Analysis of Object and Its Motion in Event-Based Videos

Abstract

In the recent years, a new generation of cameras sensitive to pixel intensity variation rather than the traditional pixel intensity value has been introduced. These cameras called Dynamic Vision Sensors (DVSs) have recently attracted significant research interest. Conventional camera captures the intensity of all pixels in the sensor and generate an entire image to produce a frame. This is then repeated at a fixed rate to produce a video stream. Being totally different with the old frame-based videos, the captured videos of Dynamic Vision Sensors are polarized events, i.e. points in 3-D spatio-temporal space, indicating the polarity, location and time properties of the pixels with variable intensities. When there is a variation in a pixel intensity, a polarized event is created in the form of a vector with three elements (t,x,y). t shows the instance of the variation and (x,y) defines the position of the pixel. Moreover, polarization of the event shows the direction of the change in the pixel intensity.

For event-based videos, some algorithms have been proposed for object tracking, optical flow extraction, human action recognition, etc. Still there are many potential capabilities of these cameras that have not been used or explored. Extracting features of these videos requires a comprehensive understanding and novel procedures accordingly.

We began our research by focusing on the existing algorithms for DVS videos. We found out that motional analysis and feature extraction are trending topics in DVS videos and we began to work on these topics. This thesis introduces two different approaches to objects' motion in event-based videos. Hough transform and edge detection are also performed in event-based videos as two important methods of feature extraction.

This research work presents a novel framework for investigating the objects motion in event-based videos and extracting the edge information subsequently. In the event-based videos, the events normally occur in the moving edge areas. We consider the events as some points in the spatio-temporal space. Ignoring noise, for each small spatio-temporal window in a moving edge area, we expect all events to be on a 3-D plane. The orientation of this plane depends on both edge direction and velocity. By approximating the object boundary as a series of linear elements, we derive a procedure based on principal component analysis to estimate their orientation and speed. According to the well-known aperture problem in machine vision, the velocity estimated at this stage is the normal portion of the actual velocity since any displacement along the edge orientation cannot be recognized in a small spatio-temporal window.

The normal velocities are utilized in a larger window which covers a whole object to estimate its actual velocity. We define a cost function based on the difference

between actual normal velocities and calculated normal velocities at the previous stage. Minimization of this cost function results in an estimation of the actual velocity which is a useful parameter in some applications e.g. object tracking. Moreover, we propose a procedure for localizing the edge based on regional exposure time and edge-dependent Gaussian filtering of the events. The regional exposure time is adjusted based on the normal velocities of edge pixels. This avoids any blurred edge which is a direct consequence of higher normal velocities. The orientation of Gaussian filter causes the maximum blurring effect along the edge direction and improve its connectivity for a better edge extraction. Any kind of discontinuity in the object texture is appeared by the local variation of the pixel intensities. When the object is moving, these variations generate many unwanted events that we should tackle them as noise. Noise is another challenge of these videos that we can suppress by detecting the outliers in many stages of our algorithm.

Another approach to motion analysis is based on well-known Hough transform for detecting straight lines. Hough transform has been widely used to detect lines in images captured by conventional cameras. We develop an event-based Hough transform and apply it to DVS output stream. The proposed algorithm is implemented in a spiking neural network to detect lines on DVS output. Spikes (events) from DVS are first mapped to Hough transform parameter space and then sent to corresponding spiking neurons for accumulation. A spiking neuron will fire an output spike once it accumulates enough input contributions and then reset itself. The output spikes of the spiking neural network represent the parameters of detected lines. An event-based clustering algorithm is applied on the parameter space spikes to segment multiple lines and track them. In our spiking neural network, a lateral inhibition strategy is applied to suppress noise lines from being detected. This is achieved by resetting a neuron's neighbors in addition to itself once the neuron fires an output spike. As an improvement to the work done, we deal with detecting small lines at the frame corners subsequently. In addition, the inhibitory window shape is optimized to suppress the lines which are close together in Cartesian space and are not necessarily close together in parameter space as assumed initially.

Finally, we perform many experiments for verification of our proposed algorithms. Some of them are performed on computer-generated videos while others performed on real DVS videos. The results show that our proposed methods have acceptable performances in recognizing the edges and estimating their velocities.