The Publication of 3D-Datasets for Interactive Access Beyond the Visualization Techniques

Gerd Kaupp, Juergen Boy, and Andreas Herrmann

Abstract-- **Data interaction is essential in all sciences. 3D data exhibit highest relevance in all fields. They are analyzed and treated with proprietary software at the site of production. However, these data are still mutilated into dead 2D-projections or at best into dead perspective snap-shots in conventional electronic publications, whereupon all 3D-information is lost. Thus, published data cannot be scrutinized, analyzed, or reinterpreted. State of the art journals are or should be threefold: paper, fulltext with multimedia, and enhancements by interactive original data. The proprietary 3D-data of hopefully all major brands of high-tech equipment shall be converted to a common exchange format and stored and held out by the publishing medium for data-interactive use. As all present VRML viewers are designed for visualizations but not for calculations, VRML shall be the broadest generally accepted exchange format for import to proprietary on-site software, as long as platform-independent analyzing 3D-viewers are lacking. For example, supermicroscopic surface data that were obtained with Digital Instruments, or Danish Microengineering, or Seiko, or Thermo, or self-built, etc. instruments should be mutually readable for verification scrutiny, or analyses, or reinterpretation. Thus, the international data exchange will be reached and real data mining will be granted to future generations. We demonstrate the feasibility with regularly published 3D-data from different vendors.**

Index Terms--**3D-data, communication, data exchange, data-interactive publishing, proprietary software, VRML**

I. INTRODUCTION

A new culture for publication of scientific results will be reached if authors and publishers start to disclose full scientific data and thus avoid costly losses just by the common publishing in print and electronic full text including colorful though dead TIFF/GIF/JPEG images. Such endeavor has not been taken up prior to the granting of funds in the GLOBAL INFO project of the German Ministry of Education and Research (BMBF) after a rather lengthy application procedure [1]. Clearly, an enormous leap will be achieved by the added value of data-interaction, far beyond the present upheaval in scientific communication that has been repeatedly claimed. The data that should be disclosed are created and available at every high-tech data collection setup at least in the natural, medicinal and technical sciences and these are handled on-site by the scientific and technical staff with full data-interaction. The major needs for data-interactive publishing in chemistry, as a representative natural science, are in the fields of fully interactive hypermolecules and crystal structures, of hyperspectra, chemometric data sets, and complex 3D-objects [2]. The preferable publication and data formats are XML/CML [3] and widespread PDB/VRML/JCAMP-DX. The presently used TIFF/GIF/JPEG formats provide dead projections of the 3D-data. Most of the advanced formats are highly developed for molecular formula in the Internet and in databases, but rarely used in journals. Exceptionally, the Cambridge Crystallographic Database (CSD/CIF) or the Protein Database (PDB) store xyz molecular data from publications of all major journals that can be checked/analyzed with appropriate software. However, the other topics mentioned continue to require development. Hyperspectra will be implemented to CML with the software package TranSpec for use in journals and on the Internet. Chemometric data will require extensions of the CML-syntax and more advanced analyses of present data-sets will be possible in the future via common exchange formats. 3D-objects are generally imaged with high-tech equipment and the exact data are interactively analyzed with proprietary software. Most of such 3D-data continue to be published as 2D-projections, even though disclosure as xyz primary data-sets would be possible and should be performed in VRML, the exchange format that enjoys the widest acceptance in the Internet. Only such full data can be viewed with free 3D-viewers. The Internet provides numerous VRML data-sets from biology, chemistry, geography, mathematics, medicine, robotics, space missions etc. [4], and numerous data-interactive 3D-surfaces have already been published in the VRML format by the authors [5]. Both interactive Internet and publication data can be viewed with CosmoPlayer, or Blaxxun, and other viewers [6]. However, such viewers have not been designed for calculations of 3D-surfaces for analyses, scrutiny, statistics, and true data-mining. Therefore, 3D-information is still thrown away by publishing/storing in TIFF/GIF/JPEC and afterwards so-called "image analysis techniques" try to guess the data with numerous assumptions and highly questionable results, and Peer Reviewers do not care: a situation subject to

Manuscript received June 15, 2001. This work was supported by the German Ministry of Education and Research (BMBF) under Grant No. 085FL017.

G. Kaupp, J. Boy, and A. Herrmann are members of the University of Oldenburg, D-26111 Oldenburg, Germany (fax: x49 441 7983409, e-mail: kaupp@kaupp.chemie.uni-oldenburg.de).

change.

II. DISCUSSION

A. 3D-Objects

The aim of this work is to overcome the age of data-less images by disclosing data, as exemplified with 3D-data. We deal with more than multimedia: data-interaction is feasible and available at all sites of data production but the original data must be made available for world-wide exchange. We will stress 3D-objects and surfaces here because hese are of highest importance in all fields of life. Points of interest are for example molecular models (e.g. pharmaceuticals), medicinal objects, microscopic or nanoscopic objects, technical constructions (like instruments, buildings, vehicles), etc. etc. They occur as huge objects like the sun, the moon, the earth, or as biological objects like beetles and flowers and human organs, or as nanoparticles or single atoms. All of these 3D-objects can be imaged, but it's more profitable to disclose the underlying data. It really covers largest to smallest objects in science, engineering, and daily life. All fields of life and sciences are concerned from archeology to zoology. Recording instruments vary enormously. Most recently, 3D-radar, 3D-laser-, digital-camera- or mechanical arm-scanners, confocal microscopes and scanning microscopes render data-sets of interest that should be used and conserved in analyzable form but not only as dead TIFF/GIF/JPEC images.

B. Nanoscopic 3D-Surfaces

We exemplify the field with timely nanoscopic data. It is just since 1986 that the new field called nanoscience and nanotechnology emerged due to the development of scanning supermicroscopy such as scanning tunneling microscopy (STM) and atomic force microscopy (AFM). These (and other supermicroscopies) record 3D-surfaces at the nanoscopic level and at the fantastic atomic resolution. Thus, single atoms are also 3D-objects. The distance between these atoms is in the range of 1.5 Angstroms that is about the seven millionth of a millimeter. From the advent of these epoch-making events mankind badly needs supermicroscopy and nano-sciences in order to analyze submicroscopic details on all kinds of surfaces. Such new knowledge provides valuable data for improvements in all branches of daily life, be they technical, medical, sustainable, or esthetic. The importance may be exemplified at the atomic level where we may have defects and interstitial sites. Creation and repair can be analyzed in real time and much can be learned for catalysis, for example in automotive prevention of pollution, or for environmentally benign production of goods [5].

Technical applications deal with nanoparticles at the nanoscopic level of some ten and hundred nanometers. For example, the color shading of dye nanoparticles depends on their aggregation. Supermicroscopic techniques reveal the types of aggregation and help the engineers to adjust the appearance of cars and so on. Another point of broad interest is the costly corrosion of steel. Submicroscopic 3D-data reveal that the rust particles do not align along the edges from polishing but form islands. These experimental data help in corrosion prevention, etc. [5]. Valuable 3D-data are produced in all of these and related investigations. These and similar data, that are created by using high-tech equipment, deserve analyses before and after publication. It's the wrong way to image them in TIFF/GIF and lose the data altogether. But what is the problem?

C. Contemporary Publishing

Contemporary publications on paper and in electronic full text with multimedia do not disclose the data and nothing useful can be done with the reported images. It is completely unacceptable, although still in common use and not halted by the Peer Reviewers, if four gif-images each are mounted horizontally per column in the two-column print (Fig.1). They

insert Figure 1

cannot be sufficiently recognized, checked, scrutinized, analyzed, used more extensively, or interpreted differently. Possible errors may stay undetected, dead-ends might be entered. Thousands and thousands of such papers passed successfully through the Peer reviewing system. Improvement of the images does not help much. Dead TIFF/GIF images of 3D-data may be enlarged or presented in color but not viewed in various perspective views (usually, no space is available for several images of one object). They cannot be reasonably analyzed any more. Mankind cannot afford any longer the loss of all scientific data by publishing dead images instead of the data. The term multimedia is confusing in this respect: even movies are not data. The next step is due and that is publication of data in a suitable exchange format that permits the interaction with the data. Suitable data formats are available in the Internet.

D. World-wide Data Exchange

World-wide exchange of interactive 3D-data from all data producers requires a generally accepted exchange format and that is presently VRML in the Internet. Therefore, publication of 3D-data should be in VRML. Fig. 2 exemplifies the world-wide exchange of interactive data. Proprietary original formats are incompatible with each other. However, they have frequently their ASCII export function. Unfortunately these ASCII data are still not compatible. But their conversion to the VRML standard opens the possibility for viewing them interactively from all sides in every orientation with freely

insert Figure 2

available VRML viewers [6]. Thus, VRML data files from everywhere can be viewed at any place on earth, if they are made publicly available in the Internet [4] or, importantly, in regular publications [5].

The scheme of Fig. 2 will be exemplified by viewing a published European data-set of an AFM surface with a freely available viewer such as Cosmoplayer, as well as by viewing, handling, and analyzing with American or Japanese on-site software. This demonstrates that published VRML data-sets disclose their full content to everybody.

E. Viewing Published VRML Data.

Fig. 3 exemplifies the viewing of VRML data from a regular publication in a modern three-fold scientific journal [5a] as an interactive image. Previously, that was only demonstrated in the Internet and in databases.

insert Figure 3

We see the perspective representation of a 3D AFM surface that is calculated from the data in [5a]. One chooses from the bibliographic details of the journal's contents page, selects enhancement (with data interaction), picks the VRML data-set for the image of interest, and views it with one of the numerous freely available VRML viewers, such as Cosmo player (all other VRML viewers work equally well). The buttons at the bottom of Fig. 3 allow for all imaginable movements over the screen and views from any direction: the 3D-surface can be rotated around the *z* axis for a view of the backside, or around any axis in space; the image can be slid and zoomed. A color height-coding was added in our applications for a better appearance. Also, the VRML-enhancement of the NIST program to show and analyze orthogonal cross-sections of mathematical functions [7] works for the natural surfaces. Clearly, we have data-interaction and may view such 3D-data by imaging of the data sets from all vendors that provide VRML export allowing authors to disclose their data in modern journals or in their home pages. The needs of the non-specialized reader will often be served by these facilities.

However, viewing is not enough and unfortunately VRML viewers were not designed for more than that. What we badly need is the full calculation capability of 3D-data that is at present only available in the proprietary software of the vendors (or self-made software) though available at every site of data production. Thus, it is also essential to create VRML import facilities for the proprietary software systems in order to fulfill the needs of the experts.

F. VRML Data for Surface Calculations, Import to on-site Software

The scheme of Fig. 2 will now be exemplified by viewing, handling, and analyzing a published European data-set of an AFM surface with American and Japanese on-site software. The published VRML data from Danish hard- and software are imported to all three on-site software packages. Fig. 4 shows the results: the same VRML-data of our publication as in Fig. 3 were imported into Danish RasterScope 4000 (DME), NanoScope 4 (DI-VEECO), and SPI3800N (SEIKO) software. They can now be imaged, treated, and analyzed.

insert Figure 4

We achieve viewing of the surface in the different software packages. It is possible to change the z-scale, to rotate in steps of 90°, to change the pitch angle, to change the weight of height-coding and to illuminate from any direction in freely defined color programs. Even more importantly, the data from the regular publication [5a] can be treated (including various filter functions) and analyzed by using the pull-down menus of the software facilities. Fig. 5 shows a typical example. Some of the techniques are shown in this paper in order to exemplify the need for the data and their benefit and power.

insert Figure 5

Distance and angle on top-views are available by mouse click and therefore not included in Fig. 5. The use of distances and angles between points of interest is self evident and fortunately these can be directly displayed.

The "Section" feature allows for the definition of cross-sections in any direction. The various possibilities of straightforward analyses are exemplified in Fig. 6. The desired cross-section (any angle) is chosen. The left two arrows were set for distance (here reading 2.708 µm for surface distance and 2.685 µm for horizontal distance at a proper display device), the light arrows for height (here reading 134.66 nm), and the right two arrows for steepness (here reading 7.069 degree). Also a spectrum along the cross-section is displayed and analyzed in Fig. 6. The display of the four measured values of all three data-pairs and of the spectrum analysis and of freely selectable statistical features is integrated and shown in additional boxes of the original image in the NanoScope software (not shown here for space limitations). That's the way all 3D-objects are and should be geometrically and statistically analyzed. The practical applications of such data cannot be overestimated.

insert Figure 6

The "Zoom" function permits the more detailed analysis of particular areas of the surface and all techniques are applicable to the enlarged, data-interactive image section of Fig. 7. More detailed analyses are clearly possible, provided that the data are stored at high resolution with a sufficient pixel density.

insert Figure 7

The "Invert" function helps to analyze hidden parts of a surface. Thus, if we wish to look into the crater depths, that are not easily seen in Fig. 8a, we simply invert the image for that purpose: the upside down inverted image is calculated and displayed in Fig. 8b. It discloses the shapes of the depths on the 3D-surface.

insert Figure 8

Another application is provided by the "bearing" analysis if volumes of hills are to be compared with those of depressions, etc.

Furthermore, the menu for analyses in Fig. 5 contains various possibilities for statistical analyses. In particular the roughness analysis is of technical importance. Most popular is the mean square roughness *Rms*. The whole image or a square box (e.g. in Fig. 7) can be statistically analyzed just by mouse click. All requirements for further analyses are met if the original data are not lost but published in the VRML exchange format. Clearly, all aspects of surface analyses can be checked with interactive data. And finally, the quality of data will improve if they are disclosed in first class publications.

III. CONCLUSIONS

Data-interactive publishing of 3D-data has just started with five publications and more than eighty data sets. Their wealth and use has been demonstrated. The global millennium needs the data, not just images in all branches of sciences and engineering. Authors, Peer Reviewers, Publishers, and Readers will certainly go ahead and cooperate. Data must not be thrown away upon publication any longer but kept and stored with their precise meaning. Multimedia does not help, we need data-interaction. Despite the principal feasibility Authors, Publishers, and Referees did not recently ask for the new tools that avoid the loss of all data. Rather it seemed to be more popular to publish images and guess the data afterwards by using image analysis techniques for (poor) "data-mining". It is, however, self-evident that real data-mining should mine data-sets rather than images.

At present we import published VRML data to proprietary software for the mathematical or statistic analyses. Furthermore, visualization of VRML data using suitable JAVA applets will provide viewers for all platforms and analysis techniques. CML still lacks a suitable 3D-viewer, however, it enables the integration of the various exchange formats into chemistry related publications.

In the long term, a general platform-independent 3D-viewer with all calculation capabilities will be essential for the general public. Such capabilities are of utmost importance as the bad news on contemporary publication habits (Fig. 1) impressively tell us. VRML may not be the ultimate exchange format, but all data-sets that are disclosed now in VRML can be used for future exchange formats, because they give a complete description of the 3D-object. Such disclosure will and must be possible for publicly funded research. Most importantly, real data-mining in data-sets will be possible and replace the mining in dead images or the guess of data therefrom. We must start now. Any delay [1] is desastros and wastes tax money. Publishers and libraries will be urged to follow if they want to keep with the state of the art. If they don´t, first class publications will rapidly switch to the Internet and enforce the global exchange from there.

The scientific community cannot afford further costly data losses. It needs quality check and improved quality of data. Future reinterpretation or efficient data-mining are only possible on the basis of the full primary data. That feature is not restricted to 3D-data. At present, PDB, CIF, VRML, and JCAMP-DX appear to be the most promising exchange formats for the large variety of further data-types. These formats enjoy the widest global acceptance.

REFERENCES

[1] G. Kaupp, M. Haak, A. Herrmann, Interaktive Kommunikation und GraphikStandards im Internet mit Online Vorführungen, *BMBF-Workshop Informations-Vermittlung, Berlin*, 1995.

[2] G. Kaupp, M. Haak, G. Gauglitz, H.-J. Schneider, R. Moll, H. Schmelz, T. Fröhlich, A plea for publishing without loss of data, http://kaupp.chemie.uni-oldenburg/global-info/plea/

[3] P. Murray-Rust, CML: The way forward, 1998 http ://ala.vsms.nottingham.ac.uk/vsms/talks/chemweb/024.html

[4] Some Websites which provide 3D-objects as VRML files: more than 60 examples of 3D-surfaces in online publications, http://kaupp.chemie.uni-oldenburg.de/; 16 examples of 3D-Insects, www.ento.vt.edu/~sharov/3d/; Interactive animated atlas of structure and function visualizations of biological structures and processes using interactive VRML virtual space worlds, www.bioanim.com/; Explore the human body as a 3D-VRML Object, www.npac.sys.syr.edu/projects/3Dvisiblehuman/VRML/VRML2.0/MEDVIS ; Virtual Reality Models and Animations of the Pathfinder Mission, www.mars.sgi.com/vrml/vrml.html; 3D/VR Images: Visualization of Remote Sensing Data; http://rsd.gsfc.nasa.gov/rsd/; VRML and Java3d Robots at the Institute of Robotics and System Dynamics of the German Aerospace Center, www.robotic.dlr.de/Joerg.Vogel/Vrml/; US EPA GIS-VIS Integration Website, Geo-VRML Visualization: A Tool for Spatial Data Mining, www.epa.gov/gisvis

[5] Overview in http://www.data-interaction.de; Original publications of interactive 3D-data of the authors´ group are found in [2] and at:
a) http://www.wiley.com/wileychi/epoc/ public/ 1999/v12i11/
b) http://kaupp.chemie.uni-oldenburg.de/spie/
c) http://www.photobiology.com/photobiology99/contrib/ kaupp/
d) http://www.wiley.com/wileychi/epoc/public/2000/ v13i7/
e) http://www.photobiology.com/photoiupac2000/kaupp/

[6] Some popular VRML-Viewers are available: Cosmo Player (Win3.11/95/98/NT, MacOS), http://www.cosmosoftware.com; Blaxxun 3D and Blaxxun Contact (Win95/98/NT, MacOS, Linux), http://www.crc.ca/FreeWRL/; More Browsers: http://web3d.about.com/compute/web3d/msub2.htm; http://www.blaxxun.com; FreeWRL (Linux), More Information on VRML: http://home.hiway.net/~crispen/vrmlworks/

http://www.edcenter.sdsu.edu/repository/nav_vrmllinks.shtml
[7] Enhanced VRML Example of the Information Access and User Interfaces Division (IAUI) of NIST, http://zing.ncsl.nist.gov/

Gerd Kaupp is full professor in chemistry at the University of Oldenburg, Germany, since 1982. He graduated from the University of Würzburg and held various positions in USA, Switzerland and Germany. He has been using electronic databases since 1970 and is engaged in the teaching of information retrieval, performs quantum mechanical calculations and organizes joint projects in the field of data-interactive publishing, where he covers the 3D-data section. His IT work is supported by the German Ministry of Education and Research.

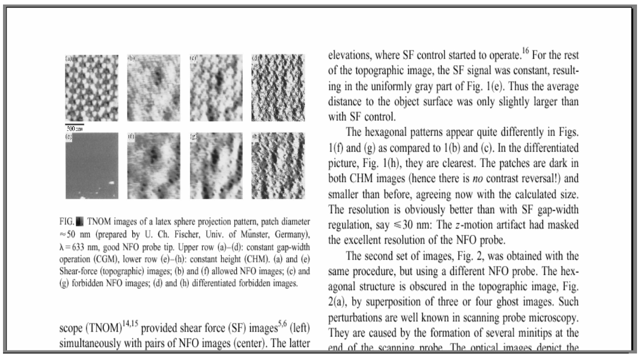

FIG.[redacted] TNOM images of a latex sphere projection pattern, patch diameter ≈50 nm (prepared by U. Ch. Fischer, Univ. of Münster, Germany), $\lambda = 633$ nm, good NFO probe tip. Upper row (a)–(d): constant gap-width operation (CGM), lower row (e)–(h): constant height (CHM). (a) and (e) Shear-force (topographic) images; (b) and (f) allowed NFO images; (c) and (g) forbidden NFO images; (d) and (h) differentiated forbidden images.

scope (TNOM)[14,15] provided shear force (SF) images[5,6] (left) simultaneously with pairs of NFO images (center). The latter

elevations, where SF control started to operate.[16] For the rest of the topographic image, the SF signal was constant, resulting in the uniformly gray part of Fig. 1(e). Thus the average distance to the object surface was only slightly larger than with SF control.

The hexagonal patterns appear quite differently in Figs. 1(f) and (g) as compared to 1(b) and (c). In the differentiated picture, Fig. 1(h), they are clearest. The patches are dark in both CHM images (hence there is *no* contrast reversal!) and smaller than before, agreeing now with the calculated size. The resolution is obviously better than with SF gap-width regulation, say $\leqslant 30$ nm: The z-motion artifact had masked the excellent resolution of the NFO probe.

The second set of images, Fig. 2, was obtained with the same procedure, but using a different NFO probe. The hexagonal structure is obscured in the topographic image, Fig. 2(a), by superposition of three or four ghost images. Such perturbations are well known in scanning probe microscopy. They are caused by the formation of several minitips at the end of the scanning probe. The optical images depict the

Fig. 1. Part of a sample page of a typical publication dealing with 3D-surfaces (on paper and fulltext with multimedia) that is unacceptable.

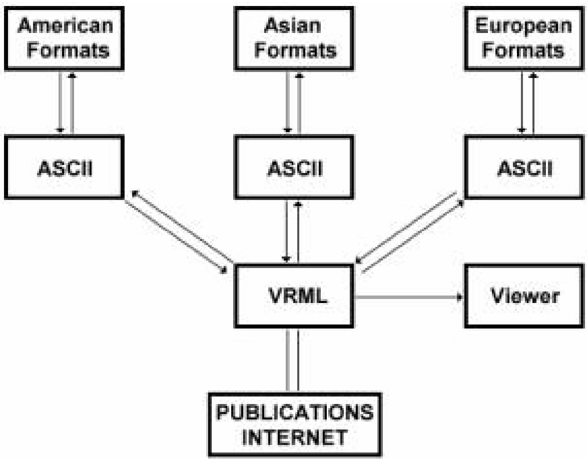

Fig 2**.** Scheme for world-wide data exchange.

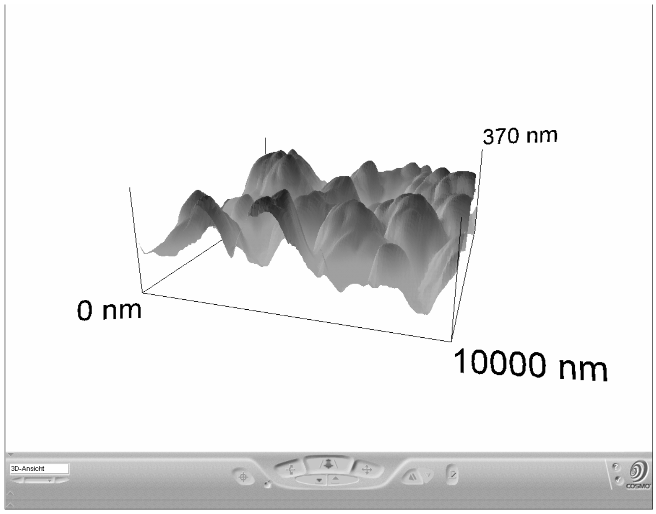

Fig. 3. Cosmoplayer screen image of a VRML data-file from [5a].

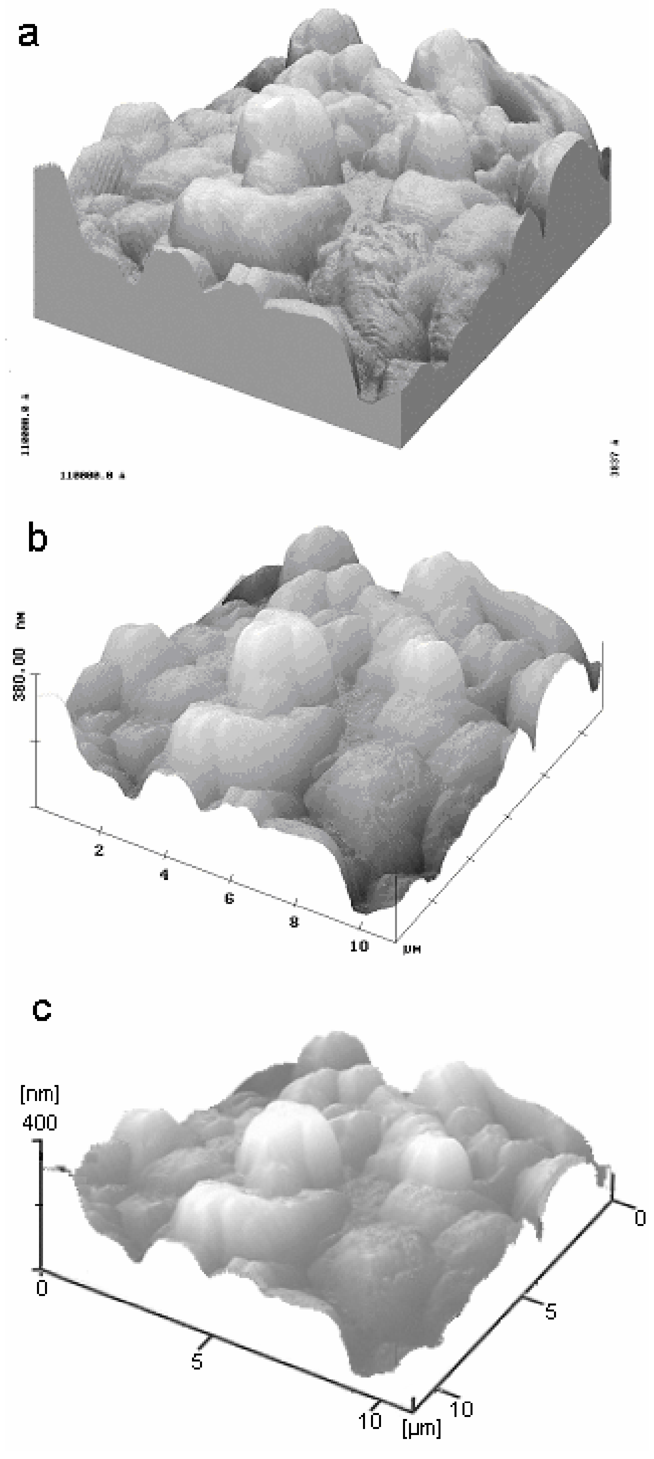

Fig. 4. VRML import from the regular publication [5a]; (a): AFM image of RASTERSCOPE 4000; (b): same image of NANOSCOPE 4; (c): same image of SPI3800N software.

Autocovariance	Grain Size	Section
Bearing	Grain Size Average	Stepheight
Bearing Compare	Particle Analysis	Auto Tip Qual
Bump Analysis	Power Spectral Density	Width
CD/DVD Analysis	PSD Compare	Zoom
Depth	Roughness	Invert

Figure 5. Typical software menu for the analysis of interactive 3D-data.

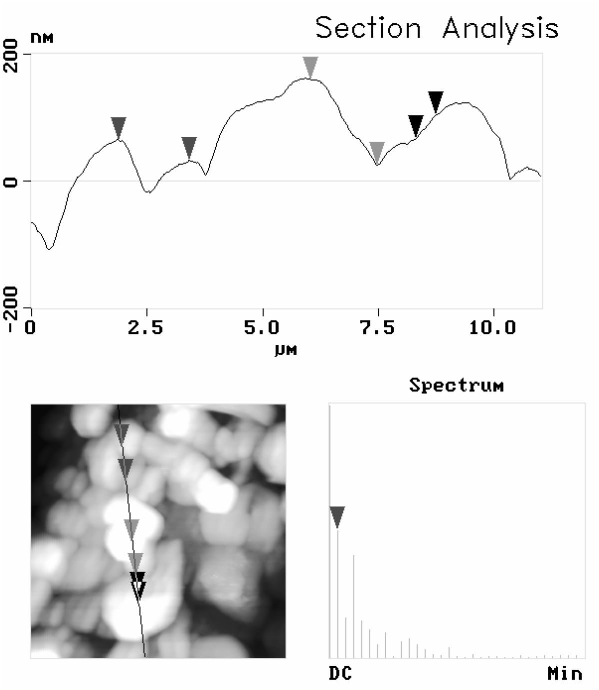

Fig 6. Section analyses of an interactive 3D AFM surface

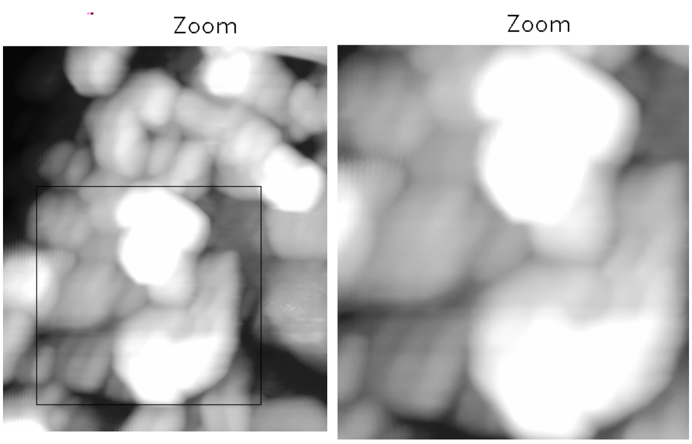

Fig 7. Zooming of 3D-images

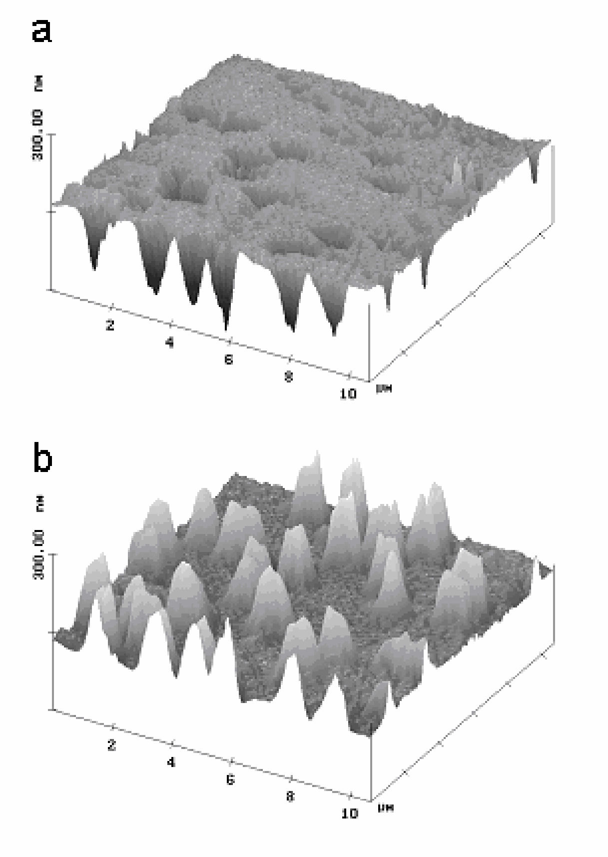

Fig. 8. AFM surface with craters; (a): original; (b): inverted image for the viewing of the crater shapes.