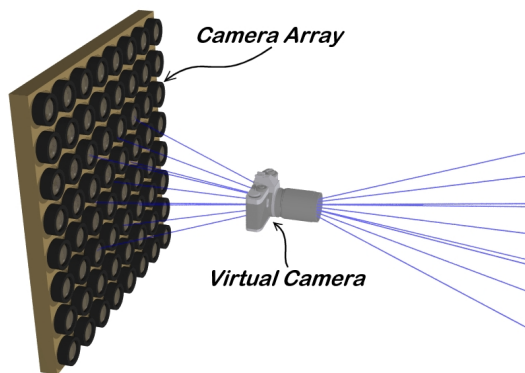


A Synthetic-Aperture Camera Array MIT 9904-14

Proposal for 1999-2000 Funding

Leonard McMillan and Julie Dorsey

One goal of computer graphics is to create compelling visuals indistinguishable from actual photographs. Ideally, these visuals would be generated in real time. Once these goals are achieved, computer graphics technology could be used to immerse users within compelling interactive environments. Thus, expanding the applications of these methods beyond the mere presentation of information to communication of visual experiences. Unfortunately, the current state-of-the-art in computer graphics is far from reaching either of these goals. Fundamentally, the two most difficult problems in traditional three-dimensional computer graphics are the generation of realistic geometric models and the simulation of the realistic physics required to render these models faithfully. We intend to side step both of these problems through the use of an alternate scene representation. We propose to construct a system in which new images are synthesized directly from a database of reference images. Furthermore, these reference images will be acquired directly from the real world. Thus, we avoid the traditional step of analyzing images in order to construct geometric and photometric models. The system that we propose would be comprised of a two-dimensional array of cameras that act together as a single device, which we call a synthetic-aperture camera.



Our synthetic-aperture camera array will be capable of generating live video streams from a wide range of different viewpoints. In addition to those video streams captured by the individual cameras of the array, it will also be able to synthesize unique video streams from other *virtual* viewpoints as if physical cameras were located there. These virtual viewpoints might correspond to the eye positions of a remote participant. The synthesized images from these viewpoints would provide a convincing sensation of a three-dimensional environment. Furthermore, several remote viewers can simultaneously view the scene captured from a single camera array, each from their own unique point-of-view.

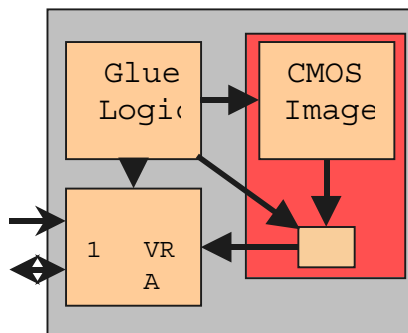
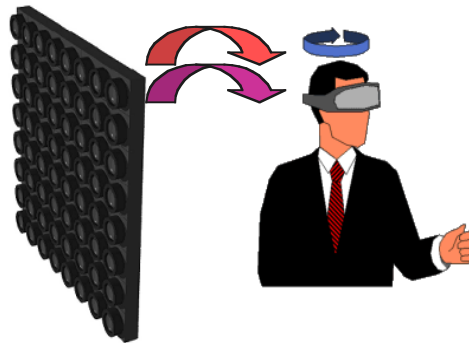
Our proposed synthetic-aperture camera array is ideally suited to a wide range of emerging applications. With its ability to synthesize multiple video streams within a large working space, it will enable a new generation of entertainment, electronic commerce, and remote-telepresence applications. For example, our camera array could be used in television and movie studios in place of dedicated camera operators to provide new and more flexible

capabilities. The studio director could freely control the position and field-of-view of a virtual camera using a joystick. Our system would also allow for a new range of special effects and camera angles that are unattainable with traditional methods. Our techniques would allow rich new capabilities for viewing and replaying sporting and other special events.

The system that we propose could also be used to provide flexible visual aids in many electronic commerce applications. In particular, it could facilitate the evaluation and purchasing of real estate, and other goods where visual inspection and spatial comprehension are essential to making purchasing decisions. Our system would allow users to freely roam within a space or around an item for sale allowing them to fully inspect or even measure items. Our system could also be employed by travel agents to market vacation packages thus allowing a perspective customer to preview nearby sights and hotel rooms.

Telemedicine is an example telepresence application that would be enabled by our camera. We imagine a scenario where one or more remote specialists could provide timely assistance to a local team of physicians. A telepresence system would allow remote participants to change their viewing position with a freedom similar to that of the local team members. This is in contrast to existing teleconferencing solutions where remote viewers are limited to a specific camera's view. Similarly, remote telepresence technologies can be applied in situations where a wide ranging set of viewpoints are required for the situation assessment or where the local conditions are too dangerous for a human participant such as when viewing chemical, or biological hazards, and space exploration. Electronic commerce, remote telepresence and virtual teleconferencing are likely to be key driving problems, or the so-called "killer-apps", of the next generation computers.

Physically, our synthetic-aperture camera will be composed of a two-dimensional array of image sensors each with its own optical path. The sensors will be laid out in a uniformly spaced planar array with approximately 5 cms separating the camera centers. Each image sensor in the array will have at least one million photosensitive elements. The external interface to this sensor chip will differ significantly from traditional CCD or CMOS image sensors. Unlike traditional video and digital imaging applications that

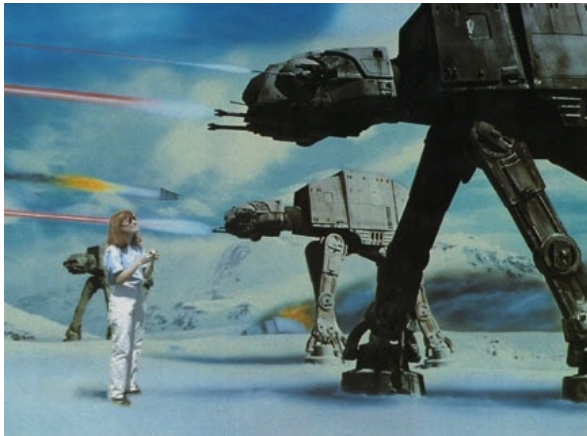


sequentially access pixels, our sensors will be randomly accessible like a typical memory chip. This capability is necessary for the synthesis of video streams from arbitrary viewpoints as described shortly. The practical construction of random access imagers has recently been enabled by the development of CMOS image sensing technologies. CMOS imagers also allow processing functions to be integrated on the same chip as the image sensors. This higher level of integration reduces system cost and allows for special purpose processing to support new applications. We plan incorporate the functions of analog-to-digital conversion, color-space interpolation, and

sub-pixel spatial interpolation and addressing, at each camera in the array.

Our synthetic-aperture camera array is able to synthesize images from viewpoints that differ from any of the array's constituent cameras. Furthermore, we intend to perform this image synthesis process in real-time for multiple streams at a target frame rates of 30 frames per second. The algorithms used in our synthetic aperture camera array do not rely on computationally expensive and fragile computer vision methods. Furthermore, we do not attempt to acquire a three-dimensional model of the scene. Instead, we are able to generate novel images by merely treating the live video streams from each camera as a database of rays from which virtual images are synthesized using optimized database queries and interpolation methods.

The simplest method for synthesizing novel views from a ray database is to query it for the closest ray to each of the desired pixels from the virtual camera. This query process can be simplified by considering it in two separate steps. First, the intersection of the desired ray with the plane of the camera array is computed and the nearest camera to this intersection is found. Next, the ray from the selected camera that is closest to the desired ray is found, and it is used as the approximation for the desired ray. Better approximations result from interpolating between multiple near-by rays from the



database. Interpolation is accomplished by combining the information from the four nearest cameras, and selecting the closest ray by interpolating between the four closest pixels. In the limit, as the spacing between cameras approaches zero and resolution of the imager sensors increases, this image synthesis method will result in an exact reconstruction. With finite camera spacing and low-resolution sensors, the resulting images will exhibit depth-of-field artifacts similar to real finite-aperture cameras. Thus, our camera array requires focusing controls, but the placement of this focal plane is achieved entirely through computation. Our goal is to build a random access camera module

using off-the-shelf parts by the end of 1999. We plan to use this camera module in conjunction with an X-Y motion platform as a proof of concept test for our system. We then hope to fabricate a motherboard with connections for an 8 by 8 camera array sometime in 2000.

There are many difficult problems on the near horizon for the field of computer graphics. While the last thirty years has seen impressive progress in the areas of photorealistic rendering and the development of specialized hardware platforms for generating graphics, the problem of generating high quality models is nearly as difficult today as it was then. Despite these difficulties, the expectations for the field are constantly increasing. Television series, such as *Star Trek*, portray futuristic *holodeck* capabilities, and movies, such as *The Matrix*, imagine virtual reality experiences that are indistinguishable from reality. Anyone familiar with the current state of the art in computer graphics is well aware that existing computer graphics technologies are a long way from accomplishing these feats, yet the expectations have not subsided.

To date, most of the focus in virtual reality research has focused on the problems of tracking and display. While these are certainly important tasks, they do not address the fundamental issue of how the underlying scene is modeled and represented in compelling detail. We believe that the synthetic aperture camera that we propose, along with its associated algorithms, is a significant new approach toward addressing these problems. Our research effort can be viewed as moving away from the traditional geometric representations of computer graphics towards an alternate image-based scene representation. We believe that sorts of representations have significant advantages over approaches used in the past. In particular, the use of acquired images simplifies the process of model acquisition, and provides photorealism by default. Thus, it is our expectation that the support of this research will have a dramatic impact on the future of computer graphics.