Performing Realism

Interactive real-time-rendered stereoscopic animations for architectural design process

Onur Yüce Gün
Massachusetts Institute of Technology
http://onuryucegun.com
oyucegun@mit.edu

Abstract. Realistic renderings contain a large amount of information about the spatial, geometric and material properties of prospective buildings that can directly affect design decisions, yet these images are most of the time used as after-the-fact representational visualizations. In this paper we propose a model to make the realistic images a very part of the design and decision making process. If we are to utilize realistic images during earlier decision making stages of the design processes, then we should develop experiential workflows in which we can create and interact with immersive realistic images real-time. We take several steps towards establishing an interactive rendering-animating-editing workflow that enables the designers to work with real-time rendered stereoscopic animations. In our system, we use realism to create an immersive exploration environment, as opposed to underutilizing it to represent a static moment.

Keywords. Visualization; stereoscopy; computation; interaction; immersion.

PROBLEM STATEMENT: DESIGN AND REALISTIC VISUALIZATIONS

Realistic Renderings in Conventional Design Processes

Architectural design processes run on visual production and evaluation. Designers cultivate the design form throughout consecutive stages of visual activities. Sketching is the first step during which designers constitute the visual substitutes of what has been evolving in designers’ mind eye as a mental image (McKim, 1972). As forms develop, the imagined transform into a composition of better-defined architectural parts. Once modeled digitally, spaces, masses, surfaces can be rendered using computational lighting simulations. Designers follow such a workflow as it appears to rhyme with the conceptual development of designs, however, doing so, designers end up isolating the realistic images from design processes. Time-consuming nature of rendering processes contributes to this isolation. What if we can create realistic images without much effort during the design processes and use what we see as an input for our design decisions?

A Novel Model for Using Stereoscopic Realistic Animations in Design Process

Designers use realistic visualizations towards the very end of the design processes, and thus photo-
realistic images remain as after-the-fact representations of potentially still-evolving designs. However, realistic images contain a large amount of information about the spatial, geometric and material properties of prospective buildings that can directly affect the decisions taken during the evolution of designs. In this paper we propose a model to shift the role of synthetically created realistic images in design processes. We introduce a workflow to create and use realistic renderings in an interactive and immersive way.

If we are to utilize realistic images during earlier decision making stages of the design processes, then we should develop experiential workflows in which we can create and interact with immersive realistic images real-time.

We take several steps towards establishing an interactive rendering workflow that enables the designers to work with immersive real-time rendered animations. Our goal is to make the realistic images a very part of the design and decision making processes.

In our system, we use realism to create an immersive exploration environment, as opposed to underutilizing it to represent a framed static moment.

We envision a seamless workflow in which we replace a computer aided design (CAD) model with a realistically rendered stereoscopic animation. Our goal is to position the designers into the very immersive environment of their designs, so they can observe and evaluate the emerging forms and spaces on the fly.

To build such a system, we define three key areas:

1. To eliminate the CPU intensive rendering processes, we use graphics processing unit (GPU) rendering so that we can create globally illuminated renderings in matters of seconds.
2. We use stereoscopic projection and active shutter goggles to visualize our images in fully immersed stereoscopy.
3. We eliminate all conventional human-computer interaction devices and utilize Kinect to interact with our realistic images.

Finally, we develop custom programs to incorporate all these technologies into one single tool. With this tool, we propose a workflow in which the designer interacts with a modifiable realistically rendered stereoscopic animation real-time in full stereoscopy (Figure 1).

**What is New**

Although there are numerous projects that incorporate some of the technologies mentioned above, there are no precedents that fuse all technologies we utilize in our system. Our system is also novel for the workflow it proposes.

<table>
<thead>
<tr>
<th>Limitations</th>
<th>Solutions</th>
<th>Technologies</th>
<th>Devices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rendering (CPU) time</td>
<td>Real-Time rendering</td>
<td>CUDA/OpenCL (GPU) Rendering</td>
<td>NVidia GeForce Mainstream GPU</td>
</tr>
<tr>
<td>Interaction with the CAD model</td>
<td>Device-less interaction</td>
<td>Motion and Voice Recognition</td>
<td>Microsoft Kinect</td>
</tr>
<tr>
<td>Immersion</td>
<td>Stereoscopic Vision</td>
<td>NVidia 3D Vision</td>
<td>Active Shutter Goggles and 120Hz DPL projector</td>
</tr>
</tbody>
</table>

Table 1: Common limitations and corresponding tools and technologies we use to address these limitations during the design and visualization processes.
We use stereoscopy in a live model: our models are rendered real-time; no textures are baked or pre-rendered. This enables us to change all materials and lighting conditions in the scene.

We use rendering engines that simulate light bounces, and this enables us to create realistic and accurate light simulations. Our results are much more accurate than the images that can be created using only ray tracing and ambient occlusion.

We use both motion and voice recognition to navigate around and modify our CAD model. We don't use any touch-operated HCI device.

Our system can be used by any number of people. Such opportunity encourages collaboration unlike head-mount display technologies (HMD), because HMD technologies are strictly limited to use of one person only (Kalay, 2004).

Our CAD model can be linked to a BIM model. We can read updates in a BIM model and register the changes into our CAD model, and thus into our animation.

BACKGROUND: MODELS FOR VISION AND VISUAL MATERIALS FOR DESIGNS

Structuring Human Vision
The way humans look at the environment and the way they structure and describe the human vision affect the way they de-construct the world into parts and construct (design) things. Alpers (1983) compares the descriptive Dutch Art and its vision of the world to the narrative art of Italy. According to Alpers, Italian art concentrates on reconstructing the image of the observed in guidance of certain rules, using the pre-conceived mechanical perspective construction. The Dutch art, however, utilizes the camera-obscura and pictures the visible with no translational reconstruction method. While the Italian art is about representing the observed, the Dutch art stresses the significance of vision (Alpers, 1983). According to Alpers (1983), the attention is drawn from the picture (observed) to the mechanical maker of the picture, the human eye itself, the observer.
Alpers differentiates the “perspectival” and the “optical” perception of the world in reference to Alberti’s framed window and Kepler’s model of the retinal image. In Alberti’s framed window, the observer always stands outside the observed, and looks at it from a distance through a phantasmal window. In Kepler’s model of the retinal image, however, the observer stands within the middle of an ever-extending landscape.

These fundamental differences in models of visual perception and representation vitally relate to one of the core issues in contemporary architectural design processes: designers work with computers to create virtual models of their designs and to visualize them, and they undertake all these tasks looking through the flat and un-familiarizing computer screen. Working with rendered volumetric representations through flat displays of computers creates a feeling of ambivalence for designers. The designers—mostly unaware—work in an uncomfortable setting: computer graphics remain in an intangible and perceptually isolated world, behind two screens: Alberti’s perspectival frame as the first translation screen, and the flattening, impermeable computer display as the second detachment screen.

**Simulating Human Vision**

Stereoscopic devices appeared as early as 1830, yet they become obsolete shortly thereafter due to the physical engagement required (Crary, 1990). (Figure 2.) Appearances of head-mount devices re-amplified the interest in stereoscopic vision in mid-20th century. Today stereoscopic vision systems have become just another mainstream technology that is built in televisions and personal computers.

Simulated stereoscopy is in no means a true replication of reality (Crary, 1990), yet stereoscopy grants the observer an enhanced proximity with the observed. Crary (1990) indicates similarities between historical stereoscope and stage design. According to Crary, in stereoscope and stage design, objects are organized into planar layers that recede in depth and such organization help visualize senses of “in front of” and “in the back of.” Today’s advanced technologies enable viewers to see simulated depth in a more continuous way. In our project, we use the amplified depth created behind and in front of the stereoscopic projection surface to generate a sense of environment while working with volumetric renderings. This very depth helps to utilize visual material to perform beyond flattened and framed renders.

**Design, Computers and Animations**

In comparison to analogue representations, digital representations are claimed to feel scale-less, and to be intangible, because they linger isolated behind the computer screen (Friedberg, 2006). Even after designers reach the high level of technical proficiency required to generate precise, descriptive and well-communicating visuals, designers’ engagement with the digital content remain beclouded.

Visualization specialists spend many hours to produce compelling realistic images, for these images to be printed, hung on the wall and edited with red pens. Designers have almost no real interaction with realistic renders, and they can hardly use such
images as feedbacks due to time consuming iteration cycles. Emerging real time rendering technologies enable designers to work with digital models via creating realistic images on the fly. These technologies also dilute the distinction between a static image and an animation: the images are produced as a by-product of the CAD scene, and any manipulation done in the CAD model affects the content of the image. Such a workflow differs from the conventional image or animation making processes, as animation becomes a product of an unplanned and intuitive navigation process. Designers rotate, pan and zoom in the scenes to see whatever is important for them at a specific moment and do modifications addressing that specific part of the model.

HCI devices and interaction techniques have remained the same for many decades; only recently alternative technologies have become accessible. Touch-less devices such as Kinect are expected to become mainstream in couple of years. The goal of these devices is to enable more intuitive interaction with computers. There are also HCI devices that are specifically designed for CAD modeling. 3D mice such as space-navigators enable designers to have a better navigation in 3D model space in comparison to conventional mouse and keyboards. Yet these devices are also table-top, hand-operated devices. Developing technologies such as Kinect and Leap-Motion enable touch-less interaction. We use Kinect and benefit from its both motion and voice recognition capabilities. Our system enables us to navigate within a realistically rendered stereoscopic 3D environment by using bodily gestures. We can update scene settings via spoken commands. These two implementations grant us the chance to work with our CAD model in an unprecedented way.

IMPLEMENTATIONS: RT RENDERING, STEREOSCOPIC AND INTERACTION

Precedents
Immersion visualization systems encourage engagement with the observed. However, as Kalay (2004) clearly states, it is a hard task to create a truly immersive environment:

“Engendering a sense of ‘being there,’ without being ‘there’ is hard to do.”

Although there are numerous projects that incorporate some of the technologies mentioned here, there are no precedents that fuse all technologies that we use in our system. Immersive CAVEs (Burnett, 2004) can be analyzed as closest matches to our system; however, CAVEs lack most of the interaction, modification and global illumination (GI) rendering capabilities. A CAVE is generally a cubic booth that measures approximately 3 meters on the edge and features three to five back-projection screens to create an immersive environment (Kalay, 2004). The user (or up to three users) can virtually navigate in a pre-prepared scene using active shutter goggles that enable stereoscopic vision. Scenes cannot be modified; and are rendered using ray-tracing and ambient occlusion only. Due to their nature, CAVEs require multiple projectors along with a designated, custom designed booth. We build our system using cost affective and easily accessible hardware that can be used with personal computers. Devices we use are top of the shelf consumer products.

Technologies, Hardware Setup and Workflow
We use both existing and developing technologies to realize our project (Table 1). Below we explain all the technologies, the way these technologies work and the impact they have in our project in detail:

- Rendering: We render a responsive 3D CAD model in Autodesk 3DS Max 2012 using two NVidia graphics processing units (GPUs). We use VRay RT (real-time) to utilize GPU rendering that can operate on both NVidia CUDA and OpenCL rendering engines. Using VRay RT, we can simulate global illumination (GI) and produce properly illuminated realistic images.
- We use one of NVidia’s GTX580 and two of GTX-560TI graphic cards. A GTX580 GPU features 512 parallel processing cores in comparison to 8 parallel processing cores of a high-end cen-
Central processing unit (CPU). Lighting simulations benefit from high number of parallel processing cores, and as a result a GTX580 GPU renders the same scene up to twenty time faster in comparison to an Intel Core i7 950 processor of the same price bracket. Higher number of cards yield lower rendering times. We distribute rendering overhead to all three GPU cards and render stereoscopic images as quick as 2 seconds. We render our scenes at the resolution of 1280x720 pixels (720p HDTV format). We render two (right and left) images simultaneously using our virtual stereo camera setting in 3DS Max and save these images in side-by-side format to a designated location on our hard-drive. We develop a custom stereo displayer in C# and use it to display our side-by-side images using the 3D projector (Figure 3).

Projection and Stereoscopic Vision: While using our system, the designer stands before a projection screen with active shutter goggles. We do not require any other device connected to the body (Figure 3.) We use NVidia’s 3D Vision technology to create our stereoscopic animations. This technology consists of four devices: a 3D projector, active shutter goggles, NVidia GTX series GPU cards and the NVidia 3D USB IR emitter. The emitter synchronizes the video card, goggles and the projector for left and right shutters of the goggles to open and close to simulate stereoscopic vision, while the 120 Hz 3D projector casts 720p resolution stereoscopic image on the screen.

Interaction: A Kinect located in front of the user can track the body and hand positions in space. We incorporate motion and voice recognition via Microsoft Kinect SDK. We use the forward backward leaning of the user to move towards or backward in the scene. We use the hand positions to move our cameras to the right or to the left in the scene. We can push the active camera in the scene up and down with corresponding hand movements. We use similar intuitive hand movements to change the places of objects or the location of the sun in the scene.
We use spoken commands for several purposes. First, we use it to switch between objects. We can switch between cameras, lights and objects (such as walls) by speaking their designated names. Once an object is activated, we can use hand gestures to modify them. We can also turn lights on or off, change textures by using voice commands. All live connections help us update the location and properties of objects in the model without an interruption in the workflow.

**Extended Capabilities**

The sequence of navigation and all modifications are all rendered and saved in the hard-drive at the time of live stereoscopic projection. This sequence of images can also be used to re-play or compile the interaction process as a video.

We can create a live connection between our visualization system and BIM models. Updates in both the BIM model and the rendered CAD model can be transferred bi-directionally.

**Live Demonstration**

Our system is fully operable. Following many tests we did a live demo of our tool and workflow during the final review of Computational Design Lab: Reinventing BIM course at the Massachusetts Institute of Technology, department of architecture. We used Barcelona Pavilion [1] in our demo. Our demo included changing artificial and day light conditions, switching between textured and clay render modes, moving objects such as the statue or the walls, navigating in and out of the building. Several stereoscopic side-by-side images can be seen below (Figure 4).

**CONTRIBUTIONS**

With this project we have both technical and conceptual contributions. Our main conceptual contribution is that we introduce a visually driven immersive design process in which the imagery acts as an explorative dynamic playground. We incorporate visualization and interaction technologies and devices in an unprecedented way, so that images and animations become responsive design environments. As opposed to static renderings, our live stereoscopic animations help designers visually evaluate, see and modify designs real-time. Massing studies, geometric modifications and visualization studies can happen within an immersive environment, in which designs and their realistic imagery nourish bi-directionally.

Our technical contribution is a working tool that incorporates state of the art motion and voice driven human-computer interaction, real-time rendering and stereoscopic vision technologies for a seamless visually driven design process.

**ACKNOWLEDGEMENTS**

I would like to thank my project collaborator Alan Tai for his support and technical input during our studies. I also would like to thank Associate Professor Takehiko Nagakura and Professor Terry Knight for their support during the development of this project. Co Architects of Los Angeles sponsored the hardware that made this project possible.

**REFERENCES**


