A Linked Visualization of Trajectory and Flow Quantity to Support Analysis of People Flow

Aya Fukute*), Masaki Onishi**), Takayuki Itoh*)
*) Ochanomizu University **) National Institute of Advanced Industrial Science and Technology
{fuku, itot}@itolab.is.ocha.ac.jp, onishi@ni.aist.go.jp

Abstract

Thanks to the recent evolution of movie- and sensor-based human tracking technologies, we can obtain and accumulate a set of walking paths ("trajectories" in this paper) of people over a long period in various places. Such people flow datasets are useful for many fields, including analyses of customer behavior, effectiveness of advertisements, and operational efficiency. This paper presents a linked visualization system to assist in the discovery of new knowledge by analyzing the accumulated people flow datasets, and a case study using this system. In this study we suppose the people flow datasets consist of a set of trajectories and temporal flow quantity. The system consists of two visualization components: classified trajectory visualization, and temporal flow quantity visualization. The former component classifies trajectories into several patterns applying the spectral clustering algorithm, and visualizes the patterns by colors on a physical space. The latter component displays temporal flow quantity of the above patterns applying a piled polygonal chart. This paper introduces a case study applying a movie-based human tracking dataset to the presented system.

1 Introduction

Recent studies and technologies for global positioning system (GPS) have been widely applied to various services. Also, recent evolution and popularization of smartphone technologies have made GPS-based applications familiar. For example, various O2O (Online to Offline) services recommend offline information specified by users' positions. Such new business models have evolved thanks to the popularization of the mobile devices; however, it is generally difficult to obtain position information of people who do not bring such devices with them.

On the other hand, cameras and laser scanners have been also widely applied to measure the position information of general people. This approach has several more advantages. It does not require the registration of the users, and therefore it does not cause serious privacy issues. Also, its accuracy is much better than GPS-based technologies. Thanks to these advantages, camera-based position measurement technologies have been applied to various applications including security management and advertisement assessment. Human trajectory measurement [1] and its integration with computer simulations [2] are typical research topics in this field.

Camera-based human tracking systems are recently equipped at various places to store the long-term information of human trajectories. Such long-term information and its analysis bring fruitful knowledge for improvement of various services and applications. We believe visualization is useful for the assistance of the analysis process of such long-term human tracking information.

This paper proposes a technique for liked visualization of trajectories and temporal flow quantity extracted from the human tracking information. Here, trajectory analysis is necessary for the understanding of purposes and behavior of the measured people. We therefore applied a clustering algorithm to discover major patterns of the trajectories, and visualized the clusters. Also, we visualized the temporal flow quantity of the clusters as a polyline chart. We also implemented a mechanism to link the two views so that we can easily discover particular patterns of human behavior observed at particular times.

Following is the processing flow of the visualization of human tracking information by our technique:

a) Clustering and visualization of trajectories.

b) Aggregation of trajectories for each cluster, and for each period.

c) Visualization of temporal flow quantity of the clusters.

The step a) groups similarly shaped trajectories by applying the spectral clustering [3], and displays the set of trajectories applying the colors corresponding to the clusters. The step b) counts the number of trajectories for each cluster, and for each period, to visualize the temporal variation of the flow quantity. The step c) displays the temporal flow quantity by applying a famous time-varying data visualization technique ThemeRiver[4]. We believe that the knowledge discovered by this technique is useful for the new service innovation including marketing support and security management.

2 Related Work

This chapter introduces related work on visualization of people flow. Several techniques visualized trajectories and flow quantity in a single view, while others visualized them in separate views.

As typical techniques of the former approach, Andrienko et al. [6] and Yabushita et al. [7] presented a technique which approximates and quantizes trajectories from accumulated people flow datasets and summarizes similar trajectories. This approach makes easier to understand overall people flow by observing a single view, however, it is often difficult to represent the time-variation of flow quantity. McArdle et al. [5] presented a technique to represent similarity of trajectories in a space-time cube. This approach provides similarity analysis from both spatial and temporal aspects; however, it often causes 3D visualization specific cluttering problems while using space-time cubes.

Visualization by Quo et al. [8] and Onishi et al. [9] are typical techniques of the latter approach. The technique presented by Quo et al. extracts user-interested trajectories with a sketch interface to specify the entrance and exit of the people flow, and the displays their temporal flow quantities and speeds on another view. The technique presented by Onishi et al. clusters the measured trajectories and visualizes the temporal flow quantity of the clusters by a polyline chart.

The technique presented in this paper is different from the above techniques, since it visualizes temporal flow quantity of overall major patterns observed from the clustering results of trajectories.

3 Trajectory Acquisition from Movies

This study applies position information of humans captured by the technique presented by Onishi et al. [10]. This section defines the data structure of the human positions used in this study.

The technique reconstructs a 3D space from a parallactic image taken by a stereo camera, and projects the feature points of the image onto a plane. It then captures the positions of humans (x, y) on the plane. This paper describes that the number of the frames of the i-th person as m_i, his/her position at the t-th frame as $\boldsymbol{x}_t$=(x_t, y_t), and his/her trajectory as $\boldsymbol{P}_i$=$\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_t, \cdots, \boldsymbol{x}_{m_i}\}$. This paper calls m as the number of dimensions of the trajectory, and $\boldsymbol{P}$ as trajectory information.

4 Trajectory Grouping by Spectral Clustering

It is meaningful to group the trajectories based on their geometric similarity, to discuss the patterns and meanings of human behavior. We apply the spectral clustering [3] to group the trajectories into several major patterns.

Our implementation firstly unifies the number of positions of trajectories as M. If the number of positions of the i-th human m_i is larger than M, it samples M of the positions from his/her trajectory. If m_i is smaller than M, it adds the interpolated positions to make the number of positions k. This process is applied to the all trajectories as a preprocessing.

We applied the spectral clustering, which applies a nonlinear dimension reduction algorithm before applying the k-means method. It solves the problem of the original k-means method, that its clustering results of high dimensional points often depend on its initial settings. Here, this section defines the trajectory of the i-th person which have M positions as $\grave{\boldsymbol{P}}_i$=$\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M\}$. Also, this section supposes the dataset $\boldsymbol{X}_{spec}$=$\{\grave{\boldsymbol{P}}_1, \cdots, \grave{\boldsymbol{P}}_n\}$, which contains n trajectories, and they are to be grouped into k clusters. Processing flow of the spectral clustering is as follows.

Step1: Calculate the element a_{ij} of the similarity matrix $\boldsymbol{A}$ from the dataset $\boldsymbol{X}_{spec}$.

$$a_{ij} = \exp\left\{\frac{\left\|\grave{\boldsymbol{P}}_1 - \grave{\boldsymbol{P}}_n\right\|}{2\sigma^2}\right\} \text{ if } i \neq j \text{, or } a_{ii} = 0$$

Step2: Calculate the diagonal matrix $\boldsymbol{D}$, by calculating the element *i, i* as the sum of the i-th columns of $\boldsymbol{A}$, and the matrix $\boldsymbol{L}$ as $\boldsymbol{L}$=$\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$.

Step3: Calculate the eigenvectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_k$ by applying the eigenvalue decomposition to $\boldsymbol{L}$.

Step4: Construct the matrix $\boldsymbol{Q}$= $[\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_k]$ from the eigenvectors, divide the rows of $\boldsymbol{Q}$ into vectors, and apply the k-means method to the vectors.

Step5: Assign $\grave{\boldsymbol{P}}_i$ to the cluster α, if the clustering result assigns the i-vector of $\boldsymbol{Q}$ to α.

Here, σ^2 is an automatically configured parameter which controls the distance between two trajectories $\grave{\boldsymbol{P}}_i$ and $\grave{\boldsymbol{P}}_j$. The algorithm maintains consistency of similarity distances among the trajectories after the dimension reduction by adequately controlling this parameter.

Figure 1 shows an example dataset of human trajectories entering to or exiting from a building, and its clustering result applying the k-means method and the spectral clustering. Figure 1(a) is an photograph of the captured space. Figure 1(b) shows all the trajectories. It is quite difficult to understand the typical patterns and behaviors from this figure, because trajectories tangle each other. Figure 1(c) shows a clustering result by the k-means method, where colors of trajectories denote clusters. It denotes an inadequate clustering result that much different trajectories into the a single cluster drawn in blue. On the other hand, Figure 1(d) shows a clustering result by the spectral clustering. It denotes that the differently shaped trajectories which are grouped into the same cluster by the *k*-means method are adequately grouped into different clusters. We expect the spectral clustering effectively groups the similar trajectories into typical patterns and behaviors.

5 Visualization of Flow Quantity

After the clustering process described in the previous section, our implementation calculates the number of trajectories in the clusters in each period of the time. We visualize time variation of the flow quantity of the clusters by applying ThemeRiver[4], a famous time-varying data visualization technique. ThemeRiver represents the time variation of multiple values like a river, by assigning time to the horizontal axis, values to the vertical axis, and painting each element as colors. Havre et al. [4] demonstrated the effectiveness of ThemeRiver by visualizing the time variation of frequency of topics in the newspapers.

Our implementation assigns clusters to the colors, and flow quantity to the vertical axis. It effectively represents the time variation of the total and cluster-by-cluster flow quantity. We think this visualization is useful for the understanding and improvement of the flow quantity.

6 Case study

This section introduces three cases visualizing the people flow datasets which we obtained by using a stereo camera located near escalators of a complex facility in Akihabara, Tokyo. Figure 4(a) shows the sight of the place. An escalator in the red square goes up to the third floor, and another escalator in the blue square goes up to the fourth floor. Figure 4(b) illustrates the all trajectories of a day at the place.

6.1 Visualization in a short period

As the first case study, we visualized and analyzed a people flow dataset between 9AM and 11PM on Saturday. We classified all trajectories into eight major patterns by applying the spectral clustering, while specifying the number of clusters experimentally. Figure 4(c) shows an example of visualization of the classified trajectories. Pink trajectories denotes some people go up to the fourth floor from the second and under floor. Purple and blue trajectories denote some people go up to the fourth floor from the third floor. Red and yellow trajectories denote some people go up to the third floor from the second floor. Light blue and orange trajectories denote some people go through near the escalators. Green trajectories denote some people walk around the escalators.

To obtain the above result, we firstly observed major trajectory patterns by applying the spectral clustering algorithms to all the trajectories and displaying the trajectories in the colors assigned to each cluster. We could deeply discuss the people flow by observing the visualization result. We visualized temporal flow quantities of each major pattern of movements by applying ThemeRiver as shown in Figure 2. The horizontal axis denotes time, and the vertical axis denotes the number of pedestrians. Colors correspond to clusters of trajectories shown in Figure 4(c). It shows that the flow quantity increased during 1PM to 6PM. We also observed time-variation of the flow quantity of each cluster, and found the flow quantity of red and yellow clusters quickly increased at that time. These clusters denote the transfer from the second or lower floors to the third floor. Here, we supposed many of pedestrians visited the second and third floor to eat lunch or dinner, because restaurant were open on these floors. We also supposed these pedestrians were not employees working in this facility, but ordinary people, because pedestrians going up to the second floor increased even though the fifth and upper floor had offices. Moreover, we supposed most of the pedestrians did not go back to the office floors, because the blue, pink and purple trajectories almost disappeared at this time.

6.2 Comparison of different periods at the same place

This system visualizes flow quantity of different periods side by side so that we can compare the flow quantity.

As the second case, we compared two datasets of different periods at the same place from 9AM to 11PM. We mixed the two datasets together, and classified trajectories into eight major patterns by applying the spectral clustering. We compared temporal flow quantities on Saturday and Tuesday, arranging ThemeRiver lengthwise. Figure 3(Upper) shows the visualization of the temporal flow quantity on Saturday, and Figure 3(Lower) shows the visualization of the temporal flow quantity on Tuesday. The figures denote that the number of pedestrians in the lunch time was larger on Tuesday than Saturday. We also observed the temporal flow quantity of each cluster at that time, and found the number of pedestrians in the pink, blue, and purple clusters especially increased. From this result, we supposed many people who worked on weekdays at the offices on the fifth or upper floors used this facility. Moreover, the peaks of the number of pedestrians around lunch time were different between the two datasets: the number moderately increased around 1PM on Saturday, while the number quickly increased around 0PM on Tuesday. We found that the peak time on Tuesday was shorter, because it seemed many of the pedestrians were employees of this facility on weekdays. On the other hand, the peak time on Saturday was longer, because it seemed many of pedestrians were ordinary people who could eat lunch anytime at this facility. Furthermore, variation of flow quantity on Saturday was totally gentler than on Tuesday. We supposed that many pedestrians on the holidays were ordinary people who can visit there various time, while many pedestrians on the weekdays were office employees who regularly walk around there in the attendance, lunch, and leaving times.

6.3 Visualization in a long period

As the third case, we visualized a datasets in a half year. We extracted measured trajectories on every Friday and Saturday during November 2010 to April 2011, and also on every Monday and Saturday during January 2012 to July 2012. We classified trajectories into five major patterns by applying the spectral clustering, and then visualized the temporal flow quantity by using ThemeRiver, as shown in Figure 5. The red and green clusters denote pedestrians going up from the second floor to the third floor. The blue cluster denotes pedestrians going up from the second floor to the fourth floor. The light blue and yellow clusters denote pedestrians going up from the third floor to the fourth floor.

We firstly observed the dataset of Friday and Saturday during November 2010 to April 2011. We found that the total flow quantity increased on March 11, and decreased on December 31, February 11, and March 18, as shown in Figure 5(a). We focused on the temporal flow quantity of each cluster on these days. On December 31 and February 11, flow quantities of the red and green clusters (which denote going up from the second floor to the third floor) did not drastically vary compared with the other Friday, while flow quantities of the blue, light blue, and yellow clusters (which denote going up to the fourth floor) decreased largely. We supposed less number of pedestrians clustered into the patterns of employees were observed on these days because they were holidays.

Next, we observed why the flow quantity increased on March 11 and decreased on March 18. Here, March 11 was the day of the Tohoku earthquake. Many transportation did not work, and therefore 3.5 million or more people could not go home. Various facilities including opened to host these people. We supposed that the flow quantity increased on that day because the people who could not go home stayed at this facility. Some of the people did not visit the restaurants and went to the floors which opened for the people, because the flow quantity of the blue cluster especially increased. Similarly, we supposed that the earthquake caused decrease in the number of pedestrians on March 18, the next Friday of the Tohoku earthquake. We found the number of visitors from outside the facility decreased because the flow quantity of the red cluster decreased. Meanwhile, Figure 5(b) shows the flow quantity on March 12, which denotes the influence of this earthquake. The flow quantity of total and each cluster did not vary usually; however, the flow quantity of all the clusters drastically decreased on March 12. It was conceivable that many people kept from going out because many media reported various topics related to earthquake like tsunami, fire, accident of nuclear plants, and radioactivity. Also, we discovered the flow quantity quickly recovered next week (on March 19).

As introduced above, we discovered the temporal variation of total flow quantity from two visualization results, and then observed the flow quantities of each cluster to find the characteristic of the days. The visualization technique presented in this paper is useful for such tasks of human behavior analysis.

Next, we observed trajectories on each Monday and Saturday during January to July 2012. Figure 5(c) shows the temporal flow quantity on each Monday. We found that the flow quantity of the blue, light blue, and yellow clusters decreased, as much as the flow quantity of the red and green clusters increased. Therefore, the total flow quantity did not drastically vary on January 2, January 9, and April 30. These days were holidays: January 2 was during new year vacation, January 9 was the coming-of-age ceremony day, and April 30 was during the golden week vacation. It was conceivable that the employees decreased on these days, meanwhile ordinary people who visited restaurants increased because of holiday.

Figure 5(d) shows the temporal flow quantity on each Saturday. We found the total flow quantity increased on May 5 and Jun 23. May 5 was during the golden week vacation, and therefore it was conceivable the number of sightseers increased. In contrast, Jun 23 was not a holiday even though the total flow quantity increased. We observed the flow quantity on that day, and found that the flow quantities of the green and yellow clusters increased. There was an event at an exhibition space located on the fourth floor of the facility, and therefore we conjectured that people who visited the exhibition space were observed when they visit restaurants.

7 Conclusion

This paper proposed a technique for linked visualization of trajectory and flow quantity to support the analysis of people flow. The technique classifies trajectories into several patterns by applying the spectral clustering, so that we can clear and analyze the major moving patterns of pedestrians. It draws the trajectories in the real space while coloring based on the clustering results. It also displays the temporal flow quantity of each cluster in another view, applying ThemeRiver. We can analyze temporal flow quantity of total and each cluster simultaneously by observing the ThemeRiver, and then discover the new knowledge regarding the people flow, such as typical patterns which often appear at particular periods. Case studies in this paper demonstrated this visualization technique is available for both short-term and long-term datasets. Also, it is useful to compare people flow datasets of different periods. We expect the new knowledge discovered by using the visualization technique can be effectual keys for new service innovation.

Following are our potential future issues:

Classify trajectories at freely walking spaces. People can walk to various directions at open or clear spaces like parks. Pedestrians may hover around there without particular purposes, and therefore it may be hard to classify trajectories into meaningful patterns. We would like to explore better ways to extract meaningful patterns from such datasets, by applying other clustering schemes and further techniques such as noise analysis.

Visualize representative trajectories. One of our target in this study is visualizing major trajectory patterns. Current visualization results are complicated because we just draw all the trajectories in the colors corresponding to clusters. We would like to ease the complexity of the visualization results by only drawing representative trajectories of the clusters.

Visualize directions of the trajectories. Our current visualization does not represent any directions of the trajectories. We would like to improve the trajectory drawing component so that we can understand their directions from the visualization results.

References

[1] J. Zhou, and J. Hoang: Real Time Robust Human Detection and Tracking System, IEEE Computer Vision and Pattern Recognition(CVPR), no.3, pp.149 (2005)

[2] K. Yamada, T. Abe, and H. Kimura: Simulation of Planned/Unplanned Purchaser Flow using an Agent Model, Annual Conference of Japanese Society for Artificial Intelligence (JSAI), vol.19, pp.3E3-02 (2005)

[3] A. Ng, M. Jordan, and Y. Weiss: On spectral clustering: Analysis and an algorithm, Advances in Neural Information Processing Systems, pp. 849-856 (2001)

[4] S. Havre, B. Hetzler and L. Nowell: ThemeRiver: Visualizing Theme Changes over Time, IEEE Symposium on Information Visualization, pp. 115-123 (2000)

[5] G. McArdle and U. Demsar: Investigating Similarity of Trajectories through Physical Decomposition of Movement in a Space-Time Cube, 14th AGILE International Conference on Geographic Information Science (2011).

[6] G. Andrienko and N. Andrienko: Spatio-temporal Aggregation for Visual Analysis of Movements, IEEE Symposium on Visual Analytics Science and Technology 2008, pp. 51-58 (2008)

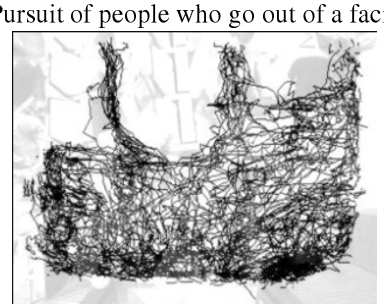
[7] H. Yabushita and T. Itoh: Summarization and Visualization of Pedestrian Tracking Data, 15th International Conference on Infromation Visualization (IV2011), pp. 537-542 (2011)

[8] H. Guo, Z. Wang, B. Yu, H. Zhao and X. Yuan: TripVista: Triple Perspective Visual Trajec- tory Analytics and Its Application on Micro- scopic Traffic Data at a Road Intersection, IEEE Pacific Visualization Symposium, 163- 170 (2011)

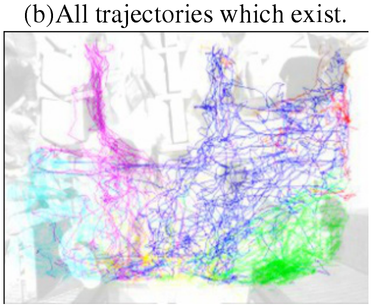
[9] M. Onishi and I. Yoda: Comparison and Visualization Method of Pedestrian Trajectories in a Complex Office for a Long Period, The Transactions of the Institute of Electronics, Information and Communication Engineers D, vol.J90-D, no.4, pp.486-493 (2010)

[10] M. Onishi and Y. Yoda: Dynamic Trajectory Extractio0n from Stereo Vision using Fuzzy Clustering, IEEJ, vol.128, no.9, pp.1438–1446 (2008)

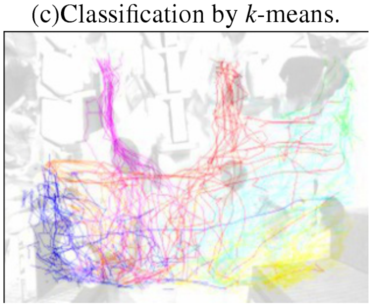
(a)Pursuit of people who go out of a facility.



(b)All trajectories which exist.



(c)Classification by *k*-means.



(d)Classification by Spectral Clustering.

Figure 1: Examples of Classification of trajectories.

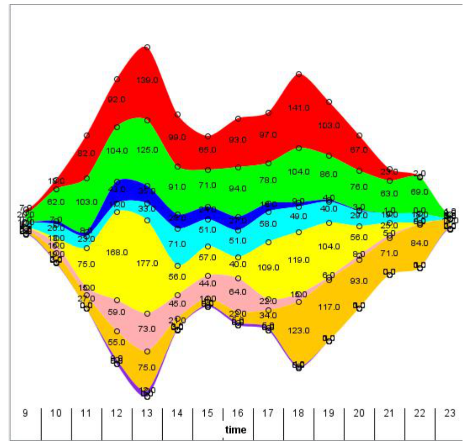


Figure 2: ThemeRiver of the flow quantity on Saturday.

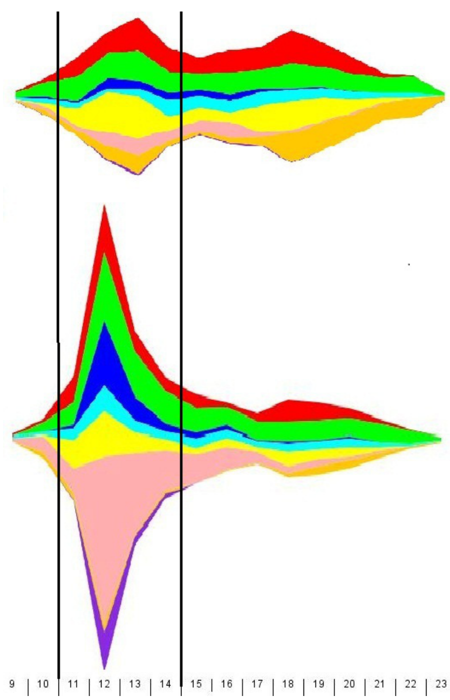
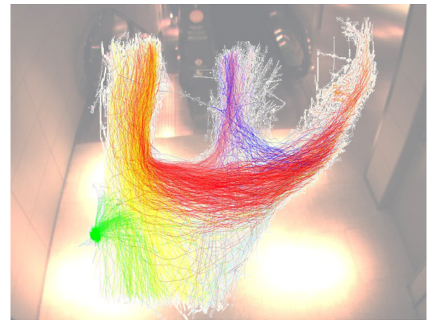


Figure 3: Comparison of temporal flow quantity of two datasets (Upper: on Saturday, Lower: on Tuesday).

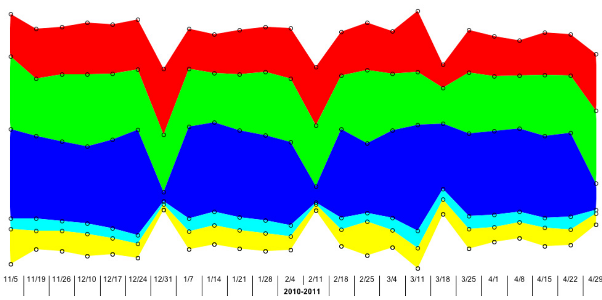
(a) The sight of the place where we obtained data.

(b) All trajectories which exist at the place where we obtained data.

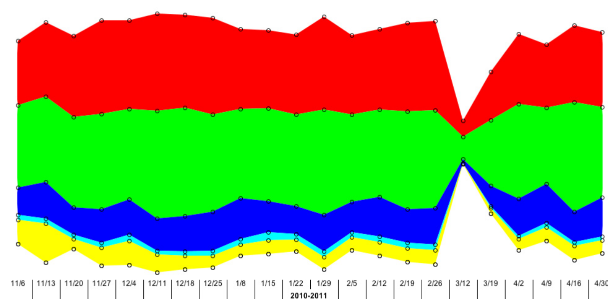


(c) Classification of trajectories by Spectral Clustering.

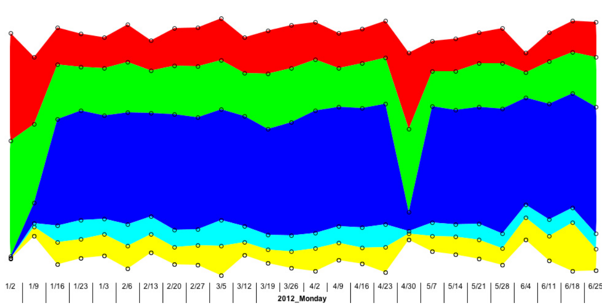
Figure 4: Examples of trajectories.



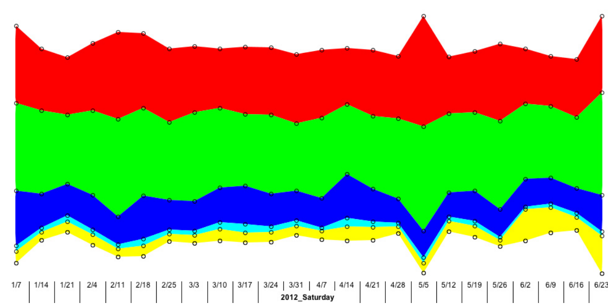
(a) ThemeRiver which expresses the flow quantity at Friday from November 2010 to April 2011.



(b) ThemeRiver which expresses the flow quantity at Saturday from November 2010 to April 2011.



(c) ThemeRiver which expresses the flow quantity at Monday from January to July 2012.



(d) ThemeRiver expresses the flow quantity at Saturday from January to July 2012.

Figure 5: Examples of ThemeRiver-based representation of flow quantity.