2014 18th International Conference on Information Visualisation

Directional Aggregate Visualization of Large Scale Movement Data

Yuuki Hyougo
Department of Computer Science,
University of Tsukuba
Ibaraki, Japan
hyougo@iplab.cs.tsukuba.ac.jp

Kazuo Misue
Faculty of Engineering,
Information and Systems,
University of Tsukuba
Ibaraki, Japan
misue@cs.tsukuba.ac.jp

Jiro Tanaka
Faculty of Engineering,
Information and Systems,
University of Tsukuba
Ibaraki, Japan
jiro@cs.tsukuba.ac.jp

***Abstract*—Widespread use of GPS terminals has made it possible to collect geospatial movement data, and visualization is an effective method to understand such data. However, for large scale movement analysis, because the data set includes complex movements of individuals, it is difficult to understand using naive visualization methods. In order to solve this problem, we developed a visualization technique that can represent large scale movement data in an aggregate manner. This visualization technique has two representations: "amoeba representation" and "amoeba colony representation." Amoeba representation represents the distance moved from a point in any direction with map scale in a geographical space, and amoeba colony representation represents movements over a wide geographical space.**

***Keywords*-Movement data, origin and destination, flow visualization, people flow, movement analysis, aggregation.**

I. Introduction

Because of the widespread use of GPS terminals, it has become possible to collect large scale movement data. Analysis of such data brings us new knowledge. With this analysis, information visualization helps the understanding of large scale data effectively.

Drawing lines that connect the positions of each object in chronological order is one naive visualization of movement data. This method can represent all of the information that the movement data contains. However, large scale movement data causes visual clutter, hence it is difficult to represent this information in a suitable manner. Thus, in order to visualize large scale data, an aggregate visualization that depicts one marker per thousand data records is an efficient method. There are currently some methods that attempt to represent movement data in an aggregate manner. However, these aggregate methods do not represent the distribution of movement distances in every direction very well.

We have developed a new aggregate visualization method for the movement of objects such as people or goods. Our method can represent the distribution of movement distance in every direction. We call our method an "amoeba representation" because of its shape. Amoeba representation makes it possible to observe a number of objects that have moved from an origin to a destination.

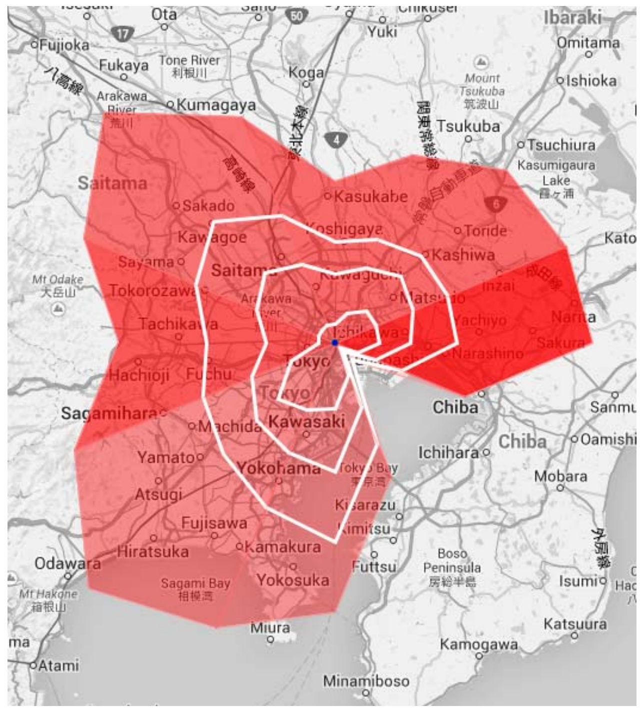


Figure 1: Visualization method in an aggregate manner (amoeba representation). This diagram represents the flow of 302 people over three hours from Akihabara station.

This paper introduces the amoeba representation (see Figure 1) and its extension. In Section II, we review related work. In Section III, we detail the problem of visualization of large scale movement data and formalize movement data. In Section IV, we explain our aggregate method for movement data. We also explain some design choices for amoeba representations. In Section V, we present the practicality of our visual representation with cases studies. Finally, we conclude with some remarks in Section VI.

II. Related Work

To visualize the flow of air and water, two-dimensional vector field visualization is generally used. This method can

1550-6037/14 $31.00 © 2014 IEEE
DOI 10.1109/IV.2014.34

also be applied to visualize movement of people and goods. For large scale geospatial data, several visualization methods in an aggregate manner have been proposed.

A. 2D Vector Field Visualization

One of the simplest representations of vector field flow is to draw arrows in the cells of a mesh. For the arrangement of these arrows, Field et al. [3] demonstrated that recognition of the flow is facilitated when arrows are continuous. Turk et al. [10] represented the flow field effectively by using a line called a "Streamline" instead of an arrow. Laidlaw et al. [4] performed a user study on the arrangement of symbols representing flow and showed advantages and disadvantages to each approach. Methods that visualize two-dimensional vector field flow only represent the movement in one direction at one point on a plane. Because of this, it is difficult to represent a plurality of objects that move various distances and directions from one point.

B. Geospatial Movement Visualization

Tobler [8] discussed computer aided visualization methods that represent large scale human migration. He showed that it is possible to represent numbers of people moving point-to-point by both continuous and discrete representations. The continuous representations that he proposed used two-dimensional vector field visualization methods such as Streamline, and the discrete representations used symbols that represent flow. Tobler suggested that it is possible to represent multiple endpoints of an object that moves away from a point. Phan et al. [5] also introduced a visual representation to represent a number of people who move point-to-point, called a flow map. A flow map can represent numbers of moving people by lines that have no intersections. The visualizations by Tobler [8] and Phan et al. [5] used table data that listed the numbers of people who moved between predefined places. Andrienko et al. [2] proposed a method that represented the numbers of moving objects between non-predefined places. These visualization methods focused on numbers of people moving between two points. As a consequence, it is impossible to represent the distance moved for each object.

Andrienko et al. [1] described the existing aggregate visualization methods that can be classified according to aggregation facets: time (T), space (S), and attributes (A). For example, aggregate visualization by Tobler is an S×S×T×T-aggregation, indicating an aggregate defined by start place (origin), end place (destination), start time, and end time. In addition, Andrienko et al. introduced two aggregate methods: S×T×D-aggregation and R×S×S×T×T-aggregation. The S×T×D-aggregation (where D indicates direction) is a method that can represent traffic volume at a point. The R×S×S×T×T-aggregation (where R indicates route) is a method that can represent the movement paths of objects. In this paper, we propose an S×T×T×D-aggregation that represents the distribution of movement distances in every direction from a start point.

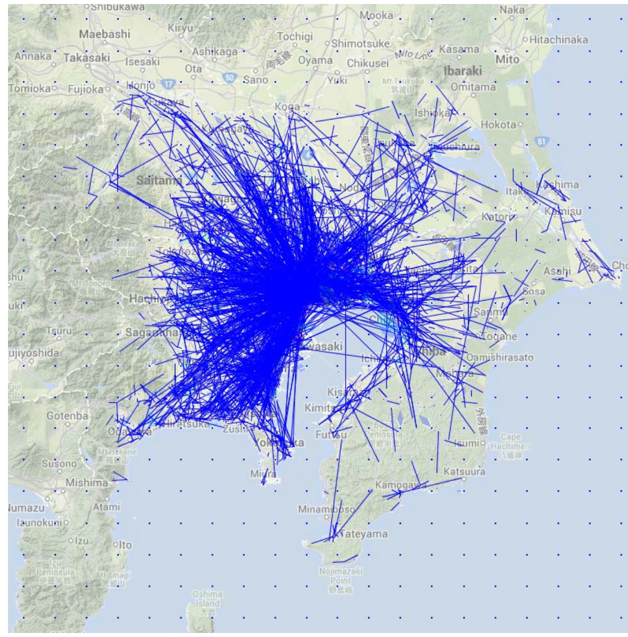


Figure 2: Movement of 72000 people in the Tokyo metropolitan area. Each blue line connects the start and end points of a movement. Visual clutter in the center area makes understanding the information difficult.

III. Data and Visualization Issues

We focus on large scale movement data in this paper. We first define some notations and then describe some issues that should be taken into consideration when visualizing such data.

A. Movement Data

We represent a set of moving objects by $O = \{o_1, \ldots, o_n\}$. In our study, n is on the order of tens of thousands or more. We also represent the position of an object $o \in O$ at time t by $p_o(t)$. This position is on the plane, that is, $p_o(t) \in R^2$. We focus on the movement of the elements of O from time s to time t, when the positions of all elements of O at time s and t ($s < t$) are given. We refer to position $p_o(s)$ as the start point and to position $p_o(t)$ as the end point. In practice, the position of the object is a point on the earth, this is a necessary consideration for the implementation.

B. Visualization of Large Scale Data

Shneiderman [6] classified visualization methods of large scale data into three categories: atomic, aggregate, and density plot visualizations. Atomic visualizations use a graphical representation for one data record. For example, a visual representation that draw lines connecting the start and end points of each object included in movement data is one atomic visualization. Figure 2 shows an example atomic visualization on actual data. This figure represents the movement of 72,000 people over three hours in the Tokyo metropolitan area. As seen in the figure, a significant number of many blue lines overlap each other. This visual clutter hide many short movements that actually exist. One of the solutions to this problem is to lower the opacity of the lines. However, we might miss lines in sparse areas.

Aggregate visualizations are an effective way to reduce visual clutter that occur in atomic visualizations. However, it is necessary to design an aggregation that is suitable for its purpose, considering the facets and size of the aggregation.

IV. Directional Aggregate

We developed an aggregate visualization method that aggregates movement data based on movement starting point, start time, end time, and direction. Our method makes it possible to quantitatively grasp movement data without visual clutter.

A. Basics of Aggregation

We determine the values to be shown in the aggregate method by dividing the moved objects as follows.

First, we determine the subset of objects included in the movement data by start point. When a position $q \in R^2$ is given, objects in vicinity of the position q at time s are the subjects of representation. In other words, the set of objects O_q that represent the subjects is defined as the following.

$$O_q = \{o \in O | \ ||p_o(s) - q|| \leq \epsilon\} \tag{1}$$

Here, ϵ is a small value that represents the neighborhood of q. Subsequently, we divide set O_q into subsets by direction of movement. We express the object direction $o \in O_q$ by $a(o)$. Here, $a(o)$ is the angle θ of the end point of object o when the start point is placed at the origin of the two-dimensional polar coordinate system (r, θ). The set O_q of the objects is divided into k sets using $a(o)$. We express the boundary angles θ_i as follows.

$$\theta_0 < \theta_1 < \cdots < \theta_k = \theta_0 + 2\pi \tag{2}$$

$$\theta_{i+1} = \theta_i + \frac{2\pi}{k} \tag{3}$$

Moreover, we represent the set $O_{q,i}$ of objects moving in the $[\theta_i, \theta_{i+1})$ direction as follows:

$$O_{q,i} = \{o \in O_q | \theta_i \leq a(o) < \theta_{i+1}\} \tag{4}$$

We refer to the number of these objects as $m_i = |O_{q,i}|$. In order to know the state of movement in each direction, we calculate the distribution of the distance moved for $O_{q,i}$. We express the distance moved of an object $l(o)$ as follows.

$$l(o) = \|p_o(t) - p_o(s)\| \tag{5}$$

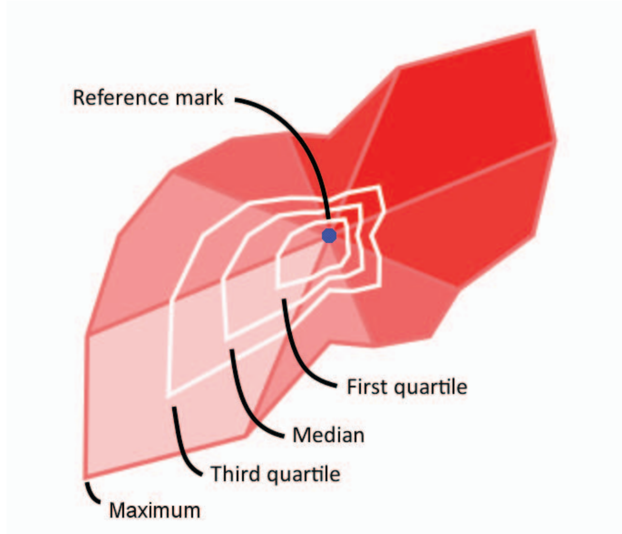


Figure 3: Overview of the amoeba representation.

We decided to use the quartile as a metric of the distances moved by the objects. We sort the objects in O by distance moved in the following manner:

$$l(o_{i,1}) \leq l(o_{i,2}) \leq \cdots \leq l(o_{i,m}) \tag{6}$$

The quartiles of the distances moved in O are obtained as follows.

First quartile	$d_{i,1} = l(o_{i,\lceil m/4 \rceil})$
Median	$d_{i,2} = l(o_{i,\lceil 2m/4 \rceil})$
Third quartile	$d_{i,3} = l(o_{i,\lceil 3m/4 \rceil})$
Maximum	$d_{i,4} = l(o_{i,m})$

B. Amoeba Representation

To represent the state of movement and place the start point near its corresponding point in geospatial space, we developed a visual representation overlaid on a map. We call it an "amoeba representation" because of its shape.

1) Amoeba Visual Representation: An amoeba representation visually represents the number of moving objects and the quartiles of the distribution of distance moved by the visual variables of surfaces and positions of lines, respectively. Once the point q in geographical space, neighborhood ϵ, division number k, and division angles $\theta_0, \ldots, \theta_{k-1}$ have been determined, the number of moving objects is $m_i = |O_{q,i}| (i = 0, \ldots, k-1)$ and the quartiles of the movement distance distribution are $d_{i,j} (j = 1, \ldots, 4)$. We can observe the movement situation for any geographical location by changing q.

For example, the amoeba representation in Figure 3 is made up of a reference point located in the center of the shape, annular lines surrounding the reference point, and

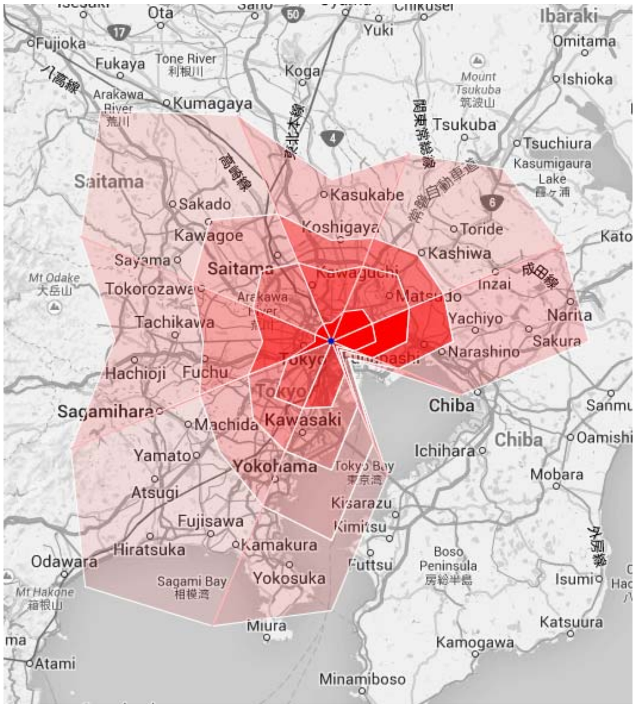


Figure 4: Amoeba representation that expresses the number of moving objects per area by opacity.

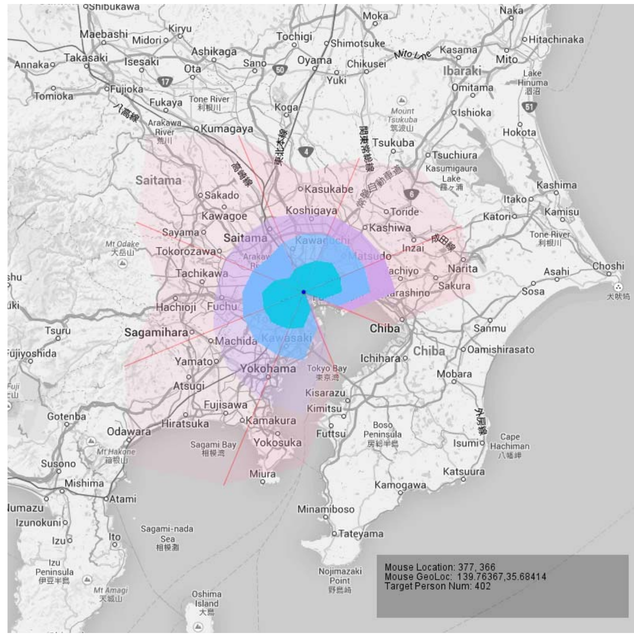


Figure 5: Amoeba representation where it is possible to grasp the quartile positions by the hue of the surface. Opacity indicates the number of objects that move per unit area.

four surfaces. The lines farthest from the reference point become the elements that make up the surface. We set $k = 8$ in this and the following figures. These shapes show the objects positioned at the start point in neighborhood ϵ of point q where the reference point is located on the background map. The surface surrounding a reference point is divided into k sections by the division angle θ_i around the reference point, and its opacity represents the number m_i of objects that move in that direction $i = 0, \ldots, k-1$. If the number of moving objects m_i is large, the opacity of the surface is high. However, if the number of moving objects m_i is small, its opacity is low.

2) Elimination of the Effect of Area: In an amoeba representation, the area of the shape may affect the observation in the same manner as a choropleth map. This is because the size of the surface that does not represent the quantity can affect the read. For example, when two faces are colored in the same way, it could appear that the larger surface area is associated with that quantity.

We designed a visual representation that represents the number of moving objects per unit area taking into account the possible misreading of quantity by area. We represent the number of moving objects per unit area of the region surrounded by quartiles $d_{i,n}$ and $d_{i,n+1}$ of direction i by $V_{i,n}$. If $S_{i,n}$ is the area of the region, $V_{i,n}$ is calculated by the following equation.

$$V_{i,n} = \frac{|O_{q,i}|}{4} \cdot \frac{1}{S_{i,n}} \tag{7}$$

This $V_{i,n}$ is represented by the visual attributes of the surface. Figure 4 is a visual representation that shows the number of moving objects per area with the same data as Figure 1 using surface opacity. In Figure 1, it appears that many of the objects moved too far towards the east. In contrast, in Figure 4, where the impact of area is considered, the objects appear to remain close to the reference point. In fact, among the objects that moved in that direction, the number of objects that moved away from the third quartile is one quarter of the total number. From this, we can say that Figure 4 expresses movement data more intuitively than Figure 1.

3) Improvement for Visibility of Quartile Position: To locate the quartiles that represent the distance distribution in each direction, we devised a method that can be grasped using the visual attributes of the surface rather than lines. The method described above indicates the quartile positions of the distance distribution in each direction by four lines. However, their visibility was not high because the lines on the shape are thin. Therefore, we devised a way to represent the position of the quartiles with visual variables on the surface rather than by lines, using the hue of the surface. We chose colors that look equally bright to human perception from the L*a*b* color space[7]. Accordingly, this separates

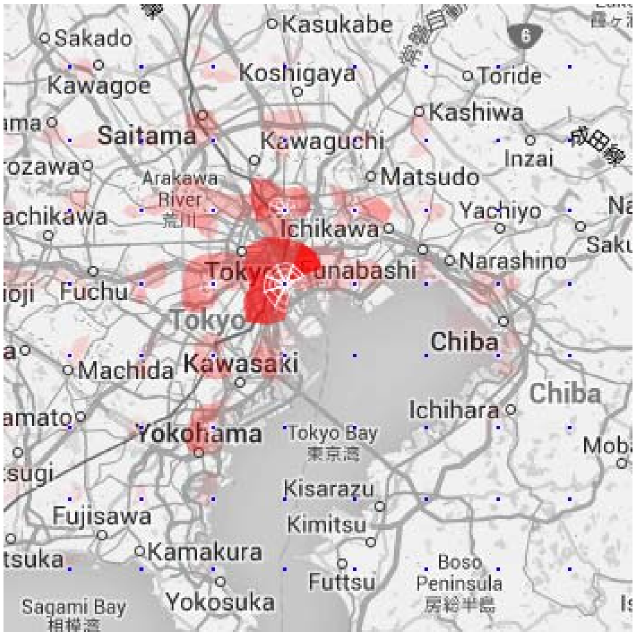


Figure 6: Overview of an amoeba colony representation.

the effect of hue from the representation of the value. We show in Figure 5 how this process is applied to the amoeba representation. In this figure, we determine the opacity using the method in Section IV-B2 as well as changing the hue of the surface. From the results, we were able to determine that the use of surface hue for grasping the quartile positions is effective. However, if opacity is low, knowing the position of the quartile is difficult. To enhance visibility even more, we should consider the hues to be used.

C. Amoeba Colony Representation

The amoeba representation shows the moving objects that depart from the vicinity of a location located on a map. We developed a representation method that is capable of representing the movement from multiple points that cannot be expressed by one amoeba representation. This representation expresses the amoeba representation using a "small multiples technique" [9] that can compare the movement from a multiple points. We call this set of amoeba representations an "amoeba colony representation."

An amoeba colony representation draws the quartile $d_{i,j}$ smaller than the scale of the map for each amoeba representation. This is because the line that expresses the quartile position overlaps the ones from other neighborhoods when the amoeba representation has quartiles $d_{i,j}$ that are drawn to the map's scale.

The amoeba colony representation makes it possible to compare the state of movement of each point because it is arranged in the same way as an amoeba representation. We show an example amoeba colony representation in Figure 6 that uses the same visual representation to as Figure 1 for each amoeba representation.

V. Case Study

We visualized large scale movement data with two methods: atomic and aggregate visualization.

In Figure 7, we show visualizations that represent the movement over three hours of 9,007 people whose starting point was within a 3 km radius of Kawasaki station in the Tokyo metropolitan area. Figure 7a represents moving people by drawing straight blue lines that connect the start and end points. In this figure, it is difficult to observe objects that move within a short distance because of the overlapping of other lines. Moreover, it seems to mainly indicate long distances.

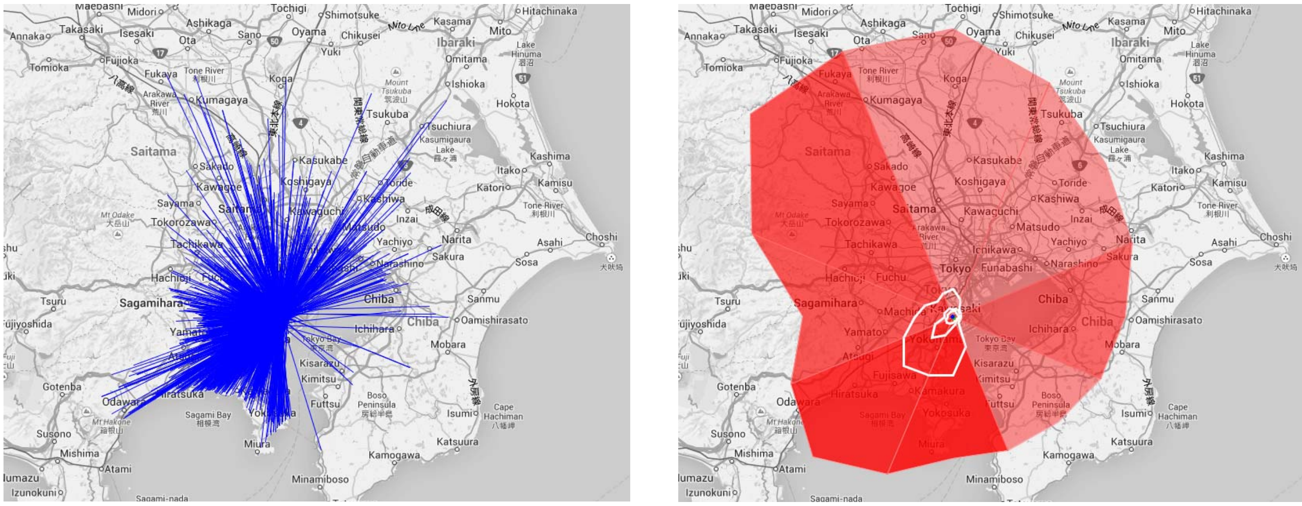
On the other hand, Figure 7b is obtained by applying the amoeba representation to the same movement data. From the figure, it can be seen that movement in the south and southwest directions happens frequently because of opacity of the surface. In addition, the people appear to remain near the reference point. We can observe that there is a longer distance travelled compared to other orientations. This type of information cannot be observed in Figure 7a, but can be observed from an amoeba representation such as the one in Figure 7b.

It is also possible to obtain useful information from the amoeba colony representation by extending the amoeba representation. In Figure 6, we found that the number of people who moved from the Tokyo station was large because the opacity of the surface of center amoeba representation was high. In addition, there were a large number of people moving to the west from Tokyo station. This is a reasonable result, considering the population of the Tokyo metropolitan area.

VI. Conclusion

We developed an aggregate visualization method to express the large scale movement data of people, goods, etc. In the visualization of large scale movement data, it is difficult to understand this information using an atomic visualization because of visual clutter.

The method we developed represents large scale movement data using an aggregate visualization method, thus the method makes it possible to express the number, direction, and distance of objects moving from a point in geographical space while suppressing visual clutter. The method aggregates large scale movement data in terms of starting point, start time, end time, and direction, and indicates quartiles of movement distance in the representation. In addition we proposed solutions to the visual problems of our representation. This method is intended for objects with an arbitrary starting point in geographical space, but can also be used for an arbitrary ending point. Furthermore, by using



(a) Atomic visualization

(b) Aggregate visualization (Amoeba representation)

Figure 7: Visualizing objects moving from Kawasaki station.

several instances of the developed method, it is possible to also express movement data over a wide geographic space simultaneously.

ACKNOWLEDGMENTS

People flow data was provided by Center for Spatial Information Science, University of Tokyo.

REFERENCES

[1] Gennady Andrienko and Natalia Andrienko. Spatio-temporal aggregation for visual analysis of movements. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology 2008 (VAST'08)*, pages 51–58, 2008.

[2] Natalia Andrienko and Gennady Andrienko. Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics*, 17(2):205–219, 2011.

[3] David J. Field, Anthony Hayes, and Robert F. Hess. Contour integration by the human visual system: Evidence for a local gassociation fieldh. *Vision research*, 33(2):173–193, 1993.

[4] David H. Laidlaw, J. Scott Davidson, Timothy S. Miller, Marco da Silva, R.M. Kirby, William H. Warren, and Michael Tarr. Quantitative comparative evaluation of 2D vector field visualization methods. In *Proceedings of the Conference on Visualization'01*, pages 143–150, 2001.

[5] Doantam Phan, Ling Xiao, Ron Yeh, and Pat Hanrahan. Flow map layout. In *Proceedings of IEEE Symposium on Information Visualization 2005 (INFOVIS'05)*, pages 219–224, 2005.

[6] Ben Shneiderman. Extreme visualization: squeezing a billion records into a million pixels. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, pages 3–12, 2008.

[7] Kiyohisa Taguchi, Kazuo Misue, and Jiro Tanaka. Visualization for spatiotemporal distribution of people's rich emotions. IPSJ SIG Technical Report 2014-GN-91(36), 2014. in Japanese.

[8] Waldo R. Tobler. Experiments in migration mapping by computer. *The American Cartographer*, 14(2):155–163, 1987.

[9] Edward R. Tufte. *Envisioning Information*. Graphics Press, 1990.

[10] Greg Turk and David Banks. Image-guided streamline placement. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, pages 453–460, 1996.