PRESIOUS

Our reference: PRESIOUS

Predictive digitization, restoration and degradation assessment of cultural heritage objects

COOPERATION

D2.1 – State of the Art Report on 3D Object Digitisation and Shape Matching/Retrieval Methods for Unstructured 3D data

Project Ref. No.	FP7-ICT-2011-9 – FP7-600533
Project Acronym	PRESIOUS
Project Start Date (duration)	1 Feb 2013 (36M)
Deliverable Due Date	31 Jul 2013 (M6)
Actual Delivery Date	24 Jul 2013 (M6)
Deliverable Leader	Ioannis Pratikakis ATHENA-RC
Document Status	Final
Dissemination Level	PU

Deliverable Identification Sheet

Project Ref. No.	FP7-ICT-2011-9 – FP7-600533	
Project Acronym	PRESIOUS	
Document Name	PRESIOUS-D2.1-24072013-v1.0	
Contractual Delivery Date	31 Jul 2013 (M6)	
Deliverable Number	D2.1	
Deliverable Name	State of the Art Report on 3D Object Digitisation and Shape Matching/Retrieval Methods for Unstructured 3D data	
Type	Report	
Deliverable Version	1.0	
Status	Final	
Associated WP / Task	WP2 / T2.1, T2.2	
Author(s)	Ioannis Pratikakis Anestis Koutsoudis Michalis Savelonas	ATHENA-RC ATHENA-RC ATHENA-RC
Other Contributors	Dirk Rieke-Zapp	BREUCKMANN GMBH
Project Officer	Philippe Gelin	
Abstract	D2.1 presents the STAR on 3D object digitisation, shape matching and retrieval methods, along with local shape descriptors for unstructured 3D data. The first part of this report focuses on current CH 3D object digitisation techniques and practices, including the handling of special CH object forms, technologies, typical acquisition problems related to the task at hand and an evaluation of relevant project results. The second part reviews current whole-from-partial shape matching and retrieval methods, along with local shape descriptors for unstructured 3D data, such as point clouds.	
Keywords	3D digitisation, unstructured 3D data, partial 3D object retrieval, 3D shape descriptors, Cultural heritage	
Sent to Internal Reviewer	17.06.2013	
Internal Review Completed	20.06.2013	
Circulated to Participants	24.07.2013	
Read by Participants	30.07.2013	
Approved by General Assembly	31.07.2013	

Document Revision History

Date	Version	Author/Editor/Contributor	Summary of Changes
1/6/2013	0.70	Michalis Savelonas, Ioannis Pratikakis	First draft.
10/6/2013	0.80	Anestis Koutsoudis	Adding CH-related text.
12/6/2013	0.90	Dirk Rieke-Zapp	Adding Breuckmann-related information.
17/6/2013	0.92	Michalis Savelonas, Ioannis Pratikakis, Anestis Koutsoudis	Text enhancements.
24/7/2013	1.00	Michalis Savelonas, Ioannis Pratikakis, Anestis Koutsoudis	Changes addressing internal review comments – Final version

Table of Contents

EXECUTIVE SUMMARY

This report presents the state-of-the-art (STAR) related to PRESIOUS WP 2. Its first part focuses on current CH 3D object digitisation techniques and practices, including the handling of special CH object forms, technologies, typical acquisition problems related to the task at hand. The investigation leads to the conclusion that the complete 3D digitisation of objects is still far from a "*single button press*" procedure. In many cases, previous digitisation experiences, such as digitisation plans or even legacy data, can be of great help to address new digitisation challenges. In addition, a review of relevant project results is provided. Among all projects reviewed, the EU-funded 3D-COFORM is probably the single most relevant project to PRESIOUS. However, 3D-COFORM, has concentrated on high accuracy 3D scanning, whereas PRESIOUS will concentrate on the development of a predictive digitisation technique that can drastically reduce acquisition times and promote the use of low-cost, low accuracy, off-the-shelf technology, such as the Microsoft Kinect-based solutions.

The second part of this report reviews current whole-from-partial shape matching and retrieval methods, along with local shape descriptors for unstructured 3D data, such as point clouds. To this end, partial 3D object retrieval methods are presented, along with 3D object retrieval methods which were originally proposed for complete queries but in some aspect they are particularly relevant to partial retrieval. Taking into account the existence of a rich literature for retrieval of structured data, this section also includes methods that were originally proposed for 3D meshes. This choice is imposed by the fact that the majority of the related literature is devoted on structured data represented by 3D meshes. However, such methods are still relevant in the context of PRESIOUS WP2, since there is an intense research activity on 3D mesh generation algorithms, which can be used to convert unstructured input data, such as point clouds.

Another conclusion drawn from the study of STAR is that most partial 3D object retrieval methods are actually 2.5D, i.e. they derive descriptors from 2D projections of the 3D model. This approach suits partial retrieval in the sense that a 2D query model can be compared with all 2D projections of each target model, in order to keep the minimum distance associated with the most similar projection. In addition, most partial 3D object retrieval methods are based on descriptors calculated over interest points, either dense, or extracted by means of a salient point detector. Interest points constitute also an intrinsic element of the bag of visual words (BoVW) paradigm. The latter is an emerging trend in 3D object retrieval, with several major works appearing in the last two years.

The second part ends with a review on local shape descriptors which have been proposed for unstructured data. Local shape descriptors are suited to the problem at hand, since a partial query and its associated full model are intuitively expected to be similar in a local fashion. A large part of the descriptors studied calculate and compare histograms over a neighborhood centered at each point of interest. Another trend is the use of point distances within a neighborhood. This type of information is the closest one to raw point coordinates provided in point cloud input data. On the other hand, surface normals, which are used either for the formation of local projections or for the calculation of feature descriptors, ask for a normal estimation algorithm. Furthermore, most local descriptors are 2.5D, calculating descriptors over 2D projections. Finally, it can be noted that with few exceptions, the rest of the local shape descriptors presented were not initially proposed for 3D object retrieval. The most usual application addressed is object recognition in scenes. In principle, there is nothing prohibitive in applying the same descriptors for 3D object retrieval.

PART A- CH 3D OBJECT DIGITISATION TECHNIQUES AND PRACTICES

A1 INTRODUCTION

PRESIOUS WP2 aims at the development of a predictive 3D digitisation platform, applicable on cultural heritage (CH) objects. This platform will facilitate a reduction in effort and time, as well as a simplification of procedures followed for CH object digitisation. This STAR survey aims to identify 3D digitisation technologies which can serve as starting points for the research of PRESIOUS WP2 and point-out current problems and bottlenecks. In that respect, apart from the presentation of 3D digitisation techniques, this STAR survey also summarizes main results of relevant past projects.

For over a decade now 3D digitisation is considered a common practice in the CH domain. It provides methods that help contribute on the important issue of digital preservation and dissemination of cultural thesaurus. The peculiarities of CH objects are reflected on the existence of a plethora of 3D digitisation methodologies. The high complexity 3D digitisation requirements of CH objects have actually contributed to the development of different approaches that attempt to successfully address particular types of objects varying from a movable artefact up to a monument. As there is still no 3D digitisation method that can be considered a panacea, the combination of multiple methods is something common for several digitisation projects[1]. An extensive study on the available methods for the 3D digitisation of the cultural heritage thesaurus has been conducted by the ‘Athena Research and Innovation Center. One of the important outcomes of this study is the construction of a nine-criteria table (Table 1), which summarises the possible parameters for choosing a 3D digitisation system for cultural heritage applications[2].

As with any digitisation project, its financial plan is highly correlated to its requirements specifications and thus, its actual breadth and scope. In many cases, the selection of the most applicable method is prohibitive due to budget issues. Furthermore, the cost criterion of a digitisation project is being examined in relation to numerous affecting parameters such: i) as the available personnel and its level of expertise, ii) the duration of the digitisation phase (e.g. availability and access time to the artefact or the monument), iii) other issues related to the actual digitisation environment such as the location of the artefact or the monument (e.g. museum, private collection, open space, archaeological excavation, etc.) and the presence of other people (e.g. curators, visitors, security, etc.) during the digitisation phase. In the case that the available time and budget are not an issue, there are three main factors that influence the suitability and the applicability of each technique for the application at hand: (i) complexity in size and shape, (ii) morphological complexity and (iii) diversity of raw materials. There are techniques suitable for microscopic objects, others for small, medium and large objects and others for monuments, open spaces and architectural ensembles. Also, there are different techniques for ceramic or metallic or glass objects.

In Section A2 of this STAR survey we review digitisation technologies and typical problems of handling special CH object forms, whereas in Section A3 we evaluate the results of previous related projects.

[1] Lin, H.-Y., Wu, J.-R., 2008. 3D Reconstruction by combining shape from silhouette with stereo, Proc. Int. Conf. Pattern Recognition (ICPR).

[2] Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., Tsioukas, V., Chamzas, C., 2007. Methods for 3D digitization of cultural heritage, Journal of Cultural Heritage, 8, 93-98.

No.	Criterion
1	Cost
2	Material of digitisation subject
3	Size of digitisation subject
4	Portability of equipment
5	Accuracy of the system
6	Texture acquisition
7	Productivity of the technique
8	Skill requirements
9	Compliance of produced data with standards

*Table 1. Criteria for the choose of an appropriate digitisation system*2

	Triangulation	Time delay	Monocular images	Active	Direct	Range	Surface orientation
Laser triangulators	X			X	X	X	
Structured light	X				X	X	
Time of flight		X		X	X	X	
Shape from stereo	X				X	X	
Shape from focus			X			X	
Shape from shadows			X	X		X	
Shape from shading			X	X			X
Photometry			X	X			X
Photogrammetry	X			X	X	X	

*Table 2. Classification of main 3D digitisation techniques*4

A2 3D DIGITISATION TECHNOLOGIES

Current 3D digitisation techniques can be classified to contact and non-contact. Contact-based techniques are not popular in the CH domain due to the fragile nature of CH artefacts. In contrast, non-contact-based techniques have been successfully used during the last decade in many CH digitisation projects[3]. Non-contact techniques can be divided into active and passive. Passive techniques use the reflectance of the object and the illumination of the scene to derive the shape information, whereas in the active form, suitable light sources are used as the internal vector of information. A distinction can also be made between direct and indirect measurements. Direct techniques result in range data, i.e. into a set of distances between the unknown surface and the range sensor. Indirect measurements are inferred from monocular images and from prior knowledge of the target properties. They result either

[3] Koutsoudis, A., Stavroglou, K., Pavlidis, G., Chamzas, C., 2012. 3DSSE – A 3D scene search engine: exploring 3D scenes using keywords, Journal of Cultural Heritage 13 (2) 187-194.

in range data or in surface orientation[4]. Apart from the application-oriented reviews of Pavlidis et al.[2] and Sansoni et al.[4], a review of 3D digitisation techniques has been provided by Blais[5].

2.1. A2.1 Laser triangulation

One of the most widely used active acquisition techniques is laser triangulation (LT). The method is based on a system with a laser source and an optical detector. The laser source emits light in the form of a spot, a line or a pattern on the surface of the object, whereas the optical detector captures the produced deformations (Figure 1).

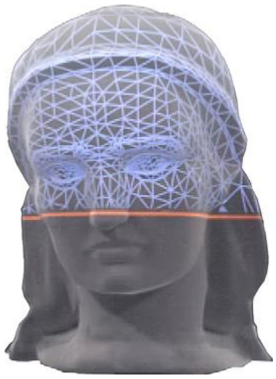


*Figure 1. Laser scanning of object*2

Depth is computed by using the triangulation principle, which is schematically illustrated in Figure 2. Points O_P and O_C are the exit and the entrance pupils of a laser source and of a camera. Their mutual distance is the baseline d. The optical axes z_P and z_C of the laser and the camera form an angle of α degrees. The laser source generates a narrow beam, impinging the object at point S (single-point triangulators). The back-scattered beam is imaged at point S' at image plane κ. The measurement of the location (i_S, j_S) of image point S' defines the line of sight $\overline{S'O_C}$, and, by means of simple geometry calculations, yields the position of S. The measurement of the surface is achieved by scanning. In a conventional triangulation configuration, a compromise is necessary between the field of view (FOV), the measurement resolution and uncertainty, and the shadow effects due to large values of angle α. To overcome this limitation, a method called 'synchronized scanning' has been proposed. Using the approach of Rioux and Blais[6], a large FOV can be achieved with a small angle α, without sacrificing range measurement precision.

Laser stripes exploit the optical triangulation principle shown in Figure 2. However, in this case, the laser is equipped with a cylindrical lens, which expands the light beam along one direction. Hence, a plane of light is generated, and multiple points of the object are illuminated at the same time. In the figure, the light plane is denoted by λ_s, and the illuminated points belong to the intersection between the plane and the unknown object (line $\overline{AB}$). The measurement of the location of all the image points from A' to B' at plane κ allows the determination of the 3D shape of the object in correspondence with the illuminated points. For dense reconstruction, the plane of light must scan the scene[7]. An interesting

[4] Sansoni, G., Trebeschi, M., Docchio, F., 2009. State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation, Sensors 9, 568-601.

[5] Blais, F. A review of 20 years of range sensors development, 2004. J. Electronic Imaging 13, 231-240.

[6] Rioux, M., Blais, F. Compact three-dimensional camera for robotics applications, 1986. J. Opt. Soc. Am. 3, 1518-1521.

[7] Trucco, E., Fisher, R.B., Fitzgibbon, A.W., Naidu, D.K., 1998. Calibration, data consistency and model acquisition with laser stripes. Int. J. Comput. Integrated Manuf. 11, 293-310.

enhancement of the basic principle exploits the Biris principle[8]. A laser source produces two stripes by passing through the Biris component. The measurement of the point location at the image plane can be carried out on a differential basis, and shows decreased dependence on the object reflectivity and on the speckle noise produced by laser beam.

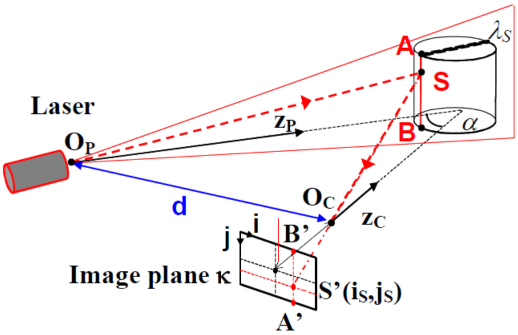


*Figure 2. Schematics of the triangulation principle*4.

Laser light is very bright and highly focused for long distances, resulting in an emitted pattern which can be always focused on the surface of the objects. One of the most significant advantages of laser scanners is their high accuracy in geometry measurements (<50μm) and dense sampling (<100μm)[9], as well as their relative insensitivity to illumination conditions and surface texture effects. On the other hand, it should be noted that in many such systems, geometry is potentially extracted without any texture information. Additionally, special attention should be paid for surface properties, such as reflectance and transparency. One other important aspect is the high cost of such devices, which renders this method useful to specific applications. Finally, the productivity of the method, as well as the portability, depends upon the used system and can vary significantly9. Some of the most broadly used LT systems are the Konica Minolta Vivid 3D Scanners[10] and the Next Engine system[11]. The latter offers a USB powered colour multi-laser solution with an integrated lighting system.

2.2. A2.2 Structured light

Structured Light (SL) is another popular active method that is based on projecting a sequence of different density bi-dimensional patterns of non-coherent light on the surface of an object and extracting the 3D geometry by monitoring the deformations of each pattern (Figure 3). This method is also based on triangulation but does not need to use specific laser sources. In many cases it is confused with the laser scanning methods and there are commercial systems that cannot be absolutely categorized to the one or the other method.

[8] Lang, S., Dorba, L., Blais, F., Lecavalier, M. Characterization and testing of the BIRIS range sensor, 1993. Proc. IMTC , 459-464.

[9] Rioux, M., 1994. Digital 3-D imaging: theory and applications, Proc. Int. Symp. Photonic and Sensors and Controls for Commercial Applications, Boston, 2650, 1994, pp. 2-15.

[10] Konica Minolta Vivid 3D Scanners, http://www.konicaminolta.com/instruments/products/3d/non-contact/vivid910/index.html

[11] NextEngine system (NextEngine, Inc., CA, USA), http://www.nextengine.com

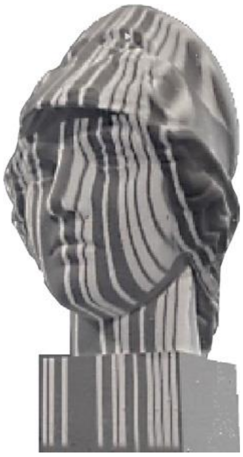


*Figure 3. Shape from structured light*2

The method works by projecting a specific predefined light pattern that covers the entire or part of the surface of the objects. This scene is then captured by a typical digital image detector and processed in order to deduce the geometry from the deformations of the pattern in the digital image. These patterns can be simple multiple fringes of different colours (Figure 4) or complex patterns with curves, either time- or space-coded. The method is accompanied by texture acquisition and can lead to very impressive results in terms of accuracy and productivity. The systems are usually portable and easy to use2.

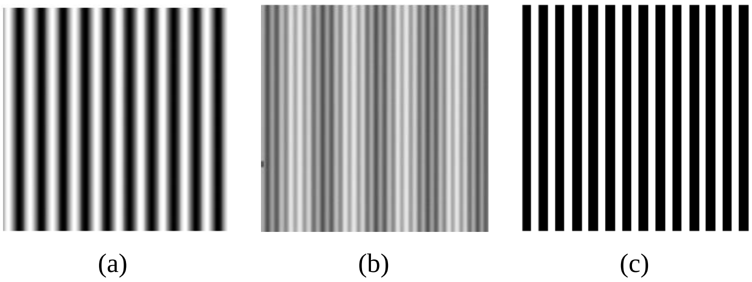


*Figure 4. Example of fringe projection schemes: (a) fringe pattern of sinusoidal fringes, (b) superposition of two sinusoidal fringe patterns at different frequencies, (c) projection of fringes of rectangular profile*4*.*

Current research is focused on developing SL systems that are able to capture 3D surfaces in real-time. This is achieved by increasing the speed of projection patterns and capturing algorithms. Breuckmann GmbH plays an important role towards this direction[12] [13]. The current generation of smart*SCAN* ^{3D} and stereo*SCAN* ^{3D} system provides rapid acquisition times (<1 sec per view). Additionally, the accuracy offered by Breuckmann's SL systems (smart*SCAN* ^{3D} offers a feature accuracy of ± 7 μm, depending on the FOV) is similar or sometimes surpasses commercial LT systems. In combination with properties such as portability, user-friendliness, straightforward calibration procedures, and the availability of color cameras, Breuckmann's systems are considered as primary solutions in several CH digitisation projects such as detailed analysis of marble-made sculptures, mould making, in-situ artefact or monument 3D digitisation, etc.

[12] Breuckmann arts and culture, http://www.aicon3d.com/breuckmann-scanner/kunst-kultur/kunst.html

[13] Breuckmann smartSCAN, http://www.aicon3d.com/products/breuckmann-scanner/smartscan/at-a-glance.html

Breuckmann scanners are operated using OPTOCAT software to control the scanning process as well as for processing of data. This software generates meshes from the acquired images and provides fast and robust alignment and merging functionality for each individual mesh. The powerful hole filling tool may be applied to create watertight 3D models. OPTOCAT is highly customizable and also supports data acquisition with turn tables, control of external sensors, i.e. external flashes or cameras, seamless integration of Aicon´s DPA photogrammetry projects, sensor tracking and control of data acquisition by means of industrial robots. Besides acquisition and processing of data, OPTOCAT includes inspection tools for taking measurements as well as quick comparison of measured to CAD data. A wide range of data export formats and interfaces to common inspection software allow for a seamless integration in quality control workflows, often used in industrial applications. One of the latest features added to OPTOCAT is the texture mapping module that allows for straightforward texturing of 3D meshes using sensor or external imagery.

2.3. A2.3 Time of flight

The time-of-flight (TOF) active method is commonly used for the 3D digitisation of architectural ensembles, such as urban areas of cultural importance, monuments and excavations. The emitter unit generates a laser pulse, which impinges onto the target surface. A receiver detects the reflected pulse, and suitable electronics measure the round-trip travel time of the returning signal, as well as its intensity. Single point sensors perform point-to-point distance measurement and scanning devices are combined to the optical head, so as to cover large bi-dimensional scenes. Large range sensors allow measurement ranges from 15m to 100m (reflective markers must be put on the target surfaces). Medium range sensors can acquire 3D data in shorter ranges. Measurement resolutions vary with the range. For large measuring ranges, TOF sensors give excellent results. For smaller objects, about one meter in size, attaining accuracy of 1/1000 with TOF radar requires very high speed timing circuitry, because the time differences are in the pico-second range. The amplitude and frequency modulated radar of Nielsen et al.[14] has shown promise for close range distance measurement. In fact, for closer range distance measurements, systems that are based on the phase-shift (PS) active method are preferred as they produce 3D point clouds of higher accuracy at lower acquisition times[15] [16]. Furthermore, apart from resolution limitations, the range may be limited by allowable power levels of laser radiation, as determined by laser safety considerations. Additionally, TOF sensors face difficulties with shiny surfaces, which reflect little back-scattered light energy, except when oriented perpendicularly to the line of sight.

As surface colour, reflectivity and transparency are factors that affect the ability of an acquisition system to capture the shape of an object, current research is aimed at the development of methodologies that are able to capture challenging objects with metallic specular or transparent surfaces. Bajard et al.[17] described a technique that works for those approaches that classic laser triangulation fails. Their technique is based on the emission of a heated pattern on the object surface. The heated zone emits an infrared radiation which is observed by an infrared camera and hence used to determine the distance between the system and a point on the object’s surface. The ability to automate the digitisation procedure has also been an important research domain. Breuckmann’s d-

[14] Nielsen, T., Bormann, F., Wolbeck, S., Spiecker, H., Burrows, M.D., Andersen, P, 1996. Time-of-flight analysis of light pulses with a temporal resolution of 100ps. Rev. Sci. Instrum. 67, 1721-1724.

[15] Alonso, J.I., Martinez Rubio, J., Fernadez Martin, J., Garcia Fernandez, J., 2011. Comparing time-of-flight and phase-shift, the survey of the royal pantheon in the basilica of San Isidoro (Leon), Proc. ISPRS Workshop ‘3D-ARCH’.

[16] Wenguo L., Suping F., Shaujun D., 2012. 3D shape measurement based on structured light projection applying polynomial interpolation technique, Int. J. Light and Electron Optics.

[17] Bajard, A., Aubretron, O., Eren, G., Sallamand, P., Truchetet, F., 2011. 3D digitisation of metallic specular surfaces using scanning from heating approach, Proc SPIE-IS&T Electronic Images 7864.

STATION offers an automated portable 3D digitisation controlled environment of small-sized objects, such as jewelry, coins, pottery shreds, etc[18].

2.4. A2.4 Image-based methods

Image-based (IM) methods can be considered as a passive version of SL. In principle, IM methods involve stereo calibration, feature extraction, feature correspondence analysis and depth computation based solely on the corresponding points. Photogrammetry is the primary IM method that can be used to extract 2D as well as 3D geometrical properties from a static scene that is depicted from different viewpoints into an image set.

Photogrammetry can be divided into two major categories: (i) aerial photogrammetry, where images are usually overhead shots captured with the help of an aircraft or more recently with an unmanned aerial vehicle (UAV) and (ii) terrestrial photogrammetry, which is based on image sets captured from positions near the surface of earth. Terrestrial photogrammetry is also called close-range photogrammetry, when the distance between the camera and object being captured is less than 100m. The accuracy of the measurements produced by photogrammetry can be considered as a function of the camera optics, sensor quality and resolution. Numerous photogrammetric software, both commercial and open source, are able today to perform procedures such as camera calibration, epipolar geometry computations, as well as 3D mesh generation and texture mapping. Furthermore, photogrammetric measurements, when combined with accurate measurements, which are derived from a total station, can produce models of high accuracy for scales of 1:100 or higher. The method is objective and reliable and can be aided by CAD software. It can be used for complex objects with high surface detail, but since it is based on photos, there is a need for adequate space. It is also useful when direct access to the monument is prohibited, as well as to record stages of the monument in time[19] 2 [20] 4. When combined with accurate measurements, it can deliver at any scale of application accurate, metric and detailed 3D information with estimates of precision and reliability of the unknown parameters from the measured image correspondences[21]. Applications of precise 3D documentation of CH have been proposed by Gruen et al.[22] and El-Hakim et al.[23]. However, images need to be acquired using satellite, aerial or terrestrial sensors, which are not an option in the context of PRESIOUS WP2.

Close to multi-image photogrammetry, the stereo vision (SV) can be considered as a simplified set-up, which is used to extract 3D information from an image-pair that has been captured by two cameras, positioned at a given distance (baseline) with slightly different viewing angles. When certain parts of the object in the scene are visible to both photographs, specific vision algorithms can be applied to extract object geometry. The external as well as the internal parameters of the optical system are used for calibration. Calibration is again critical in terms of achieving accurate measurements. The method can either be fully automated or manually operated. The final result is a depth map of the object in the scene, reflecting the distance of each recognized point on the surface of the object from the photographic sensor.

[18] D-STATION at a glance, http://www.aicon3d.com/products/breuckmann-scanner/d-station/at-a-glance.html

[19] Grussenmeyer, P., Khalil, O.A., 2002. Solutions for exterior orientation in photogrammetry: a review, The Photogrammetric Record 17 (100), 615-634.

[20] Jiang, R., Jauregui, D.V., White, K.R., 2008. Close-range photogrammetry applications in bridge measurement: literature review, J. Measurement 41 (8), 823-834.

[21] Remondino, F., 2011. Heritage recording and 3D modeling with photogrammetry and 3D scanning, Remote Sens. 3, 1104-1138.

[22] Gruen, A., Remondino, F., Zhang, L., 2004. Photogrammetric reconstruction of the Great Buddha of Bamiyan, The Photogrammetric Record 19, 177-199.

[23] El-Hakim, S., Gonzo, L., Voltolini, F., Girardi, S., Rizzi, A., Remondino, F., Whiting, E., 2007. Detailed 3D modelling of castles, Int. J. Architect. Comput. 5, 199-220.

The main advantages of SV include the ability to capture both geometry and texture, the low cost and portability. On the other hand, SV has low resolution, whereas it requires correct identification of common points between images, i.e. to solve the well-known correspondence problem. Moreover, the quality of shape extraction depends on the sharpness of the surface texture, which is affected by variations in surface reflectance. Furthermore, stereo vision based on the detection of the object silhouette has recently become very popular. The stereo approach is exploited, but the 3D geometry of the object is retrieved by using the object contours under different viewing directions. A basic requirement for this technique is that the scene does not contain moving parts. 3D urban modeling and object detection are typical applications[24].

SV has been widely applied in robotic and computer vision, where the essence of the problem is not the accurate acquisition of high quality data, but rather, their interpretation. CH modeling is also a prominent application of SV, especially when the texture information is captured together with the range[25].

Recently, the continuous increase of computation power and multi-threaded processors technology allowed the evolution of automated image-based methods. Such methods emerge from the combination of structure-from-motion (SFM) and dense multi-view 3D reconstruction (DMVR) methods. At present, several software solutions implementing SFM-DMVR methods are offered as web services or as stand-alone applications. The SFM method can be considered as the extension of SV. Instead of image pairs, the method attempts to reconstruct depth from a number of images that depict a static object from arbitrary viewpoints. Thus, apart from the feature extraction phase, the trajectories of corresponding features over the image collection are also computed.

The method relies on the positions of corresponding features that have been recognised between the different images, in order to calculate the intrinsic and extrinsic camera parameters (e.g. focal length, image format, principal point, lens distortion coefficients, location of the projection centre and the 3D image orientation)[26]. The bundle-adjustment method is used in SFM in order to improve the accuracy of calculating the cameras 3D positions, to minimise the projection-error and to reduce the error-built up introduced by the cameras' 3D positions tracking[27]. Snavely et al. [28] have recently created an open source SFM system named Bundler. Similarly Wu et al.[29] developed a version of bundle-adjustment that exploits hardware parallelism (multi-core acceleration) to provide lower computational times. Additionally, they have integrated their SFM system with a user friendly front-end and offer it as an open-source solution. Similarly, Agisoft[30] offers PhotoScan, an SFM-DMVR software solution that is able to merge the independent depth maps of all images and produce a single vertex painted point cloud that can be converted to a triangulated 3D mesh of different densities.

Shape from focus digitization is recursive and is based on taking photos of an object while continuously adjusting the focus plane. Knowing the position of this focus plane from the entire setup and system positioning, it is possible to map focused pixels in an image onto the correct position in the 3D depth map. The system, recursively, rebuilds the whole object geometry, photo by photo.

[24] DeSouza, G., Kak, A., 2002. Vision for mobile robot navigation: A Survey, IEEE T. Pattern. Anal. Mach. Intell. 24, 237-267.

[25] Baumberg, A., Lyons, A., Taylor, R., 2003. 3DSOM-a commercial software solution to 3D scanning, Vision, Video, and Graphics, The Euro-Association Eurographics Partner Event, Video, and Graphics.

[26] Robertson, D.P., Cipolla, R., Structure from motion, Practical Image Processing and Computer Vision, John Willey and Sons Ltd, 2009.

[27] Engels, C., Stewenius, H., Nister, D., 2006. Bundle adjustment rules. Proc. Photogrammetric Computer Vision Conference.

[28] Snavely, N., 2008. Bundler: structure from motion for unordered image collections, http://phototour.cs.washington.edu/ bundler/#S4

[29] Wu, C., VisualSFM: a visual structure from motion system, http://www.cs.washington.edu/homes/ccwu/vsfm

[30] Agisoft PhotoScan, http://www.agisoft.ru

Resolution as well as accuracy is limited, but the results are, in general, ‘‘reliable’’. This technique has evolved from the passive approach to an active sensing strategy. In the passive case, surface texture is used to determine the amount of blurring. Thus, the object must have surface texture covering the whole surface in order to extract shape. The active version of the method operates by projecting light onto the object to avoid difficulties in discerning surface texture2.

Most prior work in active depth from focus has yielded moderate accuracy up to one part per 400, over the FOV. The cost of this technique is relatively high, since in order to take photos with such limited depth of field, one might need specialized lens. In addition, there is a direct trade-off between depth of view and FOV; satisfactory depth resolution is achieved at the expense of sub-sampling the scene, which in turn requires some form of mechanical scanning to acquire range measurement over the entire scene. Moreover, spatial resolution is non-uniform, since depth resolution is substantially less than resolution perpendicular to the observation axis. Finally, objects not aligned perpendicularly to the optical axis and having a depth dimension greater than the depth of view will come in to focus at different ranges, complicating the scene analysis and interpretation4.

2.5. A2.5 Shape from shadows

The shape from shadow technique is a variant of structured light, rebuilding the 3D model by capturing the shadow of a known object projected onto the target, as the light is moving. The main advantage of this method is its low cost and the limited demand for computing power, at the expense of low accuracy. It can reconstruct geometry even in nonvisible object parts, under certain assumptions4.

2.6. A2.6 Shape from shading

Shape from shading requires the capturing of the object from one viewing angle. Varying the light source position causes the shading on the surface of the object to also vary. Using multiple photos of different shading conditions, the object surface geometry could be deduced. The method is simple and has low cost. It captures both geometry and texture, with a minor disadvantage in capturing texture in shaded areas. It is portable but has the disadvantage of low accuracy, especially in the presence of external factors influencing object reflectance. Moreover, the measurement software is rather complex2 4. Horn and Brooks[31] provide an earlier review of shape from shading algorithms.

2.7. A2.7 Photometry

Shape from photometry is a variant of shape from shading, where lighting conditions are varying instead of the viewing angle. In addition, the usage of reference objects, or reference lighting sources, in the scene is critical, since they are used for calibration. Calibrated lights may significantly enhance the result, but can only be found in special laboratories. Some reports favor photometry in terms of the produced data volume and the immunity to laser limitations. Generally it can be regarded as portable, easy to use and low cost. The main disadvantage is its current need for a laboratory environment2.

2.8. A2.8 Discussion

The main characteristics of optical range imaging techniques are summarised in Table 2, which is taken from Sansoni et al.4. The comments on the table are quite general in nature, and a number of exceptions are known. Strengths and weaknesses of the different techniques are strongly application-dependent. The common attribute of being non-contact is an important consideration in those applications, which are characterised by fragile deformable objects, on-line measurements or hostile environments. Acquisition time is another important aspect, when the overall cost of the measurement

[31] Horn, B., Brooks, M.J., 1989. Shape from Shading, MIT Press: Cambridge, MA.

process suggests the reduction of the operation time even at the expense of quality deterioration. Range and FOV limitations are crucial when digitising morphological complex objects. In combination with other properties such as weight, ease of transportation and power supply, they compose a complex environment in which the digitisation team has to operate in order to deliver. For example, the digitisation of a monument using a laser scanner will require in many cases the set-up of temporal scaffolding and potentially dangerous working conditions in order to perform a complete digitisation.

It is impossible to acquire a complete digital model of a monument or an artefact from just a single-scan viewpoint. The digitisation of the same object using image-based methods will also require the same temporal scaffolding or the use of an UAV. It is a fact that terrestrial range scanners are still bulky and slow when compared with a digital camera. On the other hand, marker-less image-based methods have issues when reconstructing surfaces of low features. The low frequency of colour alternations in conjunction with bad lighting in hollow areas, compose a challenging digitisation surface. Range scanners provide a more stable solution for a variety of different surface types but when it comes to colour information, image-based methods still provide higher quality sensors. In addition, the same digital image set is used to create the 3D geometry and colour information. Thus, an absolute correlation between geometrical and colour information is achieved. In case of the 3D range scanners, the above correlation cannot be taken always for granted. In order for a range scanner to acquire increased colour information than its embedded sensor, an additional digital camera is necessary. This digital camera has to be calibrated and registered in relation to the laser source of the range scanner, in order to achieve an alignment between colour and geometry. Furthermore, all 3D range scanners with an internal digital camera that has a shifted point of view from that of the range sensor, are prone to colour information ghosting, due to the different perspective projections of the same real-world scene. Colour information bleeding is another common problem found in many range scanners, which is attributed to the different sampling resolution between the digital camera and the range scanning subsystem[32].

Furthermore, active vision systems using a laser beam of a given laser class that illuminates the object, inherently involve safety considerations, and possible surface interaction effects. In addition, the speckle effect introduces disturbances that represent a serious limit to the accuracy of the measurements[33]. Fortunately, recent advances in LED technology yield a valuable solution to this problem, since they are suitable candidates as light projectors in active triangulation systems. The dramatic evolution of CPUs and memories has led to increased performances at low costs. For these reasons, techniques that are computationally demanding, such as passive stereo vision, are now more reliable and efficient.

For a given depth measurement problem, the selection of the suitable sensor type depends on: (i) the measurement time, (ii) the budget and (iii) the quality expected from the measurement. In this respect, 3D imaging sensors may be affected by missing or bad quality data. Reasons are related to the optical geometry of the system, the type of projector and/or acquisition optical device, the measurement technique and the characteristics of the target objects. The sensor performance may depend on the dimension, the shape, the texture, the temperature and the accessibility of the object. Relevant factors that influence the choice are also the ruggedness, the portability, the adaptability of the sensor to the measurement problem, the easiness of the data handling, and the simplicity of use of the sensor.

More specifically, in terms of digitisation, Boehler and Marbs[34] mention that the accuracy of the specifications provided by laser scanner manufactures are not always comparable and these should not be trusted as the accuracy of these instruments which are usually built in small numbers varies from instrument to instrument and depends on the individual calibration and the way the instruments are

[32] Case Studies for Testing the Digitisation Process – An interim report, http://3dicons-project.eu/eng/Resources

[33] Dorsch, R.G., Häusler, G., Herrmann, J.M, 1994. Laser triangulation: fundamental uncertainty in distance measurement, Appl. Opt. 33, 1306-1314.

[34] Boehler, W., Marbs, A., 2003. Investigating laser scanner accuracy.

handled. Accuracy is an important aspect of digitisation. In the general case of laser scanners there are several properties in the systems that affect the collected data[34]. Some of the most important are the following:

i) *Angular accuracy* – in the case of LT scanners, the laser pulse is deflected by a small rotating device such as a mirror or a prism positioned on a mechanical axis and angular accuracy corresponds to the stepping of the deflection line/plane. Angular accuracy directly affects the measured point density (see resolution below), especially in conjunction with the range of the target. For SL systems, it is affected by the sensor resolution and the FOV of the imaging lens system,

ii) *Range accuracy* - range is computed using time-of-flight or a phase comparison between the outgoing and the returning signal. In case of triangulation scanners the accuracy of range diminishes with the square of the distance between the scanner and the actual object. A systematic scale error is present in any spatial distance measurement. This error can be eliminated when distance differences for a given range are known. Nevertheless, systematic range errors depend on the reflective material and hence a universal error correction cannot be determined. Furthermore, it is known that this range drift uncertainty is non-uniform with respect to the imaging plane, especially in the case of dual- or multi-sensor scanners,

iii) *Resolution* – from the end user’s point of view this is the discrimination ability of the acquisition system. Technically this can be seen as the smallest possible increment of the angle between two successive digitisation points in relation to the laser spot on the object,

iv) *Edge Effects* – despite the focusing of the laser spot on the surface of the object, when the laser hits an object’s edge only a part of it will be reflected back. The rest will be either reflected from an adjacent surface or not at all. This produces a variety or erroneous measured points for both range and triangulation scanners. Such points are inevitable as the laser spot cannot be focused to point size. Edge effects are the main source of non-uniform noise, especially when the edges are not present in multiple partial scans in order to perform outlier rejection,

v) *Surface Reflectivity* - laser scanners rely on a signal that is reflected from the object’s surface. The *strength* of the return signal is affected by the: distance, atmospheric conditions, incidence angle, surface colour, size of laser spot which hits an area with two colours of extreme brightness difference, and reflective ability of the surface. Bright surfaces provide stronger reflections in relation with dark surfaces. Single spectral laser scanners are highly affected by the colour surface. Shiny surfaces are considered challenging to be captured by such devices, as the saturation and scattering of the laser spot results in erroneous measurements depicted as high levels of surface deviations, also known as spikes. Some scanners provide the ability to change the laser source power in order to handle such challenging surfaces. For objects consisting of different materials or differently painted or coated surfaces, one should expect serious errors. This can be avoided for objects that do not belong in the CH domain, by using a temporary coating,

vi) *Environmental Conditions* – the temperature and pressure variations and the interfering radiation, including sunlight and lamps, are some of the environmental conditions that affect the accuracy of the measurements,

Each of the techniques has its advantages and disadvantages and their selection is always based on the requirements of the digitisation project. In many CH digitisation projects they are used complementary. For example, in cases where terrestrial laser scanning cannot by applied to capture the top part of a monument, image-based techniques in combination with UAVs can provide an adequate solution in terms of cost and data quality. Additionally, in cases where specific parts of a monument (e.g. an inscription, a column head, a relief, etc.) need to be captured in a higher detail than the rest of the monument, then apart from the terrestrial range scanning method, a short range laser scanning triangulation technique is also used to handle the increased quality requirements. On the other hand, the 3D digitisation pipeline involves a number of procedures that are in most cases common and have to be followed in a sequential way. These are the following:

i) *Digitisation planning,*

ii) *Data acquisition,*

iii) *Editing-cleaning-hole filling of raw data of partial scans*,

iv) *Registration-alignment and merging of partial scans,*

v) *Data simplification according to requirements specification*,

vi) *Texture mapping,*

vii) *Conforming data according to application requirements*

The intensive involvement of the digitisation personnel in some of the previous procedures indicate a potential bottleneck of the given digitisation pipeline. The digitisation planning is such an example. It is a fact that a complete digitisation of a morphologically complex object is a difficult task. Callieri et al.[35] mention that an 80% of the total data is acquired in 20% of the total digitisation time. The remaining 80% of the total digitisation time is used for acquiring the challenging missing data. Nevertheless, a complete 100% sampling is in most cases impossible. In most cases a number of partials scans are produced. The partial scans population is highly affected by the morphological complexity of the object being captured. Occlusions, hollow areas and obstacles are very common when digitising movable or unmovable objects. In addition, procedures such as hole filling and partial scan registration are also considered as time-consuming steps and thus bottlenecks.

A rich arsenal of software tools usually provides the digitisation personnel with algorithmic solutions in order to complete the 3D digitisation pipeline procedures. Such tools are required to be able to handle the large amounts of 3D data and provide efficient interaction with the user, as well as processing stability. Breuckmann provides its experience in data acquisition on-site with a Breuckmann smart*SCAN* ^{3D} scanner, as well as the post-processing of the acquired scan data in Breuckmann´s OPTOCAT software. The smart*SCAN* ^{3D} is a modular scanning system based on SL technology for fast and accurate digitization. The small footprint and the little weight of the scanner allows for quick repositioning of the system for complete coverage of an object. The unique design with three different triangulation angles (10°, 20°, 30°) allows data acquisition with either camera relative to the projector, as well as stereo triangulation of the two cameras. While stereo data with 30° triangulation angle typically produce better quality data, the narrow triangulation angles allow capturing of data, even in otherwise inaccessible areas. The FOV for each scanning project is quickly adjusted on the job. OPTOCAT software supports the whole digitisation process from data acquisition to merging and hole filling. The resulting 3D models serve as the basis for the differential scanning project of PRESIOUS. Post processing of individual scans to a complete model that can be used for analysis is rather time consuming and requires manual input by experts. Intermediate steps of data processing, i.e. single scans before alignment, can also be incorporated in the workflow.

Concluding, the complete 3D digitisation of objects is still far from a "*single button press"* procedure. In fact, when compared with 2D digitisation, it is still at a very primitive level. The previous paragraphs contribute towards this conclusion. In many cases, previous digitisation experiences, such as digitisation plans or even legacy data, can be of great help when new digitisation challenges appear. Through the *on-the-fly auto-completion 3D digitisation* method that will be developed in PRESIOUS,

[35] Callieri, M., Scopigno, R., 2005. 3D scanning, improving completeness, processing speed and visualisation, Proc. Eurographics, Italian Section.

legacy data will be exploited in order to allow gradual shape prediction of partially digitised objects. Such a system will open a new range of possibilities, as well as reducing acquisition times and thus simplify the digitisation procedure. Retrieving similar 3D objects based on partial shape matching provides an instant feedback about the shape and the type artefact being digitised. Such information allows archaeologists and other experts to easily extract information and thus draw conclusions about the artefact or parts of it.

A universal rule about 3D digitisation that is valid despite the method being used has to do with the fact that if a part of the object cannot be *reached* by the acquisition device, then it will not be digitised, resulting in incomplete data. In many cases this can be cured by the auto-completion or synthetic completion, based on strong assumptions derived from the properties of the neighbouring surfaces or other morphological properties such as symmetry, surface colour, etc. PRESIOUS will investigate such situations and will provide means for automatically estimate-predict-complete the missing areas that unburden the user from the time consuming procedures of attempting to manually generate the missing parts.

Furthermore, another fact about 3D digitisation that is true despite the method being used has to do with the digital replica of the object. The digital replica depicts the geometrical and colour condition of the real artefact at a given time spot or period, whereas capturing a large monument currently requires weeks and thus different parts are captured under different conditions. These conditions are in most cases highly dependent on the lighting of the digitisation environment. PRESIOUS will contribute towards the estimation and prediction of degradation by investigating models for forward and inverse erosion. Thus, a simulation and visualisation of the artefact back and forth in time will be possible by allowing the measurement of its dynamic state.

In addition, most of the objects derived from the CH domain require restoration. 3D digitisation is the first step towards an accurate identification of the missing parts. Again, PRESIOUS will contribute on the virtual and physical restoration by proposing methods that synthesise the missing parts. The current project attempts to heal a number of problems introduced by the CH domain and 3D digitisation. The estimation and prediction of monument degradation as well as the object repair are methods that will be developed within the framework of PRESIOUS. Their importance can be strengthened by operating at the level of the actual 3D data, without being affected by the used digitisation method.

Technique	Strength	Weakness
Laser triangulators	Relative simplicity Performance generally independent of ambient light High data acquisition rate	Safety constraint associated with the use of laser source Limited range/measurement volume Missing data in correspondence with occlusions and shadows Cost
Structured light	High data acquisition rate Intermediate measurement volume Performance generally dependent of ambient light	Safety constraints, if laser based Medium computational complexity Missing data in correspondence with occlusions and shadows Cost
Time of flight	Medium to large measurement range Good data acquisition rate Performance generally independent of ambient light	Cost Accuracy is inferior to triangulation at close ranges
Shape from stereo	Simple and inexpensive High accuracy on well-defined targets	Computation demanding Sparse data covering Limited to well defined scenes Low data acquisition rate
Shape from focus	Simple and inexpensive Available sensors for surface inspection and microprofilometry	Limited fields of view Non-uniform spatial resolution Performance affected by ambient light (if passive)
Shape from shadows	Low cost Limited demand for computing power	Low accuracy
Shape from shading	Simple and low cost	Low accuracy

Table 3. Comparison of 3D digitisation techniques

A3 PREVIOUS RELATED PROJECTS

3.1. A3.1 3D COFORM

The EU-funded 3D-COFORM is probably the single most relevant project to PRESIOUS. Its aim is to establish a 3D documentation mechanism for long term documentation of tangible CH, addressing all aspects of 3D-capture, 3D-processing, shape semantics, material properties, metadata, provenance and integration with other sources. 3D-COFORM commenced in 2009 and is ongoing to this date. PRESIOUS naturally follows on from the state-of-art promised and produced by 3D-COFORM and therefore some of the data and tools will be utilized in our project as well.

3D-COFORM addressed the data acquisition of: (i) immovable heritage, (ii) movable, regular objects (in-hand scanning), (iii) movable, regular objects (dome-based acquisition), (iv) moveable, optically complicated objects and v) reflectance. Table 4 summarizes the main components comprising each of these five parts.

Immovable heritage	– Enhanced version of the image-based automatic reconstruction cloud (ARC) 3D web service, amongst others providing an automatic registration of a complete textured model out of registered depth maps and their corresponding images. – Colour-projection tool, allowing the creation of high quality textures for a given 3D mesh from uncalibrated images. – Laser-based 3D geometry and high dynamic range (HDR) texture scanner (Spheron).
Movable, regular objects (in-hand scanning)	– Modified version of a 3D scanner that can be hand-held (Breuckmann). This scanner did not become a product though. – In-hand structured light scanner that allows the digitisation of 3D-objects by manipulating them in front of the scanner (ETHZ). This version will differ from the portable Breuckmann scanner, in that it combines passive and active techniques and, as a result, can deal with large and small objects (Figure 5). – Processing software component, dedicated to on-line automatic registration and merged rendering of the sample data.
Movable, regular objects (dome-based acquisition)	– Transportable mini-dome component, capable of estimating surface reflection properties and reconstructing a 3D model through photometric stereo. – Multi-view dome component, providing a combined photometric multi-view 3D object reconstruction.
Moveable, optically complicated objects	– An extension of the previous multi-view dome set-up, to support new reconstruction techniques, capturing optically complicated objects, i.e. with reflective or transmissive materials.
Reflectance acquisition	– Extension of the ARC 3D reconstruction web service, aiming to derive additional surface reflection information from points seen in different images. – Bidirectional texture function (BTF) sampling and multi-scale synthesis component, allowing the extraction of materials from measured objects and synthesizing BTFs from photographs. – New colour data acquisition component, robust and capable of working in unconstrained environments.

Table 4. Components of 3D COFORM for 3D acquisition

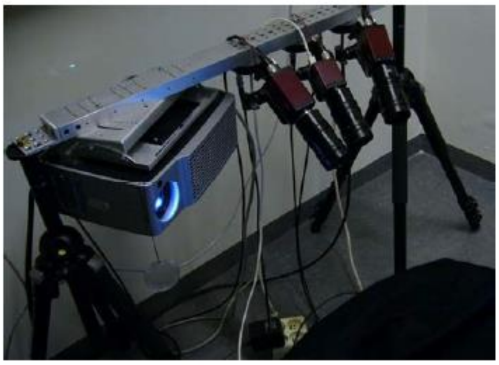


Figure 5. 3D COFORM: the original "in-hand" scanner setup[36].

The main results of 3D COFORM obtained by 2012, as reported in the third annual[36] report can be summarized as follows:

i) Immovable heritage

- Enhanced ARC 3D web service employing an automatic mesh generation algorithm that takes the depth maps as input, cleans them and generates a complete mesh (Figure 6).

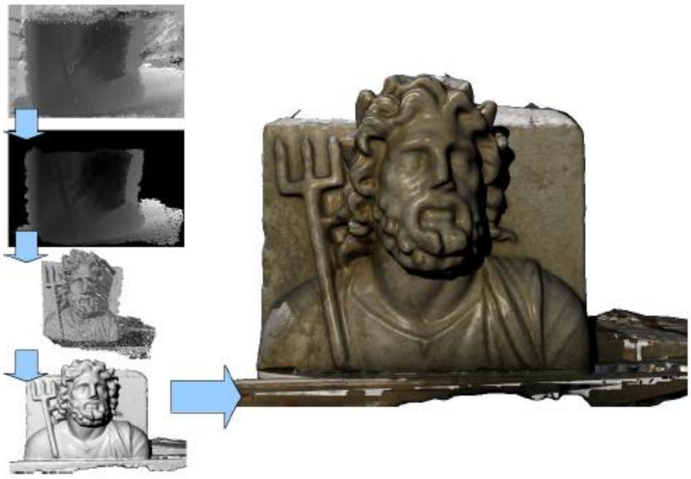


Figure 6. 3D COFORM: automatic mesh generation from ARC 3D depth maps[36]

- A view selection algorithm that selects a set of suitable views for mesh generation, according to coverage and quality, in order to avoid overlap and promote depth maps with higher quality, respectively. Figure 7 illustrates examples obtained with and without applying the view selection algorithm. As can be seen, the result is virtually identical while a significant decrease in computation time is achieved.

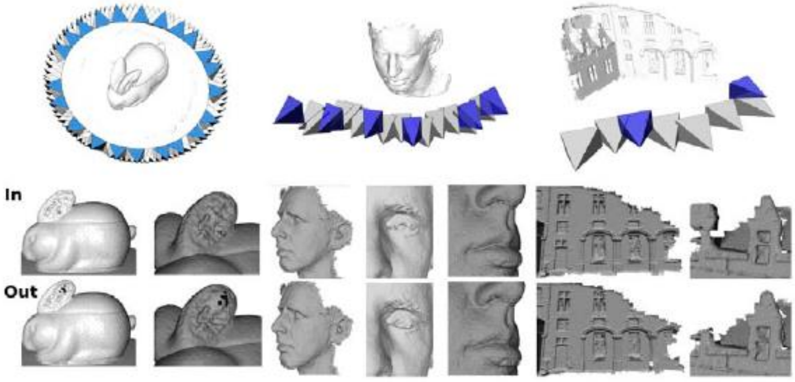


Figure 7. 3D COFORM: results from the view selection algorithm for bunny, face and corner sequence: camera position of input images with selected images highlighted in blue (top), details of meshes produced using all images (in) and the selected images only (out) [36]*.*

[36] 3D COFORM, D.4.3 – Third Year Report, WP4 – 3D Artefact Acquisition, 2012.

- Scenes with low numerical stability due to lack of overlap or feature correspondences, scenes with turntable motion, or scenes with a lot of planarity, can now be reconstructed despite their degenerate characteristics (Figure 8).

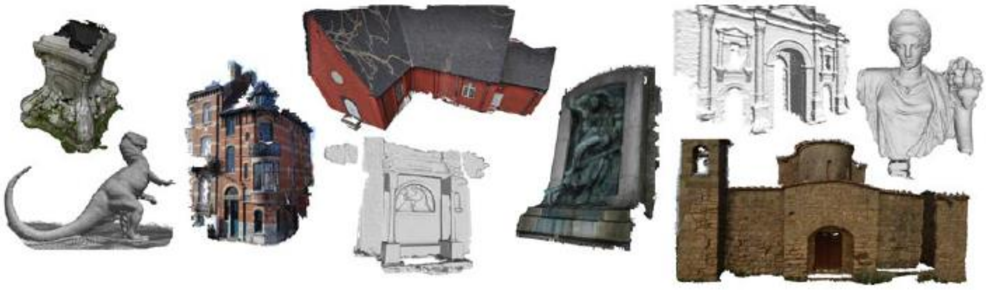


Figure 8. 3D COFORM: A few results from the ARC 3D service on recent examples[36].

- Colour mapping enhancements address the aliasing effect due to small misalignment of the images mapped on the 3D model. This effect is even more visible and critical in the case that image-based solutions, such as ARC 3D, are used to produce the geometric 3D model (Figure 9).

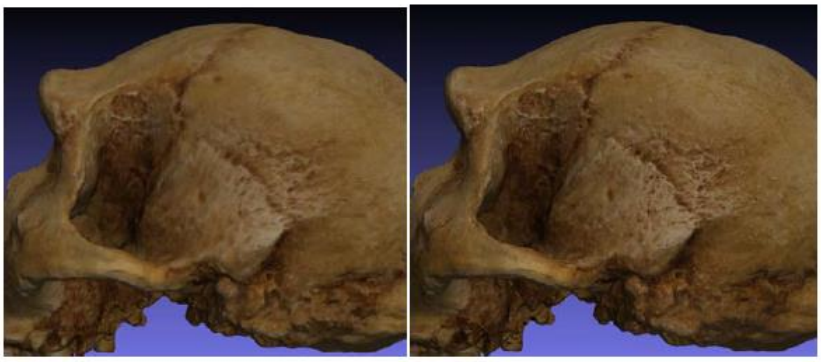


Figure 9. 3D COFORM: A snapshot of a textured model without (left) and with (right) optical flow correction[36].

- Spheron laser-based scanner hardware implementation, manufacturing and testing of the RGB sub-system have been completed.

ii) Movable, regular objects (in-hand scanning)

- Breuckmann hand-held 3D scanner (Figure 10).

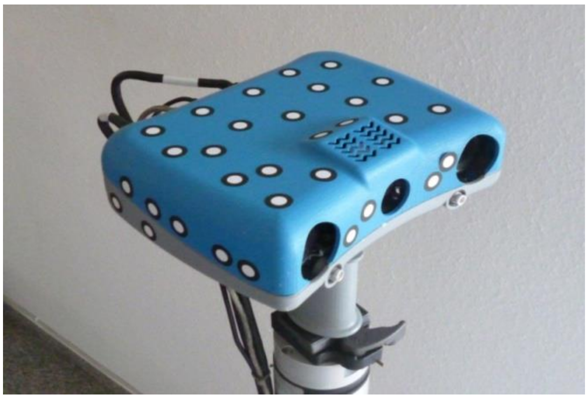


Figure 10. 3D COFORM: Breuckmann hand held 3D scanner.

- The design, setup and test of the demonstrator were realized between March and June 2012.
- In July 2012 a city model of Meersburg was scanned combining the smartSCAN^{3D} C5 system with the hand-held scanner (Figure 11).

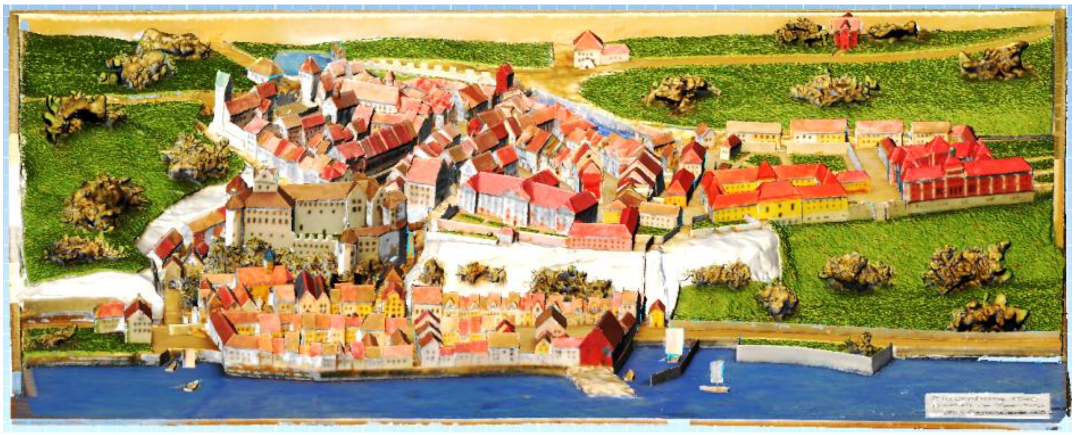


Figure 11. 3D COFORM: city model of Meersburg.

- In-hand phase-shift scanner (ETHZ) (Figure 12) is still under development. This scanning approach is non-portable and uses one projector, two b&w cameras for 3D reconstruction and additional phase detection, as well as one colour camera. The advantage of phase-sift is that it is able to show live 3D and integration of the model. The scanner allows for motion of the object during the scanning process, by properly synchronizing the projector and cameras, as well as using a clever motion compensation process between the subsequently projected patterns.

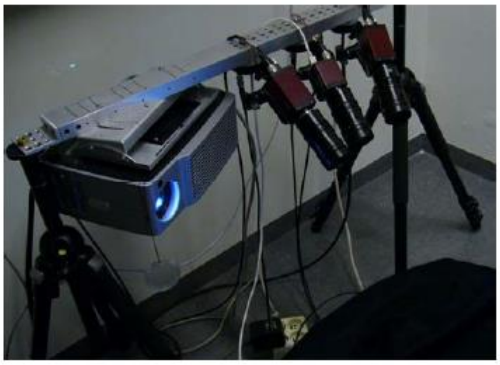


Figure 12. 3D COFORM: The original "in-hand" scanner setup[36].

- In addition, the 3D COFORM consortium is developing a "hybrid" portable approach (Figure 13), in which the object as well as the scanner will be "in-hand". In this "hybrid" scanning approach, phase-shift will be replaced by a structured light, using “one-shot” patterns. "One-shot" refers to the fact that a single projection pattern contains in itself sufficient information to create 3D information. The advantage of such an approach is that it allows for motion, since a single snapshot is sufficient to capture the data and no intricate hardware tricks are needed to create subsequent pattern projections. The “hybrid” system is in use at the moment to create collections of objects. The models illustrated in Figure 14 were acquired with a one-shot approach. The processing, however, is off-line and is still quite supervised. Nevertheless, it shows the potential to arrive at a good result. Efforts are continued to develop automated processing with live feedback.

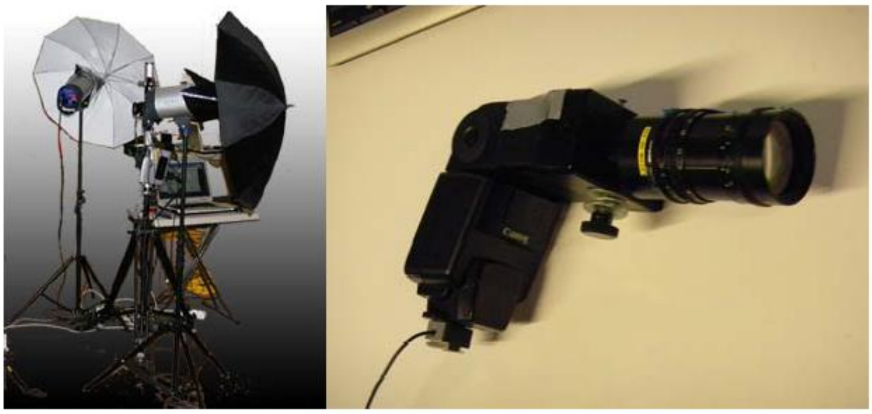


Figure 13. 3D COFORM: The new “hybrid” system. The umbrellas can be attached to the SLR to guarantee diffuse lighting. Right: Flash unit with built-in patterns and lens. When the SLR takes an image the flash fires a pattern on the object[36].

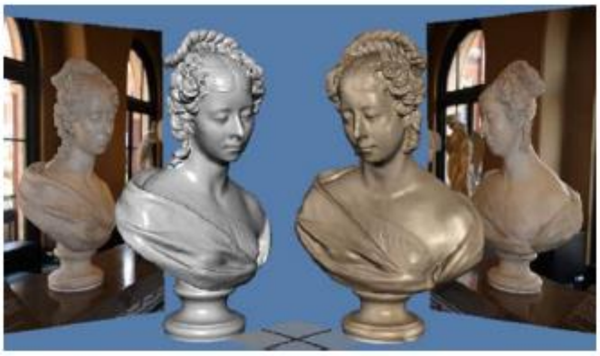


Figure 14. 3D COFORM: only one projected pattern per viewpoint was processed, simulating a one-shot approach[37].

- The output of the ETHZ in-hand scanner at the end of the second year of 3D COFORM was already a coloured 3D model, but according to CNR experience there would be potential for improving the colour encoding quality. The main source of poor quality was the low image resolution, but also the illumination provided by the projector. Hence, the colour projection pipeline of CNR-ISTI was adapted to the in-hand scanner case. Starting from the output provided by the scanner (a 3D model, a video flow and the calibration information for each frame of the video stream), it was possible to adapt the colour projection pipeline to blend the contributions of all frames in a better way. Moreover, knowing for each frame the projector position with respect to the camera and the object, it was possible to correct unwanted effects due to lighting, like shadows and highlights. The colour projection was implemented in an automatic way, showing that with a short post-processing step most of the visual inconsistencies were removed. Figure 15 shows how unwanted effects due to different illuminations, highlights and small misalignments were removed.

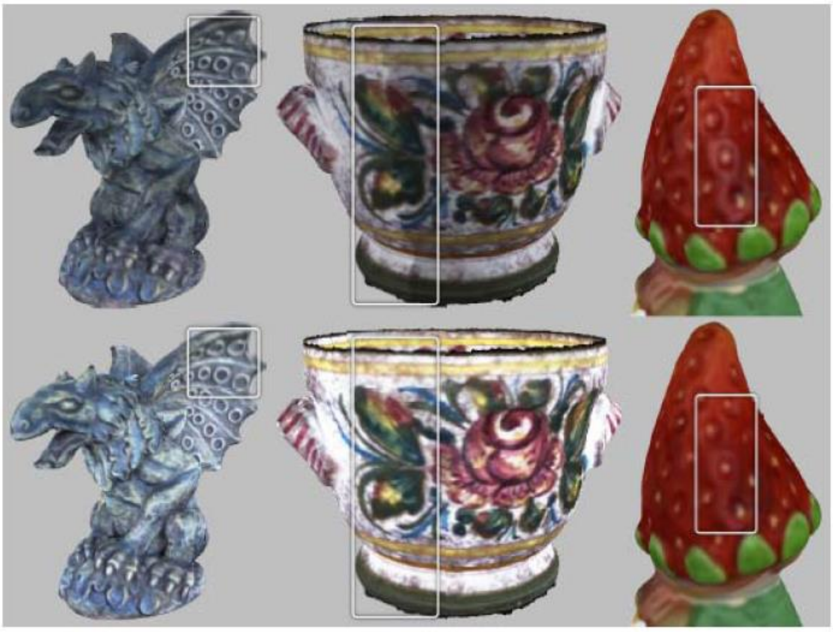


Figure 15. 3D COFORM: Colour projection improvements for in-hand scanners[36].

– Starting from the same input given by the in-hand scanner, a photometric stereo strategy was also applied to retrieve some of the fine geometric details which are usually lost during the scanning. This step has been implemented in a completely automated way. Moreover, strategies to prevent inaccurate estimation of normal vectors are used, when the scanning coverage is not accurate enough to provide data for photometric stereo. Figure 16 demonstrates examples of improvement in geometric quality.

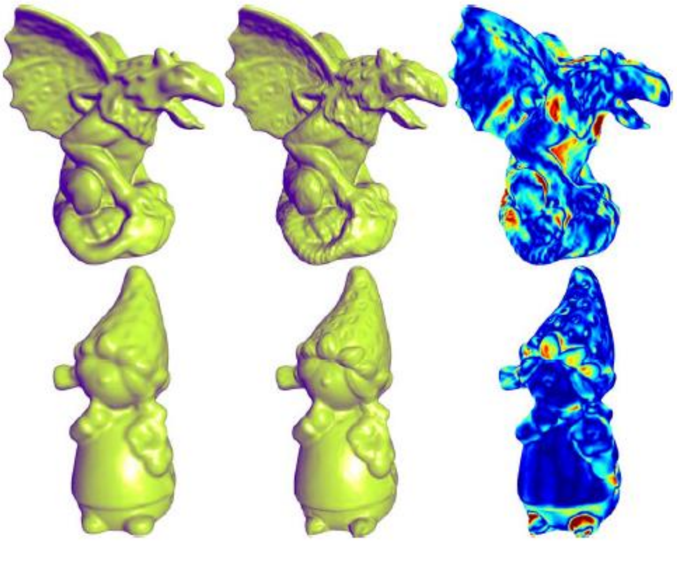


Figure 16. 3D COFORM: Detail enhancement using photometric stereo[36].

iii) Moveable, regular objects: dome-based acquisition

- A large multi-view dome (Figure 17(a)), developed by UBonn and containing multiple cameras and many light sources is targeted to moveable objects and uses both multi-view and SL techniques in order to obtain 3D reconstructions while simultaneously capturing the object's reflectance properties. The KUL mini-dome (Figure 17(b)) contains only one camera and is much smaller, limiting the size of objects that can be scanned and restricting the appearance properties captured to a single view under many different lighting directions. On the other hand, it can be brought to the object rather than the other way around.

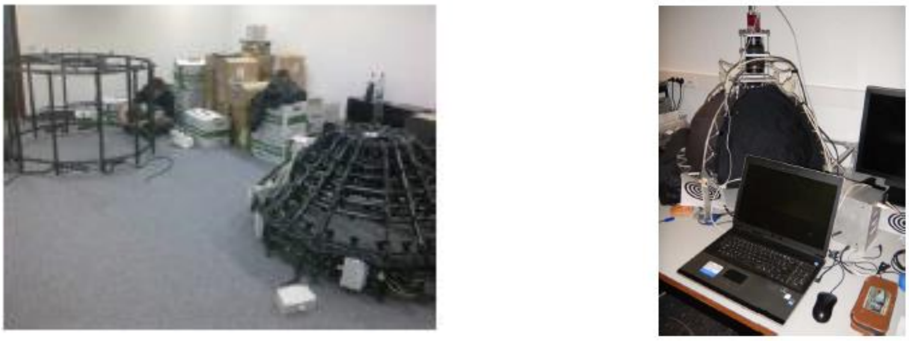

(a) (b)

Figure 17. 3D COFORM: (a) An impression of multi-view dome, (b) mini-dome setup[36]*.*

- The necessary algorithms are in place for the mini-dome to perform a full 3D reconstruction from a single camera view. However, it has been indicated that since the 3D reconstruction is based on integration of derivatives, it is sensitive to local errors in normal computation, which is affected both by the light model used for the LEDs as well as the possibly challenging surface characteristics of the object. This initially resulted in a warping effect on the final 3D model. This was addressed by including the spatially varying intensity of each LED illuminating the scene, given its correct geometric position on the outer sphere. In addition, new hardware has been designed to enforce constant control on LEDs, so as to guarantee enhanced uniformity of their light intensities.

- Several calibration procedures were set in place to compensate for possible errors. In order not to overload the setup of the mini-dome, these procedures have been optimized resulting in minimal setup time when adjusting the camera lens system or objects with different sizes. This comes as an additional advantage of the mini-dome configuration, as it would not be so easy for a multi-view setup. Figure 18 shows an oracle bone during a session at ShangDong, China. Objects varied from 5 to 20 cm.

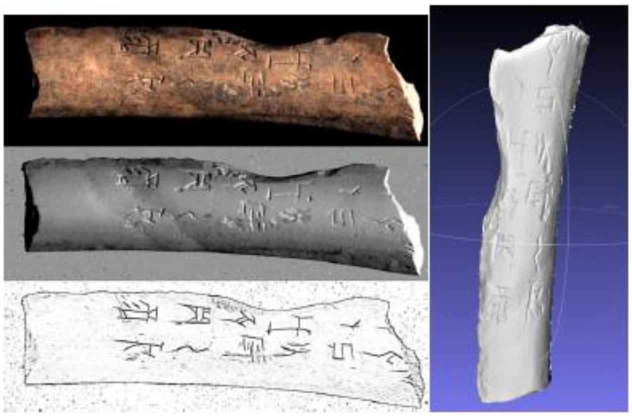


Figure 18. 3D COFORM: Jia Gu: oracle bone, ShangDong China. Left: Picture, normal-map and line-view. Right: 3D model[36]*.*

- One of the planned goals of 3D COFORM consortium with respect to the mini-dome is to try to push analysis on the local surface characteristics on a deeper level, given the current hardware configuration. Experiments and tests that have been carried out suggest that a proper analysis will benefit both the 3D reconstruction as well as the final rendering of the model. The following observations have been made: (i) the 3D reconstruction approach appears especially suited for lambertian or diffuse surfaces. For such surfaces the simple relationship between light source direction and surface normal give consistent results. Furthermore, the mini-dome is quite easy to set up for such objects since the appropriate integration time or aperture can be controlled visually, and (ii) whenever the surface characteristics get more intricate, as is the case with metallic surfaces with reflections or scattering, the standard recording circumstances have shown to fail on calculating the local surface characteristics of certain parts of the objects. Nevertheless, experiments show that there is a range of light sources (LEDs) for which the observed reflected intensities still allow to deduce a proper reflectance model and quite successfully extract proper surface normals. Typically, the proper range of sources give better results when they are recorded using aperture or integration times that would show over or under saturation for other light sources.

  The above observations also justify the use of HDRI recording of the scene. In order to increase the range in light and dark areas, one can control both the aperture or integration time. For practical reasons, the aperture is typically fixed, and the integration time of the camera is varied. At the moment, the recordings are still done separately, but an automated, parameterizable recording is on-going. Exports to .hdr and .exr for HDR data and normal information are, on the other hand, already in place.

  Based on feedback in the field, it has been suggested to provide polynomial texture maps (PTMs) for further rendering or light simulation purposes. It has been demonstrated in offline processing that the original images recorded by the mini-dome provide sufficient information to create such PTMs and simulate the local surface reflectance in a virtual environment under varying light conditions. An implementation is on-going to export the recorded data straight from the software.

  Much related to the above, for rendering purposes it is being investigated if apart from lambertian characteristics other surface reflectance maps can be deduced from the imagery.

- A multi-camera multi-projector super-resolution approach to SL is currently used within the context of 3D COFORM, in order to obtain high quality 3D reconstructions with the multi-view dome. This algorithm is able to achieve very high quality results even on challenging materials[37] (Figure 19). A first evaluation indicates accuracy within the scope of 23μm. Given this accurate geometry, significant enhancements of the quality of the material property estimation could be achieved (Figure 20).

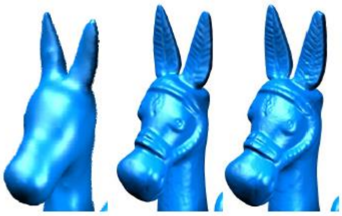


Figure 19. 3D COFORM: quality comparison of reconstruction methods, demonstrated on a metal donkey figurine: visual hull, previously used in multi-view dome (left), NextEngine laser scan (center), novel multi-view dome 3D reconstruction (center)[36].

[37] Weinmann, M., Schwartz, C., Ruiters, R., Klein, R., 2011. A multi-camera, multi-projector super-resolution framework for structured light, Proc. Int. Conf. 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 397-404.

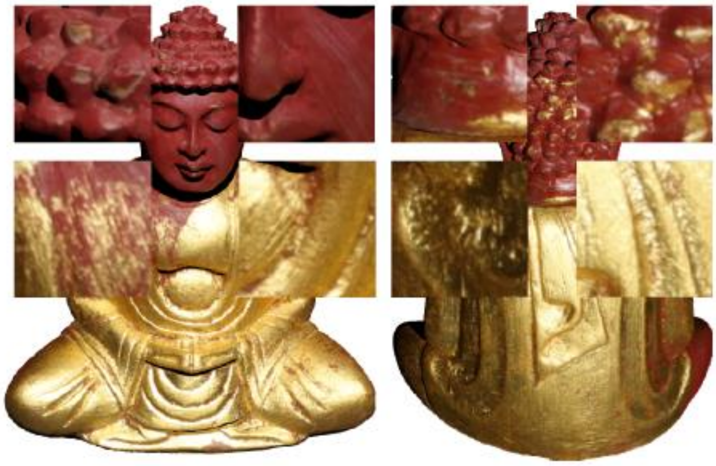


Figure 20. 3D COFORM: achieved quality of an object acquired with the new multi-view dome object acquisition pipeline , employing a 3D reconstruction obtained with the novel structured light method[37] together with BTF reflectance data[36].

iv) Moveable, optically complicated objects

3D COFORM also focused on finding ways that allow the 3D reconstruction of optically complicated artefacts, showing effects such as strong specular highlights, inter-reflection, sub-surface scattering or even transparency. This task of 3D COFORM was based on data of the same hardware as the regular object acquisition with the multi-view dome. Furthermore, this task also addresses the acquisition of spectral reflection data, especially focusing on spectral bidirectional texture functions (BTFs). Since taking spectral images is extremely time consuming and the necessary equipment is quite expensive, UBonn is looking into methods of how to register (sparse) spectral data, acquired using a gonio-reflectometer-like device, with the RGB-BTF-data of the same object, obtained with the multi-view dome and extrapolate the dense spectral information.

Following the initial experimentation, it was decided to extend the existing dome setup by installing projectors. Using these projectors, a SL approach for 3D acquisition was developed and evaluated on several test objects, exhibiting complicated surface reflectance, as well as on fresh foods. Although the latter are not CH objects, they were considered by UBonn to be an especially challenging test-case. Promising results were achieved for these objects[38].

v) Reflectance acquisition

- Research effort was devoted to the design of a new approach to extract information about reflectance properties starting from a very simple input: a short LDR video registered on a 3D model. The approach is based on a statistical analysis of the colour values of each portion of the surface in order to detect, highlight, analyse their shape and, possibly, fit a simple material properties model. Preliminary results are shown in Figure 21, where the different reflectance of portions of the same model is estimated in an accurate way.

[38] Schwartz, C., Weinmann, M., Ruiters, R., Zinke, A., Sarlette, R., Klein, R., 2011. Capturing shape and reflectance of food, ACM SIGGRAPH ASIA, 28:1-28:2.

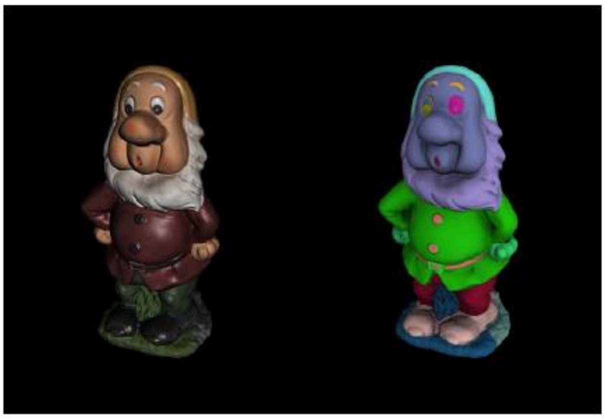


Figure 21. 3D COFORM: first result of reflectance estimation from video flows, a rendering of the object with estimated reflectance (left), the clustering of the object, showing all the surface portions with different material behavior (right) [36].

Discussion on the relevance of 3D COFORM results with PRESIOUS

Overall, 3D-COFORM, has concentrated on high accuracy 3D scanning. Although this is extremely useful in long-term acquisition, it is a very expensive and slow process. When scanning large collections for the purpose of re-assembly and not for digital reproduction, lower accuracy can be tolerated but cheaper and faster scanning is a key requirement for an efficient work pipeline. It is in this area that PRESIOUS will concentrate, using predictive digitisation that can drastically reduce acquisition times and promote the use of low-cost, low accuracy, off-the-shelf technology, such as the Microsoft Kinect-based solutions.

3.2. A3.2 3D ICONS

3D-ICONS is an EU ICT project that builds on the results of the CARARE[39] and the 3D-COFORM projects. It brings together partners from across Europe with the relevant expertise to digitise architectural and archaeological monuments and buildings in 3D and is designed to: establish a complete pipeline for the production of 3D replicas of archaeological monuments and historic buildings which covers all technical, legal and organisational aspects; create 3D models and a range of other materials, including images, texts and videos of a series of internationally important monuments and buildings; and contribute content to Europeana project (see subsection A3.7), using the CARARE aggregation service[40]. It aims to complement the collections which have been made accessible to Europeana by increasing the critical mass of 3D content currently available to Europeana's users. It covers all aspects of a 3D digitisation project from planning and obtaining permissions, selection of methods and tools, data acquisition, post-processing, publication of content online and metadata capturing to make the content available to Europeana.

One of the roles of the Athena Research and Innovation Centre in 3D-ICONS is the evaluation and testing of different digitisation processes. An interim report[41] with five different digitisation case studies varying from movable artefacts up to monuments is currently publicly available. The processing evaluation is focused on the implementation of the digitisation case studies, followed by a critical assessment and analysis of important properties and key factors of each 3D digitisation project. In some cases a number of key 3D digitisation methodologies are applied in each case and their

[39] CARARE 3D and virtual reality content to Europeana, http://www.carare.eu/eng

[40] 3D ICONS, http://3dicons-project.eu/eng/about

[41] Case studies for testing the digitisation process – An interim report, http://3dicons-project.eu/eng/Resources

advantages and disadvantages are discussed in detail. Additionally, 3D-ICONS is focused on the development of the CARARE 2.0 schema that includes additional metadata to support provenance, transformation and London Charter paradata, as well as improved mapping to the Europeana data model (EDM).

3.3. A3.3 Digital michelangelo project (DMP)

DMP was one of the first large scale CH 3D digitisation projects [39]. It was performed by the computer science departments of Stanford and Washington universities along with Cyberware Inc. It was focused on the digitisation of Michelangelo’s statues, including that of David. The technical goal of this project was to make a 3D archive of as many of his statues as they could scan in a year and to make that archive as detailed as current scanning and computer technology would permit. One of the main scopes of the project was to be able to capture Michelangelo’s chisel marks. To be able to achieve that, the digitisation resolution required was estimated around 0.25 mm. For the needs of the project a custom laser triangulation scanner with a 5mW 660-nanometer laser diode, a 512x480 CCD sensor and a motorised gantry was developed. An additional sensor (1520x1144 3-CCD) was also used for capturing colour information. These contribute towards the different and high requirements found in CH 3D digitisation projects. Some of the most important requirements met in DMP were the high resolution sampling (<1mm), a safe distance from large statues and the ability to capture the top of Michelangelo’s David (that is a 7 meters tall statue). In cooperation with the visual information technology lab of the national research council of Canada, the DMP partners analysed the unavoidable effect of subsurface scattering when laser scanning is used on marble. The 1-sigma noise introduced to their data due to subsurface scattering was estimated around 100 μm. Some important statistics regarding the digitisation of David are gathered in the following table.

Height of statue without pedestal	5.17m
Surface area	19m^2
Volume	2.2 m^3
Weight	5,800kg
Number of polygons	2 billion
Colour images producing texture map	7,000
Lossless compressed size	32 GB
Scanning Team Size	22 people
Staffing in the museum	3 people on average
Digitisation Time	360 hours over 30 days
Man-hours scanning	1,080
Man-hours post-processing	>1,500

Table 5 Statistics about Michelangelo's statue of David[42]

Table 5 indicates the need to develop new methods for representing, viewing, aligning, merging and visualising large 3D models.

3.4. A3.4 ViHAP3D

The ViHAP3D (virtual heritage: high quality acquisition and presentation) was an EU IST project that aimed at preserving, presenting, accessing, and promoting CH by means of interactive, high-quality 3D graphics. The project was focused on post-processing, data representation, efficient rendering for

[42] Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J., Fulk, D., 2000. The digital Michelangelo project: 3D scanning of large statues, Proc. ACM SIGGRAPH, 131-144.

detailed interactive displays and inspection of high density models on low cost platforms. Additionally, ViHAP3D focused on virtual heritage tools for the visualisation and navigation of high-quality digital model collections (http://www.vihap3d.org/project.html).

3.5. A3.5 Real 3D

The Real 3D[43] (digital holography for 3D and 4D real-world objects capture, processing and display) is an EU-funded ICT project which marks the beginning of a long-term effort to facilitate the entry of a new technology (digital holography) into the 3D capture and display markets. The project is focused on achieving the world's first fully functional 3D video capture and display paradigm for unrestricted viewing of real-world objects, that employs all real 3D principles. The project's outputs include functional models of four digital holographic 3D capture, processing and display scenarios, encompassing the full 360 degrees of perspectives of reflective macroscopic 3D scenes, microscopic reflective 3D scenes, transmissive or partially transmissive microscopic 3D scenes and capture of 3D scenes at infra-red wavelengths.

3.6. A3.6 PROBADO3D

The PROBADO3D project, funded by the German research foundation, ran between 2006-2011 and developed Digital Library services for 3D architectural model data. The project considered the whole workflow from 3D object indexing, storing, to user access and retrieval. A custom graph-based descriptor for indexing of building models by architectural properties including room connectivity structures was developed. The proposed 3D object retrieval method is based on the bag of visual words (BoVW) paradigm (see subsection B2.3). Figure 22 illustrates a use-case for integrating 3D models into PROBADO3D. An interactive front-end allows users to sketch the layout of a target building, and the most similar search results are visualized to the user for inspection[44].

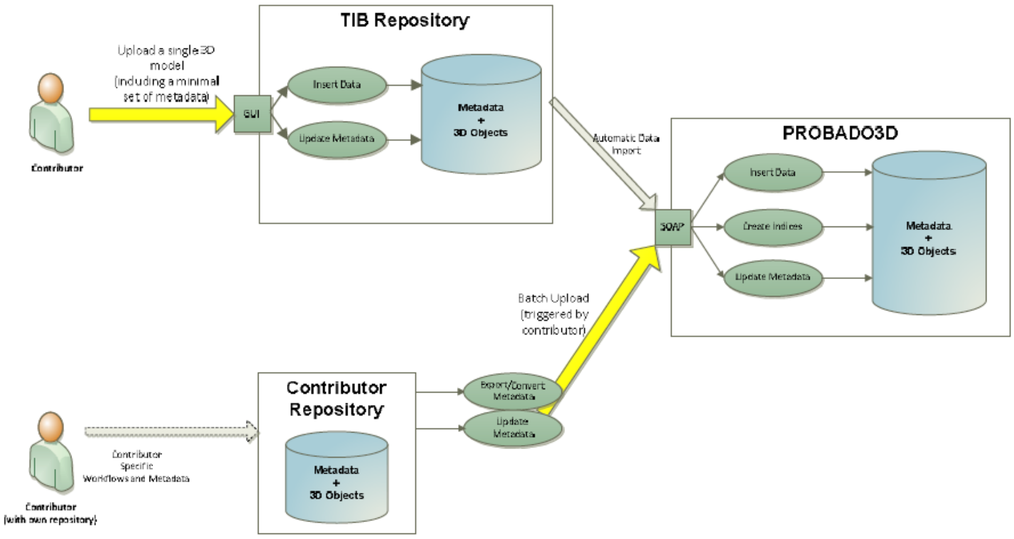


Figure 22. PROBADO3D: Use-case for integrating 3D models into PROBADO3D[45].

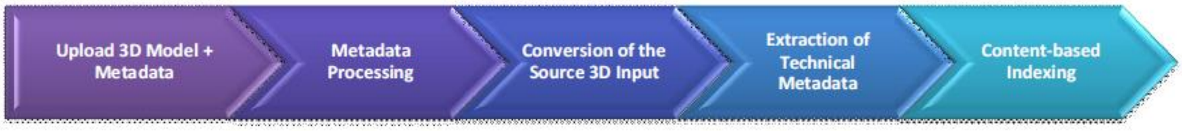
[43] Real 3D, http://www.digitalholography.eu/index.html

[44] Wessel, R., Ochmann, S., Vock, R., Blümel, I., Klein, R., 2011. Efficient retrieval of 3D building models using embeddings of attributed subgraphs, Technical Report, Institut für Informatik II, Universität Bonn, Germany.

[45] Berndt, R., Blümel, I., Wessel, R., 2010. PROBADO3D – Towards an automatic multimedia indexing workflow for architectural objects 3D models, Proc. Int. Conf. Electronic Publishing (ELPUB).

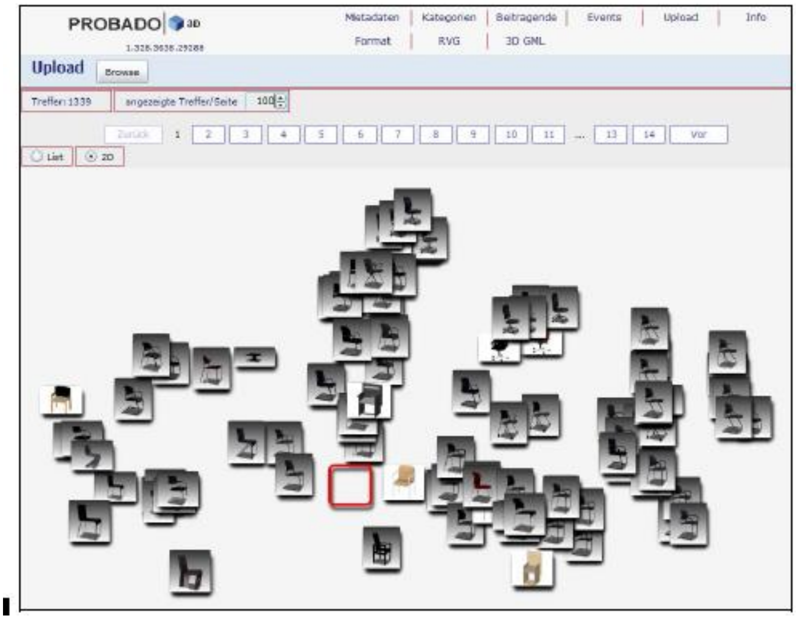
Figure 23 provides a schematic overview of the processing pipeline of PROBADO3D. One major goal of this project is to minimize the manual cataloguing work and to automatically generate the appropriate metadata wherever possible. As a 3D model normally does not bring along any describing data, if it is not catalogued in a database beforehand, the main source for first metadata is the automatic deduction.

A web-based prototype of PROBADO3D was deployed by the Technische Informations bibliothek Hannover, Germany, and is available online for testing. The user interfaces for searching, browsing and representation are based on the rich internet application (RIA) technology. The prototype implements the following query interfaces: searching in the textual metadata of the objects, browsing the repository content using different filters, uploading of a model for a query-by-example search using a 2D interface for constructing query graphs and an interactive 3D modeling environment for formulating 3D queries. In addition to the web-based user interfaces, third party modeling tools like Google™ Sketchup can also be used for accessing the PROBADO3D search services. Different result representations, such as the 2D thumbnail cloud (Figure 24), allow the user to interactively explore the result sets.



Upload 3D Model + Metadata → Metadata Processing → Conversion of the Source 3D Input → Extraction of Technical Metadata → Content-based Indexing

Figure 23. PROBADO3D: Use-case for integrating 3D models into PROBADO3D[45].



Treffer: 1339

angezeigte Treffer/Seite 100

Figure 24. PROBADO3D: Thumbnail cloud–a 2D result visualization of a content based query[46].

Rather than being a pure research project as PRESIOUS, PROBADO3D focuses to achieve long-term usage of the developed systems and workflows at the cooperating libraries. Still, parts of the PROBADO3D model repository, in particular historic building models, may be used as test data for the research in PRESIOUS.

[46] Wessel, R., Blümel, I., Klein, R., A 3D shape benchmark for retrieval and automatic classification of architectural data, 2009. Proc. Eurographics, Workshop 3DOR, 53-56.

3.7. A3.7 Europeana

Europeana is a Web portal that provides access to digital replicas of Europe's cultural heritage thesaurus. It is a multilingual common access point to millions of CH objects. This is achieved by collecting only metadata or contextual information from non-governmental heritage bodies, archaeological museums, specialist digital archives, research institutions and heritage agencies. The quantities of digital resources available and the number of institutions involved in this initiative have been increased due to the activities of the Europeana Office and those of EU funded programmes. Although the Europeana portal currently provides a keyword based search mechanism, through the ASSETS79 programme technologies, such as Linked Open Data, metadata based ranking, query suggestions, image similarity search and semantic cross linking are technologies that are being explored in order to be integrated in the near future and thereby enhance the value of service to the users. From a technical perspective, the similarity between the 3D models will be based on a *view-based*, *low-level* feature extraction algorithm, designed and implemented by CERTH-ITI (www.iti.gr). The indexing is *inverted file-based* and aims to enhance the scalability of the algorithm, as well as the ability to support the indexing of thousands of 3D models. An earlier related work, published by Daras and Axenopoulos, is reviewed in Section B2.1.3 of this survey.

PART B - WHOLE-FROM-PARTIAL SHAPE MATCHING AND RETRIEVAL METHODS

B1 INTRODUCTION

The main outcome of PRESIOUS WP2 will be the development of a predictive 3D digitisation platform, which will be capable of retrieving full 3D models from partial queries, derived from 3D scanning of CH objects. To this end, in Section B2 of this STAR survey we investigate partial 3D object retrieval methods, as well as 3D object retrieval methods which were originally proposed for complete queries but in some aspect they are particularly relevant to partial retrieval. In Section B3, we review local shape descriptors which have been proposed for unstructured data. We bring the spotlight on local shape descriptors taking into account that a partial query and its associated full model are intuitively expected to be similar in a local fashion.

1. B2 PARTIAL 3D OBJECT RETRIEVAL METHODS

PRESIOUS WP2 addresses the search and retrieval of 3D models which are similar to a query, when the available information for the latter is not complete, as it is the case with range scans. The wide availability of range scanners and 3D digitisers, as well as the emergence of next generation technologies in 3D graphics and computational equipment has significantly increased the interest for partial matching algorithms. In this context, two milestone challenges exist: (i) scanned queries can be rough and noisy; (ii) it is not straightforward to effectively match a partial query against a complete 3D model, since there is a gap between their representations. This representation gap complicates the extraction of a signature that will be similar in the case of a complete 3D model, and its partial counterpart which can be introduced as a query object.

Taking into account the existence of a rich literature for partial retrieval of structured data, as well as for 3D object retrieval with complete queries, this section also includes partial retrieval methods that were originally proposed for 3D meshes as well as a few 3D object retrieval methods which were originally proposed for complete queries. These particular methods presented, although seemingly out of the scope of this survey, encompass ideas that are relevant for partial retrieval of unstructured data. More so, if we consider that mesh generation algorithms using unstructured data as input, are embedded into modern SL scanners, as is the case with the scanners of our partner in PRESIOUS. Moreover, among others (subsection B2.4), we explore 3D object retrieval methods, which are: (i) view-based (subsection B2.1), (ii) part-based (subsection B2.2), and (iii) based on the bag of visual words (BoVW) paradigm (subsection B2.3), so as to reveal potential partial retrieval strategies. Finally, in subsection B2.5 we comparatively discuss the 3D object retrieval methods presented, indicating the main research paths to be followed in PRESIOUS WP2 in the context of object retrieval.

B2.1 View-based partial 3D object retrieval methods

View-based 3D object retrieval methods are particularly relevant to the context of partial retrieval, since an object view is closely associated with a range scan, the primary form of partial 3D object input. Accordingly, even in the case of view-based methods originally formulated for complete 3D object queries, which are used to derive object views carrying full object information, it is reasonable to assume that a modification which uses a subset of views, carrying partial information, in order to conform to the objectives of PRESIOUS WP2, is not a quantum leap. A generic scheme which can be considered in this respect, may follow the idea of Daras' and Axenopoulos' method (subsection B2.1.3), which has been proposed for both complete and partial retrieval. In the first case, a similarity metric derived from the sum of distances between all possible query-target view pairs is used, whereas in the latter case similarity is determined by the minimum of such distances.

B2.1.1 Bayesian-based 3D object retrieval based on adaptive views clustering

Ansary et al.[47] introduced a 3D object retrieval method based on the similarity of characteristic views, called adaptive views clustering (AVC). Taking into account that as the geometric complexity of a 3D model increases, its 2D views tend to differ, the authors proposed a characteristic views selection algorithm that relates the number of views to its geometrical complexity. Starting from 320 initial views, their algorithm selects the “optimal” characteristic views set that best represents the 3D model (Figure 25). The number of characteristic views varies from 1 to 40. They also proposed a new probabilistic retrieval approach that takes into account that not all the views of 3D models have the same importance, and also the fact that geometrically simple models are more probable to be relevant than more complex ones. Model selection is performed by means of Bayesian information criteria.

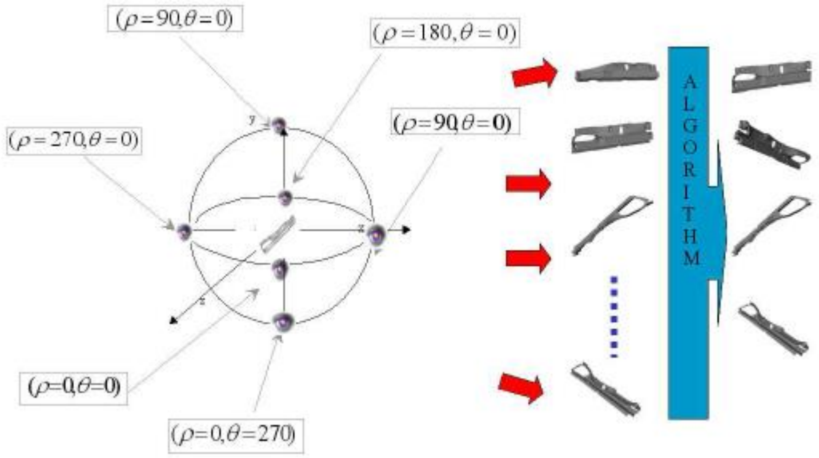


Figure 25. The views selection process[47].

The authors compared their method to twelve state of the art methods, including those of Kazhdan et al.[48] and Chen et al.[49], on the PSB database[50]. The AVC method obtains the second best results, reaching a satisfactory compromise of quality/cost, as compared to the competition. Additional results of the AVC method on a large 3D CAD models database supplied by Renault, showed that their method can be suitable for 3D CAD models retrieval. Moreover, it is shown to be robust against noise and model degeneracy and suitable against topologically ill-defined 3D models.

[47] Ansary, T.F., Daoudi, M., Vandeborre, J.P., 2007. A Bayesian 3D search engine using adaptive views clustering, IEEE Tran. Multimedia, 9 (1), 78–88.

[48] Kazhdan, M., Funkhouser, T., Rusinkiewicz, S., 2003. Rotation invariant spherical harmonic representation of 3D shape descriptors. Proc. Eurographics/ACM SIGGRAPH Symp. on Geom. Process., 156–164.

[49] Chen, D.-Y., Ouhyoung, M., Tian, X.-P., Shen, Y.-T., Ouhyoung, M., 2003. On visual similarity based 3d model retrieval. Proc. Eurographics, workshop 3DOR, 223–232.

[50] Shilane, P., Min, P., Kazhdan, M., Funkhouser, T., 2004. The Princeton shape benchmark. Shape Modeling International.

B2.1.2 View-based 3D object retrieval based on the elevation descriptor

Shih et al.[51] introduced a view-based 3D object retrieval method based on *elevation descriptor*, a feature defined to achieve translation and scaling invariance, as well as rotation robustness. First, a 3D model is represented by six gray-level images which describe the altitude information of a 3D model from six different views including front, left, right, rear, top and bottom (Figure 26). Each gray-level image, called an elevation, is decomposed into several concentric circles (Figure 27). The sum of the altitude information within each concentric circle is then calculated. To be less sensitive to rotations, the elevation descriptor is obtained by taking the difference between the altitude sums of two successive concentric circles. Since there are six elevations for each 3D model, an efficient similarity matching method is provided to find the best match for an input model.

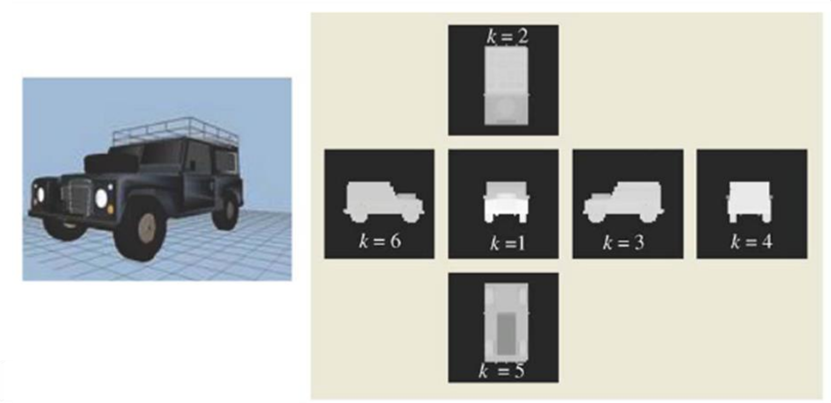


Figure 26. A 3D model of a jeep and its six elevations including front (k = 1), top (k = 2), right (k = 3), rear (k = 4), bottom (k = 5), and left (k = 6) elevations[51].

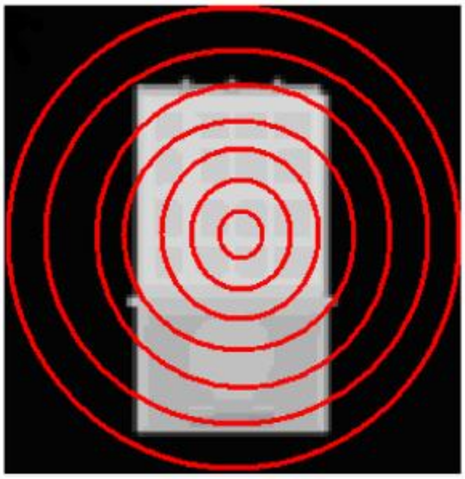


Figure 27. Top elevation of the 3D jeep model shown in Figure 26 partitioned by several concentric circles[51].

The experimental results obtained on the PSB database, as well as on a database of 3D models collected by the authors, show that for most types of 3D models, the method of Shih et al. outperforms

[51] Shih, J.L., Lee, C.H., Wang, J.T., 2007. A new 3D model retrieval approach based on the elevation descriptor, Pattern Recognition 40, 283–295.

state-of-the-art methods which include the search engine of Funkhouser et al.[52] and the method of Osada et al.[53].

B2.1.3 Hybrid view-based 3D object retrieval

Daras and Axenopoulos[54] [55] proposed a view-based 3D object retrieval method using feature vectors comprised polar Fourier coefficient, Zernike and Krawtchouk moments. These features are calculated either on binary images or on depth images extracted from 18 2D views, which are taken from the vertices of a bounding 32-hedron. A pose estimation step based on the work of Daras et al.[56] or Pu et al.[57] precedes view extraction. Their method is summarized in Figure 28.

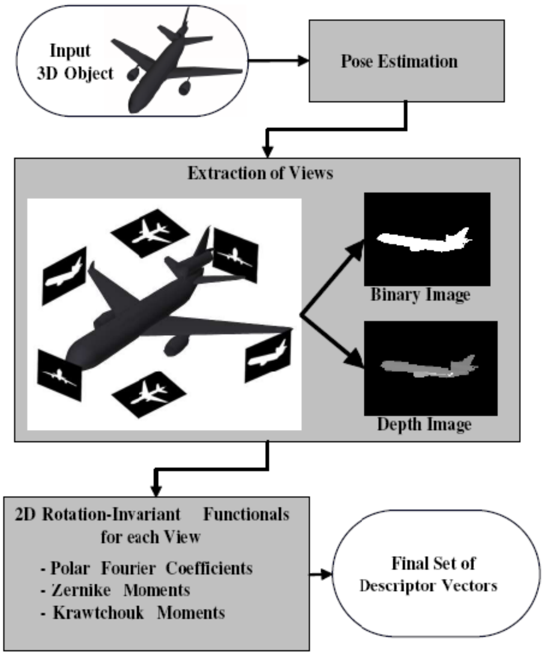


Figure 28: Block diagram of the method of Daras and Axenopoulos[54].

Daras and Axenopoulos' method can be applied for 3D object retrieval with complete or partial queries. In the first case, the utilized dissimilarity measure sums the distances of associated 2D views (Figure 29), whereas in the case of partial retrieval, the minimum distance between the query model and each 2D view is used. Experiments are performed on three databases: (i) a database of 544 3D models classified into 13 classes, as compiled by the Informatics and Telematics Institute (ITI), (ii) the

[52] Funkhouser, T., Min, P., Kazhdan, M., Chen, J., Halderman, A.,Dobkin, D., Jacobs, D., 2003. A search engine for 3D models, ACM Trans. Graphics 22 (1), 83–105.

[53] Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D., 2002. Shape distributions. ACM TOGS 21 (4), 807-832.

[54] Daras, P., Axenopoulos, A., 2009a. A compact multi-view descriptor for 3D object retrieval, Proc. CBMI, 115-119.

[55] Daras, P., Axenopoulos, A., 2009b. A 3D shape retrieval framework supporting multimodal queries, Int. J. Comp. Vis. 89, 229-247.

[56] Daras, P., Zarpalas, D., Tzovaras, D., Strintzis, M.G., 2006. Efficient 3-d model search and retrieval using generalized 3-d radon transforms, IEEE Trans. Multimedia 8 (1), 101-114.

[57] Pu, J., Ramani, K., 2005. An approach to drawing-like view generation from 3D models, Proc. IDETC/CIE, ASME.

engineering shape benchmark (ESB)[58] which contains 867 3D CAD models from the mechanical engineering domain, classified into 44 classes and (iii) the well-known PSB. Daras and Axenopoulos' method outperforms the methods of Chen et al.[49], Vranic[59], as well as the BoVW-based method of Ohbuchi et al.[60]. In the more complete journal version of their work, Daras and Axenopoulos[55] combined their feature vector with the spherical trace transform of Zarpalas et al.[61], resulting in enhanced retrieval performance. Finally, this method has participated in three tracks of SHREC'09, namely the tracks for: (i) structural shape retrieval[62], (ii) a new generic shape benchmark[63] and (iii) partial 3D retrieval[64].

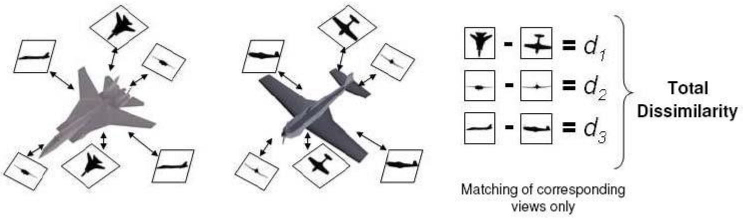


Figure 29: In the case of complete queries, the total dissimilarity between two 3D objects is the sum of the dissimilarities of the corresponding views[54].

B2.1.4 Panorama

Papadakis et al.[65] introduced PANORAMA, a 3D object retrieval method which uses Fourier and Wavelet coefficients calculated over cylindrical projections of each model. As a first step they perform rotation normalization using normal principal component analysis (NPCA) and continuous principal component analysis (CPCA)[66]. Cylindrical projections capturing surface position and orientation are extracted from both CPCA and NPCA-normalized models (Figure 30). Aiming to capture model surface position, the distances of the cylinder center to the intersections of model surface with rays

[58] Jayanti, S., Kalyanaraman, Y., Iyer, N., Ramani, K., 2006. Developing an engineering shape benchmark for cad models, Computer-Aided Design 38, 939–953.

[59] Vranic D., 2004. 3D model retrieval. PhD thesis, University of Leipzig, Germany.

[60] Ohbuchi, R., Osada, K., Furuya, T., Banno, T., 2008. Salient local visual features for shape-based 3D model retrieval, Proc. IEEE SMI, 93-102.

[61] Zarpalas, D., Daras, P., Axenopoulos, A., Tzovaras, D., Strintzis, M.G., 2007. 3D model search and retrieval using the spherical trace transform. EURASIP Journal on Advances in Signal Processing 2007.

[62] Hartveldt, J., Spagnuolo, M., Axenopoulos, A., Biasotti, S., Daras, P., Dutagaci, H., Furuya, T., Godil, A., Li, X., Mademlis, A., Marini, S., Napoleon, T., Ohbuchi, R., & Tezuka, M., 2009. SHREC 2009 track: structural shape retrieval on watertight models. Proc. Eurographics, workshop 3DOR.

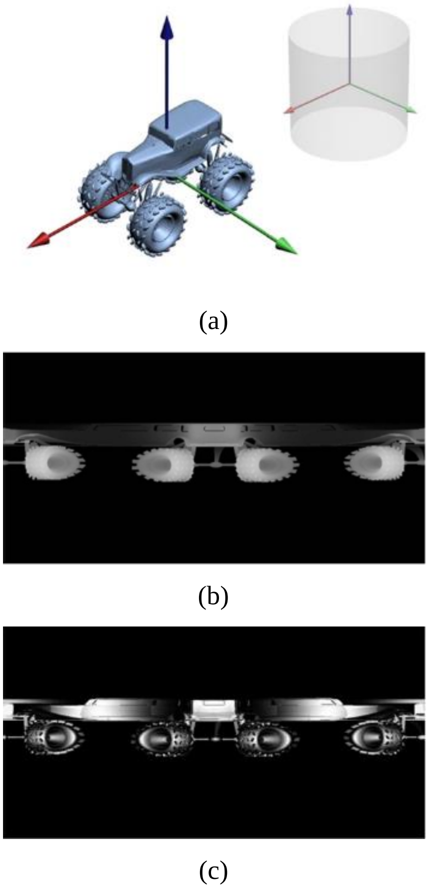
[63] Akgul, C., Axenopoulos, A., Bustos, B., Chaouch, M., Daras, P., Dutagaci, H., Furuya, T., Godil, A., Kreft, S., Lian, Z., Napoleon, T., Mademlis, A., Ohbuchi, R., Rosin, P. L., Sankur, B., Schreck, T., Sun, X., Tezuka, M., Yemez, Y., Verroust-Blondet, A., Walter, M., 2009. SHREC 2009—generic shape retrieval contest. Proc. Eurographics, workshop 3DOR.

[64] Axenopoulos, A., Daras, P., Dutagaci, H., Furuya, T., Godil, A., Ohbuchi, R., 2009. SHREC 2009—shape retrieval contest of partial 3D models. Proc. Eurographics, workshop 3DOR.

[65] Papadakis, P., Pratikakis, I., Theoharis, T., Perantonis, S., 2010. PANORAMA: A 3D shape descriptor based on panoramic views for unsupervised 3D object retrieval. Int. J. Comp. Vis. 89 (2-3), 177-192.

[66] Papadakis, P., Pratikakis, I., Perantonis, S., Theoharis, T., 2007. Efficient 3D shape matching and retrieval using a concrete radicalized spherical projection representation, Pattern Recognition 40 (9), 2437–2452.

emanating from the cylinder center are calculated. In addition, the orientation of model surface is captured by calculating the angle between the ray and the normal of the intersected surface triangle.



(a)

(b)

(c)

Figure 30: (a) Pose normalized 3D model, (b) the unfolded cylindrical projection of the 3D model capturing the position of the surface, (c) the unfolded cylindrical projection of the 3D model capturing the orientation of the surface[65].

Fourier and Wavelet coefficients are calculated over the resulting cylindrical projections of position and orientation. In the former case, only the Fourier coefficients that lie in the areas of the four corners of the 2D Fourier spectrum are maintained, since they carry most of the signal energy and cover approximately 35% of the total number of coefficients. In the latter case, standard statistical features of the Wavelet coefficients, including mean, standard deviation and skewness are derived. The dissimilarity measure used for the comparison of two 3D models which are aligned with the same alignment method (NPCA or CPCA) is the sum of the Manhattan distance of the Fourier part and the Canberra distance[67] of the Wavelet part of the feature vector. The final distance between two PANORAMA descriptors is the minimum of these two dissimilarities. Interestingly, Papadakis et al. incorporate an optional local relevance feedback (LRF) component in their retrieval strategy.

For their experiments, Papadakis et al. use several benchmark databases: (i) the classified dataset of CCCC[59], (ii) the dataset of the National Institute of Standards and Technology (NIST)[68], (iii) the

[67] Kokare, M., Chatterji, B., Biswas, P., 2003. Comparison of similarity metrics for texture image retrieval. Proc. TENCON Conf. Convergent Technologies for Asia-Pacific Region 2, 571–575.

[68] Fang, R., Godill, A., Li, X., Wagan, A., 2008. A new shape benchmark for 3D object retrieval. Proc. Int. Symp. Advances in Visual Computing, 381–392.

“watertight” track of SHREC’07[69], (iv) the MPEG-7 dataset[70], v) the PSB[50] and (vi) the ESB[58]. PANORAMA results in retrieval accuracy which is higher than the one obtained by the hybrid 2D-3D method of Papadakis et al.[71], the DESIRE descriptor of Vranic[72], the light field descriptor (LFD) of Chen et al.[49] and the Euclidean distance transform descriptor of Kazhdan et al.[48]. Moreover, it has been shown that LRF enhances the retrieval accuracy obtained in most benchmark databases.

B2.1.5 3D search and retrieval from range images using salient features

Stavropoulos et al.[73] introduced a method for identifying the correspondence between a range image and a full 3D model by searching for the camera viewpoint, orientation, scale and internal geometry that would generate an image similar to the query, as illustrated in Figure 31. Instead of attempting to match the entire image, only spatial distributions of salient points are compared. The salient points are extracted following the theory of salience, as introduced by Hoffman and Singh[74]. A coarse-to-fine, hierarchical approach is adopted for searching in the camera parameter space in order to bypass exhaustive searching.

The framework of Stavropoulos et al. is experimentally tested on the dataset used in the “watertight” track of SHREC’07[69], as well as in the PSB. The obtained retrieval results show that it outperforms the method of Germann et al.[75], including cases of noise-infused or occluded 3D models. Moreover the time dedicated for off-line preprocessing and on-line partial matching is only 1 sec and 20 msec respectively, for standard Intel-based workstations.

[69] Giorgi, D., Biasotti, S., Paraboschi, L., 2007. SHREC 2007 - shape retrieval contest: watertight models track, Proc. Eurographics, workshop 3DOR.

[70] MPEG-7, http://www.chiariglione.org/mpeg/

[71] Papadakis, P., Pratikakis, I., Theoharis, T., Passalis, G., Perantonis, S., 2008. 3D object retrieval using an efficient and compact hybrid shape descriptor. Proc. Eurographics Workshop 3DOR, 9-16.

[72] Vranic, D., 2005. Desire: a composite 3D-shape descriptor. Proc. IEEE Int. Conf. Multimedia and Expo.

[73] Stavropoulos G., Moschonas P., Moustakas K., Tzovaras D., Strintzis M.G., 2010. 3D model search and retrieval from range images using salient features, IEEE Trans. Multimedia 12 (7), 692-704.

[74] Hoffman, D.D., Singh, M., 1997. Salience of visual parts, Cognition 63 (1), 29-78.

[75] Germann, M., Breirenstein, M. D., Park, I.K., Pfister, H., 2007. Automatic pose estimation for range images on the GPU. Proc. 3D Digital Imaging and Modeling, 81–90.

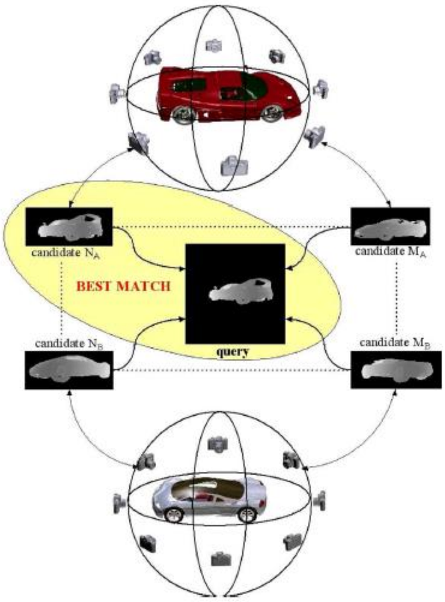


Figure 31. The algorithm of Stavropoulos et al. searches for the best match in parameter space that consists of all possible positions and orientations of the camera [73].

3.8. B2.2 Part-based partial 3D object retrieval methods

Biederman[76] suggested that humans tend to recognize objects by analyzing the semantics of their parts. This suggestion leads to the part-based paradigm in 3D object retrieval, which is based on the hypothesis that two objects are similar, if they consist of similar parts. The relevance of this approach to partial retrieval is obvious, if we consider that a part of a 3D object is actually the input of a partial retrieval method.

B2.2.1 Local surface patches (LSPs)

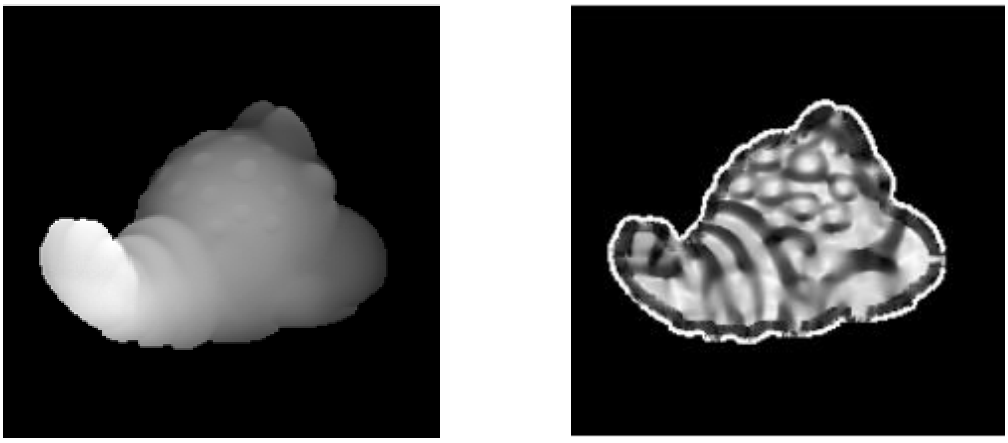
Chen and Bhanu[77] introduced LSPs for 3D object representation. Their method starts from extracting feature points in range images and defining LSP descriptors[78] for each feature point with large shape variation, as measured by the *shape index*[79] (Figure 32).

[76] Biederman, I., 1987. Recognition-by-components: a theory of human image understanding, Psychological Review 94 (2), 115-147.

[77] Chen, H., Bhanu, B., 2007. 3D free-form object recognition in range images using local surface patches. Pattern Recognition Letters 28 (10), 1252 – 1262.

[78] Chen, H., Bhanu, B., 2004. 3D free-form object recognition in range images using local surface patches. Proc. ICPR 3, 136–139.

[79] Dorai, C., Jain, A., 1997. COSMOS—A representation scheme for 3D free-form objects. IEEE Trans. Pattern Anal. Mach. Intell. 19 (10), 1115–1130.



(a) (b)

Figure 32. (a) A range image and (b) the image of its shape index. In (a), the darker pixels are away from the camera and the lighter ones are closer. In (b), the darker pixels correspond to concave surfaces and the lighter ones correspond to convex surfaces[77].

For each patch, Chen and Bhanu calculate local surface properties, which include a 2D histogram, the surface type and the centroid. The 2D histogram (Figure 33) consists of shape indices and angles between the normal of the feature point and that of its neighbors. The surface of a patch is classified into different types based on the mean and Gaussian curvatures of the feature point. For every LSP, the mean and standard deviation of shape indices are computed and used as indices to a hash table (Figure 34). Potential associations between LSPs and candidate models are hypothesized by comparing LSPs of a query and LSPs of a full 3D model, followed by casting votes for those models containing similar surface descriptors. Finally, a rigid transformation is estimated based on the corresponding LSPs, so as to enable the calculation of the match quality between the hypothesized 3D model and the query.

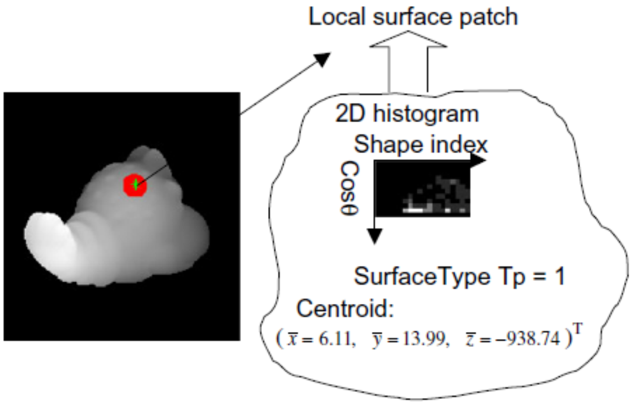


Figure 33. Illustration of a local surface patch (LSP). Feature point P is indicated in green and its neighbors N are indicated in red[77].

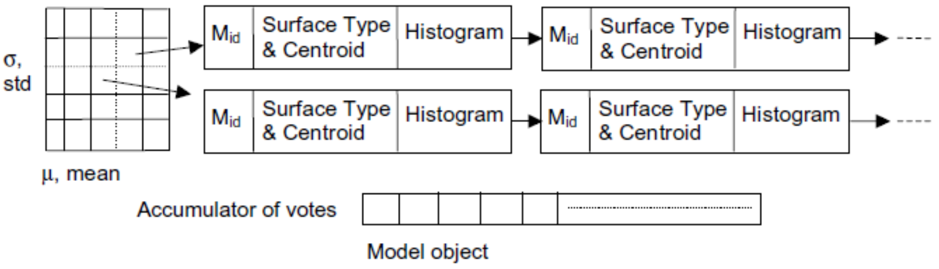


Figure 34. Structure of the hash table. Every entry in the hash table has a linked list which saves information about the model LSPs and the accumulator records the number of votes that each model receives[77].

Experiments were performed on a database of real range data, collected by Ohio State University. Comparisons with the spin-image and spherical spin-image (see subection 4.1.1) representations on real range data have shown that LSPs are as effective for the matching of 3D objects as these two representations, but are *more efficient* in finding corresponding parts between a model-query pair.

B2.2.2 MPEG-7 standard: perceptual 3D shape descriptor

Kim et al.[80] presented a method that uses convexity and graph representation for 2D and 3D object segmentation. The authors describe three main steps: (i) initial decomposition stage, (ii) the recursive decomposition stage, and (iii) the iterative merging stage. First the opening operation of mathematical morphology is performed on the binary image representing the object. This operation results in a decomposition of the object by rounding its corners. Multiple decompositions are produced by alternating the radius of the ball shaped structuring element, by which the opening is calculated. The best segmentation is selected as the one with the highest weighted convexity value, which is the sum of the convexity value of all parts weighted by their normalized volume. After the initial segmentation, each part is hierarchically decomposed with the same method until a split condition is no longer met. Following the hierarchical segmentation step, the merging stage checks for over-segmented parts. This is achieved by calculating the difference between the convexity of the merged part and the weighted convexity of the constituent parts.

The perceptual 3D shape descriptor has been defined by Kim et al.[81] [82]. An attributed relational graph (ARG) representing object features is constructed. Each part associated with a graph node is represented by an ellipsoidal blob. There are four unary attributes used to describe each node:

– the volume V of the segment,

– the convexity C of the segment defined as the ratio of the volume of the object to the volume of its convex hull,

– two eccentricity values $E_1 = \sqrt{1 - c^2/a^2}$ and $E_1 = \sqrt{1 - c^2/b^2}$, where a, b and c are the three maximum ranges along the three principal axes.

In addition, there are three binary attributes describing edges or relations between nodes:

[80] Kim, D.H., Yun, I.D., Lee, S.U., 2005. A new shape decomposition scheme for graph-based representation, Pattern Recognition 38 (5), 673-689.

[81] Kim, D.H., Park, I., Yun, I., Lee, S., 2005. A new MPEG-7 standard: perceptual 3-D shape descriptor, Lecture Notes in Computer Science 3332, 238-245.

[82] Kim, D.H., Park, I.K., Yun, I.D., Lee, S.U., 2004. A new MPEG-7 standard: perceptual 3D shape descriptor. Proc. Pacific Rim Conference on Multimedia, 238–245.

- distances between centers of ellipsoid segments,
- the angle between the first principal axes of two adjacent segments,
- the angle between the second principal axes of two adjacent segments

The comparison between two graphs is performed using the double earth mover's distance (EMD).

The proposed decomposition method is tested on several synthetic 3D objects, resulting in plausible mesh segmentation. The authors suggested that their decomposition method can be integrated in a 3D object retrieval context.

B2.2.3 Retrieval of 3D articulated objects using a graph-based representation

Agathos et al.[83] proposed a graph-based representation method that decomposes objects using the mesh segmentation method previously introduced by some of the authors[84]. Geodesic extrema of an object are considered as salient points. To this end, the integral geodesic function is used. The core partition is approximated by starting from the minimum of the geodesic function and expanding the partition. When expansion is completed, the protrusion parts are separated by the core. Boundaries are refined with a minimum cut algorithm to form the final segmentation.

After the segmentation step, each segment of the object is represented as a graph node and adjacent segments are connected in the graph with an edge. Unary and pairwise features are assigned to each node and edge, respectively. The graph matching is based on EMD of the feature vectors. Unary attributes assigned to the nodes include size, convexity, eccentricities of the ellipsoid approximating the component[84] and the spherical harmonic descriptor vector[85]. The binary attributes assigned to graph edges are the distance of the segment centroids and the angles that the two most significant principal axes of the connected components form with each other. Before the matching of two graphs, penalty nodes are inserted in the graph with the smaller number of nodes (equal to their difference of cardinality).

Experiments are performed on the McGill shape benchmark (MSB)[86] database, which consists of highly articulated shapes, and ISDB[87] database, showing that the retrieval method of Agathos et. al outperforms the part-based method of Kim et al.[82] and the hybrid descriptor of Papadakis et al.[71] in terms of retrieval accuracy.

B2.2.4 Non rigid 3D object retrieval using topological information guided by conformal factors

Sfikas et al.[88] introduced an object decomposition method using the conformal factor[89] values of the mesh. A clustering of the mesh faces is performed based on the discrete conformal factor values on

[83] Agathos, A., Pratikakis, I., Papadakis, P., Perantonis, S., Azariadis, P., Sapidis, S., 2010. 3D articulated object retrieval using a graph-based representation, The Visual Computer 26 (10), 1301-1319.

[84] Agathos, A., Pratikakis, I., Perantonis, S., Sapidis, N., 2009. Protrusion-oriented 3D mesh segmentation, The Visual Computer 26(1), 63-81.

[85] Papadakis, P., Pratikakis, I., Perantonis, S., Theoharis, T., 2007. Efficient 3D shape matching and retrieval using a concrete radicalized spherical projection representation, Pattern Recognition 40 (9), 2437-2452.

[86] McGill 3D Shape Benchmark, http://www.cim.mcgill.ca/~shape/benchMark/

[87] Gal, R., Shamir, A., Cohen-Or, D., 2007. Pose oblivious shape signature, IEEE Trans. Vis. Comput. Graph. 13 (2), 261–271.

[88] Sfikas, K., Theoharis, T., Pratikakis, I., 2012. Non-rigid 3D object retrieval using topological information guided by conformal factors, The Visual Computer 28 (9), 943-955.

[89] Ben Chen, M., Gotsman, C., 2008. Characterizing shape using conformal factors, Proc. Eurographics 3DOR, 1-8.

each vertex, which results in a decomposition of the mesh into parts. The clustering leads to the construction of an attributed graph with the following features:

- mean discrete conformal factor value of the segment,
- normalized area of the segment,
- geodesic length between borders of the segment

Graph construction is followed by graph matching. The matching first locates the core node of each graph. All other nodes are used to form strings ending at the core node. The idea is to perform a matching between strings. Consequently, a distance between two strings is defined as the sum of the distances of the corresponding nodes. The distance between two objects is the minimum distance of a correspondence between strings of the two different objects. Let m and n denote the cardinality of strings of two objects. The assignment problem is solved with the Hungarian Algorithm in the $m'n$ matrix representing the distances between all strings of two objects. The matrix becomes square by padding with a penalizing factor defined as $PF=(m-n)/(m+n)$. Figure 35 illustrates the segmentation and the associated graph obtained by the method of Sfikas et al.[88].

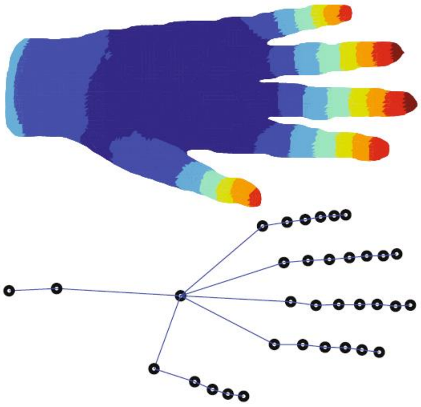


Figure 35. Segmentation and the corresponding graph using the method of Sfikas et al. [88].

Experiments on TOSCA[90] database showed that the method of Sfikas et al. outperforms the methods of Chen et al.[49], Kazhdan et al.[48] and Ben-Chen and Gotsman[91], whereas experiments on the "watertight" track of SHREC'07[69] showed that the method obtains retrieval accuracy which is comparable to the one obtained by Tung and Schmitt[92].

B2.2.5 Multiresolution Reeb graphs

The multiresolution Reeb graph (MRG) has been introduced by Hilaga et al.[93]. The normalized integral geodesic is used to construct the Reeb graph[94]. The method in Bespalov et al.[95] is an

[90] Bronstein A.M., Bronstein M.M., Kimmel R., 2006. Efficient computation of isometry-invariant distances between surfaces, SIAM J. Sci. Comput. 28, 1812–1836.

[91] Ben-Chen, M., Gotsman C., 2008. Characterizing shape using conformal factors, Proc. Eurographics 3DOR, 1–8.

[92] Tung, T., Schmitt, F., 2004. The augmented multiresolution reeb graph approach for content-based retrieval of 3D shapes, Int. J. Shape Modeling 11 (1), 91-120.

[93] Hilaga, M., Shinagawa, Y., Kohmura, T., Kunii, T.L., 2001. Topology matching for fully automatic similarity estimation of 3D shapes, Proc. Conf. Computer Graphics and Interactive Techniques, 203-212.

application of the MRG approach used for CAD models. The first step is the construction of the Reeb graph using a normalized approximation of the sum of geodesics as function. Multiple graphs of the same object are constructed with various resolutions. For each node m of the Reeb graph two attributes are defined:

$$a(m) = \frac{1}{rnum} \times \frac{area(m)}{area(M)}$$

$$l(m) = \frac{1}{rnum} \times \frac{len(m)}{\sum_n len(n)}$$

where $rnum$ is the resolution number of the MRG, $area(M)$ and $area(m)$ denote the area of the object M and the node m, respectively, and $len(m)$ is the difference of the maximum and the minimum values of the Reeb function in the segment m. The matching of two objects M_1 and M_2 is performed as described below. First a similarity measure between two nodes m and n, is defined:

$$sim(m,n) = w \times \min\big(a(m), a(n)\big) + (1 - w) \times \min(l(m), l(n))$$

where w is a weighting factor. The graph matching is performed hierarchically, meaning that first nodes at lower resolution are matched and children of the nodes are considered iteratively. When all nodes are matched, the global similarity can be computed as the sum of all matched pairs in P:

$$SIM(M_1, M_2) = \sum_{(m,n)\in P} sim(m,n)$$

Hilaga et al. performed experiments on 230 polyhedral meshes selected from selected from the various sources on the internet, the Stanford University dataset[96] and the authors' original data.

The augmented MRG has been introduced by Tung and Schmitt[92]. Each node of the MRG is enhanced with attributes so that the matching is more efficient. The graph nodes are represented in spherical coordinates. The radius and the two angles are used as attributes of the node. In addition volume, area, and curvature are used to characterize each node. Experiments are performed on three databases comprised of meshes which were collected from various sources, demonstrating that the method of Tung and Schmitt outperformed several combinations of methods and distance measures in terms of retrieval accuracy.

B2.2.6 Sub-part correspondence by structural descriptors of 3D shapes

Biasotti et al.[97] proposed a Reeb graph-based method for identifying correspondences of object sub-parts. The first step of this work involves the construction of the extended Reeb graph (ERG) of the

[94] Antini, G., Berretti, S., Del Bimbo, A., Pala, P., 2005. 3D mesh partitioning for retrieval by parts applications, Proc. IEEE ICME, 1210-1213.

[95] Bespalov, D., Regil, W., Shokoufandeh, A., 2003. Reeb graph based shape retrieval for CAD, Proc. ASME DETC.

[96] http://www-graphics.stanford.edu/data/3Dscanrep

[97] Biasotti, S., Marini, S., Spagnuolo, M., Falcidieno, B., 2006. Sub-part correspondence by structural descriptors of 3D shapes, Computer-Aided Design 38 (9), 1002-1019.

object. The ERG is directed and acyclic and resides on a surface where a finite set of contours is defined. The functions considered are the distance to the center of mass and the integral geodesic. Since the graph is directed, each node is associated with a sub-graph containing all nodes to the leaves. The signature used for each node has been defined by Kazhdan et al.[48]. A distance between two sub-graphs is defined as:

$$d(u_1, u_2) = \frac{w_1 G_s + w_2 St_s + w_3 Sz_s}{3}$$

where G_S, St_S, Sz_S are the geometric, structural and size distance between the two sub-graphs respectively. In addition, w_1, w_2 and w_3 are application-dependent weights. A matching between two graphs G_1 and G_2 is achieved by defining the following distance measure with respect to the common sub-graph G, which has been described by Biasotti et al.[98]:

$$D(G_1, G_2) = 1 - \frac{\sum_{u \in G}(1 - d(y_1(u), y_2(u)))}{\max(|G_1|, |G_2|)}$$

where y_1, y_2 denote sub-graph isomorphisms from G to G_1 and from G to G_2 respectively. $|G_i|$ denotes the cardinality of graph G_i.

Experiments were performed on a database of meshes collected from various sources which include AIMSHAPE[99] repository and PSB database. The experimental results demonstrated that the method of Biasotti et al. outperforms the MRG-based method of Hilaga et al.[93].

B2.2.7 Partial 3D shape retrieval by Reeb pattern unfolding

Tierny et al.[100] proposed a Reeb graph-based partial 3D object retrieval method, using their earlier mesh segmentation method[101]. For each segment of the object a signature is computed using its Reeb chart. Disk-like and annulus-like charts are considered. Disk-like charts correspond to one local maximum of the graph with the local maximum located in the center of the chart and the boundaries on the outer circle of the disk. Disk-like charts correspond to the fingers, whereas annulus-like chart corresponds to the palm of a hand object. Let c_i be the disk-like chart of a segment. If φ_i is the mapping of c_i to the canonical planar domain *D*, then the unfolding signature $l_{\varphi i}$ can be defined as follows:

$$l_{\varphi i}(r) = \frac{A_{ci}(r)}{A_{D(r)}} = \frac{A_{ci}(r)}{pr^2}$$

[98] Biasotti, S., Marini, S., Mortara, M., Patanè, G., Spagnuolo, M., Falcidieno, B., 2003. 3D shape matching through topological structures, Lecture Notes in Computer Science 2886, 194-203.

[99] http://shapes.aim-at-shape.net

[100] Tierny, J., Vandeborre, J., Daoudi, M., 2009. Partial 3D shape retrieval by Reeb pattern unfolding, Computer Graphics Forum 28 (1), 41-55.

[101] Tierny, J., Vandeborre, J.P., Daoudi, M., 2008. Enhancing 3D mesh topological skeletons with discrete contour constrictions, The Visual Computer 24 (3), 155-172.

where r denotes a subset of the chart, and A_{ci}, A_D denote the total area of the subset in each of the two domains. Let now c_j be the annulus-like chart of the object. The signature can be computed as follows:

$$l_{\varphi i}(r) = \frac{A_{ci}(r)}{A_{D(r)}} = \frac{A_{ci}(r)}{p(r+1)^2 - p}$$

The Reeb graph matching is performed using the above signature. A Reeb pattern is a part of the Reeb graph which contains protrusion areas. The structural signature of a Reeb pattern P_i is the couple $(n_D(P_i), n_A(P_i))$, where $n_D(P_i)$ and $n_A(P_i)$ are the number of the disk-like and annulus-like Reeb charts in P_i, which are linked by the following equation with g_{Pi} denoting the genus of the Reeb pattern:

$$n_D(P_i) = n_A(P_i) + 1 - 3g_{Pi}$$

Making use of the structural signature, the maximal common sub-graph is identified. The final step of the method is matching of the Reeb patterns using the following similarity function and a bipartite graph matching algorithm:

$$s\left(c_{Ai}, c_{Bj}\right) = 1 - L_{N1}(c_{Ai}, c_{Bj})$$

where L_{N1} is the normalized L_1 distance between the unfolding signatures of the set of matched disk charts c_{Ai} and c_{Bj}.

Experiments were performed on the partial retrieval track of SHREC'07[102] benchmark database, demonstrating that the method of Tierny el al. outperforms the methods of Biasotti et al.[97] and Cornea et al.[103] in terms of retrieval accuracy.

B2.2.8 Retrieving articulated 3D models using medial surfaces

Siddiqi et al.[104] proposed a 3D object retrieval method using medial surfaces. The medial skeleton of an object is extracted using a topology preserving thinning algorithm. The classification of the points lying on the skeleton results in an automatic segmentation of the model (Figure 36). Nodes are used to construct a graph of the object with edges connecting adjacent nodes. A bipartite graph matching method is employed to find the best match, whereas the distance used to measure similarity between nodes is the Euclidean distance of the mean curvature histogram vectors.

[102] Marini S., Paraboschi L., Biasotti S., 2007. Proc. IEEE SMI, SHREC 2007: Partial matching track, pp. 13–16.

[103] Cornea N.D., Demirci M.F., Silver D., Shokoufandeh A., Dickinson S., Kantor P.B., 2005. 3D object retrieval using many-to-many matching of curve skeletons. Proc. IEEE SMI, 366–371.

[104] Siddiqi, K., Zhang, J., Macrini, D., Shokoufandeh, A., Bouix, S., Dickinson, S, 2008. Retrieving articulated 3D models using medial surfaces. Machine Vision and Applications 19 (4), 261-274.

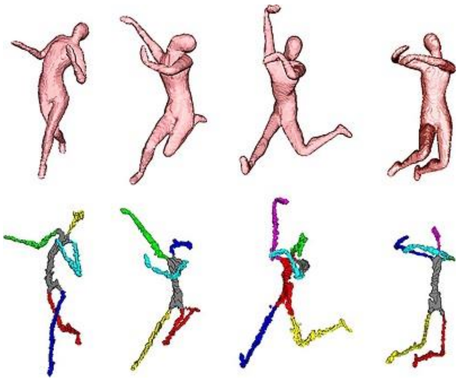


Figure 36. Segmentation using the medial surface approach[104].

B2.2.9 Shape retrieval using shock graphs

Demirci et al. [105] [106] use shock graphs to represent object shapes. The medial axis of the 2D object silhouette forms the object skeleton. Nodes on the skeleton (shocks) are processed, with each shock point being characterized by the radius of its associated maximal bitangent circle. Edges are weighted by the Euclidean distance between adjacent points. Graph embedding techniques are used to transform the shock graphs and finally the EMD is used for graph matching (Figure 37). The author performed experiments on a database of 1620 silhouettes of 9 classes, with 180 views for each, in order to demonstrate the efficiency and retrieval performance of the shock graph-based method.

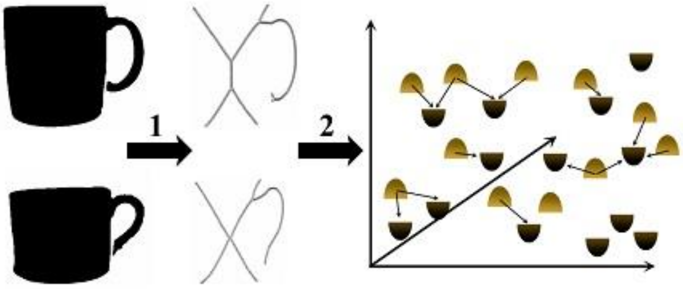


Figure 37. Overview of the shock graph based method which includes skeleton extraction and EMD matching[105].

B2.2.10 Salient geometric features for partial shape matching

Gal and Cohen-Or[107] introduced a partial 3D object retrieval method based on the segmentation of the model into patches approximated by quadric surfaces. The goal is to find salient features, i.e. regions of the surface where big difference in curvature exists. A saliency score is defined to this end for each patch F:

[105] Demirci, M.F., Shokoufandeh, A., Keselman, Y., Bretzner, L., Dickinson, S., 2006. Object recognition as many-to-many feature matching, International Journal of Computer Vision 69 (2), 203-222.

[106] Demirci, M.F, 2010. Efficient shape retrieval under partial matching, Proc. ICPR, 3057-3060.

[107] Gal, R., Cohen-Or, D., 2006. Salient geometric features for partial shape matching and similarity, ACM Trans. Graph. 25 (1), 130–150.

$$S = \sum\nolimits_{d \in F} W_1 Area(d) Curv(d)^3 + W_2 N(F) Var(F)$$

where $Area(d)$ denotes the area of the triangle d, $Curv(d)$ the corresponding Gaussian curvature, $N(F)$, $Var(F)$ the curvature variance in the patch. Weights W_1 and W_2 control the contribution of each term. Either top ten percent of the patches, or the ones that surpass some threshold are considered salient geometric features. Salient features are indexed and geometric hashing is employed to locate them in a retrieval application (Figure 38). Experiments were performed on the BSB benchmark database to support the proposed method.

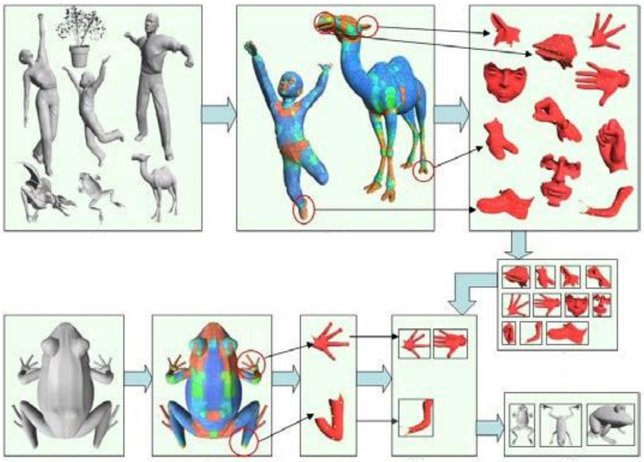


Figure 38. Salient geometric feature extraction and partial matching[107].

B2.3 BoVW-based partial 3D object retrieval methods

The past decade has seen the rise of the bag of visual words (BoVW) approach in computer vision. In relevant literature, BoVW can also be found as *bag of words*, *bag of features* or *bag of visual features*. BoVW methods have been applied for image classification, object detection, image retrieval and even visual localization for robots. They are characterized by the use of an orderless collection of image features. Although the BoVW approach in its original form lacks any structure or spatial information, it has been proved a powerful representation which matches or exceeds STAR performance in many application domains.

Abstractly, the procedure for generating a BoVW image representation is shown in Figure 39 and can be summarized as follows: (i) *build vocabulary*: extract features from all images in a training set. Vector quantize, or cluster, these features into a “visual vocabulary,” where each cluster represents a “visual word” or “term.” In some works, the vocabulary is called the “visual codebook.” Terms in the vocabulary are the codes in the codebook, (ii) *assign terms*: extract features from a test image. Use nearest neighbors or a related strategy to assign the features to the closest terms in the vocabulary, (iii) *generate term vector*: record the counts of each term that appears in the image to create a normalized histogram representing a “term vector.”[108] This term vector is the BoVW representation of the image. Term vectors may also be represented in ways other than simple term frequency, as discussed later.

[108] O’Hara, S., Draper, B.A., 2011. Introduction to the bag of features paradigm form image classification and retrieval, arXiv 1101.3354.

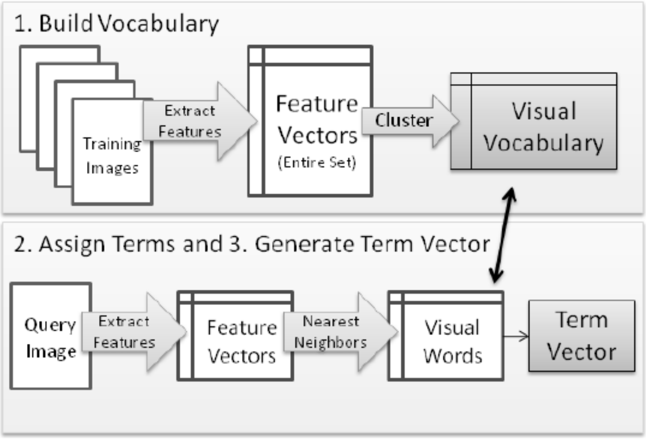


Figure 39. Process for BoVW image representation[108].

In visual information retrieval, the BoVW approach defines that each sample contains a number of local visual features. Since every visual feature, or collection of similar visual features, may appear with different frequencies on each sample, matching the visual feature frequencies of two samples achieves correspondence. Google Shape, as proposed by Bronstein et al.[109], is the one major 3D model object retrieval method in the literature. Other recent notable contributions along the same line include the methods of Furuya and Ohbuchi[110], Lavoué[111], Ohkita et al.[112], Li et al.[113] , and Atmosukarto and Shapiro[114]. It should be noted that from the BoVW-based methods presented, only the method of Lavoué has been applied for partial retrieval. However, all methods feature ideas which are potentially useful within the context of partial 3D object retrieval.

B2.3.1 3D search and retrieval from range images using salient features

Furuya and Ohbuchi[110] proposed an enhancement of an earlier 3D object retrieval method, published by Ohbuchi et al.[115], which is based on BoVW and uses dense sampling for interest point extraction, followed by the calculation of SIFT descriptors[116]. Their earlier work outperformed STAR 3D object retrieval methods on MSB database, which contains highly articulated but geometrically simple objects. However, it only equaled the retrieval accuracy of STAR methods on PSB database, which contains rigid, detailed models.

[109] Bronstein, A.M., Bronstein, M.M., Guibas, L.J., Ovsjanikov, M., 2011. Shape Google: geometric words and expressions for invariant shape retrieval ACM TOG 30 (1), 1-20.

[110] Furuya, T., Ohbuchi, R., 2009. Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features, Proc. ACM Int. Conf. Image and Video Retrieval.

[111] Lavoué, G., Combination of bag-of-words descriptors for robust partial shape retrieval, 2012. Visual Computer 28 (9), 931-942.

[112] Ohkita, Y., Ohishi, Y., Furuya, T., Ohbuchi, R., 2012. Non-rigid 3D Model Retrieval Using Set of Local Statistical Features, Proc. IEEE Multimedia and Expo Workshops, 593-598.

[113] Li, P., Ma, H., Ming, A., 2013. Combining topological and view-based features for 3D model retrieval 65, Mult. Tools Appl. 65, 335-361.

[114] Atmosukarto, I., Shapiro L.G., 2013. 3D object retrieval using salient views, Int. J. Mult. Inform. Retr. 2, 103-115.

[115] Ohbuchi, R., Osada, K., Furuya, T., Banno, T., 2008. Salient local visual features for shape-based 3D model retrieval, Proc. IEEE SMI, 93-102.

[116] Lowe, D., 2004. Distinctive image features from scale-invariant keypoint. IJCV 60 (2), 91–110.

Furuya and Ohbuchi identified that the aforementioned limitations of the earlier work of Ohbuchi et al. were connected with the quality of the SIFT-based interest point extraction of that method. Figure 40(a) illustrates an example of this issue by depicting the interest points extracted by the method of Ohbuchi et al. on a 3D model example. Furuya and Ohbuchi noted that the depth image of the potted plant produced a large number of small-scale features near the leaves. These features, being scale invariant, could match local geometrical features of other models that are similar in shape, yet completely different in scale. Consequently, the potted plant could potentially match models having completely different overall shape. On the other hand, an important large scale feature, in this case a large trapezoidal shape of the pot, is underrepresented. For simpler, less detailed shapes, e.g. those of MSB database, the original SIFT-based interest point detector worked very well. However, for the PSB database, which contains models having considerably more detail, the salient points cannot provide a balanced representation of model features.

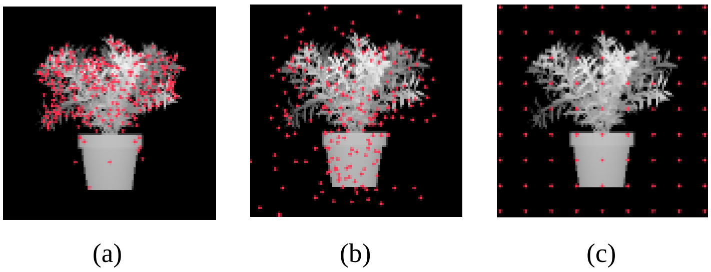


Figure 40. Example of feature points using: (a) SIFT-based interest point detector, (c) dense interest points, (c) a grid of interest points[110].

In order to cope with the limitations of the SIFT-based interest point extractor, Furuya and Ohbuchi employed dense random sampling. For each image in the multi-scale image pyramid of the SIFT algorithm, the pixels to be sampled are drawn randomly from pixels having non-zero intensity value. A range image is rendered with zero pixel value as its background, and images in the SIFT pyramid are blurred according to their scale in the pyramid. Thus, a pixel from an image in the SIFT pyramid that has non-zero value is located on or near but not far from the image of the 3D model. As non-zero pixels are different for each image in the SIFT image pyramid, the positions of samples are different across scales in the SIFT pyramid. For comparisons, Furuya and Ohbuchi also implemented another sampling strategy, which samples the image at regular grid points. Figures 40(b) and 40(c) illustrate the interest points extracted by dense grid sampling, respectively. It can be observed that that dense sampling located more samples near the pot, whereas grid sampling is uniform, regardless of the image features across image scales. Apart from dense sampling, Furuya and Obhuchi's method adopted the SIFT feature descriptors of Lowe[116], the k-means clustering approach for codebook learning and the computationally efficient ERC-tree of Guerts et al.[117] for vector quantization. Computational cost is further reduced by utilizing a GPU implementation of SIFT feature extraction.

In their experiments, Furuya and Obhuchi used the PSB, MSB and ESB databases to comparatively investigate various aspects of their BoVW-based 3D object retrieval, including codebook learning and encoding, sampling strategy and vocabulary size. Their main conclusions are: (i) although slow, *k*-means clustering is preferable, since codebook learning is needed only once, (ii) ERC-tree[117] is much more efficient than *k*-means for nearest neighbor search, with slightly worse retrieval accuracy, (iii) the dense sampling approach is much less sensitive to vocabulary size than SIFT-based sampling, (iv) Ohbuchi and Furuya's method outperforms, among others, the methods of Ohbuchi et al.[115], Chen et al.[49], Wahl et al.[118] and Kazhdan et al.[48] in all datasets.

[117] Guerts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees, Machine Learning 36 (1), 3-42.

[118] Wahl, E., Hillenbrand, U., Hirzinger, G., 2003. Surflet-pair-relation histograms: a statistical 3D-shape representation for rapid classification, Proc. 3DIM, 474-481.

B2.3.2 Shape Google

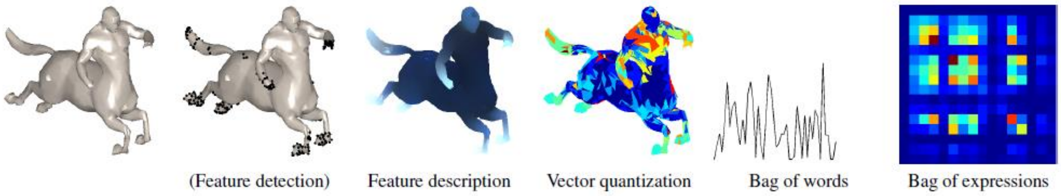


Figure 41. Overview of the Google Shape pipeline[109].

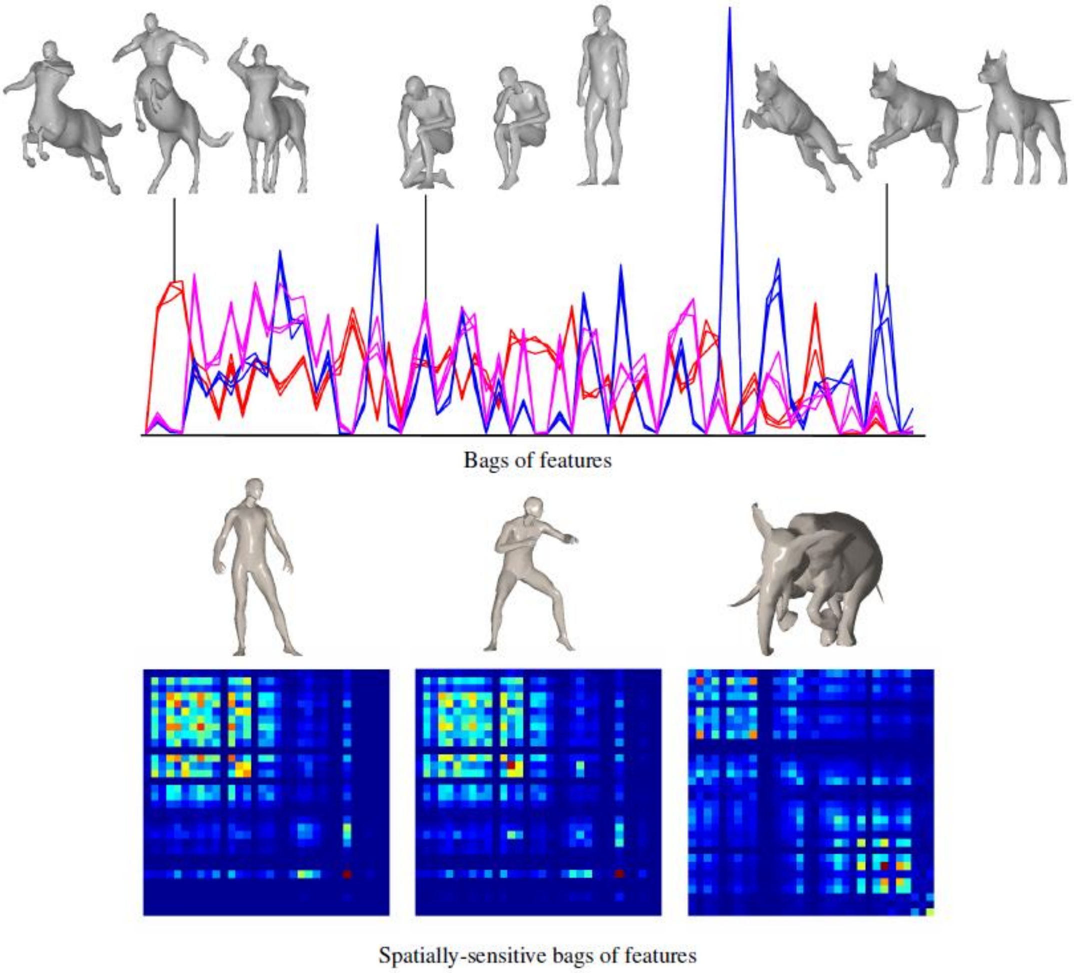


Figure 42. Top row: examples of BoVW computed for different deformations of centaur (red), dog (blue), and human (magenta). Note the similarity of BoVW of different transformations and dissimilarity of BoVW of different shapes. Also note the overlap between the centaur and human BoVW, due to partial similarity of these shapes. Bottom row: examples of spatially-sensitive BoVW computed for different deformations of human (left and center) and elephant (right) [109].

Shape Google is a 3D object retrieval method applicable on both meshes and point clouds, which focuses on the retrieval of *non-rigid objects*. It starts by calculating a feature detector and descriptor based on heat kernels of the Laplace-Beltrami operator, inspired by Sun et al.[119]. The descriptors derived are used to construct a BoVW vocabulary. This representation is invariant to isometric deformations, robust under a wide class of perturbations, and allows one to compare shapes undergoing different deformations. Bronstein et al. take into consideration the spatial relations between features in an approach similar to commute graphs[120], which has been shown to enhance the retrieval performance. Finally, they adopt metric learning techniques, widely used in the computer vision community[121] and represent shapes as compact binary codes that can be efficiently indexed and compared using the Hamming distance.

Figure 41 provides an overview of the Google Shape pipeline. The shape is represented as a collection of local feature descriptors, either dense or computed at a set of stable points, following an optional stage of feature detection. The descriptors are then represented by “geometric words” from a “geometric vocabulary” using vector quantization, which produces a shape representation as a BoVW or pairs of words, i.e. “expressions”. Finally, similarity sensitive hashing is applied on the BoVW. Figure 42 visualizes the discriminative capability of the employed heat kernel-based BoVW. Google Shape can be adopted as a framework with different descriptors and detectors, depending on the application demands.

B2.3.3 Spatially sensitive BoVW method for 3D object retrieval

Lavoué[111] has presented an alternative 3D object retrieval method which also combines standard BoVW and spatially-sensitive BoVW. His method relies on a uniform sampling of feature points based on Lloyd’s relaxation iterations. Each feature point is associated to a descriptor defined as the Fourier spectra of a local patch, which is computed by projecting the geometry onto the eigenvectors of the Laplace–Beltrami operator, so as to speed-up computations and enhance discriminative capability.

The experimental evaluation of this method for partial retrieval has been performed on SHREC 2007 partial retrieval benchmark[102]. Each of the query models is composed of sub-parts from two or three models from the testing set. A ground-truth classification of each model of the testing set as highly relevant, marginally relevant or non-relevant is provided for each query. Lavoué’s method is shown to outperform the 3D object retrieval methods of Tierny et al.[100] and Toldo et al.[122]. The author explains this on the basis that his method discards most of the structural information, hence the topological changes due to the sub-part merging do not significantly affect the BoVW. Moreover, it has been shown that standard and spatially-sensitive BoVW methods are complementary since their combination provides a significant gain with regards to their individual performances.

A weakness of Lavoué’s method, as the author admits, is that although it correctly retrieves a model from a partial query, it does not perform the precise matching between the corresponding sub-parts. A solution to perform this matching could have been to construct a graphical structure over the set of feature points and apply some kind of fast approximate subgraph isomorphism.

[119] Sun, J., Ovsianikov, M., Guibas, L.J., 2009. A concise and provably informative multi-scale signature based on heat diffusion. Proc. SGP, 1383-1392.

[120] Behmo, R., Paragios, N., Prinet, V., 2008. Graph commute times for image representation. Proc. IEEE CVPR, 1-8.

[121] Jain, P., Kulis, B., Grauman, K., 2008. Fast image search for learned metrics. Proc. CVPR, 1-8.

[122] Toldo, R., Castellani, U., Fusiello, A., 2009. Visual vocabulary signature for 3D object retrieval and partial matching. Proc. Eurographics Workshop 3DOR, 21-28.

B2.3.4 Ohkita et al. BoVW method for 3D object retrieval

Ohkita et al.[123] introduced a BoVW-based 3D object retrieval method using sets of local statistical features (LSFs). Their method natively compares 3D models in oriented point set representation. If the 3D models are in surface-based representation, e.g. manifold surface or polygon soup, an initial transformation step is required to convert them into oriented point sets.

The algorithm is illustrated in Figure 43 and can be summarized in the following steps: (i) if a 3D model to be compared is in surface-based representation, it is converted into an oriented point set consisting of m points, (ii) a set of LSFs, each one computed for one of n randomly selected feature points ($n<m$), is computed from the set of m oriented points, (iii) LSFs are integrated into a feature vector per 3D model, iv) distances among feature vectors of a query model and database models are computed. Top matches are returned as a retrieval result.

Each LSF is computed using a set of sample points within a sphere of radius r, centered at the feature point. The LSF radius r for a 3D model is set relative to the radius of the smallest bounding sphere. Let p_1 the feature center and n_1 its normal vector. Assume also that the point p_2 with an associated normal vector n_2 lies within the sphere of influence of radius r of the feature center p_1. Using p_1 and p_2, a 4-tuple (α, β, γ, δ) consisting of a distance δ and three values α, β, γ related to angles is computed. Here, point p_1 is the LSF feature point, and p_2 is a point other than p_1, which is contained in the sphere of influence. The four values α, β, γ and δ are computed as follows:

$$a = \arctan(w \cdot n_1, u \cdot n_2)$$

$$\beta = v \cdot n_2$$

$$\gamma = u \cdot \left(p_2 - \frac{p_1}{||p_2 - p_1||}\right)$$

$$\delta = ||p_2 - p_1||$$

where $u = n_1$, $\mathrm{v} = (p_2 - p_1) \times \mathrm{u}/||(p_2 - p_1) \times \mathrm{u}||$, and $w = u \times v$.

If there are k points within the sphere, a set of $(k-1)$ 4-tuples of values a, β, γ and δ are computed for each feature center. The computed 4-tuples are then collected into a 4-dimensional joint histogram. If the histogram has 5 bins for each 4-tuple value, the resulting 1D-flattened LSF is 5^4=625 dimensional.

Ohkita et al. integrate the set of features per 3D model into a feature vector by using two approaches: the BoVW and linear combination (LC). In the case of the latter, the features are calculated by simply summing, component by component, the histograms of local features into a single histogram, having the same dimension as the local features. LC may be considered as a regressed form of BoVW. Interestingly, instead of BoVW and LC, Ohkita et al. also consider an approach that does not integrate LSFs. That is, given n LSF features per 3D model, all pairs of LSF features are compared among two sets of LSF features. They name this approach “all pair comparison LSF (AC-LSF)”. AC-LSF is computationally very expensive during a search through a database, since it has O(n^2) complexity per pair of 3D models.

[123] Ohkita, Y., Ohishi, Y., Furuya, T., Ohbuchi, R., 2012. Non-rigid 3D Model Retrieval Using Set of Local Statistical Features, Proc. IEEE Multimedia and Expo Workshops, 593-598.

Experimental evaluation using two 3D model retrieval benchmarks showed that the method of Ohkita et al. is quite effective for McGill shape benchmark (MSB) database, which consists of highly articulated, yet simpler shapes. For the rigid, highly diverse set of shape models in PSB database, retrieval accuracy is modest compared to other STAR methods. Another conclusion drawn from the experimental results is that BoVW enhances retrieval accuracy when compared to LC, at the cost of increased execution time. AC-LSF, although performing admirably well, is aborted as computationally prohibitive.

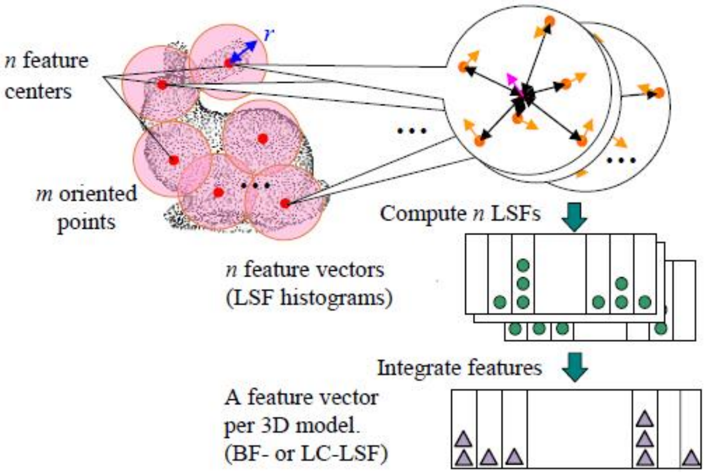

n feature centers
r
m oriented points
...
Compute n LSFs
n feature vectors (LSF histograms)
Integrate features
A feature vector per 3D model. (BF- or LC-LSF)

Figure 43. A large number of local features are extracted, and they are integrated into a feature vector per 3D model[123].

B2.3.5 Combining topological and view-based information with shell-sector BoVW

Li et al.[124] introduced a BoVW-based 3D object retrieval method which uses SIFT descriptors[116] derived from several object views. Object views are determined by spatial structure circular descriptor (SSCD) images[125] on topologically salient points.

Initially, the 3D model is pose-normalized by means of local translational invariance cost principal component analysis (LTIC-PCA)[126]. As a next step, topologically salient points are extracted as maximum critical points from the radial basis function (RBF) level set (Figure 44).

[124] Li, P., Ma, H., Ming, A., 2013. Combining topological and view-based features for 3D model retrieval 65, Mult. Tools Appl. 65, 335-361.

[125] Gao, Y., Dai Q., Zhang N., 2010. 3D model comparison using spatial structure circular descriptor. Patt. Rec. 43 (3), 1142–1151.

[126] Chaouch, M., Verroust-Blondet, A., 2008. A novel method for Alignment of 3D models, Proc. IEEE Int. Conf. Shape Modeling and Applications, 187–195.

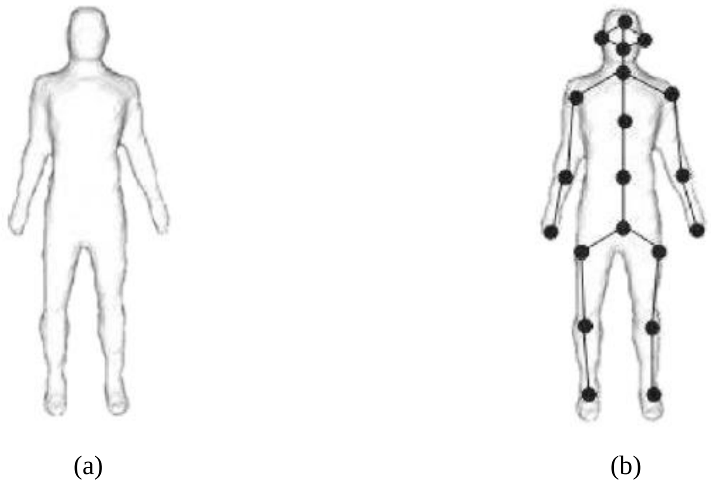


Figure 44. (a) A standing human model in the Shape Google Benchmark, (b) the topological points of this human model[124]*.*

The multi-resolutional Reeb graph (MRG) is constructed by means of the function μ on a point v in the set S of salient topological points:

$$\mu(v) = \int_{p \in S} g(v, p) dS$$

where the function $g(v, p)$ is the geodesic distance between v and the other points P in S. To exploit the full dynamic of $\mu(v)$ over [0,1], $\mu(v)$ is normalized with the value of the point to the surface center and the max value:

$$\mu_N(v) = \frac{|\mu(v) - \mu_{sc}(v)|}{\max_{p \epsilon S}[\mu(p) - \mu_{sc}(v)]}$$

where $\mu_{sc}(v)$ is the function from the point v to the surface center. A collection of SSCD images is generated for each topologically salient point (Figure 45). The optimal number of images has been experimentally determined as 20.

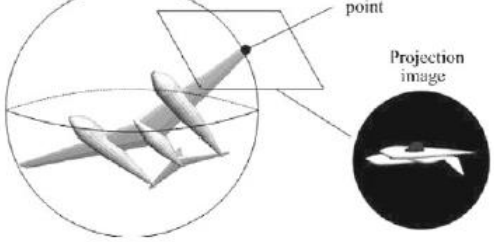
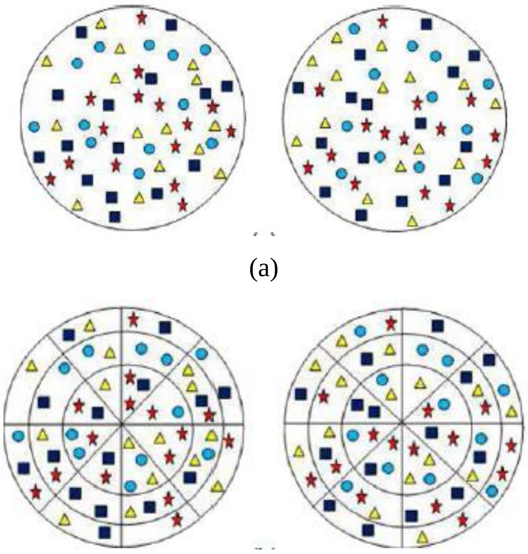


Figure 45. An example of images rendered from topological points[124]*.*

A set of SIFT features are extracted from each SSCD image by using Vedaldi's SIFT++[127]. Considering that the original BoVW method cannot preserve the spatial structure information of the view-based features, Li et al. introduce a spatially sensitive BoVW variant which combines the shell-sector model with logarithmic shell radii (Figure 46).

Li et al. performed experiments on PSB database and shape Google benchmark (SGB), comparing their method with several STAR 3D object retrieval methods, including the ones proposed by Chen et al.[49], Vranic[72], Daras and Axenopoulos[54] [55] , Papadakis et al.[71]. The authors report that their method obtains the highest retrieval accuracy in both benchmark databases.


(a)

(b)

Figure 46. Comparing the original BoVW and the combined shell-sector BoVW representation: (a) two images represented with the original BoVW method, (b) the same images represented with the combined shell-sector BoVW method[124].

B2.3.6 3D object retrieval based on salient views

Atmosukarto and Shapiro[128] proposed a 3D object retrieval method which uses silhouette-based descriptors derived from salient 2D object views. Following the surface characterization of Besl and Jain[129], Gaussian and mean curvatures are employed as low-level descriptors. These low-level feature values are then aggregated into mid-level descriptors in the form of local histograms, which represent a neighborhood. Like the local field descriptor (LFD) of Chen et al.[49], the method of Atmosukarto and Shapiro uses rendered silhouette 2D images as views to build the descriptor of the 3D object. However, unlike LFD, which extracts features from 100 2D views, this method selects only salient views, which are considered as the most useful in describing the 3D object. Since the 2D views used to describe the 3D objects are silhouette images, some of the salient points appear on the 3D object contour (Figure 47).

[127] A. Vedaldi, SIFT++, http://www.vlfeat.org/~vedaldi/code/siftpp.html

[128] Atmosukarto, I., Shapiro L.G., 2013. 3D object retrieval using salient views, Int. J. Mult. Inform. Retr. 2, 103-115.

[129] Besl, P.J., Jain, R.C., 1985. Three-dimensional object recognition, Comput. Surv. 17 (1), 75–145.

(a) (b)

Figure 47. (a) Salient points must appear on the contour of the 3D objects for a 2D view to be considered as a 'salient' view. The contour salient points are colored in green, whereas non-contour salient points are in red, (b) silhouette image[128].

A salient point is considered as a contour salient point if its surface normal vector is perpendicular to the camera view point. For each possible camera view point, from a total of 100, the algorithm accumulates the number of contour salient points that are visible for that view point. The 100 view points are then sorted based on the number of contour salient points visible in the view. The algorithm selects the final top *K* salient views used to construct the descriptor for a 3D model. The value of *K* has been experimentally determined.

A more restrictive variant of the algorithm selects the top *K* distinct salient views. In this variant, after sorting the 100 views based on the number of contour salient points visible in the view, the algorithm uses a greedy approach to select only the distinct views. The algorithm starts by selecting the first salient view, which has the largest number of visible contour salient points. It then iteratively checks whether the next top salient view is too similar to the already selected views. The similarity is measured by calculating the dot product between the two views and discarding views whose dot product to existing distinct views is greater than a threshold *P*. Figure 48 (top row) shows the top five salient views, whereas Figure 48 (bottom row) shows the top five distinct salient views for a human object. It can be observed that the top five distinct salient views provide a more complete description of the object shape.

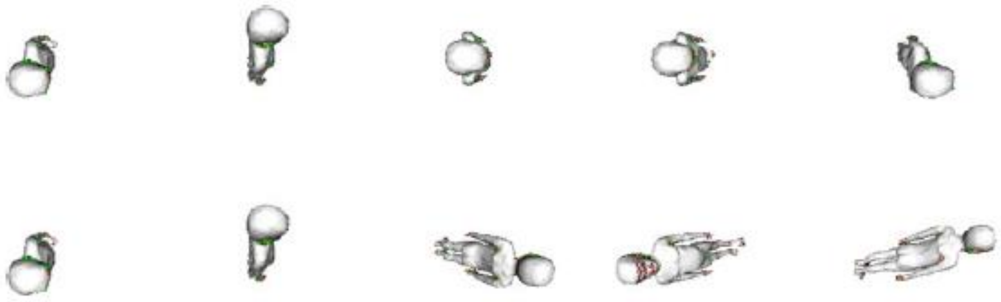


Figure 48. Top five salient views for a human query object (top row) and top five distinct salient views for the same human query object (bottom row). The distinct salient views capture more information regarding the object's shape[128].

Experiments were performed on the "watertight" track of SHREC'08[130], the PSB database and a benchmark database compiled by scanning hand-made clay toys using a laser scanner. In the case of the latter benchmark, the authors added artificial objects by deforming the original scanned 3D models

[130] Giorgi, D., Marini, S., 2008. SHREC 2008 track: classification of watertight models. Proc. Eurographics, workshop 3DOR.

in a controlled fashion, using 3D StudioMax software[131]. The results presented demonstrate that the method of Atmosukarto and Shapiro achieves higher retrieval accuracy than LFD, accompanied by a 15-fold speedup in feature extraction time.

B2.4 Other partial 3D object retrieval methods

This subsection presents partial 3D object retrieval methods that cannot be categorized as view-based, part-based or BoVW-based.

B2.4.1 Local feature histograms for range image classification

Hetzel et al.[132] explore a view-based method for the recognition of free-form objects in range images. They combine a set of local features, pixel depth, surface normal and curvature metrics in a multidimensional histogram, aiming to classify range scans, with pose-estimation as a byproduct.

Unlike luminance images, intensity is illumination invariant in range images. However, distance histograms are sensitive to the perceived depth range, whereas they are also problematic in situations where the depth clutter can be influenced by other objects or background clutter and should not be used in applications where such situations may occur. Surface normals can be easily calculated from the first image derivatives image by employing a representation based on sphere coordinates. For the representation of surface curvature, the shape index of Dorai and Jain[79], also used in the local surface patch (LSP) method of Chen and Bhanu[133] is employed.

For histogram comparison, both χ^2-test-based histogram matching and maximum a-posteriori probability estimation are used. On ideal test images, both methods produce comparable results. However, the latter method is more capable to deal with partial occlusions, whereas it also provides a measure of confidence for the obtained classification result. Experiments were performed on a collection of synthetic range images, taken from high-resolution polygonal models available on the authors' web site.

B2.4.2 3D object retrieval based on depth-buffer and silhouette relevance indices

Chaouch and Verroust-Blondet[134] have proposed a 2.5D object retrieval method which is based on relevant indices derived from silhouette or depth-buffer images. As a relevance index which depends on the outer object silhouette, they use two alternatives: the first is standard and involves the computation of the number of non-null pixels on the image, i.e. the area of the projected surface of the 3D model on the corresponding face of the bounding box:

$$R_a = card\{s_{ab} | s_{ab} = 1, 0 \leq a, b \leq N-1\}$$

[131] Autodesk, 2009. 3dsmax, http://autodesk.com

[132] Hetzel, G., Keibe, B., Levi, P., Schiele, B., 2001. 3D object recognition from range images using local feature histograms. Proc. IEEE CVPR 2, 394-399.

[133] Chen, H., Bhanu, B., 2007. 3D free-form object recognition in range images using local surface patches. Pattern Recognition Letters 28 (10), 1252 – 1262.

[134] Chaouch, M., Verroust-Blondet, A., 2006. Enhanced 2D/3D approaches based on relevance index for 3D-shape retrieval. Proc. SMI, 36.

where s_{ab} is the pixel value of the image at position (a, b). To moderate the influence of the area which in some cases may affect the retrieval performance, they consider the square root of the relevance defined in the above equation. As a second alternative for silhouette-based relevance index, they use the average cord of a 2D mesh, i.e. the average length of all possible cords connecting two contour points:

$$R_c = \sum_{a=0}^{N-1} \sum_{b=0}^{N-1} \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} \frac{\delta_{c_{ab} \cdot c_{pq}}}{L(L-1)} \sqrt{|a-p|^2 + |b-q|^2}$$

$$\delta_{x.y} = 1 \text{ if } x = y = 1 \text{ and } \delta_{x.y} = 0 \text{ otherwise}$$

Several sampling strategies have been considered for selecting the contour points used in this calculation. These strategies include using all points of the outer contour of the silhouette, points of high curvature or a subset of interest points, as in the work of Lowe[116]. However, the authors note that this type of information can well represent the relevance for some particular cases but can be much less efficient for many 3D models due to the unstable behavior of such key points.

Chaouch and Verroust-Blondet also proposed two methods to compute the relevance indices of depth-buffer images. The first one introduces the depth by taking the sum of all values of the non-null pixels of the depth-buffer image, thus computing the volume enclosed between the visible parts of the 3D object and the opposite plane of the bounding box:

$$R_d = \sum_{a=0}^{N-1} \sum_{b=0}^{N-1} u_{ab}$$

where u_{ab} is the pixel value of the depth-buffer image at position (a, b).

The second relevance index proposed for depth-buffer images is the sum of the distances between the center of mass of the 3D model and all its visible points:

$$R_g = \frac{1}{2w} \sum_{a=0}^{N-1} \sum_{b=0}^{N-1} d_{ab}$$

$$d_{ab} = \sqrt{|a - N/2|^2 + |b - N/2|^2 + 2w|u_{ab} - 1/2|^2}$$

where $2w$ is the length of the sides of the extended enclosing bounding box.

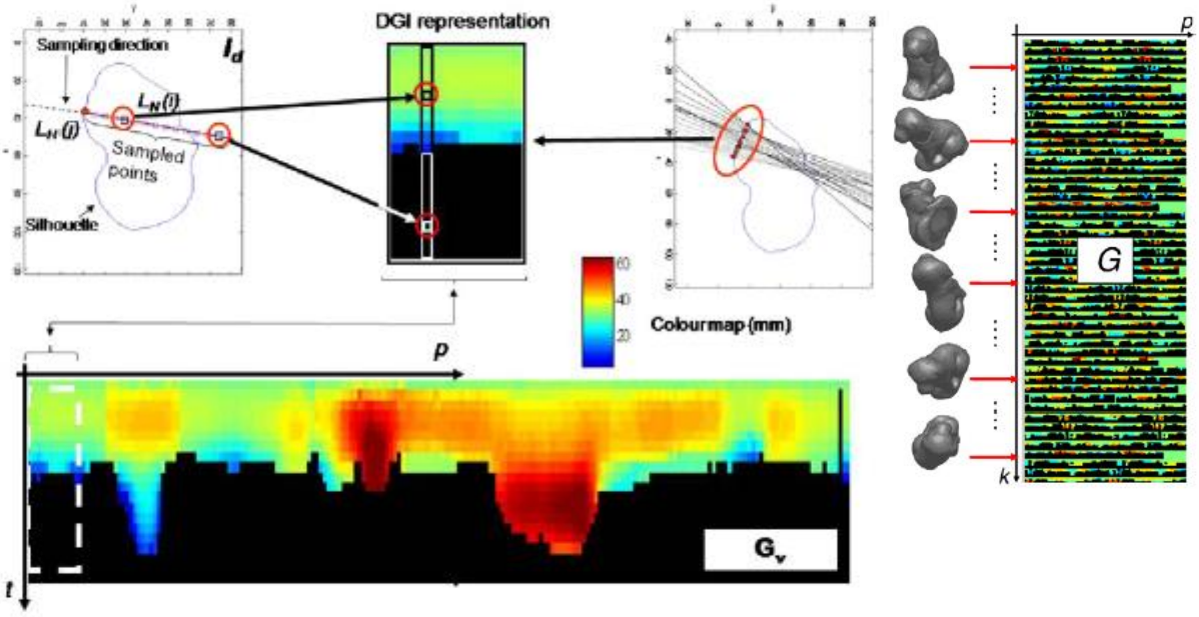
Experiments were conducted on range images artificially acquired from the Princeton 3D shape benchmark database (PSB). The obtained results have been compared with the silhouette and depth-buffer methods of Vranic[59], on the same database. Chaouch and Verroust-Blondet report enhanced retrieval quality for both silhouette and depth-buffer-based variants, without increasing the overall computational cost.

B2.4.3 Depth gradient images (DGIs)

Adan et al.[135] introduced and analyzed a 3D object retrieval strategy *for scenes*, based on depth gradient image (DGI) representation. DGI synthesizes both surface and contour information, aiming to avoid restrictions on the layout and visibility of each object in the scene. Figure 49 summarizes this strategy. Let v be an arbitrary viewpoint of the object, L be the set of pixels of depth image I_d which represent the object and L_H be the set of pixels corresponding to the object contour. In addition, let p = ord(L_H) be the number of pixel of the contour and t = dim(diag(I_d)) be the dimension of the diagonal of the depth image. The DGI from viewpoint v, G_v, is a t×p matrix:

$$G_v(i,j) = I_d\left(L_N\left(i, L_H(j)\right)\right) - I_d(L_H(j))$$

where $L_N\left(i, L_H(j)\right), i = 1, 2, \dots, t$ is the set of pixels that are in the normal direction to the contour at the point $L_H(j)$, which are sorted from $L_H(j)$ towards the object interior. L_N can be sub-sampled in order to reduce the size of the DGI representation and limit memory requirements. The depth values of a set of equally-spaced pixels in the normal direction can thus be taken in the above equation. It is clear that G_v is invariant to changes in the observer distance. G_v is a small image where the jth column contains the set of depth gradients in the normal direction in $L_H(j)$, whereas the ith row stores the gradients for points that are equidistant in the image to the contour in their corresponding normal directions. Figure 49(a) shows an example of how depth-gradient values for one contour-point (on the left) and consecutive contour-points (on the right) are calculated. Normal lines are plotted in all cases.



(a) (b)

Figure 49. Building of the DGI representation: (a) depth gradients generated over a sampling direction when a contour point is selected. DGI values inside and outside of the object are included, as well as a section of DGI corresponding to a part of the contour and its integration in the DGI, (b) the global DGI model[135].

[135] Adan, A., Merchan, P., Salamanca, S., 2011. 3D scene retrieval and recognition with depth gradient images. Pattern Recognition Letters 32 (9), 1337 – 1353.

DGI representation is suitable for characterizing both partial views and the complete object. The global DGI model can be defined as follows:

$$G(i,j) = G_{\mu}(i',j'), i' = MOD(i,t), j' = j, \mu = DIV(i,t)$$

where G_{μ} is the partial DGI obtained from the viewpoint μ. Thus, the global DGI model consists of an image of dimension $(k/t, p)$, which is duplicated in practice, so as to carry out an efficient partial-global DGI matching. Figure 3(b) shows the global DGI for one object. Note that this is a single image of 1M pixels, which synthesizes the surface information of the complete object.

Promising retrieval performance is obtained by applying DGI on various scene types, which include occlusion, injected noise and highly complex cluttered scenes. Moreover, DGI outperforms spin-images[136] (see subsection B3.1.1) on Mian's public dataset whereas it obtains comparable performance to Mian's tensor-shape method[137].

B2.4.4 Fast-reject scheme using onion descriptors

Attene et al.[138] introduced a method for 3D shape matching which is based on what they call "fast reject schema". The fast reject schema is based on the computation of monotonic *onion descriptors* for the part-in-whole matching. An onion descriptor O_M is a vector of multi-fielded descriptors of non-decreasing dimension, defined as:

$$O_M(c, r_1, \dots, r_k) = \left(S_M(c, r_1), \dots, S_M(c, r_k)\right)$$

where S_M is a *N*-dimensional multi-fielded local shape descriptor defined as:

$$S_M(c, r) = (s_M^1, s_M^2, \dots, s_M^N)$$

The descriptor S_M encodes the surface M of an object around a specific point c up to a distance r, expressed using a specific metric. Each multi-fielded descriptor S_M constituting an onion O_M is called a layer. Note that the various layers of the sequence are all referred to the same centre point c, but consider a varying size of the neighbourhood. Figure 50 illustrates an example of an incrementally-defined onion descriptor of a nose.

[136] Johnson, A.E., Hebert, M., 1999. Using spin images for efficient object recognition in cluttered 3D scenes. IEEE Trans. Pattern Anal. Machine Intell. 21 (5), 433–449.

[137] Mian, A.S., 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.

[138] Attene, M., Marini, S., Spagnuolo, M., Falcidieno, B., Part-in-whole 3D shape matching and docking, 2011. Visual Computer 27, 991-1004.

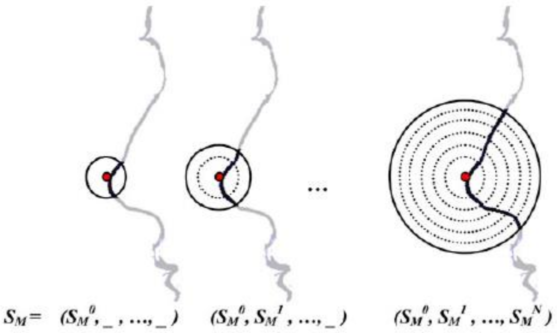

$S_M = (S_M^0, \_, \ldots, \_)$
$(S_M^0, S_M^1, \ldots, \_)$
$(S_M^0, S_M^1, \ldots, S_M^N)$

Figure 50. An example of incrementally-defined onion descriptor of a nose. From left to right: the first circle encloses a small region around the nose tip; this region defines the first layer S_M^0. The second circle encloses a slightly larger region defining both S_M^0 and S_M^1. The entire nose is enclosed by the largest circle on the right, which defines the complete onion descriptor of the shape sphere[138].

The fast reject schema starts by computing an onion descriptor O_T for the template model and discretizing the input 3D scene into a set D_I of scene points. Initially, all the points of the scene are candidate to be good matches. In the first iteration, for each scene point only the first layer of its onion descriptor is computed and its distance from the first layer of O_T is calculated. If such a distance exceeds a given threshold, the scene point is excluded from the search space or, in other words, it is rejected from the set of potential good matches. In the second iteration, for each non-rejected point, the second layer is computed and compared with the second layer of O_T. Also in this case, an excessive distance causes the rejection of the point from the search space. The process stops when the last layer is computed for all the remaining non-rejected points. Among these remaining points, the good matches are those whose last layer has a distance from the last layer of O_T smaller than the threshold. This hierarchical search procedure (Figure 51) facilitates the reduction of the search space, enabling a quick match between a template shape and one or more parts of a scene.

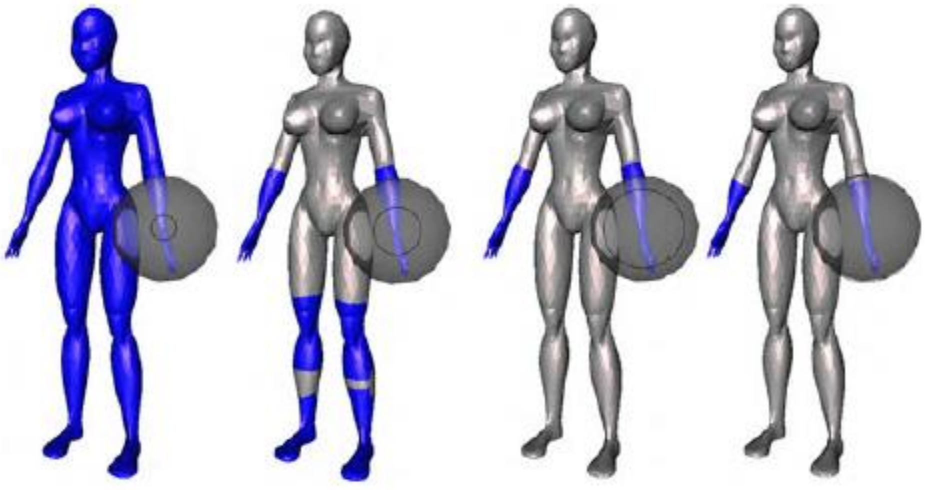


Figure 51. This image depicts the reduction in the search space while using the Gaussian curvature descriptor. From left to right: while at the beginning each vertex is a candidate for a good match, as enlarging the neighbourhood, differences become not negligible and fewer and fewer vertices have a neighbourhood sufficiently similar to the template defined by the semi-transparent sphere[138].

Attene et al. test several existing multi-fielded descriptors that act as onion descriptors and can be used in the context of the fast reject schema, including shape contexts[139] and spherical harmonics[140], as well as a coarse volumetric descriptor and a new surface descriptor based on curvature analysis. The authors perform experiments with three different descriptor configurations, on several 3D object models, including the well-known Stanford Buddha. They conclude that their method obtains comparable retrieval accuracy with the method of Gal et al.[107], but is approximately 400 times faster. However, one may notice that there are no comparisons with other well-known retrieval methods, whereas standard benchmark databases are not used in the experiments.

B2.5 Comparative analysis of methods for partial 3D object retrieval

Table 6 summarizes the STAR on partial 3D object retrieval. From the table, it can be observed that the majority of the related literature is devoted on structured data represented by 3D meshes. However, these methods are still relevant in the context of PRESIOUS WP2, since there is an intense research activity on 3D mesh generation algorithms, which can be used to convert unstructured input data, such as point clouds[141] [142]. In addition, several elements of the above methods, as is the hierarchical search of camera position and orientation, can be adopted in a point cloud context, featuring the local shaped descriptors for unstructured 3D data, which are presented in subsection B3.

Table 6 also reveals that most partial 3D object retrieval methods are actually 2.5D, i.e. they derive descriptors from 2D projections of the 3D model. This approach suits partial retrieval in the sense that a 2D query model can be compared with all 2D projections of each target model in order to keep the minimum distance associated with the most similar projection, as suggested by Daras and Axenopoulos[54] [55]. It can also be recalled that the earlier comparative study of Shilane et al.[50] concluded that the 2.5D-based retrieval methods, which were available at the time, outperformed 3D methods in terms of retrieval accuracy. Another issue which is related to the context of PRESIOUS WP2, is that creating range images or silhouettes directly from the original input data, which are in a point cloud form, cannot guarantee that there are not going to be gaps and holes present in the result. This can either be addressed by quickly constructing a mesh, provided that the ordering of the point cloud is known, or by correcting the resulting image with some form of fast in-painting.

Another conclusion drawn from Table 6 is that most partial 3D object retrieval methods are based on descriptors calculated over interest points, either dense, or extracted by means of a salient point detector. Furuya et al.[110] and Ohkita et al.[112] argue in favor of using dense points, although the appropriate strategy might be representation and application-dependent. Interest points constitute also an intrinsic element of the BoVW paradigm. It is evident that the latter is an emerging trend in 3D object retrieval, with several major works appearing in the last two years. In the case of the works of Bronstein et al.[109] and Lavoué [111], their BoVW variations employ spatially sensitive information, which has been shown to enhance retrieval performance.

Experimental comparisons featured in the BoVW publications show that these methods outperform some of the previous, non-BoVW-based methods in terms of retrieval accuracy. In particular, the comparison with the linear combination (LC) approach, presented in the work of Ohkita et al.[112], is of fundamental importance, since it actually compares two algorithms that only differ in using BoVW or not. LC omits clustering and simply sums, component by component, the histograms of local features into a single histogram, which has the same dimensions with local features. LC can be considered as a

[139] Belongie, S., Malik, J., Puzicha, J., 2002. Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 11 (6), 509–522.

[140] Shilane, P., Funkhouser, T., 2007. Distinctive regions of 3D surfaces, ACM Trans. Graph. 26 (2).

[141] Marton, Z.C., Rusu, R.B., Beetz, M., 2009. On fast surface reconstruction methods for large and noisy datasets, Proc. IEEE ICRA, 2829-2834.

[142] Zhou, K., Gong, M., Huang, X., Guo, B., 2011. Data-parallel octrees for surface reconstruction, IEEE Trans. Visualization and Graphics 17 (5), 669-681.

regressed form of BoVW. The experiments on both MSB and PSB benchmark databases show that BoVW outperforms LC. On the other hand, the non-BoVW-based method of Daras and Axenopoulos[54] [55] is shown to outperform the BoVW-based method of Ohbuchi et al.[60]. Another interesting outcome of the experiments of Ohkita et al. is that shape matching based on all pairs of interest points of a query and a target model performs admirably well and only suffers in terms of computational complexity. It is tempting to speculate that this approach could perhaps be used for matching instead of BoVW, when the number of interest points is small enough to allow calculations for all respective pairs.

Method	**Year**	**Dataset**	**Part-based**	**View-based**	**Interest points**	**BoVW**	**Methodology**	**Outperformed**	**Data type**
Hetzel et al.	2001	Authors' synthetic meshes	No	No	No	No	Multi-dim. histogr. of depth, normals and curvatures	-	Mesh
Hilaga et al.*	2001	Meshes from various sources	Yes	No	No	No	MRG, hierarchical matching	-	Mesh
Tung and Schmitt *	2004	Meshes from various sources	Yes	No	No	No	aMRG, hierarchical matching	Several combinations of distances/methods	Mesh
Kim et al. *	2005	Authors' synthetic meshes	Yes	No	No	No	Graph-based, constr. morphological decomposition	-	Mesh
Biasotti et al. *	2006	Meshes from various sources	Yes	No	No	No	ERG	Hilaga et al. (2001)	Mesh
Chaouch and Verroust-Blondet	2006	PSB	No	Yes	Yes	No	Depth-buffer and silhouette relevance indices	Vranic (2004)	Mesh
Gal and Cohen-Or*	2006	PSB	Yes	No	Yes/salient regions	No	Patch-based, quadric surface approximation	-	Mesh
Ansary et al. *	2007	PSB, Renault CAD models	No	Yes	No	No	Adaptive views clustering, Bayesian-based retrieval	Chen et al. (2003), Kazhdan et al. (2003)	Mesh
Chen and Bhanu	2007	Ohio State University	Yes	No	Yes/shape index of Dorai and Jain (1997)	No	Local surface patches based on shape indices	Johnson and Hebert (1999)	Range image
Shih et al. *	2007	PSB, meshes from various sources	No	Yes	No	No	Elevation descriptor calculated on concentric circular stripes of views	Funkhouser et al. (2003), Osada et al. (2002)	Mesh
Siddiqi et al. *	2008	MSB	Yes	No	No	No	Medial surfaces, bipartite graph matching	Osada et al. (2002), Kazhdan et al. (2003)	Mesh
Daras and Axenopoulos	2009	ITI, PSB, ESB	No	Yes	No	No	Fourier, Zernike, Krawtchouk	Vranic (2004), Ohbuchi et al. (2008)	Mesh
Furuya and Ohbuchi	2009	PSB, MSB, ESB	No	Yes	Yes, dense points	Yes	SIFT descriptor	Chen et al. (2003), Kazhdan et al. (2003)	Mesh
Tierny et al.	2009	SHREC'07 partial	Yes	No	No	No	RG	Biasotti et al. (2006), Cornea et al. (2005)	Mesh
Agathos et al.	2010	MSB, ISDB	Yes	No	Yes	No	ARG, EMD	Papadakis et al. (2008), Kim et al. (2004)	Mesh
Stavropoulos et al.	2010	SHREC'07 watertight, PSB	No	Yes	Yes/method of Hoffman and Singh (1997)	No	Hierarchical search of camera parameters	Germann et al. (2007)	Range image
Papadakis et al. *	2010	PSB, ESB, SHREC'07 watertight	No	Yes	No	No	Fourier and Wavelet of cylind. proj. of pos. and orientation	Papadakis et al. (2008), Vranic (2005)	Mesh
Demirci et al. *	2010	Authors' database	Yes	No	No	No	Shock graphs, EMD-based matching	-	Mesh
Adan et al.	2011	Mian et al. (2006)	No	Yes	Yes	No	Depth gradient images for *scene* recognition	Johnson and Hebert (1999)	Mesh
Attene et Al. *	2011	No standard benchmarks	No	No	Yes	No	Fast reject schema employing onion descriptors	Gal et al. (2006)	Mesh
Bronstein et al. *	2011	SHREC'10 large-scale	No	No	Yes/dense points/ Harris/ MeshDoG	Yes	HKS	Toldo et al. (2009), Lian et al. (2010)	Mesh, point cloud
Lavoue	2012	SHREC'07 Partial	No	No	Yes/dense points	Yes	Fourier spectra of local patches	Tierny et al. (2009), Toldo et al.(2009)	Mesh
Ohkita et al. *	2012	MSB, PSB	No	No	Yes/randomly extracted dense points	Yes	Local stat. feat. using distances and orientations	Kazhdan et al. (2003), Furuya et al. (2009)	Oriented point sets
Sfikas et al. *	2012	TOSCA, SHREC'07 watertight	Yes	No	No	No	Graph-based, conformal factors	Chen et al. (2003), Kazhdan et al. (2003)	Mesh
Li et al. *	2013	PSB, GSB	No	Yes	Yes/topologically salient points	Yes	SIFT, multi-resolutional Reeb graph, shell-sectors	Chen et al. (2003), Papadakis et al. (2010)	Mesh
Atmosukarto and Shapiro *	2013	PSB,SHREC'08 watertight	No	Yes	Yes/salient points	No	Silhouette-based descriptors derived on salient views	Chen et al. (2003)	Mesh

Table 6. Summary of STAR on methods related to partial 3D object retrieval

*: Methods originally applied for 3D object retrieval with complete queries

B3 LOCAL SHAPE DESCRIPTORS FOR UNSTRUCTURED 3D DATA

As Bronstein et al.[109] note, local feature-based methods are less common in the shape analysis community than in computer vision, as there is nothing equivalent to a robust feature descriptor like SIFT[116] to be universally adopted. However, there has been a considerable amount of recent research in shape descriptors derived from unstructured 3D data: starting from the spin-images introduced by Johnson and Hebert[136], which is a classical example of local 3D feature somehow analogous to SIFT, one can point out, among others, the point feature histogram (PFH)[143] and the normal aligned radial features (NARF)[144].

5.1. B3.1 Methods

B3.1.1 Spin-images

One of the most popular local 3D shape descriptors is the spin-image[136], which has been widely applied on both structured and unstructured data. A spin-image of an oriented point is a 2D representation of its surrounding surface, which is constructed on a pose-invariant 2D coordinate system by accumulating the coordinates of neighboring points (Figure 52). Two cylindrical coordinates can be defined with respect to an oriented point: the radial coordinate *a*, defined as the perpendicular distance to the line through the surface normal, and the elevation coordinate *b*, defined as the signed perpendicular distance to the tangent plane defined by vertex normal and position. The cylindrical angular coordinate is omitted because it cannot be defined robustly and unambiguously on planar surfaces. The spin-image is invariant to rigid transformations, since it encodes the coordinates of points on the surface of an object with respect to a local basis.

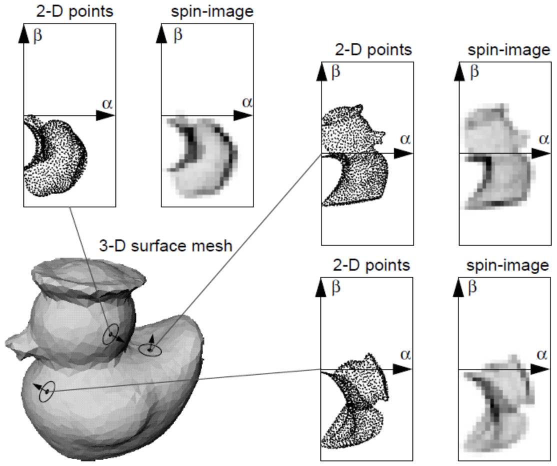


Figure 52. Original spin-images as introduced by Johnson and Hebert[136].

Some variants of spin-images[145] [146] have been proposed to enhance the discrimination capability of the original descriptor. Endres et al.[145] enhanced the comparison by applying nearest neighbor search

[143] Rusu, R.B., Blodow, N., Marton, Z.C., Beetz, M., 2008. Aligning point cloud views using persistent feature histograms. Proc. IEEE/RSJ IROS.

[144] Steder, B., Grisetti, G., Van Loock, M., Burgard, W., 2009. Robust online model-based object detection from range images. Proc. IEEE/RSJ IROS.

[145] Endres, F., Plagemann, C., Stachniss, C., Burgard, W., 2009. Unsupervised discovery of object classes from range data using latent Dirichlet allocation. Proc. Robotics: Science and Systems.

using the linear correlation coefficient. To efficiently perform the comparison of features, the authors furthermore compress the descriptor. Steder et al.[147] have shown that range value patches showed a better reliability in an object recognition system when compared to spin-images. Spin-images also do not explicitly take empty space into account. For example, in the case of a square plane, the spin-images for points in the center and for corners would be identical.

B3.1.2 3D object retrieval based on multiple orientation depth Fourier transform

Ohbuchi et al.[148] proposed an appearance-based 3D object retrieval method, using the multiple orientation depth Fourier transform descriptor (MODFD), in which model projections from 42 viewpoints are employed so as to cover all possible view aspects. Their method calculates the Fourier transform of the polar mapping of the depth buffer image and employs a shape similarity comparison algorithm. This algorithm is designed for ill-defined model representations, most notable of which is the polygon soup.

Figure 53 illustrates the steps taken to compute MODFD. The comparison of shape features is performed by using the 2D image similarity comparison algorithm by Zhang and Lu[149]. Retrieval experiments on 1213 3D models showed that the method performed quite well, despite its simple, brute force strategy. Moreover, it outperformed the D2 shape function by Osada et al.[53], as well as two methods previously published by some of the authors[150] [151], whereas it is roughly comparable to their most recent -at the time- method[152], which was based on 3D alpha-shapes.

[146] Drost, B., Ulrich, M., Navab, N., Ilic, S., 2010. Model globally, match locally: Efficient and robust 3d object recognition. Proc IEEE CVPR.

[147] Steder, B., Grisetti, G., Van Loock, M., Burgard, W., 2009. Robust online model-based object detection from range images. Proc. IEEE/RSJ IROS.

[148] Ohbuchi, R., Nakazawa, M., Takei, T., 2003. Retrieving 3D shapes based on their appearance. Proc. MIR, ACM, 39–45.

[149] Zhang D.S., Lu G., 2002. Shape-based image retrieval using generic Fourier descriptor, Signal Processing: Image Communication 17 (10), 825-848.

[150] Ohbuchi, R., Otagiri, T., Ibato, M., Takei, T., 2002. Shape-similarity search of three-dimensional models using parameterized statistics, Proc. Pacific Graphics, 265-274.

[151] Ohbuchi, R., Minamitani, T., Takei, T., 2003. Shape-similarity search of 3D models by using enhanced shape functions. Proc. TPCG.

[152] Ohbuchi, R., Takei, T., 2003. Shape-similarity comparison of 3D models using alpha shapes. Proc. Pacific Graphics.

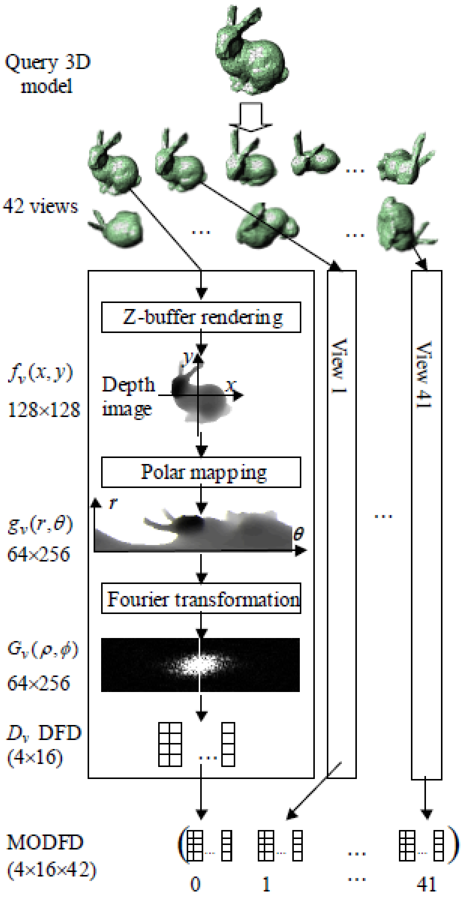


Figure 53. MODFD computational steps[148].

B3.1.3 Point feature histograms (PFH)

PFH have been proposed by Rusu et al.[143] as pose-invariant local features, aiming to represent the underlying surface model properties at a point p. Their computation is based on the combination of certain geometrical relations between k nearest neighbors. They incorporate (x,y,z) 3D point coordinates and estimated surface normals (nx,ny,nz), but are extensible to the use of other properties such as curvature and second order moment invariants. A PFH is computed as follows: 1) for each point p, all of its neighbors enclosed in the sphere with a given radius r are selected (k-neighborhood), 2) for every pair of points p_i and p_j $(i \neq j)$ in the k-neighborhood of p and their estimated normals n_i and n_j (p_i being the point with a smaller angle between its associated normal and the line connecting the points), a Darboux uvn frame $(u = n_i, v = (p_j - p_i) \times u, n = u \times v)$ is defined and the angular variations of n_i and n_j are computed as follows:

$$a = v \cdot n_j$$

$$\varphi = v \cdot (p_j - p_i))/||p_j - p_i||$$

$$\theta = \arctan(w \cdot n_j, u \cdot n_j)$$

Besides these three features, Rusu et al. [143] have considered using a fourth one characterizing the Euclidean distance from p_i to p_j. However, it has been later shown that its exclusion from the PFH introduces no significant decrease in robustness, especially when computed in 2.5D datasets, where the distance between neighboring points increases as we move away from the viewpoint. For these scans, where the local point density influences this feature dimension, omitting the fourth feature value has been proved beneficial.

Figure 54 presents an influence region diagram of the PFH computation for a query point p_q. p_q is marked with red and placed in the middle of a sphere with radius r, and all its k-neighbors are fully interconnected in a mesh.

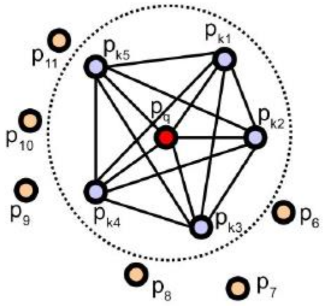


Figure 54. The influence region diagram for a PFH. The query point (red) and its k-neighbors (blue) are fully interconnected in a mesh[154]*.*

Points with histograms which are considerably dominant in the dataset tend to be less distinctive. *Persistence analysis* is performed in order to identify salient histograms for every scale and cope with this issue. The PFH selection criterion at a given scale is motivated by the fact that in a given metric space, one can compute the distances from the mean PFH of a dataset to its features. As shown by Rusu et al., this distance distribution can be approximated with a Gaussian. Using simple statistical heuristics, features which exceed a typical range can be marked as salient. To account for density variations but also different scales, the above is repeated over a discrete scaling interval and points which are marked as salient over the entire interval are marked as persistent. In particular, a point p is persistent if: (i) its PFH is selected as unique with respect to a given radius, and (ii) its PFH is selected in both r_i and r_{i+1}, that is:

$$P_f = \bigcup_{i=1}^{n-1} [P_{f_i} \cap P_{f_{i+1}}]$$

where P_{f_i} represents the set of points which are selected as unique for a given radius r_i.

Figure 55 presents the PFH signatures for points lying on 5 different convex surfaces. The confusion matrix in the figure represents the distances between the mean histograms of the different shapes, which are obtained using the histogram intersection kernel[153]:

[153] Barla, A., Odone, F., Verri, A., 2003. Histogram intersection kernel for image classification. Proc. IEEE ICIP 3, 513-516.

$$d(PFH_{\mu 1}, PFH_{\mu 2}) = \sum_{i=1}^{nr_{bins}} \min(PFH_{\mu 1}^{i}, PFH_{\mu 2}^{i})$$

It can be observed that PFHs are informative enough to differentiate between points lying on different surfaces.

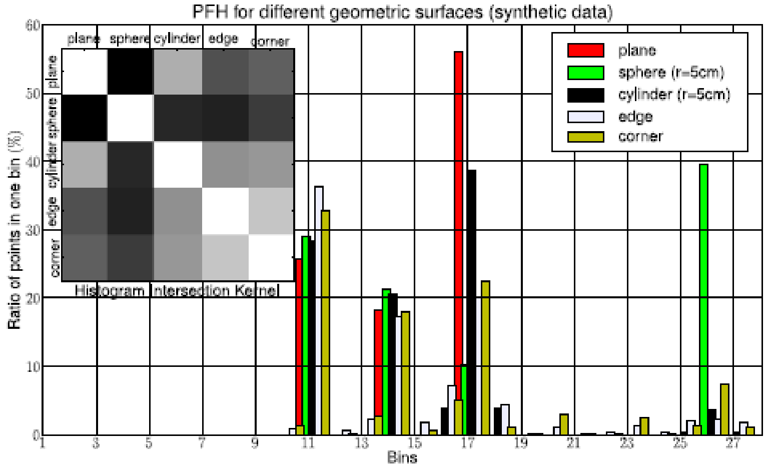


Figure 55. Example of point feature histograms for points lying on primitive 3D geometric surfaces[154].

PFH has been reported as being several orders of magnitude slower than normalized aligned radial features (NARF)[144]. Fast point feature histogram (FPFH) has been proposed by Rusu et al.[154] in order to accelerate PFH computations by employing a subset of neighboring points for histogram calculation. The theoretical computational complexity of PFH for a given point cloud P with n points is $O(n, k^2)$, where k is the number of neighbors for each point p in P. FPFH retains most of the discriminating power of PFH and has been shown to outperform spin-images in the context of registration.

An influence region diagram illustrating the FPFH computation is presented in Figure 56. For a given query point p_q, its SPFH values are first estimated by creating pairs between itself and its neighbors. This is repeated for all the points in the dataset and then the SPFH values of p_k are re-weighted using the SPFH values of its neighbors, in order to create the FPFH for p_q. As shown in the figure, some of the value pairs, which are marked with “2”, will be counted twice. The differences between PFH and FPFH are: (i) the FPFH does not fully interconnect all neighbors of p_q, and is thus missing some value pairs, which might contribute to capture the geometry around p_q, (ii) the PFH models a precisely determined surface around p_q, while the FPFH includes additional point pairs outside the r radius sphere (though at most 2^r away), and (iii) because of the re-weighting scheme, the FPFH combines SPFH values and recaptures some of the point neighboring value pairs.

[154] Rusu, R.B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (FPFH) for 3D registration. Proc. IEEE ICRA, 3212-3217.

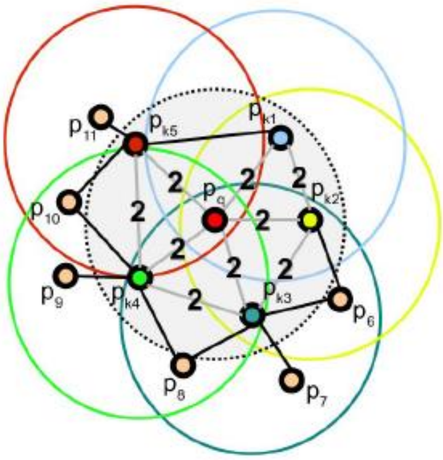


Figure 56. The influence region diagram for FPFH. Each query point (red) is connected only to its direct k-neighbors (enclosed by the gray circle). Each direct neighbor is connected to its own neighbors and the resulting histograms are weighted together with the histogram of the query point to form the FPFH. The connections marked with "2" will contribute to the FPFH twice[154].

A further PFH optimization can be pursued if the correlation problem is tackled in the feature histogram space. So far, the resulting number of histogram bins was given by q^d, where q is the number of subdivision intervals in the value range of a feature and d is the number of features selected. This can be described as a subdivided cube, where one subdivision cell corresponds to a point having certain values for its three features, hence leading to a fully correlated feature space. However, this leads to a histogram containing several zero values (Figure 55), and thus contributes to a certain degree of information redundancy. A simplification of the above is to de-correlate the values, by simply creating d separate feature histograms, one for each feature dimension and concatenate them together.

B3.1.4 Normal aligned radial feature (NARF)

NARF has been introduced by Steder et al.[144] as an interest point extraction method, along with a feature descriptor in 3D range data. The interest point extraction method has been designed with two specific goals: (i) the selected points are supposed to be in positions where the surface is stable, so as to ensure a robust estimation of the normal, and where there are sufficient changes in the immediate vicinity, (ii) *the outer shapes of objects as seen from a certain perspective are used*, considering that the focus is on partial views. The outer forms are often rather unique so that their explicit use in the interest point extraction and the descriptor calculation increase overall process robustness. For this purpose, NARFs are accompanied by a border extraction method.

Explicit handling of borders in the range data is an important requirement in the employed feature extraction procedure. Borders typically appear as non-continuous traversals from foreground to background. In this light, there are mainly three different types of points to be detected: (i) object borders, which are the outermost visible points of the object, (ii) shadow borders, which are points in the background that adjoin occlusions, and (iii) veil points, which are interpolated points between the obstacle border and the shadow border. Veil points are a typical phenomenon in 3D range data obtained by means of light detection and ranging (LIDAR) technology. Figure 57 shows an example of the different types of border points. To understand why these types of points are relevant, consider a square planar patch in a 3D range scan. The surface of the patch obviously does not provide interest points by itself. On the other hand the four corners appear as useful interest points, whereas the points on the shadow border in the background are not useful. The detection of the veil points is important, since they should be discarded from the feature extraction process.

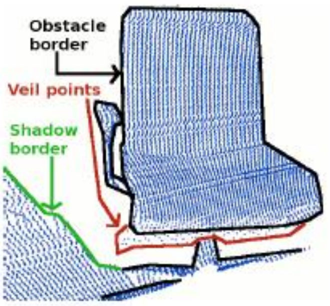


Figure 57. Different types of border points[144].

Border extraction in a range image can be benefitted by different indicators, like acute impact angles or changes of the normals. It has been experimentally observed that the most significant indicator, which is also very robust against noise and changes in resolution, is a change in the distance between neighboring points. This observation is used to classify borders by considering the neighborhood of every image point for: (i) employing a heuristic to find the typical 3D distance to neighboring points that are not across a border, (ii) using this information to calculate the probability that this point is part of a border, (iii) identifying the class of this border point, and (iv) performing non-maximum suppression to find the exact border position.

The detection of interest points is an important step to reduce the search space for feature extraction and focus the attention on informative structures. Steder et al. posed the following requirements for their interest point extraction procedure: (i) it must take information on borders and the surface structure into account, (ii) it must locate points that can be reliably detected even if the object is observed from another perspective, and (iii) the located points must be on positions that provide stable areas for normal estimation or the descriptor calculation in general.

Stable interest points need significant changes of the surface in a local neighborhood to be robustly detected in the same place, even if observed from different perspectives. This typically means that there are substantially different dominant directions of the surface changes in the area. To capture this, Steder et al.: (i) look at the local neighborhood of every image point and quantify surface changes at this position, as well as determine a dominant direction for this change, (ii) look at the dominant directions in the surrounding of each image point and calculate an interest value that quantifies the variability of these directions, as well as the surface changes near the point itself, (iii) perform smoothing on the interest values, and (iv) perform non-maximum suppression to find the final interest points.

The most important parameter of this process is the support size σ, which is the diameter of the sphere around the interest point, including all points whose dominant directions were used for the calculation of the interest value. This is the same value that is used to determine which points will be considered in the calculation of the descriptor.

Feature descriptors represent the area around an interest point in a way that facilitates efficient comparison regarding similarity. The main considerations for the NARF descriptor were: (i) capturing the existence of occupied and free space, so that parts on the surface and also the outer shape of an object can be described, (ii) robustness against noise on the interest point position, and (iii) capability of extraction of a unique local coordinate frame at the point.

The NARF descriptor enables the extraction of a unique orientation around the normal. The underlying idea is similar to what is done in SIFT and SURF[155]. Yet, unlike its 2D siblings, this orientation along with the normal defines a complete 6DOF transformation in the interest point. To

[155] Bay, H., Tuytelaars, T., Van Gool, L., 2006. SURF: Speeded Up Robust Features. Proc. ECCV, 404-417.

compute the NARF descriptor in an interest point: (i) a normal aligned range value patch is calculated in the point, which is a small range image with the observer looking at the point along the normal, (ii) a star pattern is overlaid onto this patch, where each beam corresponds to a value in the final descriptor, capturing intensity changes under the beam, (iii) a unique orientation from the descriptor is extracted, and (iv) the descriptor is shifted according to this value in order to obtain rotation invariance. The last two steps are optional, as explained above. Figure 58 visualizes an example of this process.

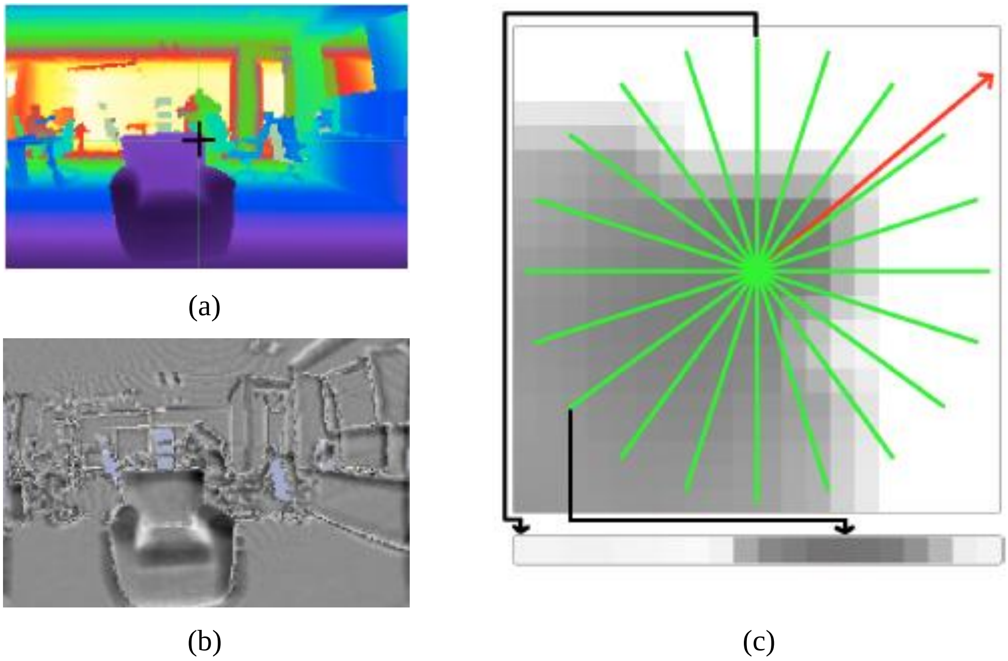


Figure 58. (a) A range image of an example scene with an armchair in the front. The black cross marks the position of an interest point, (b) visualization of descriptor calculation. The top shows a range value patch of the top right corner of the armchair. The actual descriptor is visualized on the bottom. Each of the 20 cells of the descriptor corresponds to one of the beams (green) visualized in the patch, with two of the correspondences marked with arrows. The additional (red) arrow pointing to the top right shows the extracted dominant orientation, (c) the descriptor distances to every other point in the scene, represented by grayscale intensity. Note that mainly top right rectangular corners get low values[144].

Though very successful, NARF fails to capture important cues for object recognition, such as size and internal edges[156].

B3.1.5 Depth kernels

Kernel descriptors provide a principled way to turn any pixel attribute to patch-level features and are able to generate rich features from various recognition cues. They were initially proposed for RGB images, and have not been used for depth maps and point clouds until the work of Bo et al.[156]. The match kernel framework of Bo et al. involves the following main steps: (i) define pixel attributes, (ii) design match kernels to measure the similarities of image patches based on the defined attributes, and (iii) determine approximate, low dimensional match kernels. The third step is done automatically by learning low dimensional representations and the defined kernels, whereas the first two steps make the

[156] Bo, L., Ren, X., Fox, D., 2011. Depth kernel descriptors for object recognition. Proc. IEEE/RSJ IROS, 821-826.

method applicable to a specific recognition scenario. Besides using gradient and local binary patterns in their framework, the authors have developed another three depth kernel descriptors, namely size, PCA and spin. These descriptors capture diverse yet complementary cues and their combination enhances object recognition accuracy.

The use of a size kernel can be justified by the observation that for category recognition, sizes of objects in the same category usually are constrained to some range. For example, the physical size feature can be expected to obtain perfect recognition performance in distinguishing between the apple and cap categories (Figure 59).

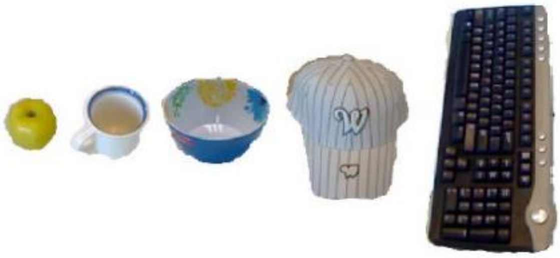


Figure 59. Sampled objects from the RGB-D dataset that are sorted according to their sizes. From left to right: apple, coffee mug, bowl, cap and keyboard[156].

In order to develop size features, depth images are first converted to point clouds by mapping each pixel into its corresponding 3D coordinate vector. The distance between each point and the reference point of the point cloud is computed, so as to capture the size cue of an object. Let P denote a point cloud and $\bar{p}$ be a respective reference point. Then the distance attribute of a point $p \in P$ is given by $d_p = ||p - \bar{p}||^2$. To compute the similarity between the distance attributes of two point clouds P and Q, the following match kernel is introduced:

$$K_{size}(P, Q) = \sum_{p \in P} \sum_{q \in Q} k_{size}(d_p, d_q)$$

where $k_{size}(d_p, d_q) = \exp(-\gamma_s \left| \left| d_p - d_q \right| \right|) \; (\gamma_s > 0)$ is a Gaussian kernel function. It can be observed that the match kernel k_{size} computes the similarity of the two sets P and Q by aggregating all distance attribute pairs.

The introduction of the Gaussian kernel results in an infinite dimensionality of the feature vector over P. Accordingly, P is projected to a set of finite basis vectors, leading to the finite-dimensional kernel descriptor:

$$F_{size}^{e}(P) = \sum_{t=1}^{b_s} a_t^e \sum_{p \in P} k_{size}\,(d_p, u_t)$$

where u_t are basis vectors drawn uniformly from the support region of distance attributes, b_s is the number of basis vector, and $\{a^e\}_{e=1}^{E}$ are the top E eigenvectors computed from kernel principal component analysis.

Another kernel descriptor employed by Bo et al. is kernel PCA. Figure 60 demonstrates the discriminative capability of kernel PCA features with an example of histograms of the top ten eigenvalues of kernel matrices formed over point clouds of two different objects. It is evident that the distributions of eigenvalues are very different, suggesting that eigenvalues can be used as 3D shape features. By evaluating kernel matrix K_P over the point cloud P and computing its top L eigenvalues, we obtain the local kernel PCA feature $\lambda_P^1, \ldots, \lambda_P^l, \ldots, \lambda_P^L$ with:

$$F_{size}^{e}(P) = \sum_{t=1}^{b_s} a_t^e \sum_{p \in P} k_{size}\,(d_p, u_t)$$

where v^l are eigenvectors, L is the dimensionality of local kernel PCA features, and $K_P[s,t] = \exp(-\gamma_k||s-t||^2)$ with $\gamma_k > 0$ and $s, t \in P$.

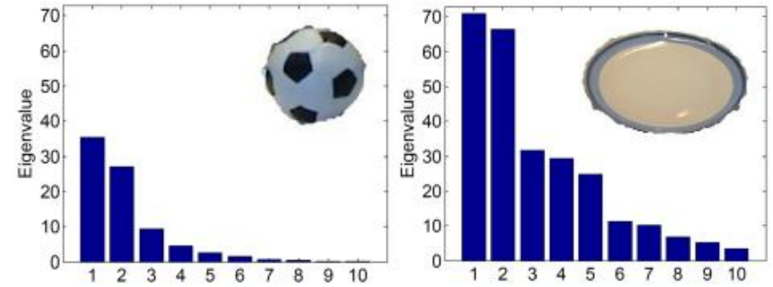


Figure 60. Top ten eigenvalues of kernel matrices formed by ball and plate[156].

A variant of the previously described spin-image[136] can also be used in the context of depth kernels, considering the angle between the normals of the reference point and its nearby points, as described by Endres et al.[145]. The elevation coordinate n_P, the radial coordinate ζ_P and the angle between normals β_P are aggregated into local shape features by defining the following match kernel:

$$K_{spin}(P,Q) = \sum_{p \in P} \sum_{q \in Q} k_a\left(\bar{\beta}_P, \bar{\beta}_q\right) k_{spin}([n_P, \zeta_P], [n_q, \zeta_q])$$

where $\bar{\beta}_P = [\sin(\beta_{\Pi}), \cos(\beta_{\Pi})]$ and P is the set of nearby points around the reference point $\bar{p}$. Gaussian kernels k_a and k_{spin} measure feature similarity. The spin kernel descriptor can be extracted from the spin match kernel by projecting the infinite feature vector to a set of finite basis vectors, in a similar fashion to the size kernel descriptor.

Gradient and local binary pattern kernel descriptors can be directly applied to depth maps in order to extract edge cues. The gradient match kernel k_{grad} is constructed from the pixel gradient attribute:

$$K_{grad}(P,Q) = \sum_{p \in P} \overline{m}_P \overline{m}_q k_o\left(\tilde{\theta}_P, \tilde{\theta}_q\right) k_S(p,q)$$

where P and Q are image patches from different images, and $p \in P$ is the normalized 2D pixel position. θ_P, m_P are the orientation and magnitude of the depth gradient at a pixel p. The normalized linear kernel $\tilde{m}_p\,\tilde{m}_q$ weighs the contribution of each gradient where $\tilde{m}_p = m_p / (\sum_{\mathrm{p} \in \mathrm{P}} m_P^2 + \varepsilon_g)^{1/2}$ and ε_g is a small positive constant to ensure that the denominator is larger than 0; the position Gaussian kernel $K_s[p,q] = \exp(-\gamma_s||p-q||^2)$ measures the spatial proximity of two pixels; the orientation kernel $K_o\left[\tilde{\theta}_p, \tilde{\theta}_q\right] = \exp(-\gamma_o||\tilde{\theta}_p - \tilde{\theta}_q||^2)$ computes the similarity of gradient orientations where $\tilde{\theta}_p = [\sin(\theta_P), \cos(\theta_P)]$. The local binary kernel descriptor K_{lbp} is developed from the local binary pattern attribute:

$$K_{lbp}(P,Q) = \sum_{p \in P} \tilde{s}_p \tilde{s}_q k_b(b_p, b_q) k_s(p,q)$$

5.2.

5.3. B3.2 Comparative analysis on local shape descriptors for unstructured data

Table 7 summarizes the STAR on local shape descriptors for unstructured data. It can be observed that three out of five methods calculate and compare histograms over a neighborhood centered at each point of interest. In this respect, the depth kernel method of Bo et al.[156] provides a framework for embedding the previous histogram-based local shape descriptors, namely spin-images, PFH and FPFH. Actually, spin-images have been already used in the original depth kernel paper, resulting in enhanced retrieval accuracy.

Another dominant attribute which is evident in Table 7 is the use of point distances within a neighborhood. This type of information is the closest one to raw point coordinates provided in point cloud input data. On the other hand, surface normals, which are used either for the formation of local projections (spin-images and NARF) or for the calculation of the feature descriptor (PFH and FPFH) ask for a normal estimation algorithm, such as the one proposed by Rusu[157]. In addition, three of the methods are 2.5D, calculating descriptors over 2D projections. It is clear that in the case of the local projections used in spin-images and NARF, the descriptive capability obtained is connected with the quality of the preceding normal estimation.

Finally, it can be noted that with the exception of MODFD proposed by Ohbuchi et al.[148], the rest of the local shape descriptors presented were not initially proposed for 3D object retrieval. The most usual application addressed is object recognition in scenes, whereas PFH was originally proposed for registration. In principle, there is nothing prohibitive in applying the same descriptors for 3D object retrieval. However, it should be pointed out that the interest point detector which precedes the calculation of NARF is formulated so as to cope with the presence of veil points and shadows, which are typical attributes of a scene. An alternative interest point detector seems a prerequisite for the use of NARF within a 3D object retrieval context.

[157] Rusu, R.B., 2009. Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments, PhD thesis, Dept. of Computer Science, Technical University of Munich, Germany.

Method	Year	Histogram-based	Information used	2.5D	Methodology	Original application	Outperformed
Johnson and Hebert	1999	Yes	Point distances and normals	Yes	Spin-images	Object recognition in scenes	
Ohbuchi et al.	2003	No	Fourier transform of depth buffers	Yes	MODFD	3D object retrieval	Osada et al. (2002), Ohbuchi et al. (2002,2003)
Rusu et al.	2008	Yes	Point distances	No	PFH, FPFH	Registration	Johnson and Hebert (1999)
Steder et al.	2009	No	Intensity gradients	Yes	NARF	Object recognition in scenes	Lai et al. (2011)
Bo et al.	2011	Yes	Depends on the used features	Depends on the used features	Depth kernels	Object recognition	Johnson and Hebert (1999), Steder et al. (2009)

Table 7. Summary of STAR on local shape descriptors for unstructured data

2. Acknowledgements

We would like to thank Dr. Dirk Rieke-Zapp for providing valuable input with respect to the scanning technologies employed by Breuckman GmbH.

Disclaimer

Any mention of commercial products or reference to commercial organisations is for information only; it does not imply recommendation or endorsement by the authors nor does it imply that the products mentioned are necessarily the best available for the purpose.