Unusual Motion Detection for Vision-Based Driver Assistance

Li-hua Fu¹, Wei-dong Wu¹, Yu Zhang¹, and Reinhard Klette²

¹College of Computer Science, Beijing University of Technology, Beijing, China

²School of Engineering, Auckland University of Technology, Auckland, New Zealand

Ijfis

Abstract

For a vision-based driver assistance system, unusual motion detection is one of the important means of preventing accidents. In this paper, we propose a real-time unusual-motion-detection model, which contains two stages: salient region detection and unusual motion detection. In the salient-region-detection stage, we present an improved temporal attention model. In the unusual-motion-detection stage, three kinds of factors, the speed, the motion direction, and the distance, are extracted for detecting unusual motion. A series of experimental results demonstrates the proposed method and shows the feasibility of the proposed model.

Keywords: Vision-based driver assistance, Salient regions, Unusual motion, Video analysis

1. Introduction

The reduction of traffic accidents and improved road safety are important research subjects in transportation-related institutions or the vehicle industry. Driver-assistance systems (DASs) aim at bringing potentially hazardous conditions to the driver's attention in real time [1], and they also aim at more driver comfort. Autonomous driving is also already reality in on-road vehicles [2].

At present, moving object detection and tracking is an important research subject of DASs. Usually, approaches of moving object detection and tracking first identify objects and then try to estimate the motion by tracking the objects. Dynamic scenes, the diversity of moving objects, including non-rigid body pedestrians and rigid body vehicles, as well as weather, light and other factors, make moving objects detection and tracking very difficult.

However, drivers seem to be more concerned about unusual motion regions than with moving objects in general. We suggest to detect the unusual motion regions at pixel-level rather than at moving object level. We estimate a collision risk for every single image point, independent of an object detection step.

Visual attention is one of the most important mechanisms of a human visual system. According to the visual attention mechanism, visual saliency can detect salient regions in image and video. The visual attention model, using a mathematical model to simulate the human visual system, became a "hot topic" in computer vision.

This article aims at using the visual attention mechanism to detect unusual motion for vision-based driver assistance. The remainder of this paper is organized as follows. The unusual motion detection framework is presented in Section 2. Section 3 describes salient region detection based on visual attention. Section 4 introduces unusual motion detection

Received: Dec. 16, 2014

Revised : Feb. 10, 2015

Accepted: Mar. 17, 2015

Correspondence to: Li-hua Fu
 (fulh@bjut.edu.cn)

©The Korean Institute of Intelligent Systems

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

within the detected salient regions. Section 5 presents the experimental results. Finally, Section 6 concludes the paper and opens perspectives for future work.

2. Unusual Motion Detection Framework

In DASs, unusual motion detection is one of the important means of preventing accidents. Motion detection techniques often rely on detecting a moving object before computing motion [3]. The performance of such methods greatly depends on the performance of moving object detection.

It is a common human experience of getting out of the way of a quickly moving object before actually identifying what it is [1]. In conclusion, a human can perceive motion earlier than form and meaning.

In this paper, we propose an unusual-motion-detection model for vision-based driver assistance. The proposed model is able to detect the collision risk for the considered image points, that is independent of an object detection step.

Figure 1 illustrates our proposed unusual-motion-detection framework. This framework contains two stages, salient region detection based on visual attention, and unusual motion detection within the detected salient regions.

Since directly computing pixel-level unusualness is computationally expensive, we first introduce a salient-region-detection method, so as to define the unusual-motion-detection areas. In the stage of salient region detection, an improved temporal attention model is proposed to detect the salient regions.

In the second stage, three different factors, the speed, the motion direction, and the distance are considered to detect the unusual motion for every pixel within the detected salient regions.

3. Salient Region Detection Based on Visual Attention

In video sequences, motion plays an important role and human perceptual reactions will mainly focus on motion contrast regardless of visual texture in the scene.

Visual saliency measures low-level stimuli to the human vision system that grab a viewer's attention in the early stage of visual processing [4]. While many models have been proposed in the image domain, much less work has been done on video saliency [5].

Zhai and Shah [6] proposed a temporal attention model to use the interest point correspondences and the geometric transfor-

Figure 1. The proposed unusual-motion-detection framework.

mations between images. The projection errors of the interest points, defined by the estimated homographies, are incorporated

Figure 2. An example of detected interest points using Scale Invariant Feature Transform (SIFT). The top figure shows the original frame. The bottom figure shows the interest points detected by a SIFT detector.

Figure 3. Flowchart of the salient region detection based on visual attention.

in the motion contrast computation.

Tapu and Zaharia [7] extended the temporal attention model. Different types of motion presented in the current scene are determined using a set of homographic transforms, estimated by recursively applying the Random Sample Consensus (RANSAC) algorithm, see [8, 9], on the interest correspondences.

In the previously developed methods, detecting the feature points is the first and most important step. Obviously, the performance of the temporal attention model is greatly influenced by the results of point correspondences [10]. The Scale Invariant Feature Transform (SIFT) is used to find the interest points and compute the correspondences between the points in video frames; see also [9] for a description of SIFT.

However, it is well known that the interest point distribution generally represents a rich texture information area. If there is less texture in the potential object regions, then there are no feature points to be detected in these regions and thus the potential object regions cannot be detected. An example of interest points, detected by SIFT, is shown in Figure 2. In this case, the region where the cat (running right to left) is located is the potential object region. As shown in Figure 2, most of the detected interest points are located in the background.

In this stage, we extend the temporal attention model, proposed by Zhai and Shah [6], to obtain dense point correspondences based on dense optical flow fields. The optical flow technique is the most widely used motion detection approach [9, 11]. Optical flow at edge pixels is noisy if multiple motion layers exist in the scene.

Furthermore, in texture-less regions, dense optical flow may return error values [6]. To overcome this problem, we use a RANSAC algorithm on point correspondences to eliminate outliers.

As shown in Figure 3, this stage consists of the following steps:

Step 1: Dense point matching - First, the dense optical flow method TV-L¹ [12] is used on two consecutive frames to calculate the dense point correspondences at 10-pixel intervals. Since the moving objects always appear on the lower part of the input frames, the dense optical flow method is applied only to the bottom two thirds of the input frames. This reduces the computation time. Furthermore, to avoid the effect of noise, we exclude a 10-pixel-wide border around every frame.

Step 2: Background / Camera motion estimation - Obviously, most of the points detected at Step 1 are located in the background. The subset of m background points can be determined with the epipolar geometry constraints.

We use the multi-view epipolar constraint which requires the background points to lie on the corresponding epipolar lines in subsequent images.

If the points are far away from the corresponding epipolar lines, then we can determine them as being foreground points.

For the spatial point correspondences detected at Step 1, we apply a RANSAC algorithm respectively to determine the fundamental matrix F .

Step 3: Different types of motion recognition - In practice, multiple motions are present that result from the moving objects, but also from background objects or camera movement.

In this case we determine a new subset of points formed by all the outliers and all the points not considered in previous step.

For the current subset, we apply a RANSAC algorithm recursively to determine multiple homographies until all the points belong to a motion class. The estimated homographies model different planar transformations in the scene.

Every estimated homography H_m has a set of points $L_m = \{p_1^m, \dots, p_{n_m}^m\}$ as its inliers, and n_m is the number of inliers for H_m .

Figure 4. An example of salient region detection based on visual attention. The top left figure shows the original frame, the top middle figure shows the dense point correspondences, and the top right figure shows the background points. The bottom left figure shows clustering results for the inliers of H_1 , the bottom middle figure shows the salient map, and the bottom right figure shows the salient region.

For every homography H_m , its inliers in L_m can be considered as being located in the same plane. However, the points may belong to a distributed region.

To avoid the problem, we use the K -means clustering algorithm to divide L_m into K subsets, $L_{m,i} = \{p_{1,i}^m, \dots, p_{n_{m,i}}^m\}$, for $i = \{1, \dots, K\}$. The spanning region of $L_{m,i}$ is denoted by $R_{m,i}$, which corresponds to a moving region.

Step 4: Saliency computing - For all the moving regions determined at Step 3, we compute now their projection errors as saliency value.

The temporal saliency value of the moving region $R_{m,i}$ is defined by

$$SalT(R_{m,i}) = \frac{\sum_{k=1}^{n_{m,i}} \sum_{j=1}^M \alpha_j \cdot \epsilon(p_k^m, H_j)}{n_{m,i}} \quad (1)$$

$$\epsilon(p_k^m, H_j) = \| \hat{p}_k^{m'} - p_k^{m'} \| \quad (2)$$

$$\alpha_j = \sum_{i=1}^K \alpha_{j,i} \quad (3)$$

where M is the total number of homographies in the scene, $\hat{p}_k^{m'}$ is the projection of p_k^m computed after applying H_j , $p_k^{m'}$ is the correspondence of p_k^m found by TV-L¹, and $\alpha_{j,i}$ is the spanning area of the subset $L_{j,i}$.

An example of salient region detection based on visual attention is demonstrated in Figure 4, where apparently the attention region in the sequences corresponds to the running cat.

4. Unusual Motion Detection within the Detected Salient Regions

In the first stage, detecting the salient regions to define the unusual motion search area can reduce the computation time. In this stage, we detect the unusual motion within the detected salient regions.

We analyze the unusualness for every pixel from the following three factors: the speed, the motion direction, and the distance. As shown in Figure 5, this stage consists of the following steps:

Step 1: The speed factor analysis - The optical flow technique can accurately detect motion in the direction of intensity gradient and is the most widely used motion detection approach [11]. The TV-L¹ method is one of the best optical flow methods proposed in recent years [13]. The optical flow features in each detected salient region are obtained using the TV-L¹ method [12].

Intuitively, the speed is a determining factor in judging the unusualness of a pixel. Let (v_x, v_y) be the motion vector at a pixel location (x, y) . Therefore, the unusualness value $U_s(x, y)$

Figure 5. Flowchart of unusual motion detection within detected salient regions.

at pixel location (x, y) can be defined as follows:

$$U_s(x, y) = 255((x - D_{min}) / (D_{max} - D_{min})) \quad (4)$$

where D_{min} and D_{max} are the maximum and minimum value of magnitude, respectively.

Step 2: The motion distance factor analysis - Since only motion within some distance range is interesting for the driver, we enhance effects of motions within an assumed distance range, and decrease the impact of motions outside of this distance range. We use the general weighted operator of [14] to calculate the weight value of the motion distance factor $w_d(\alpha_d, d)$ as follows:

$$\begin{cases} 1 - e_d - (1 - e_d)((d - e_d)/(1 - e_d))^n & d > e_d \\ 1 - e_d + e_d((e_d - d)/e_d)^n & d < e_d \\ 1 - e_d & d = e_d \end{cases} \quad (5)$$

where d is the spatial distance between the pixel location (x, y) and $(w/2, h)$, which is normalized to $[0, 1]$. Here, e_d is the threshold of the motion distance, n is $-\ln 2 / (\ln(1 - \alpha_d) - \ln 2)$ with α_d in the range of $[0, 1]$, and α_d controls the strength of

Figure 6. The motion direction for a pixel.

motion distance weighting. Larger values of α_d increase the effect of motion distance weighting so that the closer motion would contribute more to the unusualness of the current pixel.

Step 3: The motion direction factor analysis - The motion direction is another factor in considering the unusualness for every pixel. Figure 6 illustrates the motion direction for every pixel within the salient regions.

As shown in Figure 6, the host vehicle is in the bottom middle of a frame. Intuitively, for the left-half region, we should just consider those pixels with the motion directions in $[-\pi, -\pi/2]$ or $[\pi/2, \pi]$. Similarly, for the right-half region, we should just cope with the pixels with motion directions in $[-\pi/2, \pi/2]$.

In practice, to deal with the width of the host vehicle car, we will adjust the right and left region. The weight value $w_a(\alpha_a, a)$ of the motion direction factor is defined as follows:

$$\begin{cases} 1 + (a + 0.5\pi) / (\pi\alpha_a) & \text{if } x < \frac{w+w_{car}}{2} \text{ and } -\pi \leq a \leq -\pi/2 \\ 1 + (a - 1.5\pi) / (\pi\alpha_a) & \text{if } x < \frac{w+w_{car}}{2} \text{ and } \pi/2 \leq a \leq \pi \\ 1 - (a + 0.5\pi) / (\pi\alpha_a) & \text{if } x > \frac{w-w_{car}}{2} \text{ and } -\pi/2 \leq a \leq \pi/2 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where a is the motion direction at the pixel location (x, y) , w_{car} is the width of the host vehicle, and α_a controls the strength of motion direction weighting. Larger values of α_a increase the effect of motion direction weighting.

Step 4: The unusualness estimation - Based on all the factors determined in the steps before, we compute the unusualness

Figure 7. Unusual motion detection results for three different videos. Row (a) shows the original frames; row (b) shows the temporal saliency maps; row (c) shows the detected salient regions; row (d) shows the unusualness maps; and row (e) shows the regions that correspond to potentially unusual motions in the selected video (e.g. in the left column, correctly a bicyclist in the left region, and, incorrectly, motion on the ground due to the moving shadow caused by the car on the right).

value $U(x, y)$ at pixel location (x, y) as follows:

$$U(x, y) = w_d(\alpha_d, d)w_a(\alpha_a, a)U_s(x, y) \quad (7)$$

5. Experimental Results

To evaluate the performance of the proposed unusual-motion-detection model, we conducted experiments on different kinds of video. A few detailed results are shown in Figure 7. The following information is presented: the representative frames of the testing videos (Figure 7a), the temporal saliency maps of the representative frames (Figure 7b), the detected salient regions (Figure 7c), the unusualness maps of the detected salient regions (Figure 7d), and the regions that correspond to potentially unusual motions (Figure 7e).

6. Conclusions

In this paper, we have developed a model for detecting unusual motion of nearby moving regions. The model can estimate the unusualness for an output of warning messages to the driver to avoid vehicle collisions. To develop this model, two stages, salient region detection and unusual motion detection, were implemented.

Based on spatiotemporal analysis, an improved temporal attention model was presented to detect salient regions. Three factors, the speed, the motion direction, and the distance, were considered to detect the unusual motion within the detected salient regions. Experimental results show that the proposed real-time unusual-motion-detection model can effectively and efficiently detect unusually moving regions.

In our future work, we plan to extend the proposed method by taking into account not merely successive frames, but also some accumulated content (e.g. about the traffic context) of a video in order to increase the robustness of the algorithm and to incorporate an object tracking method.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

The first author thanks China Scholarship Council (CSC) for supporting her stay at The University of Auckland.

References

- [1] Cherng, C. Y. Fang, C. P. Chen, and S. E. Chen, "Critical motion detection of nearby moving vehicles in a vision-based driver-assistance system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 70-82, 2009. <http://dx.doi.org/10.1109/TITS.2008.2011694>
- [2] U. Franke, D. Pfeiffer, C. Rabe, C. Knoeppel, M. Enzweiler, F. Stein, and R. G. Herrtwich, "Making bertha see," in *Proceedings of 2013 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Sydney, 2013, pp. 214-221. <http://dx.doi.org/10.1109/ICCVW.2013.36>
- [3] R. Danescu, F. Oniga, and S. Nedevschi, "Modeling and tracking the driving environment with a particle-based occupancy grid," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1331-1342, 2011. <http://dx.doi.org/10.1109/TITS.2011.2158097>
- [4] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194-203, 2001. <http://dx.doi.org/10.1038/35058500>
- [5] D. Rudoy, D. B. Goldman, E. Shechtman, and L. Zelnik-Manor, "Learning video saliency from human gaze using candidate selection," in *Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Portland, OR, 2013, pp. 1147-1154. <http://dx.doi.org/10.1109/CVPR.2013.152>
- [6] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th Annual ACM International Conference on Multimedia*, Santa Barbara, CA, 2006, pp. 815-824. <http://dx.doi.org/10.1145/1180639.1180824>
- [7] R. Tapu and T. Zaharia, "Salient object detection based on spatiotemporal attention models," in *Proceedings of 2013 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, 2013, pp. 39-42. <http://dx.doi.org/10.1109/ICCE.2013.6486786>
- [8] J. J. Lee and G. Kim, "Robust estimation of camera homography using fuzzy RANSAC," in *Proceedings of International Conference on Computational Science and Its Applications (ICCSA)*, Kuala Lumpur, Malaysia, 2007, pp. 992-1002. http://dx.doi.org/10.1007/978-3-540-74472-6_81

- [9] R. Klette, *Concise Computer Vision: An Introduction into Theory and Algorithms*, London: Springer, 2014.
- [10] D. Liu and M. L. Shyu, "Effective moving object detection and retrieval via integrating spatial-temporal multimedia information," in *Proceedings of 2012 IEEE International Symposium on Multimedia (ISM)*, Irvine, CA, 2012, pp. 364-371. <http://dx.doi.org/10.1109/ISM.2012.74>
- [11] S. H. Zhong, Y. Liu, F. Ren, J. Zhang, and T. Ren, "Video saliency detection via dynamic consistent spatio-temporal attention modelling," in *Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI)*, Bellevue, WA, 2013, pp. 1063-1069.
- [12] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, "An improved algorithm for TV-L¹ optical flow," in *Statistical and Geometrical Approaches to Visual Motion Analysis*. Berlin: Springer-Verlag, 2009, pp. 23-45. http://dx.doi.org/10.1007/978-3-642-03061-1_2
- [13] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1-31, 2011. <http://dx.doi.org/10.1007/s11263-010-0390-2>
- [14] L. Fu, D. Wang, and J. Kuang, "Parametric analysis of flexible logic control model," *Discrete Dynamics in Nature and Society*, vol. 2013, article id. 610186, 2013. <http://dx.doi.org/10.1155/2013/610186>

Li-hua Fu is an associate professor at the College of Computer Science, Beijing University of Technology. She received the Ph.D. degree in computer software and theory from Northwestern Polytechnical University, China.

Currently, her main research interests include computer vision, image processing, and image understanding.

Wei-dong Wu is a Master student at the College of Computer Science, Beijing University of Technology. His research is in the fields of visual attention detection and pattern recognition.

Yu Zhang is an undergraduate student at the College of Computer Science, Beijing University of Technology. His research activities are related to studying optical flow.

Reinhard Klette is a Fellow of the Royal Society of New Zealand and a professor at Auckland University of Technology. He (co-)authored more than 400 publications in peer-reviewed journals or conferences, and books on computer vision, image processing, geometric algorithms, and panoramic imaging. He was an associate editor of IEEE PAMI (2001-2008).