5-2018

Visualizing the Motion Flow of Crowds

Zheng Zhao
Purdue University

VISUALIZING THE MOTION FLOW OF CROWDS

by

Zheng Zhao

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Computer Graphics Technology

West Lafayette, Indiana

May 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Yingjie Chen, Chair

Department of Computer Graphics Technology

Dr. Tim McGraw

Department of Computer Graphics Technology

Dr. Jeffrey M.Siskind

School of Electrical and Computer Engineering

Approved by:

Dr. Bedrich Benes

Graduate Program Co-Chair

Dr. Colin M. Gray

Graduate Program Co-Chair

To my loving parents,

Qing Chang and Shengbao Zhao,

who are always there to support me. Their affection, encouragement, and love help me to overcome every obstacle in my life and pursue my dream.

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my advisor, Professor Yingjie Chen, for his guidance and advices of my Master study and thesis research. His rich experience on visualization helps me a lot on the front-end programming and the design of this research. His cultivation and education also provide a high freedom of learning and choosing for me in these two years.

Secondly, I would like to thank the rest of my committee members, Professor Tim E.McGraw and Professor Jeffrey Siskind, for their advices during my thesis research. Their unique viewpoint helps me think more comprehensive about this thesis and the difficulties in the procedure of application. Without their help, I cannot successfully finish the project in this year.

Finally, I would like to thank my parents for their support of my Master study in the United States and the contributor of those remarkable open-source codes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Zhao, Zheng. M.S.
Institution: Purdue University
Degree Received: May 2018
Title: Visualizing the Motion Flow of Crowds.
Committee Chair: Yingjie Chen

In modern cities, massive population causes problems, like congestion, accident, violence and crime everywhere. Video surveillance system such as closed-circuit television cameras is widely used by security guards to monitor human behaviors and activities to manage, direct, or protect people. With the quantity and prolonged duration of the recorded videos, it requires a huge amount of human resources to examine these video recordings and keep track of activities and events. In recent years, new techniques in computer vision field reduce the barrier of entry, allowing developers to experiment more with intelligent surveillance video system. Different from previous research, this dissertation does not address any algorithm design concerns related to object detection or object tracking. This study will put efforts on the technological side and executing methodologies in data visualization to find the model of detecting anomalies. It would like to provide an understanding of how to detect the behavior of the pedestrians in the video and find out anomalies or abnormal cases by using techniques of data visualization.

CHAPTER 1. INTRODUCTION

1.1 Background

Video surveillance system such as closed-circuit television cameras is widely used by security guards to monitor human behavior and activities to manage, direct, or protect people. Often, we see a control room with a matrix of monitors showing scenes from many areas in real time. The security guard has to keep a close look at each screen so that he can keep track of all the details without missing any traces of abnormal behaviors or emergency events. Some systems record prolonged videos so that analyst can rewind, and track events happened in the past. With the quantity and prolonged duration of the recorded videos, it requires a huge amount of human resources to examine these video recordings and keep track of activities and events.

In order to solve these issues mentioned above, numerous researchers have put efforts on exploring the methods of making the surveillance system more intelligent and time-saving. The common methods of visual surveillance can be categorized into 4 parts: object detection, object tracking, data analysis and data visualization. It can also be regarded as a completed process to analyze a video. The surveillance system will detect the video and pick up the interested objects from the background video with the classified labels including human, vehicle, animals, etc. The objects of interest will be kept tracking in the following frame sequence of the video. Then the system will try to understand and describe the behaviors of the interested objects with the data grabbed from the video and the visualization abstracted from the data. A smart video surveillance system as a preferable replacement for the outdated and ineffective video surveillance system.

The object detection technique and the object tracking technique are in the limelight in the past decades because of growing needs in preventing the issues of crimes, terrorist attacks or even public disasters. What is more is the application of big data these days all come from the impressive performance of machine learning. Machine learning constructs the base of several fundamental fields in computer vision. It makes the pedestrian motion awareness feasible and simplified. However, it is still extremely difficult to process the video data even via machine learning mostly because previous research focused on tracking, instead of giving an exhaustive analysis for every pedestrian. The researchers in computer vision fields were restricted to bound every target accurately and make less mistakes on the detection, but it’s still not practical for people to get the information from millions of frames and bounding rectangles. Nowadays, the algorithm in computer vision can grab the extreme volume of data from the video efficiently. While the tracking algorithm is accurate and efficient enough, it's time to visualize the tracking data to make the result more impressive and visible. Information visualization is widely used in numerous data analysis applications to help people have a better understanding and perception of the data. The inherent relationships and the flowing trend among the data elements are easier to be cleared with the help of data visualization.

In summary, this dissertation intends to provide an efficient and effective approach to visualize the motion pattern of pedestrians and crowds directly from the untreated surveillance video. A good understanding of the crowd motion pattern can help the user of surveillance system save more time and energy to abstract the useful information from the video. In this dissertation, the researcher proposed a surveillance system which processed the raw video with an existing object detection and tracking algorithm, the researcher also proposed a visualization system to display and analysis the motion pattern of pedestrians from the processed video data.

1.2 Statement of Problem

In recent years, new techniques in computer vision field reduce the barrier of entry, allowing developers to experiment more with the intelligent surveillance video system. Different from previous research, this dissertation does not address any algorithm design issues in regard to object detection and object tracking. This study will put the efforts on the technological side and executing methodologies in data visualization to find the model of detecting anomalies. It would like to provide an understanding of how to detect the behaviors of the pedestrians in the video and find out anomalies or abnormal cases by using techniques of data visualization.

1.3 Research Questions

Is it possible to visualize a crowds' motion effectively from a prolonged video stream recorded by a surveillance system so that an administrator can be aware of the status, patterns, anomalies, and events of crowd's motion? If possible, how effective is the visualization method?

1.4 Significance

In modern cities, massive population cause problems, like congestion, accident, violence and crime everywhere. The progressively increasing population in the cities push the development of video analysis. For the sake of public safety, more than four million and five hundred thousand cameras have been installed for video surveillance in the past ten years. The quality of video has also improved rapidly. It is common now for people to watch movies with HD and even full HD quality, which could not be imagined before. Although a large number of monitor cameras are set up in the almost every corner, it's still very difficult to capture all the details from the video dataset by human eyes.

Data visualization plays a significant role to represent the complicated dataset and transform various kinds of information into proper visual display. Then the analysis and understanding of the information transmitting can be completed more directly and efficiently. Data visualization is a delicate redesign and reconstruction of the redundant numbers, letters and even millions of the pixels. It incorporates human capabilities and intuitive visualization, making people easily to capture the trend or the outline of the whole dataset. As a combination of machine intelligence and human intelligence, data visualization can be classified into three categories: scientific visualization, information visualization, and visual analytics. Scientific visualization demonstrates spatial structures and development for the physical prototypes and chemical substances. Information visualization compresses the high-dimensional data and abstract the meaning behind the disorder and uncorrelated data (Thomas, Cook, 2005). The interactive animation in the interface can provide a reasonable understanding of the Visual (Ware, 2012). Visual analytics can be used as an establishment of multi-faced human capability and flexible data analysis (Kandel, 2011).

1.5 Scope

In this study, the researcher will utilize a modern object tracking algorithm to identify pedestrians' moving data and focus on design and develop an effective visualization system to let users see and be aware of the movement pattern and abnormal behaviors.

A good data visualization must be assisted to be a completed analytic system. The process of data analysis always has several steps. The steps of data analysis are as follows: data collection, data preprocessing, data query and data analysis. Every step needs to be considered independently and solved with a specialized method. For example, data collection in Multi-Object Tracking (MOT) is always combined with the technology in the computer vision, like neural network and deep learning. Machine learning tracks and associates the object according to feedback from the training of negative example and positive example. With the existing MOT system, we can reconstruct the system for the pedestrian captured in the city monitors. As for data preprocessing, a clean and structured dataset is good to eliminate the error points and transform the useful data for downstream analysis tasks.

Furthermore, pedestrian monitoring, and pedestrian detection are the research bases of the automatic video management and information analysis. According to the procedure of existing Multi-Object Tracking and spatiotemporal visualization, the structure of pedestrian tracking data visualization could be divided into the following several parts: Recording video of pedestrian and background, moving pattern discovery and clustering, situation-aware exploration and prediction, route planning and recommendation.

1.6 Definitions

Data visualization is the science of visual representation of "data," defined as "information which has been abstracted in some schematic form, including attributes or variables for the units of information" (Friendly & Denis, 2001). According to Vitaly Friedman (2008),a great advantage of data visualization is its ability to visualize data, communicating information clearly and effectively.

Spatiotemporal visualization: It is one of the data visualization tool which can provide a universal and comprehensive view of the loosely related data sets (Zhong, 2012).

Surveillance video system: surveillance video system's are automated monitoring sensors to collect and accumulate real-time information from the specified field to increase the awareness of interested people and crowds flow (Collins et al., 2000).

Object detection: It is used as the technique searching for a specific class of objects like people, dogs ,cars ,birds or faces (Papageorgiou, 2000).

Object tracking: It identified the detected object of interest in the image plane and estimated the trajectory of the specified object moving around in the video sequences persistently (Yilmaz, Javed, & Shah, 2006).

1.7 Assumption

The assumptions included in this research include

- The data extracted from the video represents a typical long-term moving pattern of pedestrians.
- Captured video is in good quality with high definition with clear scenes.
- The object detected in this project are mainly human pedestrians.
- Object recognition algorithm used in this research can extract most of the salient information of crowds' motion.

1.8 Limitations

The limitations included in this research include

- Videos used in this research are records from the repository or taken in Purdue university. It may not contain enough types of activities and events.
- The crowd's motion information extracted from the video are limited by the algorithm we selected.

1.9 Delimitations

The delimitations included in this research include

- The research focuses only on the reconstruction of the information and the layout of visualization, not on doing research on improving methods to generate more accurate algorithm for crowd detection.
- The research focuses only on the open scene which guarantees the pedestrians shot in the video is bright and not opaque.

1.10 Summary

This chapter provides a basic introduction for the whole research, including its background, research question, scope, significance, and the statement of purpose. Furthermore, definitions of key words, assumptions, limitations, and delimitations are discussed to provide a base and scope for this research. In next chapter, the relevant work of surveillance video analysis will be introduced and summarized.

CHAPTER 2. LITERATURE REVIEW

Video analysis is a crucial task within the field of intelligent surveillance video system. This chapter gives an overview of previous research in the fields of pedestrian detection and visualization, providing both a base understanding of the methods in the subject area as well as motivation going forward to a new methodology. There are several key steps in the video analysis: shooting the object in the same image plane, detecting the interesting objects in every single frame of the video, tracking the interesting objects in a sequence of frames, analyzing tracked objects to recognize the behaviors of the crowds, visualizing the motion pattern of the people in the video. This chapter can be briefly divide into the following 2 sections: visual surveillance and information visualization.

2.1 Visual Surveillance

Surveillance video system has been concerned and discussed for several years. Especially in dynamic scenes. Visual surveillance is one of the most popular topics in computer vision research, strongly motivated by many significant and potential applications, such as crowds flow statistics analysis, long distant human identification, interactive multiple surveillance cameras (Hu, Tan, Wang, & Maybank, 2004). In general, surveillance system processes the dynamic video frame with these several steps: modeling the scenery, detecting and classifying the objects, identifying and tracking the detected objects. What makes the intelligent surveillance video system different from the traditional video monitor is the algorithm used and researched to detect and track the objects of interest.

2.2 Object Detection

Fast, robust object detection system is the base and the fundamentals of video processing applications. A good object detection system can efficiently identify the different classes of the object, such as faces, dogs, cars, pedestrians and so on. In contrast, the difficulty of the detection is how to balance the line between the accuracy and the efficiency. Since the video system can only get the information from the 2D image plane, it will lose much stereography information and increase the noise for the real environment in 3D world. At present, quite a few methods have been proven effective in the field of object detection; such as background subtraction, temporal differencing, optical flow, HOG&SVM and neural network (Hu et al., 2004).

The motion-segmentation based methods always have the feature of efficiency and inaccuracy. When it comes to object detection background subtraction is a popular motion-segmentation based method, especially for those with a comparatively stationary background. It checks the difference between foreground-background image and define the regions with bigger subtraction as the moving regions. Some previous work average the image in a period of time to stabilize the background (Stauffer & Grimson, 1991). Pixel RGB and chromaticity values of videos can also be mixed with local image gradients into background model (McKenna, Jabri, Duric, Rosenfeld, & Wechsler, 2000). In order to update the background layer, there are two different ways, pixel-based update and an object-based update (Haritaoglu, Harwood, & Davis, 2000). If people are the only objects moving in the surveillance, background subtraction can be the simplest way to detect the human. However, it is tremendously sensitive to the shadow change created by the light and has no tolerance to the overlap of the object. Thus, the result of this method highly depends on the quality of the video background and the quantity of the people in the video.

Background subtraction can also be simplified as temporal differencing. Temporal differencing doesn't have any background layer and foreground layer, it only focuses on the pixelated differences between two or three consecutive frames in the video sequences in order to detect the moving area. Compared with the background subtraction, temporal differencing is more adaptive to the light change of the environment, but generally creates more unnecessary holes on the detected motion regions. A previous work detected moving objects in the surveillance video which uses a temporal differencing method and has a pretty good identification (Lipton, Fujiyoshi, & Patil, 1998). Although this method still cannot solve the difficulties with the occlusion and termination of object motion, it avoids background layer drifting into the next calculation, and its robust identification and classification of object do not rely on any analytical temporal filter such as a Kalman filter. Santhiva (2014) has shown that the existing object detection with foreground structure is good enough for monitoring and collecting data on the traditional crowd video. They provide enough experiments to prove that the foreground has the same importance as the background in the detection subtraction-based algorithm. The feasibility and robustness shown in the experiment results make it even applicable in the field of explored semantic scene knowledge and social communications among persons.

Different from the above two methods, optical-flow-based object detection grabs the dynamic variation of the pixel motion in an image sequence with the characteristics of flow vectors of interested object. For instance, it can extract the contour and motion of articulated object with the active rays in displacement vector field (D. Meyer, Denzler, & Niemann, 1998). A good optical-flow algorithm can even track the gait features and body parts of the object (Dorthe Meyer & Niemann, 1998). This method can also process the video shot by the independently moving the camera. However, most optical flow projects need the support of a great mount of complex computations and are extremely sensitive to noise in the image sequences. This makes it hard to be put into the use of real time application with low scale of the filtering (Barron, Fleet, & Beauchemin, 1994).

As for the video including only humans, the above three motion-segmentation based can take all of the detected moving objects as the people. Situations in real life are always more complicated, the pixels of human and other objects like dogs, cars, and shadow blend and combine together subtly. The intelligent surveillance system has to classify every object detected by the video and gives them a label. The histogram of gradient (HOG) feature is collected and computed for the detected moving regions and then it is fed as the input to a SVM-HOG classifier. There are two ways to collect the hog information. One is to reuse the features of the interested regions in continuous frames, another is to compute sub-cell interpolation as the base of the HOG feature (Pang, Yuan, Li, & Pan, 2011). These SVM-HOG classifiers combined off-line training and machine learning and give a probability of every object to check whether the moving object is a pedestrian (Xu & Xu, 2013). SVM-HOG method has a very high recognition rate even in the static image situation without motion information. Most of the researches related to it sped up the algorithm without sacrificing the detecting accuracy. Unlike in the past, the SVM-HOG method doesn’t rely much on the monocular visual odometry (MVO), an old-style geometrical method. The calculation error of the detection came from the motion and background will never propagate further and bigger in the process. SVM-HOG minimizes the impact of objects on visual odometer reliability and expands the detectable types of the moving objects (Zhang,Y, 2017). Human beings are always the main moving target in the process of motion detection, and the accuracy of MVO is improved by the fuzzifying the unrelated feature of the pedestrians. The combination of featuredbased MVO and HOG + SVM methods contribute a lot to the camera poses estimation in the trajectory visualization system.

Machine learning techniques have been applied in our daily lives for a long time, and the traditional information processing mechanisms can barely satisfy today's requirement. Given an image, the machine needs to return all the objects of one or more type of targets and bound the targets with a tight contour or rectangle. As a self-learning feature classifier, the neural network needs a large amount of data to ensure the high accuracy of the architecture and avoid the destabilization of similar background, vague characteristics or target segmentation. At the beginning, time-delay neural network (TDNN) is a successful approach to analyze the feature related to the time and predict the next move of humans in the video sequences, like gesture recognition (Yang & Ahuja, 1998) and lip reading (Meier, Stiefelhagen, Yang, & Waibel, 2000). The development of the TDNN pushes the research on convolutional neutral network in field of computer vision. It can be the best deep learning architecture in processing the images and videos. In CNNs, two-dimensional data with grid topology can be directly input to the system, saving the procedure of pre-processing. What's more, the utilization of the convolution reduces the quantity of weights which simplifies the architecture of the whole network. When you lessen the number of parameters, the standard backpropagation algorithm directly better guarantees the success of the CNN (W. Liu et al., 2017). Gidaris (2015) enhanced the Multi Region CNN model with segmentation features and the corresponding feature regression under the iterative localization scheme.

2.3 Object Tracking

While surveillance video systems have already recognized all the interested regions in the video, the detected objects need to be checked whether it matches some objects in the other frames. Generally, two objects with too much intersection on regions or just features will be tracked as the same object in the video processing. In order to capture every subtle detail, lots of existing image analysis algorithms will be used for tracking, such as the Kalman filter; the Condensation algorithm; the dynamic Bayesian network; the geodesic method (Hu et al., 2004), etc. The tracking methods can also be basically categorized as two types: region-based tracking and feature-based tracking, but what object tracking does is more like matching the interested object with the same id in consecutive frames instead of classifying the objects with a different id in one static image plane. Most of the difficulties in object tracking comes from the expected sub-block image motion and the occlusion between objects and sceneries.

As for the region-based tracking method, the efficiency of the algorithm generally related to how the appearance of object is simplified in the algorithm. The appearance of the detected object can be represent by the points; primitive geometric shapes; object silhouette contour; articulated shape models; and skeletal models (Yilmaz et al., 2006). Different representations have diverse performances in real-life application. For instance, when the pedestrian looks small in the original video, it saves a lot time to track a representation point of an object instead of the whole area including the detected object. When the appearance of the object is almost the same with the rectangle or the circle, it may be the best way to use the primitive geometric shape to represent the object. Some previous work uses the circular symbol for the circle regions (Comaniciu, Ramesh, & Meer, 2003) and uses the eigenvectors to replace the rectangular shape object (Haritaoglu et al., 2000). When the objects detected in the video have a more complicated shape or appearance, contour or silhouette can be used as the representation of the object like human (Haritaoglu et al., 2000). Bera (2015) proposed an innovative real-time algorithm to track every pedestrian in tightly-knit-crowd videos with high density. The discrete point and continuous flow models are mixed into a hybrid model, which can use particle filters to compute the pedestrian trajectory. The discrete part of the model comes from the microscopic agent formulation, while the continuous model takes the crowd orientation and flow as an input. What's more, the performance and the accuracy of hybrid model in this research have been proved well in ten-level pedestrians video.

As for the feature-based tracking, the high-level features from a detected region related to the object of interest needs to be extracted and matched with these features in consecutive frames. These methods can be generally applied according to the category of the object: either global feature-based tracking, local-feature based tracking or dependence-graph-based tracking (Yilmaz et al., 2006). The global feature is usually generated from the average quality of the whole region of an object. For example, the video processed by the object detection algorithm will bound the specific region with a rectangle box, the properties of the whole box or even the centroids can represent the global feature (color, velocity, location) of the object. Even when the object overlapped with other objects or background, it can still be tracked successfully if the velocity, color, or the location of centroids changes smoothly (Schiele, 2006). The algorithm using the local feature may only take part of the lines, the curves or the points cluster. Different from these two algorithms, the dependence-graph-based algorithm utilizes the high-level physical and mathematical relations between the features (Coifman, Beymer, McLauchlan, & Malik, 1998). Most of the features like color, edge, optical flow, and even the texture is filtered manually by the researcher according to the application domain. However, manually choosing the features wastes much manpower and material resources. Researchers have much more interest on automatic feature capturing with the help of support vector machine and neural network. Milan (2017) presented an innovative method using recurrent neural networks (RNNs) to track multi-targets online. Their project focuses mostly on objects undetected before or some object detected with great time-varying change. They continuously stored the state estimation of each existing person for every frame of the video and constructed a discrete association for the collected data. This is the first time to apply the recurrent neural networks (RNNs) into end-to-end trajectory learning.

Tracking the object individually will not consider the obstruct created in the trajectory. An obstruct occurred in the trajectory can lead to the partially or even completely occluding of the detected object. Multiple object tracking (MOT) is abstract and has an increasing number of researchers on it. Most of the widely used traditional video sequences and the new challenging video sets have also be collected in the MOT Benchmark. The researchers also come up with principles to judge the solvent of these problems, such as Multiple Object Tracking Accuracy (MOTA) , Multiple Object Tracking Precision (MOTP) , the Most Track objects (MT, the object is define as MT if the trajectory of the object has been tracked for 80%), the Mostly Lost objects (ML, the object is define as MT if the trajectory of the object has not been tracked for 80% 20%), False Positives (FP), False Negatives (FN), ID Switches (IDS), Fragmented trajectory (Frag), and the processed frequency of frameworks (Hz).

Luo (2013) tried to use generic algorithm to solve the Multiple Object Tracking problems. Their method succeeds in crowds, but their target was not only on pedestrians like the previous work but also on objects which are generic. They stuck on the traditional strategies in computer vision, formulating their Multiple Task Learning (MTL) framework with combination of the detection and tracking results. Multiple trackers of MTL were trained based upon the detection result of the binary in order to track the objects of interest in video sequences. All the data with common features were considered globally in the tracker. It transformed the tracking problem into a multiple task learning problem with a generic algorithm. The proposed Mean Regularized Joint Feature Learning algorithm performed excellent in the traditional video sequence test. Another great challenge in online Multiple object tracking is matching the noisy object detection out of the video sequences. Xiang (2015) fused the traditional multiple object tracking system with the Markov decision processes of four states (Active, Tracked, Lost and Inactive). This framework takes the single object detection and tracking result as the input of MDP policy learning offline. They also divided the benchmark video sequences into training sets and tested the sets to adjust and validate the MDP learning system. Then it output the online decision for multiple object tracking from Markov Decision Processes (MDPs). Their novel fusional multi-object tracking framework outperformed most of the traditional methods only relied online calculation and offline calculation. Tang (2017) also provided a good structure to track crowds. In order to re-identify and re-associate the appearance along the long distance and time, they applied the mechanism of minimum cost lifted multi cut problem in the tracking system. Their model constructed long-range capable connections between nodes in the graph without destructing the layout of the existing dataset. It extracts the complete pedestrian representation and body pose feature from the deep neural network, and successfully matches them with the long-range links and clusters person hypotheses.

2.4 Information Visualization

This research is about figuring out how to utilize existing video sequences to make a visualization of pedestrians. With the video processed and all the information of the detected object extracted, it is time to analyze the data and understand behaviors behind the data. The use of visualization, the basic steps of visualization, and the features of different types of visualization are discussed in the following sections as guidelines for visualization design. Information visualization requires a delicate design and reasonable user interaction to compress a large amount of information and declare the nature of the data. Generally, a segment of information cannot reflect the whole content of the dataset and the pattern of the information may be hidden in the interlaced data structure. It is a bit difficult to reveal the hidden relation mingled among the dataset, which needs to be processed with some corresponding techniques and impressions. In this research, the behaviors of pedestrians should be revealed, which may include patterns, relationships, clusters, and paths. Visualization is an approach to help with this. Card, Mackinlay, & Shneiderman (1999) claimed that the main benefit of computer-based interactive visualization is to use perception and recognition to stimulate the sensory ability of human instead of remembering all information from dataset. In fact, the data in this research is almost all about the location and the time. Thus, for these sections, information visualization will be divided into three parts: visualization design, traditional traffic visualization and spatiotemporal visualization.

2.4.1 Visualization Design

The uses of visualization design generally have four progressive processes (provide overview, adjust, detect pattern, match mental model) to acquire the perception of the information visualization system (Yi, Kang, Stasko, & Jacko, 2008). The first process is to let the user get a good understanding of the whole picture of the system and have a basic summary of the data trend. A good overview can help users grasp the sense of where to focus in the dataset and make clear the relation of data elements. Adjust process means that the users need to adjust the range of the dataset selection and operate the level of the data abstraction. Then the grouped dataset can show the users more meaningful information and high-level facts, while the unnecessary information will be hidden and blurred in the visualization. Detect pattern refers to the process that the users find specific distributions, trends, frequencies, outliers, or structures in the dataset (Yi et al., 2008). People have a high probability to find the unexpected information in this procedure. In the match mental model, the gap between the data and the user's understanding has been balanced due to the reduction of cognitive load and the amplification of prehension. The data has been associated and mapped well in the information visualization.

The construction of information visualization mostly follows the pipeline of five modules: data extraction and cleaning, data mapping, elements mapping, visualization rendering, and UI controls (S. Liu, Cui, Wu, & Liu, 2014). The original dataset collected in the real life is always unsorted and unstructured. The input data also needs to be reduced and compressed, if the scale of the dataset is too large to deal with then the extracted data needs to be cleaned and categorized according to the relations of data elements and the types of the data. For unstructured data, the system may use some data mining techniques such as clustering or categorization to clear the structure of the dataset (S. Liu et al., 2014). Then the modular will use several filters to smooth the noise of the dataset and interpolate the missing point. The appropriate range of the data will be automatically or semi-automatically chosen for the formulation of the data visualization. The mapping step of the information visualization is to match the neat data with the geometric primitives (points, lines, circles, rectangles) and the attributes (color, position, size, opacity, width, length). Visualization rendering refers to the process that transforms the data into the graphic showing. The last step is to add the UI controls to the visualization system so that the users can explore and understand the data from every aspect deeply.

The model is extraordinarily important in the empirical design of visualization study. There are mainly three widely used categories: visual representation models, data-driven models, and generic models. Visual representation models are good at dealing with the situation that will have a numerous amount of the research output in the same image plane. As one of the Visual representation models, context-preserving visual can facilitate the comparison and interpretation of related elements variously (Steinberger, Waldner, Streit, Lex, & Schmalstieg, 2011). Users can also learn the important information from the visual difficulties model which can point out the trade-off balance line between efficiency and beneficial factors (Hullman, Adar, & Shah, 2011). The model is crucial for the comprehension and perception for the data visualization. In the data-driven models, the design of the visualization is driven by the need of the real-life application and the type of dataset, such as high-dimensional data, heterogeneous data, geographic data, narrative data, tables of counts, proportions, and probabilities (S. Liu et al., 2014). Researchers tend to use the generic theories and models to develop the information visualization.

Good categories and summaries of the evaluation can better conduct the design and the development of the visualization. User studies are the most widely used method to evaluate the data visualization system. Depending on the requirement and the design of the visualization, the user studies can be handled be the approach ranging from the informal surveys, to elaborate laboratory studies. Elaborate laboratory studies compare the design elements and the visual function, while the other factors in the experiment are controlled.

User interactions (UI) have the same importance as the appearance in the design of information visualization. Some old research classified the interaction methods according to the functions into seven categories: select, explore, reconfigure, encode, abstract/elaborate, filter, and connect (Yi, Kang, Stasko, & Jacko, 2007). The interaction methods can also be classified according to the mechanics into two categories: WIMP (windows, icons, mouse, pointer) interactions and post-WIMP interactions (pen, touch, gesture) (S. Liu et al., 2014). The major goal of the interactions is to let the user focus more on the specific regions and feel free to select any element in the graph. A good application of interaction like shine-through and folding can also compress the layout of the visualization and provide the overlapping views of the detail.

It's hard to draft the outline of the interactive visualization applications on the paper manually. To overcome this issue and save on designing time, researchers have come up with several design systems for the information visualization such as: Improvise, the InfoVis Toolkit, and Prefuse (S. Liu et al., 2014). Improvise can highly customize the multiple and interlaced relationship of the data elements and provide a sophisticated coordination mechanism to fix the location of any shape of the pixel clusters (Weaver, 2004). The InfoVis Toolkit is a Java-based library integrating the commonly used data-clearing algorithm and the delicate visualization model presentation, which simplifies the manufacturing flow (Fekete, 2004). Prefuse is a powerful tool library consisting of all the techniques that may be used in the pipeline of the visualization designs like data structure, layout design, and various kinds of interaction animations (Heer, Card, & Landay, 2005). In recent years, document-driven documents like javascript library has attracted the attention of data visualization researchers. As a user-friendly and webpage-based toolkit library, it supports the direct operation of the webpage elements (Scalable vector graphics, text) and simple binding between the data and the web elements. Some information-theoretic frameworks based on the webpage can provide the existing and integrated strategy for the visualization and interaction animation.

The performance of the information visualization mainly depends on the data type and the utilization of the graph. The type of data structure and the characteristic of the data elements decide the key view and focus on where the visualization displays to the users. For the data with the graph structure, the researchers usually constructed the visualization according to the topological relationships among the elements (Selassie, Heller, Heer, & London, 2011). For the textual data, the visualization tends to describe more about the segmental core content, such as the theme or the major topics, from the original documents (Cui et al., 2011). For the geographic data, the key patterns needed to be revealed are the spatial distribution of the information, the anomalies in the data clusters or even the trend in the trajectory in this research (Afzal, MacIejewski, Jang, Elmqvist, & Ebert, 2012). For the data related to time, researchers can use the timeline-based approach to visualize the data along the time axis statically or use the interaction animation to illustrate data variation over time dynamically.

2.4.2 Traffic Visualization

With the rapid development of manufacturing industries, Traffic data scientists usually tend to analyze the flow trend or the passage data of walking pedestrians, motor vehicle, non-motor vehicle, moving anomaly on the transportation sceneries such as: downtown area, uptown area, grounds, underground area, even the sea and the air. Data visualization has been proven as an effective method to reveal the deeper patterns of complicated traffic systems and present the data distribution of transportation. The advantage the travel visualization is to incorporate the intuitive visual perception of the graphic and the machine intelligence. Traffic visualization facilitates the understanding of the high-dimensional and spatial-temporal data. The traffic data visualization displays the data in the following several procedures: Visual monitoring of traffic situations, Pattern discovery and clustering, Situation-aware exploration and prediction, Route planning and recommendation (Chen, Guo, & Wang, 2015).

The traffic data is classified into three categories according to the ways collecting the original data: location-based (monitor sensor is fixed in one location and captures the information around the sensor), activity-based (monitor sensor is collected in the specific activity to record all the information around), device-based (monitor only records the information around the object that carries the device) (Andrienko, Andrienko, Stange, Liebig, & Hecker, 2012). Trajectory is the most widely collected traffic data form, which may consist of temporal information (time point of each motion), spatial information (location data of each object), and even the other attributes and properties like object type, direction, activity type, etc. According to the emphasis of the dataset, the form of the traffic visualization is also dissimilar (Ferreira, Poco, Vo, Freire, & Silva, 2013). As for time visualization, researchers can display the presentation of traffic along the linear time to show the peak and valley of the variation if the data changes with time from a start point to an end time (Pu, Liu, Ding, Qu, & Ni, 2013). The activities with periodic process can also be presented in a circle. The situations with many branch structures can use storylines to depict each branch and joint individually (Ogawa & Ma, 2010). As for the spatial visualization, the presentation of the visualization depends on the aggregation level of coordinate information (Guo & Zhu, 2014). Individual discrete data can be displayed by the point-based visualization which uses a point to show the position of some point at some time. The trajectory or the road-like traffic data can be presented with the line-based visualization with (x,y) coordinate system. The traffic data aggregated in an individual region can be solve by the region-based visualization which only summary the local trend of the data. Spatiotemporal visualization can still be used in the traffic data with the 3D coordinate system and time-based z-axis (Kraak, 2003). If the traffic data contain more attributes other than spatial and temporal information, the visualization needs to be chosen according to the additional attributes. The histogram is a good choice for time-oriented Numerical properties visualization. Color mapping is the simplest visualization for categorical properties. Text-based visualization is the appropriate method to deal with words, lexical information, or logs.

Chang-Tien Lu (2006) proposed an Advanced Interactive Traffic Visualization System (AITVS) programmed with the webpage and javascript to evaluate and understand traffic conditions. Traditional traffic visualization system only adopts the limited basic methods to explore the perception of the traffic data. This interaction visualization system is good at digging the information of critical instruments for comprehensive study and traffic condition research. It not only highlights the basic characteristic of the traffic such as velocity, time, location, orientation, but also facilitates the deep learning of the transportation data. The recent surge of network traffic can also be solved by the traffic visualization. In order to present these network traffic, Lei Shi (2009) introduced a new method based on the node-link graphs, Structural Equivalence Grouping (SEG), to condense the network and highlight the key connectivity information with the linear time complexity. They even developed SEG into a Network Security and Anomaly Visualization (NSAV) to monitor the traffic accident and anomaly.

2.4.3 Spatiotemporal Visualization

The spatiotemporal datasets are mainly visualized by graph which can highlight the fluctuation change of the time and space such as 2D geographic maps, time lines, or radial visualization. It can provide a global view for the those loosely distributed information and extract the implicit relation of the data elements deeply. Thus, it’s widely used in the complicated spatial and temporal progresses and activities to make the key decision. The applied range of Spatiotemporal Visualization is abundant and not only limited in the traffic analysis, but also includes prevention and control of infectious diseases, flood administration, road-routine change, landscape modeling, and urban construction simulation (Zhong, 2012). The development of the spatiotemporal visualization reduces the barriers of accessing the core of the data to the unprofessional users and makes it as simple as how experts do. It still takes more challenges waiting to be solved. Current technology can promise a large enough database to rebuild the details and reflect the hidden variation of the data. It also needs to balance the high dimensions of data source and the readability of view, while different disciplines and methods are combined together.

The workflow of the spatiotemporal visualization can be defined in such following pipeline processes: creating moving object database, defining the detection patterns, semantic enrichment, building conceptual data models, visualizing analyzed spatiotemporal data. The time dependent geometries information is always complicated in data categories and data depth, which need to be stored and queried into database for data-read. Static data and dynamic data may also be divided in to two columns to store (Andrienko et al., 2012). After that, raw data will be processed to extract the detection patterns, trends or relationships of object (Andrienko, Andrienko, & Gatalsky, 2003). Semantic enrichment refers to the labeling of the detected pattern of objects and makes the data easier to understand by the normal users (Spaccapietra et al., 2008). A good conceptual model can emphasize the cause and consequence of some activities or data variation in the database, instead of only displaying the numerical data (Spaccapietra et al., 2008). The last step is to study the entity and property of the raw data and extract abstract data types, which can determine how to link and manage the elements in the visualization (Erwig, Schneider, & Vazirgiannis, 1999).

Various information visualization techniques and different visual symbols (shape, size, direction, color, opacity, hue, and texture) determine the interactive ability and understandability of the spatiotemporal data (Zhong, 2012). From the static techniques to dynamic or even interactive methods, the commonly used techniques will be introduced as following. Timestamps and time labels are used for the activities marked with time information. Different specific graphic symbols and elements scatters in the same image along the timeline indicate the variation of spatiotemporal data. As for the database with more state changes, the visualization can use arrow, lines, and other base charts to represent the progress and demographic trends of activity (Kraak, 2003). As for some dynamic representations, it generally uses the timeline animations, such as zoom, drag, suspension, click, to illustrate the variation and tendencies of movements from start time to the end time. As for the complex spatiotemporal patterns, researchers use space-time cube methods combined with interactive techniques to present the high-dimension dataset (Kraak, 2003).

2.5 Summary

This chapter provides a review of the previous studies related to this research, including the background of current visual surveillance, the commonly used algorithm in object detection and object tracking, different methods of information visualization. Based on all these related works, researchers can have a deep understanding on crowds' motion visualization and some primary thoughts about the framework of this research. In next chapter, the methodology of this research will be discussed in detail.

CHAPTER 3. METHODOLOGY

3.1 Framework

In this paper, this study would like to provide an understanding of how to detect the behavior of the pedestrians in the video and find out anomalies or abnormal cases by using techniques of data visualization. This is a quantity research which will utilize the raw data of existing pedestrian videos to visualize the walking pattern of each crowd flow. In this process, this study needs to take a raw video of the pedestrian in the crowded intersection. It will focus on how to establish video visualization and design the interactive animation to detect people's pattern of their motion and trajectory. It is used to explore whether this creation can help people figure out what is happening in the crowd and what is abnormal when everyone moves naturally in the flow. The detailed design of the research will be listed in the following several parts.

3.2 Data Source

As for the data sources, this study may use the dataset from two sources. One is the video sequences from the dataset of MOT Challenge benchmark because the sequences collected by the MOT Challenge are commonly and widely used by the researchers from all around the world.

These video sequences will have less uncontrollable factors than the others, like noise, background etc. They have proven to be extremely helpful to advance the state-of-the-art in the respective research fields. Moreover, these video sequences have more abundant and complicated people relationships than in the normal cases. For instance, the people coming in and out of the frame, the people circling in somewhere and the people moving slightly are all included in one video sequence. However, in order to simulate some more real conditions, this study will also shoot some videos in the campus of Purdue to record how pedestrians move at the cross.

In the past few decades, the researchers from computer vision classified different monitor videos according to the variant detection and tracking tasks including object detection, pedestrian detection, three-dimensional environment rebuilding, optical flow, object tracking for single target, object tracking for multiple target and state estimation. The benchmark used in this study is an enormous collection of datasets focus on multi-object detection. In this benchmark, some are already in use and some are new challenging sequences. In addition, the evaluation tool provided has several useful measures, from recall to precision to running time, which can help measuring the method in this study. It's a good way to verify the performance of the algorithm used in this study, and then the best structure of detecting part will be chosen to use in this study.

Apart from that, this research still need to shoot some video for the real conditions, like the cross in Purdue. The pedestrians and the vehicles will move randomly without any deliberate arrangement. The video will be shot in a building with more than 5 floors in order to make sure that every person can be seen moving horizontally on the ground, which will make the detection more precise. Moreover, it's better to shoot in a sunny and breezeless day because there will be less shadow noise created by the wind and the pedestrians will also have a significant difference from the background. It will be difficult to deal with a high-resolution video, so a cell-phone level camera is good enough for this experiment. The camera will be fixed in a stable spot with a customized flat view parallel with the direction of the pedestrian movement. This study needs at least two such videos to test our visualization whether it can be generally used in most cases.

3.3 Data Extraction

Then we have three methods to abstract the pedestrians in our raw video, listed as follows:

1. Background subtraction (BS) is a widely used strategy to detect the moving pixels from the differences between the current frame and the reference frame. The subtraction from the current frame and the reference frame will provide two result layers: foreground layer and background layer. The foreground layer is a binary gray-scale image layer demonstrating the motion region as white pixels, while the background layer only demonstrates the stationary part in the image plane with the black pixels. The process of background modeling follows two key parts: one is to construct Background Initialization; the other is to update the background in each single iteration (It may contain several single frames according to the accuracy required by the experiment). However, the fatal shortage of this method is that it only detects the motion and the moving objects, instead of the real person. The cars, the birds, or even the shadows, all of the moving objects will be detected and labeled. It's hard to detect the pedestrians, but it still can be set up as a good pre-process method for the raw video sequence.

2. The histogram of oriented gradients (HOG) is the widely used characteristic information to detect the object with some specific features like human. One of the most popular and successful methods to detect a person from the pictures or the video is to use the HOG with the SVM method.

As one of “feature descriptor” used in image processing, HOG can generalize the object in short lines and points so that the person appearing in different sceneries will be associated and matched with the similar HOG feature, which makes the recognition and the classification of people easier. In order to work with histogram of oriented gradients, the computer still needs to train a Support Vector Machine, or “SVM”, which will provide a clear boundary to separate human-like objects from other objects. In machine learning, support vector machines can deal with the work of label classification and data regression analysis because of the advantage of supervised learning models and proven training algorithm. Support Vector Machine takes training descriptor and training ground truth as input to build the models and adjust parameters of the model with the test dataset. The model that has the best performance in the test dataset will be kept and provide a non-probabilistic linear binary determination for the object. Then the classifier can be used to classify the different objects (in this case, pedestrian) in the video sequences. The shortage of this method is that it cannot tell which person detected is moving. Some of the detections may even be a picture of man on the billboard. When the pedestrian is occluded or out of the screen, the algorithm will also fail to detect the pedestrian.

3. Neural network can also be used to train and abstract the pedestrian from raw videos, but more data and time are needed. Among the algorithms of Neural network, convolutional neural network shows outstanding performance in the field of computer vision. As a class of deep feed-forward artificial neural networks. it has successfully been applied to analyze visual imagery and video sequences. Convolutional neural network uses a variation of multilayer perceptron for minimal preprocessing. CNNs use relatively little pre-processing compared to other image classification algorithms. More training dataset and more complex network constructions bring the detection accuracy to a higher level. It can even detect the occluded person and the missing person if the construction of the network is trained well enough.

In order to extract the data more accurately, this study may use all three of these approaches. Firstly, the raw data will be converted into a gray-scale map picture to get through the background subtraction. The detected object in the first step will be bounded in the rectangle and classified by the Hog+SVM method. This is a preliminary detection of pedestrians which may only take a few human body features as reference, but the binary classification result can be easier to re-detect and shrink the range of the objects needed to be processed. Most of unrelated objects without human features such as trees, cars, animals can be eliminated from the data set and the subsequent work complexity can be simplified. Then the classified object with labels will be trained in the convolutional neural network, so that every occluded and missing object can be re-detected correctly.

3.4 Design and Visualization

When the computer extracts the data of every moving pixel from the pedestrian, the researcher needs to pre-process the raw data by computing the velocity vectors, directions and colors. The raw data extracted from the video sequences also needs to be cleaned, classified and even ranked by the dataset. Then the researcher will try to analyze data in multiple dimensions. The researcher can establish the specific dashboard of each parameter with the help of D3.js and PHP to visualize the trend and the variation of the crowd. The initial idea of this study is to imitate the water wave. Every moving pedestrian will send ripples in the screen just like a stone thrown into the water. The wave from each point or person will cross and merge into the flow, while the abnormal people show themselves in the flows. General patterns of dataset could then be explored with basic graphics. Based on the general understanding of the data and the knowledge in related areas, such as visualization and interaction design, user behavior of information re-visitation, and information management, the researcher then applied the related theories and concepts to reorganize the trajectory and visualize it. The design would also be adjusted and improved dynamically and iteratively resulting from the features and patterns found during the process of development and discussions with other professionals.

3.5 Demonstration and Evaluation

When the visualization detail and the user interaction has been designed and finished, this research will focus on the demonstration and evaluation. A good visualization needs to be user-friendly and easy to understand even without a systematic learning. Thus, the understanding and the evaluation of the whole system from the user is crucially important. In this section, the researcher will introduce every function of this system and evaluate the usability of these functions.

This research plans to compare this visualization system with the tradition surveillance system and some existed intelligent surveillance system. Data source for this research is divided into two parts. The first part is from the regular video, while the second part is from our visualization system. Both of these two parts are the results and feedbacks from the understandability and usability tests, which included their feedbacks and observations of researcher using the new visualization prototype. We will examine the visualization to see if it can provide answers to the following questions:

- Is the meaning of element encoding understandable in this visualization?
- Is the system easy to use?
- Are the functions of the system visible and conspicuous?
- Can the interactive design of system guide the users go through the whole visualization?
- Can users depict the flow of pedestrian with this visualization accurately and quickly?
- Can users easily find the anomalies in the video and perceive the reason behind it?

All answers to those questions will be evaluated by the case demonstration in different situations. The researcher will go through the whole design of the visualization and show how to depict the pattern of crowds' motion with different functions of this system. That will allow the researcher to have a general idea about how users would like to use this new visualization tool automatically, whether the way that represents the information is intuitive, and whether interactive functions are obvious and reasonable for users.

3.6 Summary

In this chapter, an overview of design thoughts and application procedure of this research is introduced. At first, it gives a design framework and layout of pedestrian motion pattern detection and visualization. Then how to collect the recording video and choose the appropriate dataset is discussed in the view of the limitation of this research. Moreover, several available object detection and tracking methods are compared according to their advantages and disadvantages. The specific procedure and relevant technology to map the data elements and graph elements is also provided as the base of the visualization. Finally, the demonstration and evaluation of this system is established to make sure the user friendliness and information readability, which can give the feedback and guideline for future improvement.

CHAPTER 4. DATA EXTRACTION THROUGH MULTI-OBJECT TRACKING

4.1 Identify User’s Needs

Pedestrian motion has been studied for a prolonged period in previous research from different view of data analysis. Previous researches can be divided in two categories, such as pedestrian’s detection, pedestrians tracking, etc. In this thesis, to figure out what information of pedestrian flows is needed and available in the real-life application, the researcher studied and compared a lot about nowadays surveillance video systems. Then the user’s need in the field of video analysis and the difficulty of corresponding solution application will directly determine the consequent algorithm chosen and visualization design.

According to previous studies researches noted that when it comes to pedestrian’s motion analysis, the researchers either focused on detecting the objects and tracking the objects in the video, or only did the analysis for the pedestrians who take the sensor to record the motion data. Few efforts were put into the data analysis based on the pedestrian detection and tracking result from surveillance video, while today’s technology has been very successful in these fields. (Figure 4.1).

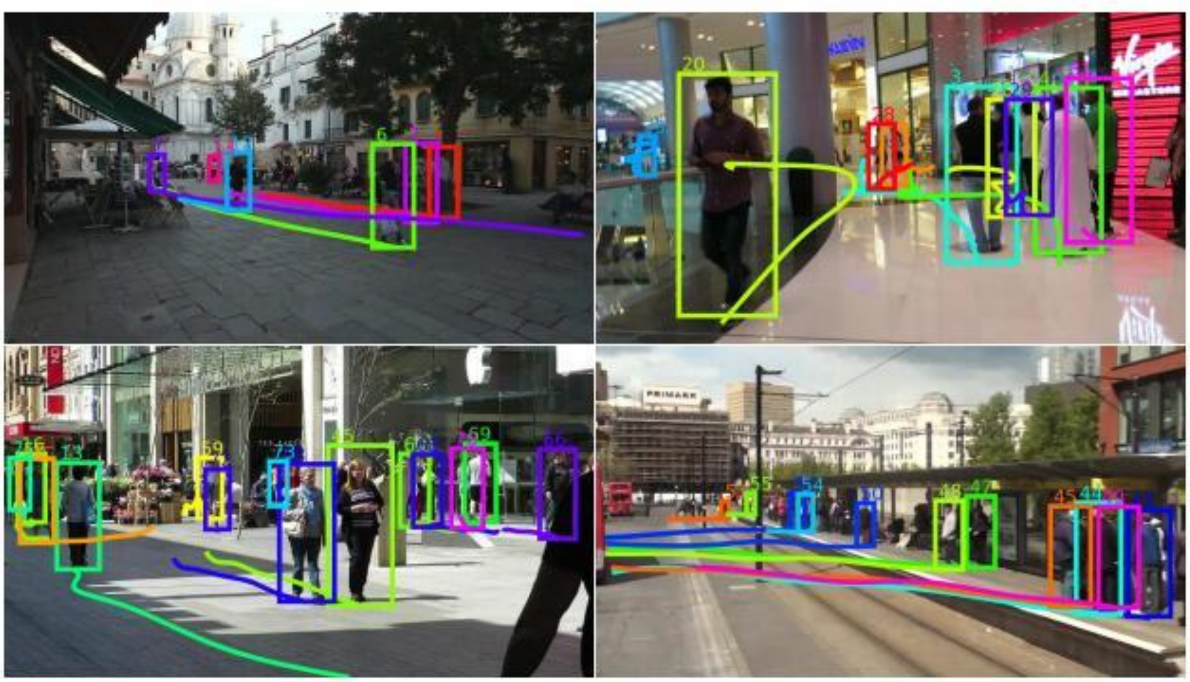


Figure 4.1 Detection and tracking of pedestrians

In surveillance videos, users may waste much energy and time to lock the destination and have no opportunity to analyze the motion or the pattern behind the motion of the pedestrians. For example, the linear trajectory of a person may turn around in a curve at some time, while some cars stop before the person in the video. Users need a comprehensive view of both the image and the data in the video. The object detection and tracking algorithms can help users filter the unimportant information and only focus the various timing in the video. Faced with a large amount of numerical data, the users may also want to get a perception of the relations among different kinds of elements, such as velocity, location, and direction (Figure 4.2).

blobs	points	frames	x	y	velocity	orientationx	orientationy
0	0	2	513.99	196.204	0	0	0
1	1	2	270.709	261.625	0	0	0
1	2	3	517.155	196.416	254.927	0.966732	-0.255793
1	3	4	514.538	197.916	3.01661	-0.867714	0.497065
1	4	5	510.336	199.592	4.52371	-0.928794	0.370596
1	5	6	504.359	201.058	6.15391	-0.971235	0.238121
1	6	7	500.637	202.025	3.84605	-0.967838	0.251573
1	7	8	497.853	202.396	2.80848	-0.99127	0.131845
1	8	9	288.593	196.672	209.338	-0.999626	-0.0273395
1	9	10	493.4	202.225	204.882	0.999633	0.027101
1	10	11	490.528	203.271	3.05621	-0.939558	0.34239
1	11	12	478.185	203.991	12.3642	-0.998304	0.0582178
1	12	13	476.08	205.891	2.83598	-0.742348	0.670014
1	13	14	470.311	208.288	6.24637	-0.923433	0.383761
1	14	15	463.844	209.978	6.68435	-0.967531	0.252752
1	15	16	459.405	210.91	4.53599	-0.978676	0.205409
1	16	17	456.322	210.943	3.08307	-0.99994	0.0109279
1	17	18	452.603	209.211	4.10279	-0.906529	-0.422143
1	18	19	451.4	204.783	4.58842	-0.262103	-0.96504
2	19	2	663.151	282.202	0	0	0
2	20	3	276.951	263.999	386.628	-0.998891	-0.0470795
2	21	4	286.726	260.475	10.3904	0.940701	-0.339238
2	22	5	289.751	259.646	3.13693	0.964453	-0.264255
2	23	6	296.246	258.036	6.69181	0.97062	-0.240618
2	24	7	301.7	257.324	5.50029	0.991592	-0.129404
2	25	8	309.726	257.106	8.02908	0.999632	-0.0271269
2	26	9	495.53	202.585	193.637	0.959543	-0.28156
2	27	10	320.479	253.806	182.391	-0.959757	0.280832
2	28	11	326.222	251.439	6.21194	0.924527	-0.381117
2	29	12	329.361	249.798	3.54157	0.886159	-0.463381

Figure 4.2 Dataset of pedestrians

It's extremely difficult to be aware of the data variation, not mention to capture the abnormal person and unusual motion. What's more, when users are watching videos, they can at most remember the things in limited few frame sequences. A good visualization of pedestrians should provide a global view for the motion patterns and trends.

User friendly animations are also vitally important in the design of this visualization. The interactive functions and interactive animations need to be obvious enough to make every user to understand it. Users can view every section and adjust every specific range of visualization themselves without any additional learning burden. In order to fully understand the visualization, users also need to be allowed to change the following visual attributes: timeline, symbols of object, zoom, drag, mouseover. In fact, the nature of these interactive animations is all going to attract user's attention to do more exploration and operation in this visualization system. It's a positive feedback system as shown in Figure 4.3. When some interactive operations trigger the animation, the animation in another part of visualization will also attract users to explore more.

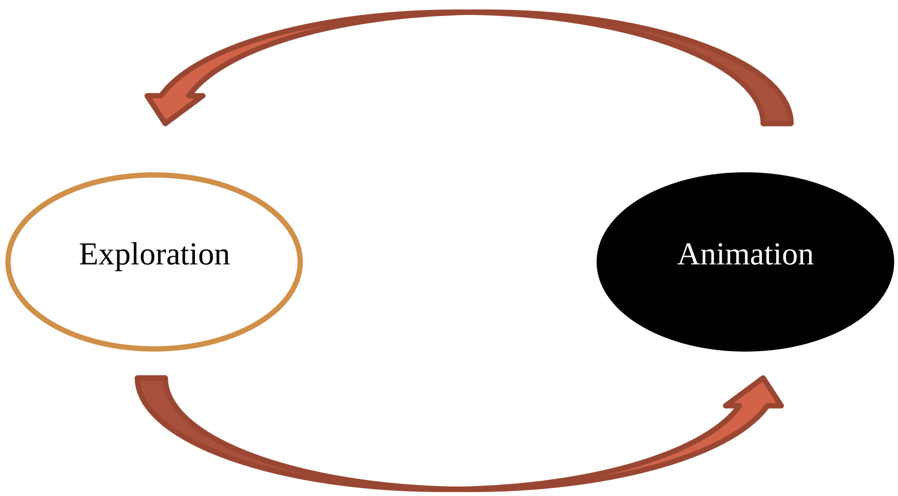


Figure 4.3 Animation and Exploration in visualization

4.2 Design Goals

Based on the general structure of visual surveillance including environment modeling, motion segmentation, object classification, tracking, behavior understanding and description the designs goals of this visualization are:

- Users can grab a global view of trends and patterns of pedestrian in the whole video sequence

- Users can detect the breakpoint or the variation of trends and patterns with the visualization interface.
- Users can see contrasting dimensions of material with different time granularities.
- Users may use this visualization to find the anomalies in the video tracking.
- Users can also consider the image plane and data visualization together.

The design of the visualization not only needs to consider the functional requirements, but also needs to provide a good user experience. According to the convenient heuristics of interactive interface created by Jakob Nielsen (1995) and the innovative architecture heuristics of information visualization from Forsell & Johansson (2010), the design goals derived from the user experience can be concluded as follows:

- The interactive design and the visualization presentation are intuitive for users.
- Users can easily be familiar with the visualization system and acquire the information they need quickly.
- With the visibility of the many functions users should be able to sensibly be aware of the main functions and comparable components.
- According to consistency and standards the interface should be designed with persistent standards to avoid inharmonious appearance.
- Valuable but not overwhelming information can be presented in each cluster of visualization.
- The design should abbreviate unrelated information and save the attention of users.
- Users will get a clear understanding of the system.
- Users can effortlessly use the interactive system.
- Users can interact with the structure flexibly.

4.3 Software Framework

In this study, the researcher implemented a user-friendly intelligent surveillance video system for pedestrian motion pattern awareness. The system processes the dynamic video frame with following three steps: pedestrians detection, pedestrians tracking, pedestrians motion visualization. For each step in this process, there are several frameworks capable of the requirement, which need to be compared and chosen carefully. In the field of motion detection and tracking, opencv and matlab are the most two commonly used software frameworks to do the image processing. Most of tracking researches on multiple object tracking used the matlab as the base.

Matlab is a full scientific software with its self-design language consisting of numerous existed image processing algorithm and modular, which leads to the cumbersome and expensive of the whole integrated develop environment (IDE). If researchers want to run their program in matlab framework, they need to write matlab code and adapt to Application Programming Interface (API) of the matlab. What's worse, researchers also need to install a more than 10GB integrated develop environment (IDE) and pay a lot on this mature product for its license. Matlab has a very simple code structure, good documentation and complete support, which is beneficial for the rapid prototyping and program debug. However, this closed complete system also has its weakness, matlab is not an open source and pretty pricey, so its program is inflexible and non-portable. As an interpreted language, matlab doesn't have a good execution performance as most programming language like C++ and python, though it's good at matrix calculation. Performance matters a lot in the field of video analysis, especially if you want to pursue the efficiency like real-time video processing. It's hard to code some customized programs with the high-level built-in function without any decrease on performance. For instance, matlab has the existed background subtraction modular, optical-flow modular, even the basic pedestrian's detection and tracking modular, but all the output API is fixed so that you can only use the data that the built-in function lets you know, instead of choosing anything you want to know in the procedure of video processing.

In recent years, opencv (Open Source Computer Vision Library) is the most widely used open source library for computer vision and it has a large user community. Opencv has more built-in functions for video analysis than matlab, while some of these functions can be implemented on GPU which can extremely accelerate the performance of video processing. What's more, as an open source library, opencv can be utilized with different popular programming language like: C, C++, Python, Java. The cooperation of C++ and opencv can generally runs faster than matlab code. However, the greatest advantage of opencv for this research is that it's easier to extract the data, even the interior data in the built-in function can be extracted by optimizing and modifying the source code. In this research, both of these two frameworks will be tested and the better one will be chosen.

As for the visualization tool, InfoVis Toolkit is a Java-based library integrating the commonly used data clearing algorithm and the delicate visualization model presentation, which simplify the manufacturing flow. Prefuse is a powerful tool library consisting of all the techniques that may be used in the pipeline of the visualization design like data structure, layout design, and various kinds of interaction animation. Tableau is also a powerful visualization system that is easy to use even for the non-trained users. It provides an abundant dashboard and visualization modular for users to choose. In recent years, document-driven documents javascript library has attracted the attention of the data visualization researchers. As a user-friendly webpage-based toolkit library, it supports the direct operation of the webpage elements (Scalable vector graphics, text) and simple binding between the data and the web elements. Some information-theoretic frameworks based on the webpage can provide the existed and integrated strategy for the visualization and interaction animation. To make the visualization enabled to be viewed in any platform, the researcher turned to the use of D3 and PHP development so that users can quickly and conveniently perform visualization in the web browser. The greatest advantage of D3.js is the high freedom of programming, which let users can not only used the existed visualization modular but also create their own customized visualization.

4.4 Data Collection and Processing

As for the data sources, this study may use the dataset from two sources. One is the video sequences from the dataset of MOT Challenge benchmark, the other is shot in campus by the researchers to simulate the real-life situation as a reference. The sequences collected by the MOT Challenge are commonly and widely used by the researchers from all around the world. It’s pretty easy to compare the results with the visualization in other researches. In order to avoid the unnecessary errors created from the shooting angle, this research only takes the video shots from the high ground rather than the video shot horizontally. The example videos of high angle shot, and low angle shot are given in the next page.

Figure 4.4 High angle shot video

Figure 4.5 Low angle shot video

The size and velocity of pedestrians detected in the high angle shot video are almost the same, while the differences between pedestrians in the low angle shot video are pretty obvious. Thus, for the video shot in the campus, this research fixed the cameral on the highest building near the crossing. Today's video and image always have a high-resolution to re-product the original scenery. This research doesn't rely much on the high-resolution video. This is because that research only focuses the relative spatial relationship of objects-background and object-object in the video. The pedestrian in this video will be simplified as a centralized circle or point. In fact, to recognize the pedestrians you do not need a large amount of pixel information, all of the pixels between the feature points can be ignored in the video processing. What the computer needs is only a small quantity of characteristic value and contour information. The gradually transiting pixels between the feature points can provide very little amounts of useful information. This research only uses the rear camera of mobile phone to shoot the video and the video captured also needs to be compressed to reach a low-resolution and reduce the subsequent computation pressure. However, the video got from the phone still has 1080 pixels in height and 1920 pixels in width, researches further get a sample of 576*768 video from the original video.

Pedestrians detection and tracking is another critical step to ensure a good visualization result. It determines accuracy of the extracted data and the final appearance of visualization. In this research, the researcher tried all the basic objected detection methods listed in the literature review. Firstly, researcher applied the motion based multiple object tracking modular in the matlab. This modular can distinguish the motion of objects in each frame and associate detections with the same ID to track the objects. This modular uses the background subtraction algorithm to detect the motion differences between frames and process the gray scale subtraction layer with the Gaussian mixture algorithm (Figure 4.6). Then it eliminates the tiny noise point on the resulting foreground mask and group the connected pixels into blobs to analyze the corresponding between moving clusters. Based on the grouped detections, Kalman filter is applied to figure out the association of motion detections. It is used to predict some detection's locations in next frame according to the motion and determine the probability of assignment for each motion detection. It's also important to take a good care of the tracking result, because the old tracking object may disappear or appear repeatedly over the time. The assigned object should be cleared from the memory to reduce the computation and re-identified as the same object in period of time. Both of the assigned and unassigned detections should be updated according to the corresponding detections over the consecutive frames. The strategy of adding new tracks simply relies on the total counts of tracking objects. In this modular, all of the detect pedestrians even the cars will be encircled in a labeled rectangle (Figure 4.6).

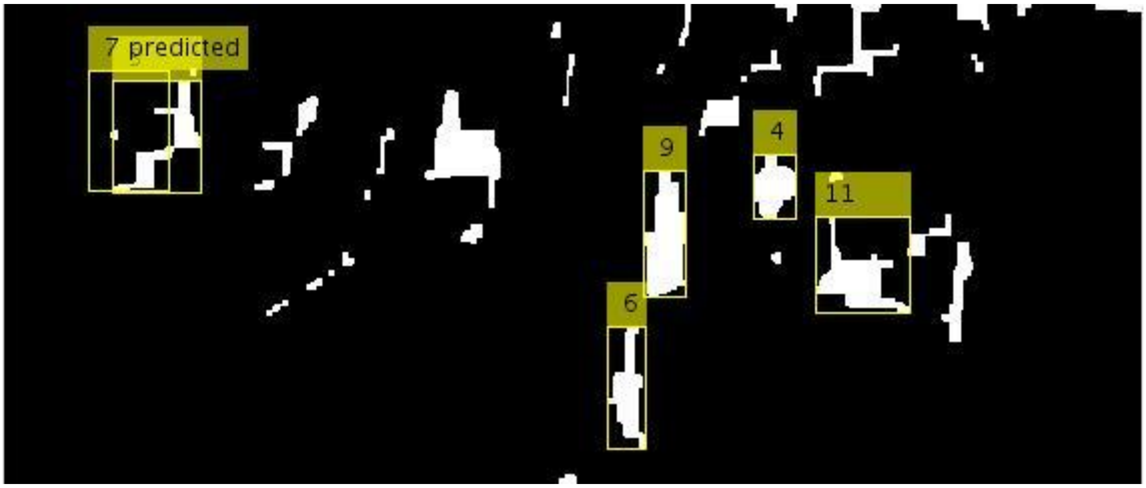


Figure 4.6 Background subtraction algorithm with Gaussian mixture algorithm

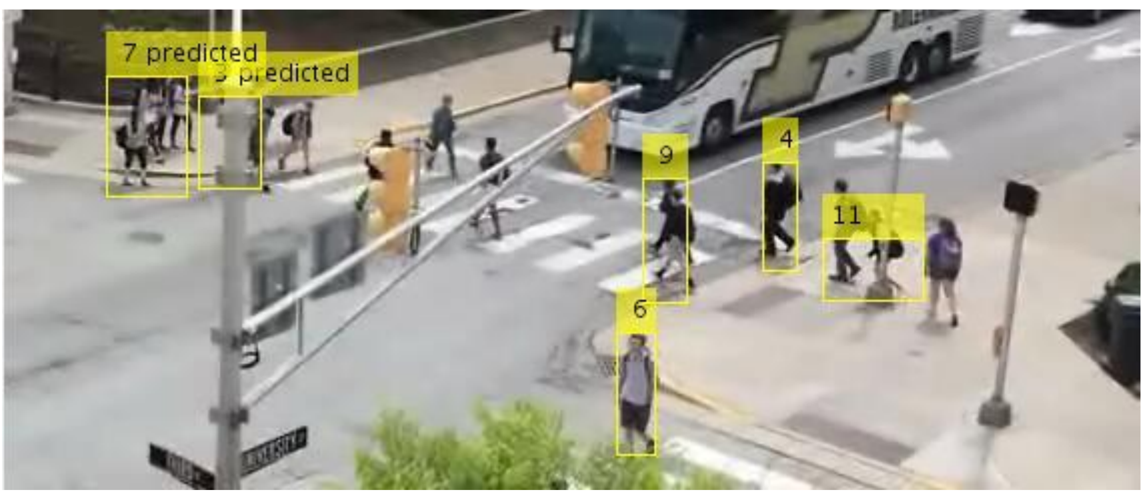


Figure 4.7 Result of multiple object tracking modular in matlab

However, it’s difficult to extract the data in the interior process of this modular, such as velocity, orientation. The velocity in this research can be described as the distance between locations of one object in consecutive two frames. If the detected rectangles in two frames are similar, the locations of the object can even be simplified as the center location of the rectangles. Although it’s easy to detect motion with background subtraction and give the roughly rectangle bound, it’s not wise to centralize the location of these variegated rectangles. If the height and the width of rectangle change dramatically in the motion, the location will be extremely unstable, which leads to the trajectory in Figure 4.8.

Figure 4.8 Unstable trajectory

As for this situation in background subtraction method, the detection rectangle changes drastically and it's not accurate to consider the center point as the location of the detection, because the center point will still move even when the detection only changes the shape of the surrounding rectangle. Thus, research needs another way to calculate velocity from the detection results. As everyone knows, optical flow is a good method to reflect the sequence of apparent motion in relation of objects, surfaces, and edges. In this way the pattern of apparent motion can be shown in every single point in the image plane. Optical flow will use a vector field to present the orientation and magnitude of the motion flow velocity at any location. The approximate motion range of the object has been restricted in a rectangle and this research only needs to get velocity estimation of optical flow in the specific region (Figure 4.9). The velocity vector of every pixel in the detection region will be summed and averaged as the total velocity of one object. Then the direction of the pedestrians can also be separated from the velocity vector. However, this solution brings out massive computation burden for the whole program, because the velocity and direction of hundreds of pixels will be computed in each frame. Especially with the matlab language, this solution needs several hours to process one video which is totally unacceptable for the preparation of data processing. Very long execution time of the program will not only affect the debugging efficiency of coding, but also delay the design process of the following data visualization.

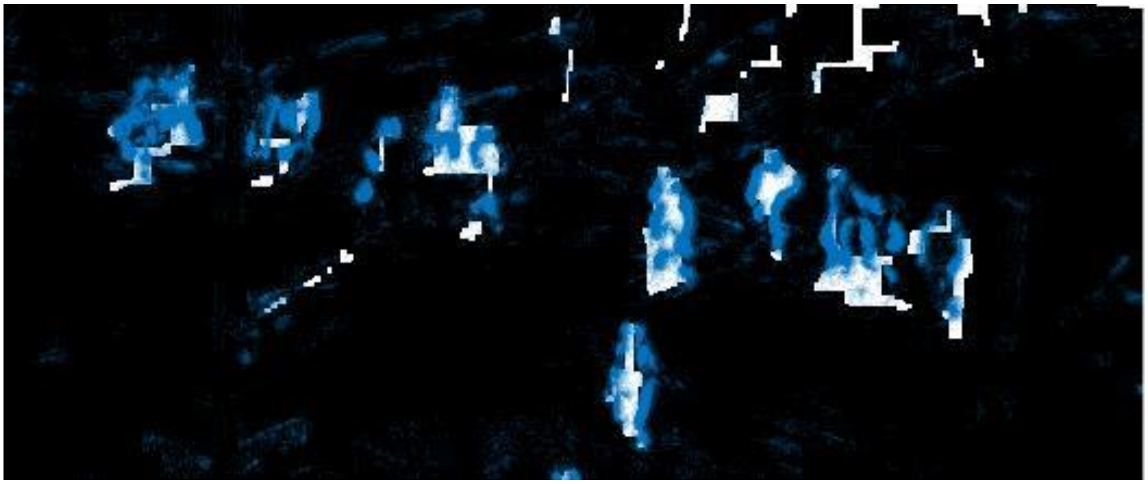


Figure 4.9 Optical flow with background subtraction

Besides the matlab solution, this research also compared the background subtraction method and SVM+HOG method in the C++ environment to check the programming performance of the SVM+HOG method and C++ language. Compared to matlab, C++ have an obviously advantage of efficiency both on the compiling and executing. The unstable trajectory is still a great problem in the background subtraction or even in the C++ environment.

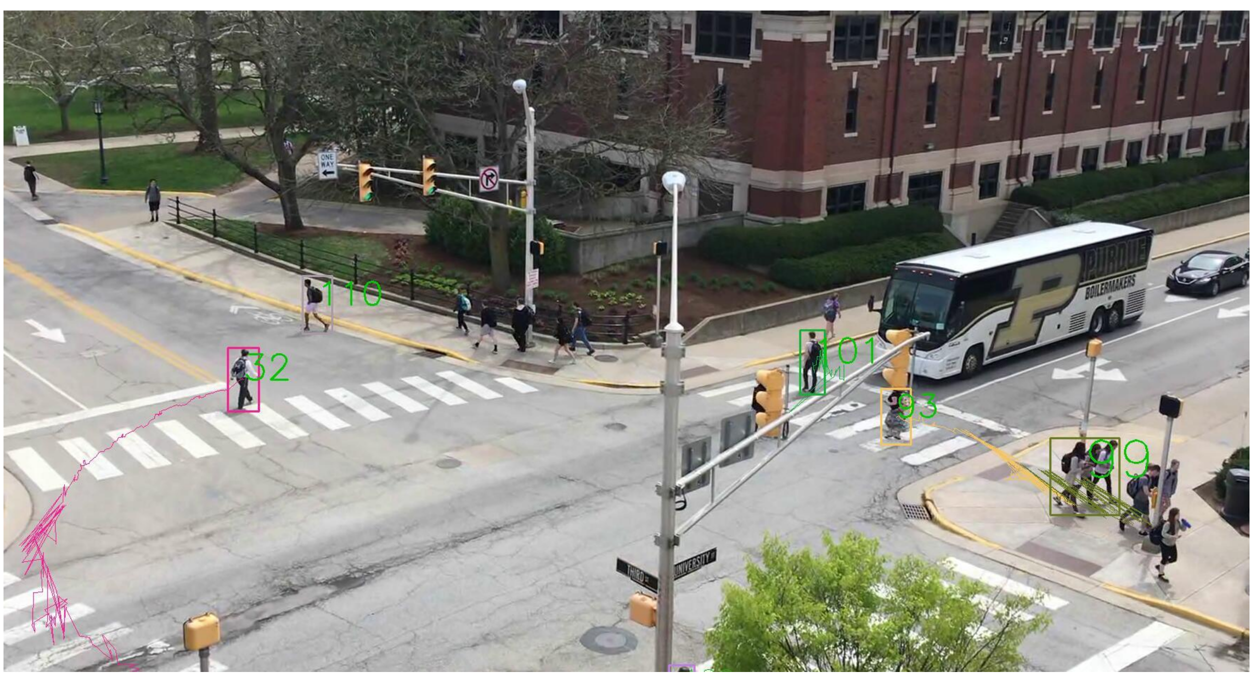


Figure 4.10 Optical flow with background subtraction

The solution of SVM+HOG method with C++ language does not perform very well in the pedestrian detection (Figure 4.11). Some bounding rectangles are set too big to block the detection of tightly-knit pedestrians, which makes the tracking trajectory furthermore tremble.

Figure 4.11 SVM+HOG in C++

Then this research tried to use YOLO, a state-of-the-art, real-time object detection system to detect the pedestrian. This system divides the image into several small images and predict the probability for each detection bounding box of different locations and scales through a single neural network. The bounding box with a high scoring predicted probability will be kept and considered as a pedestrian detection. Different from the previous methods, it focuses less on the features of local region but put more attention on the global content of the whole image. Its single network evaluation also makes it work extremely fast and accurately, especially with the help of a good GPU. Users can balance the tradeoff between speed and accuracy by choosing different models. Based on the previous basic image processing method in opencv, this system can have a pretty good detection result. All of the persons detected in the video are bounded tightly and stably over the time as Figure 4.12.

While the YOLO can provide a very good performance detection, this research can track the pedestrians based on the Simple-Online-Realtime-Tracking method (Bewley, 2016). In order to track the object, predictions of each object need to be made for the next frame. If a detection is associated with a target from the last frame, the state of this target will be updated with the processing result of last-frame prediction and current-frame detection in Kalman filter. If the detection is not associated with any target, the state of the target within next frame will be predicted with the linear model. The intersection-over-union (IOU) between each detection and the predictions can be directly used to do the assignment with the Hungarian algorithm because of the high accuracy of YOLO detection. When a section of video is processed, a series of information will be calculated and output into the online dataset like: x-coordinate, y-coordinate, velocity, x-orientation, y-orientation, person ID, frame count, object count. However, there will still be some data deviations in this tracking result, because two detections maybe considered as the same one if their trajectory has a certain extent of overlap. Thus, the data with an extremely high velocity will be discarded as a trembled outline point.

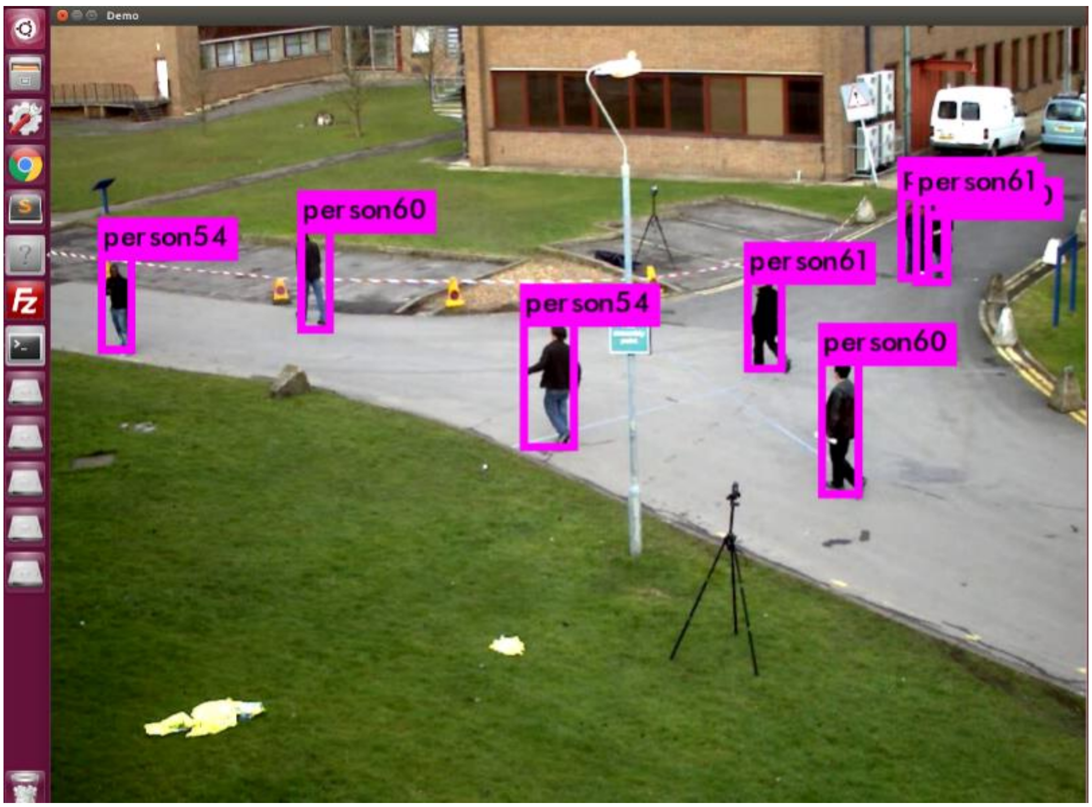


Figure 4.12 Final Tracking Detections with YOLO

4.5 Summary

In this chapter, the researcher clarified the user's need and corresponding data that will be used in the following research, including the coordinate, velocity, orientation, frame count, id count, detection count. Researchers compared different algorithms mentioned in chapter 2. According to these requirements, object detection is based on the opensource code of the YOLO object. Then the output detected object will be re-associated and tracked base on the thought in

Simple-Online-Realtime-Tracking. Finally, the data was successfully extracted and prepared for the visualization in next chapter.

CHAPTER 5. VISUALIZATION DESIGN

After processing the video and separating the results from the video, the researcher thought the main functions of this visualization were: providing an overview for the whole video, showing the pedestrians motion pattern, and allowing uses to filter the video information over the time. Considering all of these, the visualization should have 4 main sections to reflect the orientation variation and velocity variation in consecutive frames:

- The video section: show the local motion pattern and trajectory with the video to help construct the connection between the image and data.
- The timeline section: show the overview motion pattern in the view of data to help users to be aware of the anomaly in the video sequence data.
- The orientation section: mainly show the orientation variation of detection objects over the time.
- The velocity section: mainly show the velocity variation of detection objects over the time.

5.1 Visualization Explorations

According to all the design philosophies and the design goals listed in the chapter 2 and 3, the researcher started design visualization to reach different goals and compare the existed visualization design in D3. There are lots of the visualization design in the D3 gallery which provide abundant design thoughts and style. For example, bar chart can clearly show the difference of a series of objects for one dimension, then the users can perceive the fluctuation of the bar-height. The line chart can show even the subtle change in intensive discrete dataset. Similarly, the bubble charts sacrifice a certain degree of accuracy, but give more perception about the clusters and the difference between each two objects. Chord diagram focuses even more on the relations between objects. The scatterplot can overlap the objects in the same location, which is good at presenting the trend for a large number of data like this research. The calendar view presents every cell with unique color, which can help users find the anomaly in the dense dataset. Steam graph and area graph always demonstrate path transitions to compare the multiple dimension in different layers, while bi-partite graph and parallel coordinates graph construct the relations between the different dimensions. Radar graph is more useful to provide a view for the dataset that have a periodic variation. The application of different visualization can be classified according to the purpose: comparing the data, constructing the relation, finding the trends the following table 5.1. In fact, most of the successful data visualization are the aggregate of the appropriate basic visualization based on the characteristic of dataset. With the research collected for the data on the video, it possesses the following different types of information, being four in total: location, time, velocity, and orientation. Different combinations of these four parameters will require the distinctive design of visualization and reflect the different aspects of the data. Primarily, this research will choose the design framework from the two-parameter based visualization like: location-time visualization, location-velocity visualization, location-orientation visualization, time-velocity visualization, time-orientation visualization, and velocity-orientation visualization.

Table 5.1 Data visualization classification

Comparing the data	**Constructing the relation**	**Finding the trend**
Bar chart	Bubble chart	Line chart
Calendar view	Chord diagram	Scatter plot
Steam graph	Bi-partite graph	Map visualization
Area graph	Parallel coordinate graph	Trajectory visualization

5.2 Pedestrians' Trajectories

As for the video section, this researcher tried to give an intuitional view of the combination of data and video. This part should visualize the data and highlight some characteristics when users are browsing over the video back and forth. People are always easily attracted by the moving object. Thus, when users are watching the moving pedestrians in video and dragging the progress bar, how to draw the attention of users to the data visualization is vitally important. A temporal visualization is necessary here to depict the variation along the time, because the users will be sensitive to the dynamic visualization transforming with walking pedestrians at the same time. What's more, a mass of data and graph elements will be compressed when the dimension of time is expressed over the timeline dynamically. For example, the dynamic visualization can only use one point at some frames to express the change of the pedestrian, while the static visualization needs a line graph to do the same things. They don't have many differences for one group data, but it will be a tremendous comprehension burden and space wasting for the users to read hundreds of groups data in static visualization. Thus, the range of choices video section is shrunk into location-time visualization, location-velocity visualization, and location-orientation visualization. This research will go through these three sides one by one. Considering the similarities between the pedestrian flow and water flow, the researcher decided to simulate the fluxion and converge of water flow to visualize the pedestrian motion. Scatter plot is the most appropriate method to visualize the characteristic of flow-like data, because the scatter plot will draw every object as a point which is just like the water molecule in the water flow. The basic scatter plot uses the point as the symbol of every elements, which can only show the overall trend of the pedestrian flow (Figure 5.1).

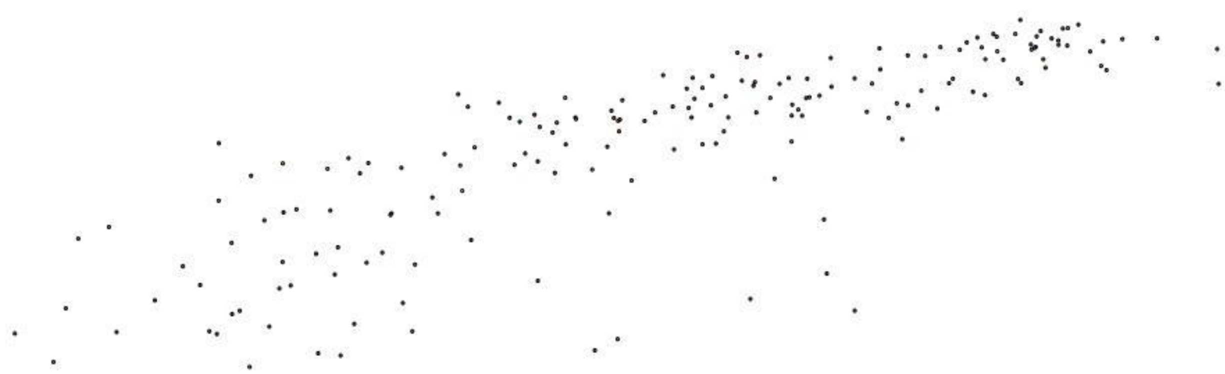


Figure 5.1 Scatter plot example 1

The overlapping point in the basic scatter plot always causes a great information loss, it is because the users will never know how many points are stacked at the same location. In order to visualize more information in this graph, researcher uses the circles instead of points as the symbol to represent every data element. When the circles are styled with a low opacity, users can observe the density change of the flow from the different depth of the color. What's more, the circles rendered with different colors can also help users tell the different categories of the objects (Figure

5.2). In this research, the color can be used to distinguish the pedestrians with different moving patterns.

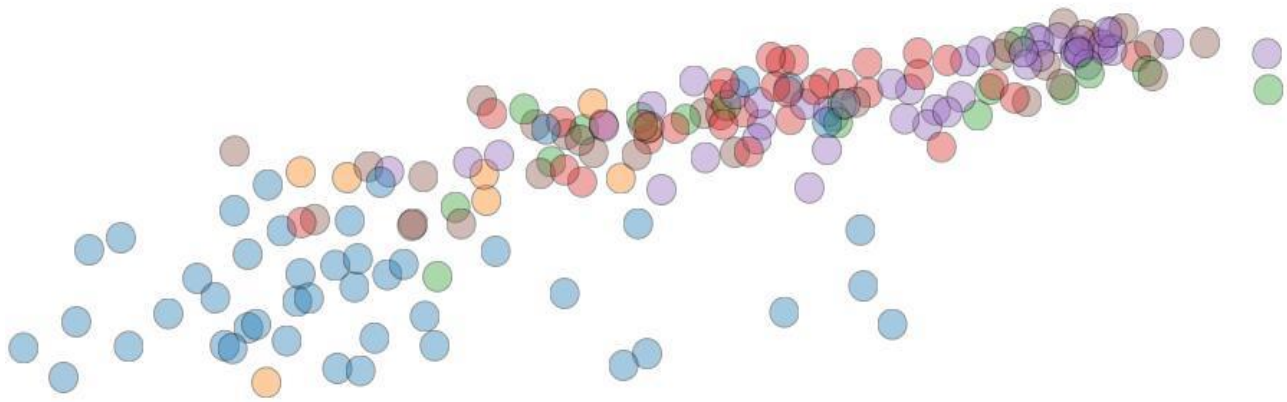


Figure 5.2 Scatter plot example2

Furthermore, the radius of circles can also be used as the carrier the information. This research associates the size of the circle with the velocity of pedestrian, which is also the only one can be compared among these four parameters in the dataset. This design has both the advantage and disadvantage. As for the advantage, the bigger circles with high velocity can be considered as anomalies instantly. However, the slow-moving pedestrians and even the static pedestrians will be neglected, because the size of these objects will be symbolled as a tiny circle even with the radius of 0 (Figure 5.3). Compared with those big circles, the users can barely notice these tiny circles, not to mention the circles are moving over the progress of the video.

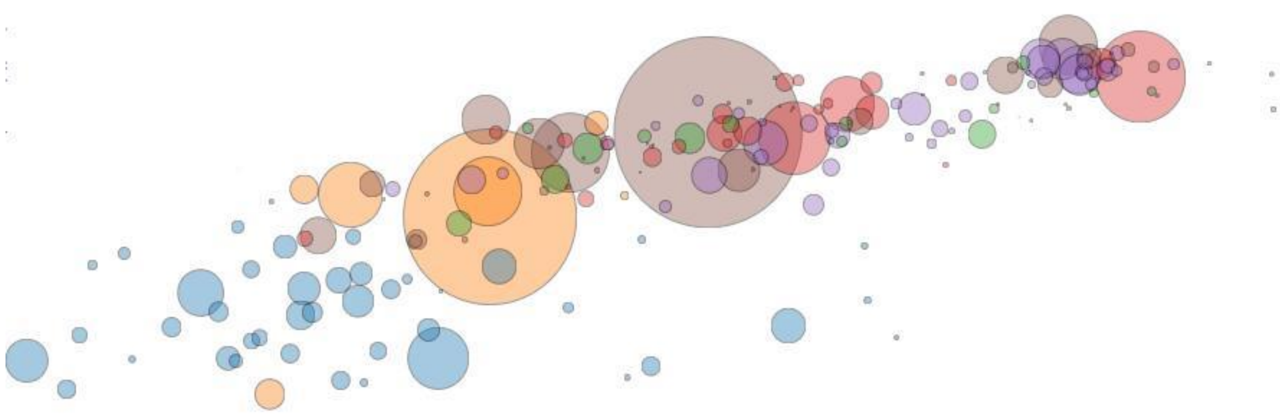


Figure 5.3 Scatter plot example3

Although the static pedestrians may not have the same importance as the high-speed pedestrians, they may also reflect some special things. For example, everyone may get stuck in some place, when an accident happened and blocked the intersection. In order to highlight the static pedestrians, researcher also came up with several solutions. It can use warm and cold colors to re-define the high and low of velocity (Figure 5.4). The radius of circle will decrease from the high speed to average speed, and the radius of circle will increase from the average speed to low speed. In this way, it can highlight both the high-speed pedestrians and low-speed pedestrians, but only neglect the pedestrians walking normally.

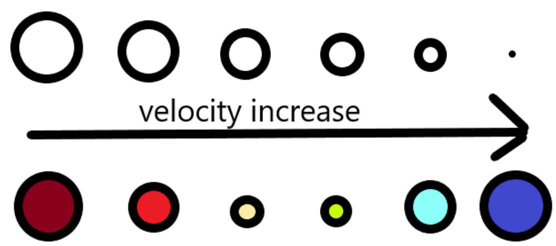


Figure 5.4 Scatter plot with warm and cold color system.

Another way is to replace the symbol of the static pedestrians from circle to another shape like square. All of the pedestrians with the velocity below a certain value will be represented with a fixed-size square (Figure 5.5). The size of the rectangle will increase with the decrease of velocity when the velocity is lower than the specified low-speed value line. However, both two methods mentioned above have a fetal weakness. They violate the cognitive rule of human. People are more likely to believe that the big symbol is associated with a high value instead of a low value. Thus, the two kinds of definition of symbol size may disturb the user's understanding of visualization.

Figure 5.5 Scatter plot with rectangle.

Besides these two methods, the researcher tried to use the local trajectory of one object to show the velocity variation trend at some time. In this visualization, the location of some pedestrian in current frame and its future locations in consecutive several frames will be covered with a circle. As shown in Figure 4.19, this scatter plot can almost show the instantaneous velocity and state of these five pedestrians. Users can not only observe the magnitude variation of velocity from the different sizes of circle, but also can observe the direction variation of velocity from the local trajectory. In this way, users can easily catch the objects which gradually become bigger or smaller.

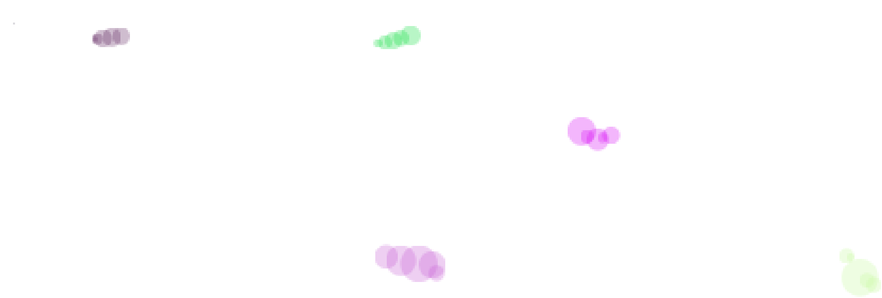


Figure 5.6 Scatter plot with local trajectory

When users find anomaly at some time, the next thing they have to do is to check further and figure out the total motion behaviors and motion pattern of this object. If the total trajectory matches the local trajectory, the pedestrian can be judged as a normal person and that anomaly at local don't mean anything at all. Only when the local trajectory and total trajectory have a big difference on motion pattern, it can be deduced that something strange happened here before. For instance, it's normal to see that a runner on the sidewalk changes the velocity and direction frequently to avoid other pedestrians, but it's abnormal to see some pedestrians walking with a constant speed suddenly speed up or slow down. Thus, it's pretty necessary and useful to combine the local trajectory with the total trajectory in these sections. The use of scatter plot is appropriate for this total trajectory as well. Scatter plot shows all the location where the pedestrian has been, and the density of the trajectory can also show the velocity variation over the whole timeline. This section takes the scatter plot of the total trajectory as the second layer graph behind the local trajectory. When the users move the mouse over the object, this visualization will show the total trajectory layer of that object (Figure 5.6).

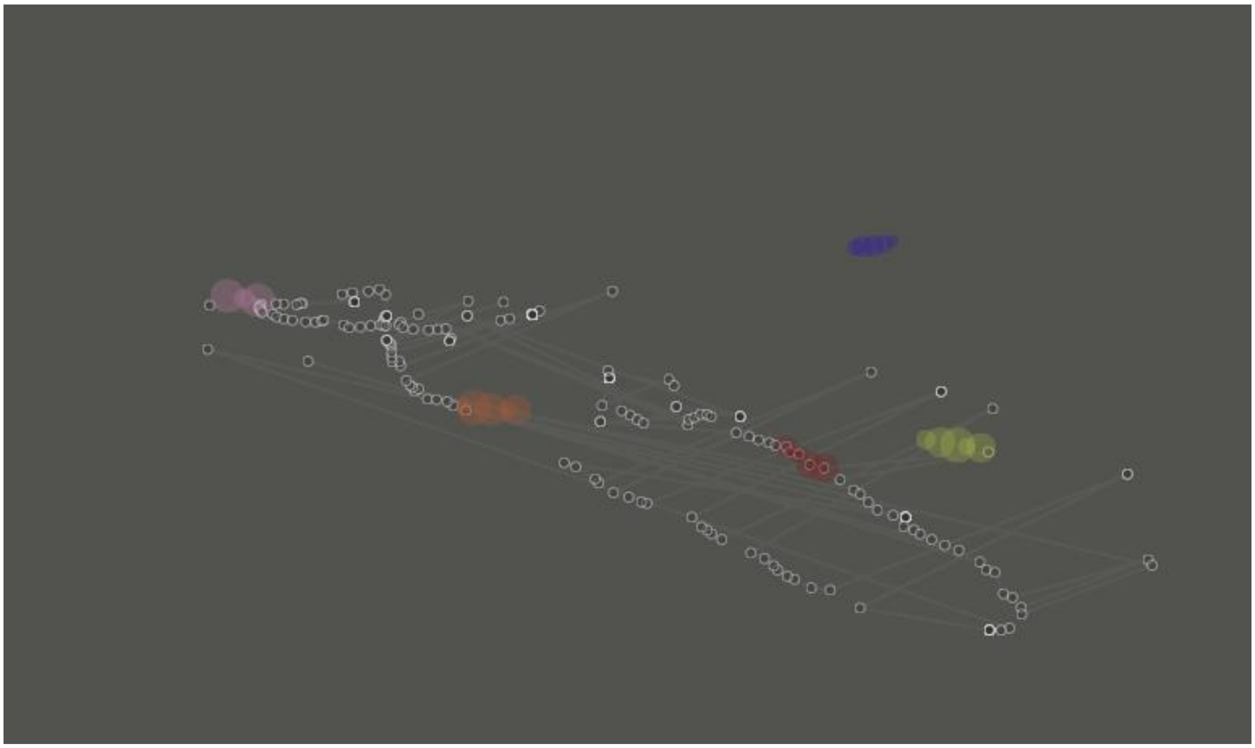


Figure 5.7 Total trajectory of one object

To construct the connections between the video and data visualization, more work needs to be done on this view. The video with the opacity of 0.5 is added as the third layer to show the additional information out of the data visualization. Then users can get the result of this three-layer combination in Figure 5.8. If the users find something unusual on the trajectory of some pedestrian, they can check what was happening at that time in this visualization.

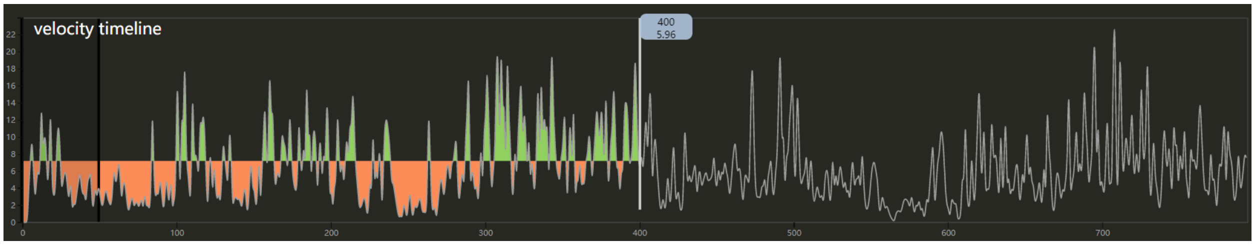
Figure 5.8 Video section.

5.3 Speed and Directions

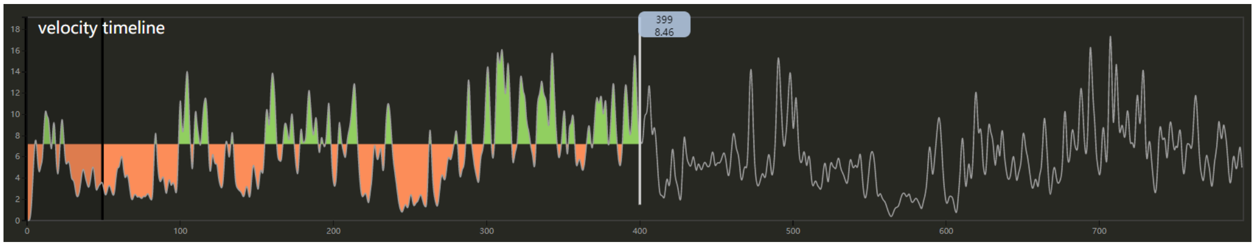
As for the timeline section, the visualization is going to solve the relation between the velocity and the time. Researcher tried to show the velocity fluctuation of every frame and help figure out which frame has anomalies in the view of whole timeline. Therefore, the x-axis is set as the timeline of the video, meanwhile, the y-axis is set as the velocity value. The problem here is how to judge the velocity of all pedestrians in some frame. Researcher provides two strategies for this visualization. When the number of the pedestrians is not so big in the video, the velocity fluctuation of the frame is computed as the average velocity of every object in this frame. When the number of the pedestrians is big in the video, the object with velocity away from median will be recorded as unusual object and the velocity fluctuation of the frame is computed as the average velocity of every object with unusual speed in this frame. In order to better help users to be able to perceive the crests and troughs of the velocity fluctuation over the timeline, the difference chart is used in Figure 5.9, which use the median velocity to divide the velocity fluctuation. However, the original graph (Figure 5.9A) has too many unreasonable fluctuations, which is caused by the estimation error in the object detection and tracking. The researcher tried to simulate the real simulation by averaging the velocity of local time range. In the seven graphs in Figure 5.9, the velocity averaged in 1 frame; 3 frames; 5 frames; 7 frames; 9 frames; 11 frames; and 13 frames in the sequence. Researcher compared several periods to compute the moving average. Researcher want to remove non-realistic anomalies but still keep the data with maximum sensitive. Finally, the graph averaged in 11 frames chosen for the better simulation performance and the less information loss. Then the users can easily observe the crest in the green regions and the troughs in the orange regions. The gray control bar in the middle can show the current state of this frame when the users move mouse over it. Moreover, dragging the control bar can the colored range of difference.

Figure 5.9

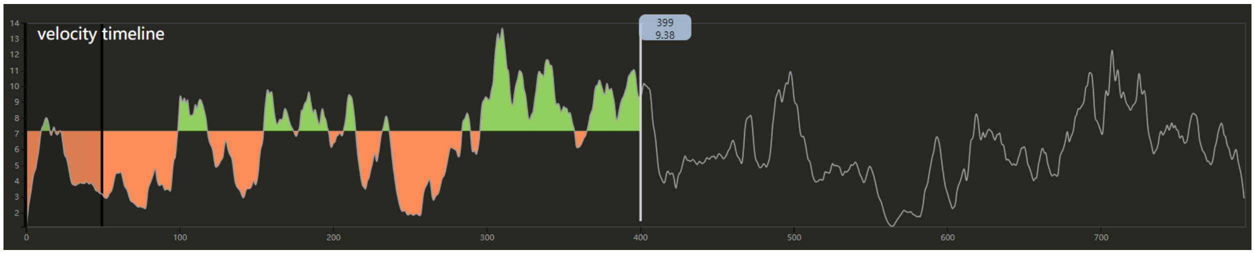
A)



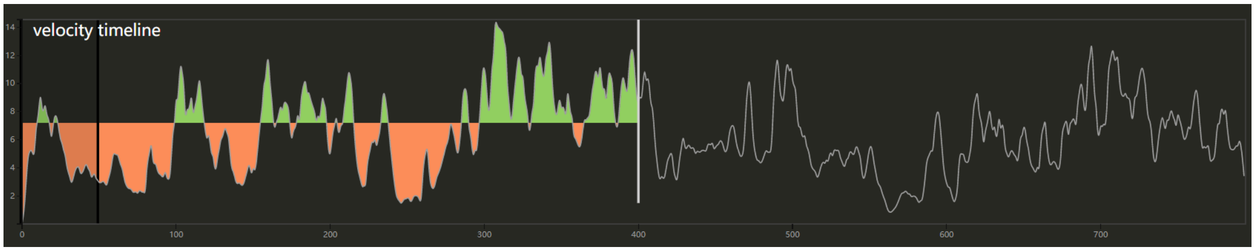
B)



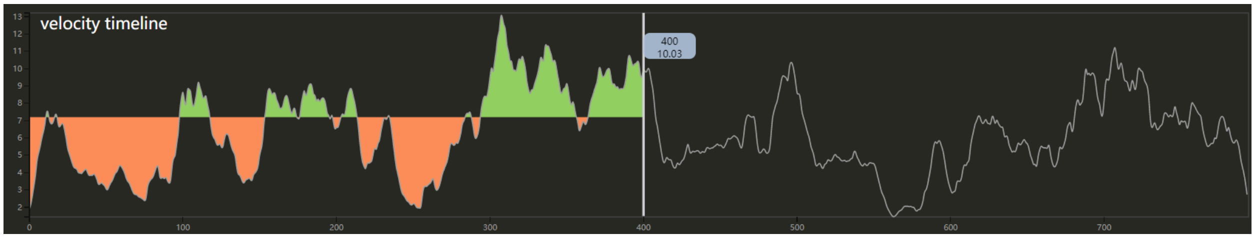
C)


velocity timeline
399
9.38
0
100
200
300
400
500
600
700

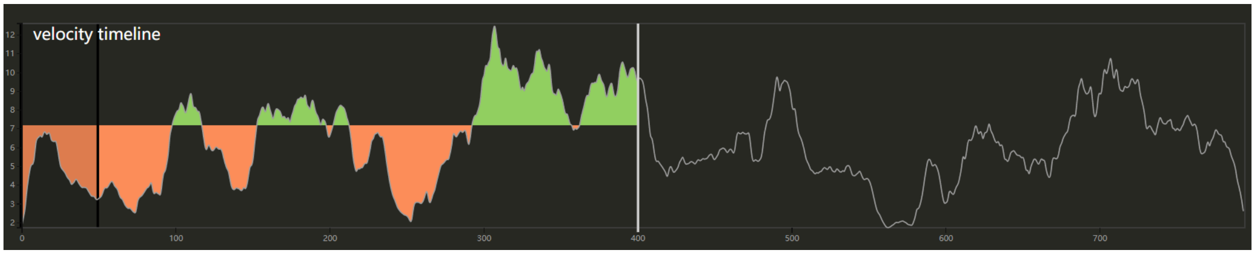
D)


velocity timeline
0
100
200
300
400
500
600
700

E)


velocity timeline
400
10.03
0
100
200
300
400
500
600
700

F)


velocity timeline
0
100
200
300
400
500
600
700

G)

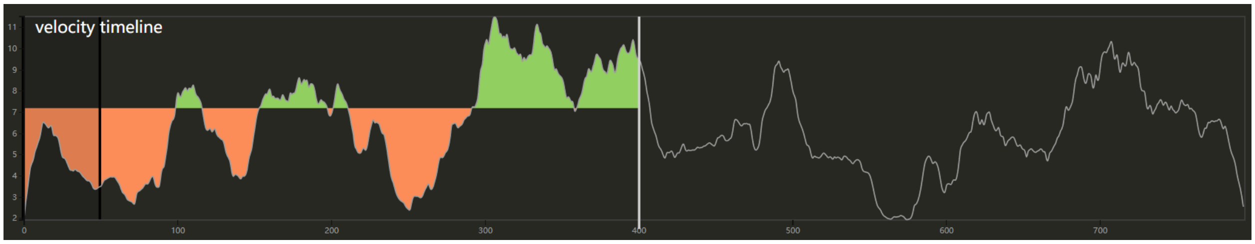


Figure 5.9 Difference chart in timeline section. A) Velocity averaged in 1 frames. B) Velocity averaged in 3 frames. C) Velocity averaged in 5 frames. D) Velocity averaged in 7 frames. E) Velocity averaged in 9 frames. F) Velocity averaged in 11 frames. C) Velocity averaged in 13 frames.

As for the orientation section, the visualization is going to solve the relations between the orientation and the time. Since the orientation is the circular type data, the radar type chart is more appropriate for this section. The traditional radar chart generally focuses three to six dimensions, but this orientation data is separated on 360-degree dimension. In order to depict distribution of the orientation, researcher modified the point-line structure of radar chart into the scatter-pie structure, because the connection between the objects don't have any meaning in the aspect of angular degree. The tiny circle symbol is scattered from interior to the exterior with the increasing of the time. The opacity of the circles is set as 0.3 to reflect the density of the object in some direction. The circles are also filled with the corresponding colors on the color wheel, which can make people more sensitive to the direction of the color. As shown in the Figure 5.10, users can easily observe that most circles gather in the red region and blue region and conclude that most pedestrians in this video was moving left or moving right. What's more, in order to show the direction variation of one pedestrian, the circles of the pedestrian with the same id will be connected into a gray curve. This curve can be considered as the direction trajectory of that pedestrian over the timeline.

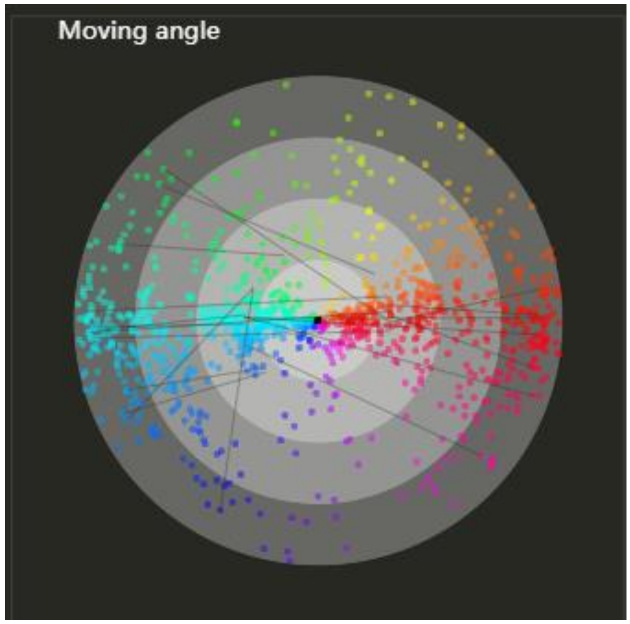


Figure 5.10 Orientation section.

As for the velocity section, researcher tried to create visualization which can reflect the variation of velocity both on magnitude and direction in detail. The velocity needs to be compared among different pedestrians and at different time of video. Thus, the visualization is designed as a mesh structure with the x-axis of time and y-axis of different pedestrians. To visualize the time-related parameter, the best way to show the variation is still the line chart. However, it's hard to compare the value on several lines of different pedestrians. For example, if users see the Figure 5.11, they can only compare the value of velocity for the pedestrian in one line, but different lines of pedestrians need to be separated to avoid the misunderstanding. Thus, a more common and reasonable judgement need to be prepared for this visualization. In two-dimension image plane, the most obvious metric for the graph elements is the size. The size of elements can be compared at any location of the image plane.

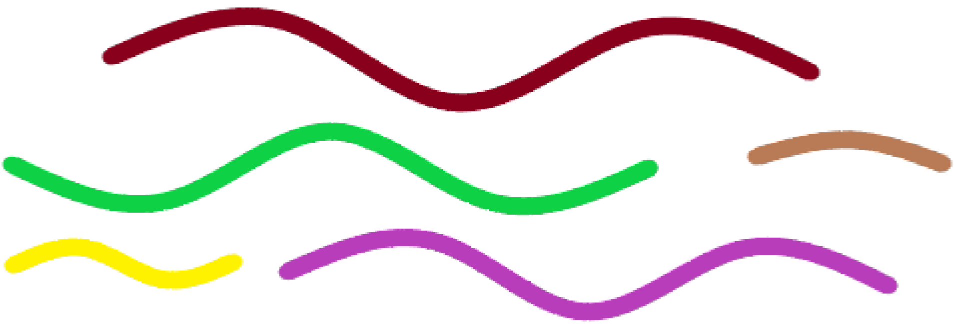


Figure 5.11 Line chart example for velocity.

How to expand the size of the lines became a big challenge here. Considering about the traditional audio waveform diagram (Figure 5.12), the size of the line can be represented as the amplitude the wave segment. However, the disadvantage of using audio waveform diagram is apparent, the intensive lines will mess the whole visualization especially when lots of pedestrians showed in the video. What's worse, it's hard to calculate the mathematical expression function of the wave needed in this visualization.

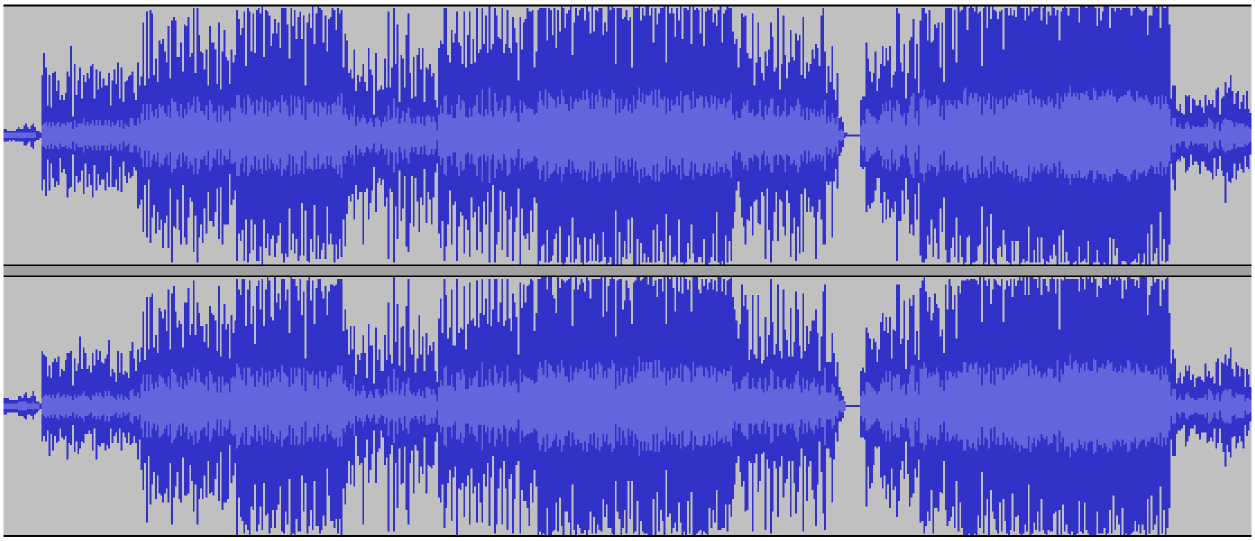


Figure 5.12 Audio waveform diagram.

Additionally, more work needs to be done on simplifying the structure of the waveform. In fact, only the contour of the waveform matters the appearance and folding curves inside the waveform don't have much meaning for this research. Therefore, this visualization can only take the envelop line of the waveform graph to show the magnitude of velocity. In order to tell the difference between the frames, the line will also be separated into the small line segments with different thickness like Figure 5.13. Each line represents a different object, and the thickness of the line segment represent the velocity of that pedestrians at some specific frame. In this visualization, these line segments are filled with different colors to represent the corresponding directions according to the color wheel. After users go through the orientation section, it will be easier for users to understand. In Figure 5.13, users can not only find the gradually changing of the color, but also find the pedestrians who turns around immediately from the contrasting colors of the line segments.

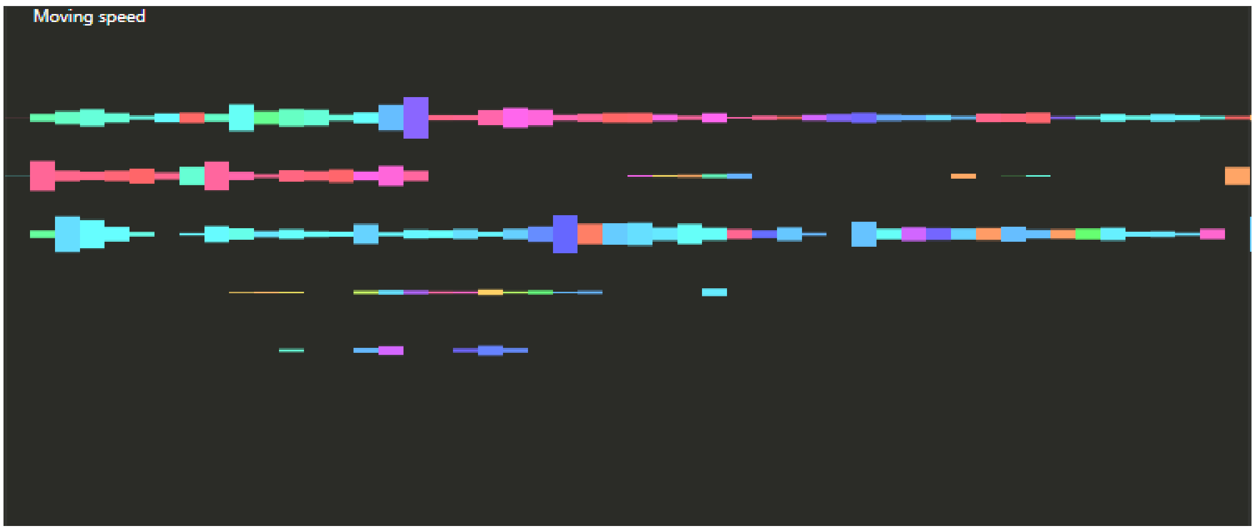


Figure 5.13 Velocity section.

5.4 Interactive Design

After all the four sections are completed, the interactive design in this visualization (Figure 5.14) needs to be considered elaborately. Interactive design is vitally important in the design of data visualization. It lights up the interest of users and guides them through the whole visualization. In this visualization, the timeline section is designed as the controller of the whole graph, which can control the progress of the video and other three sections. When the users access this visualization, this timeline section will show up with an expanding animation to highlight the function of controller. Two control modes are provided for this section. One is to drag the white lines in the middle. When users move the mouse over the white line, a tooltip will show up near the white line to demonstrate the current frame and the average velocity of this frame. The other is to drag the shadow brush bar to adjust the timeline of the whole system. The time of other three sections will always be kept the same with this time section.

When users move the shadow brush, they can also glance over the video, trajectory, or even the velocity section in the bottom. The design of every elements in these three visualizations is also vitally important, because the motion of them over the timeline is still a part of interactive design. The change on the timeline should let users quickly detect the great circle in the video section, the thick line segments in the velocity sections, which may represent the anomalies with a high speed. Thus, the scale of the circles and line segments should be not too big to overlap together and not too small so that the users can detect the difference. The change on the timeline should also let users perceive the color change in the velocity section. Since the color is set according to the color wheel, the contrasting color of turning around can easily grab the attention of the users. The lightness of the color also needs to be turned up so that the moving line segment can be more easily perceived in this dark background.

What's more, when users move mouse over any circle or line segment, a tooltip will pop out to show the current information of the corresponding pedestrian and the circles in the direction section with the same id will be connected to show the variation of direction. The circle in the video section will show the current id, while the line segment in the velocity section will show the 3 parameters, including id, velocity, and current frame. This animation can help user check the interested object between these two sections back and forth.

When users move the mouse over the circle, the total trajectory of the corresponding pedestrians will be shown with the smaller gray circle. The circles in the video section can even be dragged following the corresponding individual trajectory as another timeline controller. The local trajectory will be move to the corresponding frame at the same time. The users can have an overview of crowds' motion in this video.

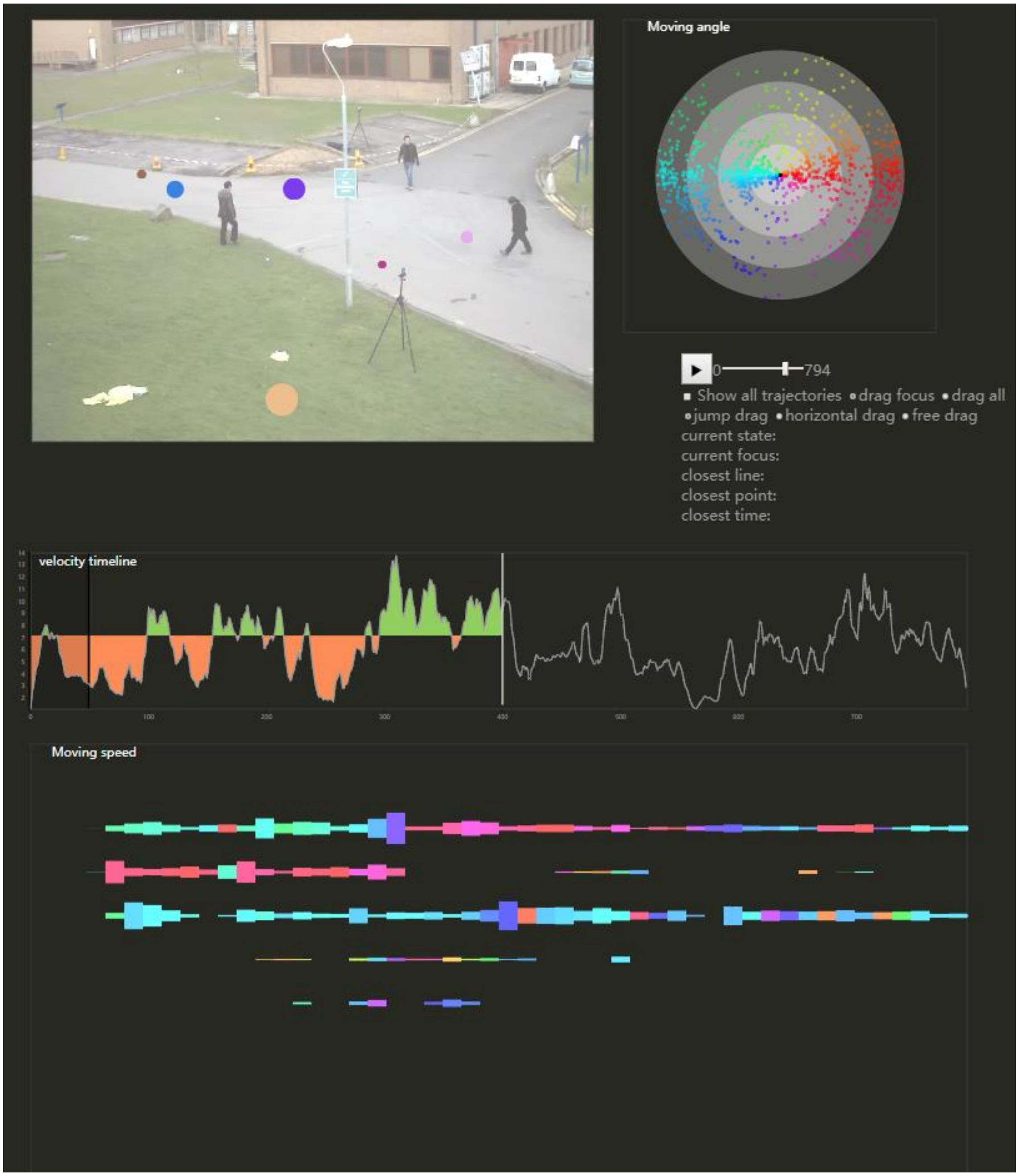


Figure 5.14 Final visualization.

5.5 Use Case Demonstration

This section intends to use this intelligent surveillance video system to analyze a specific case in the video. The video used here is still the same video shown in the Figure 5.14 and all of the analysis in this section is totally base on the final visualization in Figure 5.14. As the controller of the whole visualization, the timeline section is the first part need to be concerned by the users. The total velocity of all pedestrians in the video are overviewed over the timeline. In order to demonstrate the utility of this visualization, a prominent crest or a prominent trough will be chosen as the typical cases to analyze in detail. The crest and trough can be viewed in the graph distinctly, because the velocity line is segmented into the orange zone and green zone by the median velocity. For example, the researcher firstly moved the brush to the trough in the Figure 5.15. It’s almost the lowest trough in this video. The users may want to know what happened here and why the crowds slowed down.

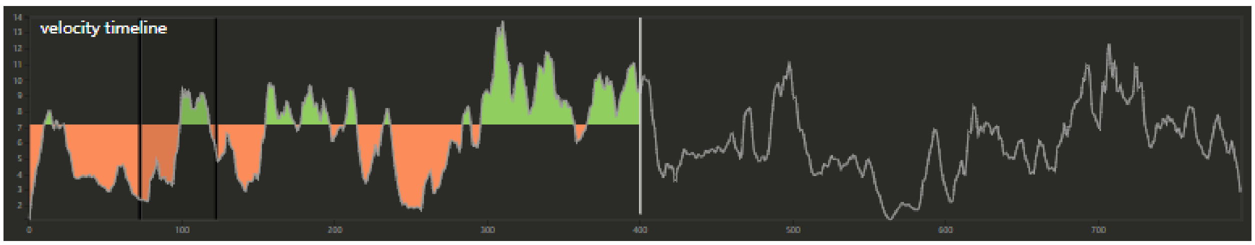


Figure 5.15 Moving the brush to the trough area.

When the users check the orientation section from the Figure 5.16, they can observe that most of the points are aggregated in the blue region and the red region, which demonstrate that most of pedestrians are moving to the left or right in this video.

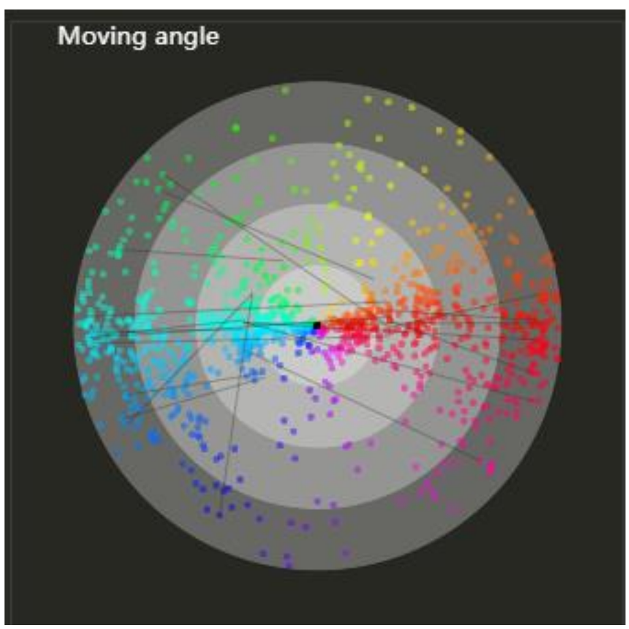


Figure 5.16 Checking the orientation section.

Then users can go through the velocity section in Figure 5.17 to acquire a detailed view of the variation of speed and direction in a range of consecutive 50 frames. In the red zone of current frame, there are four moving objects: red object, blue object, green object, deep blue object. According to the color set in orientation section, users can know that the direction of these objects are changing frequently. From the thickness of the line, users can also know that only the green one has a greater speed and the others have a lower speed. When users move the mouse over the green line segment, a tooltip will show up to depict the information of that pedestrian. In this example, 80 means the current frame, 20.08 means the current speed and 14 means the id of this pedestrian.

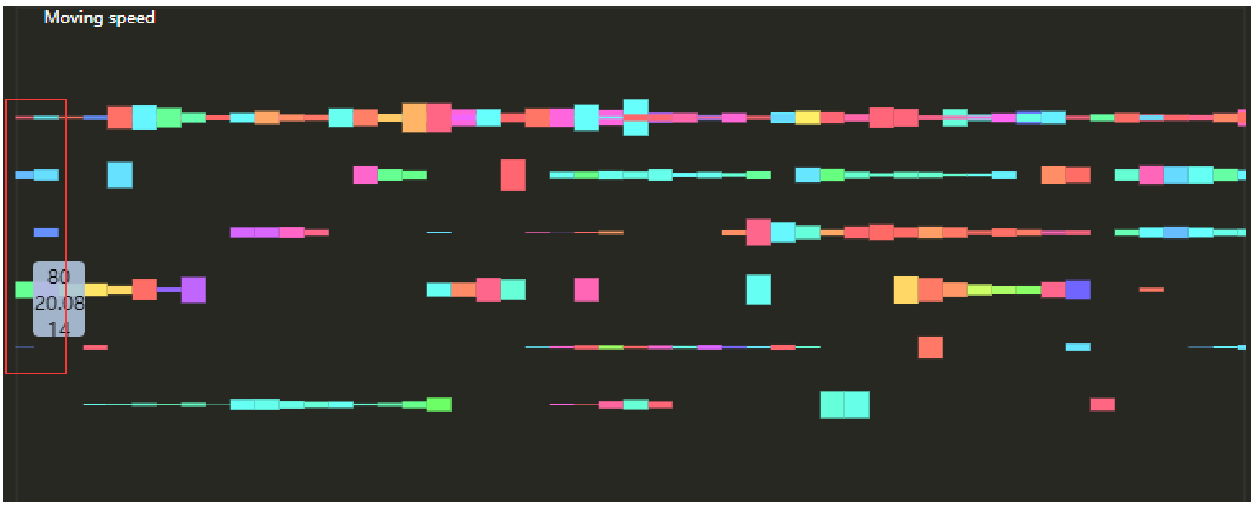


Figure 5.17 Checking the velocity section.

At the same time, the video section will also show the id around the interested pedestrian. In Figure 5.18, id 14 is attached to the interested pedestrian with a bigger brown circle, which means that this person has greater speed at this moment. The pedestrian with id-14 just sped up and left the intersection according to the brown local trajectory and the gray total trajectory. Combining the image information and the visualization, users can find that the pedestrians with the smaller circles in the yellow zone just arrived the intersection and talked with each other. That’s the reason why the people slowed down and even stopped here.

Figure 5.18 Checking the velocity section.

5.6 Summary

According to the design philosophies and design goals, this chapter clarified the complete procedure of the visualization design in this research. Firstly, researcher compared the basic visualizations widely used in daily life and classified them into different categories according to their characteristics. Then researcher designed four functional sections in this visualization, including the video section, the timeline section, the orientation section, the velocity section. After that, interactive animation design was added between the sections to guide the users and attract the attention. At last, a comprehensive case study is demonstrated to prove the effectiveness and practicability of this visualization. In the next section, the researcher will give an overall conclusion and a review of the work that will need to be done in the future.

CHAPTER 6. CONCLUSION AND FUTURE WORK

In conclusion, with this research, the researcher proposed a new intelligent surveillance video analysis system in which we can effectively support users in gaining a thorough understanding of pedestrian motion patterns. This research combined the proven object detection project and object tracking algorithm to process the raw video and extract the pedestrian information. This system can not only detect and track the pedestrians from the unprocessed video, but also provide a powerful visualization design to analyze the processed result. It helps users to find out the anomalies and perceive the motion pattern behind the behavior of pedestrians. Using the orientation section of the visualization, users can learn the orientation distribution of pedestrians. Using the timeline section of the visualization, users can learn which frame in the video has the anomaly and unusual accident. Using the velocity section of the visualization, users can observe the variation of velocity and orientation over the timeline. Using the video section of the visualization, users can gain a better view of the pedestrians' behavior with the help of original video. The interactive design in this research construct the connections between each section, which help users go through the whole visualization easier and analyze the anomalies in the video quicker. What's more, all the parts of this system applied the online real-time processing algorithm, which guarantee the shorter processing time and analyzing time.

Although this system helps users a lot on the video processing and pedestrian motion pattern analyzing, there is still further work to be done. First, a more accurate tracking algorithm needs to be applied in this system, which can smooth the final visualization. Considering the efficiency, this research used an online real-time tracking algorithm, which caused matching errors of some extent. To avoid the matching errors, this research increases the matching threshold of the tracking algorithm and filter barely matched objects, which discard a great amount of the information. In the future, researchers will try to apply the tracking algorithm with deep learning neural network on this project to optimize the tracking result.

Secondly, in this research, researchers only considered the good quality videos shot with a depression angle of almost 90 degrees. The reprojection error in this problem is neglected in this research, which means that the velocity computed from the coordinate in video may have a significant difference from the velocity in reality. More research needs to be done on the reprojection error to find an appropriate solution for this project.

Thirdly, three-dimensional interactive visualization will be added with the help of three.js, which can render the vivid flow stream effect for the trajectory. Since the trajectory in this research is simulated the flow in the water, it’s certainly better to use a better dynamic animation instead of a static trajectory. This can help users better perceive the flow of the crowds.

REFERENCES

Afzal, Maciejewski, Yun Jang, Elmqvist, & Ebert. (2012). Spatial Text Visualization Using Automatic Typographic Maps. *Visualization and Computer Graphics, IEEE Transactions on, 18*(12), 2556-2564.

Andrienko, N., Andrienko, G., & Gatalsky, P. (2003). Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages & Computing, 14*(6), 503-541.

Andrienko, Natalia, Andrienko, Gennady, Stange, Hendrik, Liebig, Thomas, & Hecker, Dirk. (2012). Visual Analytics for Understanding Spatial Situations from Episodic Movement Data. *KI - Künstliche Intelligenz, 26*(3), 241-251.

Barron, J., Fleet, L., & Beauchemin, D. (1994). Performance of optical flow techniques. *International Journal of Computer Vision, 12*(1), 43-77.

Bera, A., & Manocha, D. (2015). REACH - Realtime crowd tracking using a hybrid motion model. *Robotics and Automation (ICRA), 2015 IEEE International Conference on,* 740-747.

Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple Online and Realtime Tracking.

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D³ Data-Driven Documents. *Visualization and Computer Graphics, IEEE Transactions on, 17*(12), 2301-2309.

Chen, W., Guo, F., & Wang, F. (2015). A Survey of Traffic Data Visualization. *Intelligent Transportation Systems, IEEE Transactions on, 16*(6), 2970-2984.

Coifman, Beymer, Mclauchlan, & Malik. (1998). A real-time computer vision system for vehicle tracking and traffic surveillance. *Transportation Research Part C, 6*(4), 271-288.

Comaniciu, D., Ramesh, V., & Meer, P. (2003). Kernel-based object tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 25*(5), 564-577.

Cook, K. A., & Thomas, J. J. (2005). Illuminating the path: The research and development agenda for visual analytics.

Diansheng Guo, & Xi Zhu. (2014). Origin-Destination Flow Data Smoothing and Mapping. *Visualization and Computer Graphics, IEEE Transactions on, 20*(12), 2043-2052.

Erwig, M., Gu¨ting, R., Schneider, H., & Vazirgiannis, M. (1999). Spatio-Temporal Data Types: An Approach to Modeling and Querying Moving Objects in Databases. *GeoInformatica, 3*(3), 269-296.

Fekete, J. (2004). The InfoVis Toolkit. *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, 167-174.

Fen Xu, & Feng Xu. (2013). Pedestrian Detection Based on Motion Compensation and HOG/SVM Classifier. *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2013 5th International Conference on, 2,* 334-337.

Ferreira, N., Poco, J., Vo, H., Freire, J., & Silva, C. (2013). Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips. *Visualization and Computer Graphics, IEEE Transactions on, 19*(12), 2149-2158.

Forsell, C., & Johansson, J. (2010). An heuristic set for evaluation in information visualization. *Proceedings of the International Conference on Advanced Visual Interfaces,* 199-206.

Fradi, H., & Dugelay, J. (2016). Spatial and temporal variations of feature tracks for crowd behavior analysis. *Journal on Multimodal User Interfaces, 10*(4), 307-317.

Friedman, V. (2008). Data visualization and infographics. *Graphics, Monday Inspiration*, *14*, 2008.

Friendly, M., & Denis, D. J. (2001). Milestones in the history of thematic cartography, statistical graphics, and data visualization. *URL http://www. datavis. ca/milestones*, *32*.

Gidaris, S., & Komodakis, N. (2015). Object detection via a multi-region and semantic segmentation-aware U model. *Proceedings of the IEEE International Conference on Computer Vision, 2015,* 1134-1142.

Haritaoglu, Ismail. (2000). W4: Real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 22*(8), 809-831.

Heer, J. K., Card, S. A., & Landay, J. (2005). Prefuse: A toolkit for interactive information visualization. *CHI 2005: Technology, Safety, Community: Conference Proceedings - Conference on Human Factors in Computing Systems,* 421-430.

Hu, Y., Wu, W., & Zhou, Z. (2015). Video driven pedestrian visualization with characteristic appearances. *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*,183-186.

Hullman, J., Adar, E., & Shah, P. (2011). Benefitting InfoVis with Visual Difficulties. *Visualization and Computer Graphics, IEEE Transactions on*, *17*(12), 2213-2222.

Jain, A. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651-666.

Ji Soo Yi, Youn Ah Kang, Stasko, & Jacko. (2007). Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *Visualization and Computer Graphics, IEEE Transactions on, 13*(6), 1224-1231.

Jiansu Pu, Siyuan Liu, Ye Ding, Huamin Qu, & Ni. (2013). T-Watcher: A New Visual Analytic System for Effective Traffic Surveillance. *Mobile Data Management (MDM), 2013 IEEE 14th International Conference on, 1*, 127-136.

Kanade, T., Collins, R., Lipton, A., Fujiyoshi, H., & Duggins, D. (2000). A system for video surveillance and monitoring CMU VSAM final report final report. *STAR, 38*, STAR, 11 Aug. 2000, Vol.38.

Kandel, S., Heer, J., Plaisant, C., Kennedy, J., Van Ham, F., Riche, N., . . . Buono, P. (2011). Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization, 10*(4), 271-288.

Kraak, M. J. (2003, August). The space-time cube revisited from a geovisualization perspective. *Proc. 21st International Cartographic Conference* (pp. 1988-1996).

Kraak, M. J. (2005, July). Timelines, temporal resolution, temporal zoom and time geography. *Proceedings 22nd International Cartographic Conference, A Coruna Spain*.

Lipton, A., Fujiyoshi, H., & Patil, R. (1998). Moving target classification and tracking from real-time video. *Applications of Computer Vision, 1998. WACV '98. Proceedings., Fourth IEEE Workshop on*, 8-14.

Liu, Shixia, Cui, Weiwei, Wu, Yingcai, & Liu, Mengchen. (2014). A survey on information visualization: Recent advances and challenges. *The Visual Computer, 30*(12), 1373-1393.

Liu, Weibo, Wang, Zidong, Liu, Xiaohui, Zeng, Nianyin, Liu, Yurong, & Alsaadi, Fuad E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing, 234*, 11-26.

Luo, W., & Kim, T. K. (2013). Generic Object Crowd Tracking by Multi-Task Learning. In *BMVC*.

Mazza, R., & Dimitrova, V. (2004, May). Visualising student tracking data to support instructors in web-based distance education. In *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters* (pp. 154-161). ACM.

Mcardle, G., Demšar, U., Van Der Spek, S., & Mcloone, S. (2014). Classifying pedestrian movement behaviour from GPS trajectories using visualization and clustering. *Annals of GIS, 20*(2), 85-98.

Mckenna, Jabri, Duric, Rosenfeld, & Wechsler. (2000). Tracking Groups of People. *Computer Vision and Image Understanding, 80*(1), 42-56.

Meier, U., Stiefelhagen, R., Yang, J., & Waibel, A. (2000). Towards unrestricted lip reading. *International Journal Of Pattern Recognition And Artificial Intelligence, 14*(5), 571-585.

Meyer, D., Pösl, J., & Niemann, H. (1998, September). Gait Classification with HMMs for Trajectories of Body Parts Extracted by Mixture Densities. In *BMVC* (pp. 1-10).

Milan, A., Rezatofighi, S., Dick, A., Reid, I., & Schindler, K. (2016). Online Multi-Target Tracking Using Recurrent Neural Networks.

Ogawa, M., & Ma, K. (2010). Software evolution storylines. *Proceedings of the 5th International Symposium on Software Visualization,* 35-42.

Pang, Yuan, Li, & Pan. (2011). Efficient HOG human detection. *Signal Processing, 91*(4), 773-781.

Papageourgio, C. (1999). *A trainable system for object detection in images and video sequences* (Doctoral dissertation, PhD Thesis, Massachusetts Institute of Technology).

Qi, F., & Du, F. (2013). Trajectory data analyses for pedestrian space-time activity study. *Journal of Visualized Experiments : JoVE,* (72), E50130.

Santhiya, G., Sankaragomathi, K., Selvarani, S., & Kumar, A. (2014). Abnormal Crowd Tracking and motion analysis. *Advanced Communication Control and Computing Technologies (ICACCCT), 2014 International Conference on,* 1300-1304.

Schiele, B. (2006). Model-free tracking of cars and people based on color regions. *Image and Vision Computing, 24*(11), 1172-1178.

Selassie, D., Heller, B., & Heer, J. (2011). Divided Edge Bundling for Directional Network Data. *Visualization and Computer Graphics, IEEE Transactions on, 17*(12), 2354-2363.

Spaccapietra, Parent, Damiani, De Macedo, Porto, & Vangenot. (2008). A conceptual view on trajectories. *Data & Knowledge Engineering, 65*(1), 126-146.

Stauffer, C., & Grimson, W. (1999). Adaptive background mixture models for real-time tracking. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,2*, 246-252.

Steinberger, Markus. (2011). Context-Preserving Visual Links. *IEEE Transactions on Visualization & Computer Graphics, 17*(12), 2249-2259.

Takayanagi, H., Yamada, S., & Shibahara, H. (2015). A Study on Understanding Pedestrian Flow Using Intermittent Recording Images (PIRI). *Journal Of Asian Architecture And Building Engineering, 14*(3), 557-560.

Tang, S., Andriluka, M., Andres, B., & Schiele, B. (2017). Multiple People Tracking by Lifted Multicut and Person Re-identification. *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on,* 3701-3710.

Ware, C. (2004). *Information visualization perception for design* (2nd ed., Interactive Technologies). San Francisco, CA: Morgan Kaufman.

Watanabe, Y. Z. H. (2017). Trajectory Visualization of Ego-Motion Videos with Pedestrian Based on Monocular Visual Odometry and Machine Learning.

Weaver, C. (2004). Building highly-coordinated visualizations in improvise. *Proceedings - IEEE Symposium on Information Visualization, INFO VIS,* 159-166.

Weiming Hu, Tieniu Tan, Liang Wang, & Maybank. (2004). A survey on visual surveillance of object motion and behaviors. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 34*(3), 334-352.

Weiwei Cui, Shixia Liu, Li Tan, Conglei Shi, Yangqiu Song, Zekai Gao, . . . Xin Tong. (2011). TextFlow: Towards Better Understanding of Evolving Topics in Text. *Visualization and Computer Graphics, IEEE Transactions on, 17*(12), 2412-2421.

Xiang, Y., Alahi, A., & Savarese, S. (2015). *Learning to Track: Online Multi-object Tracking by Decision Making, 2015*, 4705-4713.

Yabushita, H., & Itoh, T. (2011). Summarization and Visualization of Pedestrian Tracking Data. *Information Visualisation (IV), 2011 15th International Conference on*, 537-542.

Yang, M., & Ahuja, N. (1998). Extraction and classification of visual motion patterns for hand gesture recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 892-897.

Yi, J., Kang, Y., Stasko, J., & Jacko, J. (2008). Understanding and characterizing insights: How do people gain insights using information visualization? *Proceedings of the 2008 Workshop on beyond Time and Errors*, 1-6.

Yilmaz, A., Javed, O., & Shah, M. (2006). Object tracking: A survey. *ACM Computing Surveys (CSUR)*,*38*(4), 13-Es.

Zhong, C., Wang, T., Zeng, W., & Müller Arisona, S. (2012). Spatiotemporal visualisation: A survey and outlook. *Communications in Computer and Information Science, 242*, 299-317.

VITA

Zheng Zhao

Career Objective

Software engineer, Machine learning (computer vision) engineer, Data visualization engineer, Front-end web engineer

Education

Purdue university *8/2016-present*

Master of Science in Computer Graphics Technology

GPA: 3.4/4.0

Coursework: Computer Graphic programming, Advanced Real-Time Graphic programming, Applications in visualization Analysis, Introduction to Scientific Visualization, Algorithms Design Analysis and Implementation, Artificial Intelligence, etc.

Shanghai Jiaotong University (UM-SJTU JI) *9/2011-8/2015*

Bachelor of Science in Electronic and Computer Engineering

GPA: 3.0/4.0

Dual Degree: Animation

GPA: 3.4/4.0

Coursework: Data Structures and Algorithms, Control Systems Analysis and Design, Introduction to Cryptography, Introduction to Operating Systems, Three-Dimension Animation Design, DV Making, etc.

Award: Innovation Project Certificate from Intel (Dec. 15, 2014)

Computer skills

C, C++, FPGA, Linux, HDL, OpenCL, OpenGL, ActionScript, Verilog, SCM C Language Programming, Cadence Schematic Editor, Python, Lua, Caffe, Torch, PHP, Javascript, CSS, OpenCV.

Publication

Sun, L., Huang, N., & Zhao, Z. 2015. Reliability Analysis for AFDX Network based on Dynamic Fault Tree. Accepted by ICMITE2015 (In English)

Wang, Q., Yu, S., Feng, S., Shan, H., Sun, J., & Zhao, Z. 2014. Analysis of Search Engine Transmission Effect on the Site Service Efficiencies of the Chinese Quality Inspection Industry. Published on E-Government (In Chinese)

Research Paper

On Coverage Issues in Partial Sensor Network *6/2014*

Classified the existing studies into two main categories: the *q*-percentage coverage and the probabilistic coverage

Work Experience & Internship

Purdue university

Master thesis *9/2017-present*

- Work on “PEDESTRIAN MOTION PATTERN AWARENESS WITH DATA VISUALIZATION：visualizing crowd’s motion flows”
- Combine the object detection project YOLO and tracking project SORT to collect the data of pedestiran.
- Visualize the tracking result on the website with PHP and D3.js.

Purdue university

Research assistant *9/2016-9/2017*

- Participated in a research program focusing on a hierarchical interactive visualization system of Purdue university career mapping with PHP, D3.js. (https://va.tech.purdue.edu/lilly/OldFiles06012017/career_mapping.php)
- Participated in a research program focusing on “Conceptual Model-based Problem Solving (COMPS): A Response to Intervention Program for Students with Learning Difficulties”, contributed to the web design and programming of this project.

National Information Center Beijing, China

Intern, Data Collection and Analysis *12/2013-2/2014*

- Employed a Big Data system, Government Dissector Professional, and analyzed the design issues by interpreting user hits on two government website pages in Guangdong Province
- Contributed to a technical paper published on E-Government

12/2015-4/2016

- Researched on some existing case of data visualization related to Three.js and D3.js (http://globe.cid.harvard.edu/?mode=gridSphere&id=SG .etc)

Capstone Design Project, Intel and UM-SJTU JI Shanghai, China

Product Owner *9-12/2014*

- Coded intensively on acceleration of real-time abstraction based on Intel GPU and OpenGL
- Employed OpenCL to condense video in real time, reducing its size by 50% and lessening the burden on CPU, achieved number of frames per second larger than 24 fps, and completed testing and adjustment on site

School of Reliability and Systems, Beihang University Beijing, China

Engineer *8-10/2014*

- Established a Dynamic Fault Tree to analyze network fault characteristics at a device-level
- Analyzed and modeled end system, switch system and the core subsystem in transmission of AFDX, besides fault of dual-network

Advanced Network Lab, Shanghai Jiaotong University Shanghai, China

Research Intern *2/2013-9/2014*

- Participated in a research program focusing on the algorithms of wireless sensor network
- Read extensively research papers in English and presented twice with papers of my choice at the weekly seminar
- Wrote a survey On Coverage Issues in Partial Sensor Network, summarizing relevant issues published in Chinese and English in recent years