network with orthogonal and non-orthogonal moments. In: 2009 5th international colloquium on signal processing its applications, pp 302–306
222. Owechko Y, Medasani S, Korah T (2010) Automatic recognition of diverse 3-d objects and analysis of large urban scenes using ground and aerial lidar sensors. In: CLEO/QELS: 2010 laser science to photonic applications, pp 1–2
223. Palossi D, Tombari F, Salti S, Ruggiero M, Stefano LD, Benini L (2013) Gpu-shot: parallel optimization for real-time 3d local description. In: 2013 IEEE conference on computer vision and pattern recognition workshops, pp 584–591
224. Pang B, Ma H (2011) An effective way of 3d model representation in recognition system. In: Proceedings of the 2011 international conference on multimedia and signal processing—volume 01, CMSP '11. IEEE Computer Society, Washington, DC, pp 107–111
225. Pang G, Neumann U (2013) Training-based object recognition in cluttered 3d point clouds. In: Proceedings of the 2013 international conference on 3D vision, 3DV '13. IEEE Computer Society, Washington, DC, pp 87–94

226. Pang G, Neumann U (2015) Fast and robust multi-view 3d object recognition in point clouds. In: Proceedings of the 2015 international conference on 3D vision, 3DV '15. IEEE Computer Society, Washington, DC, pp 171–179
227. Papazov C, Burschka D (2011) An efficient ransac for 3d object recognition in noisy and occluded scenes. In: Proceedings of the 10th Asian conference on computer vision—volume part I, ACCV'10. Springer, Berlin, pp 135–148
228. Pepik B, Stark M, Gehler P, Schiele B (2012) Teaching 3d geometry to deformable part models. In: 2012 IEEE conference on computer vision and pattern recognition, pp 3362–3369
229. Petrícek T, Svoboda T (2012) Area-weighted surface normals for 3d object recognition. In: Proceedings of the 21st international conference on pattern recognition (ICPR2012), pp 1492–1496
230. Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Chang J, Hoffman K, Marques J, Min J, Worek W (2005) Overview of the face recognition grand challenge. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol 1, pp 947–954
231. Pichler A, Bauer H, Eberst C, Heindl C, Minichberger J (2006) Towards more agility in robot painting through 3d object recognition. In: Pham D, Eldukhri E, Soroka A (eds) Intelligent production machines and systems. Elsevier Science Ltd, Oxford, pp 608–613
232. Pichler A, Bauer H, Heindl C, Minichberger J, Eberst C (2007) Recognition and 6DOF localisation of parts for Lotsize1 automation. IFAC Proc 40(3):265–270 **(8th IFAC workshop on intelligent manufacturing systems)**
233. Ping W, Wei W, Ying-hui G, Shi-fei L (2012) A recognition approach of 3-d objects based on the tsallis entropy. In: 2012 International conference on computer vision in remote sensing, pp 242–245
234. Pintilie S, Ghodsi A (2010) Conformal mapping by computationally efficient methods. In: Proceedings of the twenty-fourth AAAI conference on artificial intelligence, AAAI'10. AAAI Press, pp 557–562
235. Övünç Polat Yıldırım T (2007) Recognition of patterns without feature extraction by grnn. In: Proceedings of the 8th international conference on adaptive and natural computing algorithms, Part II, ICANNGA '07. Springer, Berlin, pp 161–168
236. Övünç Polat Yıldırım T (2008) Genetic optimization of GRNN for pattern recognition without feature extraction. Expert Syst Appl 34(4):2444–2448
237. Qian X, Ye C (2014) 3D object recognition by geometric context and gaussian-mixture-model-based plane classification. In: 2014 IEEE international conference on robotics and automation (ICRA), pp 3910–3915
238. Ramalingam S, Liu ZQ, Iourinski D (2006) Curvature-based fuzzy surface classification. IEEE Trans Fuzzy Syst 14(4):573–589
239. Rangel JC, Morell V, Cazorla M, Orts-Escolano S, García-Rodríguez J (2015) Using gng on 3d object recognition in noisy rgb-d data. In: 2015 International joint conference on neural networks (IJCNN), pp 1–7
240. Raptis SN, Koutsouris D (2006) Bayesian fusion of contour descriptions: application to 3-d object and face recognition. In: 2006 IET conference on crime and security, pp 438–444
241. Ravari AN, Taghirad HD (2013) Unsupervised 3d object classification from range image data by algorithmic information theory. In: 2013 First RSI/ISM international conference on robotics and mechatronics (ICRoM), pp 319–324
242. Ravari AN, Taghirad HD (2014) Transformation invariant 3d object recognition based on information complexity. In: 2014 Second RSI/ISM international conference on robotics and mechatronics (ICRoM), pp 902–907
243. Raytchev B, Mino T, Tamaki T, Kaneda K (2010) View-invariant object recognition with visibility maps. In: Proceedings of the 2010 20th international conference on pattern recognition, ICPR '10. IEEE Computer Society, Washington, DC, pp 1040–1043
244. University of York. NHS Centre for Reviews, Dissemination: undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews. CRD report. NHS Centre for Reviews and Dissemination, University of York (2001)
245. Ribeiro F, Brandão S, Costeira JP, Veloso M (2015) Global localization by soft object recognition from 3d partial views. In: 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 3709–3714
246. Rocha LF, Malaca P, Silva J, Moreira AP, Veiga G (2015) Development of a 3d model based part recognition system for industrial applications: main challenges. In: 2015 IEEE international conference on industrial technology (ICIT), pp 3296–3301
247. Rodner E, Hegazy D, Denzler J (2010) Multiple kernel gaussian process classification for generic 3d object recognition. In: 2010 25th International conference of image and vision computing New Zealand, pp 1–8
248. Rodolà E, Albarelli A, Bergamasco F, Torsello A (2013) A scale independent selection process for 3d object recognition in cluttered scenes. Int J Comput Vis 102(1–3):129–145
249. Rodríguez-Sánchez AJ, Szedmak S, Piater J (2015) Scurv: A 3d descriptor for object classification. In: 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 1320–1327
250. Rodrigues Ja, Lam R, du Buf H (2012) Cortical 3d face and object recognition using 2d projections. Int J Creat Interaces Comput Gr 3(1):45–62
251. Rothwell CA, Zisserman A, Forsyth DA, Mundy JL (1992) Canonical frames for planar object recognition. In: Sandini G (ed) Computer vision—ECCV'92. Springer, Berlin, pp 757–772
252. Rui Y, She AC, Huang TS (1996) Modified fourier descriptors for shape representation – a practical approach. In: Proc of first international workshop on image databases and multi media search
253. Rusu RB, Blodow N, Beetz M (2009) Fast point feature histograms (fpfh) for 3d registration. In: 2009 IEEE international conference on robotics and automation, pp 3212–3217
254. Rusu RB, Bradski G, Thibaux R, Hsu J (2010) Fast 3d recognition and pose using the viewpoint feature histogram. In: 2010 IEEE/RSJ international conference on intelligent robots and systems, pp 2155–2162
255. Rusu RB, Cousins S (2011) 3d is here: point cloud library (pcl). In: 2011 IEEE international conference on robotics and automation, pp 1–4
256. Salas-Moreno RF, Newcombe RA, Strasdat H, Kelly PHJ, Davison AJ (2013) Slam++: Simultaneous localisation and mapping at the level of objects. In: Proceedings of the 2013 IEEE conference on computer vision and pattern recognition, CVPR '13. IEEE Computer Society, Washington, DC, pp 1352–1359
257. Salgian AS (2007) Using multiple patches for 3d object recognition. In: 2007 IEEE conference on computer vision and pattern recognition, pp 1–6
258. Salgian AS (2008) Combining local descriptors for 3d object recognition and categorization. In: 2008 19th International conference on pattern recognition, pp 1–4
259. Salih Y, Malik AS, Sidibé D, Simsim MT, Saad N, Meriaudeau F (2014) Compressed vfh descriptor for 3d object classification. In: 2014 3DTV-conference: the true vision—capture, transmission and display of 3D video (3DTV-CON), pp 1–4
260. Salti S, Tombari F, Stefano LD (2011) A performance evaluation of 3d keypoint detectors. In: Proceedings of the 2011 international conference on 3D imaging, modeling, processing,

visualization and transmission, 3DIMPVT '11. IEEE Computer Society, Washington, DC, pp 236–243
261. van de Sande K, Gevers T, Snoek C (2010) Evaluating color descriptors for object and scene recognition. IEEE Trans Pattern Anal Mach Intell 32(9):1582–1596
262. Sanguino TJM, Gómez FP (2015) Improving 3d object detection and classification based on kinect sensor and hough transform. In: 2015 International symposium on innovations in intelligent systems and applications (INISTA), pp 1–8
263. Šeatović D (2008) A segmentation approach in novel real time 3D plant recognition system. In: Gasteratos A, Vincze M, Tsotsos JK (eds) Computer vision systems. ICVS 2008. Lecture notes in computer science, vol 5008. Springer, Berlin, Heidelberg, pp 363–372
264. Selinger A, Nelson RC (1999) A perceptual grouping hierarchy for appearance-based 3d object recognition. Comput Vis Image Underst 76(1):83–92
265. Shah SAA, Bennamoun M, Boussaid F (2016) A novel feature representation for automatic 3d object recognition in cluttered scenes. Neurocomputing 205:1–15
266. Shah SAA, Bennamoun M, Boussaid F (2017) Keypoints-based surface representation for 3d modeling and 3d object recognition. Pattern Recognit 64:29–38
267. Shah SAA, Bennamoun M, Boussaid F, El-Sallam AA (2013) A novel local surface description for automatic 3d object recognition in low resolution cluttered scenes. In: 2013 IEEE international conference on computer vision workshops, pp 638–643
268. Shaiek A, Moutarde F (2013) Fast 3d keypoints detector and descriptor for view-based 3d objects recognition. In: Revised selected and invited papers of the international workshop on advances in depth image analysis and applications, vol 7854. Springer, New York, pp 106–115
269. Sheta AF, Baareh A, Al-Batah M (2012) 3d object recognition using fuzzy mathematical modeling of 2d images. In: 2012 International Conference on Multimedia Computing and Systems, pp 278–283
270. Shilane P, Min P, Kazhdan M, Funkhouser T (2004) The Princeton shape benchmark. In: In Shape modeling international, pp 167–178
271. Shimamura J, Yoshida T, Taniguchi Y, Yabushita H, Sudo K, Murasaki K (2015) The method based on view-directional consistency constraints for robust 3d object recognition. In: 2015 14th IAPR international conference on machine vision applications (MVA), pp 455–458
272. Shivaswamy PK, Jebara T (2006) Permutation invariant svms. In: Proceedings of the 23rd international conference on machine learning, ICML '06. ACM, New York, NY, pp 817–824
273. Smeets D, Fabry T, Hermans J, Vandermeulen D, Suetens P (2010) Inelastic deformation invariant modal representation for non-rigid 3d object recognition. In: Proceedings of the 6th international conference on articulated motion and deformable objects, AMDO'10. Springer, Berlin, pp 162–171
274. Socher R, Huval B, Bhat B, Manning CD, Ng AY (2012) Convolutional-recursive deep learning for 3d object classification. In: Proceedings of the 25th international conference on neural information processing systems, NIPS'12. Curran Associates Inc., USA, pp 656–664
275. Soodamani R, Liu ZQ (1998) Object recognition using fuzzy modelling and fuzzy matching. In: 1998 IEEE international conference on fuzzy systems proceedings. IEEE world congress on computational intelligence (Cat. No.98CH36228), vol 1, pp 165–170
276. Soysal M, Alatan AA (2015) Joint utilization of local appearance and geometric invariants for 3d object recognition. Multimedia Tools Appl 74(8):2611–2637
277. Stasse O, Dupitier S, Yokoi K (2006) 3d object recognition using spin-images for a humanoid stereoscopic vision system. In: 2006 IEEE/RSJ international conference on intelligent robots and systems, pp 2955–2960
278. Takei S, Akizuki S, Hashimoto M (2014) 3d object recognition using effective features selected by evaluating performance of discrimination. In: 2014 13th International conference on control automation robotics vision (ICARCV), pp 70–75
279. Takei S, Akizuki S, Hashimoto M (2015) Short: A fast 3d feature description based on estimating occupancy in spherical shell regions. In: 2015 International conference on image and vision computing New Zealand (IVCNZ), pp 1–5
280. Tan AH, Godfrey KR (2002) The generation of binary and near-binary pseudorandom signals: an overview. IEEE Trans Instrum Meas 51(4):583–588
281. Tangruamsub S, Takada K, Hasegawa O (2011) 3D object recognition using a voting algorithm in a real-world environment. In: Proceedings of the 2011 IEEE workshop on applications of computer vision (WACV), WACV '11. IEEE Computer Society, Washington, DC, pp 153–158
282. Tateno K, Tombari F, Navab N (2015) Real-time and scalable incremental segmentation on dense slam. In: 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 4465–4472
283. Tateno K, Tombari F, Navab N (2016) When 2.5d is not enough: Simultaneous reconstruction, segmentation and recognition on dense slam. In: 2016 IEEE international conference on robotics and automation (ICRA), pp 2295–2302
284. Taylor G, Kleeman L (2014) Visual perception and robotic manipulation: 3D object recognition, tracking and hand–eye coordination. Springer, Berlin
285. Teh YW, Jordan M (2010) Hierarchical Bayesian nonparametric models with applications. Cambridge University Press, Cambridge, UK
286. Tombari F, Di Stefano L (2010) Object recognition in 3d scenes with occlusions and clutter by hough voting. In: Proceedings of the 2010 fourth Pacific-Rim symposium on image and video technology, PSIVT '10. IEEE Computer Society, Washington, DC, pp 349–355
287. Tombari F, Salti S, Di Stefano L (2010) Unique shape context for 3d data description. In: Proceedings of the ACM workshop on 3D object retrieval, 3DOR '10. ACM, New York, NY, pp 57–62
288. Tombari F, Salti S, Di Stefano L (2010) Unique signatures of histograms for local surface description. Springer, Berlin, pp 356–369
289. Tombari F, Salti S, Stefano LD (2011) A combined texture-shape descriptor for enhanced 3d feature matching. In: 2011 18th IEEE international conference on image processing, pp 809–812
290. Treiber MA (2010) An introduction to object recognition: selected algorithms for a wide variety of applications, 1st edn. Springer, Berlin
291. Trudeau R (1993) Introduction to graph theory. Dover books on mathematics. Dover Pub, Mineola
292. Truong HQ, Lee S, Jang SW (2008) Model-based recognition of 3d objects using intersecting lines. In: 2008 IEEE international conference on multisensor fusion and integration for intelligent systems, pp 656–660
293. Ullman S, Basri R (1991) Recognition by linear combinations of models. IEEE Trans Pattern Anal Mach Intell 13(10):992–1006. https://doi.org/10.1109/34.99234
294. Ulrich M, Wiedemann C, Steger C (2012) Combining scale-space and similarity-based aspect graphs for fast 3d object recognition. IEEE Trans Pattern Anal Mach Intell 34(10):1902–1914
295. Unel M, Soldea O, Ozgur E, Bassa A (2010) 3d object recognition using invariants of 2d projection curves. Pattern Anal Appl 13(4):451–468

296. Urdiales C, de Trazegnies C, Pacheco J, Sandoval F (2010) View planning for efficient contour-based 3d object recognition. In: Melecon 2010–2010 15th IEEE mediterranean electrotechnical conference, pp 206–211
297. Usui Y, Kondo K (2010) 3d object recognition based on confidence lut of sift feature distance. In: 2010 Second world congress on nature and biologically inspired computing (NaBIC), pp 293–297
298. Vázquez RA, Sossa H, Garro BA (2007) 3d object recognition based on low frequency response and random feature selection. In: Proceedings of the artificial intelligence 6th mexican international conference on advances in artificial intelligence, MICAI'07. Springer, Berlin, pp 694–704
299. Vázquez RA, Sossa H, Garro BA (2009) The role of the infant vision system in 3d object recognition. In: Köppen M, Kasabov N, Coghill G (eds) Adv Neuro Inform Process. Springer, Berlin, pp 800–807
300. Wan LL, Miao ZJ (2009) 3d object classification by fuzzy knn and bayesian decision. In: Proceedings of the 2009 fifth international conference on intelligent information hiding and multimedia signal processing, IIH-MSP '09. IEEE Computer Society, Washington, DC, pp 455–458
301. Wang D, Qian H (2008) 3d object recognition by fast spherical correlation between combined view egis and pft. In: 2008 19th International conference on pattern recognition, pp 1–4
302. Wang J, Lu J, Chen W, Wu X (2015) Convolutional neural network for 3d object recognition based on rgb-d dataset. In: 2015 IEEE 10th conference on industrial electronics and applications (ICIEA), pp 34–39
303. Wang S, Wang Y, Jin M, Gu X, Samaras D (2006) 3d surface matching and recognition using conformal geometry. In: Proceedings of the 2006 IEEE computer society conference on computer vision and pattern recognition, vol 2, CVPR '06. IEEE Computer Society, Washington, DC, pp 2453–2460
304. Wang S, Wang Y, Jin M, Gu XD, Samaras D (2007) Conformal geometry and its applications on 3d shape matching, recognition, and stitching. IEEE Trans Pattern Anal Mach Intell 29(7):1209–1220
305. Wang Y, Sun G, Wang C, Han D (2010) Research on 3d object recognition from wire-frame based on edge moment. In: 2010 3rd International conference on advanced computer theory and engineering (ICACTE), vol 1, pp V1-78–V1-82
306. Weiss I, Ray M (1998) Model-based recognition of 3D objects from one view. Springer, Berlin, pp 716–732
307. Westell J, Saeedi P (2010) 3d object recognition via multi-view inspection in unknown environments. In: 2010 11th International conference on control automation robotics vision, pp 2088–2095
308. Wohlkinger W, Aldoma A, Rusu RB, Vincze M (2012) 3dnet: Large-scale object class recognition from cad models. In: 2012 IEEE international conference on robotics and automation, pp 5384–5391
309. Wohlkinger W, Vincze M (2010) 3d object classification for mobile robots in home-environments using web-data. In: 19th International workshop on robotics in Alpe-Adria-Danube Region (RAAD 2010), pp 247–252
310. Wohlkinger W, Vincze M (2011) Ensemble of shape functions for 3d object classification. In: 2011 IEEE international conference on robotics and biomimetics, pp 2987–2992
311. Wohlkinger W, Vincze M (2011) Shape distributions on voxel surfaces for 3d object classification from depth images. In: 2011 IEEE international conference on signal and image processing applications (ICSIPA), pp 115–120
312. Wu J, Fukui K (2008) Multiple view based 3d object classification using ensemble learning of local subspaces. In: 2008 19th International conference on pattern recognition, pp 1–4
313. Xia S, Hancock ER (2008) 3d object recognition using hypergraphs and ranked local invariant features. In: Proceedings of the 2008 joint IAPR international workshop on structural, syntactic, and statistical pattern recognition, SSPR & SPR '08. Springer, Berlin, pp 117–126
314. Xia Y, Zhang L, Xu W, Shan Z, Liu Y (2015) Recognizing multiview objects with occlusions using a deep architecture. Inform Sci 320:333–345
315. Xiang Y, Savarese S (2012) Estimating the aspect layout of object categories. In: 2012 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 3410–3417
316. Xiang Y, Savarese S (2012) Estimating the aspect layout of object categories. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition
317. Xiang Y, Savarese S (2013) Object detection by 3d aspectlets and occlusion reasoning. In: Proceedings of the 2013 IEEE international conference on computer vision workshops, ICCVW '13. IEEE Computer Society, Washington, DC, pp 530–537
318. Xing W, Liu W, Yuan B (2006) Volumetric part based 3d object classification. In: 2006 5th IEEE international conference on cognitive informatics, vol 1, pp 405–412
319. Xing W, Liu W, Yuan B (2007) 3d object classification system based on volumetric parts. In: 2007 IEEE international conference on systems, man and cybernetics, pp 984–990
320. Xing W, Liu W, Yuan B, Lu W (2007) An integrated system for 3d object reconstruction and recognition. In: Proceedings of the 7th WSEAS international conference on simulation, modelling and optimization, SMO'07. World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, pp 281–284
321. Xing W, Yuan B, Liu M, Tang X (2006) 3D object classification by part features fusion. In: 2006 8th international conference on signal processing, vol 2
322. Xu Q, Wan W, Wang J, An X (2015) The application of local features in 3-dimentional object recognition. In: 2015 International conference on smart and sustainable city and big data (ICSSC), pp 96–100
323. Xu S, Peng Qc (2008) 3d object recognition using multiple features and neural network. In: 2008 IEEE conference on cybernetics and intelligent systems, pp 434–439
324. Xu X, Dehghani A, Corrigan D, Caulfield S, Moloney D (2016) Convolutional neural network for 3d object recognition using volumetric representation. In: 2016 first international workshop on sensing, processing and learning for intelligent machines (SPLINE), pp 1–5
325. Xu YH, Luo RH, Min HQ (2012) Label transfer for joint recognition and segmentation of 3d object. In: 2012 International conference on machine learning and cybernetics, vol. 3, pp 1188–1192
326. Yabushita H, Osawa T, Shimamura J, Taniguchi Y (2013) Mobile visual search for 3-d objects: Matching user-captured video to single reference image. In: 2013 IEEE 2nd global conference on consumer electronics (GCCE), pp 122–123
327. Yabushita H, Shimamura J, Morimoto M (2012) A framework of three-dimensional object recognition which needs only a few reference images. In: Proceedings of the 21st international conference on pattern recognition (ICPR2012), pp 1375–1378
328. Yoon KJ, Shin MG, Lee JH (2010) Recognizing 3d objects with 3d information from stereo vision. In: 2010 20th International conference on pattern recognition, pp 4020–4023
329. Yoshikawa N, Ii Y (2006) Three-dimensional object recognition using multiplex complex amplitude information with support function. In: Proceedings of the first international conference on innovative computing, information and control, volume 1, ICICIC '06. IEEE Computer Society, Washington, DC, pp 314–317

330. Yu J, Weng K, Liang G, Xie G (2013) A vision-based robotic grasping system using deep learning for 3d object recognition and pose estimation. In: IEEE international conference on robotics and biomimetics, ROBIO 2013, Shenzhen, China, December 12–14, 2013, pp 1175–1180
331. Yu X, Gao Y, Zhou J (2014) Face recognition using 3d directional corner points. In: 2014 22nd International conference on pattern recognition, pp 2802–2807
332. Zang C, Hashimoto K, Moon J (2011) A visual tracking strategy using computer graphics and edge. In: 2011 IEEE international conference on robotics and biomimetics, pp 981–986
333. Zarpalas D, Kordelas G, Daras P (2011) Recognizing 3d objects in cluttered scenes using projection images. In: 2011 18th IEEE international conference on image processing, pp 673–676
334. Zhai JH, Wang XZ, Zhang SF, Li J (2007) View-based 3d object recognition using wavelet multiscale singular-value decomposition and support vector machine. In: 2007 International conference on wavelet analysis and pattern recognition, vol 3, pp 1428–1432
335. Zhang X, Liu Y, Gao C, Liu J (2008) An isomap-eigenanalysis-regression pose estimation algorithm of three-dimentional object. In: 2008 Second international symposium on intelligent information technology application, vol 3, pp 61–65
336. Zhong Y (2009) Intrinsic shape signatures: a shape descriptor for 3d object recognition. In: 2009 IEEE 12th international conference on computer vision workshops, ICCV Workshops, pp 689–696
337. Zhou J, Cadavid S, Abdel-Mottaleb M (2010) Histograms of categorized shapes for 3d ear detection. In: 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp 1–6
338. Zhou J, Cadavid S, Abdel-Mottaleb M (2011) A computationally efficient approach to 3d ear recognition employing local and holistic features. In: CVPR 2011 Workshops, pp 98–105
339. Zhu F, Shao L (2014) Weakly-supervised cross-domain dictionary learning for visual recognition. Int J Comput Vis 109(1):42–59
340. Zhu F, Shao L, Fang Y (2016) Boosted cross-domain dictionary learning for visual categorization. IEEE Intell Syst 31(3):6–18
341. Zhuang Y, Lin X, Hu H, Guo G (2015) Using scale coordination and semantic information for robust 3-d object recognition by a service robot. IEEE Sens J 15(1):37–47
342. Zia MZ, Stark M, Schindler K (2014) Are cars just 3d boxes? jointly estimating the 3d shape of multiple objects. In: Proceedings of the 2014 IEEE conference on computer vision and pattern recognition, CVPR '14. IEEE Computer Society, Washington, DC, pp 3678–3685
343. Zografos V, Buxton BF (2007) Pose-invariant 3d object recognition using linear combination of 2d views and evolutionary optimisation. In: Proceedings of the international conference on computing: theory and applications, ICCTA '07. IEEE Computer Society, Washington, DC, pp 645–649
344. Zou F, Wang Y, Yang Y, Zhou K, Chen Y, Song J (2015) Supervised feature learning via l2-norm regularized logistic regression for 3d object recognition. Neurocomputing 151(2):603–611