

## 画像情報の再構成による人間行動の可視化

伊藤 貴之\*

### Visualization of Human Activity by Reconstruction of Image Data

Takayuki Itoh

#### 1. 緒論

ビッグデータ、IoT、人工知能といった最近の情報技術のトレンドの多くには、写真やビデオといった画像情報が深く関与している。撮影機能を有する各種機器がインターネットに直接接続することで、膨大な静止画や動画がデータとして記録される。さらに、このようにして構築された画像データベースの分析が人工知能研究の一大課題になっている。

一方で、画像情報が機械的な記録と分析の手段として活用されたのは最近のことである。言うまでもないが、画像メディアはもともと人間が鑑賞して楽しむために開発されたものである。よって、人工知能などによる機械的な知識獲得のための研究開発だけでなく、人間自身の操作と視認による知識獲得を支援するための研究開発も、依然として画像技術における重要な課題である。

本稿では画像情報を人間自身で視認するための可視化技術の動向を概観する。といっても、本稿の主題を「画像情報を可視化する」と表現してしまうと若干の違和感がある。画像はもともと視認できる情報であり、それを「見えるようにする」というのは不自然な文章表現であるとも考えられる。そこで本稿では「画像情報の再構成によって観察可能となる知識」を可視化することを考える。ここでいう再構成とは例えば、静止画や動画を並べ替える、静止画や動画から高水準な情報を抽出する、といった処理を指す。本稿では、このような処理によって再構成された画像情報を可視化する各種技術について概観する。その中でも本稿では特に、画像中に写る人間の行動に焦点をあてていくつかの技術を紹介する。

また本稿の最終章では「今後の展望」として、画像情報の可視化が従来の可視化技術と概念的にどのように異なるのか、これを抽象化して議論すると今後の可視化研究にはどのような方向性が考えられるのか、といった点

を論じる。

#### 2. 静止画像に関する可視化手法

日常生活においても専門業務においても、大量の静止画像を保有する機会は非常に多い。本章では大量の可視化結果の比較閲覧、個人写真の閲覧支援、という2つの目的を中心に、大量静止画像のブラウザとしての可視化手法を紹介する。それに続いて、著者らが発表した個人写真ブラウザについて解説する。

##### 2.1 可視化技術のための静止画像ブラウザ

一般的に可視化結果は画像として保存される。科学技術の分野では、計算機上での設計技術や解析技術の発達にともない、設計と解析を繰り返しながら可視化結果を比較して意思決定につなげる工程が広く採用されるようになった。また物理現象の測定技術の発達にともない、現実の測定結果を可視化して比較する工程も増えた。そのような大量の可視化結果を比較するための手段として、大量の静止画像を一覧するためのブラウザとしての可視化手法が有用である。

大量の可視化結果を効率的に観察するには、興味深い差異や共通性を有する可視化結果群を発見しやすくなることが重要であり、そのためには可視化結果の画面配置が重要なポイントとなる。この考えにもとづいて発表された初期のシステムに Design Gallery<sup>1)</sup>がある。Design Gallery では入力画像の特徴量にもとづいて画像間の距離を求め、次元削減手法を用いて画面上での各画像の位置を決定することにより、類似性の高い画像が画面上で近くに配置されるような可視化結果を得る。また MoireGraphs<sup>2)</sup>では、関連性のある可視化結果を稜線で連結したグラフ構造を可視化している。このように初期の静止画像ブラウザは、多次元データやグラフの可視化手法を応用したものが多く見られていた。

##### 2.2 写真閲覧支援のための静止画像ブラウザ

デジタルカメラの普及により、個人が保存する写真の枚数はフィルムカメラを用いていた時代よりも格段に増えている。このような情勢をうけて、静止画像ブラウザの研究は個人写真の閲覧支援を目的とするものが増えてきた。その中でも最も有名なソフトウェアとして PhotoMesa<sup>3)</sup>があげられる。PhotoMesa は写真群がいくつ

\* 原稿受付 2000年1月1日

\*\* 正会員 可視化情報大学 可視化技術研究所(〒114-0034 東京都北区上十条 3-29-20, E-mail: info@vsj.or.jp)

\*\*\* 正会員 可視化情報(株) 技術本部

\*\*\*\* 学生員 可視化大学大学院 ◎×専攻

† 可視化以外大学

かのクラスタに分類されていると仮定し、画面空間を分割してできる長方形領域に各クラスタの写真を配置する。PhotoMesa がすべての写真を等しい大きさで表示するのに対して、それ以降に発表された諸研究では、注視すべき写真にズームする機能が活発に議論された。Walter らの手法<sup>4)</sup>では双曲面上に写真を配置することでズーム機能を実現している。また Brivio らの手法<sup>5)</sup>では、画面の中央部分には疎に、両端部分には密に発生させた母点群にそって写真を配置することで、写真群中の注視部分を拡大表示する。著者らのグループが発表した写真ブラウザ CAT<sup>6)</sup>では、階層的に分類された写真群を入れ子状に配置することで、効果的なズーム操作を実現する。CAT によるズーム操作の概要を Fig.1 に示す。

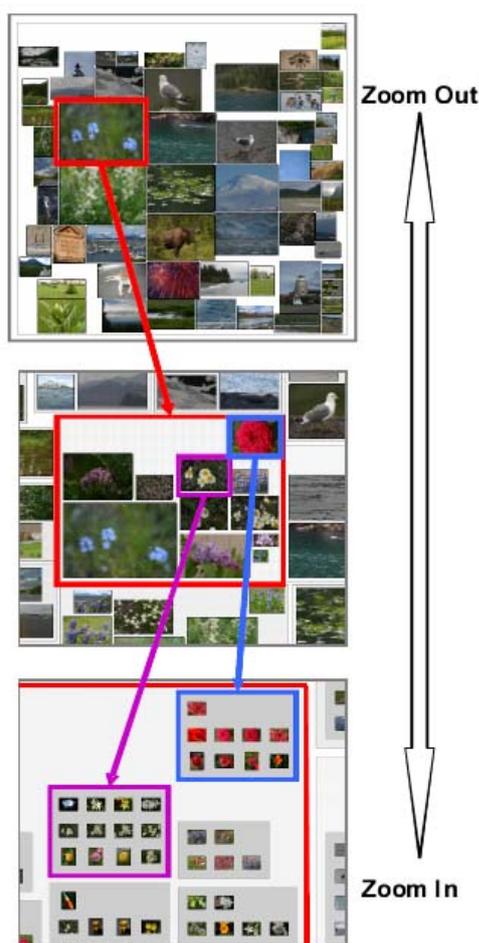


Fig.1 Zooming interface of an image browser CAT.

### 2.3 時間・場所・人物の3点に着目した写真閲覧

本節では著者らのグループが発表した写真ブラウザ MIAOW<sup>7)8)</sup>を解説する。MIAOW は写真群に撮影日時、撮影場所（緯度と経度）、被写体情報（写っている人物の ID）の3種類のメタデータが付与されていると仮定し、このメタデータにもとづいて写真群を階層的に分類し、各クラスタの画面上の位置を決定する。

Fig. 2 に MIAOW による個人写真群の表示例を示す。画面左側の LTView は XY 平面に経度と緯度が、Z 軸に時刻が割り当てられた3次元時空間となっており、ここ

にクラスタの代表写真を配置する。LTView は回転操作やズーム操作が可能になっており、ズーム操作によってクラスタ内の各写真を表示するモードに切り替わる。また LTView の背景には地図が表示され、代表写真の撮影場所を稜線で表示する。画面右側の PView には個人写真群から抽出された顔画像をその共起性でクラスタリングした結果が表示されている。

ここで、LTView と PView は連携操作が可能になっている。LTView 上で特定の代表写真をクリックすると、対応する写真クラスタに登場する人物が PView 上でハイライトされる。逆に PView 上で人物をクリックすると、その人物が登場する写真クラスタが LTView 上でハイライトされる。このような仕組みにより MIAOW は、写真所有者の人間行動を、時間、場所、登場人物の3つの視点から探索し、対話操作を楽しみながら興味ある写真を見つけられるようなユーザインタフェースを提供する。

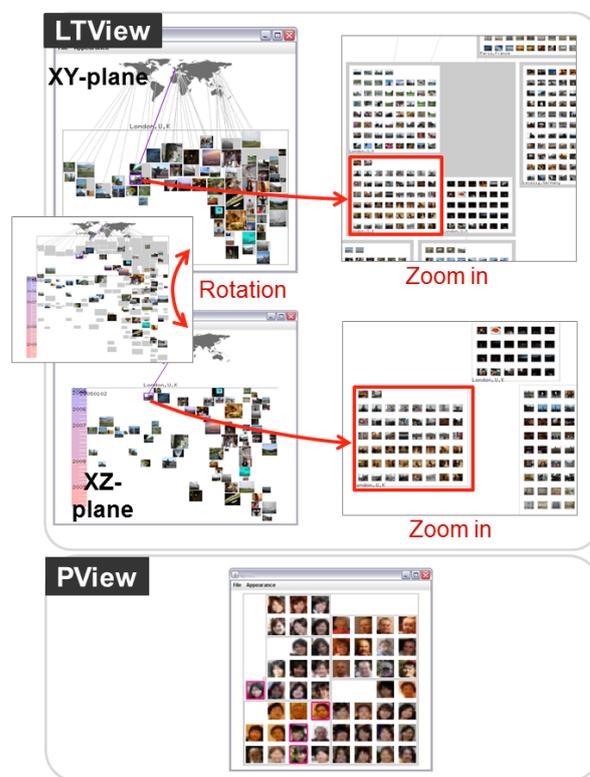


Fig.2 Visualization of personal photograph collection by a photograph browser MIAOW.

MIAOW を開発した筆頭著者は、自分自身の写真群を MIAOW で閲覧して自己の行動を振り返り、旧友と長く会っていないこと、家族旅行の形態が変わったことに気がついた、と述べている。このように写真群を探索的に閲覧する行為は、写真所有者の行動に関するライフログを振り返る行為でもありと考えられ、これにともなうさまざまな効果が期待される。

なお MIAOW を発表した論文<sup>7)8)</sup>では、LTView にて表示する代表写真の選出について詳しく言及していない。

代表写真の選出はそれ自体が活発な研究課題である。代表的な研究成果として、写真間の共通性に沿って構築したネットワーク構造に基づく Visual Rank<sup>9)</sup>や、写真中に写る人物の表情や位置関係に基づく手法<sup>10)</sup>などが知られている。著者らのグループでも、登場人物の出現頻度に基づく手法、人物の写真中の大きさや前景物体のコントラストに基づく手法、などを発表している。しかし依然として、万能な代表写真選出手法を開発することは容易ではなく、今後も写真閲覧に関する重要な課題であろうと考えられる。一方で、代表写真を1枚だけ選ぶのではなく、5~20枚程度の代表写真セットを大量写真の中から選ぶ、という課題もある。むしろ今後はこのような問題設定が重要になるとも考えられる。

### 3. 動画像に関する可視化手法

可視化技術の用途のひとつに「鑑賞に時間のかかるコンテンツの概要を理解する」という点があげられる。この観点から音楽<sup>11)</sup>や文学作品と同様に、動画像の要約表示はコンテンツの理解や検索に大きな効果をもつと考えられる。また動画像は、自然科学の各種実験記録、都市計画やマーケティング分析、セキュリティ管理といった多様な業務目的でも活躍しており、その分析結果の可視化は業務改善などの目的で有効であると考えられる。本章では動画像の要約可視化に関する動向を概観し、続いて著者らのグループによる人物歩行経路の要約結果に関する可視化手法を解説する。

#### 3.1 動画像情報の要約手段としての可視化手法

動画像を要約する直感的方法として、動画像から代表的なショットを抽出する処理を反復し、それらを静止画像の集合として一覧表示する方法が知られている<sup>12)</sup>。この静止画像群を効果的に一覧表示するためのデザインも動画像の要約における重要な課題であり、一例として漫画風に静止画像群を配置する Video Manga<sup>13)</sup>が発表されている。またビデオの題材を特定することで、例えばスポーツの要約<sup>14)</sup>やニュース番組の要約<sup>15)</sup>を効果的にした事例がある。

一方で、単一の静止カメラで撮影されたビデオに限定すれば、背景差分などの画像処理手法を用いて抽出された動体だけを重ねて表示することにより、1枚の静止画像で動画像を要約することも原理的には可能である<sup>16)</sup>。

以上の手法はいずれも静止画像の集合として動画像を表現するものであるが、それに対して、動画像から一定の非画像情報を抽出して表示する可視化手法も研究されている。その代表例が動体追跡結果として得られる移動経路の可視化手法である。汎用的な動体追跡結果に関する可視化手法が多く発表された<sup>17)</sup>一方で、追跡対象を特定することにより、単に移動経路を可視化するだけでなくその時間的傾向や因果関係なども分析できるようになる。この方針に基づく研究の例として、自動車交通流に関するデータの可視化事例<sup>18)19)</sup>が発表されている。

#### 3.2 人物歩行経路の要約と可視化

著者らのグループでも動体追跡結果の可視化に取り組んでいる。著者らの関心は移動方向が定められた自動車交通流のようなデータよりも、広場や公園、あるいはパーティー会場や展示会場といった、自由歩行に向けた空間において、大量の歩行経路をどのように要約表現するかという点にある。

この問題を解くための一つ目のアプローチとして藪下ら<sup>20)</sup>は、ビデオに撮影された空間を格子状に分割し、各格子を通過する経路群から代表的な数本を選出し、それらをエルミート曲線に変換して連結表示する手法を提案している。Fig.3(上)はビデオからの歩行経路抽出結果をそのまま描画した結果であり、これを Hermite 曲線の集合で近似して色と太さで歩行人数を表現した例が Fig.3(下)である。自由歩行に向けた空間における類似経路の集計結果を簡潔に表現していることがわかる。

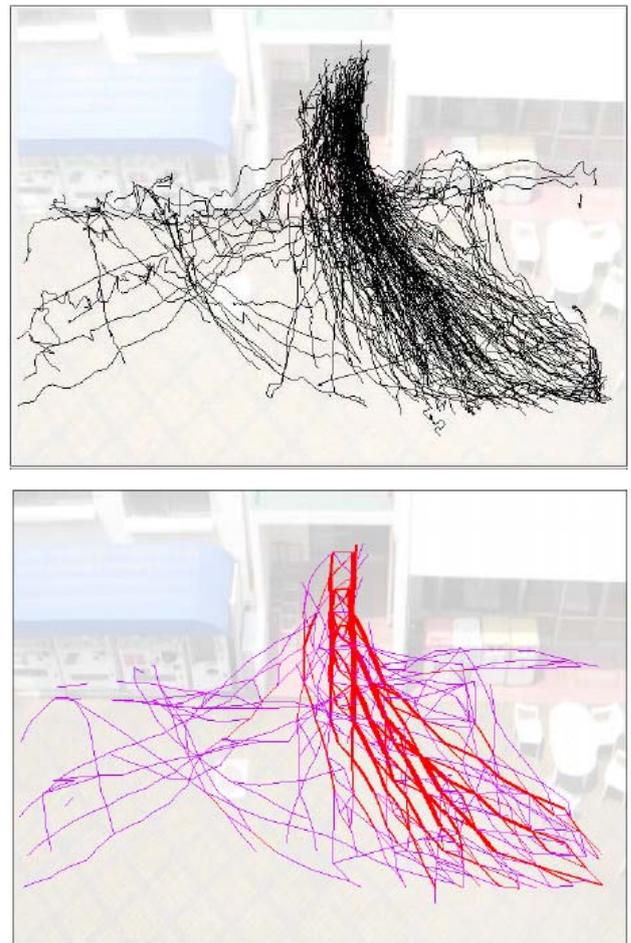


Fig.3 Visualization of trajectory of human walking path and its summarization as a set of Hermite curves.

二つ目のアプローチとして宮城ら<sup>21)</sup>は、Universal Sax というアルゴリズムを用いて経路形状を文字列に変換し、自然言語処理手法(現時点での実装では n-gram 法)を用いて頻出経路パターンを検出する手法を提案している。Fig.4はこの手法の原理を示した可視化例である。本手法ではビデオ撮影された空間を多数の多角形に分割し、そ

の各々に文字を割り当てる。そして、どの多角形領域を通過したかによって各々の経路形状を文字列に変換する。頻出経路パターンごとに色分けして移動経路を表示することで、どのような移動が多く見られるかを可視化する。

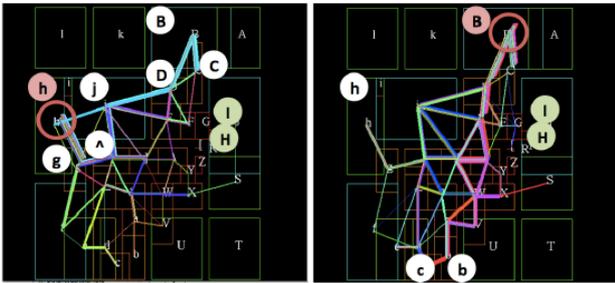


Fig.4 Visualization of summarized human walking path information. Trajectories are converted into sequences of characters.

#### 4. 今後の展望

本章では「見えない事象／見えにくい事象を見える事象に変換する」という行為を「狭義の可視化」と称する。狭義の可視化は学術的には既に長年にわたって研究されている。また科学技術系の実験環境や計算機の画面を介した可視化にとどまらず、例えばビジネスの過程や現況を共有する、捜査の過程を公開する、という意味でも可視化という単語は普及してきた。もはや狭義の可視化は「研究分野名として扱うほど先進的なものではない」と筆者は考える。そこで本章では、狭義の可視化にとどまらない可視化の概念について議論する。

本稿では画像情報の可視化技術を論じている。画像情報はもともと可視な事象である。つまり本稿で論じている可視化技術は「可視化によって見える事象を生む」のではなく、「可視化する前からもともと見える事象だったものを、計算機による変換を経て別の情報が見えるようにする」という行為に相当する。これは狭義の可視化技術と本質的に主旨が異なると考えられる。Fig.5はこれを図式したものである。図中の矢印は可視化による変換処理であり、見えない事象を見える事象に変換することを右向きの矢印で表現している。ここで狭義の可視化はFig.5の線(a)にあたり、画像情報の可視化技術は線(b)にあたる。この構図からも、画像情報の可視化技術には狭義の可視化とは異なる要件があることが示唆される。

画像情報の可視化技術を効果的に実現するためには、画像情報の再構成によってどのような潜在的な情報が抽出できるかを丁寧に議論する必要がある。そのためには、学習や論理にもとづいた機械的な手法と、人間の視覚認知能力やアフォーダンスに迎合した手法の両方が効果的であると考えられる。可視化技術の発展形として議論されている Visual Analytics という概念はまさに、機械的な手法と人間的な特性の両方を駆使したシステムの構築を目指している。言い換えれば、Visual Analytics という概念は画像情報の可視化において特に効果的であるとも考

えられる。

この論理に沿うと、「見えない事象を可視化によって変換しても依然として見えない事象」が存在するはずである。Fig.5の線(c)がそれに相当する。このような例は従来の可視化技術にもいくつか存在する。例えばネットワークを可視化するとき、一般的なノードリンク型の可視化手法では2つのノードが「リンクで連結されている」という情報を描画する。言い換えれば「リンクで連結されていない」という情報は可視化されない。仮に「リンクで連結されていない」という事象に重要な意味があるとしても、その事象は可視化されていないままである。あるいは流線でベクタ場を可視化するとき、「速度ゼロ」となる地点に流線は描かれない。例え「速度ゼロ」という事象に重要な意味があるとしても、その事象は可視化されていないままである。このような「描かれていないけど重要な事象」を可視化の研究ではどのように扱っていくべきであろうか。

以上をまとめると、狭義の可視化を拡張するキーポイントの中に「既に見える事象を別の見える事象に変換する」「可視化しても依然として見えない事象に着目する」という2点があると筆者は考える。本稿で論じた画像情報の可視化はその前者に相当する。

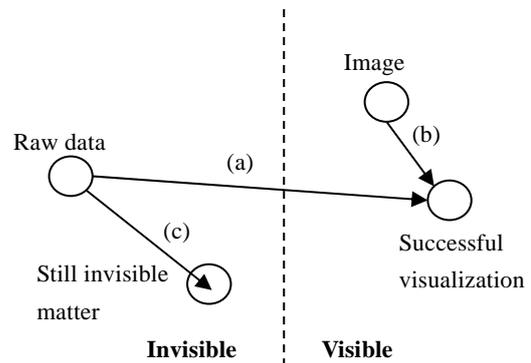


Fig.5 Concepts of visualization techniques. (a) Make invisible matters visible. (b) Convert visible matters into another style of visible. (c) Still invisible as a result of visualization process.

今後の展望としてもう1点論じておきたい点がある。人工知能をはじめとする高度な情報処理技術の発達により、今後は計算機によるビジョン技術が可視化結果を解釈できる時代になる、という点である。既に画像認識の分野では一般物体認識や一般行動認識ができるようになり、その認識対象は低水準な情報から高水準な情報に推移している。さらに高水準な画像認識事例として、人工知能にビデオゲームの画面を観察させてそのルールを理解させる、という実験も行われている。このように画像認識技術の対象が高水準化すると、当然のように可視化結果を画像認識技術で解釈可能な時代が到来することになる。そして、人間が視覚的に観察して情報を理解するために研究が進んできた可視化という行為は何だったのか、という議論が近未来的に必要なと考えられる。

## 可視化情報学会誌特集・連載記事原稿見本

しかしながら、人間と計算機には互いに得手・不得手があるという状況は今後も大きくは変わらないと考えられる。ある大手 IT 企業の人工知能システムは既に「人間との共生や分業」を一種のキャッチフレーズにしている。可視化の研究業界も同様な共生を認めるべきであろうと個人的には考える。つまり、現在は人間が観察している可視化画面の一部については、今後は人工知能などの先進的な情報処理技術でも観察可能な時代になるであろう。これを前提とした上で、逆に今後も人間の目で観察し続けるべき情報は何か、という点を議論すべきであろうと考える。例えば本稿で議論してきた「画像情報の再構成結果の可視化」においては、入力されてきた静止画像や動画像そのものは人工知能などの高度な情報処理技術によって解釈可能になるであろう。一方で、画像情報の再構成結果として視認される高水準な知識については、今後も人間が自身の目で観察すべきではなかろうか。また画像情報に限らず例えば、主観や感性を要する用途や、意思決定を要する用途、前例や正解がなく人間の経験や勘を要する用途、などにおいて人間自身で可視化結果を観察し続けるべき場面は今後もたくさん残るであろう。この点が今後の可視化研究の方向性のひとつになると筆者は考える。

### 参考文献

- 1) J. Marks, et al., Design Galleries: A General Approach to Setting Parameters for Computer Graphics and Animation, ACM SIGGRAPH, 389-400, 1997.
- 2) T. J. Jankun-Kelly, K.-L. Ma, MoireGraphs: Radial Focus+Context Visualization and Interaction for Graphs with Visual Nodes, IEEE Symposium on Information Visualization, 59-66, 2003.
- 3) B. B. Bederson, PhotoMesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps, Symposium on User Interface Software and Technology, 71-80, 2001.
- 4) J. A. Walter, D. Webling, K. Essig, H. Ritter, Interactive Hyperbolic Image Browsing – Towards an Integrated Multimedia Navigator, ACM SIGKDD, 111-118, 2006.
- 5) P. Brivio, M. Tarini, P. Cignoni, Browsing Large Image Datasets through Voronoi Diagrams, IEEE Transactions on Visualization and Computer Graphics, 16(6), 1261-1270, 2010.
- 6) A. Gomi, R. Miyazaki, T. Itoh, J. Li, CAT: A Hierarchical Image Browser Using a Rectangle Packing Technique, 12th International Conference on Information Visualisation (IV08), 82-87, 2008.
- 7) A. Gomi, T. Itoh, MIAOW: A 3D Image Browser Applying a Location- and Time-Based Hierarchical Data Visualization Technique, Advanced Visual Interface (AVI10), 225-232, 2010.
- 8) A. Gomi, T. Itoh, A Personal Photograph Browser for Life Log Analysis based on Location, Time, and Person, ACM Symposium on Applied Computing, 1250-1257, 2011.
- 9) Y. Jing, S. Baluja, VisualRank: Applying PageRank to Large-Scale Image Search, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(11), 1877-1890, 2008.
- 10) C. Li, A. Gallagher, A. C. Loui, T. Chen, Aesthetic Quality Assessment of Consumer Photos with Faces, IEEE International Conference on Image Processing, 3221-3224, 2010.
- 11) 伊藤, ユーザインタフェースとしての音楽情報可視化, 可視化情報学会誌, 35(1), 2-6, 2015.
- 12) J. Nam, A. H. Tewfik, Dynamic Video Summarization and Visualization, ACM International Conference on Multimedia, 53-56, 1999.
- 13) S. Uchihashi, J. Foote, A. Girgensohn, J. Boreczky, Video Manga, ACM International Conference on Multimedia, 383-392, 1999.
- 14) Y. Takahashi, N. Nitta, Video Summarization for Large Sports Video Archives, IEEE International Conference on Multimedia and Expo, 1170-1173, 2005.
- 15) H. Luo, J. Fan, J. Yang, W. Ribarsky, S. Satoh, Exploring Large-Scale Video News via Interactive Visualization, IEEE Symposium on Visual Analytics Science and Technology, 75-82, 2006.
- 16) C. Nguyen, Y. Niu, F. Liu, Video Summagator: an Interface for Video Summarization and Navigation, SIGCHI Conference on Human Factors in Computer Systems, 647-650, 2012.
- 17) R. P. Botchen, F. Schick, T. Ertl, Action-based Multi-field Video Visualization, IEEE Transactions on Visualization and Computer Graphics, 14(4), 885-899, 2008.
- 18) G. Andrienko, N. Andrienko. Spatio-temporal Aggregation for Visual Analysis of Movements, IEEE Symposium on Visual Analytics Science and Technology, 51-58, 2008.
- 19) H. Guo, Z. Wang, B. Yu, H. Zhao, X. Yuan: TripVista: Triple Perspective Visual Trajectory Analytics and Its Application on Microscopic Traffic Data at a Road Intersection, IEEE Pacific Visualization Symposium, 163-170, 2011.
- 20) H. Yabushita, T. Itoh, Summarization and Visualization of Pedestrian Tracking Data, 15th International Conference on Information Visualisation, 537-542, 2011.
- 21) 宮城, 大西, 渡辺, 伊藤, 記号化による人流データの圧縮と可視化, 画像の認識・理解シンポジウム (MIRU2015), SS5-9, 2015.