2012 14th Symposium on Virtual and Augmented Reality

Challenges in 3D Reconstruction from Images for Difficult Large Scale Objects
A Study on the Modeling of Electrical Substations

Francisco Simões, Mozart Almeida, Mariana Pinheiro, Ronaldo dos Anjos, Artur dos Santos, Rafael Roberto, Veronica Teichrieb
Voxar Labs, Informatics Center
Federal University of Pernambuco
Recife, Brazil
{*fpms, mwsa, mgmp, rfaf, als3, rar3, vt*}*@cin.ufpe.br*

Clarice Suetsugo, Alexandre Pelinson
Eletrobras Furnas
Rio de Janeiro, Brazil
{*clarices, aop*}*@furnas.com.br*

***Abstract*—In recent years, 3D reconstruction from images has played a major role in computer vision with a lot of improvements regarding both quality and performance. One of its main uses is the generation of 3D models of objects that are difficult to modeling. In the electrical sector, 3D reconstruction from images shows itself as a candidate to be used in specific scenarios with the advantage of its low price compared to laser scanning techniques. In fact, there are many industrial applications that can use the power of 3D reconstruction from images but no work has focused more deeply on their requirements yet. This paper analyzes the advantages and drawbacks in using 3D reconstruction from image techniques and tools in the uncontrolled environment of an electrical substation from the scenario characteristics and tools limitations points of view. Some representative available tools (commercial and academic) where evaluated and the relationship between scenario/object characteristics and reconstructed model quality could be pointed out for further improvements of the techniques in future work. This results can be used to create industrial applications with large scale difficult objects.**

***Keywords*-3D reconstruction from images; large scale objects; industrial applications; tools evaluation**

I. Introduction

In recent years, 3D reconstruction from images has played a major role in computer vision with many improvements regarding both quality and performance in model generation. It has been used from offline city generation with thousands of images to real-time scene modeling and interaction [1][2].

One of the main uses of 3D reconstruction is the automatic generation of 3D representations of objects that are difficult to model, speeding-up the model generation process for graphics applications. This technique can be used both for indoor and outdoor scenes, with some improvements needed to deal with lots of data and uncontrolled environments that commonly influence outdoor scenes [3].

In the electricity sector, there are many efforts to simulate and maintain electrical substations working properly without turning them off. In this scenario, virtual reality and simulation technologies can be applied to analyze working conditions [4] and for that, the 3D models of the equipments are a fundamental component that is not always available due to the age of electrical substations in Brazil, that sometimes were built more than 30 years ago.

Traditionally, because of its model quality and precision and the lack of an efficient cheaper technology, laser based reconstruction has been applied to many industrial applications with great results [5] and it is well accepted by engineers. Although, the main drawback of laser scanning is its price, that even with the maturity of this technique is still high. In this scenario, 3D reconstruction from images shows itself as a potential candidate to be used to generate the 3D models needed by the industry and it is vital to verify the limitations and advantages of these techniques.

In fact, there are many industrial applications in the energy sector that can take benefit of the power of 3D reconstruction. However, as far as the authors know, no work has focused on this domain yet. Recent advances in quality and performance in 3D reconstruction from images techniques will be discussed throughout the paper. This work will also discuss the advantages and drawbacks in using 3D reconstruction from images techniques and tools in the uncontrolled environment of an electrical substation, from the scenario characteristics and tools limitations points of view. Some representative available softwares (commercial and academic) where evaluated and the relationship between scenario/object characteristics and reconstructed model quality could be pointed out.

It will be shown that a scenario like an electrical substation demands from the 3D reconstruction tool the ability to deal with specular and texture less objects, dense populated environments containing large scale equipments, and uncontrolled scenarios with undesired elements for the reconstruction such as the sky and surrounding vegetation, beyond others. Our goal is to provide an analysis that can contribute to further improvements in future works to enable the use of these techniques to create industrial applications with large scale difficult objects, in the electricity sector and others with similar scenarios.

This work is structured as follows. In section II recent

978-0-7695-4725-1/25 $26.00 © 4725 IEEE
DOI 10.1109/SVR.2012.5

improvements in 3D reconstruction of large scale scenes and objects are discussed. In section III the Structure from Motion technique pipeline and the analyzed tools are briefly explained. In section IV the characteristics of electrical substations objects that challenge the reconstruction process are introduced. In section V the reconstruction results obtained from available tools are discussed. Finally, in section VI conclusions and future work are drawn about modeling of large scale objects.

II. Related Work

Most of the effort in the last few years in the area of 3D reconstruction of large scale objects and scenes is dedicated to recover city buildings and monuments [6] to be used in applications like Google Earth/Maps [7] and Microsoft's Bing Maps [8]. It is also addressed to recover historical or archaeological sites to cultural heritage maintenance [9]. The first application domain often acquires data from systems mounted on cars that go through the streets capturing the scene. The second one usually relies on photo collections available through the internet with a huge amount of data captured by different cameras and in different times of the day (web database collections).

Most of the difficulty when working with large scale objects and scenes, such as buildings and streets regards in the camera path that has to be retrieved on the Structure from Motion (SfM) algorithm, used as basis to passive 3D reconstruction from images. In order to overcome these issues, many recent works try to improve the camera tracking quality along a video/image sequence by attaching to the camera other sensors like GPSs (Global Positioning Systems) and/or inertial sensors to stabilize and reduce influence of error accumulation on the great amount of images necessary to reconstruct the scene [3]. Another possibility is to merge reconstruction from images with active range data (acquired from LiDAR - Light Detection And Ranging, for example) [10], but this approach is often more expensive and requires more power and resources to operate since it does rely as well on an active technique (an active technique adds the information that is going to be captured by the sensor), beyond the already used passive one (SfM). Obviously these two approaches do not apply to internet photo collections but to car mounted systems because of the necessity of additional equipment attached to the cameras.

One great source of information in the reconstruction of streets is the fact that the ground is orthogonal to the building walls and therefore they can be easily approximated by planes. Many pipelines were proposed to take advantage of this characteristic by using the sweeping planes approach to go from a sparse reconstruction to a dense one [11] and to labeling the objects based on that sweeping planes [12]. When applied to scenes without dominant planar parts, these approaches do not provide the same quality and show some drawbacks to be used efficiently.

In the mounted car scenario the system captures a sequence of images, and most of the techniques make use of this temporal continuity to perform feature extraction and matching between images with KLT tracker [13] or other temporal approach. The systems which use a web database of images rely on an extractor and matcher that does not depend on temporal information and often use a descriptor-based extractor and matcher such as SIFT (Scale Invariant Feature Transform) [14]. As the junction of different reconstructions is a big challenge in the process of reconstructing large scale objects, some works use a variation of SIFT (2D descriptor) for 3D descriptions that is called Viewpoint Invariant Patches (VIPs) [15]. By this, the system is capable of better relating the partial reconstructions since they rely on a 3D relationship instead of a 2D one [16].

Another important issue that needs to be pointed out about large scale objects, either for streets or for web databases is the high computational demand required for acceptable performance. In a way to achieve real-time or acquisition-time results (acquisition-time results mean that processing time is in the same order of magnitude of video acquisition time), recent systems are using GPU computing to speed-up the process [17] [18]. These works are focusing on many parts of the reconstruction pipeline, from image extraction and matching [19] to most recently bundle adjustment [18].

III. Structure from Motion Techniques and Tools

The Structure from Motion technique require only a set of images to reconstruct the scene. This process can be divided into two stages: sparse reconstruction, that recovers only a subset of points from the visible scene generating a called sparse point cloud, and dense reconstruction, which ideally recovers each point of the visible scene generating a dense point cloud. As traditional SfM pipelines - explained in sequence - are computationally too costly, often the reconstruction process is split into these two phases where the first one generates the base of the reconstruction. Thus, the SfM algorithm reconstructs only the most reliable and stable points of the images based on image processing algorithms and geometrical properties. Then, the second one adds to this first reconstruction enough points to be called a dense reconstruction. It means that the model has almost one 3D reconstructed point for each pixel on the image. This result can be accomplished either using SfM like steps [1] or other multiple view geometry algorithms [3]. An example of sparse and dense reconstruction can be seen in Figure 1.

A. SfM Pipeline for Large Scale Objects

The SfM algorithm allows the 3D reconstruction from images systems to initially recover the motion of the camera and in a second step retrieve the 3D structure of the scene.

Figure 1. Sparse 3D reconstruction on the left and dense 3D reconstruction of the same scene on the right [20].

For many interactive applications, such as augmented reality, the motion part of the pipeline is more important than the structure calculation (sometimes already known before starting tracking [21] or updated only in some frames of the sequence), allowing the system to recover camera poses in real-time in order to add virtual content to the real scene [22]. To applications that are interested in modeling the objects that are part of the scene (only reconstruct the 3D structure of the scene) the second step is the most critical, and often it is not done in real-time because of the amount of information processed. Although, as the quality of the model relies on the first step of the pipeline, these systems have to care about both motion recovery and structure retrieval. The classical SfM pipeline is represented by [23]:

- Feature Extraction and Matching;
- Camera Pose Estimation;
- 3D Point Generation (Triangulation);
- Bundle Adjustment.

In the feature extraction and matching phase, the system has to extract from the images some characteristics [13][14] that could be used to make a relationship between the input images. The relationship is possible by the matching step that looks after features correspondences through the images using detection [14] and/or temporal coherence [13] (sometimes with the use of multiple view properties as fundamental or homography relationships to improve the matching phase). With the matching of the images finished, the system chooses the best camera pair to be used for the estimation of the first camera pose [24] and triangulation [23] (real-time systems wait for the next image of the sequence that is a good pair for the first good image already chosen before starting to reconstruct [22]).

The first camera pair reconstruction is a critical step for ensuring the quality of the generated model since each of the following steps relies on its results. To do so, the fundamental matrix is calculated for the chosen pair by using one of the possible PnP algorithms (eight-point and five-point being the most usual) with some robust estimation technique as RANSAC or LMS (more information about these techniques can be found in [23]).

After that, the calibration matrix can be estimated with some auto calibration method and the essential matrix is computed (if the calibration matrix is previously known, the essential matrix is computed directly from image matches using the same approach of the fundamental). Using the essential matrix it is possible to retrieve the camera poses for the pair and after that the system uses a triangulation technique to recover the 3D points by minimizing the reprojection error on the images.

After calculating all poses and points in the sparse reconstruction stage, the system passes through the bundle adjustment step to refine both cameras and points for multiple view reprojection error minimization. This step has a great impact in the quality of the result due to the great amount of information taken into account and its execution is possible in a reasonable time because of the use of some sparse matrix properties (Sparse Bundle Adjustment - SBA). Some recent works improved the speed of this process by using NVIDIA CUDA programming [18].

B. SfM-Based Tools

The SfM pipeline has been used as basis for the development of some 3D reconstruction from images tools. Some representative tools were chosen in this work in order to perform an analysis of their applicability to reconstruct difficult large scale objects of an electrical substation scenario. The first tool is available for free use (123D Catch from Autodesk [25]), the second one is an academic tool (VisualSFM, developed by the University of Washington at Seattle [26]) and the third one is commercial (Boujou from Vicon, a trial license is used in this work [27]).

Unfortunately, during this research, it was not possible to analyze some of the state of the art techniques for large scale scenes reconstruction mentioned in the related work session because they are not available for download, being used in intern projects of the owner companies and universities.

1) 123D Catch: The 123D Catch beta is a tool released by Autodesk Inc. in 2011 that uses cloud computing to transform digital photos into photorealistic 3D models. For that, the user has to send to a web server multiple digital photos of a static scene that can portray general objects or people. These photographs are processed on the cloud and, after some minutes, the cameras and model data are generated. Differently from other tools, it is not necessary to have lots of pictures of the scene; commonly, 40 are used. More information can be found on the project website [25].

This tool generates a 3D structure of the scene (sparse point cloud), its mesh (dense reconstruction) and a 360° visualization, being possible to apply a texture to the generated model or not. This structure can be visualized within the tool or exported to other know formats, such as the Wavefront .obj format. It also allows the creation

of movies showing the user interacting with the model, selecting specific areas and other viewing functionalities. This is a powerful tool when consistent scene pictures are provided and its execution time is short due to its cloud-based characteristic.

2) VisualSFM: Developed by Changchang Wu in 2011, at the University of Washington, this tool consists of a Graphical User Interface (GUI) application that integrates and improves other works from the University, such as Noah Snavely's Bundler, that is the SfM module of the Microsoft's Photo Tourism [1], responsible for the execution of the initial steps of the reconstruction process, and the Yasutaka Furukawa's CMVS [20], which creates dense models (hundreds of thousands of points) from scenes reconstructed using the Bundler. The CMVS is an extension that converts the Furukawa's previous work, the PMVS2 [28] to be able to handle large input image collections in a more manageable cluster of images. More information can be found on the project website [26].

The main goal of this tool is to generate a visually coherent dense point cloud that could be used in visual applications with the texture associated directly to each point (each point has a color that comes from the reconstruction). Despite that, it is also possible to generate a mesh from the dense point cloud to be used in simulation applications by using simple algorithms since the point quantity is too high (more than 150k points for an object with 4 meters tall and 6 meters far from the camera).

Another positive point is that the VisualSFM tool facilitate the reconstruction process for the user by just clicking some buttons and changing a few parameters, hiding the complexity of Bundler and CMVS, like handling input configuration files and adding a more intuitive and user-friendly interface.

3) Boujou: Another computer vision application that was considered is the Vicon's Boujou [27], which is a consolidated match moving and post production software tool. However, Boujou was not suitable to perform the 3D model reconstruction analysis in this work mainly because it concerns primarily with camera features and the estimation of some scene points, optimizing this process as long as the user has previous information about the scene, i.e., a previous 3D model to estimate the camera pose. Apart from that, Boujou does not have a free academic license, so the authors could only get access to a Trial of the software, which could not have its full capabilities enabled. By that, just the 123D catch and VisualSFM were evaluated.

IV. Electrical Substations Scenes and Objects

An electrical substation is a dangerous place to work. In Brazil, for example, it is necessary to have a specific certification, obtained after a 2 weeks training and valid for only 2 years to enter the energized zone, even for visitors. Another problem when capturing inside a substation is the object's characteristics that can lead to difficulties for SfM techniques. In this section it will be discussed the problems in data acquisition inside electrical substations from the scenario point of view (section A) and objects characteristics (section B).

A. Hazardous Scenario

The scenario of an electrical substation is an extensive outdoor environment with diverse elements that have different scales (towers, transformers, transmission cables, among others) that influence both the data acquisition and the models' reconstruction process. In addition, there are also problems related to the configuration of the environment and its elements. Various equipments in electrical substations are close to each other and/or have relative high dimensions. It causes visibility problems that will trouble the visual capturing process. An example is the occlusion between objects, since it is impossible to shoot a $360\,^{\circ}$ video of the equipment due to lack of access and its dimension.

Another meaningful problem is to get closer to the equipments because of the high level of electricity involved. Some areas cannot be accessed by humans and auxiliary equipment is not allowed because of the high voltage involved. Equipment as stabilizers and rails to improve camera path estimation are not allowed except if they are made from a non-conductor material and after a rigorous inspection by engineers. By that, this work just focused on manually captured scenes using handheld cameras.

B. Difficult Objects

Being an outdoor environment scenario, image-based reconstructions can suffer direct interference of light conditions, mainly because of the sunlight, whose intensity can challenge feature extraction and matching step. Some equipment of an electrical substation are manufactured using materials such as metal and porcelain, as can be seen in Figure 2a. Due to its characteristics, these materials can cause reflection problems, since a specular feature does not obey the projective geometry properties on the object surface leading to false matches.

The greatest problem with these false matches is the difficulty to automatically verify them, even with the use of fundamental matrix relationships, because there are a lot of similar false matches that can cause a bad influence on fundamental estimation by statistical algorithms, such as RANSAC. To deal with this problem some specularity removal techniques [29] could be used to identify probably specular areas and remove them from the matching phase.

Regarding structural characteristics, most substation equipments are basically composed by many regular faces, both the structure (planar faces) and the texture (almost uniform colors), see Figure 2b. This appearance reduces the number of individual features and directly affects the sparse reconstruction result. These surfaces are not usually large

enough in relation to the object itself, leading to problems in the dense reconstruction stage since it is difficult to define a dominant facade for plane sweeping techniques. On the other hand, these planar parts are good for dense approaches that use patch expansion as basis for the process, as the CMVS through the VisualSFM.

Yet related to the equipment's structure, transmission cables and towers, for example. The Figure 2c show that they are composed mostly by thin and elongated parts, which are naturally difficult to detect and identify, causing them to be wrong classified as "noise" in the reconstruction process. Another source of noise is the background that normally is highly texturized. For safety reasons, in Brazil it is usual that electrical substations are built in somewhere isolated, normally with dense vegetation as background, as shown in Figure 2d. The relationship between conventional cameras resolution and background distance turns them into noise because the feature size cannot represent good features.

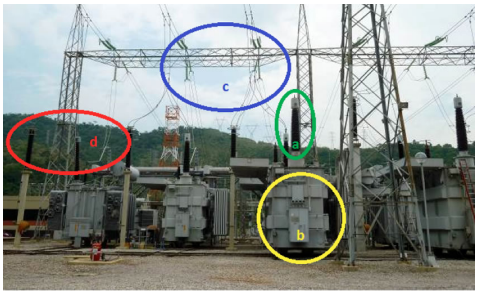

Figure 2. Typical electrical substation: (a) specular surfaces, (b) regular and texture less faces, (c) thin and elongated parts and (d) background distance too far in relation to object size.

V. Tools Analysis

This section shows the main results obtained from the three 3D reconstruction tools described in section III. All the scenes were captured using a conventional handheld camera with a 1080i image resolution and fixed calibration, without any modifications on the environment captured.

Until the time of writing this paper, the authors had no ground truth for any object considered, so the evaluation described in this section is done in form of a comparison between the tools and their capacity of giving adequate results, like number of points generated, number of noisy points generated, computing time. These are general comparison terms that make computational tools as cost-effective as possible. By adequate results is considered the quality of the generated model that should be able to be used in an electrical simulation application which needs a coherent geometrical model (relative depth of points, surface curvature and completeness of the model) and as well in a visualization application where the model needs to have a visual similarity with the real object in sense of texture and approximate geometry (a good texture could approximate the object geometry) [16].

The test cases were given numbered names as labels to easily identify and reference them when convenient. They are listed and exhibited below. For each test case, the result obtained by the 123D Catch and VisualSFM tools is commented and evaluated accordingly.

Since all the test cases were obtained from a movie file, it is natural that the amount of frames is too high to achieve a reconstruction in a feasible computing time. Besides that, the Bundler uses the SIFT feature tracker [14], whose main characteristic is that the frame baseline can be high, i.e., by using photos instead of a video. Therefore, key frames were extracted using the likelihood of each image with its predecessor and, when the images were more than 50% different we chose it as a valid key frame.

A. Test Case 298

This test case object, illustrated in Figure 3, presents a texture less body in its most part. So, for tracking purposes, this texture less regions cannot be easily detected and therefore cannot be reconstructed properly by sparse reconstruction algorithms. In the dense step it is possible to close the empty spaces left by the sparse reconstruction technique because of the great amount of points generated or by a mesh approximation.

Figure 3. Test Case 298 representative frame. There were used 68 frames from the original footage.

Another important issue to consider is the noisy background of this scene, composed essentially by vegetation and other structures not necessarily used for the reconstruction of this object.

Finally, the structure of the scene did not allow to capture a 360° view of the object, but only a partial loop of approximately 150° around the object. Therefore, only a partial reconstruction of the object is possible.

1) 123D Catch: This tool generated a dense reconstruction of the scene and was able to correctly positioning the points on the slick regions in order to generate a mesh, see Figure 4 (top). Although it is possible to observe that some wrong points belonging to the scene were found and tracked leading to unwanted points on the mesh that could be easily manually deleted, see noise

areas in Figure 4 (bottom). The reconstruction process took approximately 15 minutes and generated a mesh of 52k points. Since it was not possible to do a full loop around the object only a partial but coherent mesh is generated. Another point is that this mesh would need a lot of manual processing to be used in an electrical simulation application that needs a good representation of the object (realistic in size and shape). The dense result (texturized mesh from sparse reconstruction) shows to be adequate to be used in a visualization simulation because of the realistic texture and completeness relative to the object and background, but just for some frontal viewpoints.

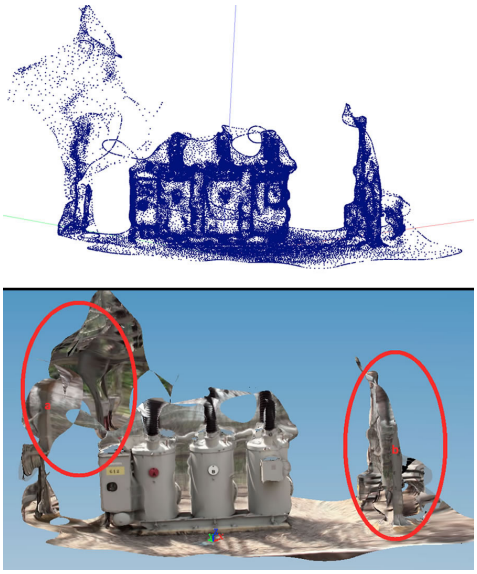

Figure 4. On the top, the final point cloud of test case 298 using 123D Catch without texturing. On the bottom, the final dense reconstruction with a texturized model. Noise areas appears in (a) and (b).

2) VisualSFM: Using this tool, it took approximately 50 minutes to obtain the reconstruction. The number of points generated in the Bundler stage, illustrated in Figure 5, was 9k points. For the dense reconstruction 176k points were generated.

Regarding the sparse reconstruction stage (Bundler) it is possible to observe the absence of smoothness on the object surface but in the dense reconstruction, the CMVS algorithm was able to fill in most of the empty spaces, leading to a realistic model. However, compared to the mesh generated by the 123D Catch (Figure 4), the generated model is not as realistic because of the smoothness on the surface of the object that is more suitable to be generated using a mesh representation instead of a point cloud.

A large quantity of wrong points were tracked and considered as part of the resulting model, since the VisualSFM could not filter those spurious points effectively. This may cause problems to a simulation application but these points can be easily filtered by manual intervention.

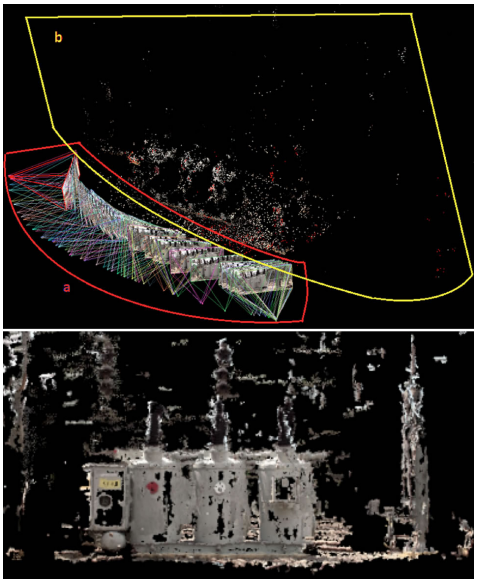

Figure 5. Top image shows the result of test case 298 after the initial reconstruction stage using Bundler. On the bottom, the results after the final dense reconstruction stage using CMVS. Camera path and frames along the footage is highlighted (a), as well as the sparse point cloud (b).

B. Test Case 304

This scene contains a lot of repeated patterns, which could confuse feature trackers, like SIFT or KLT, for instance. Besides that, objects that are not a target to the reconstruction (sky, vegetation) appear in the final model as noise, just because the intended structure cannot be isolated and filmed appropriately.

Figure 6. Test case 304 representative frame. There were used 41 frames from the original footage.

1) 123D Catch: The mesh illustrated in Figure 7 shows a 3D model of approximately 83k points generated from a sequence of 41 selected frames in 15 minutes. The reconstruction generated contains some undesired elements because of the background and complexity of the scene (many objects proportionally close to each other). The vegetation noise also worsened the tool's performance but this could be handled with some improvements on the outlier rejection method of the tool. As in the previous test case, only a partial reconstruction of the objects could be

generated according to the points of view able to be captured during shooting. This result would be a problem to be used on both visualization and simulation applications due to the difficulty in separate the objects from the noise.

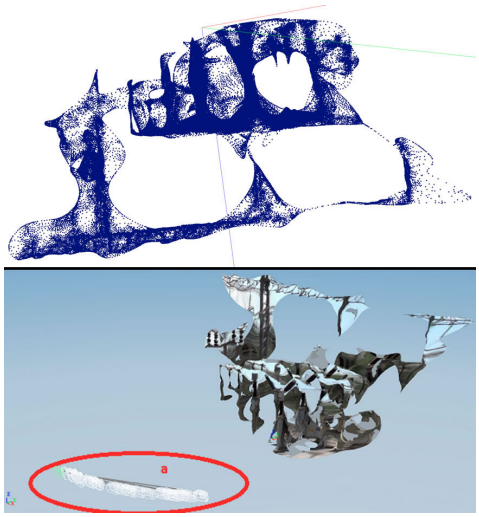

Figure 7. On the top, the final point cloud of test case 304 using 123D Catch without texturing. On the bottom, the final dense reconstruction with a texturized model. Camera path appears in (a).

2) VisualSFM: This test case took approximately 20 minutes to be reconstructed, and the Bundler stage obtained around 4k points and the CMVS increased this number to 81k points.

The Bundler stage calculated the points of the more central structures captured by the camera and some points of the noisy vegetation in the background were also included in the model, causing some difficulty to remove even with manual intervention. The CMVS added more points to the final reconstruction, as shown in Figure 8, but there was an error in the depth estimation of the algorithm in a way that some points pertaining to the sky, the clouds and the vegetation were merged to the models in the foreground.

C. Test Case 328

In this test case the object of interest, illustrated in Figure 9, is big, not only because its front size that is 6 meters by 7 meters, but also relative to the distance to the camera. Such scenarios require a large baseline in order to have a suitable triangulation. This equipment could not be surrounded because of the proximity of other structures that made impossible the complete object capture. Another important characteristic is the planarity of object's surface that gives to a mesh generation tool an advantage over a dense reconstruction one.

1) 123D Catch: The tool generated a mesh with approximately 90k points in 15 minutes. Some noisy points from the vegetation and other structures are tracked and added to the model which creates some difficulty in

Figure 8. Top image shows the result of test case 304 after the initial reconstruction stage using Bundler. On the bottom, the results after the final dense reconstruction stage using CMVS. Observe cameras trajectory as almost a linear footage at the camera path (a).

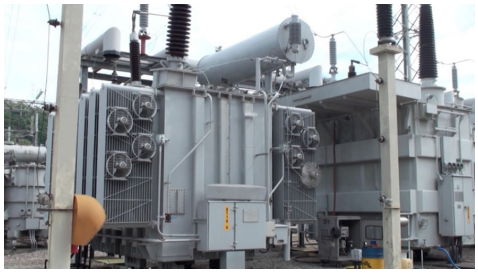

Figure 9. Test case 328 representative frame. There were used 59 frames from the original footage.

identifying the depth between the objects on the point cloud (see Figure 10). In the texturized result it is possible to observe the visual coherence of the texture mapping that results in a useful model for front-parallel visualization. Because of the absence of coherent depth resulted from the front-parallel capture done, this result is not suitable for model generation to be used in simulation applications even with manual intervention.

2) VisualSFM: The total time to execute this test case was approximately 37 minutes. The initial stage returned around 10k points and the CMVS step reconstructed a point cloud of approximately 223k points.

In this case, there is a large translation from the right to the left, trying to encompass a large quantity of information of the structures but, once again, some problems with this scenario arise. For instance, there is not enough space between the objects to enable a surrounding footage - which is more adequate to create a closed 3D model. Besides

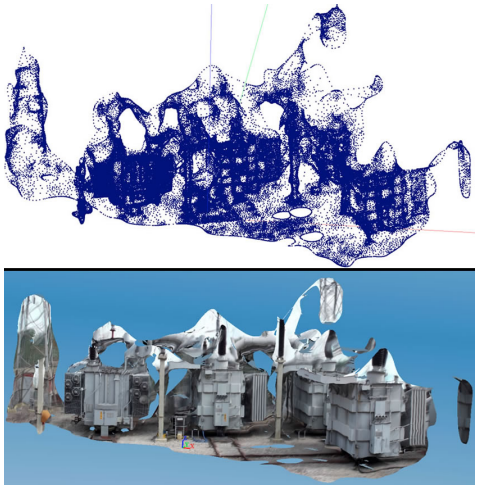

Figure 10. On the top, the final point cloud of test case 328 using 123D Catch without texturing. On the bottom, the final dense reconstruction with a texturized model.

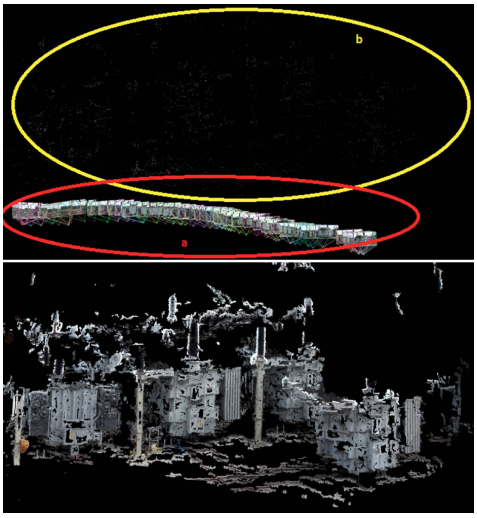

Figure 11. Top image shows the result of test case 328 after the initial reconstruction stage using Bundler. On the bottom, the results after the final dense reconstruction stage using CMVS. Camera path on this footage is almost linear (a) and the sparse point cloud is highlighted (b).

Figure 12. Test case 334 representative frame. There were used 44 frames from the original footage that was a closed loop.

that, the scenario of an electrical substation lacks of an ideal equipment that allows to shoot it isolated from the other objects and from the external world, minimizing the background influence.

The Bundler stage reconstructed some points from the main objects and took some points of the noisy background as if they were from the model. The CMVS continued the process and added more points to the model and reconstructed various points from the sky and the background vegetation merged with the 3D object. By that, this reconstruction could not be used in an electrical simulation application without much effort in a manual intervention in order to separate the interest model from background. It also has a lack of completeness of the model to be used in visualization applications because even the dense reconstruction has a lot of empty regions, requiring a post processing step to close them. Figure 11 show the results.

D. Test Case 334

In this test case, illustrated in Figure 12, it was possible to perform a 360o shooting around the equipment, leading to a very nice reconstructed model. Unfortunately this is not an interesting object from an electrical substation scenario point of view, because it is just a machine to pull electrical transformers. On the other hand, this is a good case to understand the power of 3D reconstruction from images when applied in adequate scenarios (available closed loop, non-occluded parts, textured object and planar elements for dense reconstruction).

1) 123D Catch: The 123D Catch tool used the selected 44 frames and generated a great mesh with approximately 155k points in 15 minutes. This result is the best generated model of the four test cases because of the great scene characteristics. The object has a great texture and the scene shows few sun reflections. The floor is suitable for tracking algorithms like SIFT or KLT because of its non-repetitive highly texturized surface, although it could be confused as noise. In addition, the floor is a plane surface that contributes to the mesh generation, as can be seen in Figure 13. This result could require just a few refinements through manual intervention (i.e. separate object from ground) to be used in an electrical simulation application and it is as well adequate for visualization applications in virtual and augmented reality interfaces [4].

2) VisualSFM: The approximate time of execution of this case was 10 minutes. The Bundler stage returned around 29k points, as can be seen in Figure 14 with the camera trajectory around the object correctly recovered, and the CMVS increased this number to 259k points.

As it can be noticed, the footage was performed around the object and the Bundler could get enough information to allow the dense reconstruction of the CMVS to work properly. Despite that, there are a lot of empty areas on the

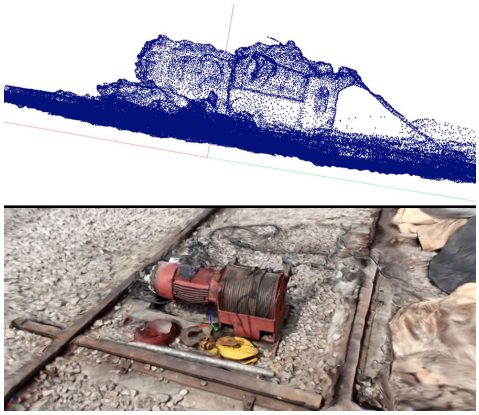

Figure 13. On the top, the final point cloud of test case 334 using 123D Catch without texturing. On the bottom, the final dense reconstruction with a texturized model.

Figure 14. Top image shows the result of test case 334 after the initial reconstruction stage using Bundler. On the bottom, the results after the final dense reconstruction stage using CMVS. Observe the camera loop around the object and the sparse point cloud in the center.

scene floor because the dense reconstruction algorithm does not use the information that the ground is a plane. With some improvements on the dense reconstruction generation, like the number of expansions in point generation, the resulting scene could be used properly in visual applications. Beyond that, the object is already well modeled for simulation and visualization applications due to its great amount of points (259k) that are very dense giving the impression to be a mesh (Figure 14).

VI. Conclusion

This work provides an original analysis about 3D object reconstruction from images applied to electrical substations modeling from the perspective of scene and object characteristics. Apart from the great advances in recent years, there are still a lot of improvements that this kind of techniques should receive in order to improve the quality of generated models. Some improvements made tackling urban scene modeling and cultural heritage applications could not be applied to this scenario because of some restrictions of the substation equipment, as the absence of great planarity on the equipment surfaces for plane sweeping. The capability of surrounding the objects of interest and separate them from background and other close objects is still a challenge to overcome. In order to solve that, projective geometry relationships could be used based on the analysis made in this paper about noise occurrence.

Some specularity removal techniques could be used to identify possible problematic regions to tracking algorithms and could be incorporated as a filter mask to improve matching phase. Another possibility is to use texture segmentation to improve matching algorithms since the objects of interest do not occupy the entire scene and are known. Some improvements using parallel processing in GPUs could also be used to speed up some algorithms, for example the VisualSFM performance that takes over 30 minutes in some scenes.

The usual available tools for 3D reconstruction from images are still not able to properly generate models of objects in an electrical substation scenario for simulation purposes because of its complexity. On the other hand, for visualization purposes the reconstruction from images is already the best solution available, being used as a complement to LIDAR techniques to capture textures. By all exposed, the authors believe that in a near future, 3D reconstruction from images could be used to achieve useful results for simulation applications in electrical substations.

Acknowledgment

The authors would like to thank Eletrobras Furnas for funding this research project. Francisco Simões and Artur Lira also thank CNPq for financial support.

References

[1] N. Snavely, S. M. Seitz, and R. Szeliski, “Photo tourism: exploring photo collections in 3d,” in *ACM SIGGRAPH 2006 Papers*, ser. SIGGRAPH ’06. ACM, 2006, pp. 835–846.

[2] R. A. Newcombe and A. J. Davison, “Live dense reconstruction with a single moving camera,” vol. 21, no. 2. IEEE, 2010, p. 14981505.

[3] M. Pollefeys, D. Nistér, J. M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. J. Kim, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewénius, R. Yang, G. Welch, and H. Towles, "Detailed real-time urban 3d reconstruction from video," vol. 78, no. 2-3. Kluwer Academic Publishers, Jul. 2008, pp. 143–167.

[4] G. R. Rey, J. M. Ibánez, J. F. Mindán, J. M. C. Becerra, M. L. M. Muneta, and A. M. C. Díaz, "Virtual reality applied to a full simulator of electrical sub-stations," vol. 78, no. 3. Elsevier, March 2008, pp. 409 – 417.

[5] C. Fröhlich and M. Mettenleiter, "Terrestrial laser scanning new perspectives in 3d surveying," *Archives*, vol. 36, no. Part 8, pp. 7–13, 2004.

[6] A. Akbarzadeh, J.-M. Frahm, P. Mordohai, B. Clipp, C. Engels, D. Gallup, P. Merrell, M. Phelps, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewenius, R. Yang, G. Welch, H. Towles, D. Nister, and M. Pollefeys, "Towards urban 3d reconstruction from video," in *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, ser. 3DPVT '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 1–8.

[7] I. Google, "Google maps website," 2012. [Online]. Available: http://maps.google.com/

[8] C. Microsoft, "Bing maps - driving directions, traffic and road conditions," 2012. [Online]. Available: http://maps.bing.com

[9] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski, "Building rome in a day," in *IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 - October 4, 2009*. IEEE, 2009, pp. 72–79.

[10] C. Früh and A. Zakhor, "An automated method for large-scale, ground-based city model acquisition," vol. 60, no. 1, Oct. 2004, pp. 5–24.

[11] D. Gallup, J.-M. Frahm, P. Mordohai, Q. Yang, and M. Pollefeys, "Real-time plane-sweeping stereo with multiple sweeping directions," in *IEEE Conference on Computer Vision and Pattern Recognition (2007)*, vol. 16, no. 5 Suppl. IEEE, 2007, pp. 171–178.

[12] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer, "Fast global labeling for real-time stereo using multiple plane sweeps," in *Proceedings of the Vision, Modeling, and Visualization Conference 2008, VMV 2008, Germany, October, 2008*. Aka GmbH, 2008, pp. 243–252.

[13] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2*, ser. IJCAI'81. Morgan Kaufmann Publishers Inc., 1981, pp. 674–679.

[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," in *Int. J. Comput. Vision*, vol. 60, no. 2. Kluwer Academic Publishers, Nov. 2004, pp. 91–110.

[15] B. Clipp, J.-M. Frahm, and M. Pollefeys, "3d model matching with viewpoint-invariant patches (vip)," in *IEEE Conference on Computer Vision and Pattern Recognition (2008)*, vol. 0, no. 6. Ieee, 2008, pp. 1–8.

[16] P. Marc, Jan-MichaelFrahm, F. Friedrich, Z. Christopher, W. Changchang, C. Brian, and G. David, "Challenges in wide-area structure-from-motion," in *IPSJ Transactions on Computer Vision and Applications(CVA)*, vol. 2, nov 2010, pp. 105–120.

[17] S. Choudhary, S. Gupta, and P. J. Narayanan, "Practical time bundle adjustment for 3d reconstruction on the gpu," in *ECCV2010 Workshop on Computer Vision on GPUs (CVGPU2010)*, 2010.

[18] K. Ni, D. Steedly, and F. Dellaert, "Out-of-core bundle adjustment for large-scale 3d reconstruction," in *Computer Vision, IEEE International Conference on*, vol. 0. Los Alamitos, CA, USA: IEEE Computer Society, 2007, pp. 1–8.

[19] J.-m. Frahm, "Gpu-based video feature tracking and matching," in *EDGE Workshop on Edge Computing Using New Commodity Architectures*, vol. 278. Citeseer, 2006, pp. 695–699.

[20] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski, "Towards internet-scale multi-view stereo," in *CVPR*, 2010.

[21] Y. Park, V. Lepetit, and W. Woo, "Texture-less object tracking with online training using an rgb-d camera," in *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality*, ser. ISMAR '11. IEEE Computer Society, 2011, pp. 121–126.

[22] G. Klein and D. Murray, "Parallel tracking and mapping on a camera phone," in *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality*, ser. ISMAR '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 83–86.

[23] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003.

[24] M. T. Ahmed, M. N. Dailey, J. L. Landabaso, and N. Herrero, "Robust key frame extraction for 3d reconstruction from video streams," in *VISAPP*, vol. 1, 2010, p. 231236.

[25] I. Autodesk, "Autodesk 123d - 123d catch turn photos into 3d models," 2012. [Online]. Available: http://www.123dapp.com/

[26] C. Wu, "Visualsfm: A visual structure from motion system," 2011. [Online]. Available: http://www.cs.washington.edu/homes/ccwu/vsfm/

[27] I. Vicon, "Boujou: The first choice for professional matchmovers," 2012. [Online]. Available: http://www.vicon.com/boujou/

[28] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32. Los Alamitos, CA, USA: IEEE Computer Society, 2010, pp. 1362–1376.

[29] H.-L. Shen and Q.-Y. Cai, "Simple and efficient method for specularity removal in an image," in *Applied Optics*, vol. 48, no. 14. OSA, May 2009, pp. 2711–2719.