Vehicle detection and tracking based on phase-correlation

Yi He (何 毅) and Xin Yang (杨 新)

Pattern Recongnition and Intelligence System Laboratory, Shanghai Jiao Tong University, Shanghai 200030

Received December 5, 2003

This paper presents vehicle detection and tracking algorithms based on real-time background (RTB) and phase-correlation (PC) in the video sequence of urban highway with fixed camera. Firstly moving objects are obtained by subtracting RTB from serial images. After classification, PC is used to determine corresponding regions of moving objects between consecutive frames. The problems of vehicles' merging and splitting, sudden stop, and restart are also considered. Experiments show that the method is practical and can realize real-time detection and tracking of vehicles on highway.

OCIS codes: 040.7290, 330.4150, 100.5070.

Although many different methods are presented for vehicle's detection and tracking, the main trends can be included as: 1) feature based methods, 2) deformable model based methods, 3) region tracking based methods[1−3]. Contrasting to the previous two methods, region tracking methods are less sensitive to occlusion, for the region can present similar features such as intensity, shape, etc..

For most of the time, the background varies slowly. The gray level existing moving objects change more violently. So object image (objImg) including only moving objects can be defined as

$$\text{objImg} = \text{cuImg} - \text{cuBg}, \tag{1}$$

where cuImg is current image, cuBg is current background. Then we can have that the more the cuBg approximates the true background, the better the segmentation result is. So the background have to be updated temporally.

If the objImg is thresholded, a binary image (obj$M$) can be obtained, in which "1" represents background and "0" represents object. One of objImgs' histogram is illustrated in Fig. 1(d). In this paper, the threshold discriminating fore-from back-ground is 10% of the peak at the right of histogram. So the threshold changes with objImg.

Two steps are taken to get dynamic background. The first is to get crude background. Let ini$M$ be initialized as zero mask. The cuBg is initialized as zero image. If

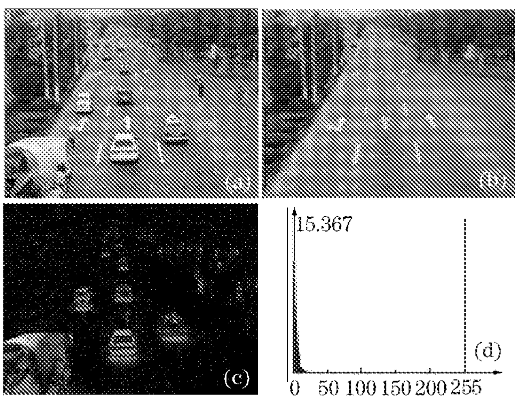

Fig. 1. Illustration of real-time background extraction to get moving objects. (a) The 80th frame; (b) the background; (c) the moving images of (a); and (d) the gray histogram of (c).

ini$M[i][j]$ is zero, cuBg$[i][j]$ is not set value. Difference image (diffImg) is defined as

$$\text{diffImg} = |\text{cuImg}(n+1) - \text{cuImg}(n)|, \tag{2}$$

where cuImg(n) and cuImg($n+1$) are the nth and the $(n+1)$th sequent images, respectively. $|\cdot|$ means absolute operator. diffImg is thresholded to get obj$M$. Then cuBg$[i][j]$ changes as

$$\text{cuBg} = \begin{cases} \alpha \cdot \text{cuBg} + (1-\alpha)\text{cuImg} \\ \quad \text{if ini}M[i][j] = \text{obj}M[i][j] = 1 \\ \alpha \cdot \text{cuImg}(n) + (1-\alpha)\text{cuImg}(n+1) \\ \quad \text{if ini}M[i][j] = 0, \text{obj}M[i][j] = 1 \\ 0 \\ \quad \text{if ini}M[i][j] = \text{obj}M[i][j] = 0 \end{cases}, \tag{3}$$

where $i = 0, 1, \cdots, R-1$, $j = 0, 1, \cdots, C-1$, $R \times C$ is image's size, and α is weight affecting update speed and set empirically as 0.1.

When all pixels of ini$M$ are set as 1, then go to the next step. From now on, cuBg is applied in Eq. (1). objImg and obj$M$ are obtained as above. If obj$M[i][j] = 0$, cuBg left unchanged, or else it is updated as

$$\text{cuBg} = \alpha \cdot \text{cuImg} + (1-\alpha)\text{cuBg}. \tag{4}$$

In once background extraction (OBE) process as above, if the gray histogram of objImg cannot tell clearly background from foreground, it will results in trail of moving objects as jet aircraft in later objImgs. The trails are formed by mis-discriminated pixels. In order to eliminate these trails and make the arithmetic be more robust, this paper takes quadratic background extraction (QBE) arithmetic. The final objImg is gotten as

$$\text{objImg} = \text{cuImg2} - \text{cuBg2}, \tag{5}$$

where cuImg2 is valued by objImg in Eq. (1), cuBg2 is the quadratic real-time background (RTB) of cuImg2, and cuBg2 is similarly derived from Eqs. (2)−(4).

Vehicles commonly have larger areas than most of pedestrians and bikes. Being practical, regions' sizes are used as classification metrication which brings rational result.

Local phase-correlation (LPC) is applied to determine corresponding objects of consecutive frames. The phase-correlation (PC)[4,5] is applied to regions only existing objects in objImg gotten by QBE. One advantage of adopting LPC is that PC primarily includes FFT and IFFT, for small region sizes, its computing cost is very low and suitable for real-time tracking. At the same time, it is enough to determine corresponding regions of consecutive frames. Two regions' correlation $c_{k,k+1}(x,y)$ is given as

$$c_{k,k+1}(x,y) = \text{preRect}[i] \times \text{cuRect}[j], \tag{6}$$

where preRect and cuRect are object regions in objImg(t_k) and objImg(t_{k+1}), respectively. $i = 0, 1, \cdots, N_1-1$, $j = 0, 1, \cdots, N_2-1$, and N_1, N_2 are number of the object regions in objImg(t_k) and objImg(t_{k+1}), respectively. According to discrete PC definition, the normal cross power spectrum of Eq. (6) is

$$\widetilde{C}_{k,k+1}(\xi,\eta) = \frac{F_k(\xi,\eta)\overline{F}_{k+1}(\xi,\eta)}{||F_k(\xi,\eta)\bar{F}_{k+1}(\xi,\eta)||}$$

$$= \exp[-2j\pi(\xi\Delta x + \eta\Delta y)], \tag{7}$$

where $C_{k,k+1}(\xi,\eta)$, $F_k(\xi,\eta)$, and $F_{k+1}(\xi,\eta)$ are the FFT of $c_{k,k+1}(x,y)$, preRect[i], and cuRect[j], respectively. $\overline{F}_{k+1}(\xi,\eta)$ is the conjugate of $F_{k+1}(\xi,\eta)$. The IFFT of Eq. (7) is

$$\widetilde{c}_{k,k+1}(x,y) = \delta(x-\xi, y-\eta). \tag{8}$$

The displacement of two corresponding moving objects is the pulse position in Eq. (8). The correlation value is normalized as

$$\bar{c}_{k,k+1}(i,j) = \frac{c_{k,k+1}(i,j)}{\max_j(c_{k,k+1}(i,j))}. \tag{9}$$

Bidirectional LPC is applied to consequent objImgs, in which regions having the maximal correlation value with each other are determined as corresponding ones. Bidirectional correlation makes every object have unique corresponding one in the other. Practically only the nearby objects of the expected positions are considered.

A Kalman filter is used to resolve vehicles' merging or splitting, stop or restart. With this filter, objects' positions are expected in the next frame. If several objects lose and reach the same position, at the same time, a new object appears at the same place, merging is declared. The big region is tracked at the following frames. Similarly, if one object cannot find its corresponding one at the next frame, and several objects appear at the expected position, splitting is declared. This is illustrated in Fig. 2. If one object disappears in the next frame, the expected position does not get across the border, and there is no merging happening, it is judged as sudden stop. When a new object appears, compute its correlation with stop ones. If the correlation surpasses the threshold, restart is determined.

Experiments are carried on video sequence with fixed monocular camera. The sample frequency is 25 frame/s. Correlation is calculated per 5 frames and tracks of moving objects are drawn per 10 frames. More than 95% moving objects can be detected in the experiment. The classification's region size just at appearing place is 32×32 pixels and more than 80% objects can be correctly classified. More than 75% vehicles after classification can be tracked. Those vehicles far from view are abandoned for track. Figure 2 illustrates split and merge case. Figure 3 illustrates detection and track of one scene.

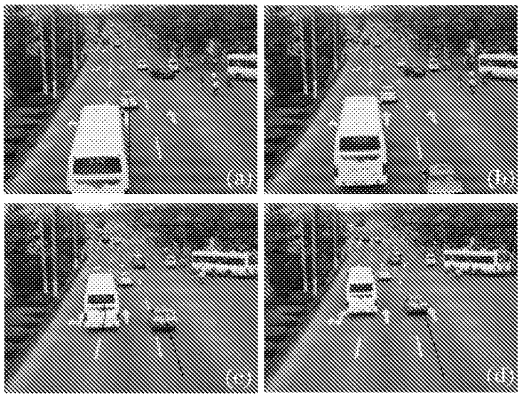

Fig. 2. Illustration of merging and splitting in the scene. (a) The 1051st frame, the bus and its frontal car are detected as two different moving objects; (b) and (c) the 1061st and 1091st frames, respectively, the bus and car have merged and been tracked as a whole region; (d) the whole region splits into two objects again.

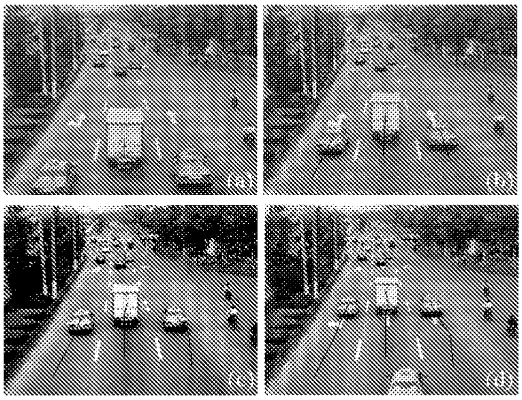

Fig. 3. Illustration of vehicle tracking of one scene. (a), (b), (c), and (d) are the 461st, 471st, 481st, and 491st frames of it, respectively.

In this paper, RTB extraction is used to detect moving objects in video sequence, bidirectional LPC is applied to determine corresponding moving objects. Merging, splitting, sudden halt, and restart are also considered. From these experiments, we can see that the arithmetic can detect almost all of moving objects and can effectively track most of them.

Y. He's e-mail address is heyihh@sjtu.edu.cn.

References

1. D. P. Kottke and Y. Sun, Pattern Analysis and Machine Intelligence **16,** 1128 (1994).
2. J. Badenas, J. M. Sanchiz, and F. Pla, in *Proceedings of 15th International Conference on Pattern Recognition* **3,** 125 (2000).
3. S. Gupte, O. Masoud, R. F. K. Martin, and N. P. Papanikolopoulous, Intelligent Transportations Systems **3,** 37 (2002).
4. H. Foroosh, J. B. Zerubia, and M. Berthod, Image Processing **11,** 188 (2002).
5. B. Tang, Y. X. Leng, H. L. Peng, Z. Q. Zhang, and L. H. Lin, Chin. Opt. Lett. **1,** 432 (2003).