Unsupervised Change Detection in Optical Satellite Imagery Using SIFT Flow

B. Awad[a,b], I. Erer[a,b]

[a]ITU, Electrical and Electronic Engineering Faculty, Electronics and Communication Department, - (awad, ierer)@itu.edu.tr
[b]34469 Maslak, Istanbul, TURKEY

KEY WORDS: Change detection, Scale invariant, SIFT flow, Satellite imagery, Very high resolution, Unsupervised learning, Unsupervised change detection.

ABSTRACT:

The process of identifying change in remote sensing images has been a focal point of research for decades now. Many classical algorithms exist, and many new modern ones are still being developed. These algorithms can be divided into supervised and unsupervised. In this work an unsupervised method is presented. This method relies on the scene alignment algorithm SIFT flow. It is shown that building upon simple principles an accurate change map can be obtained from the SIFT descriptor flow of the two input images. Furthermore, it is shown that this method despite its simplicity exceeds other unsupervised methods and comes close to supervised ones, even exceeding them in some metrics. Lastly, the advantages of SIFT flow in comparison to the supervised methods are highlighted alongside its own downsides.

1. INTRODUCTION

Change detection in optical satellite imagery has been studied extensively throughout the years. Its utility is paramount in many applications in the remote sensing filed. For example, change detection can be used to assess urban growth, urban planning, monitoring the effect of forest fires and other natural disasters, and many other applications. Methods used for change detection can be split into two categories: supervised methods and unsupervised methods. Supervised methods use labeled data in their training process, and later make predictions based on it. On the other hand, unsupervised methods make predictions directly, without the need for any training (Ban and Yousif, 2016).

Due to lack of data up until recent times, most of the methods developed for change detection belong to the unsupervised category. In (Celik, 2010)expectation maximization algorithm is carried out on wavelets extracted by dual-tree complex wavelet transform. These wavelets are extracted from the difference image, while in (Hao et al., 2014) expectation-maximization-based level set method is used instead calming to yield better results. In (Celik, 2009a) the proposed method is based on segmenting the difference image to non-overlapping blocks, applying principal component analysis, and then using k-means to obtain reference vectors for each cluster. In the end features for each pixel are extracted, compared to the reference vectors and assigned to each of the two clusters. In (Li et al., 2015a) Gabor wavelets followed by two-levels clustering was implemented. (Bruzzonel and Femluldez, n.d.) modeled the likelihood function of the data using Parazan windows to make the final decision. In 1986 an error minimizing algorithm for defining a threshold over a difference image was suggested by (Latifovic and Pouliot, 2014), this algorithm was developed an improved and used vastly later. For example, (Melgani et al., 2002) used the algorithm over a difference image of multispectral images, (Bazi et al., 2005) and (Moser and Serpico, n.d.) expanded the algorithm to deal with the intensity and amplitude images of SAR which do not have a Gaussian nature. In (Bazi et al., 2006) the same algorithm was extended to a multi-threshold case, but it causes the algorithm to be more computationally expensive. The algorithm was further improved by modifying it

to work on the log-ratio image for SAR images in(Ban and Yousif, 2012).

In general, more recent supervised or semi supervised methods such as manage to produce better results than their unsupervised counterparts. For example, (Zhang et al., 2018) created a semi supervised method, that utilizes coarse to fine detection scheme to combine the benefits of supervised learning and mitigate the lack of data. However, In recent years new datasets have been introduce such as the one (Chen and Shi, 2020), which also suggested a new supervised model based on attention Pyramid spatial–temporal attention module (PAM). This dataset opened the flood gates for deep supervised models such as the two models presented by (Diakogiannis et al., 2021). Both models are based on attention mechanism, but the more advanced CEECNET showed the most promising results. Alas, supervised methods suffer from two downsides, first, supervised methods may produce better results when the test data is either from the same dataset as the training or similar to it, however, they tend to underperform whenever the test data does not resemble the training dataset. The second problem, which is a generalization of the first one, which is that the amount of data needed to train the more advanced supervised methods such as deep networks is large, and the amount of data that would be needed to train a supervised model to generalize for large, diverse areas is almost nonexistent.

Unsupervised methods however do not need any labeled data. They depend on the information within the pair of images to make assessments and locate the area of change for example in (Celik, 2009b; Li et al., 2015b). Many unsupervised methods in change detection either use a difference image or ratio image as based approaches to produces their final result. These methods are quite successful in determining the difference between two sets of images. However, for these methods to work the pair of temporal images must go through a preprocessing step to ensure that their histograms match exactly. Furthermore, any radical change in lighting conditions or sensor inconstancy may affect the results massively. Moreover, unsupervised methods lack the

specialization that supervised methods have, they cannot be trained to pay attention on one type of change and ignore the other.

In this work, an unsupervised method for change detection is presented. This method is based on the Scale Invariant Feature Transform (SIFT) (Lowe, 1999) and its flow between to images: SIFT flow (Liu et al., 2011). In the first section of this article will be a brief introduction to SIFT features and the SIFT flow algorithm. The second section will show the simple, yet effective approach used to transform SIFT flow into a change detection method. In the third section, visual results of SIFT flow are shown and discussed, also numerical results are examined to compare the performance of SIFT flow to other unsupervised and supervised methods. The final section offers a summary of this work and suggestions for future works based on it.

2. SIFT FLOW

SIFT flow is an alignment algorithm that was inspired by optical flow. However, instead of matching pixel intensity values, SIFT feature descriptors (or SIFT descriptors) are used. For this reason, in the following subsection a short explanation of the SIFT features is conducted, followed by the matching scheme used by SIFT flow.

2.1 Scale Invariant Feature Transform

The SIFT algorithm consists of two parts: key point detection and feature extraction. The latter being the only relevant part for SIFT flow and this work. Typically, SIFT features are simple and easy to compute. Given a key point x the neighbouring $N \times N$ pixels are selected, and their orientation is calculated using (Lowe, 2004) :

$$\theta(x,y) = atan^{-1}\left(\frac{I(x,y+1)-I(x,y-1)}{I(x+1,y)-I(x-1,y)}\right), \quad (1)$$

Where $I(x,y)$ is the gradient around the pixel located at (x,y). Then this block of pixel orientations is divided into cell arrays (typically 4×4) and the histogram for each cell is calculated using 8 bins. Finally, the resulting vector is normalized and clipped where only values larger than 0.2 are kept. For a 16×16 window divided into 4×4 cells with quantization of 8 bins, the resulting feature vector has the size of $4 \times 4 \times 8 = 128$ (Lowe, 1999).

2.2 SIFT Flow

The SIFT Flow algorithm takes advantage of the scale invariant nature of the SIFT features. In SIFT flow SIFT descriptors are calculated for every pixel in the image. In other words, for every pixel location (x,y) a 128 SIFT descriptor is generated. This gives rise to what is called a SIFT image, a SIFT representation of the original image that has the size of $M \times N \times 128$ where $M \times N$ is the size of the original image. Since any SIFT descriptor in the first image may match with any SIFT descriptor in the second image (regardless of their locations) SIFT flow utilizes a top-down matching scheme over a gaussian pyramid as seen in Figure 1. This method reduces the number of possible matches and increases the quality of the matching (Liu et al., 2011).
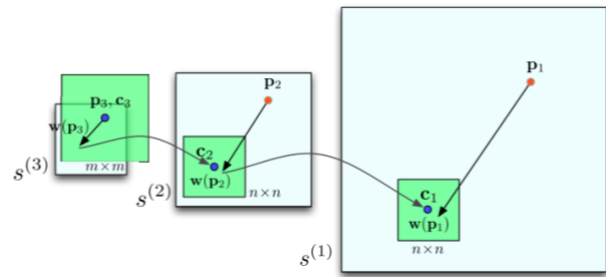
Figure 1. Top-down matching scheme used in SIFT flow (Liu et al., 2011)

The energy function seen in equation (1) used to determine the best matching SIFT descriptors is inspired by the energy function of optical flow. It has three terms to ensure smooth and consistent matching. The first term is the data term which indicates the distance between the descriptors. The second term is referred to as the small displacement term, which is there to ensure that the flow vectors remain as small as possible when no other information is available. The third term is the smoothness term, which is used to make sure that neighbouring points have a uniform flow.

$$E(w) = \sum_{\mathbf{p}} min\left(\left\|s_1(\mathbf{p}) - s_2(\mathbf{p} + \boldsymbol{w}(\mathbf{p}))\right\|_1, t\right) + \sum_{\mathbf{p}} \eta(|u(\mathbf{p})| + |v(\mathbf{p})|) + \sum_{(\mathbf{p},\mathbf{q})\in\varepsilon} min(\alpha|u(\mathbf{p}) - u(\mathbf{q})|, d) + min(\alpha|v(\mathbf{p}) - v(\mathbf{q})|, d), \quad (4)$$

In this equation: $\mathbf{p} = (x, y)$ represents the grid coordinates of any given image, $\boldsymbol{w}(\mathbf{p}) = (u(\mathbf{p}), v(\mathbf{p}))$ represents the flow vector at $\mathbf{p}$ and $u(\mathbf{p})$ $v(\mathbf{p})$ represent its components in the x and y directions respectively. s_1 and s_2 are the two SIFT images desired to be matched. t and d are thresholds for the L1 norm used in the first and last terms, they are used to increase robustness in matching outliers and prevent discontinuities in the flow field respectively. η and α are hyper parameters used to control the second and third terms of the cost function respectively. Figure 2 shows example of SIFT flow being used for image registration and scene alignment.

Figure 2. Scene alignment application of SIFT flow.

3. PROPOSED METHOD FOR CHANGE DETECTION USING SIFT FLOW

SIFT flow algorithm's goal is to find the best matching SIFT descriptors between two images. It can be and has been used in many applications such as image registration, face alignment, motion hallucination and much more. However, change detection

is very much different from scene alignment and image registration. This section presents simple yet effective method to utilize this algorithm for change detection. This novel approach is based on two basic principles and a thresholding process, both are explained in the following subsections.

3.1 Basic principles of change detection using SIFT flow

Change detection using SIFT flow is based on two simple yet strong assumptions: First, satellite images are already registered. This is a fact of life, optical satellite images used for change detection must represent the same geographical area. In other words, every pixel in each image represents the same point at different times. The second assumption is that SIFT flow for two identical images must be or is close to zero. This stems from the energy function of SIFT flow and the way it was built. If two pixels and their corresponding SIFT descriptors have not undergone enough significant change, they will match with each other.

By combining these two assumptions together: since a pair of multitemporal registered satellite images represents the same area pixel by pixel, and SIFT flow is zero or small when the pixels did not go significant change, it can be said that the regions where the flow vectors are small or close to zero are unchanged while the areas with large SIFT flow are changed.

3.2 SIFT Flow to Change Map

SIFT flow produces flow vectors in the X and Y directions of the image. In other words, for a pixel location $\mathbf{p}$ there exists a flow vector in the X direction $u(\mathbf{p})$ and a flow vector in the Y direction $v(\mathbf{p})$. To generate the change map, the intensity of the flow field must be first determined. This is done by taking the norm if the flow vectors in both directions at each pixel location:

$$Intensity_{\mathbf{p}} = \|w(\mathbf{p})\| = \sqrt{u(\mathbf{p})^2 + v(\mathbf{p})^2}, \qquad (2)$$

The change map (CM) is calculated by thresholding the intensity of the flow vectors for all pixel locations. If $Intensity_{\mathbf{p}}$ is the flow intensity at pixel location $\mathbf{p}$ and $Intensity$ is the intensity of the flow for all pixels in the image, then the change map CM is calculated as follows:

$$CM = Intensity > T, \qquad (3)$$

Where T is a threshold value determined empirically. In summary, the method proposed uses the two-dimensional flow vectors obtained from SIFT, calculates the intensity of these vectors, and applies a threshold on the resulting intensity image. This process leads the generation of change map if the pair of images has been registered. The flow chart of the proposed method can be seen in Figure 3.
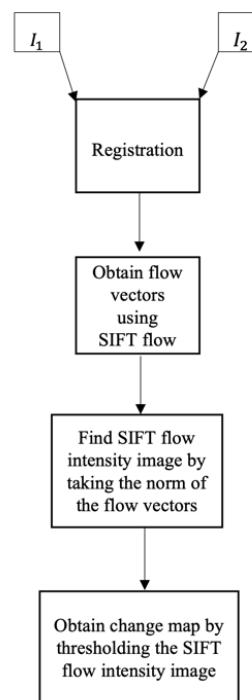

Figure 3. Flow chart for change detection using SIFT flow

4. RESULTS AND DISCUSSION

Firstly, in this section, the dataset used will be briefly described. Later, results of SIFT flow-based change detection are presented visually. Lastly, numerical results comparing SIFT flow change detection to other unsupervised and supervised methods are shown and discussed.

4.1 Dataset

The dataset used in this work is the LEVIR-CD dataset (Chen and Shi, 2020). It is made up of 637 very high resolution (VHR) image pairs. Each image has the spatial resolution of 0.5 meters. This dataset was compiled specifically to represent the change in residences through the region. Since SIFT flow is an unsupervised method, only the test images are used in the calculation of error and visualization of results. This facilitates comparability between SIFT flow and supervised methods that need the training data for training. Example images of the dataset can be seen in Figure 4.
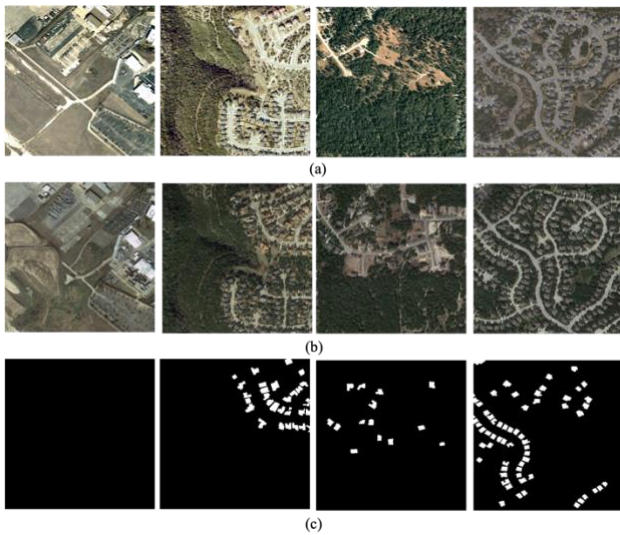
Figure 4. Samples from the LEVER-CD dataset. (a) Images at older point in time. (b) More recent images of the same location. (c) Ground truth depicting the changed area.

4.2 Results and Discussion

In this section both visual results for SIFT flow-based change detection, and numerical results are presented. These results are compared to 3 different methods: PCA-Kmeans change detection algorithm (Celik, 2009b), which a difference based unsupervised method, (PAM) (Chen and Shi, 2020), a deep supervised method that was presented alongside the dataset, and CEECNet (Diakogiannis et al., 2021), another deep supervised network that utilizes attention and residual network architecture.

4.2.1 SIFT Images

The prerequisite for SIFT flow is to generates SIFT images, these images are equal to the size of the image but have 128 dimensions instead of 3. However, SIFT images can be projected to 3-dimensional space using principal component analysis (PCA). This way SIFT images can be visualised. Figure 5 shows a few examples of SIFT images generated for pairs of images from the dataset.
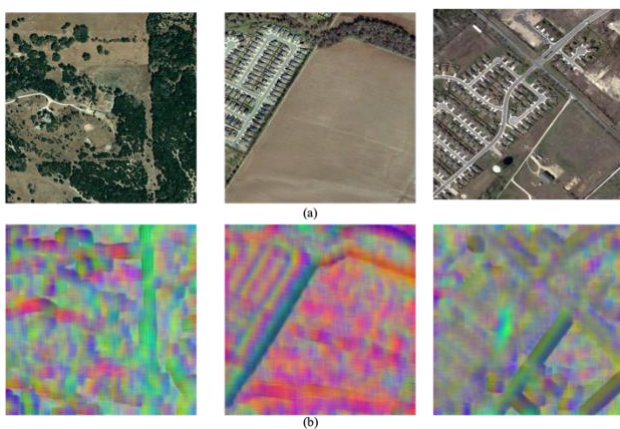
Figure 5. Visualisation of SIFT images. (a) example images. (b) corresponding SIFT image visualization.

4.2.2 SIFT Flow Visual Results:
Upon closer examination of Figure 6, It can be noticed that the resulting change map of SIFT flow is clustered around the change region. Specifically, in part (b) of Figure 6, the SIFT flow change map, although not precise, managed to encapsulate the region of change. In part (a) however, the change map is sparser, yet it still corresponds to the location of change seen in the ground truth.
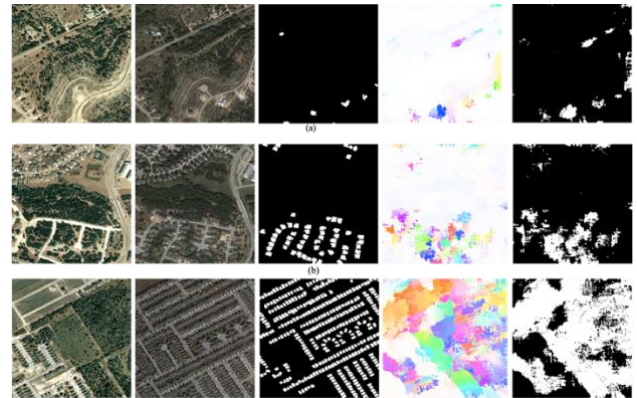
Figure 6. SIFT results for two pairs of images from the dataset. (a) shows the first pairs results and (b) shows the second pair. (c) shows the third pair. For each pair the first two images are the multitemporal pair, the third image is ground truth, the fourth image is the flow vector field resulting from SIFT flow and the last image is the change map obtained by thresholding the flow field.

To put the visual results of SIFT flow-based change detection, Figure 7, contains the visual results of two other models alongside it. The first row of images in this Figure 7, are the input image pair and their ground truth. The second row shows the results of PCA-Kmeans in (d), CEECNet in (f) and SIFT flow-based change map in (e).

Figure 7. Results each odd row is the pair of input images and their ground truth. Each even row is the results of PCA-Kmeans, CEECNet and SIFT flow respectively.

Results seen in Figure 7, put SIFT flow performance in preceptive. Although the change map produces by it is not as detailed or precise as the change map produces by CEECNet, SIFT flow's result is concentrated around the change area. Clearly it is not as good as the detailed result of CEECNet, however, it is important to remember that SIFT flow did not use any examples from the training or validation dataset. This make it very useful in cases where just general knowledge of the change area is needed, not minute details such as forest fire detection. Furthermore, since no pre-processing was conducted, the results of PCA-Kmeans has been affected massively, SIFT flow however was robust to the differences between the

histogram of the two images and still managed to find the correct area of change.

4.2.3 Numerical Results: The test dataset consists of 128 images. The supervised models are trained using the 445 images and validated on 64 images. PCA-Kmeans and SIFT flow are directly applied to the test dataset. No pre-processing was conducted for any of the models. Common metrics such as precision, precision and F1-score are used as evaluation. The results can be seen in Table 1.

$$\text{Precision} = \frac{TP}{TP+FP} \ , \tag{5}$$

$$\text{Recall} = \frac{TP}{TP+FN} , \tag{6}$$

$$F1 = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \ , \tag{7}$$

Where TP is true positive, FP is false positive, FN is false negative.

Method	Metric		
	precision	recall	F1
PAM	83.7	91	87.3
CEECNet	**93.81**	89.92	**91.83**
PCA-Kmeans	57.36	92.58	68.46
SIFT Flow	73.60	**95.42**	81.65

Table 1. Numerical results.

As seen in Table 1. SIFT flow-based change detection achieves F1-score of almost 82 percent, which mainly due to its high recall value. This is consistent with the visual results shown in Figure ß6 and Figure 7. The high recall value is due to the fact that SIFT flow identifies the regions of change accurately. However, the precision of the change maps generated by SIFT flow is not on par with supervised models. PAM and CEECNet are modern supervised models, therefore they perform well in allocating the change precisely. However, since they are supervised, they are prone to overfitting and falling victim to the problem of dataset bias. SIFT flow results rival that of PAM and are almost identical to its vanilla version which has an F1-score of %83.9 (Chen and Shi, 2020). However, it has not been trained on this dataset, and similar performance can be expected for different datasets and even in situation where no data exist, which is much more common in practice.

5. CONCLUSION

In this work a novel method for change detection in optical satellite images has been introduced. The method utilizes the SIFT flow algorithm which is typically used for image and scene alignment. By making two simple assumptions, SIFT flow's result is used to produce an accurate change map. SIFT flow rivals deep supervised models when it comes to common accuracy metrics and even exceeds them in some. Furthermore, SIFT flow-based change detection has big advantages in comparison to other models such as not needing any training data and not requiring any pre-processing. This makes it very universal and applicable in all situations regardless of data scarcity. However, SIFT flow-based change detection has its downfalls such as lower precision and lack of semantic awareness. For these reasons, it is recommended that further research to be conducted to improve the precision of the change maps produced by this method. Furthermore, its universal nature can be used with other models to produce good results when training data does not exist or is not sufficient for training.

REFERENCES

Ban, Y., Yousif, O., 2016. Change detection techniques: A review, in: Remote Sensing and Digital Image Processing. Springer International Publishing, pp. 19–43. https://doi.org/10.1007/978-3-319-47037-5_2

Ban, Y., Yousif, O.A., 2012. Multitemporal spaceborne SAR data for urban change detection in China. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5, 1087–1094. https://doi.org/10.1109/JSTARS.2012.2201135

Bazi, Y., Bruzzone, L., Melgani, F., 2006. Automatic identification of the number and values of decision thresholds in the log-ratio image for change detection in SAR images. IEEE Geoscience and Remote Sensing Letters 3, 349–353. https://doi.org/10.1109/LGRS.2006.869973

Bazi, Y., Bruzzone, L., Melgani, F., 2005. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. IEEE Transactions on Geoscience and Remote Sensing 43, 874–886. https://doi.org/10.1109/TGRS.2004.842441

Bruzzonel, L., Femluldez, D., n.d. Unsupervised Change Detection in Multisource and Multisensor Remote Sensing Images.

Celik, T., 2010. A Bayesian approach to unsupervised multiscale change detection in synthetic aperture radar images. Signal Processing 90, 1471–1485. https://doi.org/10.1016/j.sigpro.2009.10.018

Celik, T., 2009a. Unsupervised change detection in satellite images using principal component analysis and κ-means clustering. IEEE Geoscience and Remote Sensing Letters 6, 772–776. https://doi.org/10.1109/LGRS.2009.2025059

Celik, T., 2009b. Unsupervised change detection in satellite images using principal component analysis and κ-means clustering. IEEE Geoscience and Remote Sensing Letters 6, 772–776. https://doi.org/10.1109/LGRS.2009.2025059

Chen, H., Shi, Z., 2020. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. Remote Sensing 12. https://doi.org/10.3390/rs12101662

Diakogiannis, F.I., Waldner, F., Caccetta, P., 2021. Looking for change? Roll the dice and demand attention. Remote Sensing 13. https://doi.org/10.3390/rs13183707

Hao, M., Shi, W., Zhang, H., Li, C., 2014. Unsupervised change detection with expectation-maximization-based level set. IEEE Geoscience and Remote Sensing Letters 11, 210–214. https://doi.org/10.1109/LGRS.2013.2252879

Latifovic, R., Pouliot, D., 2014. Monitoring cumulative long-term vegetation changes over the Athabasca Oil Sands region. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 3380–3392. https://doi.org/10.1109/JSTARS.2014.2321058

Li, H.C., Celik, T., Longbotham, N., Emery, W.J., 2015a. Gabor Feature Based Unsupervised Change Detection of Multitemporal SAR Images Based on Two-Level Clustering. IEEE Geoscience and Remote Sensing Letters 12, 2458–2462. https://doi.org/10.1109/LGRS.2015.2484220

Li, H.C., Celik, T., Longbotham, N., Emery, W.J., 2015b. Gabor Feature Based Unsupervised Change Detection of Multitemporal SAR Images Based on Two-Level Clustering. IEEE Geoscience and Remote Sensing Letters 12, 2458–2462. https://doi.org/10.1109/LGRS.2015.2484220

Liu, C., Yuen, J., Torralba, A., 2011. SIFT flow: Dense correspondence across scenes and its applications. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 978–994. https://doi.org/10.1109/TPAMI.2010.147

Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision.

Lowe, D.G., 1999. Object Recognition from Local Scale-Invariant Features.

Melgani, F., Moser, G., Serpico, S.B., 2002. Unsupervised Change Detection Methods for Remote Sensing Images.

Moser, G., Serpico, S.B., n.d. Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery.

Zhang, W., Lu, X., Li, X., 2018. A Coarse-to-Fine Semi-Supervised Change Detection for Multispectral Images. IEEE Transactions on Geoscience and Remote Sensing 56, 3587–3599. https://doi.org/10.1109/TGRS.2018.2802785