

Spatial Multilevel Optical Flow Architecture-based Dynamic Motion Estimation in Vehicular Traffic Scenarios

Alvaro Fuentes¹, Sook Yoon², and Dong Sun Park^{1,3*}

¹Department of Electronics Engineering, Chonbuk National University, Jeonju, Republic of Korea
[e-mail: afuentes@jbnu.ac.kr] [e-mail: dspark@jbnu.ac.kr]

²Department of Computer Engineering, Mokpo National University, Jeonnam, Republic of Korea
[e-mail: syoon@mokpo.ac.kr]

³College of Computer Science and Information Engineering, Tianjin University of Science and Technology,
Tianjin, China

*Corresponding author: Dong Sun Park

*Received July 27, 2017; revised November 30, 2017; revised February 20, 2018; revised April 4, 2018;
accepted June 25, 2018; published December 31, 2018*

Abstract

Pedestrian detection is a challenging area in the intelligent vehicles domain. During the last years, many works have been proposed to efficiently detect motion in images. However, the problem becomes more complex when it comes to detecting moving areas while the vehicle is also moving. This paper presents a variational optical flow-based method for motion estimation in vehicular traffic scenarios. We introduce a framework for detecting motion areas with small and large displacements by computing optical flow using a multilevel architecture. The flow field is estimated at the shortest level and then successively computed until the largest level. We include a filtering parameter and a warping process using bicubic interpolation to combine the intermediate flow fields computed at each level during optimization to gain better performance. Furthermore, we find that by including a penalization function, our system is able to effectively reduce the presence of outliers and deal with all expected circumstances in real scenes. Experimental results are performed on various image sequences from Daimler Pedestrian Dataset that includes urban traffic scenarios. Our evaluation demonstrates that despite the complexity of the evaluated scenes, the motion areas with both moving and static camera can be effectively identified.

Keywords: Motion estimation, pedestrian detection, traffic scenarios, optical flow, intelligent vehicles.

This work was supported by the Brain Korea 21 PLUS Project, National Research Foundation of Korea. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2016R1D1A1B03936269).

1. Introduction

Human vision allows us to sense the environment surrounding us and posteriorly understand it by processing the information that it is contained in the visible light. Computer Vision aims to replicate this principle by using cameras to catch information from the environment and process it. Pedestrian detection is an important approach in the computer vision area. It has several applications such as detection and tracking of humans, pedestrian or vehicular motion, driver assistance, autonomous vehicles, objects recognition, etc. Although the conditions might be different for each application including indoor and outdoor scenarios, the development of this kind of computer vision systems is considered to be important since it is commonly concerned with activities related to people. In this paper, we propose an approach to estimate motion in dynamic environments as a solution for traffic safety in intelligent vehicles.

Traffic safety is a dominant application area of computer vision in vehicles. Many people are involved in traffic accidents every day with several fatal consequences in their lives. Those accidents can be certainly prevented with the implementation of Automatic Driving Assistance Systems (ADAS) in vehicles, which as their terms show, they might be able to help the driver in difficult situations.

In autonomous vehicles, the collision avoidance application could be estimated into two main challenges such as avoidance towards static and/or dynamic objects. Depending on the context, various objects could be part of the main scenario. Some of them might probably include characteristics that only pedestrians show, such as various poses, actions, clothes, illumination, surrounding objects, etc.

Our approach proposes an optical flow-based method for estimating motion areas in image sequences from real traffic scenarios. For the purpose of this work, we use image sequences from Daimler Dataset [1]. This Dataset includes images that are captured by a monochrome camera implemented inside a car that drives through real urban scenarios. In Fig. 1, we show a sequence of images captured by the camera when the car is moving. The representation includes a single pedestrian candidate with different postures and variations along the scene.



Fig. 1. A representation of an image sequence in an urban environment. In this example, we can evidence how the pedestrian candidate experiences various postures and variations along his way.

In this paper, we study the complex task of detecting static and moving objects, where motion is estimated in different urban traffic scenarios. Our main contributions are as follow: (1) A method for motion estimation in dynamic environments based on the principle of optical flow, namely warping multilevel optical flow. This method performs well on different urban scenarios when the camera or objects are either moving or static; (2) A multi-level pyramidal architecture of down-sampled images is built to deal with different displacements of objects moving at different speed and location from the vehicle; (3) The inclusion of a penalization function and a filtering step to make the system more robust against outliers in the final flow

field; (4) We evaluate the performance of our approach using image sequences from the Daimler Benchmark dataset [1] to estimate motion in real urban environments.

Following the motion estimation concept, this paper is organized as follows: In section 1, we introduce the background of our research. In section 2, we perform a study of different methods for motion estimation. In section 3, we describe in detail the proposed approach. The experimental results on various urban scenarios are shown in section 4. Finally, we conclude and propose our future works in the last section.

2. Related Work

Several approaches have been proposed over the last years to estimate motion areas focusing on candidates that may correspond to pedestrians. Some of them focus on camera systems, such as monocular pedestrian detection by the inclusion of features [1], background subtraction for moving cameras [2], dense stereo vision [3], or a combination of stereo dense and optical flow [4] to estimate motion. Another application that involves pedestrians has been developed by using a static camera for surveillance systems [5]. The problem becomes more difficult when the camera and objects in the scene move simultaneously. Thus, the motion produced by the camera can lead to pixel variation in the image sequence. Our motivation to overcome those difficulties is that a moving object shows different patterns between frames than a static object or the background itself.

2.1 Motion Estimation in Image Sequences

In the past few years, the use of ADAS for object detection in vehicles has become a useful driving tool and therefore, has been actively developed. One of the widely known methods was proposed by Viola and Jones [6]. In that work, the authors introduce a face detection system using the integral image method for fast feature computation and a classification approach based on the AdaBoost algorithm. The idea of using features to improve object detection results has been served as a basis of modern detectors.

The inclusion of features has generated some advances in the area, especially inspired by SIFT [7], where interesting points of the object are extracted to provide relevant feature description of the object itself. Dalal and Triggs [8] proposed the popularly known histogram of oriented gradients (HOG) for human detection, which allowed to gain performance compared to intensity based methods.

The incorporation of motion information into detectors has shown improvements when detecting objects by a moving camera. Viola et al. [9] introduce a method for pedestrian detection. That approach proposes to integrate image intensity with motion information using consecutive frames from a video sequence. The inclusion of additional information has increased the performance of individual detectors, as proposed by Wojek and Schiele [10], a method which integrates Haar-like features, shapelets, shape context, and HOG. Following that approach, Walk et al. [11] used a combination of multiple features based on self-similarity of color and motion to deal with static images and video sequences.

As an application in ADAS, Dollár et al. [12] show a detailed review of computer vision methods for pedestrian detection. In that approach, they mention that the statistics of pedestrians scale, occlusion, and location are key factors to determine the optimum conditions under which a system should operate. Another application of pedestrian detection using features extracted from single images for ADAS is referred to [13, 14]. Furthermore, the benefits of using dense stereo vision for both generation of regions of interest and pedestrian

classification are discussed in [3]. Although previous approaches have shown relatively good performance on the evaluated scenarios, there is still room to follow in order to achieve higher efficiency in real-world applications. Our study specifically focuses on algorithms based on optical flow to estimate motion in dynamic environments.

2.2 Optical Flow-Based Methods

Over the years, the original formulation of Horn and Schunck (HS) algorithm, proposed in [15] has been used as the basis of many further methods. This algorithm can deal with image sequences that are computed rather in space and time. The main idea of that approach is that optical flow is computed assuming the theory that the pixel intensity varies in every part of the image compared to its consecutive frames.

The estimation of flow vectors u and v in the Horn and Schunck algorithm [16] is called as a variational formulation, where the main goal is to minimize the energy function $E(x, y)$ in the final flow field I_f . The energy function is calculated as follows:

$$E(x, y) = (I_x u + I_y v + I_t)^2 + \lambda(|\nabla u|^2 + |\nabla v|^2) \quad (1)$$

where, the first part corresponds to the optical flow constraint with $I_x = \frac{\partial I}{\partial x}$, $I_y = \frac{\partial I}{\partial y}$,

$I_t = \frac{\partial I}{\partial t}$, which corresponds to the spatial and temporal image derivatives computed by applying mask operation from the image sequence, the second term allows smoothing the optical flow with a λ parameter that represents the Gaussian noise in the input image sequence.

From the theory of optical flow, the first term should be close to zero, or the square of the first term should be small for being considered as correct flow. Differentiating $E(x, y)$ with respect to u and v , and equating it to zero we get:

$$\frac{\partial E}{\partial u} = (I_x u + I_y v + I_t)I_x + \lambda(|\nabla u|^2) = 0 \quad (2)$$

$$\frac{\partial E}{\partial v} = (I_x u + I_y v + I_t)I_y + \lambda(|\nabla v|^2) = 0 \quad (3)$$

Let $|\nabla u|^2 = u - u_{av}$, and $|\nabla v|^2 = v - v_{av}$, where u_{av} and v_{av} are the average of the components u and v respectively taken over the four nearest neighbors of a pixel. Then,

$$(I_x u + I_y v + I_t)I_x + \lambda(u - u_{av}) = 0 \quad (4)$$

$$(I_x u + I_y v + I_t)I_y + \lambda(v - v_{av}) = 0 \quad (5)$$

As proposed in [16], solving the above equations for u and v leads to obtaining a possible solution that can be computed as follows:

$$u = u_{av} - I_x \frac{I_x u_{av} + I_y v_{av} + I_t}{\lambda + I_x^2 + I_y^2} \quad (6)$$

$$v = v_{av} - I_y \frac{I_x u_{av} + I_y v_{av} + I_t}{\lambda + I_x^2 + I_y^2} \quad (7)$$

where u and v are the flow vectors, that together represent the flow field (u, v) . The final flow field is a representation of the displacement of each pixel in the image sequence. The amount of displacement is determined in terms of the magnitude value of each flow vector. However, one of the limitations of the Horn and Schunck algorithm is that it can only deal with small displacements in the image.

Although optical flow seems to be a leading method to estimate motion in the images, it also presents some inconvenience related to high computational cost and memory consumption. For instance, in [17], one of the pioneers approaches known as Lukas-Kanade studies the cost effectiveness of the image registration techniques used for motion analysis and proposes to use the spatial intensity information to estimate the pixel that generates the best matching score in the images. Consequently, some algorithms also try to overcome those facts by the inclusion of hardware [18] or by the estimation based on the theory of warping [19], a pyramidal implementation of images [20], or by smoothing using robust kernels [21].

In order to be called efficient an algorithm should achieve a satisfactory performance, however, the relationship in terms of accuracy and efficiency has been independently pursued in optical flow methods. Liu et al. [22] analyze that trade-off by evaluating issues such as temporal filtering, subsampling, and their impact on both accuracy and speed. Sánchez et al. [23], proposes a method that is able to deal with discontinuities and noise by the use of a minimization function containing the L^1 normalization and a regularization term in the flow.

To understand the dynamic that occurs in image sequences the efficiency of motion estimation is an important fact. In [11], Walk et al. propose an optical flow-based approach to estimate motion features in images sequences for pedestrian detection. That method shows a satisfactory performance especially on low-quality image sequences and degraded flow fields. In [24], Zhang et al. introduce an optical flow-based motion segmentation approach to estimate motion generated by pedestrian candidates in image sequences.

2.3 Large Displacements Estimation

As mentioned earlier, the original formulation of optical flow in Horn and Schunck assumes that the pixel intensity remains constant over the time. Based on that assumption, Horn and Schunck method performs well for small motions in the image, but it shows lower accuracy when estimating large displacements. An important observation that we may argue is that in image sequences taken by a moving camera device, the amount of pixel displacement depends mainly on the distance to the camera. Thus, if we consider that the vehicle is moving forward, consequently, the objects located close to the camera show large pixel displacement, unlike the faraway objects with small pixel displacement.

To deal with this problem, some works have proposed the use of a pyramidal structure for computing the motion increment and flow during the interactions [20, 25, 26, 27]. For instance, in [20], motion estimation is performed by an iterative Lukas-Kanade pyramidal system that aims to detect both large and small displacements. That method shows an acceptable performance on images taken by a static camera. In [25], a combination of HS and LK uses local and global information of both methods.

In [28], an approach proposes to use region hierarchy and descriptor matching to deal with local optimizations in large displacements. Consequently, in [29], an approach estimates large displacements without a coarse-to-fine strategy. Instead, it uses a quadratic relaxation-based

technique, in which the process is held in two alternative steps, first estimating optimal data consistency, and then by discontinuity preserving regularity of the flow vectors.

The importance of motion detail preserving is introduced in [30]. In that work, the authors use a robust sparse feature matching to produce extended initialization and a selective combination of the color and gradient constraints to handle outliers during flow estimation. Prominent results were obtained, yet the cost of expensive fusion steps. In [31], a method called DeepFlow replaces the HOG matched by an approach based on similarities of non-rigid patches. Furthermore, in [32] the problems of complex motions, large displacements, and difficult imaging conditions are treated by estimating occlusions and using additional temporal information over multiple frames. Closely related to [31], an approach called EpicFlow [33] also investigates the dense matching by edge-preserving interpolation from a sparse set of matches followed by a variational energy minimization. Moreover, it introduces an approximation scheme for the geodesic distance to improve the computational cost.

Despite the success of the above methods in the evaluated scenarios, optical flow is still far from being solved. Notably, several approaches demand accurate motion estimation in challenging real applications that include several variations from the environment. Therefore the purpose of our work is to deal with those problems.

2.4 Outliers in the Flow Field

One of the drawbacks when computing the optical flow are the discontinuities in the flow field, especially affecting the direction and homogeneity of the flow vectors. In Black [34], it is demonstrated how the least-squares estimation used in Horn and Shuck algorithm makes the system more susceptible to the presence of outliers that do not correspond to the motion areas. Consequently, Muller et al. [35] introduce an alternative to penalize the differences included in the flow field by using stereo and feature correspondence to compute the vectors. In [36], Sun et al. analyses different approaches for optical flow computation and their performance for displacements estimation. They conclude that although several methods with better performances have been proposed through the years, the classical methods are still the basis of development, and so, combining them with other optimization techniques may result in competitive results.

3. Motion Field Estimation in Dynamic Environments

3.1 System Overview

As mentioned earlier, although several approaches for motion estimation on image sequences have been proposed, there are still some problems to overcome, especially related to the variations presented in dynamic environments. Our reasons to focus on motion estimation for pedestrian detection are as follows: First, moving objects are interesting for several applications such as intelligent vehicles, where pedestrians are more probable to cause an unexpected accident; Second, moving objects are usually more difficult to find in dynamic environments because of the significant variations caused by their displacements.

By estimating motion in the image, we aim to find a representation of the objects included in the scene. However, pixels in the image generate motion in the different intensities. Consequently, we identify the following causes:

- *Illumination changes*: The brightness constancy between frames can be directly affected by the intensity variation caused by illumination changes in the scene, camera movements, or image noise.

- *Outliers presence*: Because objects in the scene move into various directions and show different displacements, the presence of outliers that do not correspond to the patterns can affect the boundaries and some motion areas in the flow field.
- *Large displacements*: The amount of pixel displacement in the image is proportional to the distance respect to the camera device. In our task, all displacements represent information that can help to define different levels of accidental prevention. In Fig. 2, we show an example of various pedestrians in the image located at different distances from the vehicle.

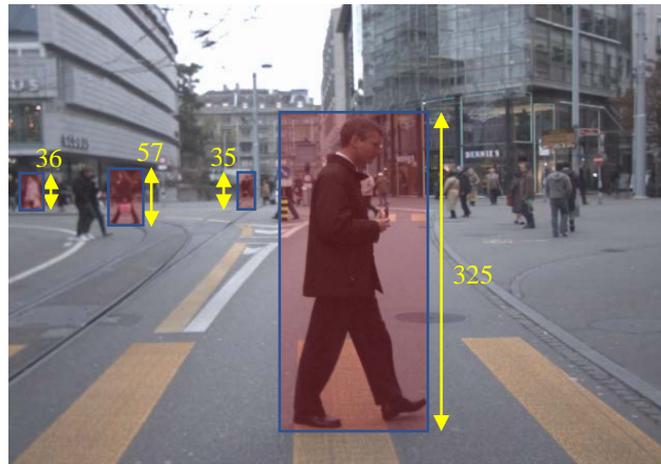


Fig. 2. In vehicular traffic scenarios, pedestrians experiment motion at various distances from the vehicle. There is a large variety of actions, location, and size of targets. Visually, large objects located close to the vehicle show larger displacements compare to the faraway ones.

In dynamic image sequences, motion estimation is an effective alternative to differentiate moving objects from the scene. Moreover, moving objects usually generate flow fields that are larger in magnitude and direction than the background and static objects. In Fig. 3, we show a flowchart of our proposed approach, and following, we describe in detail each part of the system.

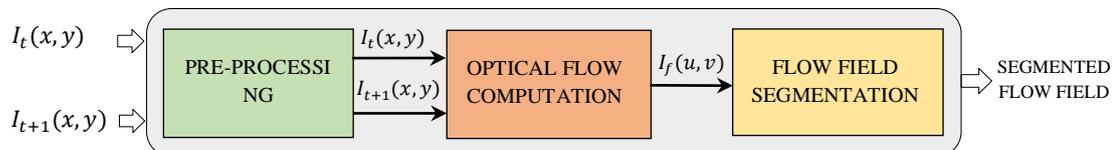


Fig. 3. System overview of the proposed approach. Two images are the inputs of the system. The images are firstly pre-processed and then compute the optical flow. Finally, the flow field is used for segmenting areas with similar motion magnitude and direction.

3.2 Input Data

Our approach firstly considers two consecutive input frames I_t and I_{t+1} respectively. Each image sequence represents a real situation in urban environments taken in real time by a camera mounted in a moving car, which enables the computation of optical flow data. An illustration of two consecutive frames of the image sequence is presented in Fig. 4. In the example, a pedestrian candidate shows a pose variation while walking across the street.



Fig. 4. A representation of two consecutive input frames. (a) Target frame I_t , (b) Current frame I_{t+1} . In this representation, we can visibly identify some variation between frames that especially correspond to the main target in the scene (pedestrian).

3.3 Pre-processing

Before estimating motion, we apply image transformation and filtering techniques to remove inconsistencies in the input frames and prevent posterior variations generated by noise or illumination changes.

First, we use the luminosity method [37] to transform the color to grayscale images, as presented in the following formula:

$$G_{value} = 0.299R + 0.587G + 0.114B \quad (8)$$

where R , G , B are the color components and the coefficients are the weights of red, blue and green colors respectively.

Second, we apply a 2D Gaussian smoothing operator to the input image sequence to remove the noise that is generated by the calibration and motion of the camera when capturing the images. The presence of noise can lead to a wrong estimation of motion that includes outliers in the scene. By the filtering application, we aim to reduce the noise while slightly influencing the region edges, by the following formula:

$$G(x, y) = e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (9)$$

where σ is the standard deviation of the Gaussian function. The values of $\sigma = 0.6$ and filter size of 15 are estimated from various experiments on image sequences. Fig. 5 illustrates an image after applying a Gaussian filter with different values of σ .

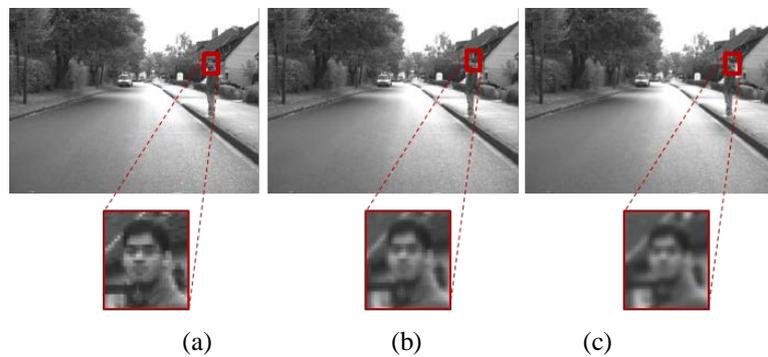


Fig. 5. A Gaussian filter with different values of σ is applied to the input images to reduce noise caused by the camera device. (a) $\sigma = 0.4$; (b) $\sigma = 0.6$; (c) $\sigma = 0.8$

3.4 Optical Flow Computation

In this part, we describe in detail the method for computing the optical flow in our proposed approach. The main goal of our system is to deal with challenges that might be presented in dynamic vehicular scenes such as different illumination conditions, outliers, constant background areas, and different pixel displacements in the image.

From its original idea [15], optical flow is mentioned as the distribution of brightness patterns in an image. The variations in the flow vectors can posteriorly help to associate similar motion regions in the flow field that include objects in the scene. The importance of estimating displacements in the image is based on the need to approximate motion without any limit of the object size in both fast and slow motions.

The idea behind our method is to create a multilevel pyramid, where the input image sequences are firstly down-sampled to make the pyramid, and then the optical flow is computed starting at lower resolution images to establish the initialization for higher resolution levels in the pyramid [38].

Motion is estimated by computing the optical flow between two consecutive frames I_t and I_{t+1} . Optical flow aims to link the target and current frames by finding the corresponding pixels that generate the motion between them. This process is divided into two parts: First, the construction of the pyramid of images and then the computation of the optical flow by a multilevel architecture, as shown in Fig. 6.

3.4.1 Pyramid of Images

This technique allows to estimate the flow field at coarser levels and then posteriorly find a solution at higher levels.

In this part, the pyramid of the images I_t and I_{t+1} is created by downsampling them from their original size by a factor $s \in \{0,1\}$. We determine the number of pyramid levels L such that at coarser level objects can be still represented. Therefore, we define the width of the last level to have a size of about 40 pixels. The size of each level is reduced to the half size of its previous level.

In order to keep information from the previous levels in the pyramid of images, the Gaussian filter is independently performed at each level during the down-sampling step. Thus, the pyramid is represented as:

$$I = G_\sigma(x, y) * I^l(x, y) \quad (10)$$

where G_σ is the Gaussian filter and I^l is the pyramid of images with different levels. The value of $\sigma = 0.6$ has been previously determined in the pre-processing step.

The input images are downsampled using bicubic interpolation. A representation of the pyramid of images $\{I_t^l\}$ and $\{I_{t+1}^l\}$ is shown in Fig. 7. The size of the largest level corresponds to the same size of the input images 640 x 480, and it is used as a reference for down-sampling the images in the next levels of the pyramid.

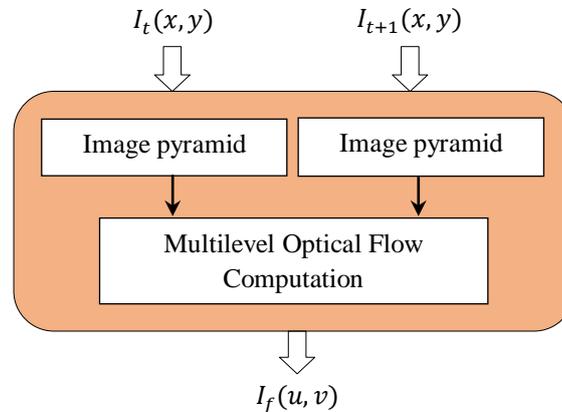


Fig. 6. A general overview of the optical flow computation

3.4.2 Multilevel Optical Flow Computation

a) Flow Estimation

To estimate the flow field that represents all pixel displacements in the image, our work proposes to compute the flow vectors u and v of each pixel using the equations (6) and (7) starting from the coarsest level in the multilevel pyramid of images. Further, the process consists in computing the flow field I_f at each level and successively until the finest level. Each intermediate flow field is used as an initialization for the next level. The resultant flow field at each level is then combined by a warping procedure, as explained in the next step.

b) Warping Flow Fields

The process for iteratively combining the resultant flow fields from each level of the pyramid follows the next steps:

- First, the optical flow is computed at the shorter level of the pyramid between I_t and I_{t+1} .
- Second, the size of the computed flow field is warped to the size of the intermediate flow field of the shorter level based on the derivative I_{t+1} of the next level in the pyramid.
- Third, compute the optical flow to the next level and combine the previously warped field with the current flow field.

Additionally, when computing optical flow at the coarsest level we consider the following facts:

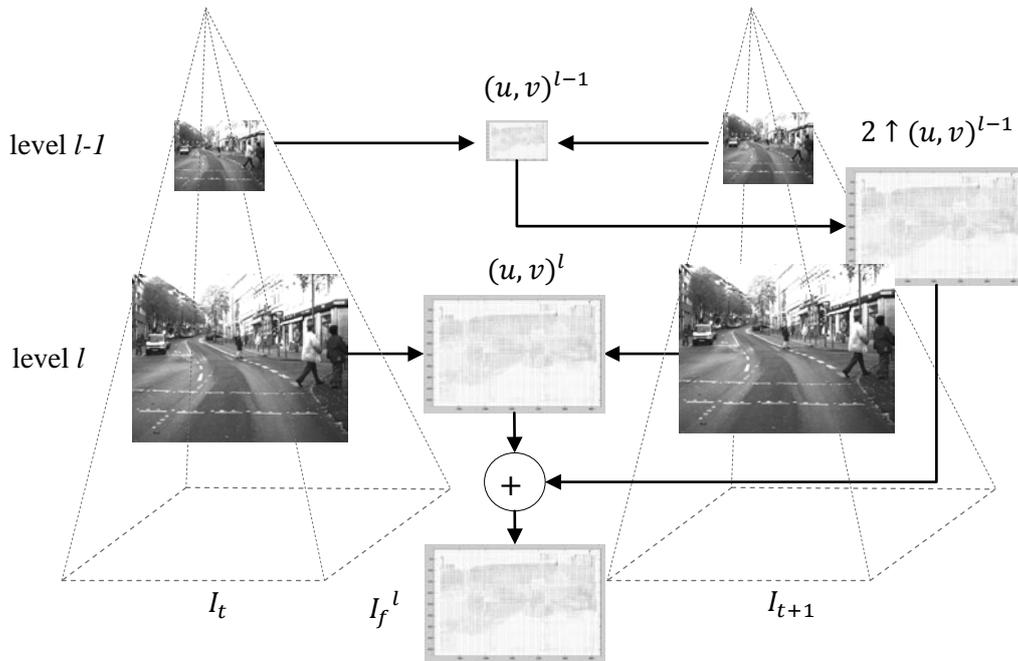
- The images are resized using bicubic interpolation to warp each flow field I_{t+1} with the intermediate level.
- Each flow field is also resized by bicubic interpolation to be able to combine the flow fields at different levels.

In **Fig. 7**, we present the optical flow computation and the warping process at each level of the pyramid. From our analysis, at coarser levels, the displacements corresponding to large moving objects are firstly detected. Those are mainly caused by the pixel size and large area covered in the image.

The final flow field has the size of the original input image and includes the addition of all warped flow fields from each intermediate level. Thus, the final flow field is computed as follows:

$$I_f(u, v) = \sum_{l=1}^L I f^l \quad (11)$$

where l represents the level of the pyramid.



1. Compute optical flow in the coarser level $(u, v)^{l-1}$
2. Warp: make a new image with flow vectors using $2 \uparrow (u, v)^{l-1}$ (up-sampling using size of next level)
3. Compute flow in the next finer level $(u, v)^l$
4. Add Flow fields from 2 and 3 $I_f^l = 2 \uparrow (u, v)^{l-1} + (u, v)^l$
5. Repeat 2 to 4 until the optical flow in the finest level is computed

Fig. 7. Warping procedure for the Multilevel Optical Flow Computation.

c) Outliers Penalization

Due to illumination variations and flow field interpolations in the multilevel architecture, some outliers in the resultant flow field are generated after the warping procedure for combining flow fields. In fact, this effect may limit the maximal motion estimation and the minimal size of the objects in the scene.

When detecting motion areas in real dynamic environments, large regions containing motion blur and noise affect the performance. The problem in the optical flow approach is that, though it can limit the effects of noise by the inclusion of a constraint parameter, it also over smooths the final flow field generating some noisy areas.

The presence of outliers can affect the estimation of the flow field in the following way:

- *The values of the brightness intensity differ from each scenario:* This could be presented in motion boundaries where the magnitude and direction of similar flow vectors change.
- The effects of denoising the flow field by using the filtering parameter can lead to an over-smoothed motion boundary.

The proposed solution leads to use a penalization function in the original optical flow formulation [15] to make it more robust against outliers in the final flow field. For that effect, we include one of the robust functions analyzed in [39], named Lorentzian, which as mentioned in its original formulation is able to differentiate inliers from outliers, based on:

$$\rho(x) = \log \left(1 + \frac{1}{2} \left(\frac{x}{\sigma} \right)^2 \right) \quad (12)$$

We use this function because its logarithmic formulation allows to recognize inconsistencies in the flow field while minimizing the energy cost and reducing the presence of outliers. Based on the application in [34], the energy minimization equation (1) is reformulated as:

$$\begin{aligned} E(u, v) = \sum_{x,y} \{ & \rho [I_t(x, y) - I_{t+1}(x + u_{x,y}, y + v_{x,y})] \\ & + \lambda [\rho(u_{x,y} - u_{x+1,y}) + \rho(u_{x,y} - u_{x,y+1}) \\ & + \rho(v_{x,y} - v_{x+1,y}) + \rho(v_{x,y} - v_{x,y+1})] \} \end{aligned} \quad (13)$$

where, ρ is the value provided by the logarithmic function, and λ is the smoothness parameter of the optical flow. The value of σ is set to $\sigma = 0.6$ and corresponds to the same σ that is previously used to build the pyramid of images.

The neighbor penalization is shown in Fig. 8. We consider the center pixel in the neighborhood as the reference point, where outliers are found on the right and top position of the flow vector u and v . This technique can help to avoid inconsistencies in the magnitude and direction of the flow vectors located especially in the boundaries.

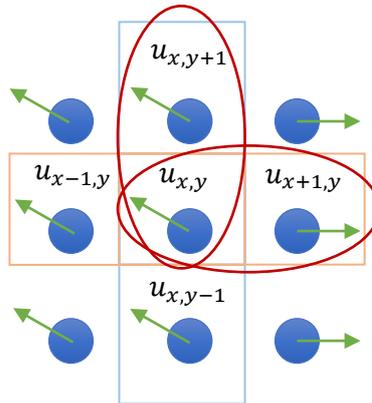


Fig. 8. Penalization function against outliers in the u direction.

Although outliers are effectively reduced after warping, some blurring areas or discontinuities on the edges of the flow field do not allow to efficiently identify objects in the scene. Because of that, motion areas that correspond to objects are often confused with parts of the background. To deal with that problem, the flow field is updated after each warping step and the filtering procedure is applied independently at each level. Thus, we reduce noise in the intermediate flow fields without losing information and preserving motion details.

After estimating the optical flow in each level of the pyramid of images and combining them by the warping process, the output of the system is the flow field $I_f(u, v)$ which represents all pixel motions in the image.

3.5 Flow Field Region Segmentation

Region Segmentation is performed to identify similar areas in the final flow field and associate them together.

We find that flow vectors of moving objects are usually more distinguishable from others. Moving and static objects show to be different in terms of both magnitude and orientation of each flow vector in the field. **Fig. 9** shows the process for segmenting the regions in the final flow field.

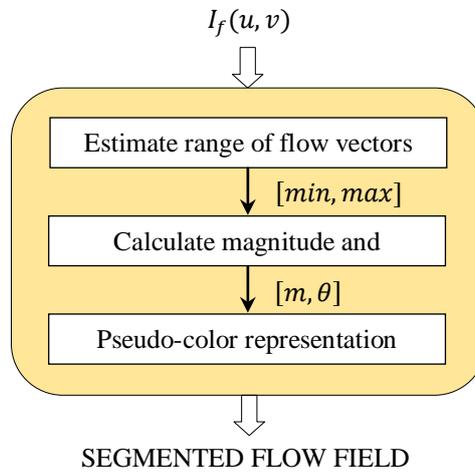


Fig. 9. Flow Field Segmentation.

The first step consists in finding the flow vectors range in both horizontal u and vertical v flow fields based on the min and max values.

By segmentation, we aim to identify flow vectors with similar properties in a neighborhood. To perform this task, the magnitude m and direction θ from each vector are calculated as follow:

- Magnitude

$$m = \sqrt{u^2 + v^2} \quad (14)$$

- Direction

$$\theta = \arctan\left(\frac{v}{u}\right) \quad (15)$$

A pseudo-color representation is used to show the similarity between flow vectors in terms of both their magnitude m and angle θ . Each flow vector is assigned with a color that represents its value according to the palette shown in **Fig. 10(a)**. The color palette used for the representation of our approach is based on the one proposed in [40]. The center of the palette corresponds to the reference point from where the color for each vector is assigned by the following rules:

- The color saturation denotes the magnitude m of the vector. The corresponding color will be closer to white if the amount of displacement is very small.
- The type of color represents the direction of the vector given by its angle θ , as shown in Fig. 10(b).

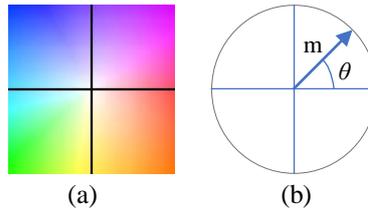


Fig. 10. Pseudo-color representation. (a) Color palette, (b) Color saturation

4. Experimental Results

4.1 Parameter Settings

The performance of our proposed approach is evaluated based on the efficiency to estimate the motion in dynamic vehicular scenes given by the final optical flow field. Our evaluation consists of qualitative experiments using image sequences from urban environments taken by a camera installed in a vehicle. Each scenario differs from others in the characteristics of the place where the images were taken and the pixel displacement as part of the main objects. For every case, we use the same parameters to compute a pyramid of images with 5 levels, and optical flow with $\lambda = 0.95$, and $\sigma = 0.6$.

4.2 Qualitative Evaluation in Real Urban Environments

To validate our method in dynamic vehicular environments, we used some image sequences from Daimler Mono Pedestrian Detection Benchmark dataset [1]. This dataset contains a large training and testing set of images. The testing set of Daimler Dataset presents the ideal conditions to evaluate the performance of our approach. It consists of an independent image sequence with about 21.790 images including 56.492 pedestrian performing various activities in urban environments. The images are captured by a camera installed in a vehicle that drives through real urban scenarios. The scenarios include images that are recorded at various daytimes and locations with different illumination conditions, pedestrian pose, or clothing styles, which let us deal with various cases.

In Fig. 11, we present some results of the optical flow region segmentation from our proposed method Fig. 11d compared to other challenging methods. Additionally, by using pseudo-color flow transformation, we are able to clearly visualize similar motion areas associated with the image, as well as to evidence how our method reduces outliers and preserve edges of the main objects (e.g. pedestrians). Our proposed method effectively shows better performance compared to other previously proposed approaches.

4.2.1 Evaluated Scenarios

We define the following evaluation scenarios based on the complexity and actions shown by moving objects:

- a) Static camera with moving objects in various directions.
- b) Moving camera with static objects in the scene.

c) Moving both camera and objects in various directions.

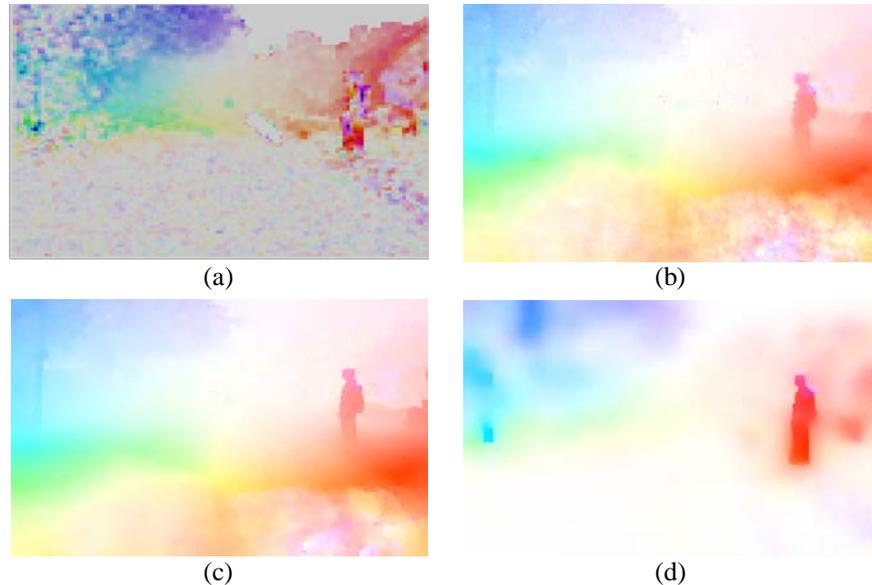


Fig. 11. Experimental results of the different optical flow approaches and segmentation using pseudo-color information. (a) Horn and Schunck baseline algorithm, (b) Multilevel Optical Flow, (c) Multilevel optical flow including a smoothing parameter, (d) Multilevel optical flow including outlier's penalization (Ours).

4.2.2 Static Camera with Moving Objects in various directions

We considered this case as the simplest one to estimate motion because the background in the image remains static due to no camera motion. This case would be an example when a vehicle remains stopped at an intersection or crosswalk. In that situation, the camera is static and only the objects perform motion in the scene. This scenario helps to clearly distinguish only moving objects from the background.

In **Fig. 12**, we show a case that includes both large and small displacement produced by the large objects located close to the vehicle and the small objects at a far distance respectively. As illustrated in the representation, the static objects are usually included as part of the background, because the optical flow only estimates motion between frames. In addition, we demonstrate with this experiment how moving objects change their position in the image over time. Furthermore, using the static camera the presence of other static objects (e.g. parked cars, buildings) does not affect the results, because they do not show any displacement.

Through the observation, we find out that moving objects can be distinguished even when they are located far away from the camera because they show motion patterns; rather the static objects that do not experience any motion, and thus usually can be confused with the background.

In dynamic vehicular environments, the amount of pixel displacement is not only related to motion in the scene but also to the distance to the camera. Therefore, objects which are located close to the camera show larger displacements than the ones that are located faraway, even in the case when they also experience motion.

4.2.3 Moving Camera with Static Objects in the Background

This scenario includes more difficulties than the previous one, because of the problem to distinguish objects when the background varies due to the moving camera. For instance, we might find this problem when the vehicle is moving and pedestrians remain static along the street. This case can help us to estimate various motion patterns of pedestrian targets that probably tend to move or walk across the street in future actions.

To evaluate this case, we choose a scene which contains 24 image pairs that correspond to 1.5 seconds that are measured by the 15 frames per second at which the dataset has been collected. As shown in [Fig. 13](#), a pedestrian stands along the way while the camera is moving together with the vehicle at an average speed of 50 Km/h in an urban environment. Our optical flow-based motion estimation method is able to detect static objects with small and large displacements at various distances from the vehicle. This can be visualized from the left image in the [Fig. 13](#) when the vehicle is moving far from the target, until the right image when the target can be clearly seen at a short distance.

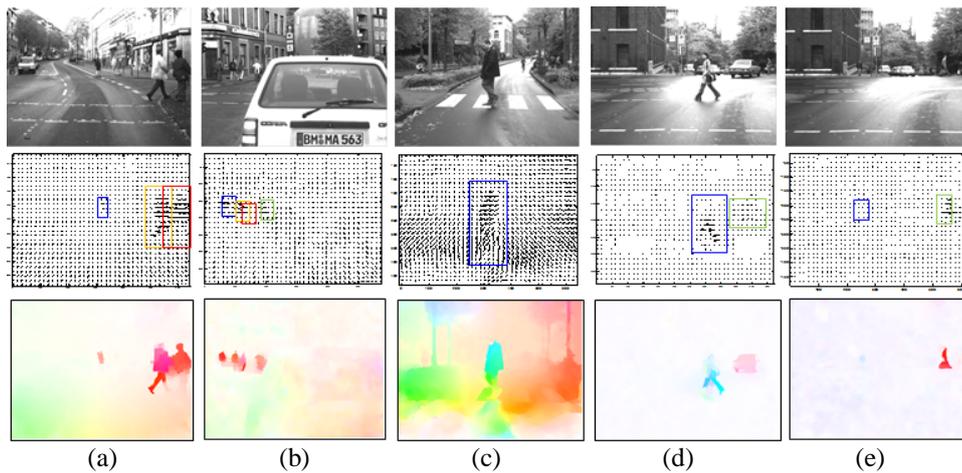


Fig. 12. Moving objects with a static camera located at various distances from the vehicle. From top to bottom: Target frame, Flow Field, and Pseudo-color representation.

From the [Fig. 13](#), we might argue that the effectiveness of detecting static objects is proportional to the distance from the vehicle. Therefore, the addition of a filtering parameter and penalization function during the optical flow computation allow to reduce the influence of outliers and to preserve edges in the final flow field. Moreover, it helps to better segment objects from the moving scenario, as represented in [Fig. 14](#).

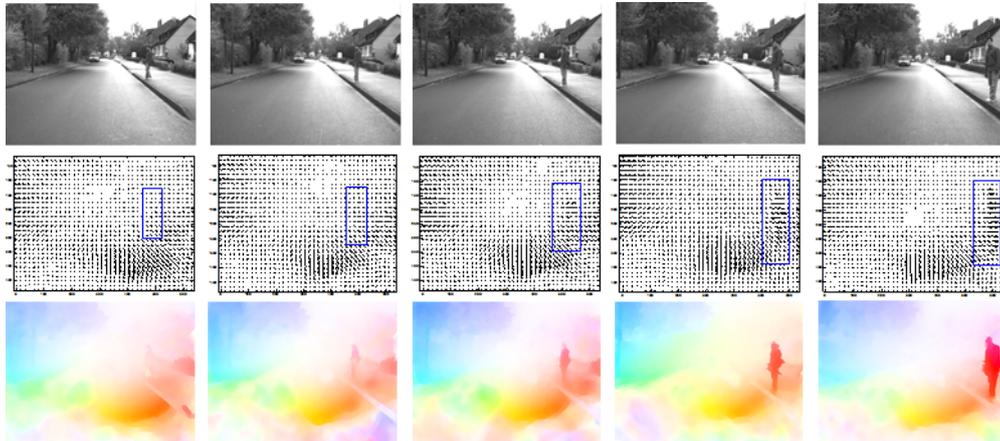


Fig. 13. Motion estimation of static objects with a moving camera. From top to bottom: Target frame, Flow Field, and Pseudo-color representation.



Fig. 14. Static pedestrian candidates in a dynamic environment. The effectiveness shows to be proportional to the distance to the vehicle.

4.2.4 Moving both Camera and Objects in Various Directions

In this case, both camera and objects move in various directions making their recognition more challenging. The most complex situation is when the objects and vehicle relatively move in the same direction at different speed. However, when the objects move along or across the street or in the opposite direction to the vehicle, those motions are mostly influenced by the pixel's displacement in the image.

The example in **Fig. 15** illustrates a case when a pedestrian is walking across the street while the vehicle and the camera are moving. The amount of pixel displacement depends on the distance from the camera and action performed by the target. The testing frames consist of 16 image sequences which represents about 1 second in time.

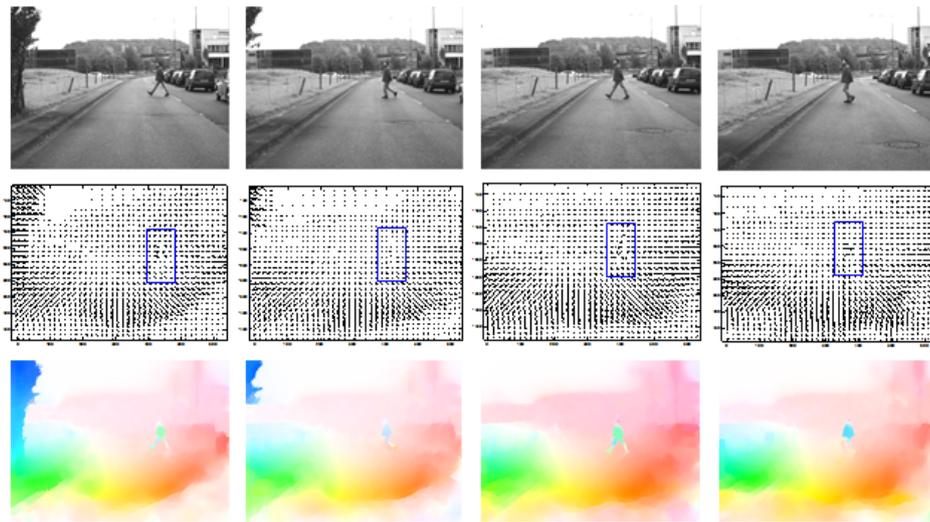


Fig. 15. Motion estimation of a moving object with moving camera. From top to bottom: Target frame, Flow Field, and Pseudo-color representation.

Visibly, moving objects show different poses when they perform an action. This can help the optical flow method to detect motion assuming that the pixel brightness remains constant between frames. Thus, optical flow is able to detect mainly motion variation. In **Fig. 16**, we present the results of various poses that a pedestrian shows over time in a sequence of images. We notice that some parts of the body such as arms or legs show more active motion since the pixel displacement of those areas is not constant in speed. Those characteristics help to estimate motion patterns that differ from the main scene in terms of magnitude and direction of the optical flow vectors.



Fig. 16. Posture variation of a pedestrian candidate walking across the street.

4.2.5 Overcoming Challenges in Urban Dynamic Environments

The evaluation of our proposed approach is supported by several experiments on different challenging scenarios specifically those with a moving camera. We choose some scenes that include changes in dynamic environments, as mentioned earlier. The system is able to detect objects performing complex actions and variations such as places with illumination and displacements variation, people walking along the street, cyclists or other vehicles moving in the same or opposite direction to the vehicle. The results of this experiment are included in **Fig. 17**.

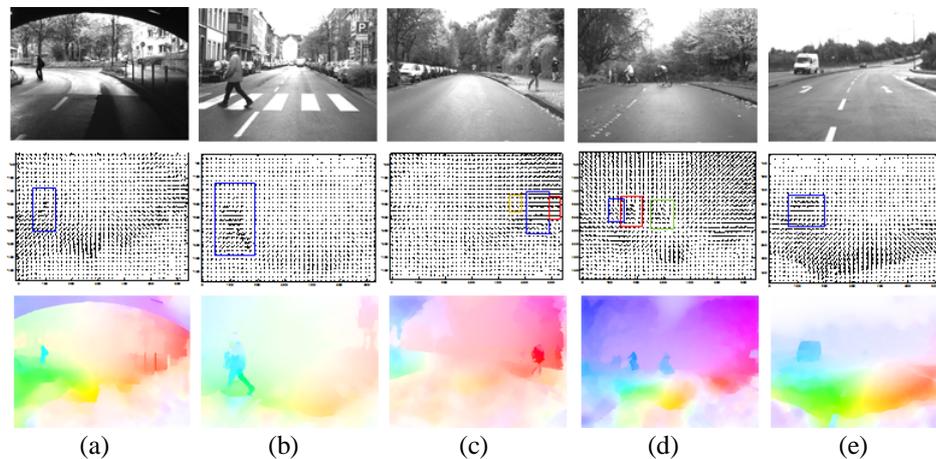


Fig. 17. Challenges in dynamic environments. (a) Illumination changes in the same scene, (b) Large displacements, (c) People moving along the street, (d) Cyclist, (e) High-speed objects moving in the opposite direction.

5. Conclusions

In this paper, we present an optical flow-based approach for motion estimation in vehicular traffic scenarios. The proposed system computes the optical flow field generated by image sequences using a multilevel pyramid architecture. The optical flow is firstly estimated at the coarsest level and then successively refined in the finer levels by a warping procedure that combines intermediate flow fields with top levels. Thus, our proposed idea is able to effectively detect small and large displacements in the image sequences.

The evaluation has also demonstrated how the inclusion of the filtering and warping steps by bicubic interpolation for up-sampling each intermediate flow field during optimization has allowed to gain performance and lead to a smoother flow field.

Objects in the scene perform various actions, and thus, they show different motion patterns in both magnitude and direction. Our optical flow method is able to detect those movements while being robust against illumination and background variations. Moreover, using pseudo-color information we visualized the resultant motion field.

Our evaluation has demonstrated that despite the complexity of urban dynamic scenarios, optical flow allows detecting motion areas that contain objects with various sizes that are located near or faraway from the camera. Therefore, this method can be applied not only to estimate motion of pedestrian targets as in our approach but also other moving objects in the image. Further research will focus on the implementation of deep learning techniques for improving computational cost and efficiency in real dynamic scenarios.

Acknowledgements

This work was supported by the Brain Korea 21 PLUS Project, National Research Foundation of Korea. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2016R1D1A1B03936269).

References

- [1] M. Enzweiler, and D. Gavrilu, "Monocular Pedestrian Detection: Survey and Experiments," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2179-2195, Dec., 2009. [Article \(CrossRef Link\)](#)
- [2] D. Zamalieva, and A. Yilmaz, "Background subtraction for moving camera. A geometric approach," *Computer Vision and Image Understanding*, vol. 127, pp. 73-85, Oct., 2014. [Article \(CrossRef Link\)](#)
- [3] C. Keller, M. Enzweiler, M. Rohrbach, D. Fernandez Llorca, C. Schnörr, and D. Gavrilu, "The Benefits of Dense Stereo for Pedestrian Detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, Dec., 2011. [Article \(CrossRef Link\)](#)
- [4] A. Talukder, and L. Matthies, "Real-time detection of moving objects from moving vehicles using dense stereo and optical flow," in *Proc. of the IEEE Int. Conf. on Intelligent Robots and Systems*, vol. 4, pp. 3718-3725, Oct., 2004. [Article \(CrossRef Link\)](#)
- [5] J. Nascimento, and J. Marques, "Performance evaluation of object detection algorithms for video surveillance," *IEEE Transactions on Multimedia*, vol. 8, no. 4, pp. 761-774, Jul., 2006. [Article \(CrossRef Link\)](#)
- [6] P. Viola, and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol 57, no. 2, pp. 137-154, May, 2004. [Article \(CrossRef Link\)](#)
- [7] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, Nov., 2004. [Article \(CrossRef Link\)](#)
- [8] N. Dalal, and B. Trigs, "Histograms of Oriented Gradients for Human Detection," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Jul., 2005. [Article \(CrossRef Link\)](#)
- [9] P. Viola, M. Jones, and D. Snow, "Detecting Pedestrians using patterns of motion and appearance," *International Journal of Computer Vision*, vol 63, no. 2, pp. 153-161, Jul., 2005. [Article \(CrossRef Link\)](#)
- [10] C. Wojek, and B. Schiele, "A performance evaluation of single and multi-feature people detection," *DAGM Symposium for Pattern Recognition*, pp. 82-91, 2008. [Article \(CrossRef Link\)](#)
- [11] S. Walk, N. Majer, K. Schindler, and B. Schiele, "New features and insights for pedestrian detection," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Aug., 2010. [Article \(CrossRef Link\)](#)
- [12] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian Detection: An Evaluation of the State of the Art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 34, no. 4, pp.734-761, Aug., 2011. [Article \(CrossRef Link\)](#)
- [13] S. Munder, and D. Gavrilu, "An Experimental Study on Pedestrian Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp.1863-1868, Sept., 2006. [Article \(CrossRef Link\)](#)
- [14] D. Geronimo, A. Lopez. A. Sappa, and T. Graf, "Survey of Pedestrian Detection for Advanced Driver Assistance Systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1239-1258, May, 2009. [Article \(CrossRef Link\)](#)
- [15] B. Horn, and B. Schunk, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, no.1-3, pp. 185-203, Aug., 1981. [Article \(CrossRef Link\)](#)
- [16] M. Shah, "Fundamentals of Computer Vision," *Computer Sciences Department, University of Central Florida*, 1997. [Article \(CrossRef Link\)](#)
- [17] B. Lucas, and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," in *Proc. of the International Joint Conference on Artificial Intelligence*, vol. 2, pp. 674-679, Aug., 1981. [Article \(CrossRef Link\)](#)
- [18] A. Garcia-Dopico, J. Pedraza, M. Nieto, A. Perez, and S. Rodriguez, "Locating moving objects in car-driving sequences," *EURASIP Journal on Image and Video Processing*, Dec., 2014. [Article \(CrossRef Link\)](#)

- [19] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. of the European Conference on Computer Vision*, vol. 3024, pp. 25-36, 2004. [Article \(CrossRef Link\)](#)
- [20] S. Tamgade, and V. Bora, "Motion Estimation of Video Image by Pyramidal Implementation of Lukas-Kanade Optical Flow," in *Proc. of the International Conference on Emerging Trends in Engineering and Technology (ICETET)*, Dec., 2009. [Article \(CrossRef Link\)](#)
- [21] A. Doshi, and A. Bors, "Smoothing of optical flow using robustified diffusion kernels," *Image and Vision Computing*, vol. 28, no. 12, pp. 1575-1589, Dec., 2010. [Article \(CrossRef Link\)](#)
- [22] H. Liu, T. Hong, M. Herman, T. Camus, and R. Chellappa, "Accuracy vs. Efficiency Trade-offs in Optical Flow Algorithms," *Computer Vision and Image Understanding*, vol. 72, no. 3, pp. 271-286, Dec. 1998. [Article \(CrossRef Link\)](#)
- [23] J. Sánchez, E. Meinhardt-Llopis, and G. Facciolo, "TV-L1 Optical Flow Estimation," *Image Processing On Line (IPOL)*, vol. 3, pp. 137-150, Sept., 2013. [Article \(CrossRef Link\)](#)
- [24] S. Zhang, C. Bauckhage, D. Klein, and A. Cremers, "Moving Pedestrian Detection Based on Motion Segmentation," in *Proc. of IEEE Workshop on Robovision*, 2013. [Article \(CrossRef Link\)](#)
- [25] A. Bruhn, J. Weickert, and C. Schnörr, "Lukas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods," *International Journal of Computer Vision*, vol. 6, no. 3, pp. 211-231, Feb., 2005. [Article \(CrossRef Link\)](#)
- [26] T. Brox, C. Bregler, and J. Malik, "Large Displacement Optical Flow," in *Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition*, Jun., 2009. [Article \(CrossRef Link\)](#)
- [27] E. Meinhardt-Llopis, J. Sanchez, and D. Kondermann, "Horn-Schunck Optical Flow with a Multi-Scale Strategy," *Image Processing On Line (IPOL)*, pp. 151-172, Jul., 2013. [Article \(CrossRef Link\)](#)
- [28] T. Brox, A. Bruhn, N. Pappenber, and J. Weickert, "High Accuracy Optical Flow Estimation Based on the Theory of Warping," in *Proc. of the European Conference on Computer Vision*, vol. 4, pp. 25-36, 2004. [Article \(CrossRef Link\)](#)
- [29] F. Steinbrucker, T. Pock, and D. Cremers, "Large Displacement Optical Flow Computation without Warping," in *Proc. of the International Conference on Computer Vision*, 2009. [Article \(CrossRef Link\)](#)
- [30] L. Xu, J. Jia, and Y. Matsushita, "Motion Detail Preserving Optical Flow Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, Sept., 2012. [Article \(CrossRef Link\)](#)
- [31] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large Displacement Optical Flow with Deep Matching," in *Proc. of the ICCV - IEEE Conference on Computer Vision*, pp. 1385-1392, Dec., 2013. [Article \(CrossRef Link\)](#)
- [32] R. Kennedy, and C. Taylor, "Optical Flow with Geometric Occlusion Estimation and Fusion of Multiple Frames," in *Proc. of the International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, vol. 8932, pp. 364-377, Jan., 2015. [Article \(CrossRef Link\)](#)
- [33] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "EpicFlow: Edge-Preserving Interpolation of Correspondences for Optical Flow," 2015. [Article \(CrossRef Link\)](#)
- [34] M. Black, and P. Anandan, "A Framework for the Robust Estimation of Optical Flow," in *Proc. of the ICCV - IEEE Conference on Computer Vision*, pp. 231-236, May, 1993. [Article \(CrossRef Link\)](#)
- [35] T. Muller, J. Rannacher, C. Rabe, and U. Franke, "Feature- and Depth-Supported Modified Total Variation Optical Flow for 3D Motion Field Estimation in real scenes," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Aug., 2011. [Article \(CrossRef Link\)](#)
- [36] D. Sun, S. Roth, and M. Black, "A Quantitative Analysis of Current Practices in Optical Flow Estimation and the Principles Behind Them," *International Journal of Computer Vision*, vol. 106, no. 2, pp. 115-137, Jan., 2014. [Article \(CrossRef Link\)](#)
- [37] R. Szeliski, "Computer Vision. Algorithms and Applications," *Springer*, 2011. [Article \(CrossRef Link\)](#)

- [38] A. Fuentes, "Efficient Motion Field Estimation in Dynamic Environments based on Warping Multilevel Optical Flow," *Chonbuk National University*, Master Thesis, 2016. [Article \(CrossRef Link\)](#)
- [39] M. Black, and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields," *Computer Vision and Image Understanding*, vol 63, no. 1, pp. 75-104, Jan., 1996. [Article \(CrossRef Link\)](#)
- [40] S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M. Black, and R. Szeliski, "A Database and Evaluation Methodology for Optical Flow," *International Journal on Computer Vision*, vol. 92, no. 1, pp. 1-31, Mar., 2011. [Article \(CrossRef Link\)](#)



Alvaro Fuentes is currently pursuing a Ph.D. degree in Electronics Engineering at Chonbuk National University, Jeonju, South Korea. He received his M.S. degree in Electronics Engineering from Chonbuk National University, Jeonju, South Korea, in 2014, and his B.S. degree in Mechatronics from Tecnica del Norte University, Ecuador, in 2012. His main research interests include deep learning, computer vision, and robotics.

E-mail: afuentes.jbnu.ac.kr



Sook Yoon is a Professor with the Department of Computer Engineering at Mokpo National University in South Korea. She received her Ph.D. degree in Electronics Engineering from Chonbuk National University, South Korea, in 2003. She was a researcher in Electrical Engineering and Computer Sciences at the University of California, Berkeley, USA, until June 2006. And she joined Mokpo National University in September 2006. She was a visiting scholar at Utah Center of Advanced Imaging Research, University of Utah, USA, from 2013 to 2015. Her main research interests include image processing and computer vision, object recognition, machine learning, and biometrics.

Email: syoon@mokpo.ac.kr



Dong Sun Park is a Professor with the Department of Electronics Engineering at Chonbuk National University in South Korea. He received his B.S. degree from the Department of Electronic Engineering of Korea University, South Korea in 1979, and his M.S. and Ph.D. degrees from University of Missouri, USA, in 1984 and 1991 respectively. His research interests include deep learning, computer vision, and biometrics.

Email: dspark@jbnu.ac.kr