Dagstuhl Seminar No. 07171
Visual Computing – Convergence of Computer Graphics and Computer Vision
22 – 27 April 2007

Organizers:
Markus Gross (ETH Zurich, CH)
Heinrich Müller (Universität Dortmund, D)
Hans-Peter Seidel (MPI für Informatik - Saarbrücken, D)
Harry Shum (Microsoft Research - Beijing, PRC)

Introduction

Due to the importance of visual information for humans, visual computing is at the very core of the technologies enabling the modern information society. New and emerging technologies such as multimedia, digital television, telecommunication and telepresence, or virtual reality further indicate the tremendous potential of visual interaction with computers in the years to come. Typical for the field is the coincidence of very large data sets with the demand for fast, if possible interactive, user-adapted high quality visual display of the results. Furthermore, the user should be able to interact with the environment in a natural and intuitive way.

In order to address the challenges mentioned above, a new and more integrated scientific view of Visual Computing is required that unifies the previously separate "visual" disciplines of computer graphics and computer vision. Computer graphics is traditionally concerned with generating visual interfaces of computers and applications to the user. Computer vision focuses on enabling computers to understand and interpret visual information from static images and video sequences.

Summary of the Seminar

The seminar considered the whole pipeline from data acquisition over processing to rendering, including perceptional issues. This approach made it possible to uncover synergies between computer graphics and computer vision research.

The seminar had three types of sessions: research talks, keynotes, and break-out sessions. Apart from concrete research problems, several fundamental questions were addressed in particular in the latter two formats:

Generation of visual content

As the methods used to generate visual content become more and more complex, and the data sets used in the modeling process grow, methods from computer vision become an integral part of the data acquisition and modeling pipeline. Here, automated methods are required that make it possible to handle huge amounts of data. Conversely, generative techniques developed in computer graphics can be used to generate auxiliary and intermediate data for computer vision tasks, where knowledge on how to create images helps in understanding images.

In computer graphics, data driven content generation has replaced model-driven techniques in many areas where the models become too complex to handle. Complex models are often acquired using learning techniques. This enables the use of more complex models without the

Dagstuhl Seminar Proceedings 07171
Visual Computing - Convergence of Computer Graphics and Computer Vision
http://drops.dagstuhl.de/opus/volltexte/2008/1503

need for designing such models from scratch. In even more complicated cases, tasks can be completed from data alone, either without an underlying model, or with only partial support from a simplified, coarse model. Such methods are being used in rendering and modeling of extremely complex scenes and materials (for instance human skin), and are also applied to physical simulations, where they lead to a reduction to a simpler model.

Analysis by synthesis
From the computer vision direction, computer graphics techniques are used in core vision tasks. In analysis by synthesis approaches, generative techniques produce hypotheses that can then be tested. In many problems, these approaches lead to more robust methods for optimization and learning tasks. Examples for such approaches include methods for face recognition, alignment tasks, tracking.

Level of resolution
Whenever we generate or analyze data in a visual form, the question arises at which level or resolution this should be done. Clearly, imposing reasonable limits on resolution is necessary, but it is unclear what level of detail and what resolution is needed for a realistic, convincing, or simply plausible result. Studying human perception can give hints, and limits in human vision and hearing can be exploited to save costs while delivering an equally convincing experience. This question is of immediate relevance in research concerning level-of-detail representations, not only considering the easier task of geometric simplification, but also model simplification and behavioral simplification.

Engineering versus science
Taking a step back, it is enlightening to ask whether future developments in the field will be due to engineering achievements or scientific insights. As computing power grows, models can become more complex, and more sophisticated numerical techniques can be applied to harder and larger problems. Such advances in engineering have contributes a great part to the rise of physical simulation in computer animation, for instance, and will continue to make important contributions to the progress of the field. However, physical simulations have evolved into a third fundamental approach to gaining scientific insight besides theory and experimentation. Thus, models created from real-world data, or created with the purpose of recreating real-world behavior, may well lead to scientific insights into the studied object, be it crowds, human behavior, or various materials, especially when the models can be verified against real-world measurements. Such insights gained in visual computing research will have an impact not only in the field itself but also in other subject areas, such as for example biomechanics.

Modeling of human characters
There are several areas in ongoing research that cannot be tackled by computer graphics or computer vision alone. One such problem discussed in this seminar is the modeling of human characters. Specifically, in order to build a believable model of a human that can be used for content generation, automated techniques are needed. Model parameters should be inferred from video data, since manually creating the model is too complex to be feasible. This inference must be able to model subtleties such as the emotional state of the character. This ability will also lead to a deeper understanding of the principles of communication of emotions, which in turn can be used in related tasks in character animation.

In appendices, thoughts and grand challenges identified by two of the break-out sessions are compiled.

Conclusions of the Seminar

It became clear during the seminar that the fields of computer vision and computer graphics are not only closely related, but are mutually dependent. As techniques are exchanged between the fields, computer graphics and computer vision are converging into a discipline of visual computing. Using the knowledge about generative and analytic techniques that is available at both ends of the spectrum leads to the development of more robust and efficient tools able to handle the huge amounts of data that are typically dealt with. The understanding of both aspects of visual data, how to analyze it as well as how to generate it, helps in identifying fundamental principles that govern the processing of visual data in a computer. This knowledge leads to the development of better representations and primitive operations on a well-founded theoretical base, allowing use to replace heuristic and fragile approaches by robust and reliable methods in visual computing.

Appendix A.
Break-out session: Capturing reality

Summary by Leif Kobbelt and Wolfgang Heidrich

In science and engineering research, numerical simulations are quickly evolving as a third fundamental approach besides theory and experimentation. In this context it is becoming increasingly clear that efficient, robust, and mostly automatic techniques are required to capture all possible modes of information on real objects, including shape, material properties, and so forth. At the same time, an ever increasing level of detail for such digital models is also driven by the continued quest for increasingly realistic display of both real-world and synthetic environments.

Hence the problem of capturing reality and handling the resulting data sets is (at least) twofold: (1) how to acquire and merge all the different aspects of a real object of scene – especially across different levels of detail/resolution and (2) how to efficiently handle the resulting huge amounts of data such that interactive response times become possible.

In the discussion, we considered three different questions:

Which modes/aspects are relevant for visual computing?

One working hypothesis would be to capture reality like it is perceived by humans (without technological support). From this principle, we could, e.g., derive the appropriate spatial resolution (no lightyears and no micoometers) and the kinds of modes. On the other hand one could argue that even not directly perceivable object properties are necessary to eventually be able to simulate its realistic behavior. This question turns out to be another instance of the more fundamental question whether visual computing should target at realistic output or rather at plausible/convincing output.

Data representation?

Here the central question is whether it is desirable to have one universal representation which can serve as a master model and from which more specialized representations can be derived. This would allow for representations which are adapted to the particular requirements in a certain application but at the same time guarantee a proper correspondence between the various modes. One the other hand, if eventually specialized representations are needed anyway, it might not be worth the extra effort to integrate all the partial information into one unified model.

More concretely, with respect to geometry representation, there are polygonal meshes as the today's established universal standard. However, what will be the representation of the future? Depending on whether flexibility or approximation power are the driving forces, polygon meshes might be replaced by unstructured point clouds or by higher order representations such as subdivision surfaces.

Besides data structures, another important question is how much individualized digital models have to be. For surgery planning, it is definitely necessary to have a model of the actual anatomy of a patient. However if one wants to model, e.g., a lawn it might be overdone to store the exact geometry of every single leaf. In this case it would be more appropriate to have one or more typicalleafs and replicate them multiple times.

Science vs. Engineering?

There are many different technological as well as algorithmic approaches to capture and reconstruct the shape, material, and other physical properties of real objects and scenes. The question is, on which level do we need to improve reality capture systems in order to make significant progress in reliability, precision, and degree of automation. One standpoint is that the existing techniques are in principle sufficient and what is needed is a better implementation and integration ("engineering approach"). On the other hand, it could very well be that many of the open problems in shape and material acquisition that we still have today are due to the fact that the reconstruction principles known today have some intrinsic issues and new approaches have to be explored ("science approach").

Appendix B.
Break-out session: Extraction and retargetting of human physical properties to synthetic characters

Summary by Eugene Fiume

The solutions for modelling humans for character animation differ from similar solutions for biomechanics due to the workflow and usability needs for character generation and animation. Furthermore, virtual actors are almost invariably bad actors. That said, the extraction of physical, behavioural and morphological parameters from real people is a grand challenge for the field. These problems break down into subchallenges such as:

1. Automatically rigging a synthetic character from a video sequence of a human or other animal.
2. Mapping human performance to a synthetic character.
3. Inferring activation sequences and biomechanical properties from video.
4. Transfer of emotional state to characters.
5. Extraction and transfer of human behaviours from video sequences of crowds to synthetic crowds.