Overlapping camera clustering through dominant sets for scalable 3D reconstruction

Massimo Mauro[1]
http://www.ing.unibs.it/~massimo.mauro/
Hayko Riemenschneider[2]
http://www.vision.ee.ethz.ch/~rhayko/
Luc Van Gool[2]
http://www.vision.ee.ethz.ch/~vangool/
Riccardo Leonardi[1]
http://www.ing.unibs.it/~leon/

[1] Department of Information Engineering
University of Brescia
Brescia, Italia

[2] Computer Vision Lab (CVL),
Swiss Federal Institute of Technology (ETH)
Zurich, Switzerland

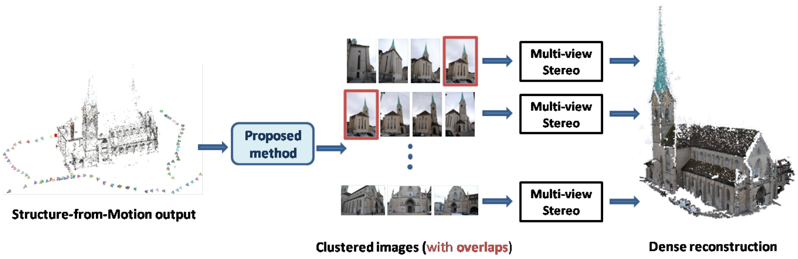

Figure 1: The overall scheme of a reconstruction using camera clustering.

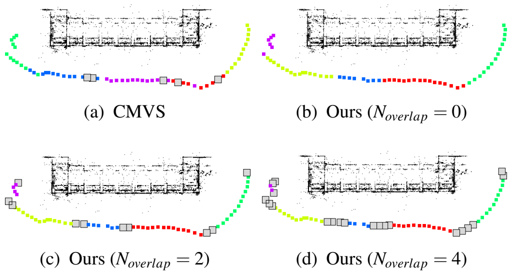

Figure 2: Camera clusters on the *Hall* dataset with overlaps in gray.

Scalability is a great issue in modern large-scale 3D reconstruction pipeline [1, 2]. Recently, *image clustering* [3, 5] and *image selection* [4, 7] methods have been developed for scaling both Structure-from-Motion (SfM) and Multi-View Stereo (MVS) algorithms. This paper focuses on MVS scalability and proposes a novel camera clustering method. Our technique produces a set of overlapping clusters suited to be processed independently by the consequent MVS (see Fig. 1). Overlaps are important to avoid the creation of holes between clusters and their benefit for the final reconstruction is analyzed in the paper.

Our clustering algorithm groups together cameras pointing to similar parts of the scene. We model our problem as a graph, where the nodes are the cameras and the edges represent the similarity between views. Given a set of N cameras and a 3D sparse point cloud obtained from SfM we define a $N \times N$ symmetric matrix W of camera similarities between as

$$w_{ij} = \frac{\sum_{p \in (V_i \cap V_j)} w_{\alpha_{ijp}}}{| V_i \cap V_j |} \tag{1}$$

where $w_{\alpha_{ijp}} = exp(-\frac{\alpha_{ijp}^2}{\sigma^2})$ depends on the angle α between the viewing directions. See full paper for details of the method.

Once the graph is constructed, we adopt the game-theoretic model of dominant sets for clustering [6]. Dominant set clustering works by iteratively evaluating the coherency of the graph nodes based on the similarity matrix W_0 (where W_0 is equal to the W matrix without any self-loops, i.e. with zeros along the main diagonal). The assignment of each camera is determined by the *participation vector* $\mathbf{x}$, which is of length N as the number of cameras. It contains the level of participation - in the range $[0, 1]$ - of each camera into clusters. The algorithm for finding a dominant set (a cluster) is the following:

1. **Initialize** all elements of $\mathbf{x}$ to $1/N$;
2. **Evolve** the system

$$x_i(t+1) = x_i(t) \frac{(W_0 \mathbf{x}(t))_i}{\mathbf{x}(t)^T W_0 \mathbf{x}(t)} \tag{2}$$

3. **Stop** when

$$\mathbf{x}(t+1)^T W_0 \mathbf{x}(t+1) - \mathbf{x}(t)^T W_0 \mathbf{x}(t) < \varepsilon \tag{3}$$

When the algorithm terminates, the vector $\mathbf{x}$ has zero values for non-relevant nodes and values above zero for nodes in the dominant set. We obtain a multi-cluster division by iteratively running the method on the remaining set of cameras, until no further separation is possible.

Dominant set clustering can be implemented with few lines of code and executed efficiently. A further main advantage of using dominant sets clustering is that additional constraints are naturally integrated, such as

- *minimum size constraint*: every cluster must contain at least three cameras, for a robust dense stereo.
- *maximum size constraint*: every cluster must be smaller than a specified N_{size}. This constraints allows to run memory expensive reconstructions on machines with limited memory capabilities.
- *overlap constraint*: every cluster must define a number $N_{overlap}$ of overlapping cameras, which lie at the boundaries of other clusters.

Contrary to related work, our method is explicitly developed with an easy integration of selection in mind: paying attention to not remove overlapping cameras (Fig. 2), any image selection technique can be computed independently on every cluster without worrying about the creation of holes between clusters. See the paper for the details about dominant sets clustering and constraints satisfaction.

In our experimental evaluation we show both quantitative (measuring the reconstruction quality and the runtimes) and qualitative results (illustrating the compactness of the clusters and the overlaps configuration). Our method leads to significant speedup factors (up to 6) for the dense multi-view reconstruction with respect to considering all images at once, while maintaining high reconstruction quality. When compared to the state-of-the-art CMVS clustering method [3], we show our clusters to have a cleaner separation and better overlaps (Fig. 2).

In summary, our method provides an alternative formulation of view clustering, which directly incorporates overlapping cameras and speeds up the subsequent Multi-View Stereo methods by providing cleaner yet well-connected clusters for parallelization.

Acknowledgments. This work was supported by the European Research Council (ERC) under the project VarCity (#273940).

[1] S. Agarwala, Y. Furukawaa, N. Snavely, I. Simon, B. Curless, S. Seitz, and R. Szeliski. Building rome in a day. *Commmunications of ACM*, 54(10), 2011.

[2] J-M. Frahm, P. Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y-H. Jen, E. Dunn, B. Clipp, S. Lazebnik, and M. Pollefeys. Building rome on a cloudless day. In *ECCV*, 2010.

[3] Y. Furukawa, B. Curless, S. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *CVPR*, 2010.

[4] A. Hornung, B. Zeng, and L. Kobbelt. Image selection for improved multi-view stereo. In *CVPR*, 2008.

[5] X. Li, C. Wu, C. Zach, S. Lazebnik, and J-M. Frahm. Modeling and recognition of landmark image collections using iconic scene graphs. In *ECCV*, 2008.

[6] M. Pavan and M. Pelillo. Dominant sets and pairwise clustering. *PAMI*, 29(1):167–172, 2007.

[7] N. Snavely, S. Seitz, and R. Szeliski. Skeletal graphs for efficient structure from motion. In *CVPR*, 2008.