# Automatic Change Detection Method for an Indoor Environment

**D. Beulah David**
Department of Computer Science, Jeppiaar Engineering College, Chennai-119
Email: beulahchrist@gmail.com

--------------------------------------------------------------------ABSTRACT--------------------------------------------------------------------

**Detecting changed regions in multiple images of the same scene taken at different times is of great interest due to a large number of applications in diverse disciplines, including remote sensing, surveillance, medical diagnosis and treatment, civil infrastructure, and underwater sensing. The goal is to identify the set of pixels that are significantly different between the last image of the sequence and the previous images. This paper presents a study of an automatic change detection method for multiple images of the same scene acquired by a mobile-camera from different positions with no illumination changes. The proposed method used for this study consists of three steps: (1) automatic image registration, (2) temporal differencing and (3) unimportant changes removal. The results in this paper show that the presented method successfully detects new objects in the scene from the multiple images.**

## 1. INTRODUCTION

A major portion of the research efforts of the computer vision community has been directed towards the study of automatic image change detection methods [1] due to a large number of applications in diverse disciplines. Important applications of change detection include video surveillance, analysis of multitemporal remote sensing images, underwater sensing [2], tracking systems of moving objects, medical diagnosis, or driver assistance systems [3]. Given a set of images of the same scene taken at several different times, the goal is to identify the set of pixels that are significantly different between the last image and a previous reference image. Changed pixels may result from a combination of underlying factors, including appearance or disappearance of objects, motion of objects relative to the background, shape change of objects or environment modifications (buildings, fires), etc., [1].

Change detection methods applied to images of the same scene captured by stationary cameras at different times are widespread, both in the literature [8] and in everyday use, but only few publications refer to image analysis from non-stationary camera positions.

Oberti et al. [6] presented a video surveillance system based on a pan/tilt mobile-camera. Their method began by creating a panoramic multi-layer background image. The background image and the camera position when capturing an input image were used to perform change detection. While the system could detect a new object in a video sequence from a non-static camera, the system depended heavily on the current position of the camera when capturing an input image. When this position of the camera

was unknown, the system failed to detect changes in the input image.

Primdahl et al. [7] described an approach that analyzes videos acquired from a moving vehicle making repeated passes through a specific, well defined corridor. The objective of their approach was to detect stationary objects that appear in the scene along the established route. Their approach began with pairing images of different videos. For each frame pair, they introduced regions of interest. Although the new objects were successfully identified out of the videos captured from different camera locations, this only occurred because the authors manually selected four points on the regions of interest that contained the new objects. The authors explicitly identified the locations of the new objects.

## 2. TECHNIQUES IN IMAGE CHANGE DETECTION

### 2.1. Image Change Detection: Types
1) Temporal Difference Models
2) Vector and Shading Models
3) Clustering Models
4) Significance and Hypothesis Tests Models

### 2.2. Contribution of This Research
In this paper, an automatic change detection method is presented for multiple indoor images of the same scene captured by a mobile-camera from different positions. The goal of the proposed system is to detect new objects present in the images. The system does not require information of mobile-camera positions and regions of interest as input.

In future, the proposed change detection method could be applied to patrol large areas such as perimeter

surveillance of defense bases and airfields. Instead of installing numerous static cameras, a digital camera installed on a mobile-robot, for example an Unmanned Ground Vehicle (UGV) guided by a Global Position System (GPS), could be sent to patrol such large areas.

## 3. AUTOMATIC CHANGE DETECTION METHOD

The automatic change detection method consists of three major steps: (1) automatic image registration (AIR), (2) temporal differencing and (3) unimportant changes removal (UCHR). Figure 1 shows flow chart of the change detection method.

### 3.1 Automatic Image Registration

Image registration is used to match (spatially align) two or more images of the same scene acquired at different times, from different positions, and/or from different sensors [4]. The following steps are usually performed in image registration [9]: (1) feature extraction, (2) feature matching, (3) determination of a transformation function and (4) image resampling and transformation.

In this study, the Scale Invariant Feature Transform (SIFT) operator [5] was used to automatically extract and match local features from both reference and input images. The local features were extracted by the SIFT operator from three templates which were chosen and cropped manually off-line in advance from the reference image.

Input Image ← Mobile Camera

↓

AIR ←

↓

Reference Image

Temporal Differencing ←

↓

UCHR

↓

Final Change Mask

↓

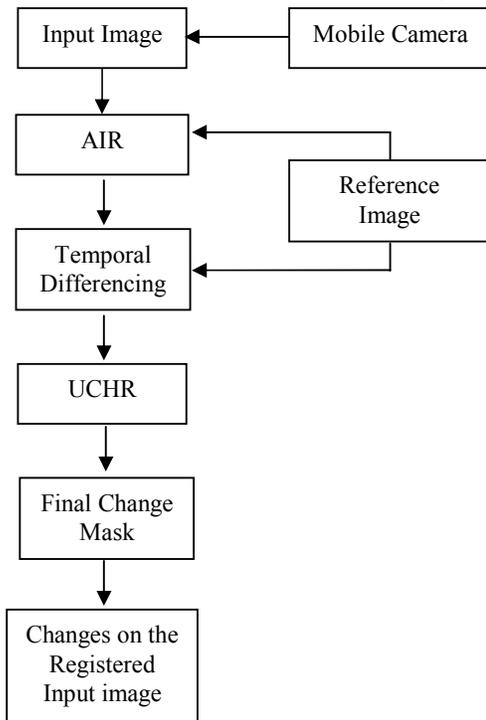Changes on the Registered Input image

Figure 1- Flow chart of the change detection method

The following were steps established in order to perform automatic image registration of an input image:

1. Read the reference image, three templates and the input image.
2. Extract three control point locations from the reference image
   2.1. Perform the SIFT operator towards the template1 and the reference image,
   2.2. Extract only matched keypoint locations in the reference image,
   2.3. Calculate an average value from the extracting matched keypoint locations and use the average value as the first control point location,
   2.4. Repeat steps 2.1 - 2.3 for template 2 and template 3 for the second and third control point locations.
3. Extract three control point locations from the input image by repeating step number 2.1 - 2.4.
4. Tune new three control point locations in the input image by using the cross-correlation method based on information of the original control point locations in the input image, the three control point locations in the reference image, the input image, and the reference image.
5. Estimate parameters of the linear conformal (similarity) transformation such as scaling, rotational, and translational differences between the reference and input images respectively, based on the three pairs of the correspondence points. The similarity transformation works based on the equation below:

$$X = sx\cos(\theta) - sy\sin(\theta) + h$$

$$Y = sx\sin(\theta) + sy\cos(\theta) + k$$

where $s$, $\theta$ and $(h,k)$ are the scaling, rotational, and translational differences between reference and input images.

6. Transform and re-sample the input image based on the estimated linear conformal transformation in order to generate the registered input image.

### 3.2 Temporal Differencing

In order to produce a preliminary change mask, the temporal difference algorithm was firstly used to subtract the reference image and the registered input image. If the reference image is assumed as R(x,y) and the registered input image is I1(x,y), then the subtracted image is :

$$D(x, y) = |R(x, y) - I1(x, y)| \qquad (1)$$

As a result of this subtraction, an overlapping region appeared on the subtracted image. Secondly, the overlapping area, referred to region of interest, was automatically cropped based on the registered input image information.

Image thresholding was finally performed to the region of interest. The result of this thresholding was a preliminary change mask (B(x,y)) created according to the following decision rule:

$$B(x, y) = \begin{cases} 1, if\ RoiD(x, y) \rangle \tau, change\ occured \\ 0, \quad if\ RoiD(x, y) \leq \tau, no\ change \end{cases} \qquad (2)$$

where τ is a global threshold value and RoiD(x,y) is region of interest of the subtracted image. To accelerate searching for significant changes, searching was only applied to the region of interest.

### 3.3 Unimportant Changes Removal

In Section 2.2, simple differencing followed by thresholding has the advantage of low computational cost. However, the result is sensitive to noise and shadow. In addition, misalignment of the input image also causes unimportant changes appearing in the preliminary change mask.

The misalignment happened because control points extracted from the three templates could not cover the whole area of the reference and input images. Hence, the following operations were designed to remove unimportant changes from the preliminary change mask.

### 3.3.1 Clearing Border Objects

In this operation, objects that touch the border of the preliminary change mask are removed. By using morphological reconstruction, in which the preliminary change mask is used as the mask, and the marker image, $f_m$ is defined as:

$$f_m(x, y) = \begin{cases} B(x, y), if\ (x, y)\ is\ not\ on\ the\ border\ of\ \mathbf{B} \\ 0, \quad otherwise \end{cases} \quad (3)$$

### 3.3.2 Erosion Using Vertical and Horizontal Lines

Erosion is a process of shrinking objects in a binary image. The manner and extent of shrinking is controlled by a structuring element. In this case, horizontal and vertical lines were used as the structuring element in order to erode objects in the $f_m(x, y)$.

### 3.3.3 Removing Small Objects

Next, all connected components (objects) that had fewer than 100 pixels were removed from a binary image $f_m(x, y)$. In this real life setting, 100 pixels represent an object with area of 480 square millimeters (size of a standard pencil) when an image is captured from a 230 cm distance.

A final change mask $f_{m1}(x, y)$ was generated as output of this operation:

$$f_{m1}(x, y) = \begin{cases} f_m(x, y), if\ (x, y) \rangle 100\ pixels \\ 0, \quad otherwise \end{cases} \quad (4)$$

## 4. REFERENCE AND INPUT IMAGES

In order to test the automatic change detection method of two images, Fig 2(a) and (b), were used as input in this study.



(a)



(b)

Figure 2- (a) The reference image and (b) The input image

The image in Fig 1(a) is referred as the reference image. It was a scene of an indoor environment that consists of objects such as books, pictures, a wall and a table. It was captured in such a way that the position of a digital camera was perpendicular towards the scene.

The distance between the digital camera and the wall of the scene was 230 cm. This current position of the mobile-camera was referred as the origin position ((0,0,0) position).

In this study, the movement of the mobile-camera was restricted only by three degrees of freedom ( $(X, Y, \Delta\theta_z)$ axes). The X axis represented translation on left (-) or right(+) sides. The Y axis indicated movement on forward (+) or backward (-) sides (zoom in / zoom out).

The $\Delta\theta_z$ axis was rotation on the clockwise (+) or anticlockwise (-) sides. Fig 3 shows a top view of the mobile-camera origin position.

After capturing the reference image, four new objects were added to the scene: a pencil, a small bolt, a conventional camera and a glasses case. The digital camera was then shifted manually 10 cm forward (zoom in) and 15 cm to the right. It was then rotated 15 degrees in

an anticlockwise direction. This current camera position, (+15, +10, -15), was referred to as the input camera position, from which another image was captured, the input image (see Fig 2(b)). Fig 4 shows a top view of the input camera position.
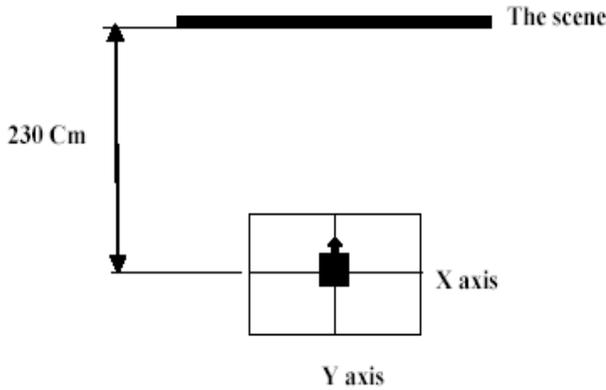


Figure 3- Top view of the mobile-camera origin position

After capturing the reference image, four new objects were added to the scene: a pencil, a small bolt, a conventional camera and a glasses case. The digital camera was then shifted manually 10 cm forward (zoom in) and 15 cm to the right. It was then rotated 15 degrees in an anticlockwise direction. This current camera position, (+15, +10, -15), was referred to as the input camera position, from which another image was captured, the input image (see Fig 2(b)). Fig 4 shows a top view of the input camera position.
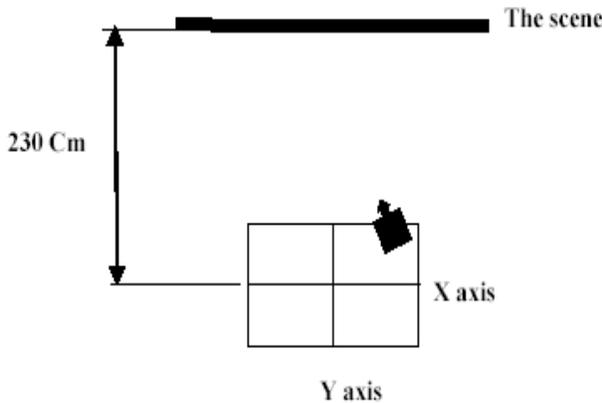


Figure 4.- Top view of the input camera position

As can be seen in Fig 3, the mobile-camera has been shifted on the three degrees of freedom direction $((X, Y, \Delta\theta_z)$ axes) . In next section, results of this study are presented.

## 5. RESULTS

In order to automatically extract control points from both reference and input images, three templates were chosen and cropped manually from the reference image. These templates are shown in Fig 5.



(a)                    (b)                    (c)

Figure 5 - (a) Template 1 (b) Template 2 (c) Template 3

Any object in the reference image can become a candidate as a template. The main reason for choosing these templates is that they contain many local features. The SIFT operator was then used in order to extract and match local features known as key points from these templates and both reference and input image.



Figure 6-A sample of matched keypoints provided by the SIFT operator

Referring to the resulting image (Fig 6), the SIFT operator extracted 691 keypoints from the input image and 116 keypoints from template 1. From these extracted keypoints, 49 were matched. One matched keypoint indicated positions of keypoints found by the SIFT operator matching on the input image and template 1 in X and Y axes.

By extracting the positions of matched keypoints only in the input image, a control point could be determined by calculating an average value in the X and Y axes.The same process was also employed for the template 2 and 3. Fig 7(a) and (b) show control points extracted automatically from both reference and input images.

A tuning process was then performed to increase, to sub-pixel accuracy, the control points on the input image. The tuning process uses the following general procedure. For each control-point pair,

1.1. Extract an 11-by-11 template around the input control point and a 21-by-21 region around the base control point,

1.2. Calculate the normalized cross-correlation (NCC) of the template with the region. The NCC equation is shown below.

$$\gamma(u,v) = \frac{\sum_{x,y}\left[f(x,y)-\overline{f_{u,v}}\right]\left[t(x-u,y-v-\overline{t}\right]}{\left\{\sum_{x,y}\left[f(x,y)-\overline{f_{u,v}}\right]^2 \sum_{x,y}\left[t(x-u,y-v)-\overline{t}\right]^2\right\}^{0.5}}$$

where $\gamma(u,v)$ is a matrix that contains the correlation coefficients, which can range in value from -1.0 to 1.0. $f$ is the 21-by-21 region around the base control point and the sum is over x,y under the window containing the 11-by-11 template $t$, positioned at (u,v). $\overline{t}$ is the mean of the template and $\overline{f_{u,v}}$ is the mean of $f(x,y)$ in the region under the template.

1.3. Find the absolute peak of the cross-correlation matrix
1.4. Use the position of the peak to adjust the coordinates of the input control point.
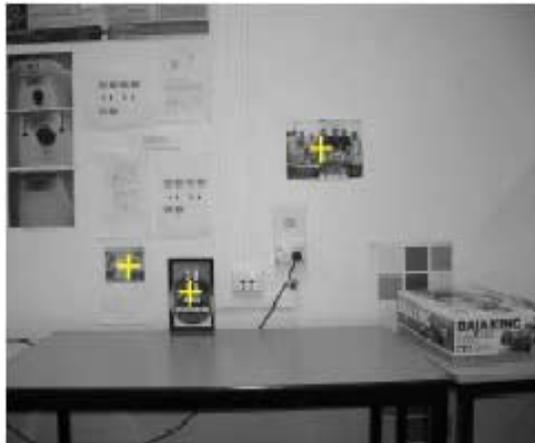


(a)
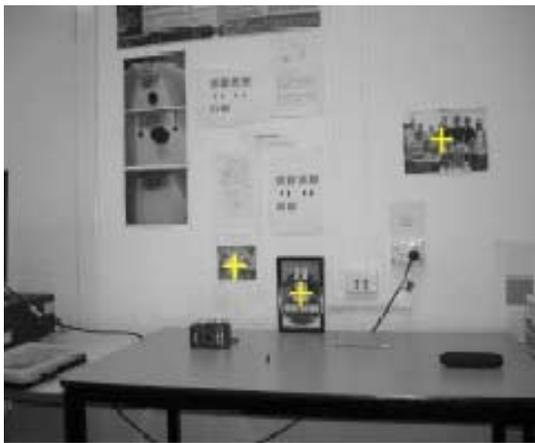


(b)

Figure 7 - Control points extracted automatically from the (a) reference image (b) input image

Table 1 below presents comparison of control point values on the input image before and after the tuning process.

TABLE I THE RESULT OF TUNING PROCESS

| Tem plate | Control point Locations in X and Y Axes | | | | | |
| | Before | | After | | Difference | |
| | Xin | Yin | Xadj | Yadj | X | Y |
|---|---|---|---|---|---|---|
| 1 | 523 | 148 | 523.9 | 149.9 | 0.1 | -1.9 |
| 2 | 356 | 317 | 354.1 | 314.7 | 1.9 | 2.3 |
| 3 | 277 | 287 | 277 | 287 | 0 | 0 |

Next, control points of the reference image and the new adjustment control points provided by the tuning process were used to estimate parameters of similarity transformation such as scaling, rotating and translating in order to register the input image into the same coordinate system with the reference image. Figure 8 shows the registered input image.



Figure 8- The registered input image

Next, a subtracted image was produced by subtracting the registered input image and the reference image. Figure 9 shows the subtracted image.

As can be seen in Figure 9, there are two areas: a black area and a gray area. The black area is an overlapping area, referred to as region of interest, as a result of simply differencing between the reference image and the registered input image.

The gray area is produced by subtracting the reference image and zero values in the registered input image. The

black area also contains significant changes. Hence, the search for significant changes was only concentrated on the region of interest.

Next, the region of interest was automatically .separated from the subtracted image by using the registered input image information. The registered input image was firstly changed into a binary image by using the rule below.
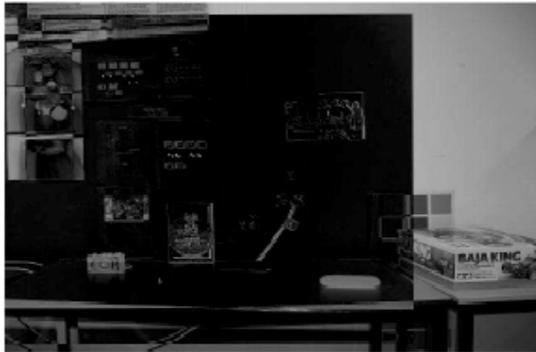


Figure 9- The subtracted image

$$RB(x, y) = \begin{cases} 1 \ if \ P(x, y) \rangle 0 \\ 0 \ if \ P(x, y) = 0 \end{cases}$$

(5)

where $RB(x, y)$ is a binary image of the registered input image and P(x,y) represents a pixel value for coordinates (x, y). As a result of the image thresholding, the binary image contained white and black areas. The white area was referred to as the bounding box.

The bounding box parameters such as the coordinate of the upper left corner, width and height were secondly extracted from the $RB(x, y)$. Based on these bounding box parameters information, the region of interest was finally cropped from the subtracted image.

By applying a global threshold value towards the region of interest, a preliminary change mask was generated. Figure 10 shows the preliminary change mask.



Figure 10-The preliminary change mask.

As can be seen in Fig 10, unimportant and significant changes apparently appear in it. Misalignment of the registered input image seems to be a major cause of unimportant changes appearing in it besides noise and shadow. Two images of the same scene captured by a mobile-camera from two different positions cause a large disparity between pixels of a same object in both images.

Registering such a pair of images by using similarity transformation parameters estimated from the three pairs of correspondence points might not accurately handle pixels which have such large disparity. However, any new object that entered to the scene apparently appeared in the preliminary change mask.

To separate significant changes while rejecting unimportant changes from the preliminary change mask, unimportant changes removal was then employed. First, any object connected to the preliminary change mask border was suppressed. Fig 11 shows the result of this suppression.



Figure 11- The result of suppressing objects that touch image border

Second, erosion was applied towards the image in Fig 11. Horizontal and vertical lines were used as structuring elements in order to reduce, horizontally and vertically, the objects in the image. Figure 12 shows the result of this erosion.



Figure - 12. The result of vertical and horizontal erosion

Third, objects with less than 100 pixels were removed from the image in Fig 12. The result of this process is the

final change mask. Fig 13 shows the result of removing these small objects.

To find the location of these significant changes in the registered input image, centroids, the center of mass of the region, of final objects in the final change mask were finally extracted.

Based on information of these centroids, circles can be drawn on the registered input image in order to highlight the location of significant changes on the registered input image. Fig 14 shows the location of significant changes on the registered input image.



Figure-13. The result of removing objects less than 100 pixels.



Figure 14. The location of significant changes

As can be seen in Fig 14, two of the four expected significant changes have been detected correctly by the automatic change detection method. Although the change detection method has successfully detected the significant changes, it has a limitation.

The limitation of this method is that the method fails to detect small objects which have less than or equal to 100 pixels or a diameter less than or equal to 1 cm.

## 6. CONCLUSION

In conclusion, an automatic change detection method is presented for multiple images of the same indoor scene captured by a mobile-camera. The method consists of (1) automatic image registration, (2) temporal difference algorithm and (3) removal of unimportant changes. The method was demonstrated to be successful in automatically detecting significant changes in the registered input image. The effects of illumination changes will be taken into consideration in the future study.

## REFERENCES

[1] Cheng-Chi Chang, Tsorng-Lin Chia, and Chen-Kuei Yang. Modified temporal difference method for change detection. *Optical Engineering 44(2)*: 1-10, February 2005.

[2] D. Edgington, W. Dirk, K. Salamy, C. Koch, M. Risi, and R. Sherlock. Automated event detection in underwater video. *In proceeding MTS/IEEE Oceans 2003 Conference, 2003.*

[3] C. Y. Fang, S. W. Chen, and C. S. Fuh. Automatic change detection of driving environment in a vision-based driver assistance system. *IEEE Transaction Neural Network 14(3)*:646-657, May 2003.

[4] A. Ardeshir Goshtasby. *2-D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

[5] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60(2):* 99-110, November 2004.

[6] Franco Oberti, Lucio Marcenaro, and Carlo S. Regazzoni. Real-time change detection methods for video-surveillance systems with mobile camera. *In XI European Signal Processing Conference 2002* (EUSIPCO'2002), pages 37-40, Toulouse, France, September 2002.

[7] K. Primdahl, I. Katz, O. Feinstein, Y. Mok, H. Dahlkamp, D. Stavens, M. Montemerlo, and S. Thrun. Change Detection from Multiple Camera Images Extended to Non-Stationary Cameras. In Proceedings of Field and Service Robotics 2005 (FSR05). Port Douglas, Australia, 2005.

[8] Richard J. Radke, Srinivas Andra, Omal Al-Kofahi, and Badrinath Roysam. Image change detection algorithms: a systematic survey. *IEEE Transactions on Image Processing 14(3)*: 1-14, 2005.

[9] B. Zitova, and J. Flusser. Image registration methods: a survery. *Image and Vision Computing 21*: 977-1000, 2003.

### Authors Biography

*Beulah David. D* was born in Tirunelveli on May 31, 1984. She graduated from the Francis Xavier Engineering College, Tirunelveli and completed her Master Degree in Dr.Sivanthi Aditanar College of Engineering, Tiruchendur. She is now working as an Assistant Professor in Jeppiaar Engineering College, Chennai in the Department of Computer Science and Engineering. Her special field of interest includes Digital Image Processing.