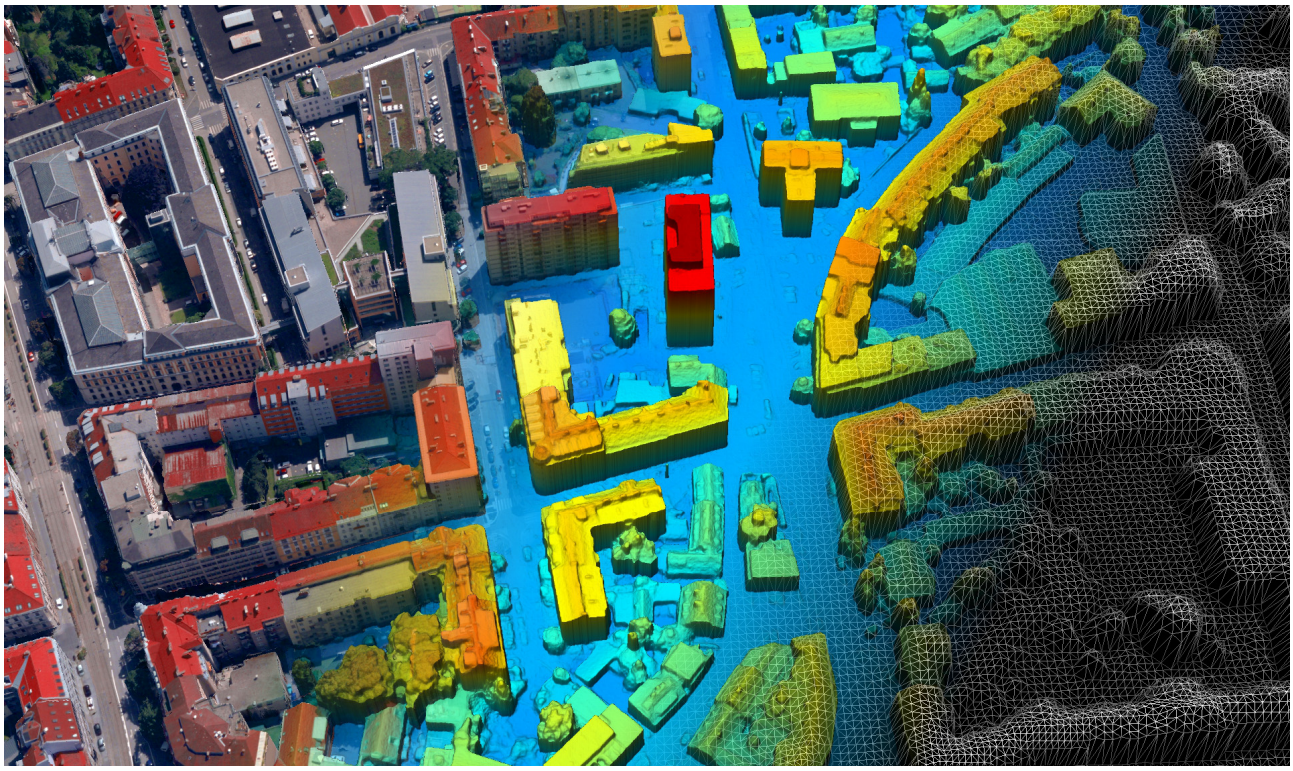
3D Mapping and Photogrammetry

STEPHEN LAWLER, Microsoft Corporation

1. 3D MAPPING AND LOCATION SEMANTICS



The real world base map is moving to 3D to provide visually real and symbolic digital experiences, to enrich your real-world view with augmented reality and to enhance the underlying data and relations. Through the combination of advances in photogrammetry and machine vision, massive scale computing resources, capable device proliferation and powerful natural user interfaces the technological landscape is now prime to embrace 3D.

Pivoting on the "where" dimension and semantically organizing data through a real world location lens requires web scale extraction of entities and all of their digital relations, attributions and properties from traditional web content as well as the vast volume of user generated data. The spatiotemporal aspects of these entities and their physical trajectories will make it a "living" 3D map evolving at the world's course and speed.

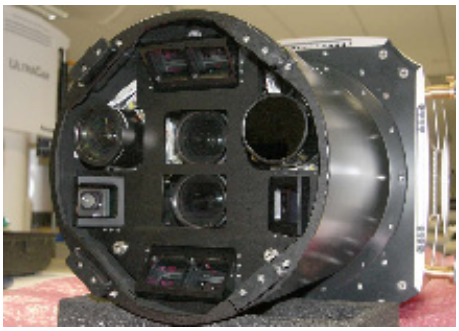
Photogrammetry and machine vision is essential to the creation of the 3D map. Specialized photogrammetric cameras are necessary for consistent, high quality, large scale programmatic collection at low cost. While imagery sources from commercial cameras, phones and wearable capture devices cast a wide net for digital capture extending the reach beyond satellites, planes and cars. Algorithms need to be tuned to leverage known camera ephemeris as well as cope with very little metadata from consumer imagery on the web.

Much of the value of the 3D world will be determined by the design and performance of 3D rendering engines. The rendering engine must provide the freedom to navigate high resolution 3D scenes at fluid frame rates. Ultimately, rendering approaches such as pre-computed 3D models or image based rendering techniques have to balance the compromise of photorealism, level of detail, freedom of movement and viewpoint with the device's memory, computing capability and bandwidth. Different devices, applications and experiences may require different approaches that optimize for their specific use cases.

The new device categories, natural user interfaces and volumes of potential metadata that could be surfaced need to be smartly displayed and optimized for form factor, application and audience. Relevance and context will be essential in determining what gets displayed and how it gets rendered. User interaction needs to be intuitive and a natural extension to the task at hand leveraging the new array of technologies and sensors now available. Transitions between immersive aerial views and first-person streetside and interior views need to be seemless and maintain context. Stylistic rendering techniques that operate against an ontological model and location intelligent semantic graph will create a personalized and contextualized visual canvas. Every urban building needs to be selectable, actionable and a container for objects and information. Rendering the prominent features of interest, enabling data overlay while removing clutter and noise or surfacing the correct actions can only be achieved through a deep understanding of intent and data.

The future 3D real-world trellis and underlying data ontology will provide the visual and data framework for augmenting your real-world view, framing your spatiotemporal exploration and analyzing the world you live, work and play in.

2. 3D MODEL CONSTRUCTION



To date production of automated, precise, photorealistic, large scale models has been costly and somewhat elusive. Precise LIDAR based models were expensive and used for smaller scale, targeted purposes. Satellite imagery based models lacked image resolution and geometric precision. Subsequent resulting artifacts from these sources and less advanced algorithms required arduous and costly manual curation to complete. Today's purposely designed aerial cameras like the UltraCam Osprey are optimized for solving these technical limitations as well as providing a platform to do large scale collection at reasonable costs.

Building an automated pipeline to create high quality visual and geometric 3D models requires a coordinated workflow including camera sensor design, calibration, flight planning, tailored camera

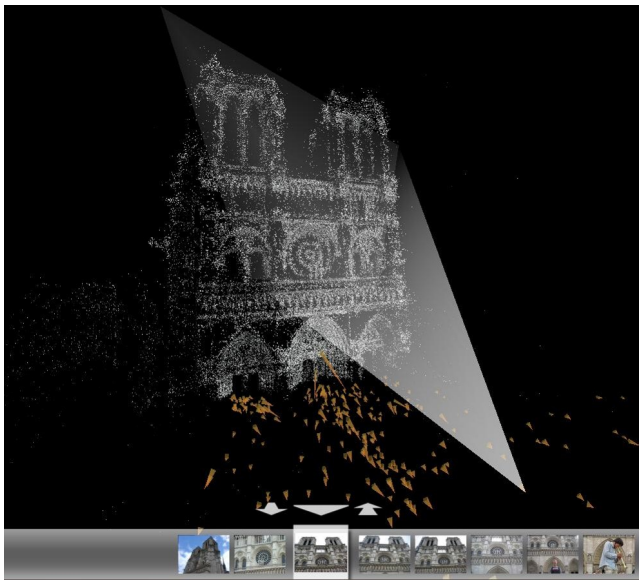
operator software, robust image processing, aerial triangulation, automated processing of point clouds/DSM/DTM, DTM/DSM ortho production, TIN or other 3D model representation.

Microsoft uses a unique dense matching technology for the automated generation of an accurate point cloud with a very high point density. A stereo pairwise matching is applied to all pixels of an overlapping set of images resulting in a highly redundant set of 3D points with multiple observations per pixel. Intelligent filter criteria reduction creates the final highly dense and accurate point cloud. Once the point cloud has been generated, the pipeline of advanced machine vision algorithms generates a high resolution DSM and DTM including high fidelity building and landscape features.

All of these algorithms and bits are processed and handled at massive scale. The pipeline deals with petabytes of data, thousands of compute nodes, thousands of concurrent connections, tens of thousands of cores, billions of individual files, hundreds of thousands of tasks per day at 99%+ uptime. By way of example we recently processed a consistent global ortho of the entire US and Western Europe totaling 12.4M square kilometers. A typical capture area of 10 square kilometers is 1600 to 2000 images and 650 to 800gb per mission. Every 3.6 seconds the sensor captures and generates 417mb of imagery. The project resulted in 1.2 petabytes of raw imagery and 238 terapixels. It was not unusual for the project output to create 10 terabytes per hour at 75-90% core utilization with peaks around 19 terabytes in an hour. The project took 30 months from collection to processing to final product and required magnitude improvements in software automation and sensor design to achieve.

In addition to the workflow, actual acquisition needs to be coordinated for different sources like aerial and streetside as much as possible with regard to season (vegetation, leaf state), lighting conditions (sun angle, cloud cover), time of day, and coverage areas while simultaneously working within country/city flying and driving regulations and dynamic changing factors like weather.

3. FILLING OUT THE REST OF THE 3D TRELLIS



More and more everyday consumers are telling the story of our changing world through the lens of their own cameras. These observations are a valuable source for filling out the 3D map. While

professional acquisition may produce a high quality consistent outcome, the reality remains the same. We need to cast a wider net to capture all of the interesting human scale spaces whether they are outdoor enthusiast spaces, event based urban settings, the countless interior spaces or the extents and reaches that people live and play that simply won't be captured professionally and programmatically. While it's true that the algorithms and models will need to deal with all of this uncertainty, there is definitely a rich source of metadata and pixels to be potentially mined.

The more you zoom in on the world from a satellite picture to aerial to streetside/outdoor to interior the higher rate of change exists. Keeping a fresh copy of the digital world will need the help of many. Whether its machine vision technology like Photosynth or other methods there is much industry advancement in creating 3D scenes from consumer photography.

Both professional, programmatic collection and consumer photography has its strengths and challenges in its contribution to our 3D map. Professional capture provides more consistent quality and predictable coverage but at a high cost. Consumer capture provides freshness, global reach and low cost of acquisition but is unpredictable, spotty and inconsistent quality. Professional sensors are optimized for high pixel count, precise geometry, specialized purposes and radiometric quality under all lighting conditions. Consumer equipment is optimized for user friendliness, low cost and plug-and-play. Professional capture is optimized for specialized vehicles and operational plans. Consumer capture is optimized for supporting a wide range of sensors, uploading and sharing. Professional capture processing is optimized for a known set of inputs. Consumer capture processing is optimized for robustness against a large variation in sensors and input data quality. Ultimately, professional capture is optimized for efficient collection and consumer capture is optimized for high rates of adoption. Leveraging both of these approaches will be key to building a 3D model that serves both head and tail applications and scenarios.

4. IN CLOSING

Ultimately, the convergence of technology between the understanding of intent, new natural user interfaces, the location semantic organization of data, a 3D model and ontology, adaptive 3D contextual rendering, professional and consumer camera technology advances and scalable machine vision algorithms set the stage for our future digital 3D world.