Rapid Scene Acquisition and Modeling with Depth Enhanced Panoramas

Gleb Bahmutov

Faculty advisors: Dr. Voicu Popescu and Dr. Elisha Sacks

Abstract

Realistic 3D models of the real-world objects or scenes are required in multiple applications of computer graphics: virtual tourism, cultural artifact preservation, etc. Obtaining such models has traditionally been a time and labor consuming task. Variety of target scenes can be divided into several levels, based on the geometrical complexity. Fragmented scenes with few large continuous surfaces present an especially difficult modeling problem. The current dominant acquisition techniques have a common major drawback: each iteration of scan / inspect / adjust loop takes a long time. Recently several researches were able to capture in real-time a geometric model of a complex small object giving the user a control over the acquisition process. The ModelCamera technique I have been working on has an interactivity of the real-time acquisition, acquires both geometry and color information, and is suitable for large fragmented environments. The current status of the project and examples of the scanned scenes are described. List of the future research objectives I would like to accomplish is included in this document.

Definitions

3D acquisition or scanning. The process of obtaining explicit rendereable geometric and photometric model of a physical object or scene.

IBR: Image-based rendering. Pure IBR methods replace rendering a novel view from an explicit model with interpolating existing views. Hybrid methods IBR methods combine interpolation with coarse geometry, usually inferred from correspondences in multiple views to generate physically correct novel views when the input set of images is sparse.

Active acquisition. 3D scanning techniques that use controlled light source such as laser or light pattern to infer the 3D shape of the scene. Passive techniques rely only on the information contained in the set of visual images.

Structured scenes. The scenes with low geometric complexity, typically with several smooth surfaces.

Fragmented scenes. The scenes with very irregular geometry difficult to capture without sampling multiple points, *i.e.* a cluttered bookshelf, an antique shop, a garden.

Introduction: Scene acquisition applications

The wide availability of powerful graphics cards and advances in rendering technologies has generated a need for more complex and realistic input models. Many applications in computer graphics require acquiring and rendering of the novel views of an existing object or scene. A list of such includes:

1. *Virtual and augmented tourism* allows anyone to visit some of the remote cultural sites. The user can explore the virtual reconstruction of the Sagalassos town in Greece with the help of a virtual guide [36]. The virtual scenery combines the detailed 3D models of the remaining building parts with the CAD modeling of the missing pieces. In ARCHEOGUIDE project, the user physically roams the Ancient Olympia site wearing a mobile computer with

augmented reality headset. The system places 3D reconstructions of ancient building and even competing athletes into the natural environment surrounding the tourist [49]. A captured 3D model of Thomas Jefferson's Virginia home Monticello has been displayed as part of the museum exhibition in New Orleans Museum of Art [50].

2. *Architectural modeling.* If an existing building is being remodeled, the building owner might want to view the result before the changes are made. A CAD model of the original structure lacks the realism of the actual scene. By capturing the 3D model of the building in its current state, the designer could better understand the impact of the future changes. City planners can inspect the virtual addition of a new building in existing city to evaluate its appearance and impact among existing ones [23].
3. *Cultural artifact preservation.* The cultural artifacts and heritage sites must both be open to the public and preserved for the later generations, two requirements often contradictory in nature. Scanning the artifacts allows the general public and the archeologists an easy and convenient access the detailed site models, simplifying the research and collaboration. The marble statues by Michelangelo enjoyed a special attention by 3D modelers [4, 26]; several very high detailed scans are available and allow the generation of the novel views impossible in reality because of the security or physical limitations. The digital reunification of Parthenon and its sculptures [46] has become possible after careful scanning and modeling of the building and the sculptures. This project is especially interesting because the sculptures are scattered physically across the world; some pieces were broken or lost and only plaster replicas survive.
4. *Medical collaboration* for remote surgery or learning has employed the acquisition and rendering of complex models of the internal organs allowing medical professionals inspect, store and share spatially complex anatomic data [9, 13].
5. *Advertisement and marketing.* Three-dimensional objects are relatively new in the e-business, but offer natural shopping experience in some retail areas such as the furniture sales [33]. Virtual 3D gallery system offers a new way to sell art pieces without the expense of the publishing catalogues [30]. In my view, there is an unrealized potential for using captured 3D models in hotel / apartment advertising and home décor sales. Perhaps this potential could be exploited with the availability of inexpensive and easy to use modeling tools targeted at the average consumer.
6. *Computer games and special effects industry.* The increasing realism of the games, and the complexity of the special movie effects drive the integration of the modeled real-world objects into the artificially generated 3D environments and vice versa [48]. The movie "Spiderman" has combined the actor's face with a 3D rendering body captured off a stunt man to create the protagonist character, later moving through a large complex CAD structure modeling the Brooklyn Bridge with the real world New York panorama in the background [20].
7. *Robot navigation and remote exploration.* The exploration of the remote environments by self-propelled robots currently is the only way to visit some of the sites on our and other planets. The Mars rover Pathfinder [32] has carried a stereo camera allowing the astronomers 3D reconstruction of the observed environment. Opportunity and Spirit rovers employ stereo vision to build a 3D model of the surrounding environment and then navigate around the obstacles [31]. The accuracy of the model and exact tracking of the robot's position in the world are the two most important factors.
8. *Accessibility devices.* The hand gesture recognition system capturing the geometric 3D model of the hand in real-time can recognize complex hand postures, such as that employed in the sign languages [27]. In this project the rendering of the captured 3D model is not required.

The 3D model acquisition has to capture both the geometry and the color information about the object to be able to render its novel realistic views. A number of approaches have been devised to obtain the geometry and color model. The scanning approaches differ in the type of scenes they can capture, the acquisition method and scanning time, the amount of operator's assistance, the quality of the output model, etc. Currently, there is no fully automatic and accurate technique that can capture all types of objects and scenes, and devising such method has been shown to be a very difficult problem. Not surprisingly, an entire new branch of computer graphics has been born lately that lessens or avoids the need for explicit 3D scene model by using pre-obtained photo images as the source for rendering novel views [29]. Using the photographs of the scene is a natural approach to the creation of the scene's novel photorealistic views that has certain advantages, and I will shortly discuss the methods employed by the image-based rendering before proceedings to traditional geometry + color modeling.

Rendering without 3D models – pure IBR

Lightfield [25] and Lumigraph [14] image based rendering techniques are very similar. The system acquires and stores a database of all possible views of an object. Model rendering is replaced by looking up and interpolating views closest to target view. Aliaga in [2] demonstrated a system where spatial relationships among the stored images were encoded using a technique similar to MPEG encoding. Pure IBR technique has demonstrated the creation of photorealistic views for some very intricate objects, such as furry toys or semitransparent objects [28], but it becomes increasingly impractical to extend the method to the larger scenes because of the sheer number of views to obtain, store and efficiently look up in the database.

QuickTime VR [6] system is applicable for larger scenes; it acquires a color panorama from a particular point in the scene. For example, see an online virtual tour of the Purdue University campus with 24 outdoor panoramas [54]. Without any geometry present, the novel view generation is constricted to rotations around the original acquisition point. Not being able to freely move around the scene a significant drawback that limits QuickTime VR technique.

Rendering with semi-explicit 3D models – hybrid IBR

The large size of the image database necessary to store all possible views of the scene has become the bottleneck of the pure IBR techniques. The hybrid approach was devised by Seitz that generates physically consistent transition novel views between sparse set of images using view morphing [7, 45]. The method relied on the user to provide manual correspondences between the images to achieve its accuracy. Automatic feature matching allowed Aliaga [3] to use view morphing on a larger scale for capturing and rendering indoor scenes. The drawbacks of the system are the long acquisition time (7 hours for a 1000 square ft room), large number of omni directional images in the database (15 000) and long post processing feature extraction and tracking step (4 to 30 hours).

Photogrammetric modeling approach [8] combines sparse images with coarse geometric model. The geometry is refined until it is consistent with the photographs. The images then provide the color samples sprayed onto the surface of the model to give realistic texture. In most applications the large site models combine photogrammetric models with other techniques more applicable to capture smaller objects on the site.

The advantages of the semi-explicit models are smaller number of input images, faster rendering compared to pure IBR methods; they offset by the long acquisition times and lack of the hardware acceleration to speedup image generation. Novel graphics architecture hardware has been proposed [37] that can efficiently generate novel views, but it is not widely available.

Automatic scene acquisition

The more traditional scene modeling acquires explicit geometrical and photometric description of the visible surfaces of the scene. The manual input of the geometric and color samples is very labor-intensive even for a simple real-world scene and is thus impractical. The color photographs or a video stream can provide the scanning device with a very dense color sampling of the scanned object's surface. The automatic capture and registration of the geometric information about the scene (the depth sampling) is the harder aspect of the problem, and I divide target scenes into three main categories in order of geometric complexity.

1. Urban scenes with simple geometry.

2. Small objects with complex geometry that can be scanned by rotating the object in from of the stationary camera.
3. Room size environments with complex geometry.

Due to the difference in the target applications, the acquisition methods should be discussed grouped by the scene type.

1. *Urban scenes with simple geometry*. For scenes where the geometry is relatively simple, the manual input of the points can be done by specifying the correspondences among the features of the source images. The color information from a video stream or color photographs is combined with the 3D samples and the domain knowledge to produce a coarse geometrical model with the real-life color (Façade [8], Hidalgo [17]). The dense color of the texture can somewhat hide from the user the lack of the geometry in the source model.

2. *Small object with complex geometry*. There are a number of techniques that can acquire a 3D model of a small object. An object is positioned on a turn table in front of a fixed camera. Multiple techniques were devised for such setup, typically called *shape from X*. Seitz has been able to reconstruct the 3D shape and color properties of the object from the color information alone [44]. Shape from stereo relies on a multi camera setup paralleling human vision and can reconstruct the 3D model of the object, assuming there are no significant occlusions [22, 23]. Structured light methods [15, 40] employ a projector casting a light pattern onto the object. The shape of the object is determined from the observed pattern distortions. Recently, the time-of-flight laser range scanning has became a popular tool capable off capturing both small [34], medium [4] and large scale objects [11]. The sensor operates by measuring the time taken by the laser beam to return after bouncing from the scanned object surface. By rotating the scanning beam, entire scene can be captured with the great precision. The drawbacks are a dependence on the reflective surface properties of the object and ambient light conditions, a long scanning time and the output unfiltered point cloud models lacking color. Structure from silhouettes [39] extracts the 2D silhouette of the object in the image plane as the object is rotated in front of a controlled background. The 2D contour defines a cone of 3D rays for each image. The intersection of the rays among all images defines the shape of the object.

Most of these techniques were not extended successfully to the rapid acquisition of a large indoor scene.

3. *Large environments* (room-size and larger) with complex geometry (such as an office, a car repair shop, or a large ancient statue) are the most difficult objects to capture. The acquisition device needs to obtain a large number of depth measurements as well as color samples to produce high quality novel views of the scene. The registration of the geometry and the color data with each other is needed. As observed from a particular viewpoint, only a part of the scene is visible and can be captured in each scan. The multiple scans have to be registered together to produce the complete model of all surfaces. The filtering the obtained data and the registration of the several 3D scans in a common coordinate system has been described by several authors as a post processing step [1, 18, 34]. The automatic acquisition of the large environments was the focus of my study and is discussed in more detail in the next section.

Automatic Scene Acquisition for Large Environments

Structure from motion [35] has been successfully applied to the large site modeling [36]. This technique uses an uncalibrated free moving hand held camera taking several photographs of the scene from the different viewpoints. The intrinsic parameters and the pose of the camera are inferred from the photographs. The color 3D model of the scene is also decided by relying on motion parallax to reveal the 3D structure of the scene. The main drawbacks of the system are its reliance on detectable correspondences between the images which is susceptible to occlusions and susceptibility to noise in the data; there is also a long delay between the data acquisition and the model inspection due to processing time.

Currently the acquisition of complex scenes is dominated by the time-of-flight laser range sensors and the structured light devices, a number of these devices are commercially available [52]. The heritage site and cultural artifact preservation have benefited from availability of such scanners: Great Buddha statues in Japan and Thailand [19], Michellangelo's Florentine Pieta [4], David and other statues [26], Parthenon [46] and Abbey of Pomposa and Scrovegni Chapel [10] archeological sites have been captured with great detail using these techniques. Both time-of-flight sensors and depth from stereo suffer from the following drawbacks [10, 41]:

- Only the geometric data is acquired in most cases. If the color of the surfaces is necessary, a camera could be attached to the acquisition rig or a separate color acquisition step can be performed and the results merged with the depth samples. The color acquisition presents the tracking of the camera and the registration of the collected information problems.
- A single scan takes a long time; the result is not visualized until it is complete, at which point the operator might want to adjust the position of the scanner and repeat the scanning. Each iteration of the scanning / inspection / adjustment loop is time consuming. The scene has to remain static during the entire acquisition.
- The scanning device is expensive, bulky and can be complex to operate
- A sensor intended for the close range is not suitable for the long range scanning and vice versa.
- The post processing step is necessary to filter the output point cloud and transform it to a triangulated surface model.

The trend toward real-time model acquisition

The complexity of the indoor scenes comes from several factors: fragmented surfaces, occlusions, high and low level detail. The large scenes require the scanning process to be repeated several times to obtain a complete model. The current methods employed in a large scene acquisition suffer from the long period necessary to perform each iteration of the scan / inspect / adjust loop. The defects in the scan are not discovered until the inspection, thus wasting a lot of time and labor.

Recently, several projects presented the interactive modeling methods. The user was shown the evolving model during the automatic acquisition process. The user was allowed to adjust the scanning process by moving the acquisition device or the object without stopping / restarting the acquisition. Structured light method has been shown to produce an accurate geometry-only reconstruction of a small object in real-time by doing a fast registration of the newly acquired points with the model [41]. In a similar research, a moving or static deformable object has been captured by adapting the projected pattern to the changing surface of the object [24].

The tracking the light producing pointer in [47] also allowed the real-time addition of the new points to the evolving point cloud. The static camera observes an object illuminated by a laser pointer freely moved by the user. The pointer produces a red dot on the object detected by camera. The orientation of the pointer is known from the green laser LEDs attached to the pointer itself and constantly detected by the camera. The 3D point corresponding to the red dot is obtained by triangulation of the video camera ray and laser pointer beam. Similarly in [12] a freely moving laser pointer projecting a thin bright line is tracked using attached LEDs, and the scene is reconstructed at interactive rates.

An interactive active stereo (trinocular) hand held system consisting of two video cameras and a projector rigidly connected together is used in [16]. The projector is used to provide a cross light pattern detected by both cameras, the patterns are matched in left and right camera, and 3D points in from the pattern are calculated from the stereo views and added to the point cloud. The camera's pose is calculated by observing a set of fiducials in the scene cast by another static projector.

In the above cases the model is continuously updated and the geometry is displayed to the user. The advantage of such approaches is the shorter acquisition times and far greater control over the scanning process. So far, all were demonstrated on small objects and acquire geometric information only. I believe that the next goal of the 3D modeling research is the real-time acquisition of both color and geometry for a large scene with the continuous model update and display to the user.

ModelCamera acquisition device for indoor scenes

We have designed a novel inexpensive scene modeling device based on a consumer grade digital camera and a rigidly attached laser pointer generating a 7 by 7 dot pattern in the camera's field of view [38]. This setup allows us to capture both the dense color and sparse depth samples in each frame. The information about the observed part of the scene is merged with the model of the scene at a rate of five frames per second. The evolving model is constantly presented to the user who can adjust the acquisition process. The device is suitable of capturing simple indoor scenes in hand-held mode, or highly complex fragmented indoor environments using a tripod. The main advantages of the ModelCamera acquisition and modeling technique are its speed, simultaneous geometry and color acquisition and the user control over the scanning process.

Research statement

The ModelCamera is a novel scene acquisition and modeling tool capable of capturing complex indoor environments. The acquisition technique we developed can be used in a variety of applications and has several advantages over existing modeling techniques such as laser range scanning or structured light. For my PhD thesis I would like to:

1. *Develop faster and more robust acquisition and modeling algorithms and data structures to work with the ModelCamera device*
2. *Apply the ModelCamera to several possible applications from the Applications list.*
3. *Compare the usability, performance and quality of the generated model to the other available scene acquisition techniques*

I will use the indoor environments with complex geometry as the target scene type. Such scenes occur frequently and present the most difficulty for the current acquisition techniques. To distinguish this technique from existing methods, I will set more specific objectives:

1. The acquisition should take less than 30 minutes to capture a complete model of a room.
2. During the scanning the user should see already acquired portions to better direct the scene capture process. A common feature to existing techniques is their full automation. The automation lessens the burden on the operator, but the technique does not use information the human operator has about the scenes. I propose using the human input for higher level tasks: for example for directing the system to obtain more depth samples on a particular complex part of the scene.
3. The captured model should contain both the high-quality color data and a sufficient (the operator can set the target level of detail) geometry detail level and should be rendereable on the existing hardware.
4. The system should be portable, robust and inexpensive to suit a large number of applications.

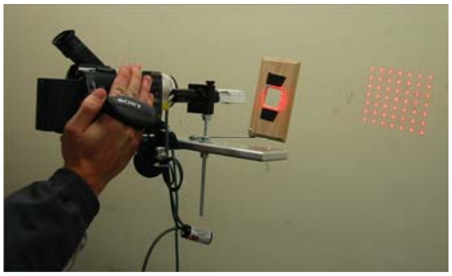
I believe that none of the existing techniques can satisfy these requirements, and the successful development of such system would have a great impact on the computer graphics field and its applications.

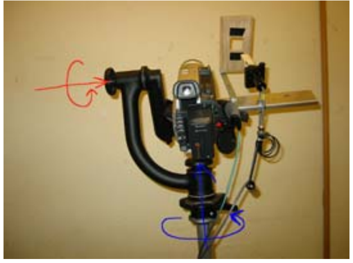
Current research status

The ModelCamera can be used in handheld mode to capture simpler scenes with low geometry fragmentation. For more complicated scenes the ModelCamera is positioned on a tripod using a special bracket that restricts camera's movement to panning and tilting around the camera's center of projection point. I will focus on the acquisition of fragmented scenes that requires using the tripod. While tripod limits the portability of the device, the fast acquisition time of the complex scenes offsets this disadvantage.

The Acquisition Device "ModelCamera"

The device consists of a digital video camera enhanced with a rigidly attached laser system casting a 7 by 7 pattern in the camera's field of view. The laser source is rated a low-end eye safe class III-a. The position of each laser ray is determined using a separate calibration procedure. The ModelCamera is positioned on a tripod using a special U-bracket (the Wimberley Head [51]) restricting its motion to pan and tilt around video camera's center of projection. The incoming color frames with the computed 3D samples are registered with the evolving model using the color match between the frame and the accumulated panorama mosaic. The panorama obtained with the ModelCamera contains multiple triangulated depth samples that provide motion parallax when viewed from a novel point of view, unlike traditional color-only panoramas.

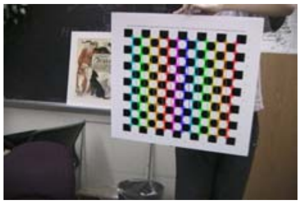




Camera calibration

The ModelCamera components (video camera, epipolar lines, laser beams and the global axis of the tripod) have to be calibrated in four steps.

1. The intrinsic optical properties of the video camera are found using camera Jean-Yves Bouguet's calibration toolkit [5] distributed as part of the Intel Open Source Computer Vision toolkit [21]. A 13 by 13 squares checkerboard is moved in camera's field of view; typically 20 captured images of the board are enough for optical calibration. We reproject the checker corners and use the distance between reprojections and the locations in the image where checker corner were detected as the error metric. Typically, the reprojection error is around 0.25 pixels on average.



2. The epipolar segments are calibrated by automatically detecting the laser dots in multiple images without using epipolar segments and then fitting a line through the grouped dots. Typically 200 frames are captured from a continuous video stream as the operator moves the camera with respect to a wall. The average distance from a detected point to the fitted line is less than 0.4 pixels.

3. The 3D position of each laser is determined using a separate projector casting a checkerboard pattern on a flat white wall. Using epipolar segments the laser dots are detected with the projector turned off. Then with the projector is turned on, the board is detected, and the 3D position of each laser dot with respect to the camera is computed using the observed checkerboard. The camera is moved and the process is repeated. Once several 3D points are accumulated for each laser (5-10 measurements is typical), a 3D line is fitted through them. This is the position and orientation of the laser ray with respect to the camera. The average distance between the laser ray and the 3D points used for fitting is around 0.1 cm.

The first three calibration steps together take less than 10 minutes and have to be done only once. The calibrated properties do not change during the acquisition process, unless the camera is zoomed in/out or the laser diode unscrewed and moved. To prevent the accidental change, the zoom button on the video camera is covered by a protective metal case.

During each acquisition process, the tripod calibration step is performed. Its goal is to discover the global panning axis. At the beginning of the scan, the user is restricted to panning of 15 degrees, and then the calibration step is complete. We believe that this is not a significant burden. The tilting axis in the camera's coordinate system does not change and is parallel to the horizontal scan line of the current image plane.

Frame processing

Each incoming video frame is undistorted to remove the intrinsic optical distortions of the lens. The projection of the laser ray onto the camera's image plane is an epipolar segment. The bright laser dot is visible in the video frame and is quickly located on the corresponding epipolar line (see Figure 1). The laser system is configured to make the epipolar segments disjoint to prevent dot detection ambiguity.

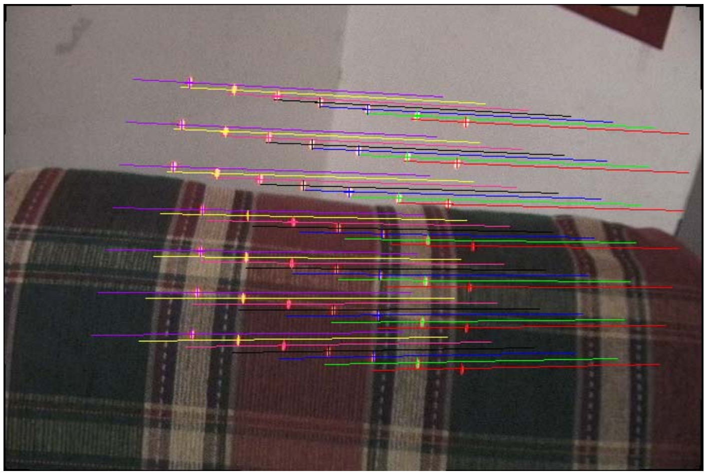
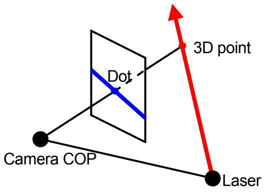


Figure 1: Laser dots are detected on the epipolar segments.



The 3D position of the dot with respect to the camera is given by the intersection of the video camera ray and the corresponding laser ray. In some fragmented scenes, the dots can be hard to detect due to laser scattering, reflection or occlusion; typically we detect 30 – 49 dots in each frame. To eliminate false positive detections, we require that the dot appear k = 3 times at a roughly same 3D position before being added to the model. The depth accuracy of the camera is 3 mm at 1 m distance.

Frame registration using color

Each incoming video frame is registered with respect to the accumulated panorama mosaic (see Figure 2). The frame to panorama registration is robust and drift-free unlike the frame to previous frame registration we used earlier, which agrees with results reported by others [42].

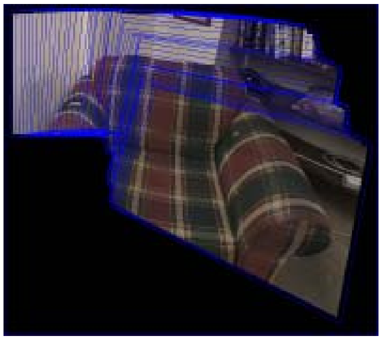


Figure 2: Registered frame sequence.

The registration discovers the pan and tilt angles that minimize the average color difference between the mosaic and projection of the incoming frame (see Figure 3). Only a small subset of all frame's pixels are used to speed up registration. Different subsets were tried; presently a pattern of vertical and horizontal segments is employed (see Figure 4). This pattern is computed for each frame; high-contrast regions of the frame are more likely to be covered by the pattern.

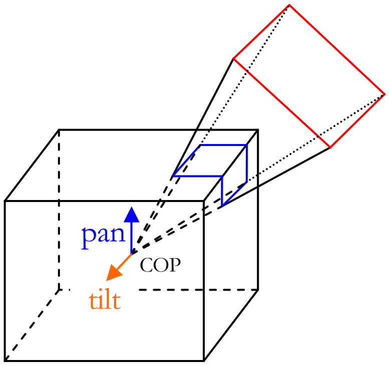


Figure 3: Frame (red) is projected (blue) onto the cubic panorama.

To make registration more robust, the incoming frame is blurred with an 11x11 raised cosine filter, and then down sampled in each dimensions (by a factor of 4 – 8 times) to make the registration faster. The average per channel squared difference is the registration error metric; the value of 20 can be used as an error threshold limit. The smaller the error corresponds to the better visual match between the registered frame and the panorama. The error is sensitive to the values of the two angles. A downhill simplex searching algorithm is employed to speed up the 2D search. We use previously found motion angles as the seed guess for the search.

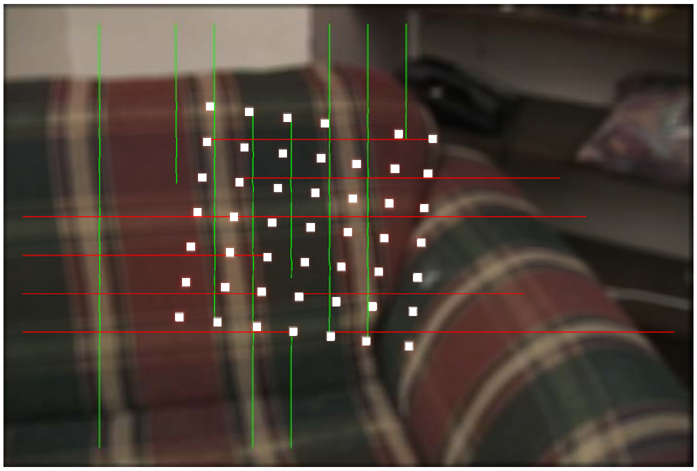


Figure 4: The registration pattern (green and red segments) in the blurred frame. The laser dots were removed from the video frame's texture (white squares) to not affect the registration and the accumulated color.

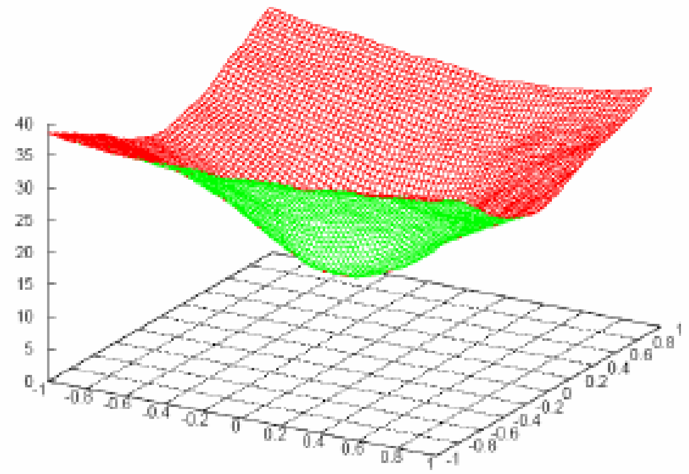


Figure 5: The search valley around the color registration solution (0, 0): the average per channel color difference versus pan and tilt angles.

Frame registration using geometry

We have observed that some parts of the environments do not have enough texture information to achieve the complete registration using the color only (see Figure 6). In this case, if the observed part of the scene is planar, we switch the registration to use the planarity to register the position of the observed patch. The planarity condition is necessary to uniquely determine the orientation of the camera. When the ModelCamera observes the wall, a plane F is fitted through the number of points (see Figure 7). The registration of the consecutive frames will force the points from the incoming frame to lie on the fitted plane F. We have found that the average distance from the points the common plane F is not a good error metric – the search valley is flat around the solution (see Figure 8). To make the registration robust, we fit a plane P through the new frame's points. The dot product between the normal of the plane P and fitted plane is sensitive to both pan and tilt angles and allows good frame registration.

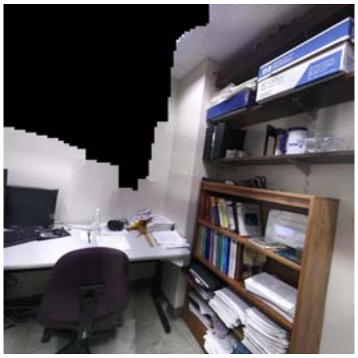


Figure 6: Bare wall above the table has no texture information to allow panorama registration. Such parts of the scene are registered using the geometry.

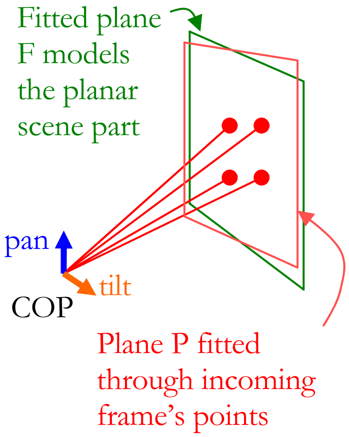

Fitted plane F models the planar scene part
pan
tilt
COP
Plane P fitted through incoming frame's points

Figure 7: The geometrical registration aligns the plane P with the fitted plane F.

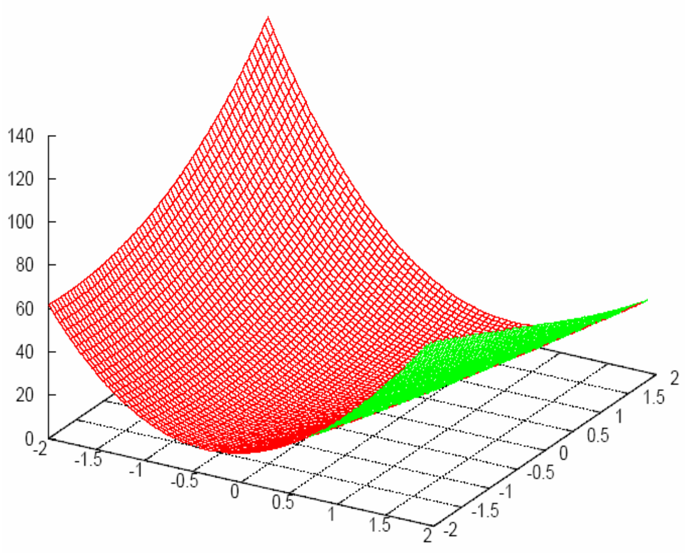


Figure 8: Search valley for geometry registration. Distance from each point to the fitted plane versus pan and tilt. The value is flat around the solution (0, 0).

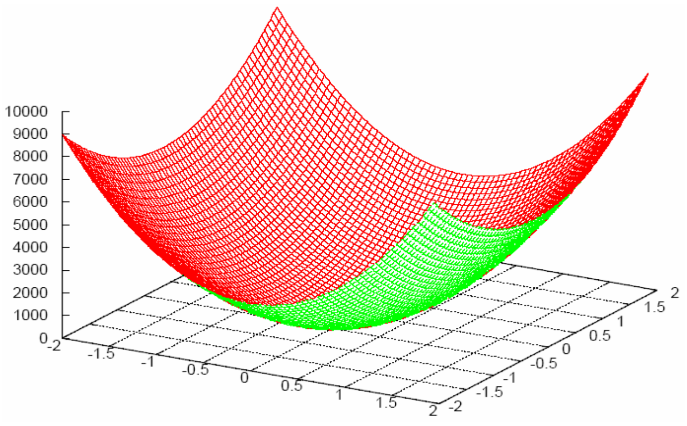


Figure 9: Search valley for geometry registration. Dot product (scaled for clarity) between the normal of the fitted plane F and plane P fitted through points used as the error metric. The search quickly converges to a unique solution.

Incremental panorama update

After the frame registration, the depth samples are added to 3D points accumulated during the scan. The color panorama is modeled as a regular cube with texture mapped faces. The color samples from the registered frame are projected onto the panorama incrementally. The cube's texture is constantly updated with the new color samples from the video frames. Each texture face of the cube consists of smaller square tiles (We use 8x8 pixels tiles for a 1024x1024 pixels cube faces). The registered frame's color samples are used to fill the empty tiles only which speeds up texture update. The video frames demonstrate changing frame to frame contrast and white balance due to the video camera adjusting exposure times. We use weighted color blending to lessen the tiling artifacts due to the changing lightning conditions (see Figure 10).

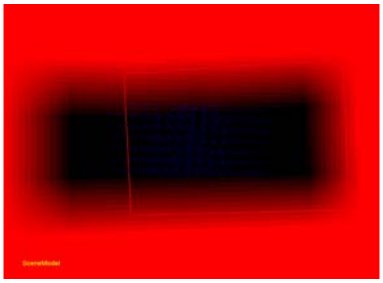
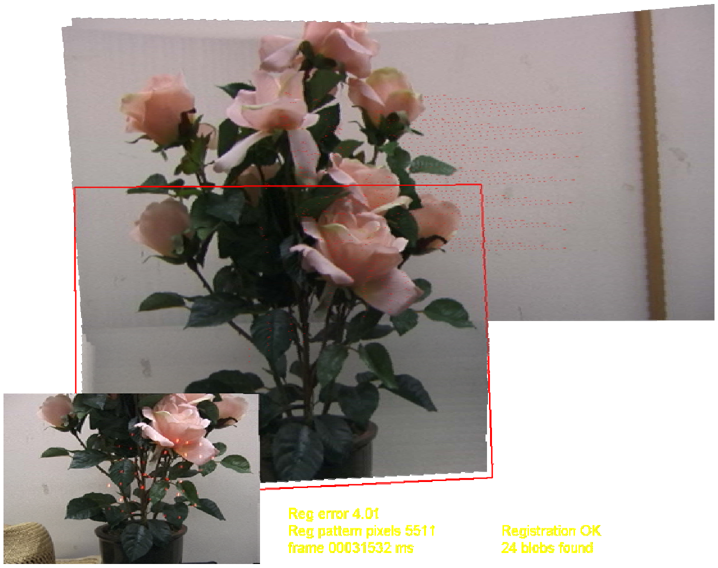


Figure 10: Filled tiles on a face of the panorama cube. Red tiles are empty, black are completely filled. The transition tiles will be blended as new frames are registered.

User feedback during scanning

During the acquisition process the user is shown the texture mapped cube and the 3D point cloud using OpenGL rendering (see Figure 11). We display to the user several parameters: the registration error of the last frame, the size of the last registration pattern, the timestamp of the current frame, and the number of the laser dots found in the current frame.

The user can freely move around the scene and inspect the acquired geometry. The last registered frame is highlighted with the red rectangle. The user can see the camera's current frame in the left bottom corner of the screen; if registration fails (the user moved the camera too fast, or there is not enough texture information to correctly position the new frame) the user can maneuver the ModelCamera to match the camera view with the last registered frame and continue registration.

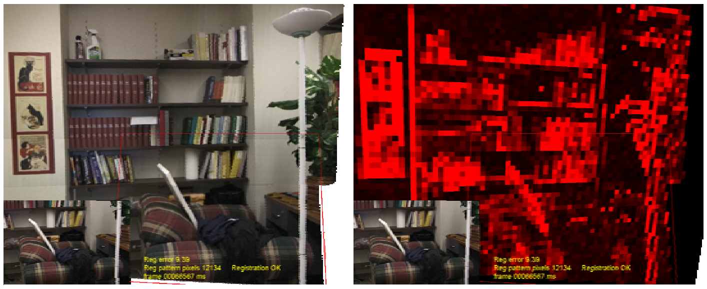


Reg error 4.01
Reg pattern pixels 5511
frame 00031532 ms

Registration OK
24 blobs found

Figure 11: The monitor screen during the acquisition. The current camera frame is in the left bottom corner. The last registered frame is highlighted with red. The small red dots are the registered 3D points.

To better assist the user in panorama registration, there is a feedback mode that displays texture variation of the filled tiles (see Figure 12). For each filled tile of the panorama face we compute maximum pixel-to-pixel change on the middle vertical and horizontal segments. The regions with larger texture variation are better suited for color registration.



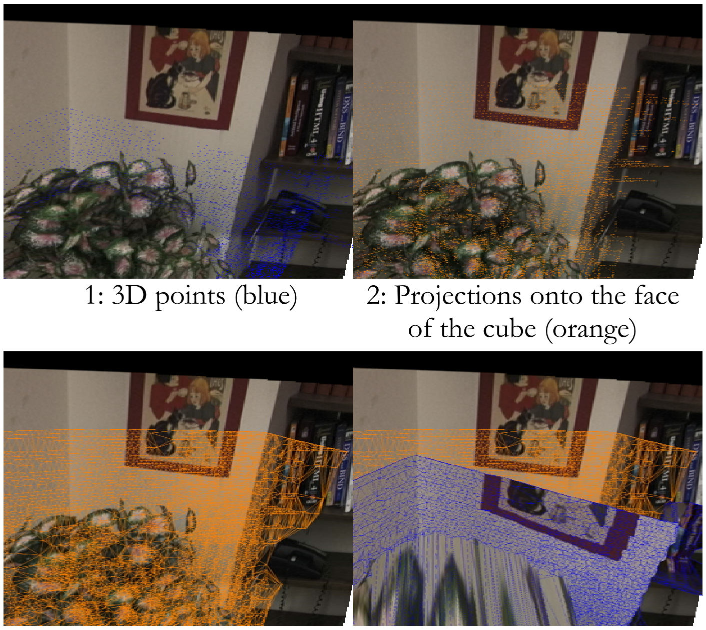
Reg error 9.39
Reg pattern pixels 12134 Registration OK
frame 00066567 ms

Reg error 9.39
Reg pattern pixels 12134 Registration OK
frame 00066567 ms

Figure 12: Panorama color texture (left) and tile texture variation (right). The brighter tiles correspond to larger texture variation in the tile's color. The frame registration will be more robust on the poster on the wall than on the clothes on the armchair.

Modeling with Depth Enhanced Panoramas

After the scanning process is complete the 3D samples are projected onto the faces of the cube. These projections are connected together using 2D Delaunay triangulation (using Triangle, see [43]) and the obtained mesh is used for the corresponding 3D samples (see Figure 13). The errors in the laser dot detection manifest themselves as sudden spikes in geometry and are filtered as a post processing step.



1: 3D points (blue)

2: Projections onto the face of the cube (orange)

3: 2D triangulation on the face

4: Corresponding texture-mapped 3D mesh.

Figure 13: 2D triangulation of the points on the panorama faces.

The depth enhanced panoramas combine the best qualities of the traditional color panoramas (fast, inexpensive acquisition, high quality rendering) without constricting the user to stay at the original center of projection.

Scanned Models

By using ModelCamera in the panoramic mode we have acquired several types of scenes: parts of rooms, furniture, flowers, and clothes. The results can be viewed as VRML models in the results gallery on the project's website [53]. Below are the screenshots of some of the captured models.

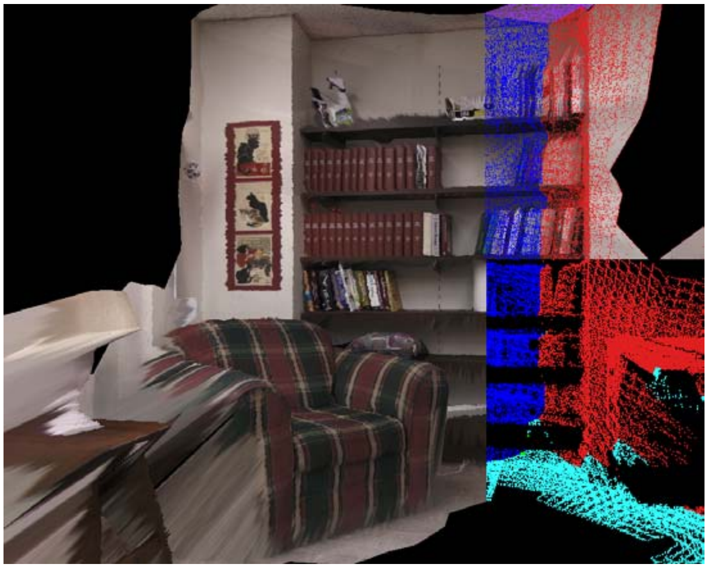


Figure 14: Bottom right: accumulated 3D points assigned to panorama faces. Top right: mesh connecting 3D points. Left: the texture mapped model. Acquisition time: 15 minutes. This model contains 220 000 vertices connected by 420 000 triangles.

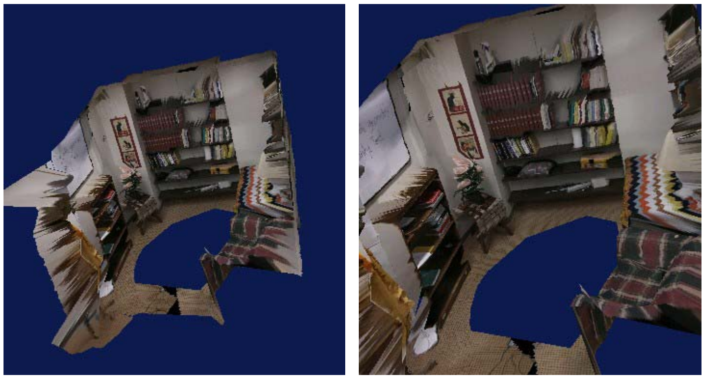


Figure 15: Math 409 with the installed carpet.

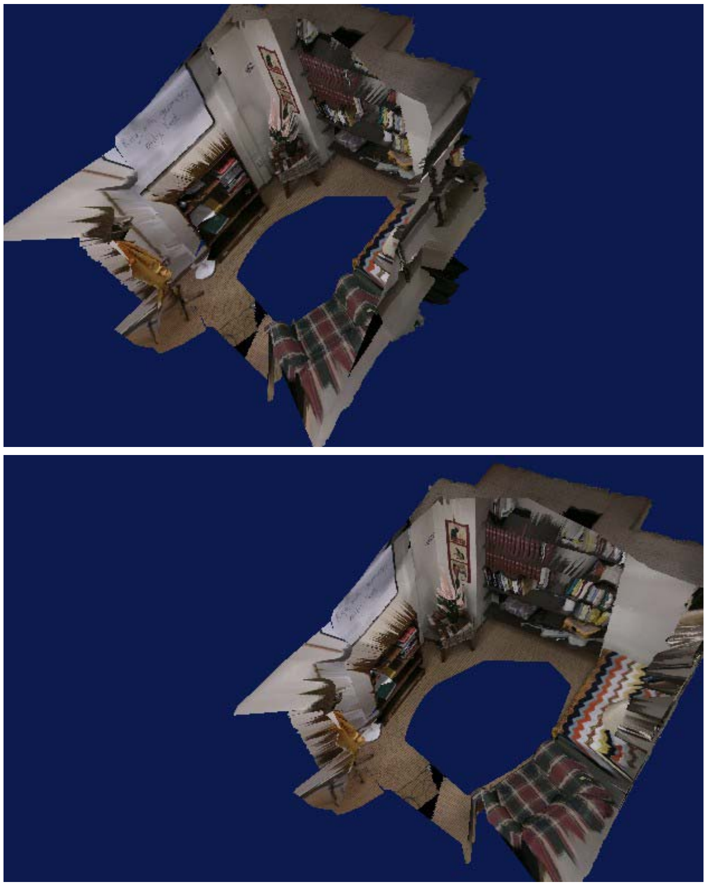


Figure 16: Two views with lateral translation.

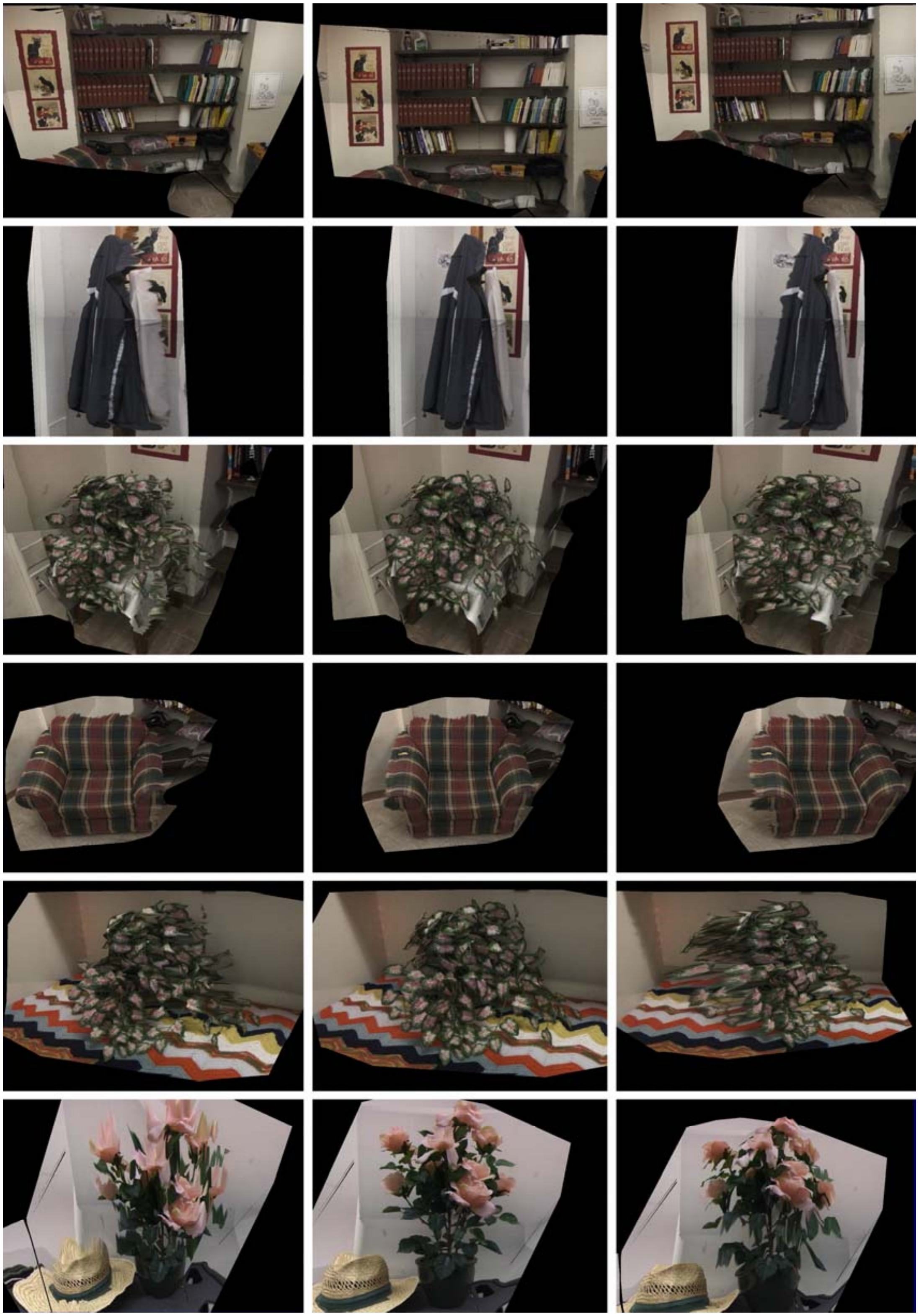


Figure 17: Depth enhanced panoramas of the real world scenes captured with ModelCamera. The center column shows a view close to the camera's original view. The left and right columns show the model from a translated novel view. Each model has been acquired in less than 5 minutes.

Research plan

I have described the current status of the ModelCamera scene acquisition technique. I would like to continue working on this technique to improve it and apply it to the real-world applications. The following list states some of the things that need to be done. There is no doubt the future progress will reveal other research questions that are not evident now.

1. **Offline drift-free panorama registration.** The current online registration suffers through the registration error accumulation over the full revolution. The global offline registration should redistribute the color error among all registered frames in the sequence and eliminate visible mismatches between the last / first frames.
2. **Dynamic range panorama acquisition.** The video camera is constantly adjusting its exposure to accommodate various lighting conditions. The registration algorithm should collect this information and allow the creation of the novel views with the dynamic color range. There are several issues: detecting the change in the camera's exposure, blending the overlapping frames to create seamless transitions, visualization of the dynamic range panoramas.
3. **Merging several depth enhanced panoramas.** A single scan cannot capture all surfaces of the scene, and thus several scans are required to be merged together to produce the complete model. The relevant questions are:
    a. How to register multiple scans together?
    b. How to visualize them efficiently? Can the model be rendered with splats or should the scans be merged to create a single geometrical mesh?
    c. Can view-dependent effects be modeled and visualized with the depth enhanced panoramas?
4. **Registration without texture.** Regions of the scene with uniform color (such as bare walls, ceiling and floor) cannot be registered using the color information alone. The research questions is:
    a. Can the regions without texture be registered by relying on the geometry information alone?
5. **Modeling a single room** (such math 409 office) **with multiple panoramas in less than 1 hour.**
6. **Large scale model acquisition and efficient walkthrough visualization.** Obtain the models of 10 different rooms from the same building in less than 1 day. Position them in the same coordinate building and visualize the walkthrough.
    a. How to combine a coarse model of the building with the depth enhanced panoramas?
    b. How to render such model efficiently on the commodity graphics hardware?

The principle of the ModelCamera work can be used in several applications besides 3D modeling, such as:

1. Real-time capture and modeling of the moving object with the static ModelCamera.
    a. Capture the 3D color model of a simple moving object (ball, cube). While the camera is static, the ball is moving. The camera sees the object from different views, and in real-time registers the visible pieces together using geometrical and color information (the object moves slow enough to provide camera with the views of the overlapping object pieces). Both the 3D geometry of the object and its color and position with the respect to the ModelCamera should be reconstructed in real-time. Successful experiment will lead to the feasibility of a system where several static ModelCamera devices monitor the moving object and reconstruct its position in real-time without need for background blue-screen or object attached fiducials.

2. Hand held interactive color acquisition for fragmented geometrical meshes.
    a. Suppose only the geometry of the fragmented scene has been acquired (with the help of the laser range scanner). The task is to acquire the corresponding color information.
    b. Scan the same scene with the ModelCamera. Determine the camera position by matching the geometry samples with the geometric model.
    c. Spray the color information given by the ModelCamera on the part of the scene observed.
    d. Such interactive color acquisition is feasible if the pose of the ModelCamera can be found precisely from the geometry of the model and depth samples in each frame. The research challenge is to find the pose in the presence of errors in both the model and the depth samples.

3. Navigation device for the visually impaired person.
    a. Attach the solid state miniature video camera (could be black & white camera only) to the laser diode. In each frame the camera will obtain 49 depth samples in its local coordinate system.
    b. This setup could be worn on the blind's persons' wrist. The camera feeds the depth samples to a haptic device, such as a mechanical glove.
    c. A person with the visual disability can thus "touch" the part of the scene that ModelCamera observed. The frame registration is unnecessary; the human operator performs this task instead.

References

1. Agathos, A. and Fisher, R., Colour Texture Fusion of Multiple Range Images. in *4th International Conference on 3D Digital Imaging and Modeling (3DIM)*, (2003).
2. Aliaga, D., Funkhouser, T., Yanovsky, D. and Calbom, I., Sea of Images. in *IEEE Visualization*, (2002), 331-338.
3. Aliaga, D., Yanovsky, D., Funkhouser, T. and Calbom, I., Interactive Image-Based Rendering Using Feature Globalization. in *ACM Symposium on Interactive 3D Graphics*, (2003).
4. Bernardini, F., Martin, I., Mittleman, J., Rushmeier, H. and Taubin, G. Building a Digital Model of Michelangelo's Florentine Pieta. *IEEE Computer Graphics & Applications*, *22* (1). 59-67.
5. Bouguet, J. Camera calibration toolkit for Matlab, 2004, http://www.vision.caltech.edu/bouguetj/calib_doc/.
6. Chen, S., Quicktime VR - An Image-Based Approach to Virtual Environment Navigation. in *ACM SIGGRAPH*, (1995), 29-38.
7. Chen, S. and Williams, L., View interpolation for image synthesis. in *ACM SIGGRAPH*, (1993), 279-288.
8. Debevec, P., Taylor , C. and Malik, J., Modeling and Rendering Architecture from Photographs: A Hybrid Geometry and Image Based Approach. in *ACM SIGGRAPH*, (1996), 11-20.
9. Dech, F. and Silverstein, J., Rigorous Exploration of Medical Data in Collaborative Virtual Reality Applications. in *6th International Conference on Information Visualization (IV'02)*, (2002).
10. El-Hakin, S., Beraldin, J. and Picard, M., Effective 3D Modeling Of Heritage Sites. in *4th International Conference on 3D Digital Imaging and Modeling*, (2003).
11. Fisher, R., Solving architectural modelling problems using knowledge. in *4th International Conference on 3D Digital Imaging and Modeling (3DIM)*, (2003).
12. Furukawa, R. and Kawasaki, H., Interactive Shape Acquisition using Marker Attached Laser Projector. in *4th International Conference on 3D Digital Imaging and Modeling (3DIM'03)*, (2003).
13. Giachetti, A., Tuveri, M. and Zanetti, G., Distributed quantitative evaluation of 3D patient specific arterial models. in *1st International Symposium on 3D Data Processing Visualization and Transmission (3DPVT'02)*, (2002).
14. Gortler, S., Grzeszczuk, R., Szeliski, R. and Cohen, M., The Lumigraph. in *ACM SIGGRAPH*, (1996), 43-54.
15. Gruss, A., Tada, S. and Kanade, T., A VLSI Smart Sensor for Fast Range Imaging. in *IEEE Int. Conf. on Robots and Systems*, (1992).
16. Hebert, M., A Self-Referenced Hand-Held Range Sensor. in *3rd International Conference on 3D Digital Imaging and Modeling*, (2001), 5-12.
17. Hidalgo, E. and Hubbold, R., Hybrid geometric-based rendering. in *Eurographics*, (2002), Computer Graphics Forum, 471-482.
18. Huber, D. and Hebert, M., Fully Automatic Registration Of Multiple 3D Data Sets. in *IEEE Computer Society Workshop on Computer Vision Beyond the Visible Spectrum (CVBVS 2001)*, (2001).
19. Ikeuchi, K., Nakazawa, A., Hasegawa, K. and Ohishi, T., The Great Buddha Project: Modeling Cultural Heritage for VR System through Observation. in *2nd International Symposium on Mixed and Augmented Reality (ISMAR'03)*, (2003), 7-17.
20. Imageworks, Spider-man: Behind the Mask. in *Special sessions at ACM SIGGRAPH*, (San Antonio, TX, 2002).

21. Intel. Open Source Computer Vision Library, 2002, http://www.intel.com/research/mrl/research/opencv/index.htm.

22. Kanade, T., Yoshida, A., Oda, K., Kano, H. and Tanaka, M., A Stereo Machine for Video-rate Dense Depth Mapping and Its New Applications. in *IEEE Computer Vision and Patten Recognition Conference (CVPR)*, (San Francisco, 1996), 196-202.

23. Koch, R., Automatic Reconstruction of Buildings from Stereoscopic Image Sequences. in *Proceedings of the EUROGRAPHICS*, (Barcelona, Spain, 1993).

24. Koninckx, T., Griesser, A. and Van Gool, L., Real-time Range Scanning of Deformable Surfaces by Adaptively Coded Structured Light. in *4th International Conference on 3D Digital Imaging and Modeling (3DIM)*, (Banff, Canada, 2003), 296.

25. Levoy, M. and Hanrahan, P., Light Field Rendering. in *ACM SIGGRAPH*, (1996), 31-42.

26. Levoy, M., Rusinkiewicz, S., Ginzton, M., Ginsberg, J., Pulli, K., Koller, D., Anderson, S., Shade, J., Curless, B., Pereira, L., Davis, J. and Fulk, D., The Digital Michelangelo Project: 3D Scanning of Large Statues. in *ACM SIGGRAPH*, (2000).

27. Malassiotis, S., Aifanti, N. and Strintzis, M., A Gesture Recognition System Using 3D Data. in *1st Internation Symposium on 3D Data Processing Visualization and Transmission (3DPVT'02)*, (2002).

28. Matusik, W., Pfister, H., Ngan, A., Beardsley, P. and McMillan, L., Image-Based 3D Photography Using Opacity Hulls. in *ACM SIGGRAPH*, (2002).

29. McMillan, L. and Bishop, G., Plenoptic modeling: An image-based rendering system. in *ACM SIGGRAPH*, (1995), 39-46.

30. Muller, A., Leissler, M., Hemmje, M. and Neuhold, E., Towards the Virtual Internet Gallery. in *Proceedings of IEEE Multimedia Systems*, (1999), 214-219.

31. NASA. Mars Exploration Rover Mission, 2004, http://marsrovers.nasa.gov/home/index.html.

32. NASA. Mars Pathfinder, 1997, http://marsprogram.jpl.nasa.gov/MPF/.

33. Paquet, E., Viktor, H. and Peters, S., The Virtual Boutique: a Synergetic Approach to Virtualization, Content-based Management of 3D Information, 3D Data Mining and Virtual Reality for E-commerce. in *1st Internation Symposium on 3D Data Processing Visualization and Transmission (3DPVT'02)*, (2002).

34. Park, S. and Subbarao, M., A Range Image Refinement Technique for Multi-view 3D Model Reconstruction. in *4th International Conference on 3D Digital Imaging and Modeling (3DIM)*, (2003).

35. Pollefeys, M. and Van Gool, L. From Images to 3D Models. *Communicatons of the ACM*, *45* (7). 50-55.

36. Pollefeys, M., Van Gool, L., Akkermans, I. and De Becker, D., A Guided Tour to Virtual Sagalassos. in *Virtual Reality, Archaeology and Cultural Heritage (VAST2001)*, (2001).

37. Popescu, V., Eyles, J., Lastra, A., Steinhurst, J., England, N. and Nyland, L., The Warpengine: An architecture for the post-polygonal age. in *ACM SIGGRAPH*, (2000), 433-442.

38. Popescu, V., Sacks, E. and Bahmutov, G., The ModelCamera: A Hand-Held Device for Interactive Modeling. in *4th International Conference on 3D Digital Imaging and Modeling (3DIM)*, (2003).

39. Potmesil, M., Generating octree models of 3D objects from their silhouettes in a sequence of images. in *CVGIP*, (1987), 1-29.

40. Proesmans, M., Van Gool, L. and Defoort, F., Reading Between the Lines - A Method for Extracting Dynamic 3D with Texture. in *Proc. ICCV*, (1998).

41. Rusinkiewicz, S., Hall-Holt, O. and Levoy, M., Real-Time 3D Model Acquisition. in *ACM SIGGRAPH*, (San Antonio, TX, 2002).

42. Sawhney, H. and Kumar, R., True multi-image alignment and it application to Mosaicing and Lens Distortion Correction. in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1999).

43. Schewchuk, J. Triangle: A Two-Dimenstional Quality Mesh Generator and Delaunay Triangulator, 2004, http://www-2.cs.cmu.edu/~quake/triangle.html.

44. Seitz, S. and Dyer, C.R., Photorealistic Scene Reconstruction by Voxel Coloring. in *Computer Vision and Pattern Recognition Conf.*, (1997), 1067-1073.

45. Seitz, S.M. and Dyer, C.R., View Morphing. in *ACM SIGGRAPH*, (1996), 21 - 30.

46. Stumpfel, J., Tchou, C., Yun, N., Martinez, P., Hawkins, T., Jones, A., Emerson, B. and Debevec, P., Digital Reunification of the Parthenon and its Sculptures. in *4th International Symposium on Virtual Reality, Archeology and Intelligent Cultural Heritage*, (Brighton, UK, 2003).

47. Takatsuka, M., West, G., Venkatesh, S. and Caelli, T., Low-cost interactive active monocular range finder. in *Computer Vision and Pattern Recognition*, (1999), 444-449.

48. Vacchetti, L., Lepetit, V., Papagiannakis, G., Ponder, M., Fua, P., Thalmann, D. and Magnenat-Thalmann, N., Stable Real-Time Interaction Between Virtual Humans and Real Scenes. in *4th International Conference on 3D Digital Imaging and Modeling (3DIM)*, (2003).

49. Vlahakis, V., Karigiannis, J., Ioannidis, N., Tsotros, M., Gounaris, M. and Sticker, D., 3D Interactive, On-Site Visualization of Ancient Olympia. in *1st International Symposium on 3D Data Processing Visualization and Transmission (3DPVT'02)*, (2002).

50. Williams, N., Hantak, C., Low, K., Thomas, J., Keller, K., Nyland, L., Luebke, D. and Lastra, A., Monticello Through the Window. in *4th International Symposium on Virtual Reality, Archaelogy and Intelligent Cultural Heritage (VAST2003)*, (2003).

51. Wimberley. Wimberley gimbal-type tripod heads, 2004, http://www.tripodhead.com/.

52. WWW. 3D Sensing and Modeling Links, Signal Analysis and Machine Perception Laboratory, 2004, http://sampl.eng.ohio-state.edu/~sampl/data/3DDB/links.htm.

53. WWW. ModelCamera project webpage, 2004, http://www.cs.purdue.edu/cgvlab/modelCamera/modelCamera.html.

54. WWW. Purdue University Virtual Visit 2.0, 2004, http://www.tech.purdue.edu/resources/map/mapv2/.