FULLY AUTOMATED IMAGE ORIENTATION IN THE ABSENCE OF TARGETS

C. Stamatopoulos [a, *], T.Y. Chuang [b], C.S. Fraser [a], Y.Y. Lu [a]

[a] Department of Infrastructure Engineering, University of Melbourne, Victoria 3010, Australia
[b] Department of Civil Engineering, National Taiwan University, 1, Roosevelt Rd., Sec. 4, Taipei 10617, Taiwan
c.stamatopoulos@unimelb.edu.au, d95521008@ntu.edu.tw, c.fraser@unimelb.edu.au, yylu@student.unimelb.edu.au

Commission V, WG V/4

KEY WORDS: Feature-based Matching, Target-less Orientation, Relative Orientation, Image-based Modelling

ABSTRACT:

Automated close-range photogrammetric network orientation has traditionally been associated with the use of coded targets in the object space to allow for an initial relative orientation (RO) and subsequent spatial resection of the images. Over the past decade, automated orientation via feature-based matching (FBM) techniques has attracted renewed research attention in both the photogrammetry and computer vision (CV) communities. This is largely due to advances made towards the goal of automated relative orientation of multi-image networks covering untargetted (markerless) objects. There are now a number of CV-based algorithms, with accompanying open-source software, that can achieve multi-image orientation within narrow-baseline networks. From a photogrammetric standpoint, the results are typically disappointing as the metric integrity of the resulting models is generally poor, or even unknown, while the number of outliers within the image matching and triangulation is large, and generally too large to allow relative orientation (RO) via the commonly used coplanarity equations. On the other hand, there are few examples within the photogrammetric research field of automated markerless camera calibration to metric tolerances, and these too are restricted to narrow-baseline, low-convergence imaging geometry. The objective addressed in this paper is markerless automatic multi-image orientation, maintaining metric integrity, within networks that incorporate wide-baseline imagery. By wide-baseline we imply convergent multi-image configurations with convergence angles of up to around 90°. An associated aim is provision of a fast, fully automated process, which can be performed without user intervention. For this purpose, various algorithms require optimisation to allow parallel processing utilising multiple PC cores and graphics processing units (GPUs).

1. INTRODUCTION

Automated network orientation via coded targets commonly leads to a precise calculation of the object space coordinates due to the integrity and highly accuracy of the image point correspondence determination. In addition, and contrary to the currently developed feature-based matching orientation techniques, there are no factors that limit the development of a highly convergent and consequently geometrically strong photogrammetric network. It has been shown by Jazayeri (2010) and Barazzetti (2011a) that the matching accuracy achieved by feature-based matching (FBM) procedures is approximately 0.3 pixels. Additionally, Remondino et al. (2006) has shown that the image coordinates of the homologous points extracted using FBM can be further improved through use of least-squares matching (LSM) techniques. An increase to approximately 0.25 pixels (Barazzetti, 2011a) can be anticipated in such cases, but this will not have a significant impact in the resulting accuracy of the photogrammetric network.

Considering that the accuracy of the current state of the art FBM algorithms is unlikely to be further improved, the focus should instead be placed on the design and optimisation of the network geometry. Thus, the topic of this paper concerns the development of algorithms that will allow the successful orientation of highly convergent multi-image networks. The approach developed starts with the familiar feature extraction stage via algorithms such as SIFT (Lowe, 1999) and SURF (Bay et al., 2008) to provide possible common points among images. This produces both valid matching points and a large percentage of outliers. The prime objective then is to develop a photogrammetric methodology to effectively filter out these 2D image point outliers within image pairs, such that the remaining valid matches can be used to perform a robust multi-image RO to sub-pixel accuracy. A secondary objective is to try and make this procedure more efficient in order for it to be performed in a timely manner. For this purpose, selected algorithms are implemented to run in parallel either in the CPU or GPU. General-Purpose computation on Graphics Processing Units (GPGPU) has helped to solve various computationally intensive problems due to the high-performance of the GPUs, which are comprised of many core processors capable of very high computation and data throughput.

2. POINT CORRESPONDENCE CALCULATION

The proposed methodology, shown in Figure 1, starts with feature point extraction and description through the use of an algorithm such as SIFT or SURF. It should be noted that image detectors and descriptors do not provide matches but their purpose is limited to the detection and description of feature points in one image. An additional step is required to find point correspondences for two or more images. Feature descriptors calculate a vector of 64 or 128 dimensions to describe an interest point. The descriptor vector can be utilised to match points in different images. For an interest point that is visible in a pair of images, the feature descriptor algorithm is expected to return almost identical descriptor vectors. Typically, the Euclidean length of the descriptor vector is used to quantify such a relationship. Thus, the process of finding point correspondences involves the calculation and comparison of all

* Corresponding Author

the descriptor lengths in order to find the best possible matches. However, since in many cases images contain repetitive patterns or features, the descriptor vectors are not unique. As a result, this procedure produces both valid matching points and a large percentage of outliers, as shown in Figure 2.

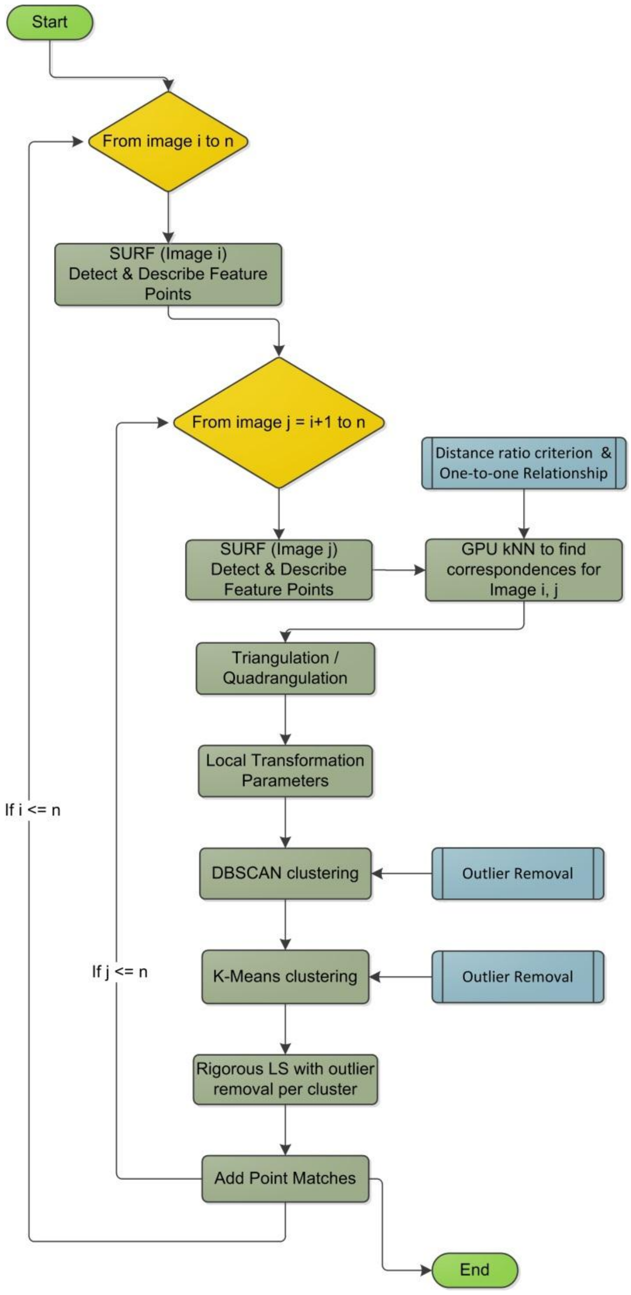
Start
From image i to n
SURF (Image i) Detect & Describe Feature Points
From image j = i+1 to n
Distance ratio criterion & One-to-one Relationship
SURF (Image j) Detect & Describe Feature Points
GPU kNN to find correspondences for Image i, j
Triangulation / Quadrangulation
Local Transformation Parameters
If i <= n
DBSCAN clustering
Outlier Removal
If j <= n
K-Means clustering
Outlier Removal
Rigorous LS with outlier removal per cluster
Add Point Matches
End

Figure 1. Flowchart of the proposed matching methodology

Indisputably, the computational cost of finding potential homologous points can be very high and it is highly dependent on the number of interest points. For example, to find the best match of an interest point in another image having N interest points, N descriptor vector length calculations and searches are required. This problem is commonly known as nearest neighbour (NN) search and its computational cost is O(N) if all the distances among descriptor vectors have already been calculated. One solution is to limit the number of points so that the calculations can be performed in a timely manner. This can be optimally achieved by using a threshold in the interest point detector stage in order to obtain points with high affine invariance response only. While this is expected to provide sufficient matches for narrow baseline imagery due to the small angle variation and consequently high similarity of features, it can certainly pose a problem in highly convergent imagery where the resultant points in each image might not be common. Thus, in order to increase the number of possible matches for wide baseline imagery, it is imperative that no threshold is used for the interest point detection stage.

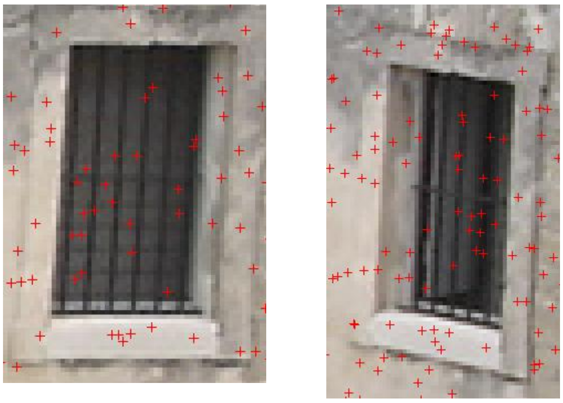

Figure 2. SURF matching results for an image pair

However, as already mentioned, an increased number of points will lead to an extensive number of calculations. Barazzetti (2011a) proposed the use of a *kd-tree* to organise the data and optimise the calculations of the NN search. Due to the way that the *kd-tree* stores data in memory, it is able to perform this search efficiently by using the tree properties to quickly eliminate large portions of the search space. The computational cost of a NN search after building the *kd-tree* is O(logN) compared to O(N) when a simple linear search is performed. In order to further optimise this procedure an implementation of the *kd-tree* that runs in parallel within the available CPU cores was developed. The *kd-tree* is a highly scalable algorithm and the availability of each extra CPU core can halve the calculation time. For example, in an 8-core CPU the computation of corresponding matches for a pair of images where the number of extracted SURF points exceeded 20 thousand per image was approximately 15s.

In the initial matching, a large number of outliers emerge, especially in situations where images contain similar patterns or features. This can preclude a successful RO. An additional advantage of the *kd-tree* is the ability to perform *k*-nearest neighbour searches (*k*NN). This allows the calculation of not only the best match for a point pair, but also potentially the *k* best matches. This information can be used in the point correspondence calculation stage to filter similar descriptor vectors. This can easily be achieved by setting a criterion so that the difference in the descriptor length of the first and second best match for each point is more than 60%, for example. It should be noted, however, that this does not constrain one point from being uniquely matched, and consequently an additional filtering process has to be performed so that a one-to-one relationship between the feature point matches is ensured. Figure 3, shows the SURF results that were previously presented in Figure 2 after the application of this filtering procedure.

While the *kd-tree* algorithm offers various advantages it can also suffer from what is known the curse of dimensionality. For a large number of points and especially with high dimensionality, in this case 64 or 128, the speed of the algorithm is only slightly better than a linear search of all of the points. For this reason investigations in the field of GPGPU were initiated. The computer science community has researched extensively on this problem but even though GPU *k*NN solutions have been presented, there are certain limitations associated with them, mainly related to the size of the datasets that can be processed.

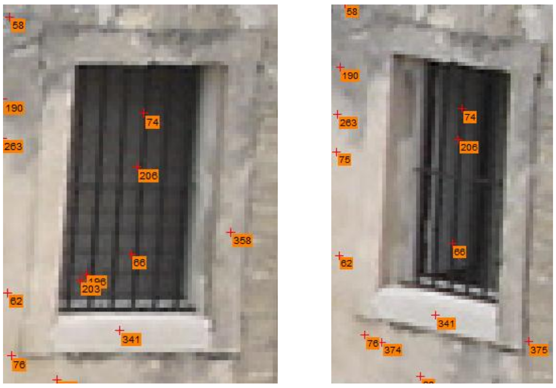

Figure 3. SURF matching results after distance threshold and one-to-one relationship restrictions applied.

The solution developed for this research is based on that proposed by Garcia et al. (2008), with the difference that various optimisations have been applied in order to address any shortcomings. The algorithm developed is written in the CUDA programming language and can easily be employed as part of any program developed in the C or C++ programming languages. CUDA has been actively developed by NVIDIA and as a consequence it is only available for NVIDIA GPUs. While alternative options exist, CUDA is certainly the only one that has achieved wide adoption and usage.

With the use of the GPU the calculation time of a *k*NN search is reduced dramatically. The required time for a pair of images containing approximately 300,000 points each, is two minutes when performed by the GPU. Such calculation times would be impossible to reach using the CPU, even when parallel *k*NN implementation is employed. Additionally, in order to provide a direct comparison to the parallel CPU *k*NN implementation, a search for a pair of images where the number of extracted SURF points exceeded 20,000 per image is a matter of a few hundred milliseconds when performed in the GPU. The above calculations were performed with an NVIDIA GPU using the available 192 cores. A significant performance gain could be anticipated when a state of the art GPU with 1536 cores is employed due to the parallel execution of the calculations.

3. POINT CORRESPONDENCE REFINEMENT

At this stage it is considered that the point correspondence has been calculated and that a one-to-one relationship has been established, as previously mentioned. It should be noted that even if a higher distance threshold between the first and second best match is used, it can never be ensured that outliers will be absent from the data and this is why an additional refinement step is required. For this additional stage of filtering, two similar quasi-RANSAC approaches are proposed, one based on the affine and the other on the projective model. This refinement procedure is premised on the assumption that point correspondences for pairs of images have been successfully established and consequently transformations such as the affine or projective can be used to transfer points from one image to the other. In the next paragraphs it will be explained how this process can be used to filter outliers. The process can be split into four distinct phases, as indicated in Figure 1. The first phase is different, depending on the transformation model used while the remaining phases are identical regardless of the transformation used.

While an affine transformation cannot model the pinhole camera model accurately at whole-image scale, it can adequately do so for smaller areas. For an optimal geometry and utilisation of all the currently matched points, a 2D Delaunay triangulation takes place. This triangulation need only be performed in one image of the pair. As the affine transformation requires a minimum of three homologous 2D points, the formed triangles are used to calculate all the local affine transformations.

The projective model can also account for image skew and can thus handle image transformations better than the affine model. However, the projective transformation has eight unknown parameters so four identical points are required for the calculation of each of the local transformations. In a similar manner to the triangulation process, a quadrangulation process takes place. In order to calculate the quadrangles, a 2D Delaunay triangulation is initially performed. Then, pairs of triangles that share a common line are merged to form a quadrilateral.

The second step of this methodology is the same for both transformation models and is based on data mining algorithms. Its purpose is to identify the clustering structure of the local transformation parameters and the consequently the outliers. Following a careful literature review, an algorithm known as DBSCAN (Ester et al., 1996) was selected for both its efficiency and speed. DBSCAN was proposed to cluster data based on the notion of density reachability. Basically, neighbouring points are merged in the same cluster as long as they are density reachable from at least one point of the cluster. Additional criteria such as minimum points per cluster can easily be assigned. As a result, transformations that are not part of any cluster are considered as outliers and removed. A variation of this algorithm, known as OPTICS (Ankerst et al., 1999) can also be used for this purpose. OPTICS aims to deal with DBSCAN's major weakness, the problem of detecting clusters of varying density.

The third phase of the filtering approach employs one of the simplest clustering algorithms, the k-means algorithm. The k-means algorithm clusters the filtered transformation parameters into k clusters so that each belongs to the cluster with the nearest mean. In this case the transformation parameters are clustered using the Euclidean distance as a metric. Figure 4, highlights the basis on which this phase is applied. Erroneous matches, marked in red in Figure 4, will certainly result in affine or projective parameters that are not similar to any of the correct local transformations. As a consequence they are not expected to be grouped with any other transformation parameters.

However, in some cases it occurred that a few erroneous transformations were grouped together due to neighbouring mismatches. This resulted in another criterion being added so that each group would contain a minimum of eight transformations. Consequently, clusters containing less than eight transformations were deemed to be outliers. It is important to note that the removal of outliers at any of the aforementioned phases does not necessarily disregard a feature point, since each point can take place in many transformations. Figure 4, for example, shows that the central feature point participates in a total of six affine transformations. The biggest problem of the k-means algorithm is that there is no notion of outliers and this is the main reason that the previous clustering phase was employed along with the additional criteria to ensure that the majority of outliers are removed.

In this research, a variation of the typical k-means was used where the number of groups returned, which is not necessarily k, is optimised depending on the input data. The advantage of this algorithmic variation is that a large value can be set as the number of the required clusters without worrying about data/cluster fragmentation, since the value can change dynamically. In addition, a different algorithm, known as k-means++ (Arthur et al., 2007), was used for the calculation of the initial values of the k-means clustering algorithm in order to avoid poor clustering results. This algorithm is also reported to offer speed and accuracy improvements compared to the traditional k-means. Parallel implementations of both k-means and k-means++ algorithms were developed in order to make the calculations even more efficient.

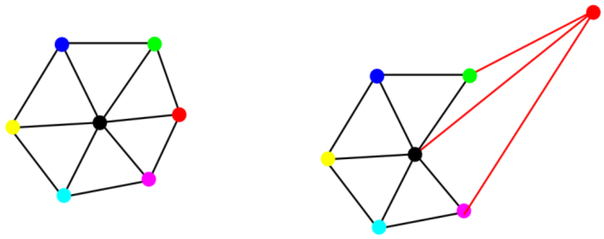

Figure 4. An erroneous 2D point and the resultant triangulation are displayed.

The fourth and final step of the filtering procedure involves a rigorous least squares (LS) adjustment with outlier detection. This is performed for each cluster. In this phase, the actual 2D points that comprised every transformation are used to calculate the affine or projective parameters for the whole group. A rejection threshold is set for the process in order to remove any points that have a high residual and are potentially outliers. This threshold is calculated dynamically during every iteration of the LS adjustment while the value of one pixel is set as the maximum. The dynamically calculated value is set as three times the RMS of the transformation. This ensures the integrity of the transformation for groups containing either good or bad matches. While this threshold might seem strict, experimental evaluation has shown that it performs very well, even when uncalibrated cameras are used. The reason for this lies in the clustering of the similar transformation parameters.

4. AUTOMATED NETWORK ORIENTATION

With the procedure described above, correct 2D correspondences among pairs of images can be found. By doing so, the orientation of an arbitrary number of images is reduced to the same process as orienting image networks that employ targets. The difference with this technique compared to previously reported approaches, for example Barazzetti (2011a), is that there is no need to orient the images in a sequential order, since 2D point matches among images are known a-priori. This circumvents the need for a sequential orientation of the photogrammetric network. Figure 5, shows a photogrammetric network comprised of 14 images, where a 7-ray point is highlighted.

An optional procedure can be implemented to strengthen the network after the automated orientation process is complete. Additionally, this procedure can be utilised for the addition of new 2D point correspondences to the network. In the point correspondence determination process for non-coded targets, which employs geometric constraints, candidate 3D points obtained via initial 2-ray intersection are back-projected to each image in order to find additional corresponding 2D observations in close proximity to the back-projected points. If an existing 3D point is found, the 2D measurement is added as a new observation. If no 3D point is found, a new 3D point is added to the network.

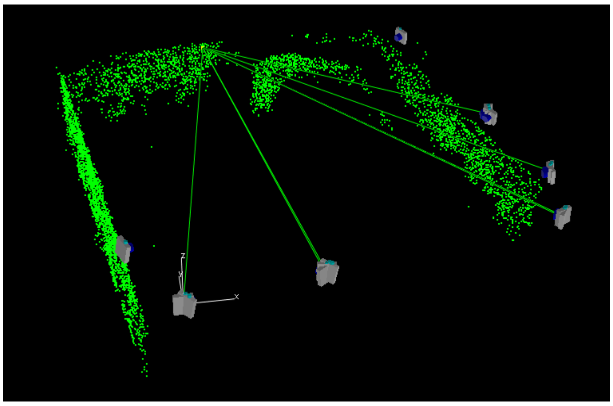

Figure 5. Automated targetless orientation of a photogrammetric network comprised of 14 images.

This procedure does not cause any problems for the structured scenes typically present in close-range photogrammetric applications. On the other hand, feature extraction algorithms provide numerous points so on many occasions the back projection of multiple 3D points can occur within the area covered by a single pixel due to the ability of sub-pixel matching. Thus, in order for the geometrically constrained point correspondence determination to be successful, highly accurate interior orientation (IO) parameters have to be known. Additionally, this procedure can be further optimised to work with uncalibrated cameras by using the feature descriptor information that is available.

Figure 6 shows the proposed algorithm where new 2D point correspondences are added to the network, or additional rays are added to existing 3D points sequentially. This procedure has also been implemented to run in parallel. The use of the feature descriptor allows the search to be limited to a specific number of points, making this procedure much more efficient and reliable for cases where the IO is not known. It is noteworthy that due to the nature of typical close-range photogrammetric measurements employing targets in the object space, a specific number of 2D points usually exist in an image. However, when feature extraction algorithms are employed a few hundred thousand 2D points can be anticipated. Searching for matches using a geometric constraint can be quite slow in such cases as all the 2D points have to be traversed.

5. EXPERIMENTAL EVALUATION

A number of experimental tests were conducted to evaluate the proposed methodology. Software was developed in C++ and CUDA and integrated into the iWitnessPRO (Photometrix, 2012) software package to allow for easier testing. The scope of the testing covered both image pairs, with convergence angles of up to 90^o, as well as existing multi-image photogrammetric data sets. During this testing procedure both the affine and the projective model were evaluated. As anticipated, the projective model performed better than the affine model, the latter often rejecting a number of correct point matches as well.

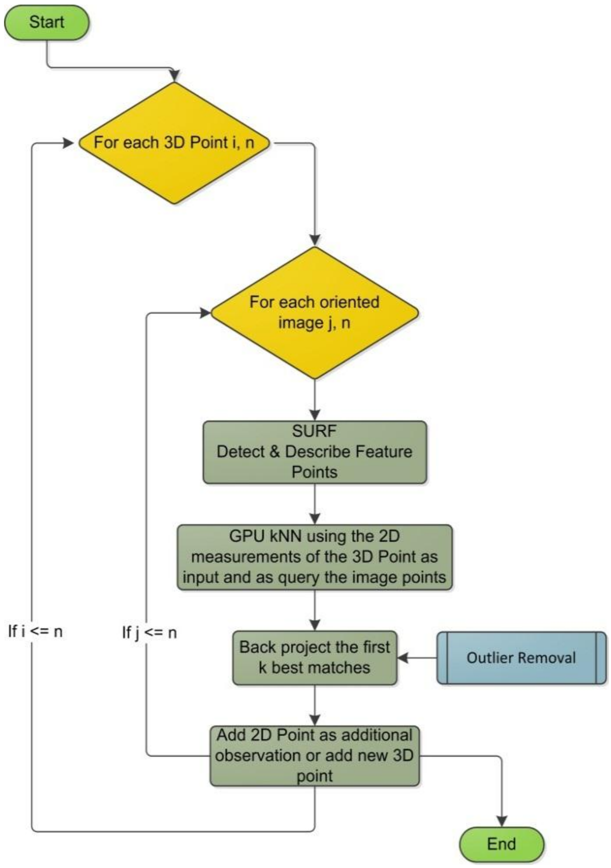
Start
For each 3D Point i, n
For each oriented image j, n
SURF Detect & Describe Feature Points
GPU kNN using the 2D measurements of the 3D Point as input and as query the image points
If i <= n
If j <= n
Back project the first k best matches
Outlier Removal
Add 2D Point as additional observation or add new 3D point
End

Figure 6. Flowchart of the proposed algorithm for adding new point or additional rays.

Figures 7 shows a characteristic image of one photogrammetric network, as well the resultant geometry of the object space. It aims to highlight that the present methodology does not have any issues with repetitive patterns. Figure 8 displays the amount of information that can be extracted from narrow baseline imagery, even though repetitive patterns are present. Additionally, in order to ascertain the reliability of the algorithm in cases of repetitive patterns, various tests with images displaying different parts of the same object were performed. In such cases the algorithm correctly returned no point correspondences.

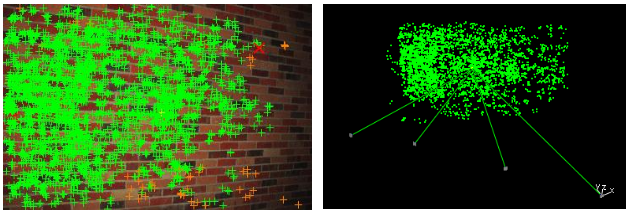

Figure 7. A characteristic image of the photogrammetric network along with the resultant object space.

Finally, existing photogrammetric networks that employed coded targets were solved using both the proposed and the 'standard' automated approach in order to compare the computed interior orientation (IO) and exterior orientation (EO). This investigation showed no significant variation in the results obtained, either in IO or EO determination. Barazzetti et al. (2011b) also performed a similar investigation for a targetless camera calibration procedure and reported a similarly successful outcome.

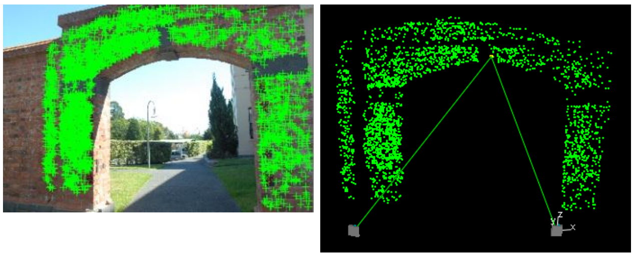

Figure 8. Matching results for a pair of images. The resulting object array contained 2706 points.

6. CONCLUSION

Through adoption of the filtering process developed in the reported research, it was possible to filter all mismatched points and successfully perform a RO using the coplanarity condition equations. Results presented in the paper demonstrate that the approach developed can orient pairs of highly convergent images without the need for targets. This highlights the benefits of using the proposed filtering approach, as it allows subsequent automatic orientation of networks comprising an arbitary number of images. The performance of the methodology is impressive in terms of the time required to detect, describe and match the feature points, along with the filtering methods introduced. Parallel processing implementations in the CPU as well the GPU provided a significant performance boost to the algorithms, as the filtering and RO stage could be performed in less than a second in cases where the number of extracted SURF points exceeded 40,000 per image. Finally, it is noteworthy that an image matching accuracy of better than a 0.3 pixels was attained, not atypical for many moderate-accuracy close-range photogrammetric measurement applications.

7. REFERENCES

Ankerst, M., Breunig, M.M., Kriegel, H.P., Sander J., 1999. OPTICS: ordering points to identify the clustering structure, In: *Proceedings of the 1999 ACM SIGMOD international conference on Management of data*, pp. 49-60, Philadelphia, Pennsylvania, USA

Arthur, D. and Vassilvitskii, S., 2007. k-means++: the advantages of careful seeding. In: *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1027–1035.

Barazzetti, L., 2011a. Automatic tie point extraction from markerless image blocks in close-range photogrammetry. Ph.D. thesis, Politecnico di Milano.

Barazzetti, L., Mussio, L., Remondino, F., Scaioni, M., 2011b. Targetless camera calibration. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 38.

Bay H., Ess A., Tuytelaars T., Gool L.V., 2008. SURF: Speeded Up Robust Features. In: *Computer Vision and Image Understanding (CVIU)*, Vol. 110, No. 3, pp. 346-359

Ester, M., Kriegel, H.P., Sander J., Xu X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, AAAI Press, pp. 226–231.

Garcia, V., Debreuve, E. & Barlaud M., 2008. Fast k nearest neighbor search using GPU. In: *Proceedings of the CVPR Workshop on Computer Vision on GPU*, Anchorage, Alaska, USA.

Jazayeri, I., 2010. Image-based modelling for object reconstruction. Ph.D. thesis, The University of Melbourne.

Lowe, D.G., 1999. Object recognition from local scale-invariant features". In: *Proceedings of the International Conference on Computer Vision*, pp. 1150–1157, Corfu, Greece.

Photometrix, 2012.
http://www.photometrix.com.au (19/03/2012).

Remondino, F., 2006. Detectors and descriptors for photogrammetric applications. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36(3), pp. 49-54.