

WIDE-BASELINE IMAGE CHANGE DETECTION

Ziggy Jones* Mike Brookes* Pier Luigi Dragotti* David Benton †

* Department of Electrical and Electronic Engineering, Imperial College, London, UK

†L-3 TRL Technology, Tewkesbury, UK

ABSTRACT

We present a fully automated method for the detection of changes within a scene between a reference and a sample image whose viewing angles differ by up to 30° . We also describe an extension to the SIFT technique that allows extracted feature points to be matched over wider viewing angles. Matched correspondences between reference and sample images are used to construct a Delaunay triangulation and changes are detected by comparing triangles after affine compensation using a dense SIFT metric. False positives are reduced by using a novel technique introduced as local plane matching (LPM) to match mean-shift segments in unmatched areas using the homographies of local planes to compensate for perspective distortions. The method is shown to achieve pixel-level equal error rates of 5% at a 10° azimuth view angle difference.

Index Terms— Change Detection, SIFT, Wide Baseline, Affine Compensation, Feature Points, Segmentation, Image Matching, Local Features

1. INTRODUCTION

This paper discusses the problem of wide-baseline image change detection and presents a method for identifying areas that have changed when a sample image is compared with a reference image taken from a different viewpoint. Change detection is important in both civilian and defence applications for example, the analysis of surveillance images from unmanned aerial vehicles (UAVs). The capability of automated change detection allows operator attention to be focused on regions of possible change. Our goal is to identify any regions of the reference image that have changed due to the addition, removal or movement of one or more objects.

If the observed scene is unchanged, images taken from widely spaced viewpoints may appear different due to (a) the occlusion or disocclusion of distant objects by nearer objects, (b) the self-occlusion or disocclusion of objects by themselves, (c) perspective distortions arising from the change in viewpoint. Occlusion occurs when surfaces in view in one image are hidden behind another object in a different image while disocclusion is where surfaces are revealed. In addition, images taken at different times will include small

differences that are normally unimportant such as those arising from lighting changes and slight movements such as leaves rustling in the wind.

The appearance of any planar region of the scene in two different images is related by a 2D projective transformation, or homography [7]. The corresponding inverse homography can therefore be used to compensate for its change in appearance between images. A general 2D homography has eight degrees of freedom and is uniquely defined by determining four matching points, or correspondences, between the images. If the depth variation across the planar region is small compared with the distance from the camera, the 2D homography may be well approximated by an affine transformation which, with only 6 degrees of freedom, is therefore uniquely determined by three correspondences [7].

Attempts have been made to incorporate affine correction into feature points [9, 14, 17]. These approaches tend to produce fewer matches and require a larger computational effort. This paper utilises a novel adaptation of the Scale Invariant Feature Transform (SIFT) [12] which incorporates affine correction into the algorithm in a way that allows for different levels of distortion at each point. A similar density of feature points are matched at large angles as SIFT achieves at small angles with only a small increase in computational load at the point that the sample image becomes available.

In this paper we extend an approach to change detection that was introduced in [10]. Correspondences between the reference and sample images are used to partition them consistently into small triangular regions and we assume that the depth variation within the observed scene is sufficiently smooth that the conditions for the affine approximation hold for most of the regions. Corresponding regions are then matched between images based on a local pixel intensity gradient metric; any regions that match are classified as change-free. The remaining regions, which contain potential changes are then segmented into patches on the basis of the image content. Patches that are consistent with a local homography determined from adjacent triangular regions are classified as change-free while the remaining patches are classified as containing changes.

2. BACKGROUND

The majority of change detection algorithms described in the literature address the detection of change in images taken from overhead such as those captured from satellites or surveying aerial flights [16, 15]. In these cases the reference and sample images can be aligned by a translation, scaling and/or rotation. It is important to achieve precise registration since misregistration has been found to produce a substantial degradation in the accuracy of remotely sensed change detection. For example registration errors as small as 0.2 pixels in standard satellite imagery introduced changed detection errors of up to 10% compared with perfectly registered images [3]. This means that it is infeasible to attempt wide-baseline change detection by means of image registration and direct comparison.

An alternative approach to wide baseline change detection is to use additional sensors to capture depth information. LASERS can be used in systems such as in light detection and ranging (LIDAR) [1, 6] or airborne laser scanning (ALS) [8, 18, 13]. By using the depth information the 3D position of each pixel can be found in each image which allows for the comparison of the pixel information without the ambiguities introduced by projective distortion or occlusion.

This paper presents an approach for wide-baseline change detection that does not require range data or aligned images, allowing for the use of standard imaging equipment. This allows for the use of a larger range of input images in a larger range of scenarios which increases the scope of change detection.

3. AFFINE COMPENSATED SIFT

In order to compare two unregistered images, it is necessary to locate points in the two images that correspond to the same positions in the scene. A widely used technique for finding such correspondences is SIFT [12] which both identifies distinctive points within an image and also provides a mechanism for matching them between images using a descriptor based on the local pixel intensity gradient.

A modified SIFT algorithm known as ‘Affine SIFT’ (ASIFT) was introduced in [10]. In ASIFT a number of anisotropic scaling factors are applied to the reference image in a range of orientations to create a set of modified reference images. SIFT points and descriptors are collected from the original and modified images to form an extended set of dictionaries. These dictionaries can be created offline before the sample image is available. A SIFT descriptor from the sample image is matched to each of the dictionaries using the SIFT matching procedure. If more than one match is found for a sample point across the reference dictionaries the match with the smallest matching distance is used. In some applications the positions and orientations of the cameras are accurately known. In this case, the appropriate scaling

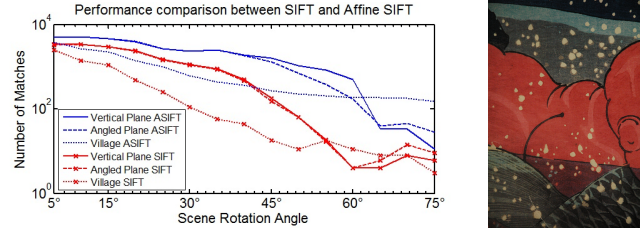


Fig. 1. Left: Effect of affine compensation of SIFT matching. Matching performance improves with affine compensation of SIFT in the planar image and more realistic village scene. Right: Planar image used for evaluation.

orientation for each feature point is determined by the corresponding epipolar line and the number of dictionary sets can be correspondingly reduced.

3.1. Affine SIFT performance

Figure 1 illustrates, on a logarithmic scale, the performance gain of ASIFT over SIFT in three test image sets. The red lines with crosses show the number of matches obtained by SIFT as a function of scene rotation angle. The blue lines show the corresponding performance of the ASIFT algorithm which always exceeds that of SIFT. The solid lines show the performance when the planar image shown in Fig. 1 is rotated around a vertical axis. We see that the number of matches obtained by SIFT falls rapidly for rotation angles above 40° whereas ASIFT maintains a good performance for angles up to 60°. The dashed lines show the performance when the axis of rotation is at a worst-case angle relative to the compensation orientations used in ASIFT. We see that although the performance of ASIFT degrades slightly, it still comfortably exceeds that of SIFT for rotation angles above 30°. The dotted lines show the performance on the more realistic model village scene that is shown in Fig. 2(b). This scene includes surfaces at many different orientations and the number of matches falls off more rapidly with rotation angle. In all three cases ASIFT significantly outperforms SIFT and achieves good performance at angles up to 50°.

4. IMAGE CHANGE DETECTION SYSTEM

This section describes an image change detection system which produces consistent segmentation of the reference and sample images using ASIFT points to construct a Delaunay triangulation in combination with mean-shift [2] segmentation. The first stage of the algorithm compares corresponding pairs of triangles between the two images and flags any triangles that include a scene-change. The second stage of the algorithm refines this initial classification by segmenting the flagged triangles into patches on the basis of image content. Patches from the reference image are mapped onto the sam-

ple image and classified individually. The mapping between reference and sample images is a homography determined from adjacent triangular regions. Each step of the system is described in more detail below.

4.1. Delaunay Triangulation Segmentation

The first step is to partition both the reference and sample images into consistent triangular regions. A set of matched ASIFT points described in Sec. 3 are used to construct a Delaunay triangulation [4, 11] to partition the reference and sample images. A initial set of high matching threshold feature points are used to find the fundamental matrix between the two images using a RANSAC approach [5]. A second pass collects a larger set of correspondences by using a lower matching threshold. Errors are reduced by checking for consistency with the fundamental matrix. Delaunay triangulation is performed on the matches in the reference image and a consistent triangulation applied to the sample image. In rare cases, triangles in the sample image will overlap due either to a correspondence error or to a large variation in depth within a triangle; where this occurs, the corresponding feature points are removed.

The results of the partitioning are illustrated in Fig. 2(b) and 2(c) where we see that it results in dense triangles in most regions of the image. Where a change has occurred, such as the missing building on the left of the image, no matches are found in the changed region, so the triangles are larger.

4.2. Stage 1: Region matching with dense SIFT

The above segmentation approach produces segment areas within the two images that approximately coincide. The three corners of each triangle allow for affine compensation by mapping the triangle in the reference image to the triangle in the sample image and interpolating pixel values. If the scene within a triangle in the reference and sample images is approximately planar a comparison of the areas after affine compensation can determine if they contain changes.

Even if the scene is unchanged, corresponding triangles will not match perfectly for a number of reasons. First, the localisation errors of the ASIFT feature points result in small misalignments of the compensated triangles. Second, the inaccuracies from using an affine approximation to represent the homography between the segments will also produce additional misalignments. Third, the scene area within a triangle may not be exactly planar and so a single transformation will not be correct for the entire triangle. Because of these errors, it is not possible to match the segments using techniques, such as autocorrelation, that require near perfect alignment.

In our algorithm we use the SIFT descriptor to compare each mapped triangle from the reference image with the corresponding triangle from the sample image because it is robust to lighting changes and also to small pixel misalign-

ments. SIFT descriptors with a 4 pixel window size are computed for both triangles on a dense grid with approximately 2-pixel spacing. To detect changes, the euclidean squared distance between corresponding descriptors is found. The regions are said to match if the number of descriptor distances above a first threshold is below a second threshold. The first threshold is set high so that only significant differences between the images are highlighted. The second threshold is set so that triangular regions with a significant number of non-matching 4 pixel descriptor windows are marked as changed.

4.3. Stage 2: Local plane matching

The procedure described in Sec. 4.2 identifies the triangular regions that potentially include scene changes. However, the procedure may include a substantial number of pixels that are falsely classified as having changed for two reasons. First, a genuine change within a triangular region will normally only occupy a portion, perhaps a small portion, of the entire region. Second, if the scene within a triangular region is very far from planar, then perspective effects will prevent a match between the reference and sample images even if no scene change is present. In the second stage of our algorithm, we divide the triangles into smaller patches on the basis of the image content and assess whether or not each of the patches represents a genuine change in the scene.

The first step is to form groups of adjacent triangular regions that are approximately coplanar in the scene. This is done using a growing cluster algorithm. The three correspondences that form the vertices of a triangular region are used to define an affine mapping between the reference and sample images. This mapping is then applied to the vertices of each adjacent triangular region in the reference image. If their correspondences in the sample image lie within a defined threshold distance of the mapped reference-image vertices, the adjacent triangle is added to the current cluster. This procedure is then iterated and results in clusters of contiguous triangles that are approximately coplanar.

A modified RANSAC [5] procedure is now used to find a homography for each local plane. The random samples used in the procedure used to define a candidate homography are constrained to exclude image points that are very close to each other. This constraint is introduced to increase the stability of the resulting homography.

A segmentation algorithm such as mean-shift [2] is now used to subdivide triangles identified by the method in Sec. 4.2 as potential changes. If the triangle is within or adjacent to a clustered plane the homography of the plane is applied to the segment which is then matched against the corresponding region of the sample image. The matching is conducted using the dense SIFT matching method used in Sec. 4.2, if the two regions match this the change classification of this segment is marked as negative.

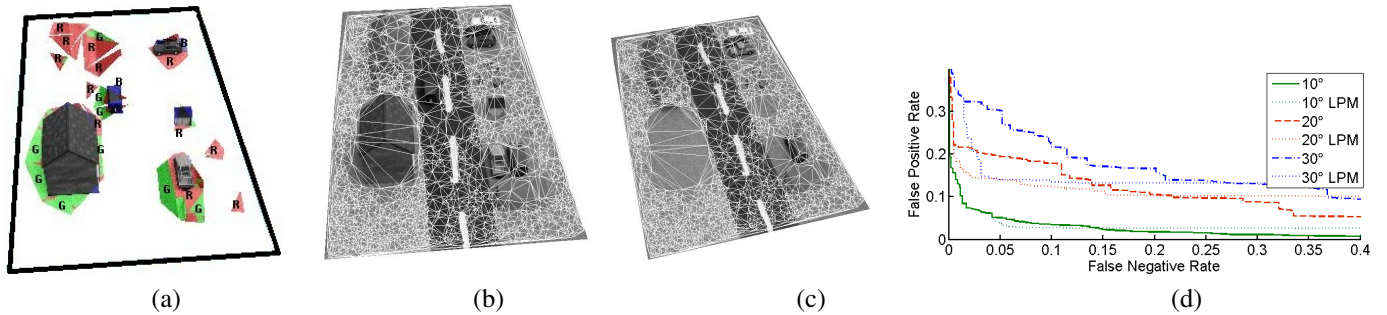


Fig. 2. Delaunay segments formed using ASIFT points on images at a 10° with scene changes visible shown in reference image (b) and sample image (c). Change mask shown in (a) with true negatives highlighted in white, false positives and negatives in red (labelled ‘R’) and blue (labelled ‘B’) respectively and false positives that are re-classified as negatives using the technique described in Sec. 4.3 in green (labelled ‘G’). Areas of change incorrectly re-classified as not containing change are too small to be visible. (d) shows the ROC curve of pixel classification performance for scene rotations of 10° , 20° and 30° using the model scene shown in Fig. 2. Benefit gained from local plane matching marked as LPM shows the performance boost gained from the technique in Sec. 4.3.

5. PERFORMANCE

In this section, we illustrate the algorithm performance using a model village scene for which reference and sample images are shown in Fig. 2(b) and 2(c) respectively. There are five changes between the two images: two buildings have been removed (one from bottom left and one from middle right), two jeeps have been removed (one from the road and one from bottom right) and a car at top right has been replaced by a jeep.

The first stage of the algorithm classifies all the pixels within a given triangular region as changed (positive) or unchanged (negative) and these decisions are compared with the ground truth. The second stage then reclassifies some of the positive pixels as negative. In Fig. 2(a) false positives from the first stage are labelled G (shaded green) if they are correctly reclassified after the second stage or R (shaded red) otherwise. False negatives from the first stage are labelled B (shaded blue) and are never reclassified. True negatives are shown as white. We see that all the true changes have been correctly identified and that the false negative triangles occur only at the boundaries of true changes. All the false positives that remain after the second stage of the algorithm arise because the scene is far from planar within the affected triangle.

The majority of triangles overlaying areas of change have been detected and no changes have been missed completely. All of the false positives are caused by the triangle not lying on the plane of the scene.

The ROC curve showing system performance is shown in Fig. 2(d) and is based on a false positive and false negative rate judged per pixel in the reference image. As the algorithm functions on a triangle region basis, the false positive rate includes any unchanged pixels that lie within the triangles correctly identified as changed. The local plane matching

described in Sec. 4.3 reduces these false positives as shown in Fig. 2 by the areas in green marked ‘G’. 43% of the false positives and a large majority of the false positives on the ground level are re-classified as negative. No significant areas of change have been marked as negative (these are too small to be visible in Fig. 2). The effect can also be seen in the ROC curves labelled LPM in Fig. 2.

As Fig. 2(d) shows, the performance of the system is dependent on the angle between the sample and reference images and the number of false positives increases at higher angles. The reason for this is partly a reduction in the density of feature point matches and partly an increase in occlusions. The system works best on scenes with low depth variation where occlusions occur less often.

6. SUMMARY

We have presented a method for detecting and locating changes in a scene from two images taken from widely separated viewpoints. The method partitions the reference image into triangular regions and, in a first stage, identifies the triangles that contain scene changes. A second stage refines this initial classification by detecting any unchanged patches within the triangles identified in the first stage. We have shown that the method performs well when comparing views with an angular separation of up to 30° and that the procedure effectively combines the low false negative rate of the first stage with the low false positive rate of the second.

Overall the approach allows for change detection using images that are not registered and without the use of additional input data such as depth information.

7. REFERENCES

- [1] CHEN, L. -C., HUANG, C. -Y., TEO, T. -A. 2012 Multi-type change detection of building models by integrating spatial and spectral information. *Intl J. Remote Sensing*. **33**, 1655–1681.
- [2] COMANICIU, D. AND MEER, P. 2002 Mean shift: a robust approach toward feature space analysis. *IEEE J. Pattern Analysis and Machine Intelligence*. **24**, 603–619.
- [3] DAI, X. 1998 The effects of image misregistration on the accuracy of remotely sensed change detection. *IEEE Trans. on Geoscience and Remote Sensing*. **36**, 1566–1577.
- [4] DELAUNAY, B. 1934 Sur la sphere vide. *Otdelenie Matematicheskikh i Estestvennykh Nauk*. **7**, 793–800.
- [5] FISCHLER, M. A., BOLLES, R. C. 1981 Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*. **24**, 381–395.
- [6] GIRARDEAU-MONTAUT, D., ROUX, M., MARC, R., THIBAUT, G. 2005 Change detection on points cloud data acquired with a ground laser scanner. *Proc. ISPRS Workshop*. **36**, 30–35.
- [7] HARTLEY, R., ZISSERMAN, A. 2004 Multiple view geometry in computer vision. *Cambridge University Press*.
- [8] HEBEL, M., ARENS, M., STILLA, U. 2011 Change detection in urban areas by direct comparison of multiview and multi-temporal ALS data. *Proc. ISPRS*. **6952**, 185–196.
- [9] KADIR, T. AND ZISSERMAN, A. AND BRADY, M. 2004 An affine invariant salient region detector. *Computer Vision – ECCV 2004* **3021**.
- [10] JONES Z., BROOKES M., DRAGOTTI P. L., JAHANGIR M. 2013 Wide-baseline image change detection. *3rd IMA Conference on Mathematics in Defence* **6952**.
- [11] LEE, D.T., SCHACHTER, B. J. 1980 Two algorithms for constructing a delaunay triangulation. *Intl J. Computer and Information Sciences*. **9**, 219–242.
- [12] LOWE, D.G. 2004 Distinctive image features from scale-invariant keypoints. *Intl J. Computer Vision*. **60**, 91–110.
- [13] MATIKAINEN, L., HYYPA, J., AHOKAS, E., MARKELIN, L., KAARTINEN, H. 2010 Automatic detection of changes from laser scanner and aerial image data for updating building maps. *Remote Sensing*. **2**, 1217–1248.
- [14] MIKOLAJCZYK, K. AND SCHMID C. 2004 Scale and affine invariant interest point detectors. *Intl. J. Computer Vision*. **60**, 63–86.
- [15] RADKE, R., J., ANDRA, S., AL-KOFAHI, O., ROYSAM, B. 2005 Image change detection algorithms: a systematic survey. *IEEE Trans. Image Processing*. **14**, 294–307.
- [16] SINGH, A. 1989 Digital change detection techniques using remotely-sensed data. *Intl J. Remote Sensing*. **10**, 989–1003.
- [17] TUYTELAARS, T. AND GOOL, L. V. 2004 Matching widely separated views based on affine invariant regions. *Intl J. Computer Vision*. **59**, 61–85.
- [18] YU, X. 2008 Methods and techniques for forest change detection and growth estimation using airborne laser scanning data. *Ph.D. dissertation, Aalto University [Online]*. Available: <http://urn.fi/URN:ISBN:978-951-711-270-3>.