Literature review

I have chosen the 2 papers from ACM digital Library :

1. BundleFusion: Real-Time Globally Consistent 3D Reconstruction Using On-the-Fly Surface Reintegration
2. 3D Lite: Towards Commodity 3D Scanning for Content Creation

The second paper was cited in the first paper as per the Assignment Instruction.

The First paper that I have selected deals with designing and constructing a efficient global pose optimization algorithm that can be applied in unison with a large-scale, real-time 3D reconstruction.
At every frame, they continuously run optimization and update the reconstruction as per the newly computed pose estimates. They allow free-form camera paths, instantaneous re-localization, and frequent revisiting of the same scene region. This makes the process more robust toward sensor occlusion, fast frame-to-frame motions, and featureless regions.

The input taken is through a RGB-D stream captured by a commodity depth sensor. To perform global alignment, sparse-then- dense global pose optimization is conducted along with the usage of sparse feature correspondence to obtain a coarse global alignment, as they state that sparse features inherently provide for loop closure detection and re-localization.This alignment is then refined by optimizing for dense photometric and geometric consistency.

Scattered correspondences are established through pairwise Scale-Invariant Feature Transform (SIFT) feature between all input frames , detected SIFT key-points are matched against all previous frames and carefully filtered to remove mismatches.To make real-time global pose alignment tractable, a hierarchical local-to-global pose optimization using the filtered frame correspondences is performed.

On the first hierarchy level, every consecutive n frames compose a chunk, which is locally pose optimized under the consideration of its frames.
On the second hierarchy level, all chunks are correlated with respect to each other and globally optimized. However, instead of analyzing global connectivity once all frames are available, the new method forms chunks based on the current temporal window.

This two-stage optimization strategy reduces the number of unknowns per optimization step and ensures that the method scales to large scenes. Pose alignment on both levels is formulated as an energy minimization problem in which both the filtered sparse correspondences and the dense photometric and geometric constraints are included. They use a fast data-parallel GPU-solver tailored to the problem to solve this highly nonlinear optimization problem.A dense scene reconstruction is obtained using a sparse volumetric representation and fusion, which scales to large scenes in real-time. The continuous change in the optimized global poses necessitates continuous updates to the global 3D scene representation . In order to update the pose of a frame with an improved estimate, remove the RGB-D image at the old pose with a new real-time de-integration step and reintegrate it at the new pose.

Thus, the volumetric model continuously improves as more RGB-D frames and refined pose estimates become available if a loop is closed.

The Steps involved are as follows :

1. GLOBAL POSE ALIGNMENT

1. Feature Correspondence Search :
They compute SIFT key-points and descriptors at 4−5 ms per frame and match a pair of frames in ≈0.05ms
2. Hierarchical Optimization :

Local Intra-chunk Pose Optimization : Intra-chunk alignment is based on chunks of N-chunk = 11 consecutive frames in the input RGB-D stream.

Global Inter-chunk Pose Optimization : The global pose optimization computes the best global alignments for the set of all global keyframes, thus aligning all chunks globally.

2. DYNAMIC 3D RECONSTRUCTION

Key to live, globally consistent reconstruction is updating the 3D model based on newly optimized camera poses. We thus monitor the continuous change in the poses of each frame to update the volumetric scene representation through integration and de-integration of frames. Based on this strategy, errors in the volumetric representation due to accumulated drift or dead reckoning in featureless regions can be fixed as soon as better pose estimates are available.

1 Scene Representation

Scene geometry is reconstructed by incrementally fusing all input RGB-D data into an implicit truncated signed distance (TSDF) representation,. The TSDF is defined over a volumetric grid of voxels; to store and process this data.. This approach scales well to the scenario of large-scale surface reconstruction, since empty space needs to be neither represented nor addressed the TSDF is stored in a sparse volumetric grid based on spatial hashing. Following the original approach, we also use voxel blocks of 8 × 8 × 8 voxels. we allow for RGB-D frames to both be integrated into the TSDF and be de-integrated. In order to allow for pose updates, we also ensure that these two operations are symmetric; that is, one inverts the other.

2 Integration and De-integration

We can thus update a frame in the reconstruction by de-integrating it from its original pose and integrating it with a new pose. This is crucial for obtaining high-quality reconstructions in the presence of loop closures and revisiting, since the already integrated surface measurements must be adapted to the continuously changing stream of pose estimates.

3 Managing Reconstruction Updates

Each input frame is stored with its associated depth and color data, along with two poses: its integrated pose and its optimized pose. The integrated pose is the one used currently in the reconstruction and is set whenever a frame gets integrated. The optimized pose stores the result of the pose optimization.

They have presented a online real-time 3D reconstruction approach that provides robust tracking and implicitly solves the loop closure problem by globally optimizing the trajectory for every captured frame. They combine online SIFT feature extraction, matching, and pruning with a novel parallel nonlinear pose optimization framework, over both sparse features and dense correspondences, enabling the solution of the global alignment problem at real-time rates. The continuously changing stream of optimized pose estimates is monitored and the reconstruction is updated through dynamic integration and de-integration. The capabilities of the proposed approach have been demonstrated on several large-scale 3D reconstructions with reconstruction quality and completeness .We believe online global pose alignment will pave the way for many new and interesting applications.

The Second paper discusses on how to improve the content of 3D scanned image. The breakdown of the Article is that it takes and input from a RGB-D video. Then 3D Lite computes a primitive-based abstraction of the scene, and then leverages this representation to optimize for high-quality texture maps, as well as complete holes in the scene with both geometry and color.

They capture the input RGB-D stream with a handheld, consumer-grade RGB-D sensor. Using a modern RGB-D reconstruction system (i.e.Paper 1), compute initial camera poses for each frame and a truncated signed distance (TSDF) representation of the scene, from which an initial mesh is extracted. Primitive-based abstraction generations are done by detecting planes for each frame, and then merging them into scene primitives according to the camera poses ,Then optimize for the global geometric structure by favoring primitives to support a Manhattan world assumption, so as to encourage orthogonal and parallel structures.

From this lightweight representation of the scene optimization for texture maps over the geometry, directly addressing the issues of motion blur and misalignments. Application of exposure correction to achieve consistent color across the input images, which may vary with auto-exposure and white balancing ,we introduce sparse color feature and geometric primitive constraints to help bring the optimization into the basin of convergence of a dense photometric consistency energy .In order to account for motion blur in input color images, we introduce a method to sharpen color projected from input frames to the model. Rather than select sharpest keyframes throughout the input video, which may still select relatively blurry frames or lose color information by filtering out too many frames, we seek sharp image regions, from which we formulate a graph-cut based optimization for image sharpness and coherence .While general, high-resolution scene completion is a very challenging task, primitive-based abstraction enables effective hole filling for both geometry and color on scene representation ,then complete the color in these regions through image in painting, following Image Melding. This produces a clean, complete, lightweight model mapped with sharp textures.

The Steps involved are as follows :

TEXTURE OPTIMIZATION

Color-based Primitive Refinement : This helps preserve the variance of the color. Since this optimization depends on the quality of the camera transforms, we iterate this color correction step with the color image alignment optimization.

Refined Texture Map Alignment : Key to achieving high-quality texture is obtaining precisely aligned color frames.

Texture Sharpening : Using direct color sampling can still result in noticeable artifacts when transitioning from sampling from one frame to another.

SCENE COMPLETION

Geometry Completion :

- Two planes should be extrapolated to the intersection if the extrapolated area is unobserved.
- If three planes will intersect each other, extrapolate them to meet at a corner if the extrapolated area is unobserved.

- No planes should be extrapolated into open space.
- Holes self-contained in a plane should be filled if they are in unobserved space.

Texture Completion : when there are incomplete regions of a texture which cover a distinct foreground and smooth background, or when there are shadows on the background. In those cases, they detect the background using image segmentation and extend it to the entire image using laplacian smoothing. Then combine the synthesized background with the foreground, and synthesize pixels less than 10 pixels from the foreground using Image Melding.

Finally generate a Mesh.

They have presented a new approach to generating visually compelling 3D reconstructions, 3D Lite focuses on generating high- quality textures on abstracted geometry through texture optimization and sharpening, generating a consistent 3D model with textures that can be sharper than the original RGB images. I believe that this is a first step towards commodity 3D scanning for content creation, and hope that this will pave way to handheld 3D scanning in production.

LIMITATIONS

If there is geometry entirely unseen in the original scan, we cannot generate it from scratch to complete these regions. Additionally, small objects are often projected onto planar primitives where the coarse resolution, noise, and distortion of a commodity depth sensor can also make small objects difficult to distinguish geometrically. so relatively non-planar objects are not captured in geometric abstraction.