Pedestrian Detection and Tracking Using a Dynamic Vision Sensor

Israel Ruelas[1], Gustavo Torres-Blanco[1], Susana Ortega-Cisneros[3], E. Ulises Moya-Sánchez[1,2]

[1] Computer Science Posgraduate Department,
Uniersidad Autonoma de Guadalajara, Jalisco,
Mexico

[2] Barcelona Supercomputing Center, Barcelona,
Spain

[3] Electronic Design Laboratory,
CINVESTAV Guadalajara,
Mexico

elisrael2@hotmail.com, gustavo.blanco@edu.uag.mx,
sortega@gdl.cinvestav.mx, eduardo.moyasanchez@bsc.es

Abstract. Neuromorphic sensors such as the Dynamic Vision Sensor (DVS) emulate the behavior of the primary vision system. Its asynchronous behavior makes the data processing easier and faster due to the analysis is only in the active pixels. Pedestrian kinematics contains specific movement patterns feasible to be detected, like the angular movement of arms and feet. Some previous methodologies were focused on pedestrian detection based on the static shapes detection like cylinders or circles, however, they do not take into account the kinematic behavior of the body by itself. In this paper, we presented an algorithm inspired in K-means clustering and describes the analysis of the human kinematics based on DVS in order to detect and track pedestrians in a controlled environment.

Keywords. Dynamic vision sensor, pedestrian detection, pedestrian tracking.

1 Introduction

The ability to extract local features in a visual scene is the fundamental building block in a wide range of computer vision solutions. Dynamic Vision Sensors (DVS) detect local features, using a change in luminance above a certain threshold in log scale, these sensors include a high dynamic range and power efficiency, making it ideal for outdoor usage on embedded systems in autonomous systems [3, 2].

Automatic pedestrian detection and tracking have a great interest in some areas, like security and protection of walking people in any wheater condition fog, rain, among others [1, 9, 14, 18]. For that reason, the analysis of human kinematics plays a crucial role in the development of computing vision algorithms [1, 17, 14]. Due to the physical characteristics, the extraction and the complex movements in these tasks, the algorithms can be expensive in time processing and memory consumption [13, 15].

In general, computer vision methods use the human shapes with an color camera [13, 17, 5]. The main difference between the Dynamic Vision Sensors DVS and RGB cameras is that the DVS reports only the stimulated pixels by the object movement even in bad weather conditions [9]. The DVS is the electronic implementation of the mammalians primary vision algorithm. DVS128 from INILABS3 and it has a matrix of 128x128 pixels. DVS has a dynamic range of operation greater than 120dB or 6 decades, which makes it an ideal sensor for poor light applications[8].

Some algorithms which use RGB cameras have proposed include the base-point extraction of

Computación y Sistemas, Vol. 22, No. 4, 2018, pp. 1077–1083
ISSN 1405-5546
doi: 10.13053/CyS-22-4-3080

silhouette and shape [17]; however, they don't take into account the kinematic behavior of the body by itself. In addition, other models were able to analyze the kinematic behavior of objects, using Bayesian filters with DVS [15].

In this work, we propose an algorithm to detect and track pedestrians using an asynchronous analysis generated by the movement detected by the DVS, using Java Address Event Representation (jAER), and DBscan (Density-Based Spatial Clustering with Application Noise). Additionally, we present some of the body parts detection (arms, legs, head) through the identification of the movement pattern of each of those. One advantage of this approach is having a fast response sensor (up to 8 M events per second), even in poor illumination conditions and we do not need any preprocessing algorithm.

2 Materials and Methods

2.1 Dynamic Vision Sensors

The mammalian visual system functioning is sometimes compared, incorrectly, with the acquisition process of a camera, which transduces light intensities of the visual field in a point-to-point correspondence array. The primary visual system, in contrast, parses scenes into different components separating the foreground from the background, in order to determine which light stimulus belongs to one object and to the background [10].

The retina is the light (biological) sensor, which provides the capacity of vision to mammalians. The rods and cones, transduce the luminosity signals into electro-mechanic ones that are then decoded and interpreted by the brain [10]. After the retina, there are different layers that contain specialized neurons interconnected to decode and interpret the scene.

The main difference between the DVS and any other technologies for vision is that the DVS reports only the stimulated pixels by the movement or contrast changes in an asynchronous manner. While using other technologies, the whole frame is captured and sent on a given sample rate. In the DVS, the pixels that were not stimulated do not report any data, so a static image will not produce any output, separating the foreground or the object from the background [6, 16].

Each pixel can be stimulated separately, with 14 bits intensity range, for a (x, y) matrix position, time base or timestamp, and polarity of the stimulation. The photoreceptors on the DVS produce a logarithmic response, and the Asynchronous Event Representation protocol (AER) is used to transfer the information from the sensor to the central unit processor. AER is a robust high-speed protocol, that transports the information generated by the pixels [11, 5]. The computed speed of the neuromorphic sensor, can be at least 2 M of events per second and a maximum of 8 M of events per second [11, 5].

2.2 Fundamentals of Human Kinematics

The automatic detection and tracking of pedestrians are very important in many scenarios such as airports, cross street, borders, and crowds [19]. For that reason, the analysis of human kinematics plays a crucial role in the development of computing vision algorithms. Additionally, the human kinematics contains enough information that can be used to detect identity and the emotional state, among others [12]. However, the background, illumination, and kinematics of a human being together complicate the detection and tracking.

The pedestrian kinematics models for computer vision usually are based on methods like volumetric models or 2D contour identification, using predefined shapes [4]. In our case, we take several regions (shapes like circles, cylinders) and compute centroid in order to obtain different regions of the body, the direction of the movement, angular speed, and position.

The pipeline of the pedestrian detection could be easy using the DVS128. The shape is getting with the activation of the pixels (events), clustering the whole shape, and after that, searching of the of the body waves of the extremities and head around the center of mass see Fig 1. As observed during the gait cycle analysis, these movements are repeatable across different steps.

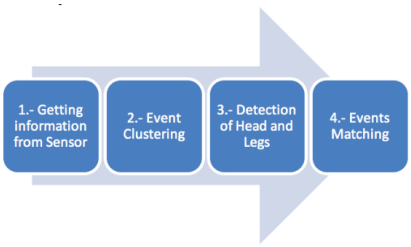

Fig. 1. Pipeline for pedestrian detection

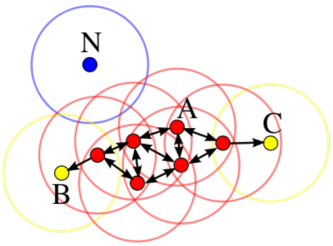

Fig. 2. The distance between point N and other points A, B, C is calculated

2.3 Centroid and Clustering Event Detection

The event detection through a DVS remove the background because the sensor by itself only reports moving objects (contrast changes). The centroid of all points N (see Fig 2) can be calculated through the use of equations 1, where X_c, Y_c provide the position of the centroid, and X_i, Y_i is a pixel of the image, also it is the number of pixels in the pattern [7, 4]:

$$(X_c, Y_c) = \frac{1}{N_b} \sum_{i=1}^{N_b} (X_i, Y_i). \tag{1}$$

The centroid is used to track objects, specifically using X_c, Y_c. However, even that the sensor reports only the dynamic pixels, can be is some photoreceptor that activates by biases currents into the circuitry. These pixels should be removed by using a filtering event or even only consider the denser sides of the picture (event) that represents the core movement. The points A, B, C in Fig 2 represent an example botom points (legs points) tracking using N (general centroid of the shape) as a reference. Using this information, is possible to compute the angular velocity with the difference of the points B and C.

Our algorithm is an adaptation of K-means with asynchronous input events, the pseudocode for DBscan (Density Based Spatial Clustering with Application Noise) divided into two main processes, events inside a radius, and minimum events in the cluster See Fig 3.

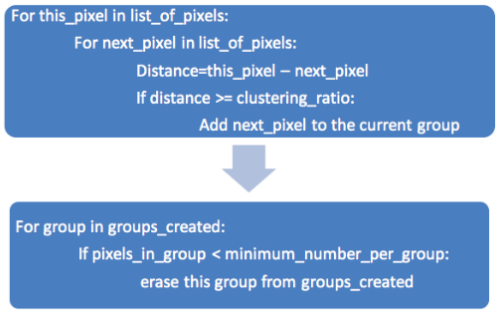

Fig. 3. Pseudocode for DBscan divided into two main processes. Events inside a radius, and minimum events in the cluster

2.3.1 Detection of Head and Legs

For kinematic pedestrian detection, the body parts need to be detected through the identification of the movement pattern of each of those. In the literature there are different patterns observed during the human walking that seems to be easy to implement in a first instance, however, they become complicated when starting to analyze a lot of data, some examples are shown in Fig. 4. The events are mixed due to some noise and in the kinematic gait some parts are occluded. By taking advantage of the main centroid N the separation of upper and lower side of the body is straightforward, as shown in Fig 5. The upper centroids correspond to the head/arms moving and the lower centroids corresponding to the legs in the step by step. The head detection is performed by using the standard deviation computed with equation 2.

$$\sigma_{head} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y - y_i)^2}. \qquad (2)$$

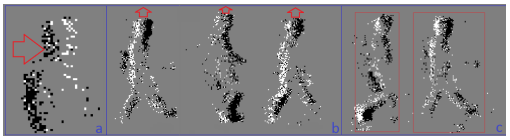

Fig. 4. The events have noise and in the kinematic gait some parts are occluded. Some examples are: a) Noise concavity corresponds to the neck, b) Vertical oscillation of head when human is walking and c)Change of the width while the human walking action

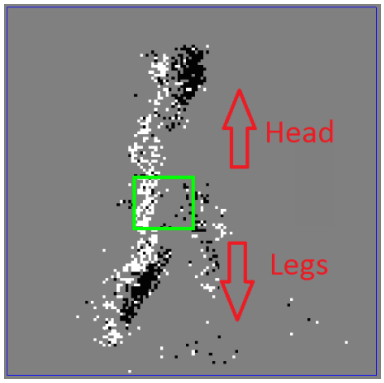

Fig. 5. Centroid of human walking around

To detect the legs is necessary to consider the centroids located on the lowest side of the picture, by analyzing them these are mainly integrated into 2 groups that correspond to each leg. Assuming that the legs move separately, is possible to compute angular velocity. The proper distribution identification of the centroids corresponding to the legs is a key factor of pedestrian identification, as shown in Figure 6. The Figure 7 shows the complete algorithm jAER.

3 Experimental Setup

The experiment was held on a room with 2 windows, with several objects in the background, with normal illumination conditions (60 W bulbs, white light, and natural light), with 4x4 m of

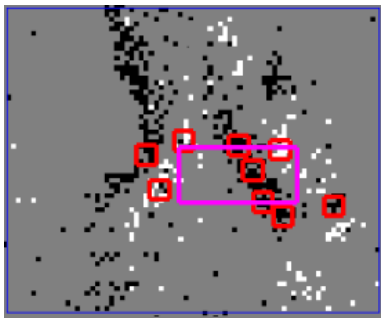

Fig. 6. Centroids corresponding to the legs are grouped for each leg

minimum space. The camera was placed in a perpendicular position of the walking direction of the persons.

This experiment tries to sense people on one walking direction only and one perspective (sagittal). Given this situation, the DVS must be placed on the perpendicular side of the walking direction, allowing the full body picture to be shown while walking across the scenario, see Fig. 8.

3.1 DVS Characteristics

The sensor used in this study is the DVS128 (tpmdiff128) INILABS3 with 128x128 pixels. The lens included on the sensor has a Focal length of 4:5mm and Amplitude of Field of View (AFOV) of 60 degrees. The sensor must be placed at a distance of 1.15 m in order to have a Field of View (FOV) of a 2x2 meters height/width that allows us to obtain a full frame of 3 to 4 steps of the person, as shown in Figure 8.

The proper environment setup allows the human body to stride 3-4 times on the scenario. The sampling videos were acquired on the environment described above, containing one human walking in front of the DVS. The changes in the illumination do not affect the experiment setup.

4 Results

The main result of our algorithm is shown in Fig 9, asynchronous adaptation of K-means is possible and effective. First, a centroid is computed, and up to 15 frames need the algorithm to detect a

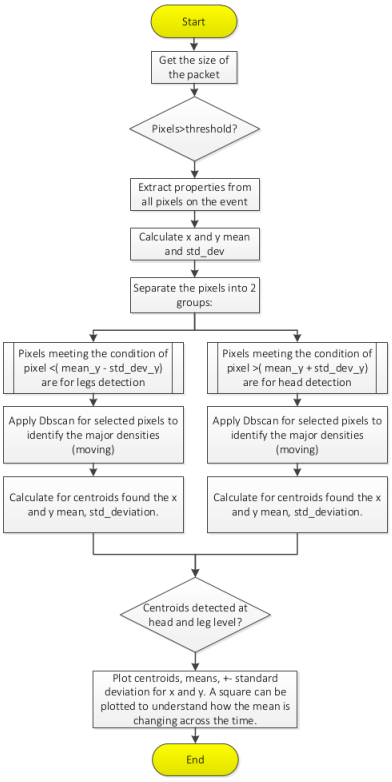

Fig. 7. Diagram of pseudocode and implementation in jAER

pedestrian, (see Fig 9) due to the kinematics of pedestrian not show all points of the legs at the walk beginning.

When the head and legs are detected a pedestrian is recognized. In addition with the centroid information (position, timestamp, and sign) we can compute, the pedestrian waking direction, the speed, and even the angular velocity. Head and legs detection are shown in Fig 10 (see squares and the lines).

One advantage of our approach is the fast response sensor (up to 8 M events per second) due to we do not need a preprocessing algorithm to segment, even in poor illumination conditions. In the future work, we want to improve this algorithm to detect and track multiple pedestrians and increase the performance of our algorithm.

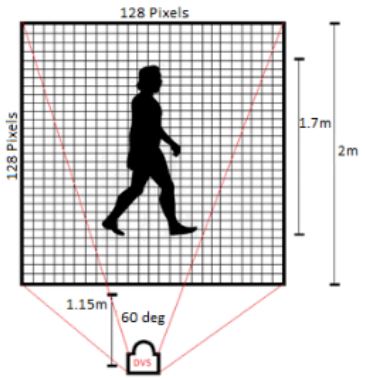

Fig. 8. Camera mounted in front of the human walking (this is only an illustrative picture)

5 Conclusions and Future Work

We presented an implementation of a sagittal pedestrian detector based on the walking kinematics behavior using the neuromorphic retina (DVS). The kinematics of the human gait was detected using the DBScan algorithm together with the centroid detection, inspired by K-means clustering.

Our algorithm can detect the head, center, and legs through the identification of the movement pattern. Our adaptation of K-means for asynchronous events allows us to track the pedestrian and compute information of the pedestrian such as the direction and speed.

One advantage of our approach is the fast response, because of the DVS sensor (up to 8 M events per second), in addition, we can detect the pedestrian, even in poor lighting conditions and we do not need a preprocessing algorithm to segment the pedestrian. In the future work, we want to improve this algorithm to detect and track multiple pedestrians and increase the performance of our algorithm.

Acknowledgements

This work is partially funded by CONACYT with PNPC grant and SNI Grant.

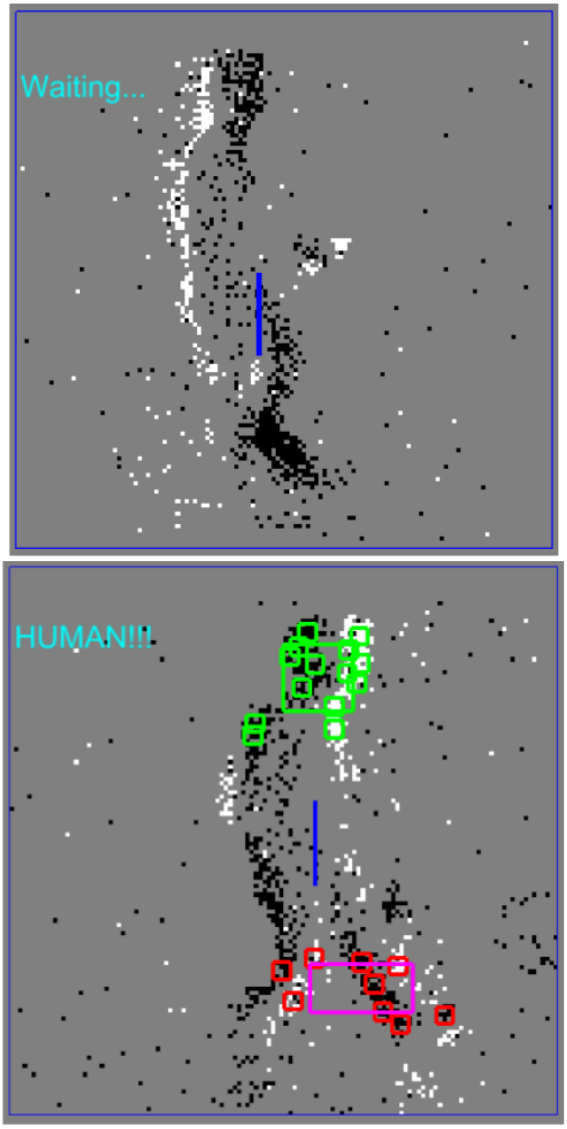

Fig. 9. Pixel densities detected (green squares) as well the horizontal mean (red line) on legs and head

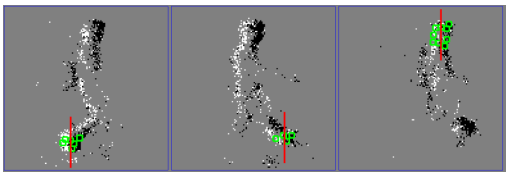

Fig. 10. Pixel densities detected (green squares) as well the horizontal mean (red line) on legs and head

References

1. **Aggarwal, J. K. & Cai, Q. (1999).** Human motion analysis: A review. *Computer vision and image understanding*, Vol. 73, No. 3, pp. 428–440.
2. **Chen, N. F. (2018).** Pseudo-labels for supervised learning on dynamic vision sensor data, applied to object detection under ego-motion. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 644–653.
3. **Cohen, G. K. (2016).** *Event-based feature detection, recognition and classification*. Ph.D. thesis, Paris 6.
4. **Dai, Q., Qiao, J., Liu, F., Shi, X., & Yang, H. (2012).** A human body part segmentation method based on markov random field. *Control Engineering and Communication Technology (ICCECT), 2012 International Conference on*, IEEE, pp. 149–152.
5. **Eibensteiner, F., Kogler, J., Sulzbachner, C., & Scharinger, J. (2011).** Stereo-vision algorithm based on bio-inspired silicon retinas for implementation in hardware. *International Conference on Computer Aided Systems Theory*, Springer, pp. 624–631.
6. **Franco, J. A. G., del Valle Padilla, J. L., & Cisneros, S. O. (2013).** Event-based image processing using a neuromorphic vision sensor. *Power, Electronics and Computing (ROPEC), 2013 IEEE International Autumn Meeting on*, IEEE, pp. 1–6.
7. **Fujiyoshi, H., Lipton, A. J., & Kanade, T. (2004).** Real-time human motion analysis by image skeletonization. *IEICE Transactions on Information and Systems*, Vol. 87, No. 1, pp. 113–120.
8. **Gangwar, D. S., Tiwari, T., & Singh, B. (2008).** Electronic implementation of biologically inspired neuromorphic vision sensor. *Modeling & Simulation, 2008. AICMS 08. Second Asia International Conference on*, IEEE, pp. 410–414.
9. **Han, W.-S. & Han, I.-S. (2014).** All weather human detection using neuromorphic visual processing. In *Intelligent Systems for Science and Information*. Springer, pp. 25–44.
10. **Kandel, E. R., Schwartz, J. H., Jessell, T. M., of Biochemistry, D., Jessell, M. B. T., Siegelbaum, S., & Hudspeth, A. (2000).** *Principles of neural science*, volume 4. McGraw-hill New York.

11. **Kogler, J., Sulzbachner, C., & Kubinger, W. (2009).** Bio-inspired stereo vision system with silicon retina imagers. *International Conference on Computer Vision Systems*, Springer, pp. 174–183.
12. **Lee, C.-S. & Elgammal, A. (2012).** Style adaptive contour tracking of human gait using explicit manifold models. *Machine Vision and Applications*, Vol. 23, No. 3, pp. 461–478.
13. **Li, M., Yang, T., Xi, R., & Lin, Z. (2009).** Silhouette-based 2d human pose estimation. *2009 Fifth International Conference on Image and Graphics*, IEEE, pp. 143–148.
14. **Litzenberger, M., Kohn, B., Belbachir, A., Donath, N., Gritsch, G., Garn, H., Posch, C., & Schraml, S. (2006).** Estimation of vehicle speed based on asynchronous data from a silicon retina optical sensor. *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*, IEEE, pp. 653–658.
15. **Piatkowska, E., Belbachir, A. N., Schraml, S., & Gelautz, M. (2012).** Spatiotemporal multiple persons tracking using dynamic vision sensor. *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, IEEE, pp. 35–40.
16. **Sulzbachner, C. & Kogler, J. (2010).** A load balancing approach for silicon retina based asynchronous temporal data processing. *Software Engineering and Advanced Applications (SEAA), 2010 36th EUROMICRO Conference on*, IEEE, pp. 431–435.
17. **Tong, M.-l. & Bian, H.-q. (2011).** 3d human motion tracking by using interactive multiple models. *Journal of Shanghai Jiaotong University (Science)*, Vol. 16, No. 4, pp. 420–428.
18. **Wachs, J. P., Kölsch, M., & Goshorn, D. (2010).** Human posture recognition for intelligent vehicles. *Journal of real-time image processing*, Vol. 5, No. 4, pp. 231–244.
19. **Wang, C.-W. & Hunter, A. (2010).** Robust pose recognition of the obscured human body. *International journal of computer vision*, Vol. 90, No. 3, pp. 313–330.

Article received on 20/01/2018; accepted on 30/04/2018.
Corresponding author is Israel Ruelas.