View Clustering of Wide-Baseline N-Views for Photo Tourism

Aveek Shankar Brahmachari
Computer Science and Engineering
University of South Florida
Tampa, Florida, USA
Email: abrahmac@mail.usf.edu

Sudeep Sarkar
Computer Science and Engineering
University of South Florida
Tampa, Florida, USA
Email: sarkar@cse.usf.edu

***Abstract*—The problem of view clustering is concerned with finding connected sets of overlapping views in a collection of photographs. The view clusters can be used to organize a photo collection, traverse through a collection, or for 3D structure estimation. For large datasets, geometric matching of all image pairs via pose estimation to decide on content overlap is not viable. The problem becomes even more acute if the views in the collection are separated by wide baselines, i.e. we do not have a dense view sampling of the 3D scene that leads to increase in computational cost of epipolar geometry estimation and matching. We propose an efficient algorithm for clustering of such many weakly overlapping views, based on opportunistic use of epipolar geometry estimation for only a limited number of image pairs. We cast the problem of view clustering as finding a tree structure graph over the views, whose weighted links denote likelihood of view overlap. The optimization is done in an iterative fashion starting from an minimum spanning tree based on photometric distances between image pairs. At each iteration step, we rule out edges with low confidence of overlap between the respective views, based on epipolar geometry estimates. The minimum spanning tree is recomputed and the process is repeated until there is no further change in the link structure. We show results on the images in the 2010 Nokia Grand Challenge Dataset that contains images with low overlap with each other.**

***Keywords*-Computer Vision, Image Collection, Epipolar Geometry, Photo Organization**

I. Introduction

Community photos on web have been the subject of computer vision research lately. These datasets usually contain images of a monument or geographical region sometimes on city scale or even at world scale. The focus of current vision algorithms has been to use densely sampled scene photo collections, with high overlap with each other, for large scale reconstruction, camera pose estimation, and 2D panorama stitching [1], [2], [3], [4], [5], [6], [7], [8], [9]. Works on 2D panoramic stitching assume the scene to be far from the camera, or to have essentially negligible translations between camera centers. While multi-view 3D reconstruction algorithms assume that we have more than two views of the same scene content for reconstruction via bundle-adjustment. Bundle adjustment is a well known final refinement step in almost all shape-from-motion problems and can work with two views, but in practice it works the best if each scene point is seen by more than two cameras. For instance, Snavely [2] in his work on photo-tourism uses tracks of more than 20 keypoints across multiple images that are consistent with pairwise epipolar geometries between consecutive views in the track. These kind of algorithms typically exploit the high overlap in scene content between closely spaced views and can have problems when the images are widely spaced in 3D space, i.e. camera positions are widely separated. In such collections it is rare to have more than two views of the same scene content. In this work, we address the problem of finding connected clusters of views in a photo collection of weakly connected photos. We exclusively focused on the Nokia Grand Challenge Dataset [10] (see Fig. 1 for samples). No result on this dataset has been reported so far, although the dataset has been open for research for more than 2 years now. This dataset is relatively much harder than those already handled in the state-of-the-art vision research. It covers very large geographical area as much as covered by much larger datasets in the state-of-the-art research works making it an even harder problem for vision based research. To locate connected cluster of views, we need to compute similarity between two views. This can be done in two ways: photometrically and geometrically. Photometric distance between images can be computed based on appearance of image features, without considering geometric consistency. Geometric similarity can be computed by using the epipolar geometry between two views to constrain the matches between image features. Most current approaches to epipolar geometry estimation, especially for widely varying view points, typically rely on some variation of random sampling and hence tend to be computationally expensive. Once the clusters have been identified, they can be used either for photo tourism or as input to bundle adjustment for scene reconstruction.

Given the sparse nature of the views in the dataset, we structure our search for connected sets of views as a search for tree structures connecting the views. The nodes are the views and the links denote view overlap capable of estimating pose with some minimum confidence. We are looking for minimum number of such tree structures over the views. Ideally we want a single spanning tree but it might not be always possible. Snavely [1] also proposed similar skeletal graphs for efficient structure from motion but for photo collection with significant overlap, unlike our case. In another work by Schaffalitzky [11] results of graph based connectivity in images have been shown

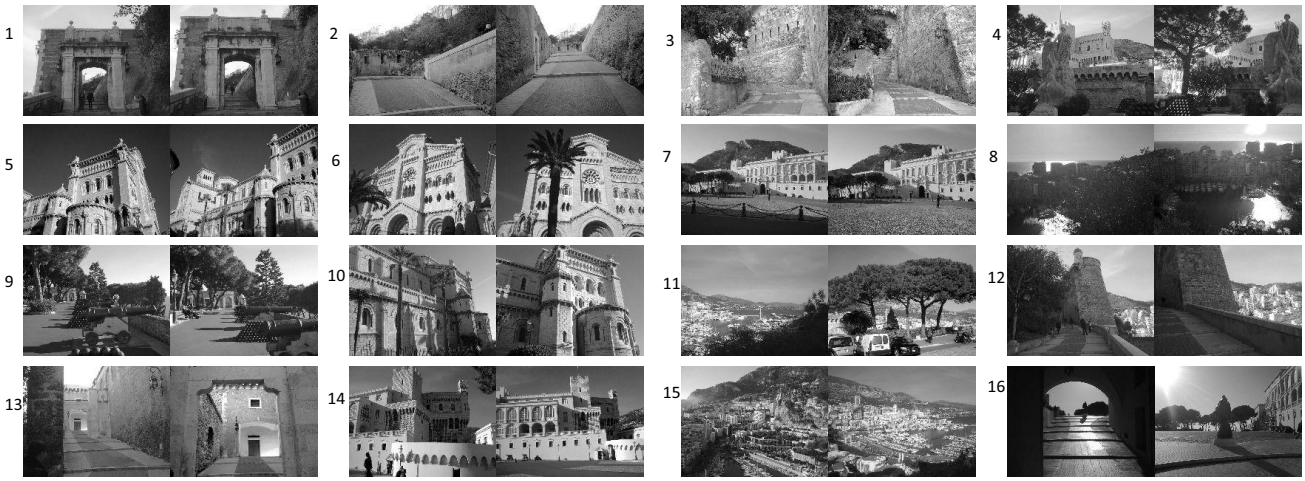

Fig. 1. Image pairs from the Demoset in Nokia Grand Challenge Dataset used to demonstrate the accuracy of our photometric distance measure. Notice the significant viewpoint changes for some pairs.

on small datasets, but again for densely spaced images. Heath *et al.* [12] also built a graph structure over a large collection of images, but the goal was different – linking was sought between image subregions across images, and not matching of entire images. Li *et al.* [13] has a similar philosophy to organizing large image datasets as ours. Like them, we also view the task of finding view clusters as a precursor to other possible tasks such as 3D reconstruction or browsing through collections. 3D reconstruction might not be needed for many tasks. Li *et al.* [13] has a similar approach as ours, however, the problems considered are of different flavors. They start with a large collection of images that are then grouped into clusters. Iconic images that represent the clusters are then selected. Our starting point is more like these iconic images. The views we consider also have large viewpoint and appearance changes. Their approach requires a training step to restrict costly geometric matching only to visually similar subset of iconic images, ours do not require any such training.

In our work, we use both the photometric and geometric similarities to decide on image similarities. Just photometric distance is not sufficient in identifying pairs of images that have common scene content. Two different scenes can have low photometric distance. For instance, see Fig. 2 that shows image pairs that were judged to have low-photometric distances but clearly not do not have overlap. And, it does not make sense to compute geometric consistency measures between pairs of images that do not have common scene content. The pure geometric approach is actually as accurate as we can get as it uses the strongest constraint (the epipolar constraint) on two-view geometry known in vision. The photometric method uses weak constraints like appearance and uniqueness. However, exhaustive geometry estimation is very time consuming while exhaustive photometry estimation is faster, but less accurate. We attempt to get the benefits of both the methods by photometric estimation based initialization and geometry based verification to perform an iterative search for the optimally connected tree structure. We start with a minimum spanning tree based on photometric distances between all pairs of views. Tree edges are tested for geometric consistency and are removed if found to be inconsistent. Some views with low computed photometric distances might not be geometrically consistent. The minimal spanning tree(s) are computed again on the modified graph, which might have multiple connected components. The process is repeated. The process stops when there is no further change in the spanning tree.

In this paper, we compared our algorithm with the exhaustive geometric method that allows us to measure our relative performance in terms of speed gain and accuracy loss, independent of the coding platform. However, other known algorithms have shown scalability of their algorithm without reporting similar metrics as we do. While exhaustive geometric algorithm is $O(n^2)$ in terms of geometric estimations, our algorithm needs $O(n)$ geometric estimates. This answers the scalability issue theoretically in our case.

II. Objective and Search Mechanism

The goal is to construct a tree structure, τ, connecting the camera nodes, where the links between two nodes denote overlap between the corresponding views. We want the tree structure with the largest total overlap between image pairs. Camera views are not calibrated nor do we use any GPS information. This is unlike [14] which uses GPS information for geo-clustering. We will use the concept of photometric distance constrained by epipolar geometry estimates to search for the optimal tree structure.

A. Photometric Distance

Photometric distance between two images I_i and I_j is defined based on the similarity between the sets of the point features found in these images. We used the SIFT [15] features in

Fig. 2. Few image pairs corresponding to the geometrically rejected edges. These pairs were photometrically similar based on local similarity in appearance around the putative correspondence but their geometric consistences were low.

this work, but other features could also work. Let the features found in I_i and I_j be denoted by the sets $\{u_1, \cdots, u_M\}$ and $\{v_1, \cdots, v_N\}$, respectively. A correspondence (u_i, v_j) will be denoted by x_k. We take the similarity of a correspondence to be the reciprocal of the distance between SIFT features. We have an M by N similarity matrix, ρ, computed for all possible correspondences.

The accepted putative correspondences, $\{x_1, \cdots, x_n\}$, are those that have highest similarity along both row and column in the photometric feature similarity matrix, ρ. For these putative correspondences, one feature of the pair is the best match for the other and vice-versa. Let the similarity of the k-th such putative match be denoted by ρ_k. And, ρ_{kr} and ρ_{kc} be the second highest values in the row and column of this match, respectively. The confidence in a putative correspondences could be related to how different these second maximums are from the similarity of the putative correspondences, which is the maximum along the corresponding row and the column. The more the difference, more the confidence in the match being correct. We use the following combination to result in a normalized similarity for the k-th putative match.

$$w_{ij}(k) = (1 - \exp^{-\rho_k})^2 (1 - \frac{\rho_k}{\rho_{kr}})(1 - \frac{\rho_{kc}}{\rho_k}) \qquad (1)$$

The combination results in high values for putative matches with high similarities *and* those for which the next best matches have low similarities. This photometric weight is same as that used in [16]. Using these weights over the K putative matches, we define the photometric distance between the two images.

$$\mathbf{G}(i,j) = \exp\left(-\kappa \log K \frac{\sum_{k=i}^{K} w_{ij}(k)}{K}\right) \qquad (2)$$

where the photometric distance between two images are exponentially related to the average similarity over the putative matches, $\frac{\sum_{k=i}^{K} w_{ij}(k)}{K}$, and the number of putative matches, the $\log K$, term. This distance decreases with the increase in number of putative correspondences and their average similarity. The constant term κ is fixed based on training data. In the experiments, we present some results with alternative combination forms. The above form resulted in the best performance.

B. Vision Based Geometric Consistency

Given a set of putative correspondences between two views, identified as described in the previous section, consistency of these correspondences with respect to the epipolar geometry can be used as a geometric measure. There are many such epipolar geometry estimation algorithms [17], [18], [16], [19]. Some require high overlap between views and others can work with widely varying views, with little overlap [16]. We use the latter since many view pairs in our problem could be widely separated.

Given an estimate of the epipolar geometry in terms of the fundamental matrix ($\mathbf{F}$), we take its fits to the putative correspondence set $\mathbf{X}$ as the geometric consistency measure between the two images. Let $\mathbf{u_k}$ and $\mathbf{v_k}$ be the homogeneous coordinates of the k-th putative correspondence. We denote by $\delta_{ij}(\mathbf{F}, \mathbf{x_k})$ the Sampson's distance of the k-th putative correspondence, which is an excellent approximation of the re-projection geometric error – the gold standard [20].

$$\delta_{ij}(\mathbf{F}, \mathbf{x_k}) = \frac{(\mathbf{v_k}^T \mathbf{F} \mathbf{u_k})^2}{||\mathbf{F}\mathbf{u_k}||^2 + ||\mathbf{F}^T\mathbf{v_k}||^2} \tag{3}$$

These individual errors of the putative correspondences need to combined into one overall error measure. However, instead of simply summing them, we consider an robust combination form that allows us to weigh down outliers. We choose Welsh's [21] weight function as our robust combination kernel to arrive at the overall geometric consistency between images I_i and I_j.

$$\gamma_{ij} = \sum_{k=1}^{K} \exp\left(-\frac{\delta_{ij}(\mathbf{F}, \mathbf{x_k})}{\sigma}\right) \tag{4}$$

The negative exponential in the Welsh's weight function suppresses the effect of outliers on the evaluation of the quality of the fundamental matrix. Correspondences with very large Sampson errors will not contribute to the overall geometric consistency between the two images. The sum of such exponentials gives an M-estimate of the fundamental matrix fitting quality. In our experiments, we have fixed $\sigma = 10^{-4}$. It must be noted here that we do not multiply probability measures obtained from residual errors as done by others [22] because a multiplicative cost function would primarily be determined by the low probabilities associated with the outliers. Even one outlier would weigh down the multiplicative form value. Additive cost function would instead allow for suppressed fitting values of outlier correspondences without getting effected by them.

C. Algorithm

The algorithm to search for the tree structure is based on a greedy approach that uses the expensive geometric consistency computation in an opportunistic manner. It starts from a spanning tree of a complete graph structure, weighted by photometric similarities, and gradually refines it based on geometric consistency. The specific steps of the algorithm are shown in Algorithm 1.

Note that the final tree could be disconnected. Geometric consistency is computed only for the pairs of images with low photometric distance, i.e. high similarity. In practice the total number of iterations needed are small around 10 to 15 for a dataset of around 400 images. So, the number of times geometric consistency is computed is determined by the number of links in the spanning tree, which is proportional to the total number of images in the dataset.

Although, theoretically our work does not depend on a specific kind of feature, but we relied on SIFT point features in our work. This might be a possible limitation of our work.

Algorithm 1 CLUSTER-VIEWS ($I_1, \cdots, I_N$)

Require: : N images

Ensure: : $N \geq 2$

Preprocessing:

1) Downsample images to about 200 by 200 pixels
2) Compute starting graph $\mathbf{G}(i, j)$ based on photometric distances (Eq. 2)
3) τ^0 = Minimum spanning tree of $\mathbf{G}$, $n = 0$

Iterative Optimization:

4) $\forall (i, j) \in \tau^n$ estimate geometric consistency using γ_{ij} (Eq. 4)
5) $\forall (i, j) \in \tau^n$ with $\gamma_{ij} < t_\gamma$ set $\mathbf{G}(i, j) = \infty$, i.e. remove the edge.
6) τ^{n+1} = Minimum spanning tree(s) of $\mathbf{G}$. Note $\mathbf{G}$ might get disconnected.
7) If $(\tau^{n+1} \neq \tau^n)$ then $n = n + 1$ and goto step 4.

Also, the upper limit of the accuracy of our algorithm (or any possible vision based algorithm for this problem) is defined by the exhaustive geometry estimation.

D. Convergence of our optimization strategy

First of all, it is worth noticing in our work that our photometric initialization is very good (see Fig. 6) and so we do not need many iterative geometric optimization steps. However, the quality of the results would improve with the number of consecutive iterations we would continue to wait for a change even after no change in the MST (Minimum Spanning Tree) occurs for those iterations. If we continue till all edges are exhausted, this would clearly converge to the result of the exhaustive geometric approach but would not save time.

E. Time Complexity

The time-complexity is measured in terms of the number of time most expensive sub-step involved in the algorithm is performed for a given input size. The most expensive sub-step in our algorithm is epipolar geometry estimation. The number of epipolar geometry estimations done in our case is $O(n)$ in the average case, since the estimations are done only along the spanning tree and is initially restricted using the photometric similarity matrix. We can ignore matches below a threshold on photometric similarity leaving a sparse photometric similarity matrix.

III. Experiments

The Nokia Grand Challenge dataset consists of widely spaced images over a large geographic area. It also has GPS tags with each image, which we use only to visualize the results. GPS information is not used to compute the view clusters. The dataset has two sub-datasets - the Demoset and the Lausanne set. The Demoset has 105 images, which we use for training. And, the Lausanne dataset has 243 images on which we show the clustering results.

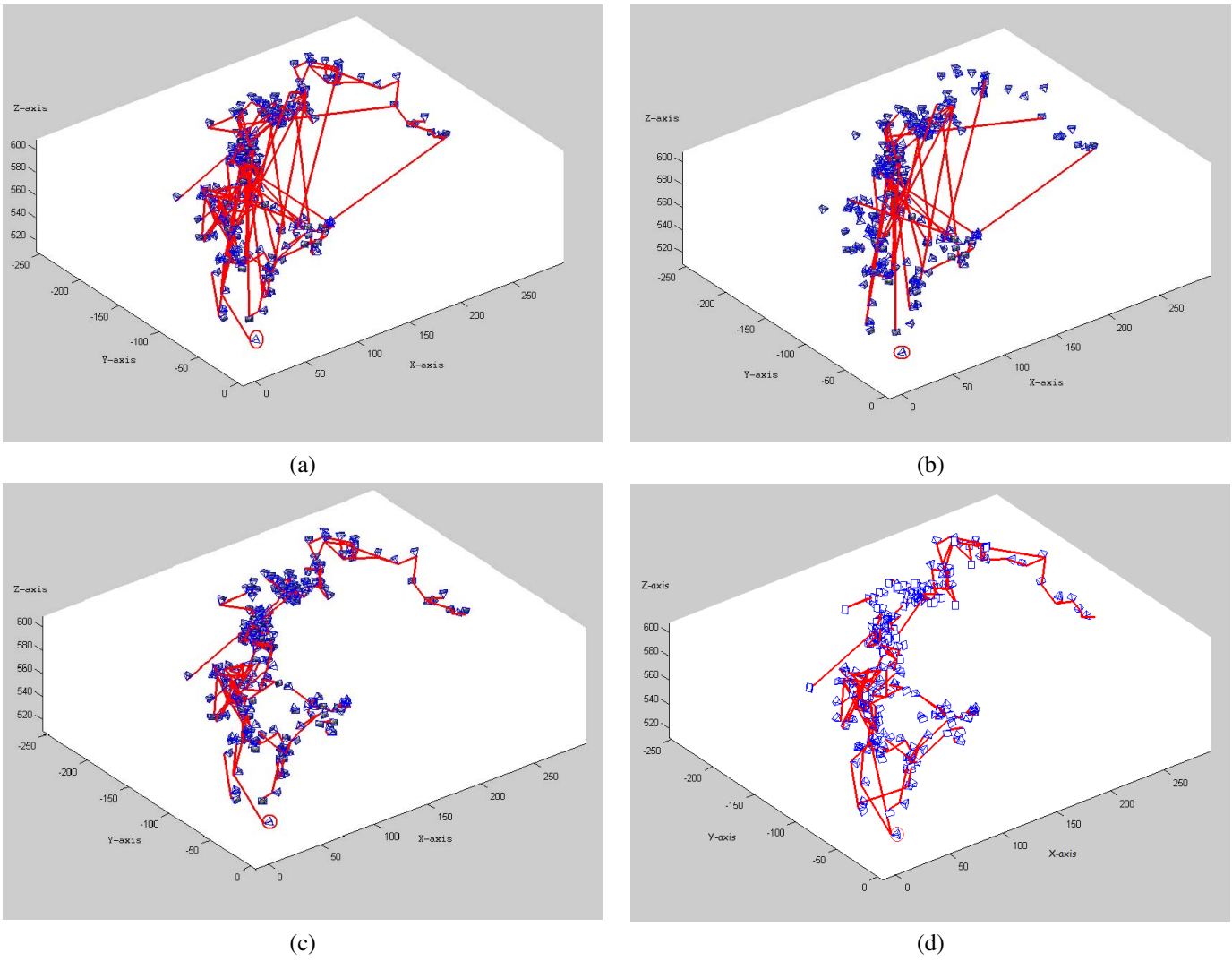

Fig. 3. Tree computation for the Nokia Challenge Dataset. Each image is represented by a point at the corresponding GPS tagged locations, in local ENU (East North Up) coordinates, with negative Y-axis as north and positive X-axis as east. The reference camera at origin is encircled. We use the GPS tags just for visualization of the results. They are not used in the computation of the tree. (a) Initial photometry based minimum spanning tree. (b) Geometrically rejected edges from initial photometric MST. (c) Final tree structure with our approach. (d) Final tree structure with a purely geometric approach – the standard to compare against.

A. *Training*

We used as training set 16 image pairs from the Demoset (shown in Fig. 1) and 4 image pairs from images used in [16] for diversity. This training data was used to train the value of κ in Eq. 2 and the geometric consistency threshold t_γ used in step 5 of the Algorithm 1. We also used this training dataset to experiment with other forms of the photometric distance functions shown below.

$$\mathbf{G}_1(i,j) = \exp\left(-\frac{\sum_{k=i}^{K} w_{ij}(k)}{K}\right) \quad (5)$$

$$\mathbf{G}_2(i,j) = \exp\left(-\log K\right) \quad (6)$$

$$\mathbf{G}_3(i,j) = \exp\left(-\log K \frac{\sum_{k=i}^{K} w_{ij}(k)}{K}\right) \quad (7)$$

$$\mathbf{G}_4(i,j) = \exp\left(-3\log K \frac{\sum_{k=i}^{K} w_{ij}(k)}{K}\right) \quad (8)$$

We used each of functions to select the inliers from a set of putative correspondence and compared the inlier rate with the ground truth inlier rate, which was computed manually for the training image set. The inlier rate is the fraction of the putative correspondences that are deemed to be correct correspondences based on the chosen distance function. In Fig. 4, we show the correlation of the computed inlier rates using the four forms with the ground truth inlier rate. We see that the fourth form is the best correlated one with the ground truth rates. It lies closest to the diagonal. This is the form we pick for our experiments.

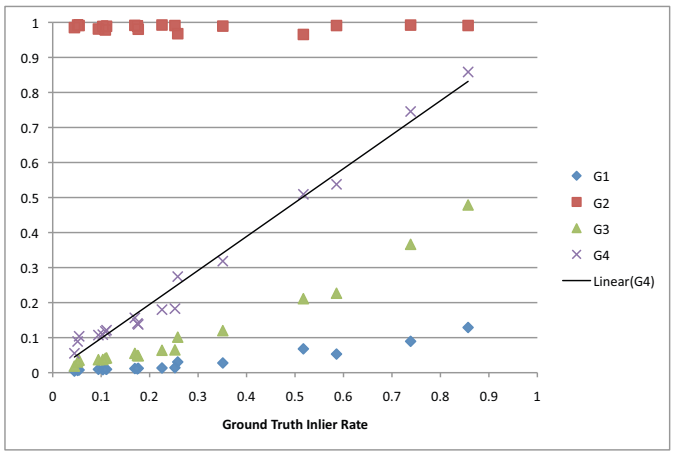

Fig. 4. Correlation between the inlier rate computed by each of the four photometric distance functions, $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$, and $\mathbf{G}_4$, with the ground truth inlier rate on a set of 20 training images. The fourth similarity function, $\mathbf{G}_4$, is most correlated with actual rates and is the one we use in our research.

B. Test Results

In the Lausanne dataset, we had 242 edges in the initial MST. Many of these initial edges based on photometric similarities were wrong as they connected views that are widely separated. Out of 242 initial edges, 45 were rejected by initial pose estimate verification, leaving us with 197 well connected edges. In the next iteration, 6 edges with good geometric consistency were added and so on until we do no better in subsequent iterations. Fig. 6 shows this progress in the number of edges in the tree with iteration.

In order to give an idea of the timing benefit we achieved from our algorithm, we show a wall-clock timing comparison of our method compared to the exhaustive geometric method in Table I. The actual running time of our algorithm executed in MATLAB for the 'Lausanne' dataset of 243 images on a single machine with Intel Dual Core processor @1.80 GHz for the exhaustive method and our algorithm are shown in Table I. We achieved 45 times speed-up compared to the exhaustive geometric case. Our algorithm takes around 15 minutes to generate the photometric similarity matrix and then takes another 15 minutes for geometric verification and updation of the similarity matrix for the 'Lausanne' data in the Nokia Grand Challenge dataset.

TABLE I

COMPARISON OF THE TIME TAKEN BY THE EXHAUSTIVE GEOMETRY VERIFICATION METHOD AND OUR ALGORITHM. THE TIME TAKEN BY PHOTOMETRIC PRE-PROCESSING AND GEOMETRIC PROCESSING HAS ALSO BEEN SHOWN FOR OUR ALGORITHM

Vs	Hours	Minutes	Seconds
Exhaustive Geometric Case	21	38	25
Our Algorithm			
a. Photometric Preprocessing		14	42
b. Geometric Processing		14	12
c. Total		28	54

In Fig. 3(a), we visualize the initial MST τ^0 using East North Up (ENU) coordinates present in the associated GPS tags. We use the GPS tags just for visualization of the results. They were not used in the computation of the tree. In the figure, East is along X-axis, North is along the negative Y-axis, and Z-axis points upwards. The code for this display was available as a toolkit in Nokia Grand Challenge website [10], but was modified to display a spanning tree through the camera nodes. In Fig. 3 (b), we show the edges that were rejected by the geometric consistency measure. In Fig. 2, we show 6 examples of such image pairs. These were photometrically similar but were rejected due to low geometric confidence. We had 213 edges in the final tree, whose edges are shown in Fig. 3(c). The tree appears connected, but on careful observation it can be noticed that it is disconnected. Our results are consistent with the GPS information in the images. Views that are close in the GPS are also connected in the final tree structure. This shows that our algorithm performs well in clustering the views.

As comparison, in Fig. 3(d) we show the tree structure based on using purely geometric consistency measures instead of photometric distances. The tree structures in Fig. 3(c) and (d) are similar. Purely geometric method resulted in 219 edges in the final tree, which is just marginally more than 213 edges that we got using our fast method.

In Fig. 5, we show an example of our tree cluster of images. There are several aspects that are worth pointing out. First, notice how widely different viewpoints of the same scene structures are associated. For instance, on first glance it is not obvious why 1 and 3 linked, however, on closer scrutiny we see that the building in 1 is actually seen in the middle of image 3, viewed at a distance. The geometric consistency measure is able to account for such drastic scale changes. Similarly, images 3 and 4 both have the right building in image 3 as common content. We also see the significant rotation between images 1 and 38, 14 an 16, and 16 and 17. Images 38, 39, 40, 41, and 42 all are views of the building in 1 viewed from different locations and angles. There are other instance of significant viewpoint change, e.g. notice 26 and 27, and also 35 and 20, both have only small portion of the common scene.

Second, depth first traversal through this view clustering tree do result in meaningful path through the scene. For instance, consider the path 1, 3, 4, 5, 10, 13, 14, 16, 17, 18. It represents path around the circular structure with step. Whereas the branch 1, 3, 4, 5, 6, 7, 8, and 9, is a path that diverges away from this structure. The path 1, 3, 4, 19, 20, 21, 22, 25, 26, 27, 28, 29, 30, 31, 32, 33 take you through a different route. We are able to recover these meaningful paths through the dataset without the use of any GPS information.

Third, notice that we are able establish matches even in the presence of some amount of scene content motion between views. For instance, between 20 and 21, or between 29 and 34.

Fig. 5. One of the clusters formed as an output of our algorithm shown as a tree.

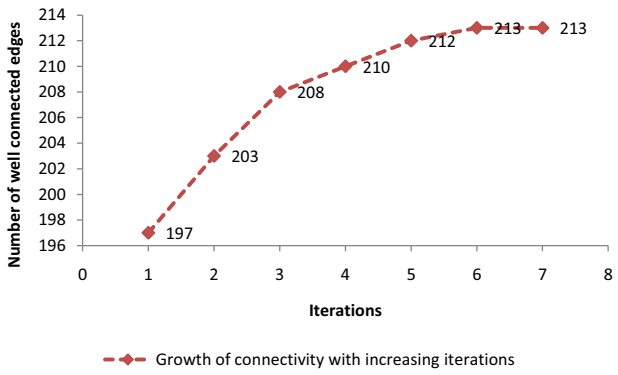

Fig. 6. Figure showing the number of good edges in the tree structure after every iteration till convergence

IV. Conclusion

In this paper, we have considered the problem of establishing vision-based connectivity between widely spaced views with low overlaps. We found the Nokia Grand Challenge dataset suitable for this research. This dataset is much difficult in terms of 3D spacing and view-overlapping than other commonly used datasets in large scale view organization research. We model the vision-based connectivity between images as a search for a tree structure over the images and define the best tree as the one with maximum edges connecting image pairs that are photometrically similar and geometrically consistent. In this paper, there are two main contributions. One is the photometric distance measure and another is the iterative optimization technique that we have used. Both these strategies are fairly simple, yet powerful, and potentially scalable to very large collections. The initial MST found using the photometric distance measure is a good initialization to the iterative optimization as shown in the paper. We found that most of the edges were correctly initialized and there were reasonable number of edges replaced during iterative optimization as well. Both of these methods together have shown good results on dataset with widely spaced images. The computed tree structure matched well with the GPS information in the images. GPS information was not used in the algorithm. It was only used for visualization. For the 'Lausanne' data in the Nokia dataset, we achieved a speed up of about 45 with respect to the exhaustive algorithm. Our algorithm took less than 30 minutes to generate view clusters for the 'Lausanne' dataset, while the exhaustive algorithm takes about 22 hours. The computed visual clusters of images could be used for photo tourism to organize and to navigate through a large collection of photos. It could also be used to refine magnetometer and GPS information in the images [23]. 3D scene reconstruction by bundle adjustment would benefit for the identification of image subsets with common scene content, as captured in the view clusters.

References

[1] N. Snavely, S. Seitz, and R. Szeliski, "Skeletal graphs for efficient structure from motion," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[2] N. Snavely, S. Seitz, and R. Szeliski, "Photo tourism: Exploring image collections in 3D," pp. 835–846, 2006.

[3] N. Snavely, *Bundler: Structure from Motion for Unordered Image Collections*. Online, May 2009.

[4] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. I: 519–528.

[5] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. Seitz, "Multi-view stereo for community photo collections," in *International Conference on Computer Vision*, 2007, pp. 1–8.

[6] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1362–1376, August 2010.

[7] Y. Furukawa, B. Curless, S. Seitz, and R. Szeliski, "Towards internet-scale multi-view stereo," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1434–1441.

[8] S. Agarwal, N. Snavely, I. Simon, S. Seitz, and R. Szeliski, "Building rome in a day," in *International Conference on Computer Vision*, 2009, pp. 72–79.

[9] M. Brown and D. Lowe, "Recognishing panoramas," in *International Conference on Computer Vision*, 2003, pp. 1218–1225.

[10] "Nokia Challenge 2010: Where was this Photo Taken, and How?" http://comminfo.rutgers.edu/conferences/mmchallenge/2010/02/10/nokia-challenge/.

[11] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets, or 'how do i organize my holiday snaps?'," in *European Conference on Computer Vision*, 2002, p. I: 414 ff.

[12] K. Heath, N. Gelfand, M. Ovsjanikov, M. Aanjaneya, and L. Guibas, "Image webs: Computing and exploiting connectivity in image collections," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, june 2010, pp. 3432 –3439.

[13] X. Li, C. Wu, C. Zach, S. Lazebnik, and J. Frahm, "Modeling and recognition of landmark image collections using iconic scene graphs," in *European Conference on Computer Vision*, vol. 8, 2008.

[14] Y.-T. Zheng, M. Zhao, Y. Song, H. Adam, U. Buddemeier, A. Bissacco, F. Brucher, T.-S. Chua, and H. Neven, "Tour the world: Building a web-scale landmark recognition engine," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 0, pp. 1085–1092, 2009.

[15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[16] A. Brahmachari and S. Sarkar, "BLOGS: Balanced local and global search for non-degenerate two view epipolar geometry," in *International Conference on Computer Vision*, 2009, pp. 1685–1692.

[17] M. Fishler and R. Boles., "RANdom SAmple Concensus: A paradigm for model fitting with applications to image analysis and automated pages cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[18] H. Chen and P. Meer, "Robust regression with projection based M-estimators," in *International Conference on Computer Vision*, 2003, pp. 878–885.

[19] P. Torr, "Bayesian model estimation and selection for epipolar geometry and generic manifold fitting," *International Journal of Computer Vision*, vol. 50, no. 1, pp. 35–61, 2002.

[20] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, NY, USA, 2000.

[21] A. Welsh, "On M-processes and M-estimation," *Annal of Statistics*, vol. 17, no. 1, pp. 337–361, 1989.

[22] P. Torr and D. Murray, "The development and comparison of robust methods for estimating the fundamental matrix," *International Journal of Computer Vision*, vol. 24, no. 3, pp. 271–300, 1997.

[23] Anonymous, "Fast detection of noisy gps and magnetometer tags in fast detection of noisy gps and magnetometer tags in wide-baseline multi-views," in *Submitted to ACM Multimedia Conference, 2011*.