Wide-Baseline Image Change Detection

By Ziggy Jones* Mike Brookes* Pier-Luigi Dragotti* Mohammed Jahangir**

Department of Electrical and Electronic Engineering, Imperial College, Exhibition Road, 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. The method is shown to achieve pixel-level equal error rates of 5% at a 10° azimuth view angle difference.

1. Introduction

This paper discusses the problem of wide-baseline image change detection and presents a method for identifying areas of a reference image that have changed in a sample image taken from a different viewpoint. Change detection is important in both civilian and defence applications including, for example, the analysis of surveillance images from UAVs. The capability of automated change detection obviates the need for time-consuming manual inspection. In this paper we compare a sample image of a scene with a reference image taken at a different time and from a different viewpoint. 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 appear different due to (a) the occlusion or disocclusion of far objects by nearer objects, (b) the self-occlusion or disocclusion of objects by themselves, (c) projective distortions arising from the change in viewpoint. 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. 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. At differing angles different surfaces of an object come into and out of view giving rise to self occlusion or self disocclusion.

The appearance of a planar region in two different images is related by a 2D projective transformation, or homography. A homography can therefore be used to compensate for the change in appearance of planes between images. A general 2D homography has eight degrees of freedom and can be uniquely defined by finding four matching points, or correspondences, between the images. These 8 degrees of freedom result in a large range of possible distortions when defining a range of possible homographies a plane could undergo. 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 has 6 degrees of freedom can be uniquely specified with only three correspondences Hartley(2004). In this paper, we segment the observed scene into small triangular regions and assume that the conditions for the affine approximation hold for most of the regions.

2. Background

The majority of change detection algorithms discussed in the literature concern detecting change in images taken from overhead such as those captured from satellites or surveying aerial flights Singh(1989), Radke(2005). In these cases the reference and sample images would at most require a translation and/or rotation to align the scenes. Precise registration is important to achieve since misregistration is shown to produce a substantial degradation in the accuracy of remotely sensed change detection. It is shown that, with standard satellite imagery, a registration accuracy of less than 0.2 pixel is required to achieve a change detection error below 10% when compared with perfectly registered images Dai(1998). 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) Chen(2012), Girardeau-Montaut(2005) or airborne laser scanning (ALS) Hebel(2011), Yu(2008), Matikainen(2010). 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 produced by projective distortion or occlusion.

This paper presents an approach for wide-baseline change detection that does not require additional 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 points is the Scale Invariant Feature Transform (SIFT) Lowe(2004) which both identifies distinctive points within an image and provides a mechanism for matching them between images. The first step of the SIFT algorithm locates stable feature points at the centre of regions that are lighter or darker than their surroundings at a particular scale. The second step associates with each feature point a 128-element descriptor that characterizes the texture within a window that is centred on the feature point. The size of this descriptor window is proportional to the scale of the feature point. Each vector is added to a list to form a dictionary of the feature points present in the image. To find matching feature points, descriptors from the sample image are compared with the dictionary entries to find those that are closest in Euclidean distance. The corresponding points are assumed to match if the ratio of the closest distance to the second closest distance exceeds a threshold Lowe(2004). This method enables the reliable matching of points between two images and will typically match thousands of points in a 6M pixel image. Using a higher threshold value reduces the number of matches found but increases their reliability. Although the feature point locations found in the images frequently include small localization errors, typically a few pixels, the descriptors are constructed in a way that is tolerant to these.

The SIFT algorithm is robust to translation, rotation, scale changes and lighting changes but not to changes in camera viewpoint. In this paper, we therefore extend the algorithm to make it less sensitive to viewpoint changes. Robustness of a feature point to changes in the angle of the camera line is critical for its use in wide baseline image change detection. When a planar surface is viewed by a camera whose optical axis makes an angle θ with the surface normal, the image of the plane is foreshortened by a factor $\cos\theta$. If the camera viewpoint is moved, the value of θ, and hence the amount of foreshortening, will change and this may prevent the matching of SIFT descriptors from feature point on the surface. To overcome this, we use 'affine SIFT' (ASIFT), a modified SIFT algorithm in which a range of foreshortening factors is applied to the reference in a range of orientations to create a set of modified reference images in addition to the original. SIFT points and descriptors are collected from the set of images from the entire

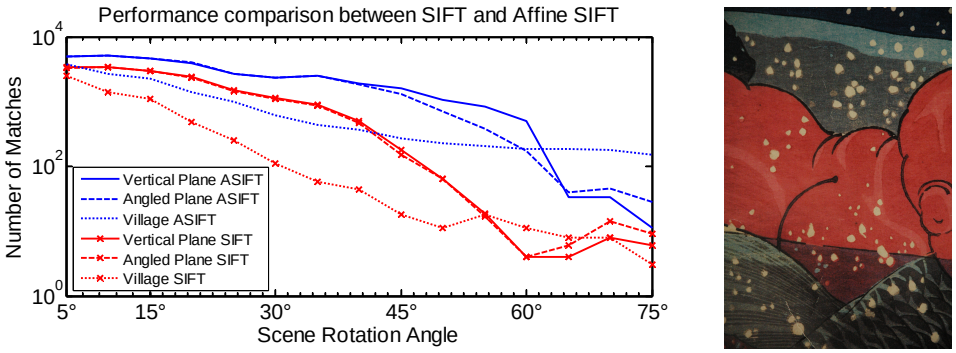


FIGURE 1. Left: Effect of affine compensation of SIFT matching. Right: Image of textured planar image used to evaluate ASIFT.

set of images to form an extended set of dictionaries. The formation of these dictionaries can be carried out offline before the sample image is available. A SIFT descriptor from the sample image is matched to each of the dictionaries as in the SIFT algorithm. 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 many applications the positions and orientations of the cameras will be accurately known. In this case, the orientation of foreshortening for each feature point can be calculated in advance and the number of dictionary sets can be correspondingly reduced.

3.1. *Results*

In the experiments below, we form the extended set of dictionaries by applying four different foreshortening factors in each of four orientations to give a total of sixteen modified versions of the reference image in addition to the original. The four orientations are at $\{0°, 45°, 90°, 135°\}$ relative to the horizontal and the foreshortening factors equal $\cos\theta$ where θ is in $\{20°, 40°, 60°, 80°\}$.

To evaluate the effectiveness of the compensation on surfaces that have undergone a projective transformation, we rotated the textured planar image shown in Figure 1 about an axis orthogonal to the camera line. For the reference image, the plane is orthogonal to the camera line and for the sample images, the plane is rotated by up to 75° in steps of 5°. The experiment is conducted twice: first with the rotation direction aligned vertically and second with the rotation aligned at 22.5° to the vertical to provide a worst-case test.

The two solid lines in Figure 1 show, on a logarithmic scale, the number of reliable matches obtained as a function of scene rotation angle for both SIFT (red line with x markers) and our proposed ASIFT algorithm (blue line with no markers) using a vertical rotation axis. We see that even at small rotation angles, ASIFT achieves more matches than SIFT and that it is much more robust to scene rotation of up to 60° where it obtains 200 times as many matches. The dashed lines show the corresponding results when the rotation angle is changed to be at 22.5° to the vertical; this is the worst case in relation to the foreshortening orientations used to form the extended set of dictionaries. We see that ASIFT now performs slightly worse at scene rotation angles above 45° but that it is still much better than SIFT whose performance is unaffected. Finally, the dotted lines, show the performance on the more realistic model village scene that is shown in Figure 2(b). This scene includes surfaces at many different orientations and the number of matches falls off more rapidly with rotation angle. We see that ASIFT algorithm consistently outperforms the SIFT with over ten times as many matches at scene rotation angles above 45°.

The method does not require additional processing to compute descriptors at the point where the sample images becomes available as the compensation takes place using only the reference image and so can be done beforehand however matching the descriptors will take longer due to

the requirement to match to a number of descriptor dictionaries.

4. Image Change Detection System

This section describes a image change detection system which utilises the ASIFT algorithm described above. The system uses ASIFT points to construct a Delaunay triangulation that results in a consistent segmentation of the reference and sample images. Corresponding triangular segments of the two images are compared by by aligning them with an affine transform and interpolating the pixel values onto a rectangular grid using cubic interpolation. A dense SIFT descriptor is obtained from the corresponding areas and used to determine if the areas appear the same between viewing angles. 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. The ASIFT points described in Section 3 are used to construct a Delaunay triangulation Delaunay(1934), Lee (1980) to segment the reference and sample images. A set of matched points is found using a high matching threshold in order to ensure the reliability of the detected matches. These initial matches are used to calculate the fundamental matrix that relates the positions of matching points on the two images using a RANSAC Fischler(1981) approach. A second set of matches are then found using a lower matching threshold which results in a increased number of less reliable matches. These matches are filtered by rejecting any that are not consistent with the estimated fundamental matrix.

Delaunay triangulation is conducted on the matches in the reference image and a consistent triangulation applied to the sample image. Occasionally, this results in overlapping triangles in the sample image, due either to false matches or to large deviations from planarity in the scene. In either case, the corresponding points are removed from the triangulation. This reduces the possibility of incorrect matches which might remain if the incorrect match happens to fall along the correct epipolar line and ensures areas of the scene are uniquely assigned to triangle areas.

The results of the segmentation are illustrated in Figs. 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. *Dense SIFT Region Matching*

Segmenting the images with Delaunay triangulation results in two useful properties, the areas within corresponding triangles in the two images approximately coincide and the three corners of a triangle allow affine compensation by mapping one triangle onto the other, both of which make registration of the contained area easier. If the area within the triangles in the reference and sample images are approximately planar a comparison of the areas after affine compensation can determine if they contain changes.

For a number of reasons, corresponding triangles will not match perfectly even if the scene is unchanged. Firstly the localisation error of the feature points used results in misalignments of the compensated triangles. Secondly the inaccuracies from using an affine approximation to represent the homography between the segments will also produce additional misalignments. Thirdly 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 those errors, techniques for matching the triangular segments that require perfect alignment such as autocorrelation do not produce optimal results.

The SIFT descriptor is used to compare triangular segments because it is robust to lighting changes and also to small pixel misalignments. Using a window that is stepped along the x and y axis SIFT descriptors can be collected across an area larger than the descriptor window. The window can be set to the desired size for the granularity of the comparison, experimentally

it was found that a single scale of a few pixels per bin with a step size equal to the bin size produce the best result. If two images have been registered a dense SIFT descriptor obtained from the two images will produce individual SIFT descriptor pairs that correspond to the same regions of the scene. To detect changes the euclidean squared distance between corresponding descriptors can be found. The region is said to match if the number of descriptor distances above a first threshold is below a second threshold.

4.3. *Evaluation*

The algorithm has been evaluated using the model village scene for which reference and sample images are shown in Figs. 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 algorithm classifies each triangle as changed (positive) or unchanged (negative) and these decisions are compared with the ground truth. Figure 2(a) shows false positives in red ('R'), false negatives in blue ('B') and true positives in gray; true negatives are omitted. We see that all the true changes have been correctly identified and that the false negative triangles ('B') occur only at the boundaries of true changes. False positives ('R') occur in three small regions and in all cases arise because the scene is far from planar within the affected triangle. When this happens, the affine compensation applied to the triangle does not correctly compensate for the perspective distortion in the image. The performance at 20° and 30° is affected by a reduction in the number of matched feature points which results in larger triangles which do not follow the scene contours as well moreover, performance is also affected by the increased levels of occlusion present.

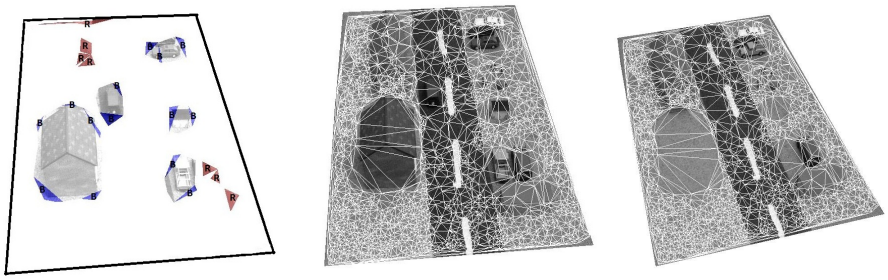
The majority of triangles overlaying areas of change have been detected. All of the false positives are caused by the triangle not lying on the plane of the scene. When the triangle does not lie on the plane of the scene the affine compensation is not correct for the areas within the triangle, also the area is likely to suffer from occlusion or disocclusion which the current method is not robust to. The larger a triangle the more likely it is to overlap two planes and so increasing the number of triangles by increasing the number of feature point matches will boost performance, indeed the key factor that causes performance to decrease at 20° and 30° is the reduction in matching feature points.

The ROC curve showing system performance in Figure 3 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. As the figure shows the performance of the system is dependent on the angle between the sample and reference images as the performance is sensitive to the density of feature point matches and levels of occlusion. The system works best on scenes with low depth variation where the levels of occlusion are small.

5. Summary

We have presented a method for detecting and locating changes between two images of a scene taken from widely spaced viewpoints. The method produces good performance at angles up to 30° with all true changes detected. False negatives are restricted to triangles that contain a small amount of change and false positives only appear where the triangles do not align to the image planes.

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. The method has been illustrated on images with azimuth change in the camera position but could equally be applied to situations with a change in elevation. This will allow for change detection to be applied to a far larger range of source data in a wider range of application and situations where the data capture is not necessarily optimised to this application.



(a) (b) (c)

FIGURE 2. Delaunay segments formed using ASIFT points on images at a 10^o with scene changes visible shown in reference image (b) and sample image (c). Change mask shown in (a) with correct positives highlighted in white, false positives and negatives in red (labelled 'R') and blue (labelled 'B') respectively.

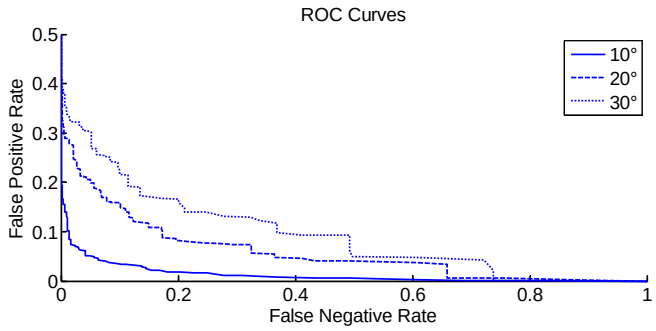


FIGURE 3. ROC curve of pixel classification performance for scene rotations of 10°, 20° and 30° using the model scene shown in Figure 2.

REFERENCES

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.l Remote Sensing.* **33**, 1655–1681.

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.

DELAUNAY, B. 1934 Sur la sphere vide. *Otdelenie Matematicheskikh i Estestvennykh Nauk.* **7**, 793–800.

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.

GIRARDEAU-MONTAUT, D., ROUX, M., MARC, R., THIBAULT, G. 2005 Change detection on points cloud data acquired with a ground laser scanner. *Proc. ISPRS Workshop.* **36**, 30–35.

HARTLEY, R., ZISSERMAN, A. 2004 Multiple View Geometry in Computer Vision. *Cambridge University Press.*.

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.

LEE, D.T., SCHACHTER, B. J. 1980 Two algorithms for constructing a delaunay triangulation. *Intl J. Computer and Information Sciences.* **9**, 219–242.

LOWE, D.G. 2004 Distinctive image features from scale-invariant keypoints. *Intl J. Computer Vision.* **60**, 91–110.

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.

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.

SINGH, A. 1989 Digital change detection techniques using remotely-sensed data. *Intl J. Remote Sensing.* **10**, 989–1003.

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.*