

Visual Change Detection on Tunnel Linings

Simon Stent · Riccardo Gherardi · Björn Stenger ·
 Kenichi Soga · Roberto Cipolla

Received: September 2013 / Revised: July 2014 / Accepted: October 2014

Abstract We describe an automated system for detecting, localising, clustering and ranking visual changes on tunnel surfaces. The system is designed to provide assistance to expert human inspectors carrying out structural health monitoring and maintenance on ageing tunnel networks. A three-dimensional tunnel surface model is first recovered from a set of reference images using Structure from Motion techniques. New images are localised accurately within the model and changes are detected versus the reference images and model geometry. We formulate the problem of detecting changes probabilistically and evaluate the use of different feature maps and a novel geometric prior to achieve invariance to noise and nuisance sources such as parallax and lighting changes. A clustering and ranking method is proposed which efficiently presents detected changes and further improves the inspection efficiency. System performance is assessed on a real data set collected using a low-cost prototype capture device and labelled with ground-truth. Results demonstrate that

The authors gratefully acknowledge the support by Toshiba Research Europe.

S. Stent (✉) · K. Soga · R. Cipolla
 Department of Engineering,
 University of Cambridge,
 Trumpington Street, Cambridge, CB2 1PZ,
 United Kingdom
 Tel.: +44 1223 332 748
 Fax: +44 1223 332 662
 E-mail: sistent@cantab.net

R. Gherardi · B. Stenger
 Toshiba Research Europe Ltd.,
 Cambridge Research Laboratory,
 208 Cambridge Science Park,
 Milton Road, Cambridge, CB4 0GZ,
 United Kingdom

our system is a step towards higher frequency visual inspection at a reduced cost.

Keywords Change Detection · Anomaly Detection · Structure from Motion · Structural Health Monitoring · Infrastructure Inspection · Tunnel Inspection

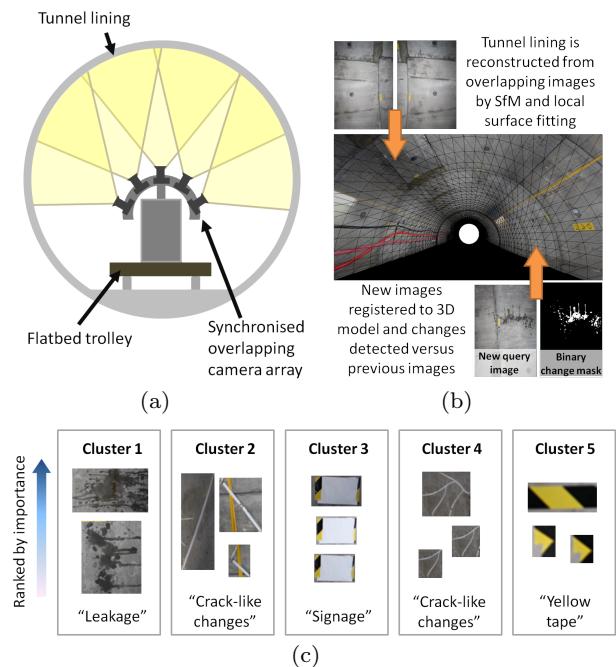


Fig. 1 Illustration of the system. (a) Hardware for data capture; (b) changes detected in new images by localising within a reconstructed reference model; (c) sample output: detected changes are clustered by appearance and ranked within each cluster according to a user-defined importance measure.

1 Introduction

Efficiently monitoring the structural health of large-scale tunnel infrastructure is a socially important challenge. Tunnel infrastructure is commonly located in urban areas and provides essential functions to society such as transport, electricity and communications. As urban populations grow, tunnels are often worked beyond their original design specifications for both function and lifespan, and hence face a growing risk of structural failure.

A critical requirement for effective structural health monitoring is the early detection of any visual changes in the tunnel surface, such as leakages, cracks and corrosion. Early detection allows early intervention, keeping the cost of remedial measures low and reducing the risk of unexpected failure. Detecting such changes is often the work of human inspectors, but given the sometimes adverse working environments and the extensive coverage areas, this is a costly and time-consuming process and is subject to human error. More recently, digital camera and laser technology have been used to capture data to improve the efficiency of visual inspection [16, 20], but systems remain expensive and routine inspections are typically conducted only once every few years.

We present a vision-based change detection system which is a step towards low-cost, high-frequency monitoring. The system's goal is to automatically detect, localise, cluster and rank visual changes on tunnel surfaces in newly acquired images. The system automates the time-consuming process of visual defect discovery, reducing the workload and increasing the effectiveness of expert human inspectors.

This article is an extended version of the conference paper [17]. Section 2 describes the main contributions of the paper in the context of related literature. Section 3 details the theory behind the change detection framework. Section 4 provides an extended description of the complete system and its constituent components. Section 5 describes practical experiments performed, including new results on pixel-level accuracy and on the clustering and ranking of changes. Section 6 concludes with a description of the system's limitations and areas for future work.

2 Related work

The main contributions of this work are twofold. Firstly, we describe a low-cost means of collecting and organising large-scale visual datasets of tunnel linings. Secondly, we devise a framework for change detection on

newly captured, unregistered images. These two contributions are further described below, in the context of related literature.

Data acquisition and reconstruction. Existing automated approaches for tunnel surface inspection tend to make use of more expensive or bespoke visual capture systems such as laser scanners [16] or calibrated laser/camera hybrids [10]. While the use of high-precision depth sensors can enable more accurate and robust geometry estimation than can be achieved from images alone, the increased cost of the sensors makes the systems less economical in situations where such high precision is not necessary.

Alternative approaches using only CCD sensors, such as [19, 20, 25], avoid dealing with geometric information by assuming an annular 2D world. Such systems rely on accurate camera positioning to maintain a constant distance to the tunnel surface. The approach in [6] describes a means of overcoming this by inferring geometric information from the images, but the use of this information is limited to quantifying the scale of cracks rather than to facilitate comparison with previous images.

We opt for a fixed but otherwise unconstrained array of synchronised, overlapping, consumer-grade digital cameras (fig. 1(a)). The low cost of the capture device allows for the possibility of assigning one or several devices to monitor individual tunnels continuously, rather than using a single expensive device to monitor many tunnels sporadically as is common. From the captured image set, we use Structure from Motion (SfM) techniques [15] to recover 3D geometry, and model the tunnel surface by locally fitting quadric surfaces to the resulting point cloud [2]. A 3D wire-frame surface model, texture-mapped with captured images, is shown in fig. 1(b).

Change detection. Change detection in 2D images is a well-studied problem, particularly in the fields of remote sensing, video surveillance and medical imaging [9, 12]. Systems exist for similar applications to ours, such as pattern-matching for concrete crack detection or road surface condition monitoring [16, 20]. However, the detection of general changes on tunnel surfaces does not seem well explored, despite its importance. We identify the three main challenges and related work from the computer vision literature:

- i. *Query image registration.* Accurate registration is an essential prerequisite for changes to be detected without a large number of false positives. In remote sensing, the standard means of registration

is by coarse localisation via GPS, then feature-matching and homography estimation (assuming a planar world geometry). In the tunnel environment, GPS is unavailable and the presence of 3D relief necessitates a geometric model. Recent techniques have adopted voxel-based [11, 18] and mesh-based [4] geometric models for 3D change detection in cluttered scenes with many occlusions. The geometric change detection system in [14] adopts a probabilistic rather than deterministic geometric representation, citing the difficulty of producing sufficiently accurate deterministic models in many realistic scenarios e.g. due to the limited variety of camera poses or insufficient textures. The tunnel environment in our scenario is in general uncluttered and well-defined, and as we are specifically interested in detecting visual changes on its surface, we opt for a simple, scalable, local quadric surface model. The benefit of such an approach is that surfaces can still be recovered with sufficient accuracy for fine-grained change detection even in areas with little texture, given a suitable model.

- ii. *Nuisance variability.* Fig. 2 illustrates some typical sources of nuisance variability in the tunnel environment. One source of false changes between the registered images is image parallax from unmodelled geometry, such as textureless cables and poorly-lit panel anchor holes. This can be avoided by explicitly modelling all geometry [4, 18], however this is challenging in areas of poor texture or limited visibility. Instead, since the tunnel surface is our main concern, we circumvent the problem using a nuisance mask, in the style of [13], which downweights regions of the change image depending on their adherence to the fitted surface model. A further source of nuisance variability is illumination, amplified by the enclosed and poorly lit nature of the tunnel environment. In surveillance applications, background modelling is used to mitigate this variability, but is not feasible with limited temporal information. We investigate single image colour-normalisation and colour-constancy techniques such as Multi-Scale Retinex (MSR) [7], to counter both high and low frequency illumination variability.
- iii. *Many modes of relevant changes.* Many existing systems for large-scale infrastructure monitoring focus on pattern matching to detect specific features such as cracks in concrete [6, 20, 24, 25]. Our main concern is to capture all visual outliers which are not accounted for by understood modes of nuisance variability. Unlike all of the mentioned approaches, we aim to do so by statistical comparison

against previous images rather than by creating a set of heuristics for performing detection of an explicit type of change such as a crack. Clustering the outliers that we detect based on their appearance establishes groups of features such as cracks or leaks, as illustrated in fig. 1(c), but without enforcing any prior knowledge on what types of changes we detect.

3 Theory

We denote the query image by I^q , a function $D^q \rightarrow \mathbb{R}^3$ which maps pixels from location \mathbf{x} in the query image domain D^q to RGB values $I^q(\mathbf{x})$. The set of matching images is given by $\{I_i^m\}_{i=1,\dots,M}$. This is the set of images taken at a previous time instance which have non-zero intersection with the query image. The matching images are registered to the query camera viewpoint such that $D_i^m \subset D^q$ for $i = 1, \dots, M$.

We are interested in obtaining a change map $C : D^q \rightarrow \{0, 1\}$, which maps a location to 1 in case of a change and 0 otherwise. The goal is to achieve invariance to nuisance variability as described above and return a change map of only the relevant changes for pixel-level or image-level classification.

3.1 Change detection

We first consider the case of estimating a single change map C_i from the query image I^q and one of the matching images I_i^m from the database:

$$p(C_i | I^q, I_i^m) = \frac{1}{Z} \mathcal{L}(I^q | I_i^m, C_i) p(I_i^m | C_i) p(C_i) \quad (1)$$

$$\propto \mathcal{L}(I^q | I_i^m, C_i) p(C_i), \quad (2)$$

where the normalising constant Z and the matching image prior $p(I_i^m | C_i)$ are disregarded as they are constant with respect to the query image. This leaves a likelihood term $\mathcal{L}(I^q | I_i^m, C_i)$ and a prior term $p(C_i)$ for the probability of change at any given pixel. In our experiments we set this prior to a constant value, but in a working system it might be varied depending on the location of the image pixel within the tunnel. This would allow a user to bias the system to detect changes with more sensitivity in areas of structural importance (such as where the tunnel passes nearby other critical infrastructure).

We define a distance function d between the query and the matching image, and the likelihood term is then expressed as the distribution of values of d given



Fig. 2 Sources of nuisance variability. From left to right: new query image with no relevant change; warped matching image from database; absolute difference image, with brighter areas indicating larger changes. All differences observed in this final image are caused by nuisance variability rather than relevant change. Such nuisances include low frequency contrast changes across the tunnel lining, hard shadows around off-lining geometry and glare from specular surfaces. They are caused here by the changes in camera position and lighting.

Table 1 Subset of distance functions examined.

Feature	Dist.	Formula/Reference
RGB	L2	(R, G, B)
Chromaticity	L2	$(R, G, B)/R + G + B$
Gray-world	L2	$(R/\bar{R}, G/\bar{G}, B/\bar{B})$
Multi-scale Retinex	L2	See [7]
Dense SIFT	Angle	See [8, 21]
Grayscale	NCC	$\frac{1}{n} \sum_{x,y} \frac{(c_{x,y}^q - \bar{c}^q)(c_{x,y}^m - \bar{c}^m)}{\sigma_{c,q} \sigma_{c,m}}$
Grayscale dist.	Regression error	$ c_{x,y}^m - g(c_{x,y}^q) $, see [5]

whether or not a change has occurred at a particular location:

$$\mathcal{L}_d(I^q | I_i^m, C_i) = \begin{cases} \exp(-d(I^q | I_i^m)/\sigma^2) & \text{if } C_i = 0 \\ \mathcal{U}(d) & \text{if } C_i = 1 \end{cases}, \quad (3)$$

where $\mathcal{U}(d)$ is a uniform distribution over the range of values of d . The smoothing constant σ is set as the mean value of d over the whole query image set. Each matching image, I_i^m , provides information for changes to be identified in its areas of overlap with I^q .

3.2 Choices for distance function

The distance function, d , maps corresponding query and matching image pixels into a feature space and returns a value, $d(I^q | I_i^m)$, using some distance metric. A good choice of function is one that detects relevant changes, yet is invariant to changes due to nuisance variables. Table 1 details a subset of the functions that we examined. These include: colour-normalisation techniques such as chromaticity and gray-world; colour-constancy techniques such as Multi-Scale Retinex (MSR); a spatial histogram of gradients technique in the form of dense Scale Invariant Feature Transform (SIFT); a measure of textural similarity

in the form of grayscale Normalised Cross-Correlation (NCC); and a measure of violation of the smooth relationship, g , between query and matching image intensities fitted in local windows using polynomial regression.

3.3 Combination of multiple change maps

In many cases the query image contains regions which are visible in multiple matching images. In these areas, we can combine the outputs of the individual change maps using a probabilistic OR function:

$$p(C(\mathbf{x})) = 1 - \prod_{\{i: \mathbf{x} \in D_i^m\}} (1 - p(C_i(\mathbf{x}))), \quad (4)$$

where D_i^m is the domain of the matching image I_i^m and the dependencies on I^q and I_i^m are dropped for clarity.

3.4 Geometric prior

We use the information available to us from our SfM reconstruction to form a geometric prior, $p(C|\mathcal{G})$, included as follows:

$$p(C_i | I^q, I_i^m, \mathcal{G}) = p(C_i | I^q, I_i^m) p(C | \mathcal{G}). \quad (5)$$

The prior makes use of the recovered scene geometry, \mathcal{G} , which maps image locations to corresponding 3D points: $D^q \subset D^m \rightarrow \mathbb{R}^3$. The objective of the prior is to mask out nuisance changes caused by geometry or poorly reconstructed features. It can thus also be thought of as an inverse ‘nuisance map’ [13].

To construct the prior, we first group the image interest points into an inlier (on-surface) and outlier (off-surface) set, based on the distance of their corresponding 3D points to the nearest point on the locally fitted surface. Given the relatively sparse nature of \mathcal{G} (fig. 3(b)), we next apply mean-shift segmentation to

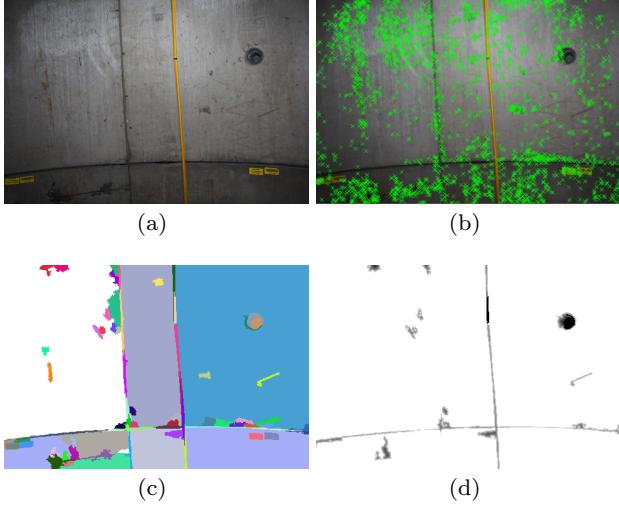


Fig. 3 Geometric prior. (a) Query image (b) Distribution of reconstructed SIFT features (green) (c) Mean-shift segmented image with colour-coded segments (d) Final geometric prior (black areas indicate off-surface or uncertain geometry).

the query image [3]. This delineates the image into pixel groups of similar colour and texture (fig. 3(c)).

Inliers and outliers contained within a pixel group vote towards its overall classification. Pixel groups containing only outliers are classified as off-surface and assigned a prior probability of zero, i.e. changes in those regions are considered to be nuisance variability and are ignored. Pixel groups containing more inliers than outliers are classified as on-surface and assigned a prior probability of one. For pixels lying in groups which contain no points, or fewer inliers than outliers, the prior depends on the distance of the pixel to the nearest inlier. The prior is therefore expressed as:

$$p(C(\mathbf{x})|\mathcal{G}) = \begin{cases} 1, & \text{for on-surface groups.} \\ 0, & \text{for off-surface groups.} \\ \exp\left(\frac{-||\mathbf{x}-\mathbf{x}_{in}||}{\sigma_{\mathcal{G}}^2}\right), & \text{otherwise.} \end{cases} \quad (6)$$

where $\sigma_{\mathcal{G}}$ controls smoothness in uncertain regions and \mathbf{x}_{in} signifies the nearest inlying 2D SIFT feature.

The geometric prior thus downweights changes where the geometry is either known to be off-surface or known to be unreliable. The latter is important in the tunnel environment, where off-surface features such as cables and boxes tend to have matte, featureless surfaces and are therefore reconstructed poorly using feature-based SfM. Fig. 3(d) shows an example of a geometric prior mask, downweighting changes along the yellow cable and in the panel anchor (surface hole).

4 System description

A flowchart illustrating the main processes of our system is shown in fig. 4. We now describe each process in turn.

Reference images acquisition. At time t_0 , we use a prototype capture system consisting of five cameras with synchronised shutters and flash units, arranged in a semi-circular array as shown in fig. 1(a) to capture a stream of reference images $\{I_i^r\}_{i=1,\dots,R}$. Each reference image overlaps with its immediate neighbours by 50%, both radially and longitudinally along the tunnel.

Structure from Motion (SfM). SIFT feature descriptors [8] are extracted from each reference image using a GPU implementation [22]. The image set is split into smaller overlapping subsets and reconstructed in parallel using standard SfM [15, 23]. Local reconstructions are stitched into a global coordinate frame, centred at the first image in the sequence, using overlapping feature correspondences. The resulting reconstruction is illustrated in fig. 5. Note that the global geometry recovered by the reconstruction process suffers from drift as loop closure is not possible in linear tunnels. However, such drift has little consequence on the change detection system, which relies only on local geometry.

Atlas (3D database) builder. Following SfM, the pose of each reference image is known in a global coordinate system. Each reference image is stored along with its intrinsic and extrinsic parameters, the set of descriptors for its N largest SIFT features over scale and its corresponding subset number in an “atlas” database. This is used later for query image localisation.

Geometric primitive (surface) fitting. The tunnel surface is modelled locally with quadrics, fit to each point cloud using robust non-linear optimisation with outlier removal. As in [2], we find that in the tunnels we consider, a piecewise cylindrical representation is sufficient, though the system can be trivially extended to any extruded shape (e.g. square or rectangular tunnels). This representation does not limit the system to straight tunnels, but can be used for any curved tunnel provided its gradients are smooth and shallow enough to be accurately approximated as locally straight.

Query image(s) acquisition. Query images, I_i^q , are acquired at time $t_1 \neq t_0$ using either the same capture device or a new device such as a human inspector’s camera. In the former case, the query image data will

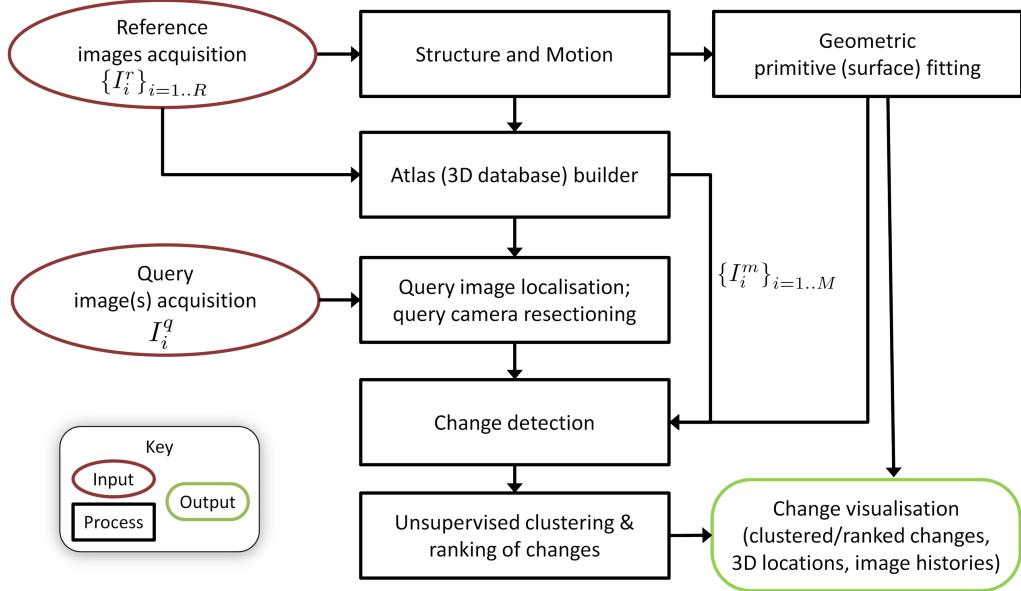


Fig. 4 Flowchart of main system processes.

be dense and overlapping and hence can be used as a new reference dataset for t_1 ; in the latter case, the query image data will be sparse and unordered. In our evaluation, we assumed the latter, making localisation more challenging.

Query image localisation/camera resectioning.

Approximate k -nearest neighbours matching using a k -d tree is used to match the descriptors of the N largest 2D SIFT features in the query image to the atlas database. In our experiments we set $k=5$ and $N=300$. Each match is a weighted vote and votes are aggregated to find the highest scoring reference image subset. Within the subset, RANSAC-based registration is performed over all SIFT features and the query image camera is accurately resectioned with radial distortion estimation. This method of localisation was found to be sufficiently discriminative on concrete tunnel surfaces to correctly register all of our query image dataset.

Change detection. Images from the reference dataset which overlap with the query image are then back-projected onto the recovered tunnel surface and re-projected into the query image. This provides a set of matching images, $\{I_i^m\}_{i=1,\dots,M}$, for change detection, as described in sec. 3.

Unsupervised clustering and ranking of changes.

To present the detected changes to the user in an efficient manner, we employ an unsupervised clustering and ranking approach. The benefit of clustering, even if the number of clusters is large, is that it can remove the

need for the user to address each image change individually. This is especially useful as entire groups of real but unimportant changes can be quickly disregarded (e.g. the addition/removal of a cable along the complete length of the tunnel or the addition of a yellow chainage marker on every panel along the tunnel, which would appear as a change in many images).

The change probability maps are first thresholded with hysteresis to give a discrete set of connected changes. Each connected change is then represented as a 6D point in a simple colour and shape-based feature space:

- mean colour of (MSR-corrected) change as it appears in the query image (3D),
- perimeter to area ratio (1D),
- ratio of principal axes (1D),
- morphological Euler number (1D) - the number of unchanged connected components surrounded by the change.

The feature space is normalised and mean-shift clustering is used with an adaptive bandwidth in order to over-cluster the changes. In our experiment, we set the number of clusters at 100, far greater than the ~ 10 types of changes applied. The motivation for over-clustering is to ensure that clusters remain homogeneous to avoid grouping together different types of changes.

Changes within each cluster are then ranked by a user-defined importance measure. In our experiment, we choose this measure as the sum of the pixels within the change weighted by their change probabilities.

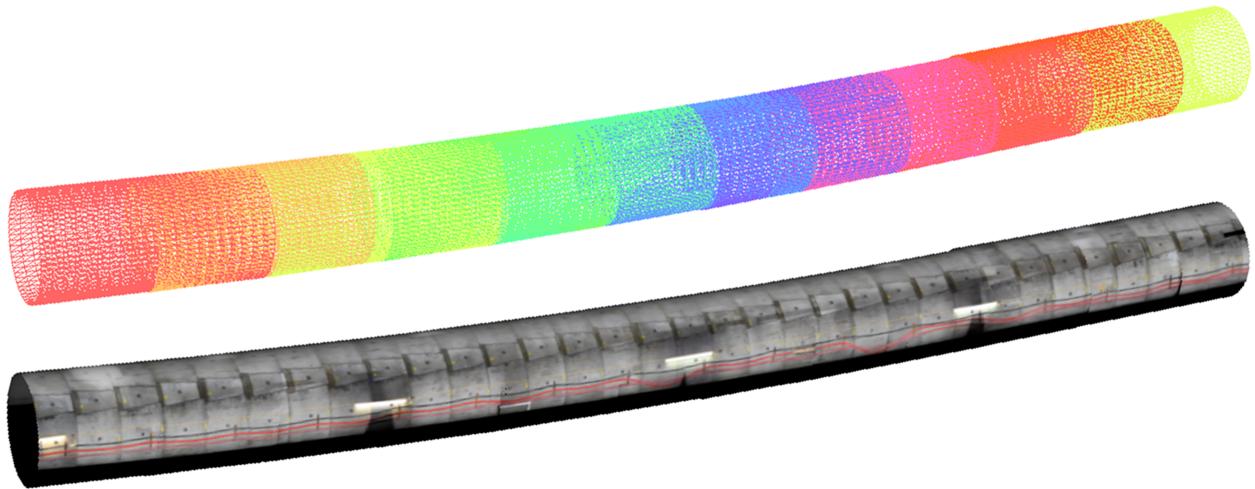


Fig. 5 Tunnel surface reconstruction from overlapping image subsets. Above: overlapping reconstructed subsets are shown in different colours; below: after texture mapping.

Large, high probability changes therefore appear before small, low probability changes.

5 Experiments

We captured data covering 180° of a 3m diameter tunnel section of 100m length. This comprised 1,000 images at a resolution of $3,888 \times 2,592$ pixels. Next, artificial changes were applied to the concrete tunnel surface to simulate the visual changes that might be observed in a real environment - such as leaking, cracking and spalling. A query set of 232 images was taken, of which 131 contained relevant changes. All 232 images were labelled with ground truth for the presence of absence of change, 60 of which were labelled at the pixel-level.

5.1 Qualitative results

Fig. 6 shows three sample queries as well as different distance functions, the geometric prior mask and final change detection results. Relevant changes in seq. 1 and 2 include leaking, fine chalk markings, discolouration and objects attached to the surface. The three illustrated distance functions – gray-world, regression and NCC – pick out changes with different degrees of success. Gray-world tends to amplify changes and has good resolution, comparing each pixel individually without taking into account its neighbourhood, but the model we use is a global one and hence illumination effects are also undesirably amplified. Polynomial regression, implemented here as cubic regression with a 9×9 window, picks out fine changes such as cracks which disrupt the smooth relationship between query and matching image

intensities, but predictably fails to detect larger changes such as water leaks where the entire window (and therefore relationship) is transformed. NCC, implemented here with a 5×5 window, reaches something of a compromise, highlighting both fine and coarser changes by taking into account intensity and spatial information, but at the cost of reduced resolution of the resulting distance image. All methods falsely detect changes from the lighting units and cabling. This is especially evident in seq. 3, where there is significant parallax and specularity in the scene.

The geometric prior in all three cases correctly identifies and removes most of the nuisance change caused by off-surface features. The final column shows a probabilistic output change mask, formed by a combination of gray-world and NCC features, multiplied by the geometric prior as per Eqn. (5). In seq. 1 and 2, its performance is close to ground truth. Seq. 3 illustrates a failure case, caused by the unusual presence of some thread on the normally featureless red cable. The large downweighted area in the geometric prior of seq. 3 corresponds to an area of unknown geometry, as the query image is at the edge of the reconstructed area.

5.2 Quantitative results

5.2.1 Pixel-level performance

The pixel-level detection performance of various combinations of features is compared in the ROC curves of fig. 7(a). Distance functions which explicitly take into account local spatial information (NCC and DSIFT) performed better than methods which compare indi-

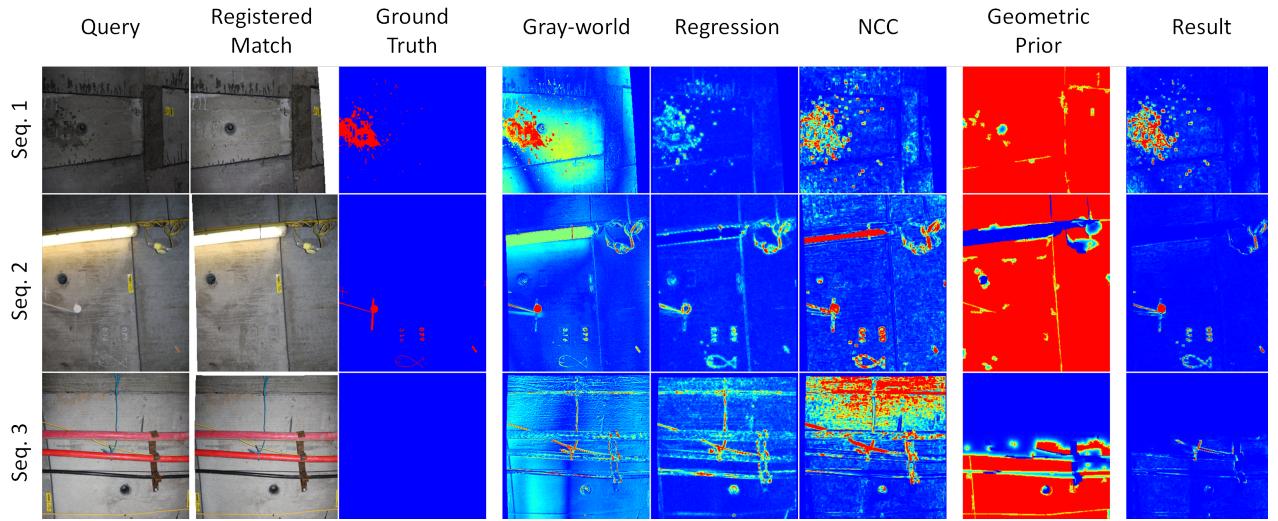


Fig. 6 Illustrative results for three cases. Two sequences (1 and 2) feature relevant change: water leakage, chalk marks, added features; all three sequences feature nuisance change: lighting change (significantly in 1 and 3) and cables, fixtures and other off-surface geometry (significantly in 2 and 3).

vidual query pixels against individual reference pixels (MSR and gray-world) or against a locally fitted relationship (regression). Grayscale NCC returned the best performance, detecting 98% of positive change pixels at a 20% false positive rate. NCC was also tested on MSR and gray-world normalised images, although no significant difference in performance was observed.

5.2.2 Effect of geometric prior

We compared the image-level detection performance of two feature combinations, gray-world and gray-world NCC, in three scenarios: without a prior; using the mean-shift based geometric prior described in sec. 3.4; and using an alternative SLIC superpixel based geometric prior [1]. The SLIC superpixel prior is calculated in the same manner as described in sec. 3.4, but replacing the mean-shift algorithm with SLIC superpixelisation in the segmentation stage. ROC curves are shown in fig. 7(b). Classification performance after the introduction of the geometric priors increases substantially. With no prior, gray-world is initially far more discriminative than NCC, which is more prone to detect changes across nuisance areas of the image space. When a prior is introduced however, nuisance regions are masked out and NCC can safely exploit local spatial information solely in the regions of interest (i.e. the tunnel surface), allowing it to outperform gray-world. Finally, the quantitative performance of our proposed mean-shift prior is shown to be improved versus the more local, SLIC-based prior. We tried several parameter settings for each and found that qualitatively, mean-shift per-

formed better than SLIC. Despite larger computational expense, it was able to capture both irregularly shaped thin structures (e.g. cables) and large flat structures (the tunnel surface) of non-uniform size at the same parameter setting, thus returning a more semantically meaningful and useful segmentation. SLIC, in comparison, could not capture such different structures with a given region size and regularisation parameter.

5.2.3 Image-level performance

The image-level detection performance of various combinations of features is shown in the ROC curves of fig. 7(c). Consistent with our pixel-level results in sec. 5.2.1, NCC performed best, detecting 81% of true positives at 20% false positive rate. Furthermore, running NCC on MSR and gray-world normalised images returned no significant quantitative difference in performance.

One difference of note when comparing Figs. 7(a) and 7(c) is that while gray-world and regression have worse pixel-level performance than MSR, their image-level performance is notably better. In the case of regression, this can be attributed to its failure to detect large areas of change such as the centre of the white circle in seq. 2 of fig. 6. This reduces pixel-level performance, but it still detects the boundaries of such areas accurately (where there is overlap with an unchanged area) and therefore correctly flags the image as containing change. Gray-world returns a higher rate of false positive pixels because it offers little invariance to local lighting changes, but there are relatively few images

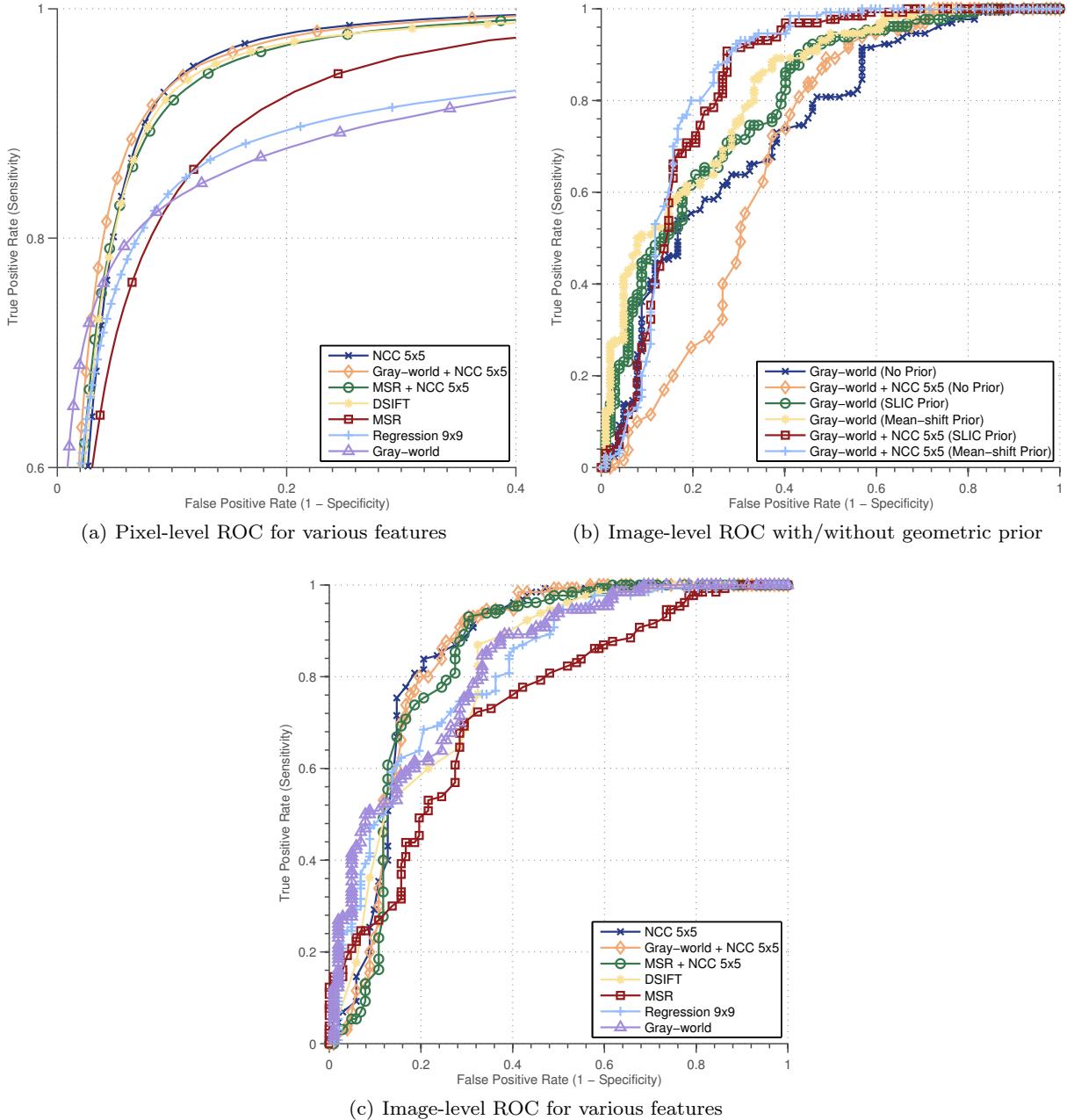


Fig. 7 ROC comparison. (a) Pixel-level performance of various features; (b) image-level performance with and without geometric prior; (c) image-level performance of various features with mean-shift geometric prior. Area-based methods gave the best performance and image-level performance was significantly improved with the introduction of a discriminative geometric prior.

in which this is a problem so image-level performance is not significantly degraded. Conversely, MSR resolves lighting change by using local rather than global image statistics, and so has improved pixel-level performance but is more susceptible to artefacts at sharp local boundaries e.g. between gray concrete and brightly coloured cables. This results in a higher number of false positive images.

5.3 Clustering and ranking results

The top-ranked changes in a subset of the clusters returned after the unsupervised clustering and ranking stage described in sec. 4 are shown in fig. 8. The method employed showed good qualitative performance at picking out groups of similar changes, although it was found that due to the variable visual nature of the false positives, larger clusters with more variation would often

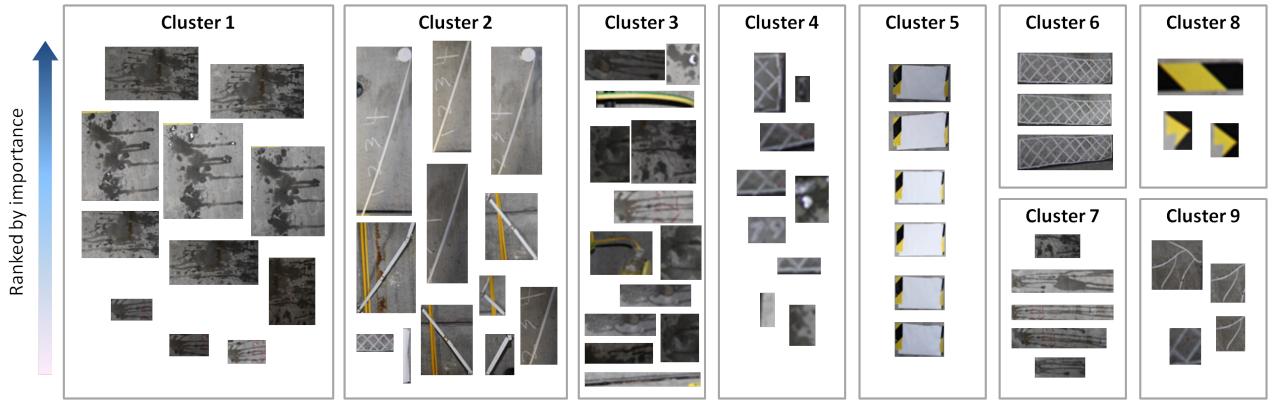


Fig. 8 Top-ranked changes in a subset of the clusters returned after unsupervised clustering and ranking.

contain some contamination (e.g. two instances of yellow cable in cluster 3). Smaller clusters such as clusters 5-9 were generally more homogeneous, although offer less benefit in terms of reducing the workload for the human inspector. It should be noted that in a real system, adding location to the feature space should enhance the results of the method, by allowing for example all crack-like changes in the crown of the tunnel or all leakages in a particular tunnel segment to be grouped together.

6 Discussion

We have presented a system which is suitable for the automated monitoring and detection of general visual changes on smooth, unpainted, concrete tunnel surfaces. Our system is inexpensive to implement and reduces the workload for visual inspection, enabling higher frequency, more effective tunnel inspections and better use of visual inspection data. The change detection framework we present is broadly applicable to any situation where an accurate geometric model can be recovered of the area of interest, such that reference images may be accurately synthesised from the viewpoint of a query image.

6.1 Limitations

A key limitation of the proposed system is that there must be sufficient texture on the tunnel surface to allow reconstruction at a single time instance, and sufficient stable texture to allow registration of images between time instances. Our experience is that concrete and cast iron tunnels are sufficiently textured for both, provided they have not been painted or panelled. However, we have thus far tested in relatively static (utility) tunnels, not in more dynamic environments such as road or subway tunnels, where the build up of dust and dirt over

time might mask the image texture used for localisation. Similar problems may occur between wet and dry environments or at tunnel extremities. The variability introduced by such factors could be mitigated to some extent by using odometry and/or 3D information from the reconstruction to improve query image registration. However, further tests are needed.

A second limitation is that we assume the tunnel geometry has a locally uniform cross-sectional shape, which can be retrieved from the 3D reconstruction. Our experience is that this is a fair assumption in modern pre-cast concrete tunnels which are precisely fabricated, but does not hold true of all tunnels. A significantly varying tunnel geometry would require a more precise approach to surface reconstruction.

Finally, despite the inclusion of the geometric prior, many false positives are still detected around areas such as cables. Performance might be further improved by adding domain-specific knowledge such as segmenting out all cable-like structures which appear as a certain colour in the images. A more generally applicable method would be to segment out cable structures by first reconstructing them using a model-based approach.

6.2 Future work

We are currently developing our capture device to acquire much larger volumes of data automatically. In the future, we plan to test our system in an active tunnel environment to detect real changes on a larger scale and to make a more direct comparison of our system against existing manual inspection techniques.

We also plan to further explore nuisance invariant features and extend our system to more complex tunnel geometries. Another interesting avenue for research is designing the system to scale efficiently in time as well

as space, so that historical data may be stored and used efficiently.

References

1. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC superpixels. *EPFL Tech. Report No. 149300*, 2010.
2. Krisada Chaiyasarn, Tae-Kyun Kim, Fabio Viola, Roberto Cipolla, and Kenichi Soga. Image mosaicing via quadric surface estimation with priors for tunnel inspection. In *ICIP*, pages 537–540, November 2009.
3. D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *Trans. PAMI*, 24(5):603–619, 2002.
4. Olivier Delaunoy, Nuno Gracias, and Rafael Garcia. Towards detecting changes in underwater image sequences. In *OCEANS 2008-MTS/IEEE Kobe Techno-Ocean*, 2008.
5. Rouhollah Dianat and Shohreh Kasaei. Change detection in optical remote sensing images using difference-based methods and spatial information. *Geoscience and Remote Sensing Letters, IEEE*, 7(1):215–219, 2010.
6. Mohammad R Jahanshahi, Sami F Masri, Curtis W Padgett, and Gaurav S Sukhatme. An innovative methodology for detection and quantification of cracks through incorporation of depth perception. *Machine Vision and Applications*, 24:227–241, 2013.
7. D J Jobson, Z Rahman, and G A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.*, 6(7):965–76, 1997.
8. DG Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.
9. D. Lu, P. Mousel, E. Brondízio, and E. Moran. Change detection techniques. *International Journal of Remote Sensing*, 25(12):2365–2407, June 2004.
10. Thomas Ortner, Gerhard Paar, Gerd Hesina, Robert F. Tobler, and Bernhard Nauschnegg. Towards True Underground Infrastructure Surface Documentation. *Real Corp Proceedings 2010, Vienna*, pages 18–20, 2010.
11. Thomas Pollard and Joseph L. Mundy. Change Detection in a 3-d World. In *CVPR*, June 2007.
12. R J Radke, Srinivas Andra, Omar Al-Kofahi, and Badrinath Roysam. Image change detection algorithms: a systematic survey. *IEEE Trans. Image Process.*, 14(3):294–307, 2005.
13. A. Ravichandran and S. Soatto. Long-range spatio-temporal modeling of video with application to fire detection. *ECCV*, pages 329–342, 2012.
14. Ken Sakurada, Takayuki Okatani, and Koichiro Deguchi. Detecting changes in 3D structure of a scene from multi-view images captured by a vehicle-mounted camera. In *CVPR*, 2013.
15. Noah Snavely, Steven M. Seitz, and Richard Szeliski. Modeling the world from internet photo collections. *Int. Journal of Computer Vision*, 80(2):189–210, 2008.
16. SPACETEC, 2012.
17. Simon Stent, Riccardo Gherardi, Björn Stenger, Kenichi Soga, and Roberto Cipolla. An image-based system for change detection on tunnel linings. In *MVA*, 2013.
18. Aparna Taneja, Luca Ballan, and Marc Pollefeys. Image based detection of geometric changes in urban environments. In *ICCV*, pages 2336–2343, 2011.
19. tCrack system by terra vermessungen ag, 2012.
20. Masato Ukai. Advanced Inspection System of Tunnel Wall Deformation using Image Processing. *Quarterly Report of RTRI*, 48(2):94–98, 2007.
21. A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms, 2008.
22. Changchang Wu. SiftGPU: A GPU implementation of scale invariant feature transform (SIFT), 2007.
23. Changchang Wu. VisualSfM: A Visual Structure from Motion System, 2011.
24. Tomoyuki Yamaguchi and Shuji Hashimoto. Fast crack detection method for large-size concrete surface images using percolation-based image processing. *Machine Vision and Applications*, 21:797–809, 2010.
25. Seung-Nam Yu, Jae-Ho Jang, and Chang-Soo Han. Auto inspection system using a mobile robot for detecting concrete cracks in a tunnel. *Automation in Construction*, 16(3):255–261, 2007.



Simon Stent received the B.A. and M.Eng. degree from the University of Cambridge in 2007. He is currently working towards a Ph.D. in computer vision at the Machine Intelligence Lab, University of Cambridge and is funded by Toshiba Research Europe.



Riccardo Gherardi obtained his Master's (2005) and PhD degree (2009) from the University of Verona (Italy). His research interests include 3D reconstruction and geometrical computer vision, real-time binocular vision and computer-aided infrastructure inspection. He joined Toshiba Research Europe in 2010. He was one of the recipients of the company's outstanding service award for his work on the Fukushima power plant.



Björn Stenger received the Diplom-Informatiker degree from the University of Bonn in 2000 and the Ph.D. degree from the University of Cambridge in 2004. From 2004-06 he was a Toshiba Fellow at the Toshiba Corporate RD Center in Kawasaki, Japan, where he worked on gesture recognition and 3D body tracking. He joined Toshiba Research Europe in 2006, where he is currently leading the Computer Vision Group.



Kenichi Soga is a Professor of Civil Engineering and the Head of the Geotechnical Group at the University of Cambridge. He is a Fellow of the Royal Academy of Engineering as well as a Fellow of the Institution of Civil Engineers. He obtained his B.Eng. and M.Eng. from Kyoto University in Japan and Ph.D. from the University of California at Berkeley in 1994. His current research activities are innovative monitoring and long-term performance of geoinfrastructure, energy geomechanics, and modeling of geotechnical construction processes. He has published more than 250 journal and conference papers and is co-author of Fundamentals of Soil Behavior, 3rd edition with Professor James K. Mitchell. He is a recipient of many awards including the George Stephenson Medal (2006) and the Telford Gold Medal (2010) from the Institution of Civil Engineers and the Walter L. Huber Civil Engineering Research Prize (2007) from the American Society of Civil Engineers.



Roberto Cipolla received the B.A. degree in engineering from the University of Cambridge, Cambridge, U.K., in 1984, the M.S.E. degree in electrical engineering from the University of Pennsylvania in 1985, and the D.Phil. degree (computer vision) from the University of Oxford, Oxford, U.K., in 1991. His research interests are in computer vision and robotics and include the recovery of motion and 3D shape of visible surfaces from image sequences; object detection and recognition; novel man-machine interfaces using hand, face and body gestures; real-time visual tracking for localisation and robot guidance; applications of computer vision in mobile phones, visual inspection and image-retrieval and video search. He has authored 3 books, edited 8 volumes and co-authored more than 300 papers.