CYKLS: Detect Pedestrian's Dart Focusing on an Appearance Change

Masahiro Ogawa, Hideo Fukamachi, Ryuji Funayama, and Toshiki Kindo

Future Project div., Toyota Motor Co.
1200, Mishuku, Susono, Shizuoka, Japan
{masahiro@ogawa,fuka@machi,ryuji@funayama,toshiki@kindo}.tec.toyota.co.jp

Abstract. We propose a new method for detecting "pedestrians' dart" to support drivers cognition in real traffic scenario. The main idea is to detect sudden appearance change of pedestrians before their consequent actions happen. Our new algorithm, called "Chronologically Yielded values of Kullback-Leibler divergence between Separate frames" (CYKLS), is a combination of two main procedures: (1) calculation of appearance change by Kullback-Leibler divergence between descriptors in some time interval frames, and (2) detection of non-periodic sequence by a new smoothing method in the field of time series analysis. We can detect pedestrians' dart with 22% Equal Error Rate, using a dataset which includes 144 dart scenes.

1 Introduction

We propose a new method for detecting "pedestrians' dart" to support drivers cognition in real traffic scenario.

The main idea is to detect sudden appearance change of pedestrians before their consequent actions happen.

To confirm that there exist such an appearance change as a premonitory phenomenon of dart, we first checked several drive recorder data. (Unfortunately, we don't allowed to process the drive recorder data. In addition to that, if we were allowed to process the data, the resolution is too coarse.) Those consist of 2099 near-miss scenes involving pedestrians. About 25% of these scenes include sudden acceleration or direction changes of the pedestrians, unrelated to the car motion. Then we investigated whether these scenes include any appearance change of the pedestrian, and we confirm there exist some appearance changes, such as "upper body turning toward traveling direction", "head bending forward", "increasing the length of stride" and "broadening legs". So if we can detect these appearance changes, we could predict pedestrian's dart and it has the potential to reduce car-pedestrian accidents.

Until now, variation detection has not been studied so extensively, especially for car mounted camera. As far as we know, in the field of automatic risk detection involving pedestrians, this is the first study of these types.

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part II, LNCS 7584, pp. 556–565, 2012.

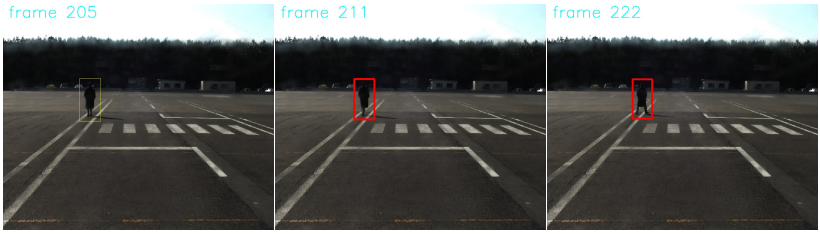


Fig. 1. Typical detection result of CYKLS. Red rectangle indicates dart detection result.

1.1 Related Work

Gesture Recognition by Supervised Learning. There have been many works concerning detection of specific postures and gestures. A.Efros et al [1] developed a method to detect low resolution human movement. It tracks humans and gets the local motion using optical flow. The recognition is performed using a nearest neighbor framework. I.Laptev [2] introduces local spatio-temporal image features, which do not need spacial segmentation. Recognition is done by linear discrimination. E.Shechtman et al [3] also propose a method which does not need detection or tracking. It uses space-time patch and detects similar movements in 2 video sequences. L.Yeffet et al [4] introduce LTP feature which extend LBP [5] to time dimension , and recognize actions using linear SVM. M.Andriluka et al [6] introduce a 3D pose estimation method based on 2D inference and dynamical model, and apply it to the real world. All above-mentioned methods need learning/prior data set, which are sometimes hard to create, especially for all dart postures or gestures.

Abnormal Behavior Detection. There also have been many attempts to detect abnormal behavior in scenes [7][8]. In recent years, F.Nater et al [9] proposed a method which can build hierarchical activity model automatically. However in a real traffic scenario, pedestrians can be seen only for a few seconds at most, so there is only about one-third of a second to make a decision. All these methods use static background assumption, so they can not be applied directly to car mounted camera.

Optical Flow Based Method. Daimler researchers develop a 6D-Vision [10][11] to estimate objects 3D position and 3D motion in traffic scene using car mounted camera. It successfully combines a dense variational optical flow with Kalman filters at every single pixel. This method cannot detect premonitory phenomenon of dart. On the other hand, our method use appearance change to catch premonitory phenomenon of dart so that we can detect dart as early as possible.

2 Dart Detection Method

2.1 CYKLS

In this section, we give an explanation of our dart detection method which we named CYKLS (Chronologically Yielded values of Kullback-Leibler divergence between Separate frames).

CYKLS, is a combination of two main processes:

1. Calculation of appearance change by Kullback-Leibler divergence [12] between descriptors in some time interval frames, and
2. Detection of non-periodic sequence using a smoothing method

The second part is a totally new change detection algorithm in the field of time series analysis. The detailed algorithm of CYKLS is described below.

Calculation of Appearance Change. After detecting pedestrians, we catch the pedestrian's shape:

$$\mathbf{v}(t) : \text{some descriptor inside the ROI} \tag{1}$$

using some descriptor of a region of interest (ROI) inside the pedestrian's image window at each frame t , and L1 normalize the descriptor into:

$$\mathbf{p}(t) := \frac{\mathbf{v}(t)}{\|\mathbf{v}(t)\|_{L1}}. \tag{2}$$

We then consider the normalized SIFT feature vector [13] $\mathbf{p}(t)$ as a probabilistic distribution, and compute the difference of the distribution between frame t and t-n using KL divergence as:

$$d(t,n) := D_{KL}(\mathbf{p}(t)||\mathbf{p}(t-n)). \tag{3}$$

Detection of Non-periodic Sequence. Next, we create a l-dimension vector:

$$\mathbf{u}(t,n,l) := (d(t-l+1,n),\ldots,d(t,n)) \tag{4}$$

from the time series data $d(t,n)$, and sum up backward to get:

$$\mathbf{U}(t,n,l,K) := \sum_{k=0}^{K} \mathbf{u}(t-k,n,l). \tag{5}$$

Then, we compute the cost:

$$\text{cost} := \cos\Theta(t,n,l,K) := \frac{\mathbf{U}\cdot\mathbf{I_l}}{\|\mathbf{U}\|_{L2}\|\mathbf{I_l}\|_{L2}} = \frac{\sum_i U_i}{\sqrt{l}\|\mathbf{U}\|_{L2}} \tag{6}$$

as the cosine of the angle between $\mathbf{U}(t, n, l, K)$ and the constant vector:

$$\mathbf{I_l} := (1, 1, ..., 1)\ ;\ l-\text{dim}. \tag{7}$$

If a pedestrian's movement is cyclic, like walking, and parameter K is constant factor of walking cycle, then the sum, $\mathbf{U}$, become constant factor of $\mathbf{I_l}$. So if we compare $\mathbf{U}$ and $\mathbf{I_l}$ in angle, we can suppress periodic movement. This algorithm is a new smoothing method in the field of time series analysis.

Finally, the cost $\cos\Theta$ is compared to a detection threshold T:

$$\text{cost} < \text{T} \Rightarrow \text{detect dart} \tag{8}$$

The total flow of CYKLS is in Fig.2, and output examples are in Fig.4.

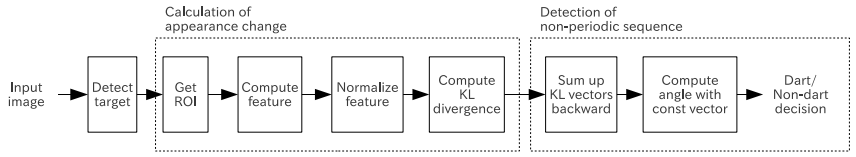


Fig. 2. Flow chart of CYKLS

2.2 Notes

The first part of the process ("Calculation of appearance change") allows to detect sudden direction change, while, in addition to that, the second part ("Detection of non-periodic sequence") deals with sudden acceleration. This is illustrated in the Fig.4c.

3 Evaluation

3.1 Dataset

We created a new dataset to evaluate our method. We depicted the typical 4 dart pattern which cause the danger from the drive recorder data. The dataset includes 6 pedestrians(A-F), 6 environmental settings(a-f) and 4 dart patterns(1-4), so the total is 144 sequences. The image size is XGA, and each sequence is about 10 seconds long with 15 fps and only includes one person and one dart. In order to evaluate only the dart detector, we annotate all pedestrians by hand, and use these annotated windows instead of the output of a pedestrian detector. The composition of our dataset is listed below.

– pedestrians: A:woman, B:man wearing long sleeve shirt, C:man wearing coat, D:man wearing black clothes, E:man wearing short sleeve shorts, F:man wearing similar color clothes to the background (see Fig.3)

– environmental settings:

	a	b	c	d	e	f
Background	plain	plain	natural	natural	natural	plain
Movement before dart	walk	stop	walk	stop	walk	walk
Distance at the dart moment [m]	22	22	22	22	10	10

– dart patterns: 1:change of direction, 2:change of direction followed by running, 3:running, 4:zigzag slightly (see Fig.4)

(a) A (b) B (c) C (d) D (e) E (f) F

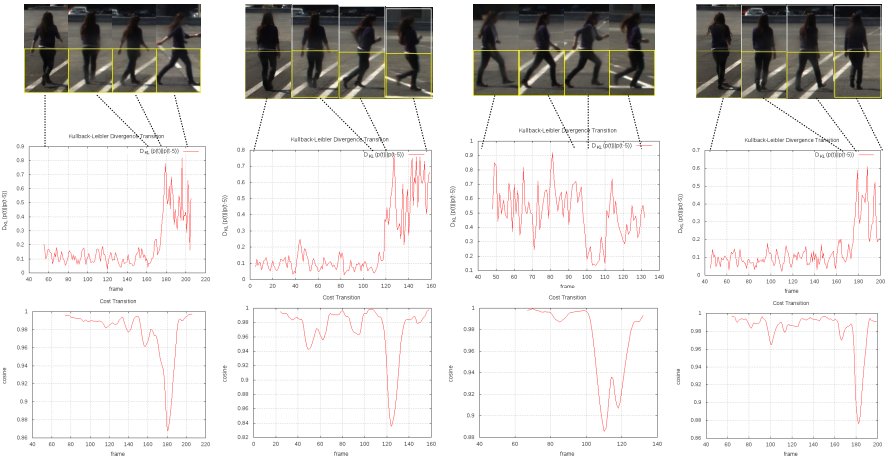
Fig. 3. Pedestrians of our dataset

3.2 Evaluation Scheme

We annotate every pedestrian with dart or non-dart, and evaluate the result by drawing standard DET(False Positive(FP)-False Negative(FN)) curve. Once the parameter set $\{n, l, K, T\}$ is given, we can calculate FP and FN.

3.3 Evaluation Results

Using the dataset mentioned in section 3.1, we compute the FP-FN rates for each 7980 parameter sets($1 \leq n \leq 20, 2 \leq l \leq 20, 0 \leq K \leq 20$). We compare CYKLS with other methods. Dart detection is somewhat new problem, and we cannot find out state-of-the-art method in this field, so we compare CYKLS with below baseline detectors. 3 norm options(KL divergence(same as CYKLS),L1,L2) and 5 smoothing methods(no smoothing, same as CYKLS, moving average with weight vector=const,x,exp(x)), so there are 15 detectors in total for comparison. The result is shown in Fig.5, in which we abbreviate each norm to "KL,L1,L2", and each smoothing method to "no,CY,const,x,exp". As a result, CYKLS get the lowest FP-FN rate (about 22% EER). One of the best parameter set we found is $(n, l, K) = (2, 19, 8)$, and using this parameter, we can detect dart with FP 20%, FN 26%. The percentage is not as bad as it may seems. We want to detect dart as early as we can, so we annotate the pedestrian who has very slight appearance change as dart, for example second from left pedestrian image in Fig.4. This causes large FN.



(a) pattern1 (b) pattern2 (c) pattern3 (d) pattern4

Fig. 4. Dataset examples of 4 dart patterns with result graphs. Top row is annotated windows with ROI (yellow rectangle). The second row is the graph of KL divergence transition. The third row is the graph of cost transition. We can see that all 4 dart patterns are detectable in the graph of the third row.

3.4 ROI Test

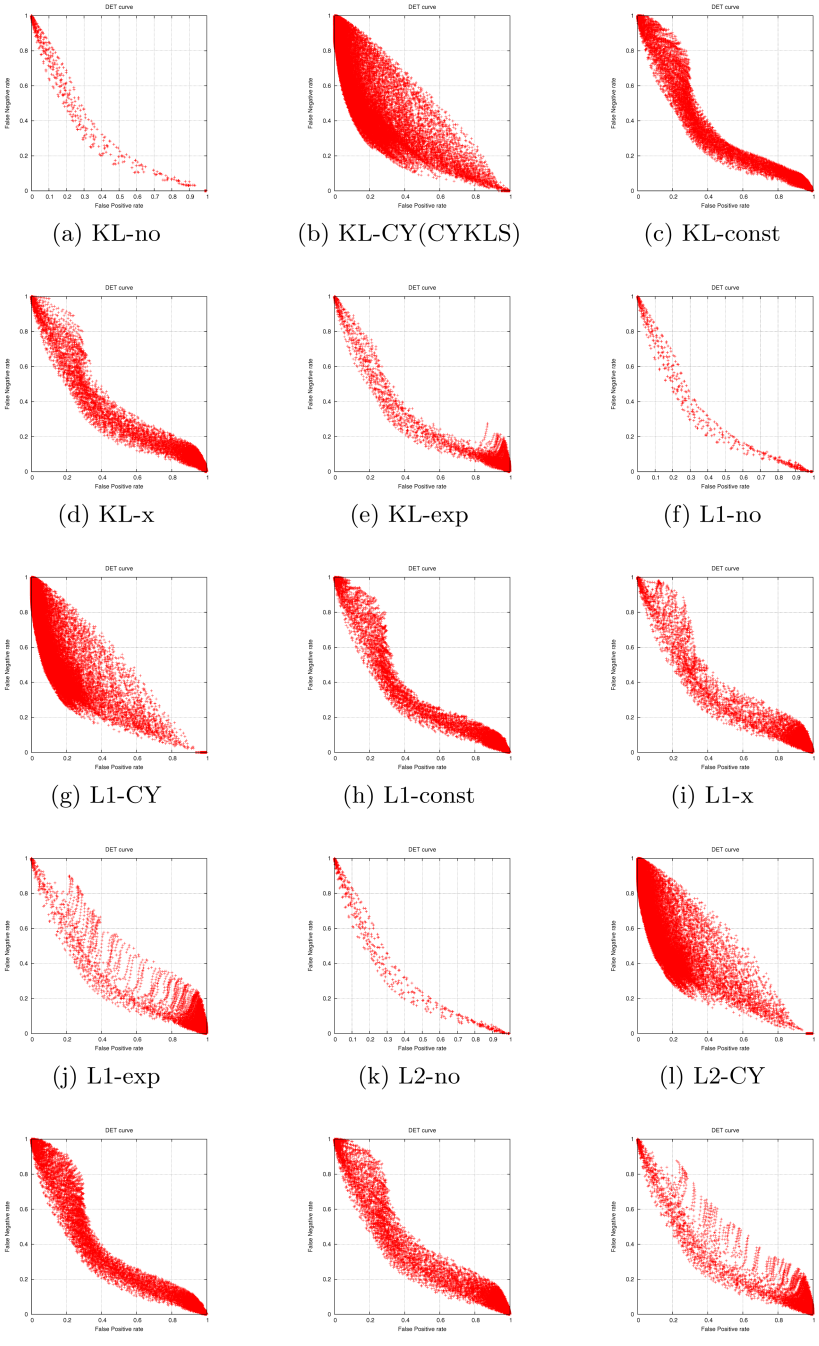
Actually there is no necessity that the ROI used for dart detection coincides with the window used for detecting pedestrians. We then test several ROIs inside the detected pedestrian's window, for which we compute $\mathbf{v}(t)$ in (1).

We first create 7 ROI as shown in Fig.6. For the latter 4 ROI, we move the ROI inside the detector window such as it minimizes the appearance change using SIFT descriptor and a template model. The template model is computed using the first frame. Finally, the total number of tested ROI is then 11.

We evaluate those 11 ROI by the method mentioned in section 3.2 using the one of the best parameter set which is found in section 3.3($(n, l, K) = (2, 19, 8)$). As a result, block0 (whole window) give the lowest EER of the 11 ROI (see Fig.7).

3.5 Effect of Tracking Error

In practice, before detecting dart, we first need pedestrian detection and tracking. The detected pedestrian's window has some motion noise. But as mentioned above, as we want to divide the problem of dart detector and pedestrian tracker, we simulate scale and center position errors as a Gaussian distribution around the hand annotated center. We fix the standard deviation of the center position error to 3 pixels, and of the width error to 3 pixels (aspect ratio is fixed at 1:2).



(a) KL-no (b) KL-CY(CYKLS) (c) KL-const

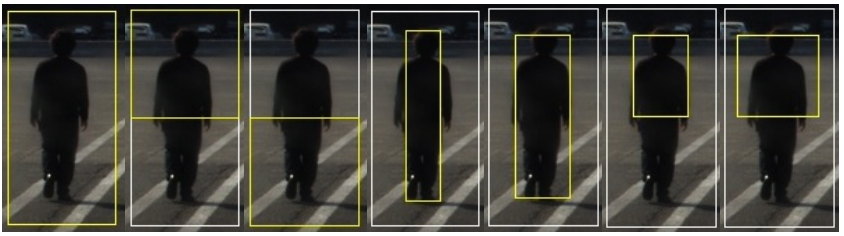
(d) KL-x (e) KL-exp (f) L1-no

(g) L1-CY (h) L1-const (i) L1-x

(j) L1-exp (k) L2-no (l) L2-CY

(m) L2-const (n) L2-x (o) L2-exp

Fig. 5. Comparison of methods with 7980 parameter sets. Every title represents norm-smoothing method. CYKLS gets the minimum Equal Error Rate(EER).



(a) block0 (b) block1 (c) block2 (d) block3 (e) block4 (f) block5 (g) block6

Fig. 6. ROI images. Yellow rectangle suggests the ROI, and white one suggests the hand annotated rectangle.

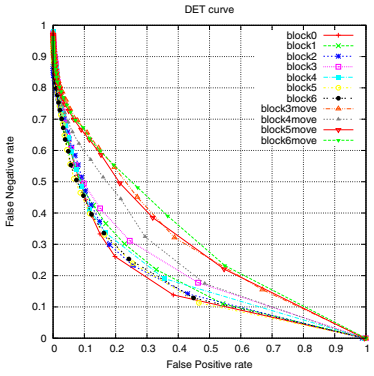


Fig. 7. ROI test results. "block0"(whole window) give the lowest EER.

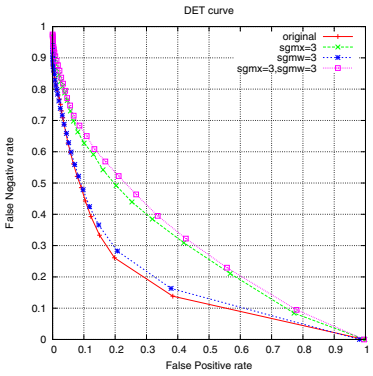
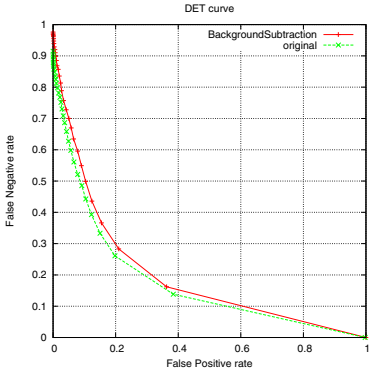


Fig. 8. Simulation result of tracking error

(a) Input image example



(b) Evaluation result

Fig. 9. Evaluation for background subtraction

The result is shown in Fig.8(Again, we use the one of the best parameter set $(n, l, K) = (2, 19, 8)$). As we can see, while width error hardly slow down the performance, position error cause performance degration in some degree.

4 Discussion

4.1 The Relationship with Fourier Transformation

CYKLS contains a new algorithm to detect non-periodic movement, but one might think it can be detected using Fourier Transformation for time series data. To check this, we carry out FFT to the time series data of output from the first part ("Calculation of appearance change") of CYKLS. However, we cannot see walking period peak shifting at dart point because pedestrians walk so short before dart in our dataset.

4.2 Background Subtraction

All of our dataset have white lines in the background of the pedestrian window(see Fig.1), so CYKLS may actually detect the presence of a white line. To check this, we did a background subtraction for our dataset in order to remove the white lines (see Fig.9a). We use median image following erosion and dilation method, because background is almost static. As it turns out, CYKLS can also detect dart without white lines in the background, though performance goes slightly down. This result is shown in Fig.9b (We fix $(n, l, K) = (2, 19, 8)$).

5 Conclusion

We have developed a new algorithm to detect dart by focusing on an appearance change, which we named CYKLS. We compared 15 baseline detectors including CYKLS, and found CYKLS marked the best score. And we also tested 11 ROI to compute CYKLS, and found the whole window marked the best score. As a result, we can detect pedestrians' dart with 22% Equal Error Rate, using a dataset which includes 144 dart scenes.

CYKLS can also be applied to detect other object's non-periodic movement, including cars and bicycles.

Future Work. In this paper we use static images, but we are ready for testing this algorithm on a moving car. In this case, tracking error and background change will become a big issue. To treat such issue, we will search for better features and smoothing methods. Also, dynamically adapting the parameters during the sequence may improve the performance.

References

1. Efros, A.A., Berg, A.C., Berg, E.C., Mori, G., Malik, J.: Recognizing action at a distance. In: ICCV, pp. 726–733 (2003)
2. Laptev, I.: Local Spatio-Temporal Image Features for Motion Interpretation. PhD thesis, Department of Numerical Analysis and Computer Science (NADA), KTH (2004)
3. Shechtman, E., Irani, M.: Space-time behavior based correlation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 405–412 (June 2005)
4. Yeffet, L., Wolf, L.: Local trinary patterns for human action recognition. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 492–497. IEEE (2009)
5. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24(7), 971–987 (2002)
6. Andriluka, M., Roth, S., Schiele, B.: Monocular 3d pose estimation and tracking by detection. In: The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, June 13-18, pp. 623–630. IEEE (2010)
7. Adam, A., Rivlin, E., Shimshoni, I., Reinitz, D.: Robust real-time unusual event detection using multiple fixed-location monitors. IEEE Trans. Pattern Anal. Mach. Intell. 30(3), 555–560 (2008)
8. Yin, J., Meng, Y.: Abnormal Behavior Recognition Using Self-Adaptive Hidden Markov Models. In: Kamel, M., Campilho, A. (eds.) ICIAR 2009. LNCS, vol. 5627, pp. 337–346. Springer, Heidelberg (2009)
9. Nater, F., Grabner, H., Van Gool, L.: Temporal relations in videos for unsupervised activity analysis. In: British Machine Vision Conference (2011)
10. Franke, U., Rabe, C., Badino, H., Gehrig, S.K.: 6D-Vision: Fusion of Stereo and Motion for Robust Environment Perception. In: Kropatsch, W.G., Sablatnig, R., Hanbury, A. (eds.) DAGM 2005. LNCS, vol. 3663, pp. 216–223. Springer, Heidelberg (2005)
11. Rabe, C., Müller, T., Wedel, A., Franke, U.: Dense, Robust, and Accurate Motion Field Estimation from Stereo Image Sequences in Real-Time. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 582–595. Springer, Heidelberg (2010)
12. Kullback, S., Leibler, R.A.: On information and sufficiency. Ann. Math. Statist. 22, 79–86 (1951)
13. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)