DEEP LEARNING BASED OPTICAL FLOW ESTIMATION FOR CHANGE DETECTION: A CASE STUDY IN INDONESIA EARTHQUAKE

H. J. Qiao $^{\rm 1,\,2},$ X. Wan $^{\rm 2,\,3,\,*},$ J. Zh. Xu $^{\rm 1},$ Sh. Y. Li $^{\rm 2,\,3},$ P. P. He $^{\rm 4}$

¹ School of Remote Sensing and Information Engineering, Wuhan University, China – (qiaohj, jz_xu)@whu.edu.cn
² Technology and Engineering Center for Space Utilization, Chinese Academy of Science, China
³ Key Laboratory of Space Utilization, Chinese Academy of Sciences, China – (wanxue, shyli) @csu.ac.cn
⁴ School of Surveying and Geographic Information, North China University of Water Resources and Electric Power, China – hepei@ncwu.edu.cn

KEY WORDS: Change Detection, Natural Disaster, Optical Flow Estimation, Deep Learning, FlowNet 2.0

ABSTRACT:

Real-time change detection and analysis of natural disasters is of great importance to emergency response and disaster rescue. Recently, a number of video satellites that can record the whole process of natural disasters have been launched. These satellites capture high resolution video image sequences and provide researchers with a large number of image frames, which allows for the implementation of a rapid disaster procedure change detection approach based on deep learning. In this paper, pixel change in image sequences is estimated by optical flow based on FlowNet 2.0 for quick change detection in natural disasters. Experiments are carried out by using image frames from Digital Globe WorldView in Indonesia Earthquake took place on Sept. 28, 2018. In order to test the efficiency of FlowNet 2.0 on natural disaster dataset, 7 state-of-the-art optical flow estimation methods are compared. The experimental results show that FlowNet 2.0 is not only robust to large displacements but small displacements in natural disaster dataset. Two evaluation indicators: Root Mean Square Error (RMSE) and Mean Value are used to record the accuracy. For estimation error of RMSE, FlowNet 2.0 achieves 0.30 and 0.11 pixels in horizontal and vertical direction, respectively. The error in horizontal error is similar to other algorithms but the value in vertical direction is significantly lower than them. And the Mean Value are 1.50 and 0.09 pixels in horizontal and vertical direction, which are most close to the ground truth comparing to other algorithms. Combining the superiority of computing time, the paper proves that only the approach based on FlowNet 2.0 is able to achieve real-time change detection with higher accuracy in the case of natural disasters.

1. INTRODUCTION

Rapid detection and visualization of change in natural disaster regions are vital for swift response to rescue and relief. As one of the key technologies in disaster evaluation, change detection refers to identifying the set of pixels that are significant and possibly subtle changes between the image sequences (Fernàndez-Prieto et al., 2011). The basic principle of change detection methods takes multi-temporal images as input and outputs a binary image B, where a set of different pixels x between the pre- and post-image of the sequence would be valued according to the following generic rule: If there is a distinct change at pixel x in the last sequence, B(x) could be assigned a value of 1, otherwise, it is 0.

Change detection can be divided into two categories: appearance detection and motion detection. Appearance change includes newly built objects and destroyed objects, while motion change detection means that the appearance of the object remains the same but the position of the object has been changed. In general, changes happen continuously and gradually, therefore motion change can lead to appearance change owing to the different time intervals.

According to our literature review, most change detection studies focus on appearance change (Hussain et al., 2013; Jin et al., 2013; Tomowski et al., 2010). Because the time difference between two satellite images is usually quite long and the appearance of objects has been greatly altered in that period of time. Due to the development of video satellites and small satellites constellation technologies, the temporal resolution of remote sensing images have been greatly increased, which enables the record of whole change process during natural disasters (Toth et al., 2016) and allows us to do motion detection in these cases. Thus, it has brought a great challenge to traditional appearance change detection methods.

In this paper, the optical flow estimation method based on FlowNet 2.0, an end-to-end algorithm based on deep learning will be introduced to do motion change detection based on video data. The pixels' changes in image sequences are extracted by optical flow estimation methods from frame to frame and then a change map can be generated. Specially, this paper will focus on two key issues:

• How to implement deep learning based optical flow estimation for disaster change detection from high resolution video satellite sequence data?

• Can deep learning based optical flow estimation achieve the efficiency and accuracy for quick response in natural disasters?

2. RELATED WORK

Various motion detection methods have been studied for several years, whose main purpose is to divide the changed object from unchanged part and then track the movement. The most famous

^{*} Corresponding author

methods can be classified into background subtraction and optical flow estimation methods.

Background subtraction is a widely used real-time method for moving object detection, which generally builds an appropriate model for a reference background based on pixel distribution by averaging frames over time and then compare the model with objects in the current frame to detect the differences (Sun et al., 2006), that is to say these techniques separate the image into background (unchanged parts) and foreground (changed parts) and then segment the changed ones. Based on this idea, many adaptive background model methods have been proposed with segmentation strategies, for example, using MOG (Mixture of Gaussians) to construct the model, applying a linear predictive model in the buffer, non-parametric model, eigenvectors combining with PCA (Principal Component Analysis) and the universal ViBe (Visual Background Extractor) collecting background samples to process the background subtraction (Barnich and Van Droogenbroeck, 2010; Stauffer and Grimson, 1999; Toyama et al., 1999). These methods are simple and easy to realize, also do not require previous knowledge of moving objects such as land cover types or movements (Prajapati and Galiyawala, 2015). However, they are sensitive to the change of the so-called background so that they are difficult to discriminate changed objects from backgrounds being significantly changes. But this is most frequent in natural disaster scenes (Suresh et al., 2014). Therefore, it is difficult to put this kind of methods to use in our research.

Optical flow, which represents change of the pixels' displacement vectors between image frames, is the most widely used in motion estimation. The optical flow can be regarded as instant velocity of each pixel on imaging plane and then obtain approximately motion field which is unable to directly get from image sequences. Horn and Schunck (1981) have introduced the optical constraint equation based on the combination of velocity field and grey to build a basic algorithm of optical flow estimation(Horn and Schunck, 1981). For nearly 40 years of research, optical flow estimation methods have obtained improvements in reliability and accuracy based on the original Horn and Schunck (HS) formulation(Sun et al., 2010). Thus, it has been widely used in various areas such as, gait recognition (Lam et al., 2011), visualization of 3D cell migration (Kappe et al., 2015), reconstruction of dynamic objects in medical images (Ruymbeek et al., 2020).

Given that optical flow estimation methods have great advantages in continuous changing background, this paper have a try to adopt optical flow estimation based on deep learning to detect the motion change in natural disasters.

3. OPTICAL FLOW ESTIMATION FOR CHANGE DETECTION

Optical flow estimation, which regards the change in multiple remote sensing images of the same scene taken at different times as a movement, utilizes a new correlation to describe the ground feature change and allows to achieve both the efficiency and accuracy in critical situations (Wan et al., 2018; Ye et al., 2016).

3.1 Optical Flow Estimation

The optical flow estimates the displacement of every pixel in a sequence of images or from a video frame to another and the most common method is Horn Schunck (HS) (Horn and Schunck, 1981). The computation of the displacement in X and Y direction is as Figure 1 showed.



Figure 1. The illustration of the displacement's computation

To compute the optical flow, the following optical flow constraint must be solved as equation (1) showed:

$$I_x u + I_y v + I_t = 0 \tag{1}$$

where I_x , I_y and I_t are the spatial-temporal image brightness derivatives and u and v are the horizontal and vertical optical flow to be estimated, respectively. For HS, the optical flow is assumed to be smooth over the entire images. The method minimizes the equation (2) to compute an estimate of the velocity field $[u, v]^T$.

$$E = \iint (I_x u + I_y v + I_t)^2 dx dy + \alpha \iint \{ \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 \} dx dy$$
(2)

where $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$ are the spatial derivatives of the optical velocity component, u and α scales the global smoothness term. HS method further minimizes equation (2) to compute the velocity field [u, v], as follows:

$$u_{x,y}^{k+1} = \bar{u}_{x,y}^k - \frac{I_x[I_x\bar{u}_{x,y}^k + I_y\bar{v}_{x,y}^k + I_t]}{\alpha^2 + I_x^2 + I_y^2}$$
(3)

$$v_{x,y}^{k+1} = \bar{v}_{x,y}^k - \frac{I_y[I_x \bar{u}_{x,y}^k + I_y \bar{v}_{x,y}^k + I_t]}{\alpha^2 + I_x^2 + I_y^2}$$
(4)

In these equations, $[u_{x,y}^k, v_{x,y}^k]$ is the velocity estimate for the pixel at (x, y), and $[\bar{u}_{x,y}^k, \bar{v}_{x,y}^k]$ is the neighborhood average of $[u_{x,y}^k, v_{x,y}^k]$. For k=0, the initial velocity is 0. When achieving the estimated optical flow filed, the next step is the visualization using the color coding by Butler et al. (Butler et al., 2012). In this rule, hue is for the direction of the motion vector, the intensity of colors mean the magnitude grades of the displacement vector and white corresponds to no motion.

The majority of state-of-the-art methods are derived from the original formulation of HS. Sun et al have synthetically defined a series of baseline algorithm, named began with 'Classic', which methodically change the model and method with different techniques from the art (Sun et al., 2014). These models include Classic+NL-Fast, Classic+NL, Classic+NL-Full, Classic++, Classic-C, Classic-L and it is worthwhile to mention that these classic optical flow estimation algorithms retain powerful competitive results on the Middlebury optical flow benchmark.

3.2 FlowNet 2.0 based optical flow for change detection

Booming deep neural network technology in recent years, established approaches have shown that optical flow estimation can be naturally regarded as a supervised learning problem and can be directly solved with a simple convolutional neural

ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume V-3-2020, 2020 XXIV ISPRS Congress (2020 edition)

networks (CNN). FlowNet is the first end-to-end optical flow estimation model with CNN in 2015 (Dosovitskiy et al., 2015). It uses an encoder-decorder architecture, where the encorder module, consisting of 9 convolutional layers and ReLU (Rectified Linear Unit) active function layer, computes abstract features from receptive fields of increasing size and the decorder module, including 4 deconvolutional layer and ReLU active function layer, reestablishes the original resolution via an expanding upconvolutional architecture. The whole network, resembling fully convolutional networks (FCN), is made up of convolutional and deconvolutional layer with additional crosslinks between these contracting and expanding networks rather than any fully connected layer. However, this first new idea is not unmatched by the fine-tuned existing methods like all new ideas, which greatly limits its widespread use. In the winter of 2016, the research team has modified and advanced the network, generating an enhanced version FlowNet 2.0, as shown in Figure 2 (Ilg et al., 2017).



Figure 2. Schematic view of FlowNet 2.0 (Ilg et al., 2017)

The key improvements of FlowNet 2.0 are made in the following areas: adding multiple dataset and a learning schedule of the training dataset order, stacking two networks for flow refinement, specified network for small displacement and fusion. These contributions play an important role in the close accuracy with state-of-the-art methods while running orders of magnitude faster and the new-born network has been marked as a milestone by using CNN for optical flow estimation (Hui et al., 2018; Sheng et al., 2019).

For motion change detection during natural disasters, the input images for FlowNet 2.0 are video satellite image frames. If the area remains unchanged in some natural crisis, the pixel will have zero displacement, nevertheless, the corresponding pixels would hold none-zero values in the optical flow estimation results. Then, the displacement data from FlowNet 2.0 would be divided into changed and unchanged part based on Otsu (Khan and Communication, 2014; Vala et al., 2013).

4. EXPERIMENTS AND RESULTS

4.1 Data

The study area is located in Petobo, Indonesia where a 7.5 magnitude earthquake trigged a tsunami that killed 1600 people and destroyed more than 70,000 homes on Sept.28, 2018. Digital

Globe's WorldView captured these before and after satellite images and transformed into a video (Digital, 2018). The video gives a glimpse at the damage in the worst-hit area where soil liquefaction causes the ground to boil. 58 frames of the video have been extracted as the input image sequence for change detection and visualization. In the following sections, we will present the optical flow estimation result and change map taking two frames as an example.

4.2 Flow field visualization

In this section, the selected image frames are tested based on classic optical flow estimation methods and FlowNet 2.0. The requirement of using FlowNet 2.0 is the input image size must be an integer multiple of 64, thus a subset of 1920 x 1024 pixel are clipped (the original image size is 1920 x 1080 pixel). Afterwards, the optical flow of the image sequence can be generated by different optical flow estimation methods as shown in Figure 3. In order to demonstrate the detailed difference between the used optical flow estimation methods in Figure 3, one area, shown using a red box in Figure 3(a), is extracted and compared in Figure 4. From Figure 3 and 4, we can get that optical flow estimation results generated by most of the classic optical flow methods are not satisfactory. However, the result based on FlowNet 2.0 has a clear advantage in producing smooth flow fields with crisp motion boundaries and is enough robust to show motion blur, which can satisfy the goals of change detection.



ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume V-3-2020, 2020 XXIV ISPRS Congress (2020 edition)



Figure 3. Initial flow fields from optical flow estimation methods on images from Digital Globe's WorldView



Figure 4. Zoom-in areas of the initial optical flow fields based on the red box in Figure 3(a)

4.3 Accuracy comparison

To evaluate the accuracy of the optical flow estimation results, 105 corresponding feature points i have been manually selected as shown in Figure 5, and their image coordinates XY, X'Y' are recorded to generate the ground truth data. The ground truth data U, V for displacements are generated by the difference of the corresponding points' coordinate:

$$U_i = X_i - X_i' \tag{5}$$

$$V_i = Y_i - Y_i' \tag{6}$$

U and V represent horizontal and vertical displacement of each feature point and the mean U and V of all feature points is 1.53 and 0.05 pixel, respectively.



a. Frame 1



Figure 5. Profiles of feature points on experimental images

Meanwhile, different optical flow estimation results generated by 7 state-of-the-art methods and FlowNet 2.0 are compared by using two indexes of RMSE and Mean Value and the results are shown in Table 1. A comparison of RMSE data in horizontal and vertical direction shows that RMSE values in vertical direction of all optical flow estimation methods are significantly lower than their counterparts along the horizontal direction, wherein the values are at least 0.1 lower than their horizontal counterparts. Along the horizontal direction, all the RMSE of the classic optical flow estimation methods have values of 0.29, while the FlowNet 2.0 method has a horizontal RMSE value of 0.30 and the basic HS method has a slightly higher horizontal RMSE value of 0.31. The vertical RMSE of the optical flow estimation methods have values that range from 0.11 to 0.18, where FlowNet 2.0 has the lowest value of 0.11 and the HS method has

the highest value of 0.18. Generally, the basic HS method has both the highest horizontal and vertical RMSE value, therefore it is the least accurate optical flow estimation method. Meanwhile, the FlowNet 2.0 method is just as accurate as, if not more accurate than, the classic optical flow estimation methods that are studied in this paper because it has the lowest vertical RMSE value and its horizontal RMSE value is similar to the values of the other classic optical flow estimation methods. The Mean Values of the 105 corresponding feature points are 1.50 and 0.05 pixels in horizontal and vertical direction, respectively. From Table 1, the Mean Value calculated by FlowNet 2.0 is approximately similar to other optical estimation methods in horizontal direction, however, is the most close to the ground truth in vertical direction. The advantage of optical flow estimation results based on FlowNet 2.0 benefits from the introducing a subnetwork donated by FlowNet2-CSS-ft-sd specializing on small displacement or subpixel motions, which does not lose performance on large displacements at all. Thus FlowNet 2.0, for the natural disaster video data, can reach optimal performance on arbitrary displacements and can be compared to other state-of-the-art classic optical flow estimation algorithms.

	DMCE		Maan Valaa	
Method	RMSE		Niean Value	
	Х	Y	Х	Y
	direction	direction	direction	direction
HS	0.31	0.18	1.50	0.15
Classic+NL- Fast	0.29	0.13	1.49	0.10
Classic+NL	0.29	0.12	1.49	0.10
Classic+NL- Full	0.29	0.12	1.49	0.10
Classic++	0.29	0.14	1.50	0.12
Classic-C	0.29	0.12	1.49	0.10
Classic-L	0.29	0.15	1.50	0.13
FlowNet 2.0	0.30	0.11	1.50	0.09

Table 1. The comparison of RMSE and Mean Value for different optical flow estimation methods (unit is pixel)

4.4 Running time comparison

Except for the accuracy, the time complexity is also important for real-time change detection of natural disasters. In the experiments, the mean operating time are recorded in Table 2, which shows that FlowNet 2.0 using CNN for optical flow estimation runs orders of magnitude faster than the other classic methods. Benefiting from the deep learning and GPU, FlowNet 2.0 is the ideal optical flow estimation algorithm in this research.

Method	Average Running time(s/f)	Operation environment	
HS	4052.63		
Classic+NL-Fast	4871.98		
Classic+NL	5852.56	Intel Xeon(R)	
Classic+NL-Full	9565.48	CPU E5-2407	
Classic++	3608.74	96G	
Classic-C	3765.46		
Classic-L	3356.96		
FlowNet 2.0	12.04	TITAN X (Pascal) Mem 100G	

Table 2. Speed comparison for all methods

4.5 Binary image

Although the optical flow estimation result based on FlowNet 2.0 is clear to distinguish the changed and unchanged area, we should finish the final result according to the basic rule of change detection. Firstly, the actual displacement based on the displacements on horizontal and vertical direction from FlowNet 2.0 can be calculated by the following equation:

$$\mathbf{D} = \sqrt{U^2 + V^2} \tag{7}$$

Then OTSU, a typical global threshold selection method, is used to evaluate the optimal threshold for dividing the changed and unchanged part. Afterwards, binarization processing of images is processed based on the rule, namely the changed part is valued by 1 and the other is given the value of 0 as showed in Figure 6.



Figure 6. Image binarization for change detection

5. CONCLUSION

In this paper, a deep learning based optical flow estimation method, FlowNet 2.0, is used for change detection in 2018 Indonesia Earthquake using image sequence from Digital Globe WorldView. Comparison experiments are carried out using FlowNet 2.0 and the 7 typical optical flow estimation methods, wherein the results show that only FlowNet 2.0 can achieve change detection in real time while maintaining equivalent optical flow estimation accuracy. Based on the optical flow estimation results of all video sequence, it is possible to track the dynamic change of the whole crisis area or the motion of some critical buildings and further evaluate the anti-risk capability in certain natural disasters.

ACKNOWLEDGEMENTS

The work is supported by National Key Research and Development Program of China (Grant Nos. 2018YFD1100405) and National Natural Science Foundation of China (Grant Nos. 41901285).

REFERENCES

Barnich, O., Van Droogenbroeck, M.J.I.T.o.I.p., 2010. ViBe: A universal background subtraction algorithm for video sequences. 20, 1709-1724.

Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J., 2012. A naturalistic open source movie for optical flow evaluation, European Conference on Computer Vision. Springer, pp. 611-625.

Digital, G., 2018. Digital Globe Data in INdonesia Earthquake Available on Youtube: Https://www.youtube.com/watch?v=-41ENJF0wVwx

Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., Brox, T., 2015. Flownet: Learning optical flow with convolutional networks, Proceedings of the IEEE International Conference on Computer Vision, pp. 2758-2766.

Fernàndez-Prieto, D., Marconcini, M.J.I.T.o.G., Sensing, R., 2011. A novel partially supervised approach to targeted change detection. 49, 5016-5038.

Horn, B.K., Schunck, B.G.J.A.i., 1981. Determining optical flow. 17, 185-203.

Hui, T.-W., Tang, X., Change Loy, C., 2018. Liteflownet: A lightweight convolutional neural network for optical flow estimation, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8981-8989.

Hussain, M., Chen, D., Cheng, A., Wei, H., Stanley, D.J.I.J.o.p., sensing, r., 2013. Change detection from remotely sensed images: From pixel-based to object-based approaches. 80, 91-106.

Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T., 2017. Flownet 2.0: Evolution of optical flow estimation with deep networks, IEEE conference on computer vision and pattern recognition (CVPR), p. 6.

Jin, S., Yang, L., Danielson, P., Homer, C., Fry, J., Xian, G.J.R.S.o.E., 2013. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. 132, 159-175.

Kappe, C.P., Schütz, L., Gunther, S., Hufnagel, L., Lemke, S., Leitte, H.J.I.t.o.v., graphics, c., 2015. Reconstruction and visualization of coordinated 3D cell migration based on optical flow. 22, 995-1004.

Khan, M.W.J.I.J.o.F.C., Communication, 2014. A survey: image segmentation techniques. 3, 89.

Lam, T.H., Cheung, K.H., Liu, J.N.J.P.r., 2011. Gait flow image: A silhouette-based gait representation for human identification. 44, 973-987.

Prajapati, D., Galiyawala, H.J.J.I.J.o.C.A., 2015. A review on moving object detection and tracking. 5, 168-175.

Ruymbeek, K., Vanroose, W.J.J.o.C., Mathematics, A., 2020. Algorithm for the reconstruction of dynamic objects in CTscanning using optical flow. 367, 112459.

Sheng, L., Pan, J., Guo, J., Shao, J., Wang, X., Loy, C.C.J.a.p.a., 2019. Unsupervised Bi-directional Flow-based Video Generation from one Snapshot.

Stauffer, C., Grimson, W.E.L., 1999. Adaptive background mixture models for real-time tracking, Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149). IEEE, pp. 246-252.

Sun, D., Roth, S., Black, M.J., 2010. Secrets of optical flow estimation and their principles, 2010 IEEE computer society

conference on computer vision and pattern recognition. IEEE, pp. 2432-2439.

Sun, D., Roth, S., Black, M.J.J.I.J.o.C.V., 2014. A quantitative analysis of current practices in optical flow estimation and the principles behind them. 106, 115-137.

Sun, J., Zhang, W., Tang, X., Shum, H.-Y., 2006. Background cut, European Conference on Computer Vision. Springer, pp. 628-641.

Suresh, S., Deepak, P., Chitra, K., 2014. An efficient low cost background subtraction method to extract foreground object during human tracking, 2014 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2014]. IEEE, pp. 1432-1436.

Tomowski, D., Klonus, S., Ehlers, M., Michel, U., Reinartz, P., 2010. Change visualization through a texture-based analysis approach for disaster applications. na.

Toth, C., Jóźków, G.J.I.J.o.P., Sensing, R., 2016. Remote sensing platforms and sensors: A survey. 115, 22-36.

Toyama, K., Krumm, J., Brumitt, B., Meyers, B., 1999. Wallflower: Principles and practice of background maintenance, Proceedings of the seventh IEEE international conference on computer vision. IEEE, pp. 255-261.

Vala, H.J., Baxi, A.J.I.J.o.A.R.i.C.E., Technology, 2013. A review on Otsu image segmentation algorithm. 2, 387-389.

Wan, X., Liu, J., Li, S., Dawson, J., Yan, H.J.I.T.o.G., Sensing, R., 2018. An Illumination-Invariant Change Detection Method Based on Disparity Saliency Map for Multitemporal Optical Remotely Sensed Images. 1-14.

Ye, S., Chen, D., Yu, J.J.I.J.o.P., Sensing, R., 2016. A targeted change-detection procedure by combining change vector analysis and post-classification approach. 114, 115-124.