

Precise Deterministic Change Detection for Smooth Surfaces

Simon Stent^{*†}

Riccardo Gherardi^{*‡}

Björn Stenger[‡]

Roberto Cipolla[†]

[†]University of Cambridge

{sais2, rc10001}@cam.ac.uk

[‡]Toshiba Research Europe

{firstname.lastname}@crl.toshiba.co.uk

Abstract

We introduce a precise deterministic approach for pixel-wise change detection in images taken of a scene of interest over time. Our motivation is for applications such as artefact condition monitoring and structural inspection, where a common problem is the need to efficiently and accurately identify subtle signs of damage and deterioration. The approach we describe is designed to compensate for the three most common sources of nuisance variation encountered when tackling the problem of change detection, namely: viewpoint variation due to camera motion between images, photometric variation due to lighting differences, and changes in image resolution/focal settings.

To tackle viewpoint variation, particularly in areas of low texture, we propose the use of the generalised PatchMatch (PM) correspondence algorithm to compute a dense flow field. The flow field is regularized using a Thin Plate Spline (TPS) model which assumes a smooth underlying geometry and allows registration to be interpolated precisely through areas of low texture or uncertain flow. To compensate for low-frequency lighting variation, we fit a second TPS model to the photometric differences between registered images. Finally, to account for changes in focal settings, we estimate and apply a blurring kernel via optimisation over image differences.

We provide a thorough evaluation of the performance of our method on an illustrative toy dataset and on two recent, real-world inspection datasets. Our approach performs favourably versus state-of-the-art baselines in both cases, while remaining relatively transparent to understand and simple to compute.

1. Introduction

Image-based change detection is a fundamental low-level vision problem with application in areas as diverse as artefact monitoring, remote sensing, urban management,

^{*} First two authors contributed equally.

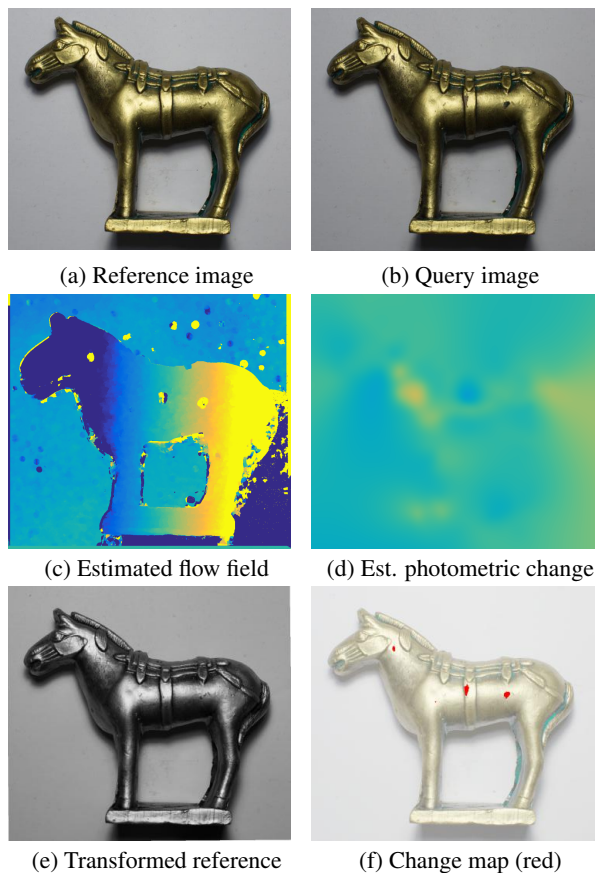


Figure 1: Illustration of our method on an image pair (a,b) from [3]. To counteract viewpoint variation, we estimate sub-pixel flow using the generalised PatchMatch correspondence algorithm (c). To counteract lighting differences, we fit a low-frequency photometric change mask (d). To counteract lens/focal differences, we perform a gradient adjustment (not shown). A registered image (e) is produced in which these three nuisances (viewpoint, lighting, focal) are attenuated. The resulting change map (f) highlights fine-grained changes (red) that may otherwise be hard to spot. Our approach is widely applicable, from the condition monitoring of high-value artefacts to the inspection of civil structures and human skin.

tamper detection, medical imaging and industrial inspection [3, 5, 6, 7, 10, 11]. The problem involves the estimation of a change map between two images or sets of images of a scene, taken at different times. In real-world datasets, precise change map estimation is complicated by “nuisance” changes caused by varying illumination conditions, image capture characteristics (focal, lens, sensor), and camera viewpoints. When the true changes are relatively fine-grained compared to the nuisances, discounting the nuisances to reveal the true signal is especially challenging.

In this work we make one key assumption so as to tackle the problem efficiently: we assume that the unknown scene geometry can be modelled as a collection of smooth, continuously differentiable surfaces. The advantage of this assumption is that two of the largest sources of change – those of viewpoint variation and illumination – may then be robustly modelled and discounted. While the assumption restricts our approach to smooth surfaces, such scenarios abound in practice [3, 10, 11].

Fig. 1 illustrates our approach on the example of high-value artefact monitoring, using an example image pair from the Statues dataset of [3]. The reference image (a) and query image (b) are taken some time apart with approximate but inexact alignment and changing lighting. Our approach is designed to directly compensate for three common sources of “nuisance” change: registration error, lighting changes and focal variation. First, we warp the reference image according to a dense flow field (c), estimated using the generalised PatchMatch correspondence algorithm [1] and a Thin Plate Spline (TPS) model. Next, low-frequency photometric variation (d) is estimated using a second TPS. Finally, focal differences between the images are accounted for by smoothing one image to minimise the sum of absolute difference in the gradients. The combination of these three processes results in a transformed reference image (e) in which changes due to geometry, lighting and focus have been attenuated. Panel (f) shows the detected change, overlaid on the reference image (a). This change map has the same resolution as the reference image and highlights fine-grained changes that may otherwise be hard to spot.

In addition to artefact monitoring, our PatchMatch - Thin Plate Spline (PM-TPS) approach is motivated by its possible application to structural inspection, since many large concrete structures such as tunnels, bridges and dams must be routinely inspected for fine-grained changes such as hair-line cracks on a regular basis [11]. A further possible application is the monitoring of skin lesions to provide early warning or tracking of melanocytic skin cancers [14].

2. Background

Let $I = I_c^t$ and $I' = I_{c'}^{t'}$ be two images of a scene, taken with different cameras, c and c' , at different times t

and t' . Assuming a Lambertian reflectance model with uniform diffuse lighting and ignoring the camera’s point spread function, the formation of the initial image at a pixel \mathbf{x} may be modelled by:

$$I(\mathbf{x}) = \int R(\mathbf{x}, \lambda) L^t(\lambda) S_c(\lambda) d\lambda \quad (1)$$

where $R(\mathbf{x}, \lambda)$ is the scene reflectance, $L^t(\lambda)$ the illumination spectrum at time t , and $S_c(\lambda)$ the spectral sensitivity of camera c . A second image, captured from a different position, under different illumination and with unknown changes in the scene is then formed by:

$$I'(\mathbf{x}') = \begin{cases} \int R(\tau(\mathbf{x}), \lambda) L^{t'}(\mathbf{x}', \lambda) S_{c'}(\lambda) d\lambda & \text{if } \mathbf{x}' \notin \mathcal{C} \\ \int R_u(\mathbf{x}', \lambda) L^{t'}(\mathbf{x}', \lambda) S_{c'}(\lambda) d\lambda & \text{if } \mathbf{x}' \in \mathcal{C} \end{cases} \quad (2)$$

where τ represents a geometric mapping from \mathbf{x} to \mathbf{x}' due to viewpoint change, R_u is the spectral reflectance of an unknown object and \mathcal{C} is the unknown set of changed pixel positions in the image. Detecting changes requires the estimation of the probability $\mathcal{P}(\mathbf{x}' \in \mathcal{C} | I, I')$, but this is challenging since none of the functions τ , L^t , $L^{t'}$, R , R_u or S can be known with certainty.

Related Work. Most change detection approaches start by estimating the geometric mapping, τ . A common assumption, which holds true in many remote sensing and CCTV scenarios [4], is to assume static or purely rotating cameras [6] or planar scenes, reducing τ to a homography or simple model whose parameters can be efficiently estimated using standard feature-based techniques.

It is common to decompose the calculation of τ into the estimation of the camera perspective projection matrices P_c and $P_{c'}$ and the scene geometry $d(\mathbf{x})$, since a pixel \mathbf{x} in I can be projected into I' via the reprojection equation:

$$[\mathbf{x}', 1]^T = s P_{c'} (\hat{P}_c^{-1} [\mathbf{x}^T, 1, d]^T) \quad (3)$$

where \hat{P}_c denotes the matrix obtained by squaring P_c with a 1 on the diagonal and s is an unknown scale factor. The main difficulty lies in estimating the pixel-wise scene depth d ; other parameters can be determined more robustly since evidence can be accumulated across the entire image. The structural inspection work of [12] models d using piecewise quadric surfaces, estimated over a wider field of view using structure-from-motion. The works of [2] and [13] use more general multi-view stereo methods to estimate d . In all such cases, the achievable accuracy is generally not sufficient for very fine-grained change detection, since the models do not hold well in practice. Recently, [8] propose a change detection approach which avoids the explicit determination of d , by integrating for change over depth uncertainties. The authors show this improves change detection in sparsely tex-

tured scenes where geometry is difficult to estimate. However their method is expensive to compute for applications requiring high precision.

In our approach, we avoid concerning ourselves with camera calibration or pose estimation by estimating τ directly, firstly by aggregating evidence for it where available using the efficient generalised PatchMatch correspondence algorithm [1] and secondly by propagating the evidence via a TPS model. This achieves a natural trade-off in reliance between local high-frequency flow data where available and the smooth spline model in sparsely textured or changed regions.

Following the estimation of τ , a common approach is then to compare registered images using patch-based similarity methods such as Normalized Cross Correlation (NCC), census transform or mutual information [6]. Recently, [7] and [11] showed the effectiveness of using deep learned features for this comparison; such approaches may learn invariance to L , S and even errors in τ , but their patch-based nature reduces their ability to detect and localise changes very precisely.

Instead of building such invariance into our similarity function, we explicitly model illumination, again using a TPS for illumination estimation [9]. Our final cost is a simple pixel-differencing. Unlike [7] and [11], our method is deterministic, requires no parameter learning with labelled datasets and maintains the same resolution of output change map as the input image, thus making it ideal for very precise change detection.

Finally, a recent method that is closely related to ours in spirit is that of [3], who jointly optimize over geometry correction, lighting difference and a change mask in a coarse-to-fine manner and apply rank minimization to achieve a final change decision. Unlike our approach, this requires redundancy in the capture set up; images must be captured under multiple illuminations with rough alignment between lighting and capture points. Such redundancy is impractical in many settings, such as large-scale infrastructure inspection, where capture time and cost may be critical. We compare directly with [3] on their dataset in section 4.

3. PM-TPS Method

Our PatchMatch - Thin Plate Spline method is composed of three deterministic steps: geometric registration, photometric matching and gradient compensation. A key assumption underlying our method is that of local smoothness. Without any form of regularization, eqn. 2 could be trivially satisfied for every $\mathbf{x}' \in \mathcal{C}$ by an arbitrary transformation τ , while accounting for any difference through functions L^t , R , and S . For this reason, we impose local smoothness on the estimated geometric and photometric mappings.

This assumption may be violated by edges, sharp geom-

etry or illumination, and in general by any discontinuity in an underlying physical parameter, such as a material with an uneven bidirectional reflectance distribution function. However, local smoothness is one of the defining characteristics of natural images and is prevalent in human-made environments. In practice, this assumption holds for a large range of real-world scenarios. In the following, we describe each step of our method in more detail.

3.1. Geometric Registration

Geometric registration accounts for image differences arising from extrinsic properties of the scene such as shape, camera position or perspective effects. The combination of all these factors, under the assumption of a local smoothness, creates a continuously differentiable flow field between the images.

Such a transformation, if we are not to rely on any scene or geometry priors, can only be modelled by a non-rigid mapping between 2D images. The task must be resilient to untextured regions, outliers, illumination and scale changes. Because of all of these reasons and in the absence of a global motion or transfer model, our choice is to compute sub-pixel, semi-dense correspondences and densify them using an interpolating spline.

3.1.1 Semi-dense correspondences

Semi-dense correspondences are computed with the generalised PatchMatch (PM) [1] correspondence algorithm. PM is an efficient randomized approach to find, for every pixel in image I , the approximate nearest neighbour pixel in image I' , by minimising a matching cost between local image patches. We modify our implementation to search for sub-pixel correspondences over displacement, rotation and scale. We employ Normalized Cross Correlation (NCC) as our cost function, since the locations to be matched could still be affected by absolute and relative biases due for example to illumination variations.

$$\text{NCC}(\mathbf{x}, \mathbf{x}') = \frac{1}{N} \sum_i^N \frac{(\mathbf{x}_i - \mu_{\mathbf{x}}) \cdot (\mathbf{x}'_i - \mu_{\mathbf{x}'})}{\sigma_{\mathbf{x}} \sigma_{\mathbf{x}'}} \quad (4)$$

In the above equation, μ and σ indicate respectively the mean and variance of the vectors \mathbf{x} and \mathbf{x}' . Fig. 2 shows the intermediate results computed during this stage, on the example of Fig. 1. Panels (a-c) show the flow fields for x displacement, y displacement and rotation while (d) shows the NCC cost map of the final PM solution after three complete iterations (one iteration includes 4 diagonal scans, one from each corner). Each map shows gross errors in the presence of change or discontinuities.

3.1.2 Dense flow field inference

To create a continuously varying displacement field, we compute the gradient (e) of the NCC cost map, in which we find the largest candidate set of regions with smoothly varying gradients (f). This region is discretised for computational efficiency over a sampling grid (g), which adapts to fit the boundaries of the region with the aim of reducing extrapolation error. The nodes selected are used to fit an interpolating spline in order to create a smoothly varying sub-pixel flow field (h,i).

The choice of the interpolating spline is not critical to the overall performance of the system; we employ Thin Plate Splines (TPS) due to their well-defined behaviour regarding surface continuity and tendency not to introduce sharp creases. TPSs explicitly minimise the energy:

$$E(\tau) = \sum_{i=1}^N NCC(I(\mathbf{x}_i), I'(\tau(\mathbf{x}_i))) + \alpha_\tau E_{bend}(\tau) \quad (5)$$

where N is the number of samples in the grid, NCC is the normalised cross-correlation matching cost (adjusted from its usual definition to vary between 0 for positive and 1 for negative correlation), α_τ is a smoothing parameter and E_{bend} is the bending energy:

$$E_{bend}(\tau) = \iint \left(\frac{\partial^2 \tau(\mathbf{x})}{\partial \mathbf{x}^2} \right)^2 d\mathbf{x} \quad (6)$$

In all experiments, we set α_τ equal to the average of the diagonal entries of the matrix A , with $A + \alpha_\tau \mathbb{1}(m)$ the coefficient matrix of the linear system for the m coefficients of the spline to be determined. This ensures a balance between the freely bending least squares solution and the constrained TPS energy and is found to work well in practice. For the NCC sampling pattern, we adopt a circular pattern of 1,025 uniformly sampled points. We oversample the pixel grid in order to reduce aliasing errors; the choice of the absolute pattern size is problem specific, but was set to a disk of radius 10 pixels for all our experiments. The same value was used for the resolution of our TPS sampling grid.

3.2. Photometric Matching

From the previous step, we obtain a sub-pixel displacement map τ for the registration of the input image pair. We next account for illumination variation between them.

As with geometric registration, bias and scale variations in colour can occur for several, independent reasons such as differences in ambient illumination, camera transfer functions or surface geometry. Without resorting to scene or image priors, we again model the variation in bias as a continuous per-channel TPS, sampling the difference between the registered images over the same nodes as in fig. 2(g).

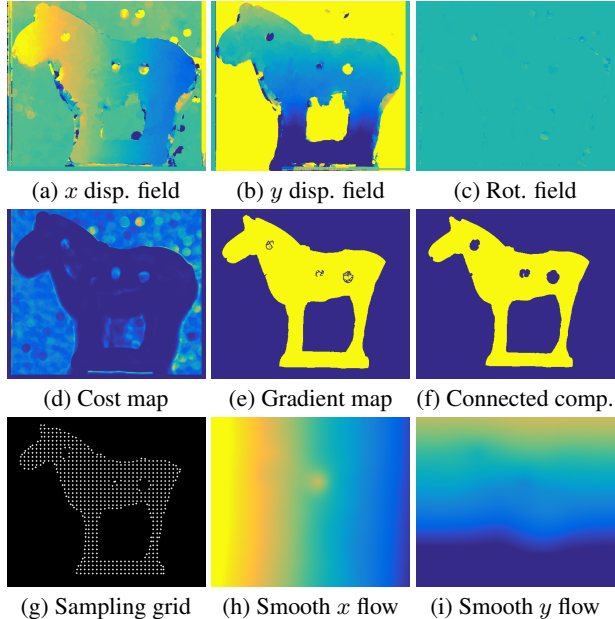


Figure 2: Intermediate stages of geometric registration. See Section 3.1 for further details.

Taking as an example the red channel r , the TPS energy of the bias of that channel, b_r , may be expressed:

$$E(b_r) = \sum_{i=1}^N \left\| G_\sigma(r(\mathbf{x}_i)) - (r'(\tau(\mathbf{x}_i))) - b_r(\mathbf{x}_i) \right\|^2 + \alpha_{col} E_{bend}(b_r) \quad (7)$$

where $r = r_c^t$ is the value of the red channel of image I and G_σ is a Gaussian filter with standard deviation σ , applied on I to compensate for the smoothing caused by interpolating $I'(\tau(\mathbf{x}))$. We set σ to 0.5 pixels and assign α_{col} with the same method as α_τ in all experiments.

Photometric registration does not explicitly account for nuisances like cast shadows and highlights, but can model them if they conform to the local smoothness assumption (similarly, geometric registration does not account for certain image variations like foreshortening and perspective, but can model them). Note that in our approach we only explicitly address bias (and not scale) illumination variation. A simple approach to do so could be to capture it explicitly could be to use a scale illumination TPS in sequence with the existing bias illumination TPS. We will make this clear in the final version.

Fig. 1(d) illustrates the resulting photometric bias that minimises the colour difference between the registered input couple, shown in fig. 1(b) and 1(e).

3.3. Gradient Compensation

The major source of differences left after geometric and photometric alignment are due to sharp gradients in the images. The magnitude of gradients can vary depending on the camera position and transfer function. Different lenses or zoom levels can make gray-level transitions non-uniform across images. The procedures described in subsection 3.1 and 3.2 may also affect gradient magnitudes, resulting in the appearance of artificial differences around sharp edges.

We perform a gradient compensation step to minimise gradient mismatch by applying appropriate low-pass filtering to one of the input images. The optimal amount is selected by energy minimisation with gradient descent, where the energy is calculated as the sum of squared gradient differences computed over the set of candidate regions, \mathcal{S} (fig. 2(f)):

$$E(\sigma) = \sum_{\mathbf{x}_i \in \mathcal{S}} \left\| \frac{\partial G_{\max(0,\sigma)}(I)}{\partial \mathbf{x}} - \frac{\partial G_{\max(0,\sigma)}(J')}{\partial \mathbf{x}} \right\|^2 \quad (8)$$

where $J'(\mathbf{x}_i) = I'(\tau(\mathbf{x}_i)) + b(\mathbf{x}_i)$ is the geometrically and photometrically corrected image, $|\sigma|$ is the standard deviation of the Gaussian blurring kernel and $sign(\sigma)$ determines which image to blur.

3.4. Change Detection

Finally, we output a change map by simple absolute differencing with the final corrected image from its partner:

$$m(\mathbf{x}) = |G_\sigma(I(\mathbf{x})) - G_\sigma(J'(\mathbf{x}))| \quad (9)$$

3.5. Implementation

The bottleneck of our approach is the correspondence computation; to accelerate this we implement PatchMatch in CUDA and achieve further gains using the jump flooding technique of [15]. Our final implementation takes approximately 0.8 seconds for an NVIDIA Titan X GPU on a pair of 720×515 px images. The main parameter affecting the computational complexity is the conservative size of the sampling pattern (1,025 sample points over a 10 px radius circle) used for geometric registration, chosen in order to keep the same parameters across all experiments for exposition of a general working method. For time-critical applications, reducing the density of the sampling pattern (for both PatchMatch and spline fitting) would accelerate the method with little or no loss in performance, depending on the dataset. PatchMatch could be further accelerated by initialisation using sparse feature matches (rather than random). This is possible in applications where feature matches have already been computed during an earlier approximate registration step, which is true for both datasets examined [3, 11].

4. Experiments

4.1. Performance Analysis Using Control Dataset

We evaluate our method on a real-world dataset taken in a controlled environment. Using this dataset, we perform a qualitative performance breakdown of our method to justify the inclusion of each of the three steps - geometric registration, photometric matching and gradient compensation - in our approach. Fig. 3 illustrates a set of results from this analysis. The progressive improvement in change map estimation with each additional step of our method is evident. Even after geometric corrections are applied, changes can be very hard to detect in certain scenarios such as sequence S3, becoming more evident only after photometric correction is applied, and even more so once high frequency artefacts are attenuated via the gradient compensation step.

The image pairs here were chosen to reflect the possible application of the method to useful tasks. In sequence S1, PM-TPS is successful in registering the images despite the relative lack of skin texture, and providing a highly accurate measurement of the growth of a dark mole. Sequence S2 illustrates change detection in a scenario where it is difficult to determine scene change by eye due to the presence of strongly distracting texture. Finally, sequence S4, which shows a strand of hair against the blank page of a book, illustrates that it is possible to precisely detect, for example, the growth of a hairline crack in an environment with limited texture. Such a scenario may be of particular interest to the structural health monitoring community [16], where the rate of crack evolution is often a more informative indicator of structural health than crack detection.

4.2. Evaluation on Real-world Datasets

We next demonstrate the performance of our method on two recent, real-world change detection datasets: the structural inspection dataset of [11] and the artefact inspection dataset of [3].

4.2.1 Structural Inspection Dataset

The large-scale dataset of [11] contains 45 realistic cracks and other subtle changes spread over 61 high-resolution image mosaics, each of typical size $34,000 \times 9,000$ pixels, taken inside a concrete tunnel. The dataset breaks down into \mathcal{D}_{short} and \mathcal{D}_{long} , where the latter contains images captured over a longer time period and hence more severe visual change. Due to the scale of the dataset, we subdivide each mosaic into overlapping $1,024 \times 1,024$ pixel patches to be processed independently, piecing them back together during evaluation.

Using the same evaluation methodology as [11], we present a quantitative comparison versus their CNN-based approach in Fig. 4. While our method performs slightly

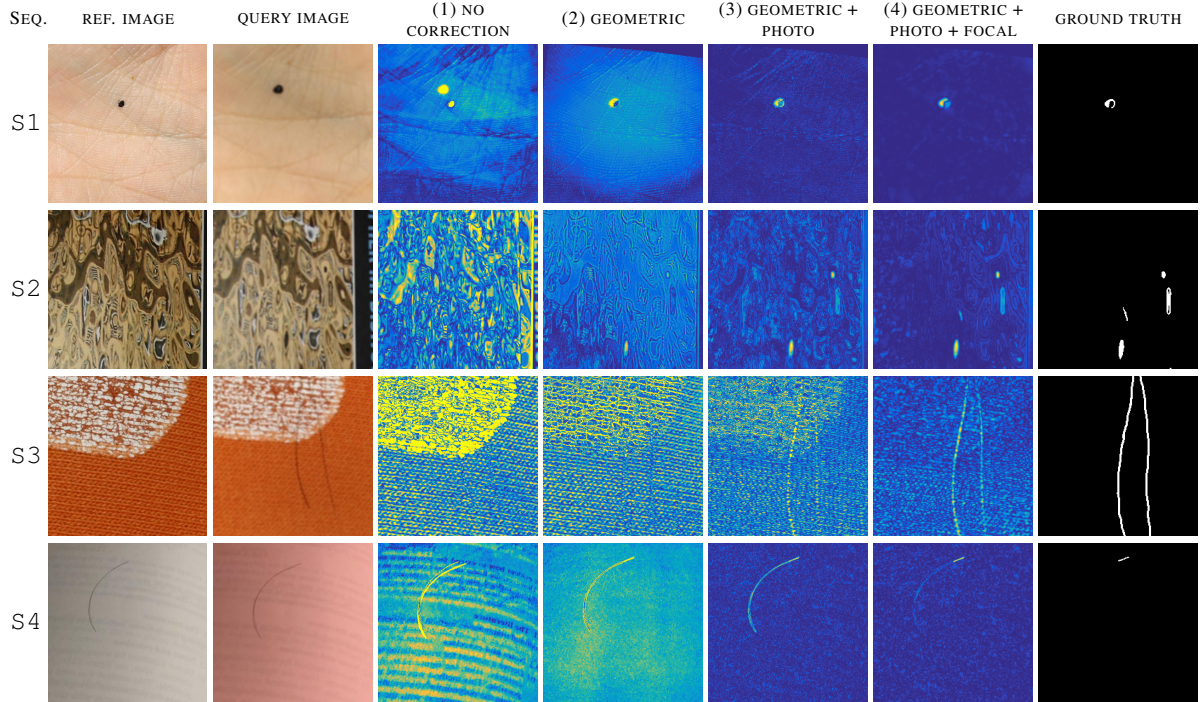


Figure 3: Illustration on control dataset examples showing the effect of sequential stages of our method, from (1) the raw image difference between reference and query image, to (2) difference after geometric compensation, to (3) difference after geometric and photometric compensation, to (4) final output following gradient compensation. All three corrective steps are necessary to produce a precise change map. The sequence pairs show, in row order: human skin; strong texture which distracts from the change; fabric with loose thread; the change in length of a strand of hair against the blank page of a book.

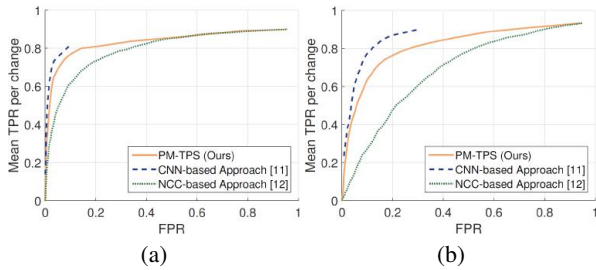


Figure 4: Change detection performance on a structural inspection dataset from [11] on (a) \mathcal{D}_{short} (b) \mathcal{D}_{long} . Our method performs comparably with the CNN approach of [11], despite being entirely deterministic, not requiring any training (time or data), and the significant disadvantage of not being taught what constitutes a change of interest as described in section 4.2.

worse in both \mathcal{D}_{short} and \mathcal{D}_{long} datasets, this is expected since the provided ground truth only labels changes which are considered “interesting” to the inspector, while other changes are unlabelled. In this scenario, a learning-based

method such as [11] has an advantage, since it can be taught to detect the former and be invariant to the latter. The effect is particularly strong in \mathcal{D}_{long} , where many real but uninteresting changes are detected by our method but are unlabelled and therefore penalised by the performance metric. Our method being entirely physically based, does not try to estimate the importance of change from its appearance. One possibility for future work could be to combine the two methods by, for example, using a CNN-based similarity metric to perform the final change map estimation.

Fig. 5 shows that our method provides a significant qualitative improvement. The compared approaches of [12] and [11] both suffer from being patch-based, and therefore incapable of producing precise change masks. In addition, since our method is not learning-based, it is not susceptible to design factors such as training set sizes and model sizes, which may be the cause of errors in the CNN change mask, in row 2 for example. The accuracy of PM-TPS suggests that it may be a useful tool to monitor and specify the evolution in crack width over time - an application of great interest to structural engineers.

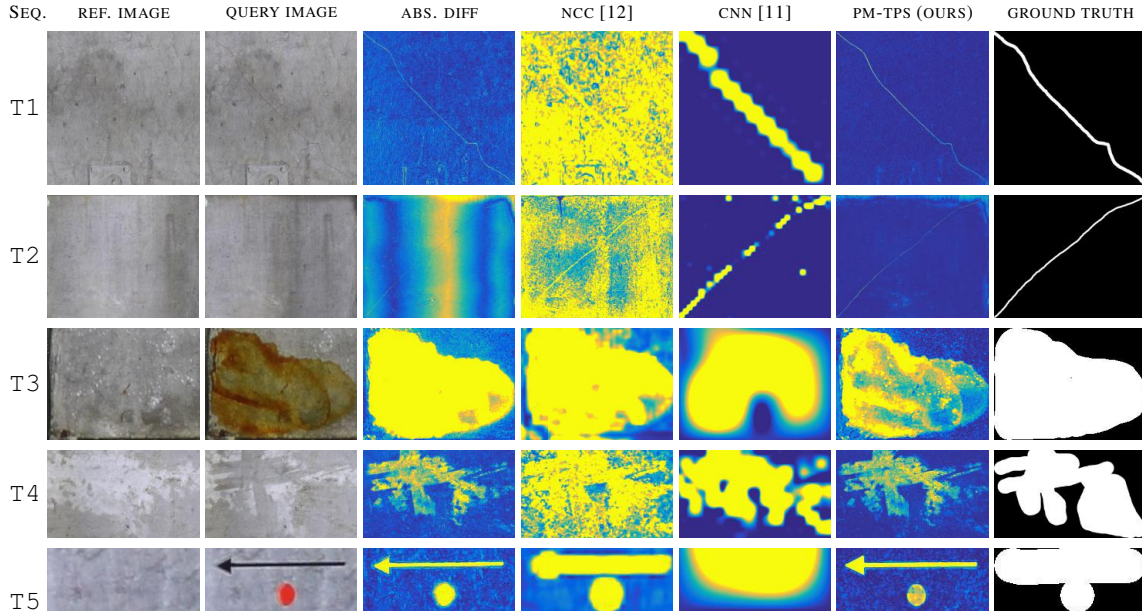


Figure 5: Example changes from the tunnel dataset. Each row represents a different change. Columns from left to right: reference image; query image; ground truth change mask; absolute difference image; NCC-based difference image [12]; CNN-based difference image [11]; PM-TPS difference image (our method). Note that images have been rescaled for the purpose of illustration, so the granularity of the ground truth labelling appears to vary greatly. Note that some of the changes are quite fine and are best viewed on screen.

4.2.2 Artefact Monitoring Dataset

Finally, we examine the recently released artefact condition monitoring datasets of [3]: the Statues dataset, D_s , containing images taken of 4 small statues and artefacts; the Palace dataset, D_p , of two outdoor sculpture scenes; and the Briquettes dataset, D_b , a sequence of 10 sets of laboratory images simulating the deterioration of a mural over time. Since each image set provides 7 approximately aligned image pairs – 1 environment lighting (EL) and 6 directional side lighting (DSL) pairs – we display results from our method using a single pair (EL), 3 pairs (EL + 2 DSL), and 7 (EL + 6 DSL) pairs. We combine results across pairs for a given set using a simple approach: we normalise the change map per image pair to sum to 1 and average the results. Note also that for each dataset, we report quantitative figures with a fixed change threshold over the entire dataset. Designing an adaptive threshold mechanism is expected to improve our results further and is a topic for future work. As in [4], we use the same 7 metrics for comparison: F1-measure (F1), recall (Re), precision (Pr), specificity (Sp), false positive rate (FPR), false negative rate (FNR) and percentage of wrong classifications (PWC).

Table 1 and Fig. 6 illustrate our quantitative and qualitative performance respectively. Our method outperforms

competing approaches on the Statues dataset for F1 score, precision and percentage of wrong classification (first three rows of Fig. 6), and performs close to the state-of-the-art on the remaining two datasets. Inspecting the change masks reveals that in the majority of cases, our method produces less noisy masks than [3], with a few failure cases such as that shown in the bottom row of Fig. 6. This occurs with greater frequency in the Sculpture and Briquette datasets, where local smoothness assumptions may be violated more frequently. However, we believe that an adaptive change threshold will boost our method’s performance noticeably. A final point of note is that our method requires only a single approximately aligned image pair. In Table 1(c), the addition of extra pairs makes performance slightly worse, since the additional available (DSL) images are taken under quite diverse illumination settings. The speed of our approach in terms of both image acquisition and processing time makes it attractive for any inspection task requiring large volumes of data.

5. Discussion

In this work we have introduced PM-TPS, an efficient, effective approach to precise change detection over smooth surfaces. We have illustrated the performance of our

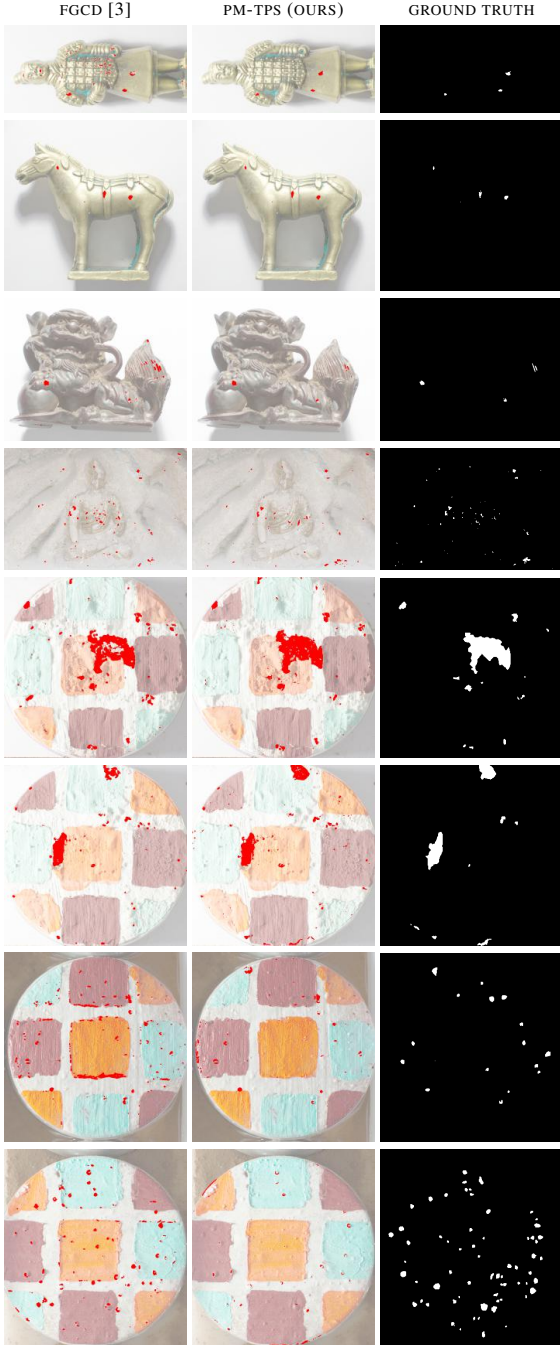


Figure 6: Qualitative comparison with [3]. We superimpose changes in red over whitened versions of the original images.

method both visually and quantitatively over a variety of applications and highlighted in particular its potential use for the structural inspection of engineering structures and the condition monitoring of high-value artefacts. One limitation of our approach is that the assumption of smooth surfaces restricts the number of suitable applications. In the

Table 1: Performance comparison with alternative approaches on three artefact monitoring datasets from [3]. Unlike the method of [3] which uses 7 image pairs per test, we show our results using 1, 3 and 7 pairs of images.

(a) Average results on Statues dataset D_s

Method	F1	Re	Pr	Sp	FPR	FNR	PWC
FGCD (D&T) [3]	0.53	0.78	0.43	1.00	0.00	0.22	0.28
FGCD (SVM) [3]	0.51	0.86	0.39	1.00	0.00	0.14	0.29
Ours (EL)	0.65	0.66	0.76	1.00	0.00	0.34	0.08
Ours (EL+2DSL)	0.76	0.74	0.79	1.00	0.00	0.26	0.06
Ours (EL+6DSL)	0.79	0.76	0.84	1.00	0.00	0.24	0.05

(b) Average results on Briquettes dataset D_b

Method	F1	Re	Pr	Sp	FPR	FNR	PWC
FGCD (D&T) [3]	0.45	0.40	0.56	1.00	0.00	0.60	1.23
FGCD (SVM) [3]	0.53	0.62	0.48	0.99	0.01	0.38	1.41
Ours (EL)	0.40	0.49	0.39	0.99	0.01	0.51	1.40
Ours (EL+2DSL)	0.45	0.56	0.43	0.99	0.01	0.44	1.24
Ours (EL+6DSL)	0.45	0.56	0.45	0.99	0.01	0.44	1.13

(c) Average results on Palace dataset D_p

Method	F1	Re	Pr	Sp	FPR	FNR	PWC
FGCD (D&T) [3]	0.34	0.28	0.52	1.00	0.00	0.72	0.92
FGCD (SVM) [3]	0.51	0.53	0.47	0.99	0.01	0.47	1.02
Ours (EL)	0.37	0.38	0.66	1.00	0.00	0.62	0.96
Ours (EL+2DSL)	0.32	0.33	0.68	1.00	0.00	0.67	0.98
Ours (EL+6DSL)	0.28	0.24	0.76	1.00	0.00	0.76	0.97

future, relaxing this assumption and investigating joint geometric and photometric registration may help to improve precision in harder cases.

Acknowledgements The authors would like to thank Toshiba Research Europe for funding the project and the conference reviewers for their helpful feedback.

References

- [1] C. Barnes, E. Shechtman, D. Goldman, and A. Finkelstein. The generalized PatchMatch correspondence algorithm. In *ECCV*, pages 29–43. Springer, 2010.
- [2] O. Delaunoy, N. Gracias, and R. Garcia. Towards detecting changes in underwater image sequences. In *OCEANS 2008-MTS/IEEE Kobe Techno-Ocean*, 2008.
- [3] W. Feng, F. Tian, Q. Zhang, N. Zhang, L. Wan, and J. Sun. Fine-grained change detection of misaligned scenes with varied illuminations. In *ICCV*, 2015.
- [4] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. Changedetection. net: A new change detection benchmark dataset. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 1–8. IEEE, 2012.
- [5] D. Lu, P. Mausel, E. Brondizio, and E. Moran. Change detection techniques. *International Journal of Remote Sensing*, 25(12):2365–2407, June 2004.

- [6] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: a systematic survey. *IEEE Trans. Image Process.*, 14(3):294–307, 2005.
- [7] K. Sakurada and T. Okatani. Change detection from a street image pair using CNN features and superpixel segmentation. In *BMVC*, 2015.
- [8] K. Sakurada, T. Okatani, and K. Deguchi. Detecting changes in 3D structure of a scene from multi-view images captured by a vehicle-mounted camera. In *CVPR*, 2013.
- [9] L. Shi, W. Xiong, and B. Funt. Illumination estimation via thin-plate spline interpolation. *J. Opt. Soc. Am. A*, 28(5):940–948, May 2011.
- [10] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin. Sub-sense: A universal change detection method with local adaptive sensitivity. *Image Processing, IEEE Transactions on*, 24(1):359–373, 2015.
- [11] S. Stent, R. Gherardi, B. Stenger, and R. Cipolla. Detecting change for multi-view, long-term surface inspection. In *BMVC*, 2015.
- [12] S. Stent, R. Gherardi, B. Stenger, K. Soga, and R. Cipolla. Visual change detection on tunnel linings. *Machine Vision and Applications*, pages 1–12, 2014.
- [13] A. Taneja, L. Ballan, and M. Pollefeys. Image based detection of geometric changes in urban environments. In *ICCV*, pages 2336–2343, 2011.
- [14] J. Vogel, A. Dului, Y. Oyamada, J. Gardiazabal, T. Lasser, M. Ziai, R. Hein, and N. Navab. Towards robust identification and tracking of nevi in sparse photographic time series. In *SPIE '14*, 2014.
- [15] P. Yu, X. Yang, and L. Chen. Parallel-friendly patch match based on jump flooding. In *Advances on Digital Television and Wireless Multimedia Communications*, pages 15–21. Springer, 2012.
- [16] Z. Zhu, S. German, and I. Brilakis. Visual retrieval of concrete crack properties for automated post-earthquake structural safety evaluation. *Automation in Construction*, 20(7):874–883, 2011.