

IMAGE MOSAICING VIA QUADRIC SURFACE ESTIMATION WITH PRIORS FOR TUNNEL INSPECTION

Krisada Chaiyasarn, Tae-Kyun Kim, Fabio Viola, Roberto Cipolla, Kenichi Soga

Department of Engineering, University of Cambridge, Cambridge, UK

ABSTRACT

In this paper, a system which constructs a mosaic image of the tunnel surface with little distortion is presented. The tunnel surface is typically composed of a roughly cylindrical surface and protuberant regions containing objects such as pipes, pans and tunnel ridges. Since the true surface is neither planar nor quadric, existing mosaicing methods, which assume either homography or quadratic motion models, suffer from distortion. The proposed system obtains a sparse 3D model of the tunnel by multi-view reconstruction. Then, the Support Vector Machine (SVM) classifier is applied in order to separate image features lying on the cylindrical surface from those of the non-surface. The reconstructed 3D points are re-projected into images to retrieve the priors given by the SVM classifier for accurate cylindrical surface estimation. The final mosaic image is obtained by flattening the estimated textured surface onto a plane. The results suggest that the mosaic quality depends critically on the surface estimation accuracy and the proposed system is able to produce the mosaic image that preserves all physical sense, e.g. line parallelism and straightness, which is important for tunnel inspection.

Index Terms— Visual inspection, Mosaicing, Multiple view geometry, Support Vector Machine

1. INTRODUCTION

Much of the underground infrastructure, especially in London, was constructed over half a century ago and has shown signs of deterioration which causes problems in structural integrity. Maintenance works are regularly carried out and visual inspection is a common practice that is used in detecting and monitoring anomalies (e.g. cracks, spalling, and staining). Photographs are used as a mean of recording anomalies although over years they become large and difficult to organise. Therefore, an automatic tool such as image mosaicing will assist inspectors in organising images and in examining the tunnel surface much more effectively.

In recent years, image mosaicing technology has been greatly improved which is attributed to better algorithms in



Fig. 1: The result from a state of the art mosaicing method based on the planar motion model [12]. The parallel lines (tunnel ridges) diverging along the horizontal axis of the image and mis-registrations due to protuberance are evident (within the red box).

feature detection and matching [1], and the use of the advanced motion models. The homography-based model, which assumes the planar surface, is widely adopted due to its simplicity. However, it suffers from perspective distortion [1], especially in our setting in which photos are taken near the tunnel surface (see Fig.1). To allow the non-planar surface, it is often assumed that a camera undergoes pure rotation since estimating the full perspective model is generally difficult. Although this assumption is a popular choice for large-scale image stitching, it is inherently limited when the camera motion is arbitrary. Another important line of mosaicing is to map images to cylindrical or spherical surfaces with different camera motion constraints [2, 3]. The hierarchical estimation of the motion model has been proposed to relieve the motion constraints for the quadric surface mosaicing [3]. The global consistency, in which more than a single pair of images are registered, also helps to improve the quadric mosaic quality [2, 3]. However, the true tunnel geometry is not quadric but typically a mixture of a cylindrical surface and protuberant regions such as pipes, pans and tunnel ridges. Hence, the methods of the quadric surface mosaicing still cause distortion subject to the amount and degree of protuberance from the cylindrical surface.

We present a system that mosaics the tunnel images via robust quadric surface estimation. The multi-view reconstruction and classification of the cylindrical surface points by Support Vector Machine (SVM) enable accurate surface estimation which results in a mosaic image with little distortion. The rest of the paper is organised as follows. Section 2 explains

The research is funded by the Engineering and Physical Sciences Research Council (EP/E003338/1 Micro-Measurement and Monitoring System for Ageing Underground Infrastructures (Underground M3))

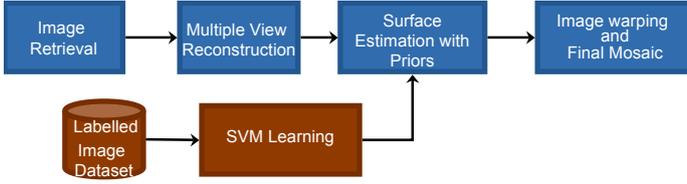


Fig. 2: System outline.

the outline of the proposed system. Multi-view reconstruction is explained in Section 3. Section 4 describes classification of points on the tunnel images and its use in surface estimation, and texture mapping. Section 5 contains results and discussion. Conclusions and future works are included in Section 6.

2. SYSTEM OUTLINE

Figure 2 shows an outline of the proposed system. The first component is the image collection. A camera which is calibrated from the Caltec calibration toolbox [7] is employed to obtain images from the real underground environment. The images are then input to the sparse multi-view reconstruction algorithm based on the work of Snavely *et. al.* [5] as explained in Section 3. It is important to note that, at this stage, points from both the surface (hereinafter, the quadric or cylindrical surface is simply written as surface) and the protuberant regions (see Fig. 1) are useful for estimating the 3D model and the camera projection matrices, although, the latter will need to be removed in the subsequent steps for reliable quadric surface estimation. For SVM classification, some input images are manually labelled to separate the surface to the protuberant regions. The labelled images with their associated image points, which is represented by Scale Invariant Feature Transform (SIFT) descriptors [6], are used for training and evaluating the SVM classifier as explained in Section 4.1. The reconstructed 3D points are then classified into two regions by applying the learnt SVM classifier to their associated image points (each 3D point has more than two corresponding image points). The priors obtained by classification of the 3D points help the quadric surface estimation step as explained in Section 4.2. The surface is correctly estimated by removing the non-surface points that can badly skew the quadric surface estimation. The proposed system approximates the true tunnel surface geometry to be cylindrical, this assumption is exploited for mosaicing. The images are registered to the estimated cylindrical surface and the textured surface is then flattened to yield a mosaic image, see Section 4.3. The sparse 3D model reconstructed by the proposed system is also a useful bi-product for tunnel inspection.

3. MULTIPLE VIEW RECONSTRUCTION

The algorithm in [5, 8] is modified to obtain a sparse 3D reconstruction of the tunnel. The algorithm starts with the

SIFT algorithm to extract features and their associated descriptor vectors for each image [6]. It has been shown that this feature is invariant to scale and robust to affine transformation. The 128-dimensional descriptor vectors are then pairwise matched by the k-nearest neighbour method for each pair of images. Incorrect matches are filtered out by the geometrical consistency constraint by the Random Sampling Consensus (RANSAC) algorithm with the 8-point algorithm. The fundamental matrix \mathbf{F} obtained by the RANSAC is used to obtain the essential matrix, $\mathbf{E} = \mathbf{K}_2^T \mathbf{F} \mathbf{K}_1$, where \mathbf{K} is the camera calibration matrix obtained by [7]. The essential matrix is decomposed by Singular Value Decomposition (SVD) to produce the relative camera rotation and translation matrices, hence the projection matrices, between each camera pair. The ambiguity of the projection matrices is solved by the cheirality constraint [9]. To register all cameras in a global scale, the scale ambiguity is solved in a triplet-wise fashion. The scale ratio of each triplet is found by the SVD method.

Once all camera projection matrices in the global frame are found, 3D points can be recovered from any pair of the projection matrices by triangulation. In our algorithm, a 3D point is triangulated from the first pair of the cameras that see this point. The DLT algorithm is used in the triangulation step [4]. All 3D points, the camera rotation and translation matrices, and the calibration matrix are then used to initialise the Bundle Adjustment (BA) algorithm. This algorithm can greatly improve the reconstruction accuracy by global registration in which all parameters (i.e. 3D points and all camera projection matrices) are optimised together by minimizing a suitable cost function. The implementation of the BA algorithm is obtained from [10].

4. QUADRIC SURFACE ESTIMATION FOR MOSAICING

4.1. Implementation of Support Vector Machine

The SVM classifier is applied to discriminate the tunnel surface points from the non-surface points. The interest points detected by the SIFT algorithm on or near the protuberant regions are collected as the non-surface class and the other interest points as the surface class, as shown in Fig. 4 (Right). The image patches centered on the interest points of the two classes exhibit distinctive appearances as shown in Fig.3, thus they are separable. The training data set is composed of $\{\mathbf{x}_i, y_i\}$ where \mathbf{x}_i is the 128-dimension SIFT descriptor vectors [6] and $y_i \in \{-1, +1\}$ is the class label. The SVM classifier [11] that optimally separates the positive class from the negative class is obtained as

$$f(\mathbf{x}) = \sum_{i=1}^{N_s} \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + b \quad (1)$$

where $\{\mathbf{s}_1, \dots, \mathbf{s}_{N_s}\}$ is the set of support vectors, which is a subset of the training vectors, and α_i is the weight for the i -th

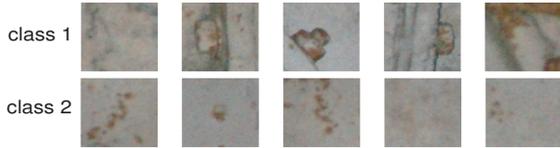


Fig. 3: An example of two classes. The interest points on and near the protuberant regions are Class 1 and on the surface are Class 2.

support vector and b is the bias. The kernel function that we use in the system is $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2}$ where σ is the kernel variance. The variance σ and the penalty constants are set using a validation set.

4.2. Quadric Surface Estimation with Priors

The tunnel surface geometry is assumed to be a cylinder in which a 3D coordinate \mathbf{X}_i is described by:

$$(\mathbf{X}_i - \mathbf{C})^T \frac{(\mathbf{I} - \mathbf{A}\mathbf{A}^T)}{r^2} (\mathbf{X}_i - \mathbf{C}) = 1 \quad (2)$$

where the centre \mathbf{C} , the directional vector \mathbf{A} and the radius r are used to parameterise the cylinder. Since not all of the 3D points are equally reliable in the surface estimation, the weighted cost function is proposed as

$$\sum_i w_i d(\mathbf{X}_i)^2, \quad w_i = \sum_j P(\mathbf{x}_{ji}) \quad (3)$$

where d is the distance of the 3D point to the estimated cylinder and P is the probability that the corresponding image point belongs to the surface class. The SIFT vector of the j -th image point of \mathbf{X}_i is denoted by \mathbf{x}_{ji} and $P(\mathbf{x}) = 1/(1 + \exp(-f(\mathbf{x})))$ where the function f is the SVM classifier in (1). That is, each 3D point is re-projected into the images that see the 3D point and the probabilities are summed over the associated images coordinates. The surface parameters are estimated by minimizing the weighted cost function using the Gauss-Newton method. Using a more robust estimator such as the RANSAC algorithm with the SVM priors is an interesting future work.

4.3. Image Warping and Final Mosaicing

A cylinder is the surface of the zero Gaussian curvature so it is possible to define a local isometry for flattening the curved surface onto a plane. Moreover, given the constraints on the image collection process, cameras are put inside the cylinder and each ray intersects the surface in only a single visible point, defining a bijection between an image sample and a point on the surface. These facts allow us to define a warping producing the flattened versions of the input images. The warped images can finally be mosaiced with any standard stitching algorithms using the planar projective motion model. The Microsoft Image Composite Editor [13] is used to obtain the final mosaics in both Figure 1 and 6(b).

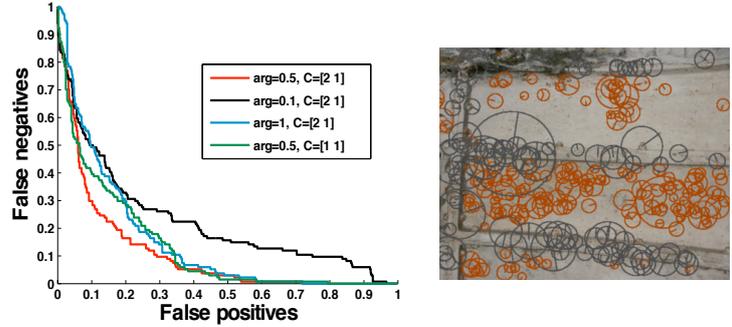


Fig. 4: SVM classification results. The ROC curves for the different kernel parameters and the penalty constants (Left); a test example: the points in gray are classed as non-surface points or class 1, and the points in red are surface points or class 2 (Right).

5. RESULTS AND DISCUSSION

For the SVM classification, the data set contains 1369 training points and 635 test points from 19 labelled images. The ROC curves for the different variance and the constants are shown in Figure 4 (Left). The best performance is achieved when the kernel variance and the penalty constants are set as $\sigma = 0.5$ and $[C_1, C_2] = [2, 1]$, respectively. Figure 4 (Right) shows an example of image points classified by the learnt classifier.

Figure 5 shows the sparse 3D reconstruction model of the tunnel without the global registration by the BA, (b), and with the BA, (c). The tunnel linings are clearly seen after the BA is performed. The convergence graph of the BA algorithm, (a), quantitatively shows significant improvement in the global registration as the cost function converges to a correct minimum. The 3D points classified as the surface points by the SVM classifier are marked as red in Figure 5(c). The estimated surface with (in blue) and without (in red) the SVM classification is shown in Figure 5(d).

The top figure in Fig. 6(a) shows the result before the BA is applied. Misalignment in the overlapping regions are clearly seen due to the errors in the camera registration. The bottom figure in Fig. 6 (a) illustrates the result after the BA is used but without the SVM classifier employed. Skewness in the surface estimation induced by the non-surface points causes noticeable distortion in the mosaic image shown as the curvature in the tunnel ridges. In this result, parallelism of the tunnel linings is somewhat preserved, but not the line straightness. Similarly, the physically incorrect mosaic is seen in the homography-based mosaic in Fig. 1 where the parallel lines diverge. Figure 6(b) shows the result after the cylindrical surface is corrected by the SVM classification. Noticeably, both line parallelism and straightness are preserved. The angle between the vertical and horizontal lines is 90° which indicates the tunnel in the correct physical sense. This is important for tunnel inspection. The region within the red box indicates a small misalignment compared with significant distortion shown in Fig. 1.

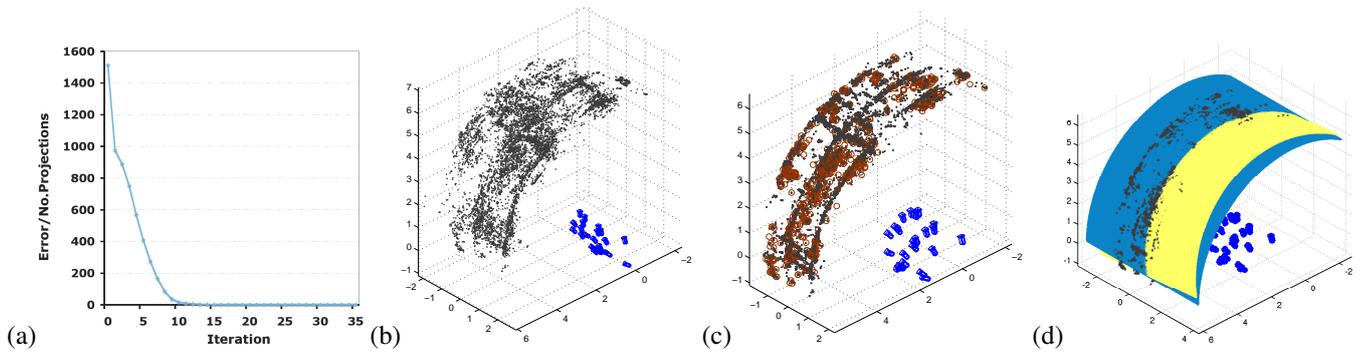


Fig. 5: The sparse multi-view reconstruction models and the surface estimation. The convergence graph of the BA algorithm (a). The reconstructed models without and with the BA (b and c, respectively). The estimated surfaces, (d).

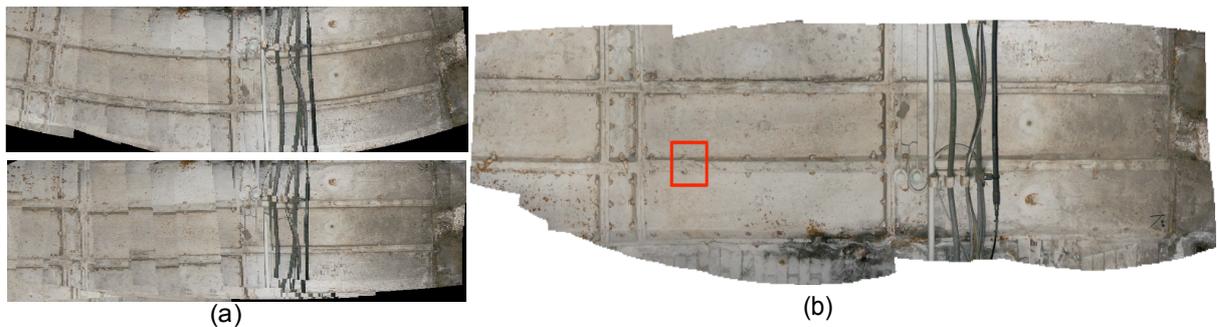


Fig. 6: Mosaic image results. (a) The top image is the result with BA but without SVM, the bottom image is the result before BA. (b) The final result with BA and SVM applied.

6. CONCLUSIONS AND FUTURE WORKS

The system for mosaicing images of the tunnel is presented. The quality of the mosaic image depends critically on the estimation of the surface geometry. The SVM classifier is able to separate the points belonging to the tunnel surface the non-surface. The priors obtained by the SVM classification greatly improves the quality of the mosaic. Although, a small misalignment is still observed in one region, this can be improved by locally correcting the region.

In the future, it is planned to conduct further validation on more datasets that were recently acquired. Furthermore, the prototype of the system including the software and the apparatus for acquiring images (e.g. 20–30 metres of the tunnel length) is currently being developed so that it can be practically adopted in the tunnel inspection procedure.

7. REFERENCES

- [1] R. Szeliski, Image Alignment and Stitching: A Tutorial, *Handbook of Mathematical Models in Computer Vision*, 2005.
- [2] H.Y. Shum and R. Szeliski, Construction and refinement of panoramics mosaics with global and local alignment, *Proc. of CVPR*, 953–958, 1998.
- [3] A. Can, C. Stewart and B. Roysam, Robust hierarchical algorithm for constructing a mosaic from images of the curved human retina, *Proc. of CVPR*, 2:289–292, 1999.
- [4] A. Zisserman and R. Harley, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 1999.
- [5] N. Snavely, S. Seitz and R. Szeliski, Photo Tourism: Exploring Photo Collections in 3D, *Int. Conf. on Computer Graphics and Interactive Techniques*, 835–846, 2006.
- [6] D.G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, *IJCV*, 60(2):91–110, 2004.
- [7] J.Y. Bouquet, <http://www.vision.caltech.edu/bouquetj/>, 2008.
- [8] M. Brown and D.G. Lowe, Unsupervised 3D object recognition and reconstruction in unordered datasets, *Int. Conf. on 3-D Digital Imaging and Modeling*, 56–63, 2005.
- [9] D. Nister, An Efficient Solution to the Five-Point Relative Pose Problem, *Proc. of CVPR*, 756–777, 2004.
- [10] M. Lourakis and A. Argyros, *Technical Report 340* <http://www.ics.forth.gr/~lourakis/sba/>, 2004.
- [11] Christopher J. C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery* 2:121–167, 1998.
- [12] Kolor Autopano, <http://www.autopano.net>, 2008.
- [13] Microsoft Image Composite Editor, 2008 <http://research.microsoft.com/en-us/um/redmond/groups/ivm/ICE/>