Real-time Detection of Wearable Camera Motion Using Optical Flow

Ola Younis*, Waleed Al-Nuaimy*, Fiona Rowe† and Mohammad H. Alomari*,‡
*Department of Electrical Engineering and Electronics
University of Liverpool, Liverpool, United Kingdom
†Institute of Psychology Health and Society
University of Liverpool, Liverpool, United Kingdom
‡Electrical Engineering Department
Applied Science Private University, Amman, Jordan

***Abstract*—The efficient use of image sensors has been one of the top challenges for computer vision researchers for several years. Detecting and tracking objects, video surveillance, navigation, and many other real-time applications depend on motion estimation for moving camera. In this paper, a real-time method for the detection and classification of the motion of a wearable, moving monocular camera is proposed. This approach was adopted to be used with smart glasses to assist people with visual field defects. Five main motion classes (corresponding to the five primary degrees of freedom) were detected using optical flow and motion velocity vectors calculation. These classes cover different degrees of freedom including rotation and translation. The proposed method classifies the type of camera motion as static, translation/rotation left, right, up or down. This classification is important for object detection and tracking that can alert the user to the potential hazards outside their field of view.**

The proposed approach has been tested on a real first-person perspective video captured by a wearable camera. The experimental results demonstrate that the proposed method classifies the type of motion successfully in real-time and can be used as part of low-cost wearable solutions for various forms of vision loss assistive technologies.Promising performance results of 84% correct states for camera motion detection were obtained.

***Index Terms*—Camera motion detection, optical flow, assistive technology, wearable camera, vision loss, egocentric vision**

I. Introduction

Healthcare systems are developing significantly due to the huge jump in wearable technologies and smart embedded systems. These technologies are designed to cover several body parts such as head, chest, fingers, feet, and ears. Vision defect assistive technologies are a special type of these healthcare systems to help people with various degrees of vision loss to improve the quality of their life [1]. The development of head-mounted cameras (HMC), or wearable cameras, and data processing functionalities allows computer vision algorithms to perform real-time processing of video streams and still images for object detection, object tracking, visual augmentation and many other applications. Although these algorithms have helped in designing reliable and accurate assistive smart systems, new challenges have arisen in terms of egocentric video analysis and processing [2].

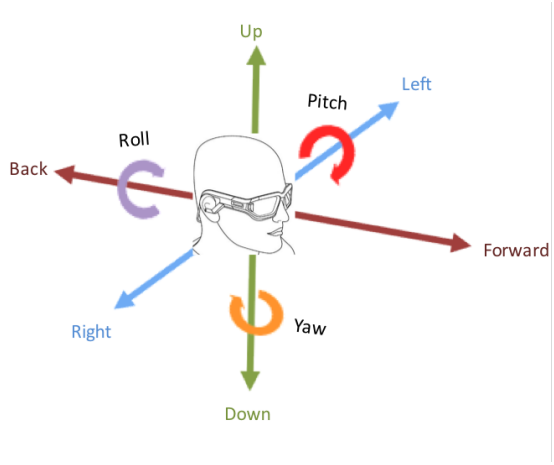

Fig. 1. Degrees of freedom for wearable camera

Egocentric vision (or first perspective person vision) is a new field of computer vision and machine learning algorithms which refers to information extraction from visual data in order to enhance human visual field by adding new computer-generated information into it [3]. Several types of applications have been developed to be applied on egocentric vision like object detection [4], visual field extension [5], social interaction analysis [6] and others [7].

One of the main challenges in this field is the camera movement itself. Because the camera is wearable, its movement is random and unpredictable. Therefore, head (or camera) motion detection is a key step before starting other video processing procedures [8]. Camera motion type detection phase is important to decide if the camera is moving or not (stationary) and detects the motion type. The output of this will be used in object detection, object tracking, and many other computer vision applications. In the case of a wearable camera, six degrees of freedom are expected based on head movements as shown in Figure 1.

Considerable research work has been developed to detect camera motion and determine its type over the past years. LK optical flow method was used by Wang et al. [9] to

present an analysis model of video shot motion with the combination of a segmentation method and feature extraction. Several motion types were analyzed like: push, pull, follow, rejection, shift, shake and others. Starting by extracting video frames and applying Canny edge detection, frame subtraction was performed to divide the video into many shots. The following step is to calculate motion trajectory and velocity using LK optical flow to classify the shot motion type.

Narayana et al. [10] proposed a segmentation method for clustering pixels with similar real-world motion regardless of their depth in the scene. Their work is capable of defining the number of foreground motions automatically. Based on the assumption that if the camera is translating, optical flow orientations will be independent of object depth, the authors build their work using the calculation of these orientations instead of the whole motion vectors. Their main drawback is that they focus only on translation and are liable to suffer from errors when the camera rotates.

Erdem et al. [11] addressed the problem of camera ego-motion estimation using a feature-based approach. Using feature selection helped in lowering the computational complexity since the number of key-points is limited. The experimental results demonstrated the potential usages of their approach.

Nguyen et al. [8] extended the work which in [11] by introducing a powerful method for detecting moving video camera. The purpose of their work was to analyze objects in video streams in the movie industry. Using explicit analyses of optical flow vectors and the idea of sub-regions, the authors calculated magnitude and direction for these sub-regions for camera motion classifications.

Wearable cameras introduced new challenges for egocentric vision and camera motion detection techniques for object detection and tracking in particular. This is due to the complex motion changes within an unstructured environment and moving background scenes. In this case, moving object detection and tracking algorithms should address the problems of occlusion, big moving objects, scaling, unstructured and moving background and many other issues in addition to compensating for motion of the wearable camera. All these approaches need to keep track of the camera motion while operating in order to update the moving scene and compensate camera motion to differentiate between dependent moving objects (objects that appear to be moving due to the user motion) and independent moving objects (the ones which are really moving). These challenges have not yet received adequate coverage in the literature and need further attention.

In our work, we tried to solve the egocentric vision camera motion detection using the optical flow calculation. We use fixed location key-points close to frame corners to avoid moving objects error assuming that most of the motion displacement appears in the inner part of the frame.

Since this is a part of a larger project for hazard detection in smart assistive technology, camera motion detection result is not required for each frame. Therefore, we calculated the average of the motion type for each half a second. This filters out errors that occur between consecutive frames. Then, to test our method, we used real street video captured by a head-mounted camera.

Fig. 2. Orcam MyEye 2.0 wearable camera [16]

II. Wearable cameras

Wearable computers and head-mounted display devices and technologies are steadily gaining publicity. Smart glasses in specific are more popular due to their entertainment goals and techniques [12]. Younis et al. suggested the use of augmented reality concepts to help people with visual field defects using smart glasses [13] [14].

Starting with the announcement of Google glass™ in 2012, several smart glasses were developed to use both artificial and augmented reality concepts. The first scientific review for the clinical and surgical applications of smart glasses in healthcare systems was presented in [15].

OrCam MyEye™ 2.0 [16] is a smart technology that uses computer vision algorithms with the addition of wearable platforms to help people with vision problems [16]. Their main goals are to improve individuals independence and help visually impaired people to read by themselves. The design is very simple, lightweight, efficient and could be clipped onto a pair of glasses. Utilizing any surface, the attached camera can read text instantly using the person's gesture and generates a loud voice using a small speaker for the user. The system also can recognize faces, products, and money notes in real time. Figure 2 shows the design for Orcam.

Much like Microsoft HoloLens™ [17], Daqri™ smart glasses [18] implement augmented reality concepts in manufacturing, medical remote experts, field services, maintenance, and repair sectors. Their design is powerful and consists of a wearable head helmet that contains a video camera, a display unit and Mini portable computer including Intel core m7 processor as shown in Figure 3.

Fig. 3. Daqri smart glasses [18]

Fig. 4. EyeTrek Insight EI-10 glassess attachment [19]

The EyeTrek Insight EI-10™ [19] is the latest generation of Olympus optical solutions for smart glasses. Inspired by the Google glasses [20] design, the tiny display unit superimposes the user's field of view by computer-generated information without blocking normal vision. The difference between this design and the Google glasses design is that EI-10 can easily be connected to any pair of normal eyeglasses. Its lightweight, powerful operating system, and efficient display unit make it a good choice for business applications as shown in Figure 4.

Table I shows a comparison between the mentioned smart glasses specifications. Although we provide a quick review about the some of the available smart glasses in the market currently, it is important to mention that there are several other options such as Moverio BT-300™ [21], Vuzix M300™ [22] and others.

A. Proposed system

In the case of wearable cameras and smart glasses in particular, camera motion is often synonymous with head motion. The head can move in a forward/backward, left/right and up/down translation, although rotation is more common. In terms of rotation, pitch motion represents the rotation around the x-axis, yaw rotation is a movement around the y-axis and finally, a roll is a rotation around the z-axis as shown in Figure1. In our project, we will cover all translation motion types (left/right, up/down and forward/backward). Pitch rotation is considered to be similar to up/down type, and yaw rotation is

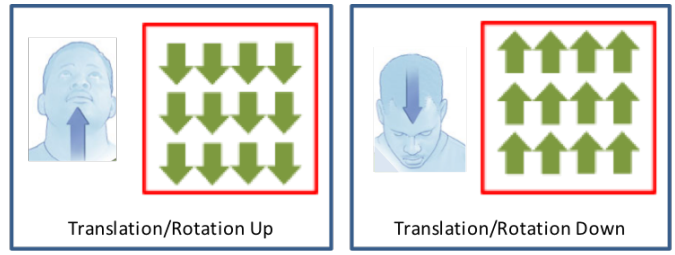

Fig. 5. Two movement types example showing up and down panes

considered to be the same as left/right motion. In this paper, we are presenting all mentioned degrees of freedom except forward/backward translation and roll rotation which will be presented in our future work. Figure 5 demonstrates graphical representations for some of the mentioned motion types. In general, camera motion can be summarised as:

1) Stationary camera (S): static background, moving objects.
2) Translation/Rotation Right (TRR), Moving Translation/Rotation Left (TRL): background change in horizontal direction.
3) Translation/Rotation Up (TRU), Moving Translation/Rotation Down (TRD): background change in vertical direction.
4) Moving Forward (MF) or Moving Backward (MB) : fast changes in the background and foreground.

Optical flow is one of the motion estimation techniques used to compute the motion of the pixels of an image sequence or video stream. It shows the real displacements of the objects in terms of velocity and direction (motion vector) [23].

A wide range of computer vision application depends on optical flow calculations for real-time processes such as motion estimation [24], video processing [25], moving object detection [26], camera motion classification [27] and other applications. The main aim is to compare the location of the pixel/region of pixels between two consecutive frames/images to compute the motion vector for that pixel/region.

Figure 5 shows the optical flow for up and down panes. In this figure, the red rectangle represents the video frame and the black arrow represents the direction of movement.

In this work, we address the problem of detecting the type of motion for a wearable, free moving monocular camera. This is part of a larger project to develop a smart assistive technology system for people with visual field defects that involves object detection and tracking, hazard classification and alert displays using smart glasses.

One of our main goals behind the development of this wearable assistive technology is to detect moving objects while the camera is moving. Hence, camera motion type will determine the best detection algorithm to be used. In the case of the stationary camera, moving objects can be detected using

TABLE I
COMPARISON OF 3 SMART GLASSES AVAILABLE ON THE MARKET: ORCAM MYEYE 2.0 [16], DAQRI [18] AND EYETREK INSIGHT EI-10 [19]

Product Name	Functionality	Sensors	Processor and operating system	Battery	Price	Output	Notes
Olympus	business applications: Logistics, field services and maintenance and task management	- Forward Camera touch-sensor bar	Android 4.2 Open-source	30-60 minutes	$1,500	1 x 640 x 400 OLED	Attached to normal glasses
Daqri	Solve complex problems for manufacturing, field service and maintenances	Camera (44-degree VF) Depth sensing HD colour camera	Intel core m7, 3.1 GH 90 frame/second	4 hours	$ 4,995	1360 x 768 screen	Weight: 0.7 pound
Orcam	Text-object recognition	Camera Microphone	i.MX 6 quad	4 hours	£2,600	Bone-conduction earpiece	Figure gesture Not open-source

traditional foreground segmentation methods. In the case of moving camera, a motion compensation step is needed before background subtraction in order to distinguish between real and artificial movement. Finally, forward/backward moving camera scenarios need more advanced motion estimation and compensation algorithms before the object detection phase which will be addressed in our future work.

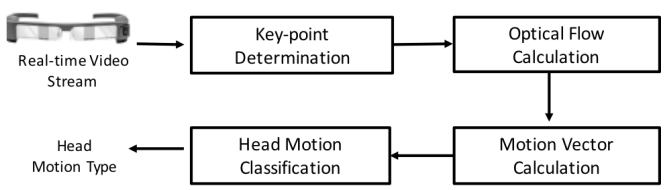

Fig. 6. Block diagram for the proposed system

Figure 6 shows the overall work-flow for the proposed method. It consists of four main steps starting by defining key-points in the first frame (step 1) to be used in optical flow calculations (step 2). Motion vectors will be extracted from step 2 to calculate velocity and direction in step 3. Finally, the final camera motion type will be classified based on the average motion vectors using half of the frame rate (step 4). Further details will be presented in the following sections.

III. OPTICAL FLOW AND CAMERA MOTION CLASSIFICATION

The first step is to define a set of key-points in the previous frame I_{t-1} and looking for their corresponding location in the current frame I_t using optical flow methods. A dense feature optical flow method was applied using (10p x 10p) sliding window. Figure 7 (c) shows blue lines that represent the calculated optical flows at each pixel using (10p x 10p) sliding window. It is clear that the camera rotated to the left from the uniform blue right arrow lines. The irregular optical

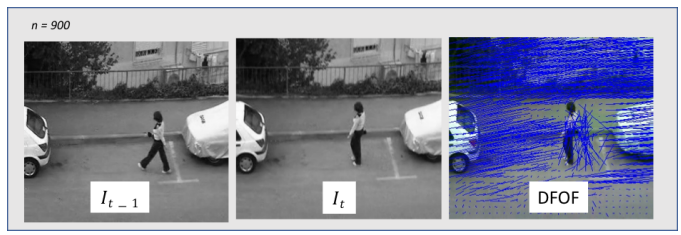

Fig. 7. Dense feature optical flow (DFOF) example. (a) the previous frame (b) the current frame and (c) DFOF optical flows.

flow directions around the lady in the middle of the image represent moving object.

To lower the computation load, different values for the number of key-points (N) were tested and the points were defined manually and located at the frame corners (10%-30% from boundaries). The minimum value for N is set to 4 to represent the possible motion direction as shown in Figure 8. The red lines in Figure 8 (c) represent the calculated optical flows at the four corners. The magnitude of motion of the detected points are represented by the length of the red line and the direction of motion for each point is represented by the direction of the red line. The trade-off between computation speed and the global optical flow result accuracy should determine the number of detected corners.

The Lucas-Kanade with pyramid method [28] was used to repeatedly calculate the optical flow for sparse features over a time window T where $T = FR/2$ (half the video frame rate). The magnitude ($|V_n|$) and direction (θ_n) for the key-points velocity (V_n) were calculated for each frame then averaged per frame to find ($|V|$) and (θ). Then (V) values were averaged over T to determine the current camera motion type (CMT).

Given

$$|V| = \frac{\sum_{n=1}^{N} |V_n|}{N} \tag{1}$$

and

$$\theta = \frac{\sum_{n=1}^{N} |\theta_n|}{N} \tag{2}$$

then, the CMT is considered stationary (S) at time t if the mean over T of the average velocity of key-points motion V_n satisfies the condition:

$$\frac{\sum_{t=1}^{T} |V|}{T} < V_{th} \tag{3}$$

where V_{th} is a pre-defined velocity threshold, $t = 1, 2, ..., T$(frames), and $n = 1, 2, ..., N$(key-points).

If the velocity exceeds V_{th}, the camera is considered to be moving. Thus, a motion type should be calculated. If the key-points in the top half of the frames are considered, we assumed that translation/rotation happens when the direction of motion (θ) of the detected points falls along the horizontal axis between -45°to 45°for TRL or between -135°to 135°for TRR.

For TRU class, we assumed that the direction of motion (θ) of the detected points falls along the vertical axis between -45°to -135°. Finally, we assumed that TRD class happens when the direction of motion (θ) of the detected points fall along the vertical axis between 45°and 135°.

Therefore, the CMT at time t can be determined as following:

$$\begin{aligned} TRR: &\quad \frac{\sum_{t=1}^{T} |V|}{T} > V_{th} \quad AND \quad \frac{\sum_{t=1}^{T} \theta}{T} > \theta_1 \\ TRL: &\quad \frac{\sum_{t=1}^{T} |V|}{T} > V_{th} \quad AND \quad \frac{\sum_{t=1}^{T} \theta}{T} < \theta_2 \end{aligned} \tag{4}$$

In the case of moving up and down, we defined two directions of motion using the whole set of key-points as following:

$$\begin{aligned} TRU: &\quad \frac{\sum_{t=1}^{T} |V|}{T} > V_{th} \quad AND \quad \theta_3 < \frac{\sum_{t=1}^{T} \theta}{T} < \theta_4 \\ TRD: &\quad \frac{\sum_{t=1}^{T} |V|}{T} > V_{th} \quad AND \quad \theta_5 < \frac{\sum_{t=1}^{T} \theta}{T} < \theta_6 \end{aligned} \tag{5}$$

Fig. 8. Sparse features optical flow (SFOF). (a) the previous frame. (b) current frame and (c) SFOF optical flows. Note: Magnitude was magnified for demonstration purpose only.

The described algorithm for camera motion type detection can be summarised as follows:

1) Calculate the average magnitude $|V_n|$ for every N key-points in each frame.
2) Calculate the average magnitude $|V_n|$ over T frames.
3) If $|V_n|$ is greater than or equal to a predefined threshold V_{th}, proceed to step 4.
4) Check for motion type based on equations (4) and (5).

These steps will be performed at the beginning of each second of the video stream. Half of the frame rate will be used in this algorithm to calculate the CMT which will determine the best motion compensation algorithm to be used in the subsequent object detection phase.

IV. Evaluation and System Output

A 4K street view video [29] was used to represent the actual environment around people while walking in streets. Our system was applied to 9616 frames in this video to test the performance of the CMT detection method. This video has a frame rate of 30 Fps and has been chosen due to its strong correlation with a normal situation of a person who is walking in a street with the existence of other pedestrians, moving cars and objects. Five different values for V_{th} (10, 20, 40, 60 and 70 pixel/frame) have been tested using four different sets of key-points N (4, 8, 16 and 36). All these scenarios were compared with the actual CMT detected by a personal observer to evaluate the performance of the proposed method as shown in Table II.

Based on these results, it is found that the best camera motion type classification accuracy of 84% can be achieved with V_{th}=20 and N=16. These configurations were used to implement our system and test it with many videos for a moving camera.

A. Conclusion

In this work, optical flow was used to detect and classify the motion of a wearable camera in real time. Camera motion type (CMT) was classified as either stationary, moving left, moving right, moving up and moving down. A wearable, free-moving monocular camera was used in this work as one of several smart glasses modules that will be used in future work in developing smart assistive technology to help people suffering vision loss to avoid hazards while walking. The output of this will be used in moving object detection and tracking in addition to the reduction of the required processing loads. Promising performance results of 84% correct states for CMT detection were obtained.

References

[1] V. Pradeep, G. Medioni, J. Weiland, Robot vision for the visually impaired, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, IEEE, 2010, pp. 15–22.

TABLE II
THE PERFORMANCE ACCURACY FOR THE CAMERA MOTION STATE SYSTEM WITH DIFFERENT SPEED THRESHOLDS AND NUMBER OF KEY-POINTS.

V_{th}	N=4	N=8	N=16	N=36
10	83%	83%	84%	81%
20	83%	82%	84%	81%
40	79%	78%	79%	74%
60	72%	68%	70%	63%
70	69%	65%	67%	58%

[2] B. F. Katz, S. Kammoun, G. Parseihian, O. Gutierrez, A. Brilhault, M. Auvray, P. Truillet, M. Denis, S. Thorpe, C. Jouffrais, Navig: Augmented reality guidance system for the visually impaired, Virtual Reality 16 (4) (2012) 253–269.

[3] S. Alletto, G. Serra, R. Cucchiara, Egocentric object tracking: an odometry-based solution, in: International Conference on Image Analysis and Processing, Springer, 2015, pp. 687–696.

[4] M. Funk, R. Boldt, B. Pfleging, M. Pfeiffer, N. Henze, A. Schmidt, Representing indoor location of objects on wearable computers with head-mounted displays, in: Proceedings of the 5th Augmented Human International Conference, AH '14, ACM, New York, NY, USA, 2014, pp. 18:1–18:4. doi:10.1145/2582051.2582069.

[5] K. Fan, J. Huber, S. Nanayakkara, M. Inami, Spidervision: Extending the human field of view for augmented awareness, in: Proceedings of the 5th Augmented Human International Conference, AH '14, ACM, New York, NY, USA, 2014, pp. 49:1–49:8. doi:10.1145/2582051.2582100.

[6] A. Fathi, J. K. Hodgins, J. M. Rehg, Social interactions: A first-person perspective, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 1226–1233. doi:10.1109/CVPR.2012.6247805.

[7] C. Forster, M. Pizzoli, D. Scaramuzza, Svo: Fast semi-direct monocular visual odometry, in: 2014 IEEE International Conference on Robotics and Automation (ICRA), 2014, pp. 15–22. doi:10.1109/ICRA.2014.6906584.

[8] N.-T. Nguyen, D. Laurendeau, A. Branzan-Albu, A robust method for camera motion estimation in movies based on optical flow, International Journal of Intelligent Systems Technologies and Applications 9 (3-4) (2010) 228–238. doi:10.1504/IJISTA.2010.036578.

[9] P. Wang, C. Chen, C. Dong, H. Xu, F. Tian, The analysis method of video camera's motion based on optical flow and slam, in: 2016 International Conference on Audio, Language and Image Processing (ICALIP), 2016, pp. 62–66. doi:10.1109/ICALIP.2016.7846577.

[10] M. Narayana, A. Hanson, E. Learned-Miller, Coherent motion segmentation in moving camera videos using optical flow orientations, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 1577–1584.

[11] Y. S. Erdem, F. Galip, I. F. Ince, M. H. Sharif, Estimation of camera ego-motion for real-time computer vision applications, International Journal of Scientific Research in Information Systems and Engineering (IJSRISE) 1 (2).

[12] W. Barfield, Fundamentals of Wearable Computers and Augmented Reality, CRC Press, 2015.

[13] A. Al-Ataby, O. Younis, W. Al-Nuaimy, M. Al-Taee, Z. Sharaf, B. Al-Bander, Visual augmentation glasses for people with impaired vision, in: Developments in eSystems Engineering (DeSE), 2016 9th International Conference on, IEEE, 2016, pp. 24–28.

[14] O. Younis, W. Al-Nuaimy, M. A. Al-Taee, A. Al-Ataby, Augmented and virtual reality approaches to help with peripheral vision loss, in: 2017 14th International Multi-Conference on Systems, Signals Devices (SSD), 2017, pp. 303–307. doi:10.1109/SSD.2017.8166993.

[15] S. Mitrasinovic, E. Camacho, N. Trivedi, J. Logan, C. Campbell, R. Zilinyi, B. Lieber, E. Bruce, B. Taylor, D. Martineau, et al., Clinical and surgical applications of smart glasses, Technology and Health Care 23 (4) (2015) 381–401.

[16] My eye 2.0, Orcam, last accessed: 2018-1-15.
URL https://www.orcam.com/en/

[17] Microsoft hololens, Microsoft Corporation, last accessed: 2018-1-15.
URL https://www.microsoft.com/en-gb/hololens

[18] Daqri smart glass, Daqri, last accessed: 2018-1-15.
URL https://daqri.com/products/smart-glasses/

[19] Eyetrek insight, Olympus Corporation, last accessed: 2018-1-15.
URL http://www.getolympus.com/smartglasses?ref=CJs

[20] Glass 2.0, X Company, last accessed: 2018-1-15.
URL https://www.x.company/glass/

[21] Moverio BT-300, Epson Seiko Epson Corporation, last accessed: 2018-1-15.
URL https://www.epson.co.uk

[22] Vuzix m300, Vuzix Firm, last accessed: 2018-1-15.
URL https://www.vuzix.com/

[23] D. Fortun, P. Bouthemy, C. Kervrann, Optical Flow Modeling and Computation: a Survey, Computer Vision and Image Understanding 134 (2015) 1–21.

[24] T. Brox, J. Malik, Large displacement optical flow: Descriptor matching in variational motion estimation, IEEE transactions on pattern analysis and machine intelligence 33 (3) (2011) 500–513.

[25] A. Plyer, G. Le Besnerais, F. Champagnat, Massively parallel lucas kanade optical flow for real-time video processing applications, Journal of Real-Time Image Processing 11 (4) (2016) 713–730.

[26] L. Sevilla-Lara, D. Sun, V. Jampani, M. J. Black, Optical flow with semantic segmentation and localized layers, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3889–3898.

[27] D. D. Doyle, A. L. Jennings, J. T. Black, Real-time camera tracking system using optical flow feature points, uS Patent 9,696,404 (2017).

[28] B. D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'81, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1981, pp. 674–679.

[29] 4k street view, korea, iLuvTech YouTube channel, last accessed: 18 Nov 2017.
URL https://youtu.be/qA2W4hLh6Gc