Practical Real-Time System for Object Counting based on Optical Flow

Jacky Baltes[1], Amirhossein Hosseinmemar[1], Joshua Jung[1], Soroush Sadeghnejad[2], and John Anderson[1]

[1] University of Manitoba, Winnipeg, MB, R3T 2N2, Canada,
jacky@cs.umanitoba.ca, http://www.cs.umanitoba.ca/~jacky
[2] Amirkabir University of Technology, No. 424, Hafez Ave., Tehran, Iran,
s.sadeghnejad@aut.ac.ir, http://autman.aut.ac.ir
[3]

Abstract. This paper describes a simple and effective system for counting the number of objects that move through a region of interest. In this work, I focus on the problem of counting the number of people that are entering or leaving an event. I design a pedestrian counting system that uses a dense optical flow field to calculate the integral of the optical flow in a video sequence. The only parameter used in the system is the the estimated integral flow for a single person. This parameter can be easily calculated from a short training sequence. Empirical evaluations show that the system is able to provide accurate estimates even for complex sequences in real-time. The described system won 2nd place in the pedestrian counting computer vision competition at the IEA-AIE 2014 conference.

1 Introduction

This paper describes a practical real-time system for counting objects that pass through a region of interest ROI. There are many applications for these type of systems in security, urban planning, and crowd control. This paper focuses on the specific problem of counting the number of people that enter or exit via a doorway (i.e., a sub-problem of pedestrian counting). Estimating the number of people entering or leaving a major event or venue (e.g., Disneyland) is important to ensure the safety of visitors. Overcrowding may lead to injuries or even deaths when trying to evacuate (e.g., in case of a bomb threat or fire) or when visitors panic into a stampede. Unfortunately, there are many cases in recent years where people have been injured or died due to overcrowding at an event. Famous examples of human stampedes due to overcrowding are the Who Concert in Cincinnati, U.S.A. on 3rd Dec. 1979, the Hillsborough Stadium, Sheffield, England, on 15th April 1989, and the Love Parade in Duisburg, Germany, on 24th July 2010 [1].

Figure 1 shows some images of the type of pedestrian counting that I address in this research. The first row shows a simple scenario where people enter from the top of the image through a well defined gate area and exit at the bottom.

The people move in different sized groups (shown are a single person (left), two people (middle), and three people (right).

The second row shows a more complex scenario. The region of interest is larger and the perspective distortion of the camera is more pronounced. Some people enter or leave on the edges of the field of the view of the camera, which results in only legs or arms being seen in the video. Furthermore, the sun casts long and hard shadows across the ground. There are several non-people objects that cross the scene. For example, baby carriages are shown in the middle and left image and an umbrella in the right image. There are also people moving in the opposite direction as shown in the right image.

The images shown in Fig. 1 were taken from sample videos provided by the organizers of the computer vision competition of the 27th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA-AIE) [2, 3], held on June 6th 2014 in Kaohsiung, Taiwan.

The images in the first row were taken at the Sun Yat Tse Memorial Hall in Taipei, Taiwan. The images in the second row were taken at the Flora Expo in Taipei, Taiwan.

Fig. 1. Sample scenes for pedestrian counting as used in this research. Images taken from the competition website. Videos were sample videos for the IEA-AIE-2014 Computer Vision Competition.

2 Design

As can be seen from the sample images, the difficulty of the problem can vary greatly between image frames and a 100% accurate solution would require an extensive collection of background knowledge to deal with the various objects that may be present in the scene such as baby strollers, umbrellas, and packages.

The large number of possible models leads to increased processing times for such systems and they often are not able to maintain real-time performance, i.e., in and out counts can be updated at 15 frames per second or higher frame rate. Furthermore, these models would require extensive calibration and tuning of the algorithm, which makes them unsuitable to be used in situations where setup needs to be done quickly and where only untrained personal is available.

However, the main goal of our system was to help in crowd control during large events, so I believe that: (a) real-time monitoring, (b) fast setup and (b) setup by untrained workers are essential in those circumstances.

I therefore decided to focus on a system that provides accurate estimates in a majority of likely scenarios while maintaining real-time performance and to ignore errors introduced by outliers.

The processing pipeline of our system consists of four steps:

1. **Pre-processing**: the image is cropped to the region of interest and the images are converted into grey-scale.
2. **Optical Flow**: a dense optical flow field of the images is calculated.
3. **Noise Removal**: flow vectors below a threshold are discarded.
4. **Optical Flow Field Discretization**: The flow field direction is discretized into the two directions perpendicular to the gate axis.
5. **Optical Flow Field Integration**: The flow field is integrated according to the two directions perpendicular to the gate axis.
6. **Pedestrian Counting**: The average integral of the flow field is used to estimate the number of people that entered or left the area.

Each processing step is described in more detail in the following subsections.

2.1 Cropping and Conversion to Grey-Scale

In the first part of my system, the images are cropped to a region of interest representing the gate area. Furthermore, the axis of the gate is input as a parameter thus defining the In and Out directions of my pedestrian counter.

Next the image is converted into grey-scale to be used as input for the optical flow computation.

In this step, I have also considered the possibility of adding a pre-processing step to enhance the contrast of images. In practice, that did not improve the results in our evaluation.

2.2 Farnebäck Optical Flow Computation

At the heart of my system is the computation of dense optical flow field. There are many algorithms described in the research literature to compute a dense optical flow field starting from seminal work by Horn and Schunk [4, 5].

In this system, we used a dense optical flow algorithm developed by Farnebäck [6]. The main reason for choosing this algorithm is that a robust and efficient implementation of this algorithm is available in the OpenCV library [7]. OpenCV is a open-source library which includes implementations of many different image capture, low level processing, and higher level processing algorithms. OpenCV also provides bindings to many different programming languages such as C++ and Python.

In my pedestrian counting system, I used the OpenCV library with the Python bindings.

The images in Fig. 2 show the output of the Farnebäck optical flow algorithm on the sample video sequences. The direction of the flow is indicated by the direction of the small lines and the magnitude of the flow is given by the length of the line segment where a 10 pixel line segment corresponds to a 1 pixel inter frame flow.

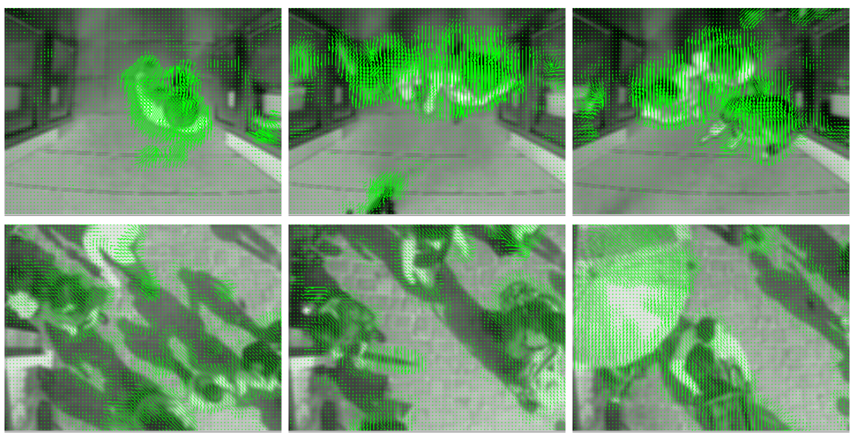

Fig. 2. Dense optical flow field calculated from the sample sequences. Direction and magnitude of the optical flow is shown by the direction and length of the line segments.

2.3 Noise Removal

As can be seen in the images, the direction and magnitude of the optical flow is more prominent in the simple images in the first row. The images in the second row show that the shadows of the people move as well and therefore generate an optical flow.

In some of the images, noise introduced by the camera resulted in small movement of features. I added a noise removal step where optical flow fields that were below a threshold were removed.

2.4 Optical Flow Field Discretization

In this step, the direction of the optical flow vectors are being discretized into one of eight directional buckets with major axis (0 deg., 45.0 deg., 90 deg., 135 deg., 180.0 deg., 225 deg., 270 deg., and 315 deg.). The width of each bucket is 45 degrees which leads to the following boundaries for the buckets: (-22.5 deg. to 23.0 deg., 22.5 deg. to 68.0 deg., 67.5 deg. to 113.0 deg., 112.5 deg. to 158.0 deg., 157.5 deg. to 203 deg., 202.5 deg. to 248.0 deg. 247.5 deg. to 293.0 deg., 292.5 deg. to 338.0 deg.). The bucket boundaries are overlapping by 0.5 deg. to avoid sudden cutoffs during the integration along the boundary of a bucket.

2.5 Optical Flow Field Integration

The integral of the optical flow field of bucket is denoted as R_i. R_i is calculated given the following formula where i corresponds to a bucket in the optical flow field discretization, and t is time. B_i is the major axis of bucket i (e.g., 0 deg., 45 deg., ...). B is the set of all buckets. T is the length of the entire video sequence. The dot product calculates the projection of the flow onto the major axis of its corresponding bucket(s).

$$\forall i \in B : R_i = \int_0^T F_i(t)\,\mathrm{d}t$$

where

$$F_i(t) = F(t) \bullet B_i \text{ if } F(t) \text{ assigned to bucket } i, 0 \text{ else.}$$

The current value of the integral flow field for all buckets for the video sequences shown in Fig. 1 are shown in Fig. 3.

2.6 Pedestrian Counting

As can be seen, individuals or groups of people result in a sharp increase in the integral of the optical flow. The top figure of Fig. 3 shows timeline for a video showing three people entering (90 deg.) and 23 people leaving the gate. The timeline in the top figure shows that the three people that in two groups: a single person and a group of two.

Furthermore, I estimate the average integral flow per person by diving the total integral of the flow field by the number of people that crossed the gate. In this case, the average flow of the people moving through the gate.

The estimate of the integral flow per person for video 1 (E_{V1} is given as the average flow for the up and down directions (I_{Up}

$$E_{V1} = \frac{I_{Up} + I_{Down}}{N_{Up} + N_{Down}} \approx \frac{0.5 * 10^7 + 3.5 * 10^7}{3 + 23} = 1.5 * 10^6$$

The second video shows approximately 20 people entering and approximately 64 people leaving. The exact number of people is impossible to determine since some are hidden behind an umbrella and some are only indicated by their shadow. I use a similar method to determine an average integral flow for the second video: $E_{V2} = (0.1 * 10^8 + 0.9 * 10^8)/(20 + 64) \approx 1.1 * 10^6$.

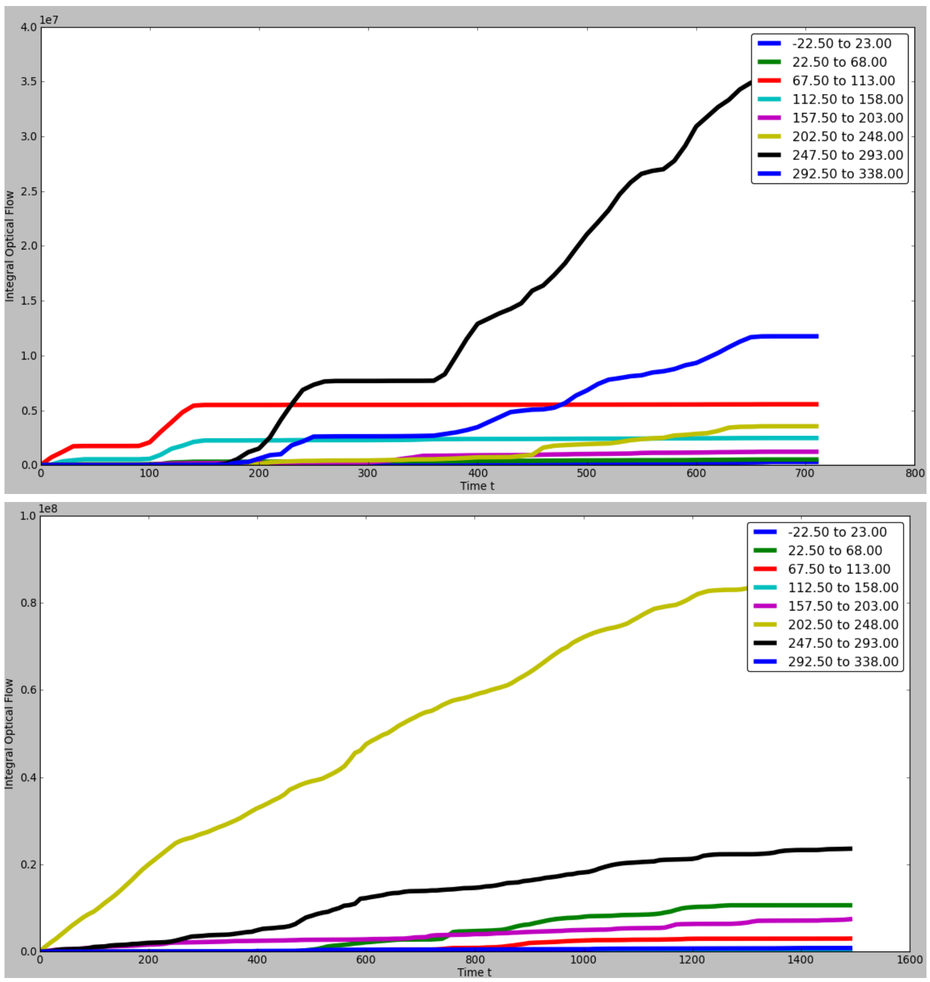

Fig. 3. Integral Optical Flow Field for sample video sequences. Top figure shows 3 people entering and 23 people leaving. Bottom figure shows approximately 20 people entering and approximately 64 people leaving.

So for each video, a short training sequence is used to determine the average integral flow field E for a person.

The total number of pedestrians that entered or left the gate are determined by dividing the total integral flow field by the estimated flow field for each person. This simple measure turned out to be a reasonable good estimator for our pedestrian counter as is described in the next section.

3 Evaluation

I evaluated the system on the benchmark videos provided by the organizers of the IEA-AIE 2014 computer vision competition.

Video	Error in percent
SYS Memorial Hall normal	3.8
SYS Memorial Hall advanced	8.00

The system was also used in the IEA-AIE-2014 Computer Vision Competition on June 6th 2014 in Kaohsiung, Taiwan. My system won 2nd place in the competition against twelve other teams.

4 Conclusion

This paper describes the implementation of a practical system for counting objects (e.g., pedestrians or cars) as they traverse a ROI. The system is based on calculating the integral of the optical flow perpendicular to traversal axis. Some limitations of the system may be addresses in the future.

One assumption of the system is that the objects have approximately the same size, so that the average of the integral of the optical flow field can be used as an estimator. This assumption was warranted when the system was used to count pedestrians since there was not much difference in the integral of the flow field for small children or large adults. However, a more sophisticated classification algorithm may be necessary when the system is used to count vehicles that pass through an intersection. In this case, the integral of the optical flow field would be very different for a car or a truck. Note that the ability to distinguish between types of vehicles is not always important. The standard way of estimating traffic uses induction loops (wires) that are placed on top of the lane. Similar to our system, these systems are also unable to distinguish between one truck with a trailer or two cars driving closer together. In many cases, manual counting is used when the type of traffic is important.

Another weakness of the system is the fact that the integral of the optical flow for an object remains constant. This assumption is problematic in some circumstances. As can be seen in the images in the second row of Fig. 2, a person that casts a shadow on the ground will have a larger optical flow integral as one that follows closely behind a second person. The additional component of the integral of the optical flow due to the shadow is not taken into consideration in our system. Improvements to our system could be made by removing optical flows that are most likely associated with shadows. This may require providing some information about the lighting condition in the scene (e.g., direction of the sun) to the system.

However, as can be seen in our evaluation, the pedestrian counting system worked reasonable well over a large set of benchmark problems without dealing with shadows separately.

References

1. Patrick Weidinger. 10 tragic human panics and stampedes. WWW, November 2010. Visited 1st Aug. 2014.
2. 2014 international computer vision competition with iea-aie 2014. WWW, June 2014. Visited 1st July 2014.
3. The 27th international conference on industrial, engineering & other applications of applied intelligent systems (iea-aie). WWW, June 2014. Visited 1st July 2014.
4. Berthold K Horn and Brian G Schunck. Determining optical flow. In *1981 Technical Symposium East*, pages 319–331. International Society for Optics and Photonics, 1981.
5. John L Barron, David J Fleet, and Steven S Beauchemin. Performance of optical flow techniques. *International journal of computer vision*, 12(1):43–77, 1994.
6. Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Image Analysis*, pages 363–370. Springer, 2003.
7. The opencv library. WWW, 2014. Visited 1st Aug. 2014.