Digital Visual Information Processing: Adding Vision and Graphics

Franz W. Leberl
Institute for Computer Graphics and Vision
Graz University of Technology, Inffeldgasse 16, A-8010 Graz, Austria
Leberl@icg.tu-graz.ac.at

Abstract

Models of the real world are at least 3-dimensional and require the technologies of image processing and computer graphics. We see the processing of images, be they of the real world or of models in a computer, as one and the same basic technology, namely "digital visual information processing". This paper seeks to illustrate and justify this point-of-view by means of examples taken mostly from the work at the author's institute.

1. Introduction

"Graphics meets vision" is becoming an increasingly popular slogan in the literature of imaging and graphics (Computer, 1998; Solina and Leonardis, 2000). This paper is an attempt to justify a position that the boundaries between these two fields of computer vision and computer graphics are artificial, historical and should be abandoned. This argument is being supported by several examples of recent work performed at the intersection between these two traditionally separate fields. Figure 1 is a common reminder that terrain models are both created from images and visualized or rendered using images. This technology is attributed to classical photogrammetry where the view exists that the field concerns itself with "terrain images", irrespective of the application to modeling by shape from stereo, or to rendering by means of orthophotos and digital elevation models DEM.

(a) (b) (c)

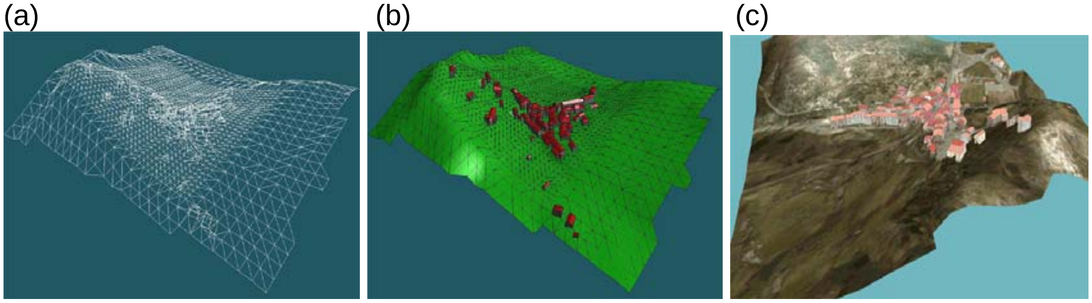
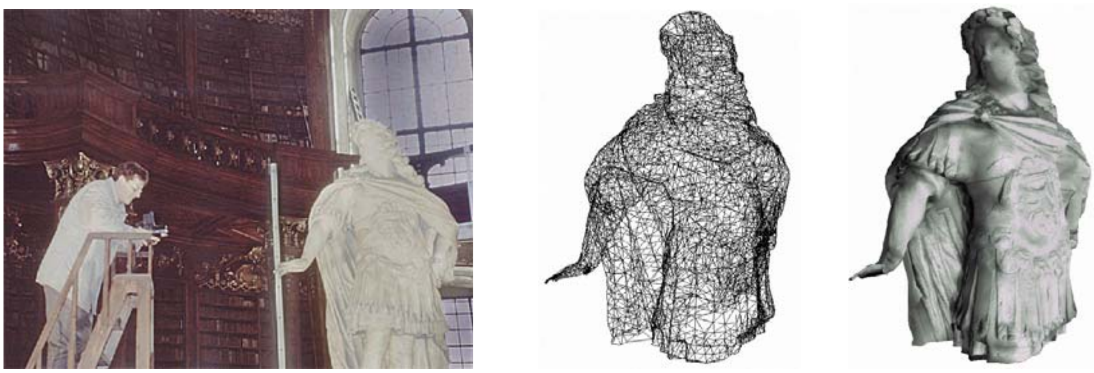


Figure 1: Modeling an alpine village in (a) by a triangulated mesh, in (b) as a flat shaded object, and in (c) as a photo textured model useful for simulation and training.

Figure 2, in analogy to Figure 1, illustrates the same issue with the help of a model of a statue of an emperor in which a point cloud was created to describe the statue's geometry. The triangulation of the points considered the image context to avoid that erroneous arcs get defined between points that are geometrically close but topologically unconnected. Finally, rendering the statue of course employs the photographic material as a major element.



(a) Acquiring the images (b) The point cloud (c) Rendering

Figure 2: Image based modeling and rendering of a statue of Habsburg Emperor Karl V from the National Library in Vienna, Austria.

We will in the following define the areas of overlap between graphics and vision and then illustrate by examples from robotics, geographic information systems, augmented reality, and as-built documentation how the separation into two fields makes little sense. We discuss one algorithmic issue surrounding texture to further make our case, and conclude with a class schedule adopted at Graz University of Technology that abandons the traditional separation into two fields.

2. Overlap between Vision and Graphics

2.1 Basing One's View on the Objects of Interest

Figure 3 presents the idea of "vision" on one hand, and "graphics" on the other, to overlap where the real world is concerned as the object of interest, and where the objects are three-dimensional. Vision of cancer cells, or graphics of virtual worlds, fall outside the overlap between vision and graphics.

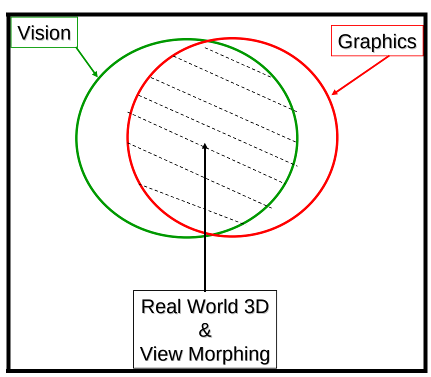


Figure 3: Where "Vision" meets "Graphics" – in modeling and rendering 3D real world objects and environments.

2.2 Basing One's View on the Process Flow

One often refers to the sequence of work steps in an extended process as a "pipeline". The image as an information source traditionally is seen at the root of the image analysis pipeline. In graphics, the image is seen as the outcome of the pipeline. Figure 4 describes the traditional view of image analysis and graphics pipelines. While they may appear separate, they are interconnected since the processes are not purely sequential. Instead they often represent a closed loop process going back from the "rendering" to "analysis", in Figure 4 from viewing and interacting with a rendered object to he object extraction. Pinz (1994) captures this closed loop in a circular model of the total pipeline, consciously abandoning the separation between graphics and vision (see also Figure 5).

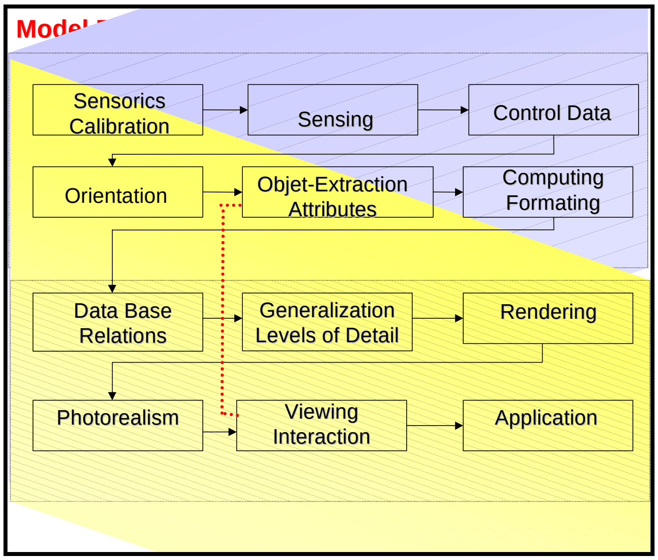


Figure 4: Sensor-based modeling of the real world in a model-building pipeline versus rendering the model in a graphics pipeline, and the need to se a closed loop approach by connecting the "viewing and interaction" step with he "object extraction and attributing" box.

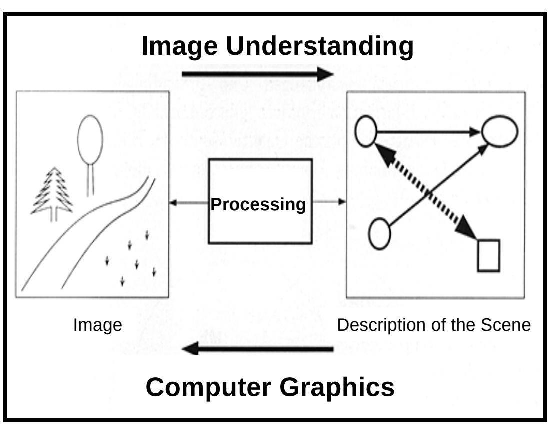


Figure 5: Relationship between vision and graphics (from Pinz, 1994)

2.3 Basing One’s View on Hardware
Of course, the view of a combined field of digital visual information processing could be discussed not only from a consideration of the objects of interest, but from the methodologies and technologies. Historically, image processing used specialized computer hardware based on raster processors, whereas graphics employed vector displays and light pens. Those differences have fallen by the way side, hardware is now identical, whether it is for graphics or for image processing.

2.4 Basing One’s View on the Mathematical and Procedural Tools
We hope to show along this paper, that even a view at the procedures and mathematics cannot justify that here are two fields, but only one. Too many issues are identical, be it transformations, resampling, color, texture or other elements of the fields.

3. A Review of Examples

3.1 Robotics
Figure 6 illustrates the operations of a robot arm used to place parts on a conveyor belt. The parts are presented the robot in a heap, typically in a box. The part is grabbed nearly randomly by the robot hand and needs now to be oriented properly for placement on the conveyor belt.

(a) Robot configuration with parts in a pile | (b) Mismatch between CAD model and image | (c) Matching complete

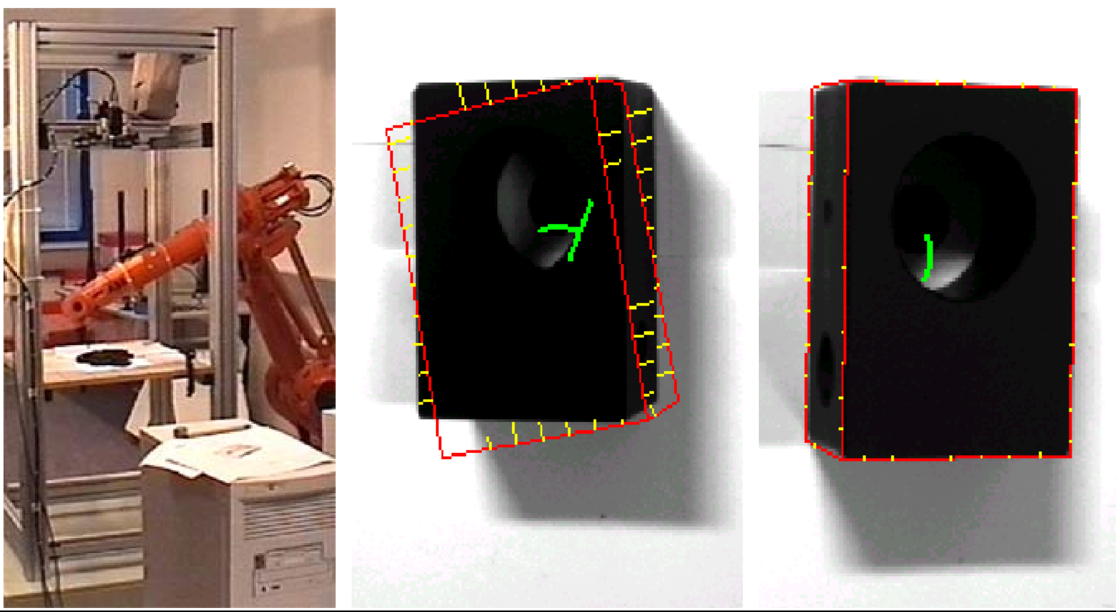


Figure 6: Robotics example with the need to match a CAD-model of a part with the image of the part as seen with a robot camera.

A CAD-model of the parts exits in the control computer, and multiple images are being taken of the part as the robot reaches for it and holds it. The task is thus to match the CAD description of the part with its images in order to establish the knowledge about location and orientation (the pose) of each part (object). We argue that the CAD model is “graphics”, the robot camera is “vision”.

3.2 Augmented Reality

There is hardly a better case than the new concept of *augmented reality* to make the point that graphics and vision are one, not two fields. Figure 7 illustrates the idea that humans se the real world through transparent goggles/monitors, and simultaneously see also virtual objects that appear on the monitors. The reality gets augmented by the virtual objects. Applications are many, from training, cooperative design to medical discussions. Since the virtual objects need to “interact” with the real world, this real world needs to be modeled. Since the position and attitude of the human observers needs to be fed into the computer continuously, “tracking” is an important component of augmented reality systems.

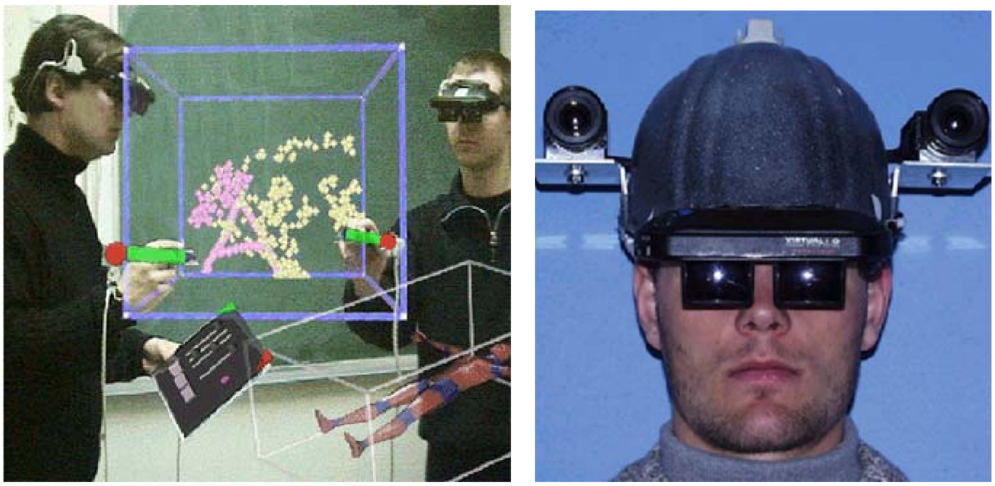
(a)  (b)

Figure 7: Augmented reality combines vision and graphics. The real world is seen stereoscopically and is merged with a virtual world, as shown in (a), while sensors track the human participant as shown in (b), noting the magnetic tracker on the helmet in (b)

The cameras on the helmet in Figure 7 serve to optically improve the tracking that would otherwise be performed with magnetic trackers alone. At the same time can the cameras be used to collect images of the objects in the environment to support the computer’s understanding of the environment in which he operator exists and moves.

3.3 As-Built Documentation

Either an object has been designed and is then built so that such a built reality needs to get compared to the design, or an object’s design is so old that it is lost or useless, so that the design needs to be recreated by a form of reverse engineering. There exists in both cases the task of creating a model of a 3D real environment and of its objects.

Figure 8 illustrates an augmentation of the commonly used MicroStation or AutoCAD software in which a so-called “block” of many overlapping images taken with a digital camera is the source of geometric information about the objects of interest. This results in classical case of image-based modeling of complex real world structures, and the instantaneous rendering of the results to obtain an assessment of the quality by superimposing the models with the photographic realities.

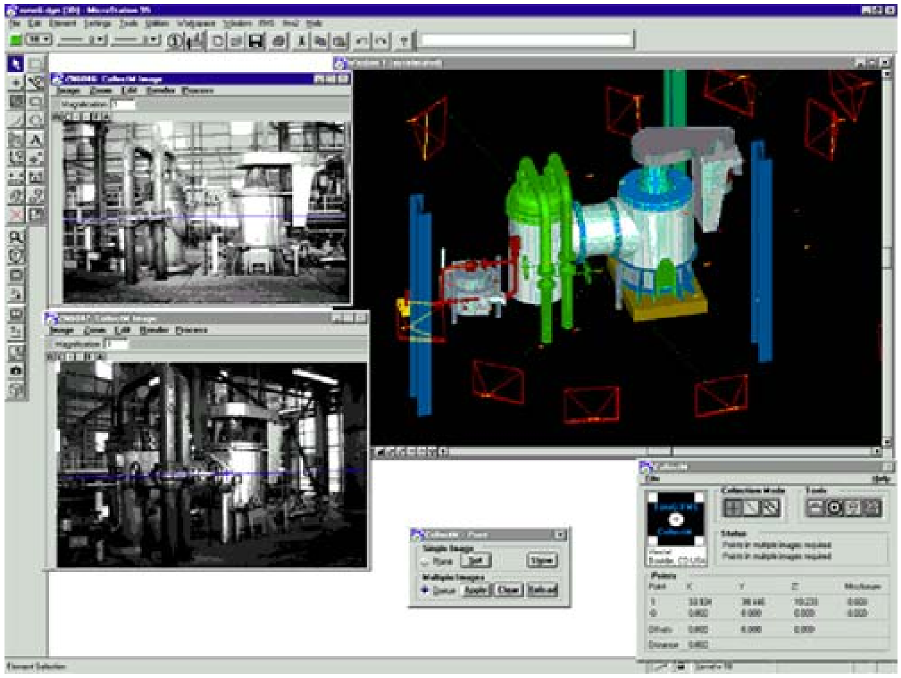


Figure 8: Photo-based “As-Built-Documentation” of existing plants or objects combines graphics, in this case the commercial software package *MicroStation*, with photogrammetry and image analysis. Either a model is needed of an existing entity for planning, or reality needs to be checked against previously drawn-up plans.

3.4 <u>CyberCities</u>

Urban areas have long been modeled in 2-D in the form of so-called Geographic Information Systems GIS, as shown in Figure 9(a). The third dimension is not part of the data base except as an attribute of some objects.

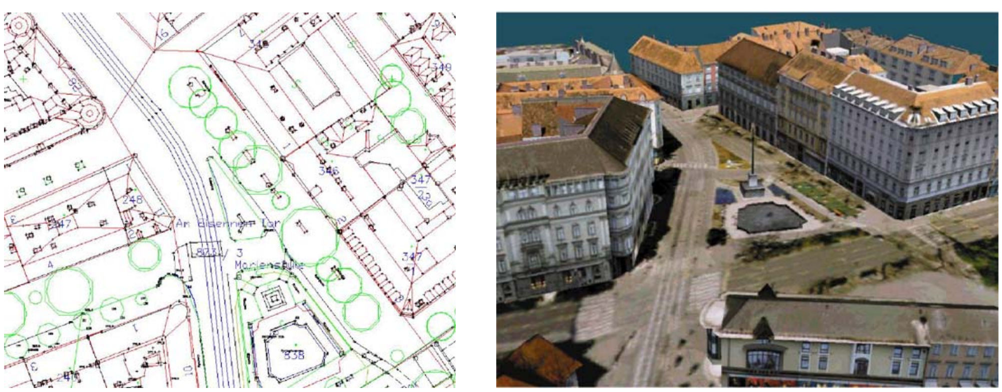
(a)  (b)

Figure 9: From the 2-D Geographic Information System GIS in (a) to Cyber Cities in (b) by modeling entire cities and make hem available over the Internet to he “armchair tourist”.

Recently an interest has emerged to describe urban areas in their full three dimensions. The initial impetus derives from the need of broad-band wireless access systems in which one needs to understand the limitations of intervisibility of buildings in cities.

Figure 10 shows an example of a city model used for telecom signal propagation analyses. Cities with 400,000 buildings and more are being modeled per city at a ±1 m accuracy in XYZ.

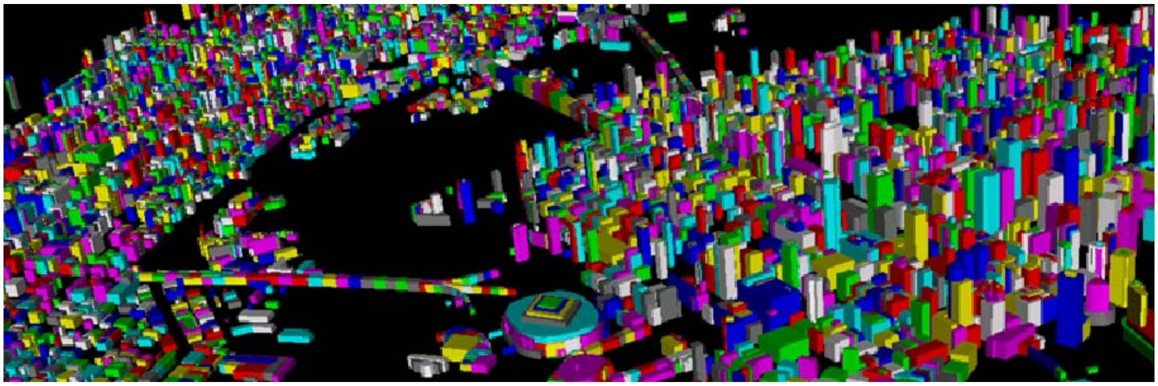


Figure 10: Typical building data set or telecom wireless broadband access studies. Shown is Vancouver, Canada (courtesy Vexcel).

4. A Look at Algorithms

A review of computer graphics text books and vision literature very quickly reveals that the same algorithms find applications in both areas. We review two examples of approaches where it is unclear where vision ends and graphics begins.

4.1 Texture

Another view of graphics and vision is obtained when one reviews specific algorithms. One quickly finds that one and the same approach exists in both fields, vision and graphics. Texture is an example. Figure 11 reviews a method of describing and modeling a surface property called "texture" which is being used to segment image regions, and which is also an important factor in achieving a realistic looking synthesis of computer generated images.

Texture can be described by an equation as shown in Figure 11, and the unknown parameters in the equations can be estimated for each specific texture using images taken of a sample surface and a reference pattern as illustrated in Figure 12. The reference permits one to separate out the low frequency parts of the surface material and the specular reflection effects, so that the signal from the test pattern is the basis of estimating the "noisy", thus the texture part of the surface pattern.

Based on the parameters found from the test images, any surface can now b synthesized with very little stored data, namely merely the parameters of the texture equation.

4.2 Affine Matching

In many instances of dealing with models and data of the real world, here is a need to match data that are in different data structures and geometries. An example is shown in Figure 13, where a 2D GIS data set consisting of the footprint of buildings needs to be matched with aerial imagery showing the roof lines of buildings. The idea is to find the roof-line automatically for use in a 3D model of buildings, but employing the known footprint of each building as a map to make it easier to define the building's roof line.

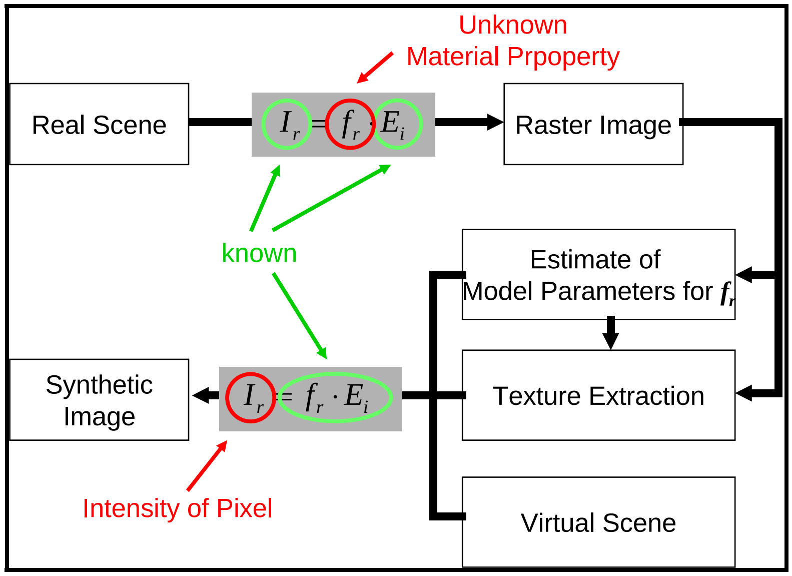
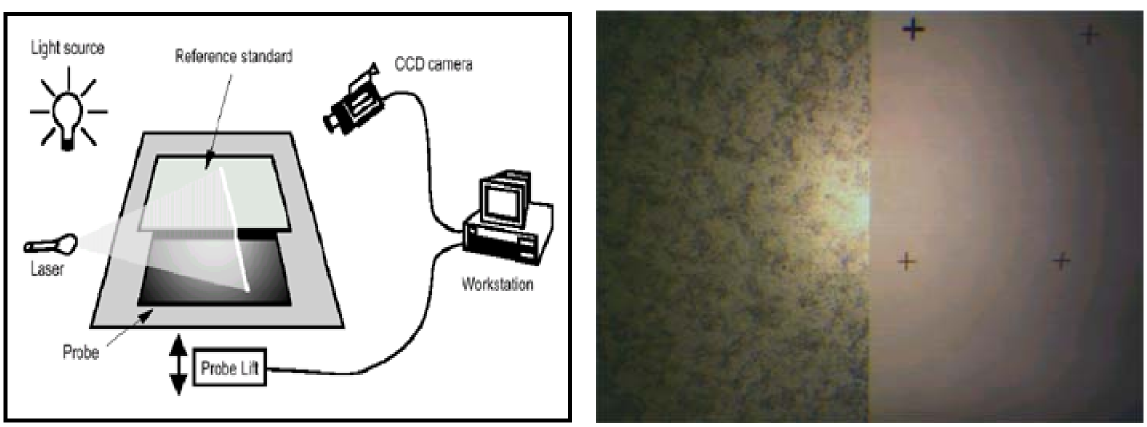


Figure 11: Working with texture cuts across the boundaries between vision and graphics. Unknown parameters of a texture equation get estimated in a least squares sense by means of images (vision) and are then used to create a synthetic image (graphics), employing the same mathematical model of texture that was used to describe texture in the real image



(a) (b)

Figure 12: Experimental set-up in (a) to create sensor data shown in (b) from which to determine the parameters of the texture equations shown in Figure 10 (Mayer, 1999).

The process begins with the creation of edge images from the input aerial photographs. These edge images are still in a raster data format. The 2D GIS footprint of the building of interest is converted also to a raster format. A distance transform is being made from the edge image, and the GIS footprint is matched with the distance transform, thereby changing the position, orientation and geometry of the footprint until an optimum match is found.

Matching via a distance transform has been called "chamfer matching". The simultaneous warp of the reference geometry augments this approach into the so-called affine matching method. The example in Figure 13 derives from Gruber et al. (1995).

(a) (b)

(c) (d)

Figure 13: A vector data set from a 2D GIS in (a) gets matched with raster data from aerial photos (b) extracting first edges in (c) and then using the technique of "affine matching" (also "chamfer matching"). The match result is shown in (d)

5. Curriculum Development

A final look shall be taken at a curriculum that takes a unified view of vision and graphics. Figure 14 summarizes the program for a class taught at Graz University of Technology where the idea is to abandon the thin line between vision and graphics. It becomes quickly apparent that there are numerous overlaps and subjects that would have to be addressed twice, were one to separately teach vision and graphics.

Each of the 20 Chapters contains material that will traditionally be found in both vision and graphics text books. This may be in "color" as much as in "transformation", even in unexpected contexts such a “motion”, where it occurs as motion blur both as an undesired effect in natural images, as well as a realistic effect in renderings of moving objects or cameras.

Basics
- 1 Image Description
- 2 Sensorics

Binary Images
- 3 Raster-Vector-Raster
- 4 Morphology

B&W and Color Images
- 5 Color
- 6 Image Quality
- 7 Filtering
- 8 Texture

Processing Spatial Data
- 9 Transformations
- 10 Data Structures

Spatial Objects
- 11 3D Objects and Surfaces
- 12 Interactions Between Objects and Light
- 13 Stereoscopy

Decisions
- 14 Object Descriptions and Classification

Changes
- 15 Resampling
- 16 Simulation, VR and AR
- 17 Motion

Systems
- 18 Systems and Pipelines
- 19 Man-Machine Interactions
- 20 Image Representations

Figure14: Schedule of the class on “Processing of Digital Visual Information” at University of Technology.

6. Conclusion

We advocate that a fresh look be taken at processing digital visual information in two separate fields, namely those of vision and of graphics, instead of thinking basically of only one field, along Figure 15 suggesting that image based modeling and image based rendering naturally form a unity of technology, of purpose and of methodology. This innocent basic concept, when taken seriously, will have to lead to various changes in the way digital visual information is being managed, its technologies are being studied and taught, software is being organized and used.

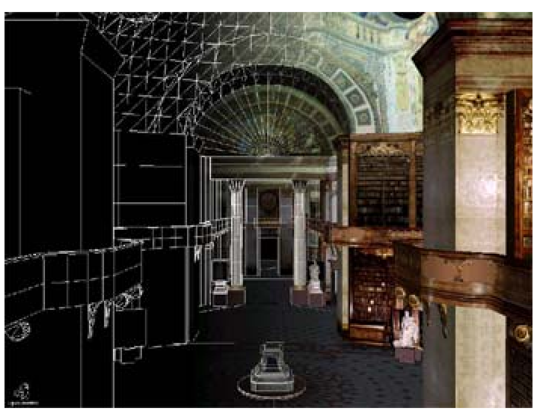


Figure 15: Image based modeling of the representation Hall of the National Library in Vienna, and image based rendering of the model, shown in a creative visualization that combines the two views (courtesy M. Grabner, Graz).

The issues are mostly with the structured academic institutions, instead of industry. Market forces have for some time enforced a broader view of vision and graphics in the commercial worlds, so that most commercial systems today cross the order lines between vision and graphics, as it should be.

References

Computer (1998) Special topic. Computer Society, July 1998 Issue.

Gruber M., M. Pasko, F. Leberl (1995) *Geometric versus Texture Detail in 3D Models of Real World* Buildings. In: "Man-Made Structures from Aerial and Space Imagery", Edited by A. Grün, O. Kübler, P. Agouris. Birkhäuser-Verlag, Basel, pp. 189-198.

Mayr H. (2000) Prosnal communication

Pinz A. (1994) *Bildverstehen*. Springer Wien-New York, 235 pages

Solina F., A.Leonardis, eds. (2000) *Confluence of Computer Vision and Computer Graphics*. Proceedings of a NATO Advanced Workshop, Ljubljana, Slovenia. Kluwer, Holland.