Please fill in the following information after you read the paper.

[Paper tile] Towards Internet-scale Multi-view Stereo

[Summary] Describe the key ideas, experiments, and their significance.

The paper describes a new approach to scale up the existing multi-view stereo algorithms to work on image collections found on the Internet which are usually unstructured. The authors specifically focus on the multiview stereo problem which suffers from scalability issues if it is to work on internet-scale. For that they propose to solve it using overlapping view clustering and specifically how to do overlapping view clustering? Their goal in this paper was to divide the huge set of images in clusters which have small overlap between individual clusters. They also propose to design filtering algorithms which shall be used during reconstruction (TBD using MVS) to solve quality issues and reconstruction errors.

The authors propose a clustering algorithm and define a view clustering problem which helps in clustering images based on a set of constraints. The authors also assume that the input images have already been processed by a SfM algorithm like the one describe in Agarwal et al's Building Rome in a Day paper. This in turn gives them camera poses and SfM points which are useful in defining constraints as seen in the paper. This constraints that govern this cluster ensure that there are no redundant images or *compactness*, they also ensure that the cluster only contains a certain number of images so that they can be processed efficiently or *size constraint,* the images in the cluster also ensure that there is minimal loss in detail and the visibility is not compromised also termed as *coverage*, in the paper. The significance of all this constraints are that these ensure overlapping between the clusters, it also minimizes redundant images (which are discarded) and it incorporates image quality factors as images of poor quality would anyway have lesser SfM points. The final step before applying MVS is to run the view clustering algorithm. Firstly, it merges SfM points, which in turn helps in reducing the number of SfM points to be considered for reconstruction. This is done to take into account images which are not properly matched because of undetected features. Second step is to remove redundant images from the image set until the coverage constraint remains valid. The third and fourth steps are performed in a loop, they again account for size constraint and coverage until a cluster set is formed which satisfies both the constraints. The significance of this approach is that splitting the clusters first and adding images in each iteration tends to rapidly converge and satisfy the constraints like coverage and size while achieving moderate compactness. This is verified in their experiments as this approach helps validate constraints in few iterations.

The MVS is done using the PMVS software to reconstruct 3D points. The authors propose two filters to merge the reconstructed points. First is the quality filter, since same regions can be viewed in multiple cluster of images with varying quality depending on distance from the object. This filter ensures that the cluster which produces dense and accurate 3D points is considered and the rest of the points are filtered out. The visibility filter counts the inconsistencies of 3D points (or MVS points) between clusters, if a 3D point conflicts >3 times it is filtered out. This is done using Depth maps generated for each image in a cluster.

The algorithm is tested on several datasets which reconstruct the Basilica, Trevi Fountain, Colesseum, Piazza San Marco, results of which are printed in the paper. The experimental results show that the authors gain a significant time improvement in processing the images and merging the reconstructions. To further validate their clustering approach they test their software by replacing the PMVS with Depth Based MVS algorithm and achieve an improvement of almost 50% in execution time, which further validates their image selection and clustering approach. Their last experiment iterates over different values of cluster size upper bound alpha, α. It is observed as alpha decreases or becomes too small the execution slows down because of excessive overlap between the images.

[Strengths] Consider the aspects of key ideas, experimental or theoretical validation.

The main strengths of this algorithm are described as follows:
1)Addresses the challenge of scalability, which is critical for using large unstructured internet applications
2)The algorithm is designed to be run out of core and parallelizable which reduces execution time significantly.
3)The algorithm removes redundant images which might be taken from almost similar view points which have relatively low baseline between the cameras which in turn results in removal erroneous depth points.
4)The algorithm prevents under sampling of surfaces near cluster boundaries by confirming through coverage constraint that there is an overlap between cluster of images.
5)The main advantage of the algorithm specifically the view clustering algorithm is that it is compatible and capable of handling variety of MVS algorithms as demonstrated in experimental results by the authors. (Used Depth Based MVS from Goesele's method)

[Weaknesses] Consider the aspects of key ideas, experimental or theoretical validation, writing quality, and data contribution (if relevant). Explain clearly why these are weak aspects of the paper

1)Although it is claimed in the paper that the algorithm achieves high quality point cloud in lesser execution time, the author seems to have missed the consideration of time required to compute SfM points and Camera poses which they assumed they had in advance.
2)The paper doesn't address other challenges like handling complex Bidirectional reflectance distribution function and lighting variations which could be another quality feature in a architectural reconstruction
3)There is no benchmark dataset on which this view clustering algorithm was tested.

[Reflection] Share your thoughts about the paper. What did you learn? How can you further improve the work?

The paper overall presents the first approach for large scale unstructured image data based 3D reconstruction at city scale. This paper also led me on many detours to explore techniques explored in cited papers like SfM, Patch based MVS and I learnt an overview

of this classical 3d reconstruction algorithms. The algorithm isn’t tested on benchmark dataset which could be a further improvement on this paper.