

GSLetterNeo vol.126

2019年1月

距離が近いデータをグループにして表示する

松原 伸人 matubara@sra.co.jp

はじめに

hakeimodoki は、軸上の距離が近いデータをグループにして表示する、データの範囲と量を表す方法のプロトタイプです。vol.123 画像群の可視化方法のプロトタイピングで紹介した、函館公園内を2018年10月9日に歩いて撮影した画像セットを hakeimodoki で表示してみます。[図1]

本内容は、科学技術振興機構(JST) 戦略的創造研究推進事業(CREST) 「データ粒子化による高速高精度な次世代マイニング技術の創出」プロジェクト (研究代表者: 宇野毅明 教授 (国立情報学研究所)) で行なっている研究開発の一部です。今回使用している画像は、筆者自身と、共同研究者の山本恭裕 特定教授 (京都大学)、およびKTL所長でもある中小路久美代 特定教授 (京都大学)の協力により得られた研究データの一部です。両者の同意のもと本誌に掲載させていただきました。

横が時間軸で、縦に4段あり、上の段からそれぞれ異なるカメラで3名が同じ2018年10月9日のほぼ同じ時間帯にほぼ同じ経路で函館公園内を散策して撮影した写真の撮影時刻を5pxの黒い四角形で表してプロットしています。3段目のデータは、右肩あたりに取り付けたカメラで15秒間隔で自動撮影したデータです。それぞれ



図1 hakeimodokiで画像セットを読み込んだ画面

最初の撮影時刻と最後の撮影時刻を表示しています。撮影した写真が上の段から134枚、99枚、406枚、220枚ありました。

どの期間でどれくらい撮影されていたかを見ようとすると、黒い四角形が集まっているあたりでたくさん撮影されていることはわかりますが、個数まではわかりません。

hakeimodoki は、時間軸上にプロットしている四角形が重ならないように、時間的に近いデータをグループにして縦に並べて表示します。

グループ化

hakeimodokiのグループ化の目的は、データを時間軸上に置いて表示した際に重ならないようにすることです。そこで、四角形同士が重ならない距離をグループ化の距離としています。四角形の中心からの距離が、指定したグループ化の距離よりも短いデータをグループにします。そのため、四角形の大きさにより、グループ化の距離の最小値が決まります。図1の場合、四角形の幅を5pxに設定しているので、グループ化の距離の最小値は5pxになります。

画面上端の3つの数値はグループ化の距離を表しています。左端が最小値0.00px、中央が現在値0.00px、右端が最大値1440.00pxです。最大値は表示エリアの幅 = 時間軸の見た目の長さです。現在値の上で左右にマウスを動かしてグループ化の距離を5pxずつ変えられるようになっていきます。

左端の写真の撮影時刻と右端の写真の撮影時刻から時間軸の長さは6304秒、表示エリアの幅から、5ピクセルあたりの時間はおよそ 21.89秒です。

$$6304 / 1440 * 5 \approx 21.89 \text{ [sec/px]}$$

グループ化は、データ間の距離を全て求めたデータ列を作成し、データ列の中からデータ間の距離が短いデータから順に、グループ化の距離より短いデータをペアにします。ペアにしたデータの平均時刻をグループの時刻とします。見つけたペアをデータ列に加え、ペアになったデータをとりぞいたデータ列を作成し、このデータ列から再びペアを見つける処理を行い、ペアが見つからなくなるまでこれを繰り返します。

時間軸の長さを1とすると、グループ化の距離 dw は $21.89 / 6304$ 、撮影時刻 dTx は、時間軸の開始時刻 T_s 、終了時刻 T_e 、データの時刻 T_x として次の式で表せます。

$$dTx = (Tx - Ts) / (Te - Ts)$$

2つのデータ dTx_1 と dTx_2 をペアにする条件は、2つのデータの距離が dw より短い場合です。

$$|dTx_1 - dTx_2| < dw$$

グループになったデータは、縦に四角形を積み上げて表しています。図1はグループ化を行う前の状態です。図2はグループ化の距離5、図3はグループ化の距離10、図4はグループ化の距離15の時の結果を表しています。

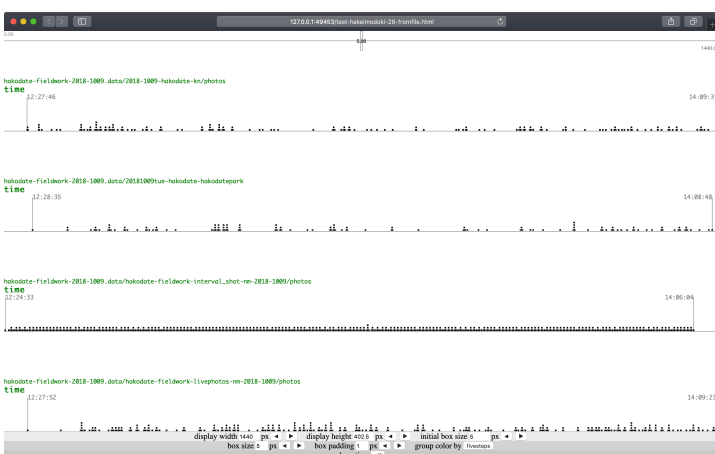


図2 グループ化距離 5.00

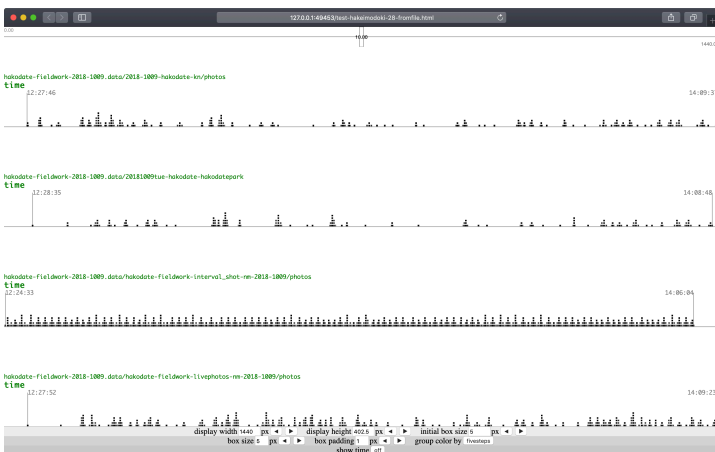


図3 グループ化距離 10.00

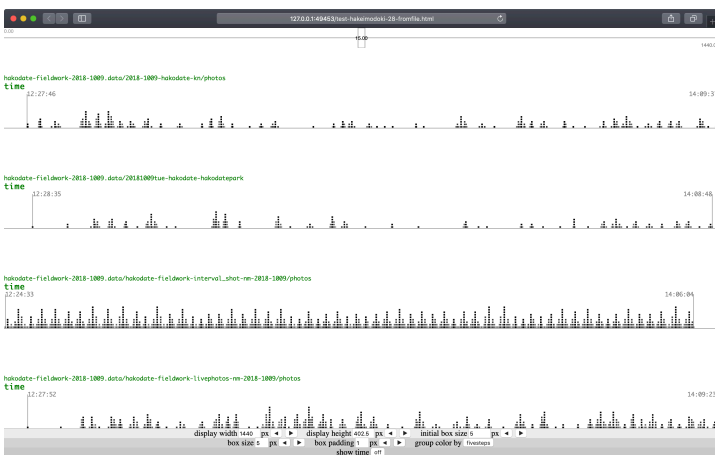


図4 グループ化距離 15.00

グループにすることでデータが重なることがなくなり、データを積み上げて表すことで数えられるようになりました。

図4を拡大表示すると僅かに見える程度ですが、グレーの四角形は、1段階前のグループ化の状態を表していて、もともとどこにデータがあったのかを表し、グループの期間が分かるようにしています。

データの重なるの可視化

軸をクリックすると、クリックした時刻に赤い線を表示します。[図5] 赤い線を含む期間にあるグループを薄橙色で表しています。同時に、他の軸にある期間が重なるグループも薄橙色で表します。

図5は、1段目の軸の左から4つ目のグループの上をクリックした様子です。1段目のグループの撮影期間と、2段目の軸の左から3つ目のグループの撮影期間が重なっていて、3段目の軸の左から7つ目と8つ目のグループの撮影期間が重なっていて、4段目の軸の左から4つ目と5つ目のグループの撮影期間が重なっていることが分かります。

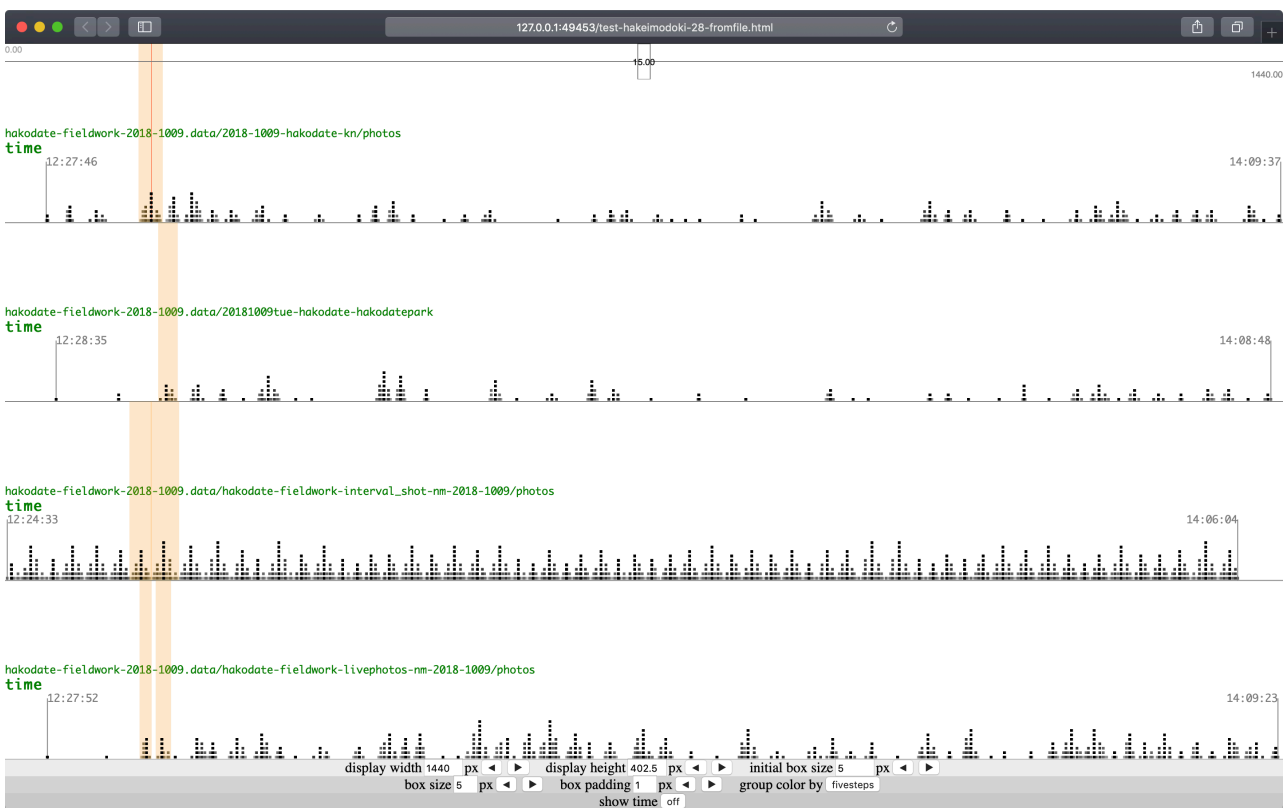


図5 1段目の軸の左から4つ目のグループを選んだ様子

見た目の違いを試すプログラミング

画面の下部の7つのパラメータは、画面の状態を変えてグループ化の結果、見た目の違いを試すためのものです。

display widthとdisplay heightで表示エリアの大きさを変えます。initial box size は四角形の大きさの初期値を変えます。[図6]

box size と box padding は四角形の見た目の大きさを、グループ化の計算と関係なく変えて見てみるためのものです。



図6 表示エリア800x400 四角形の大きさ10にした様子

group color by は、グループを表す見た目を変えます。グループの見た目としてどれがいいか考えた候補を全てプログラムして書き溜めておいて、切り替えて見比べられるようにしています。いくつか試したグループの見た目を掲載します。

livestepsはグループ化の現在の状態から4段階前までの状態をだんだんと色が薄くなるようにして表します。

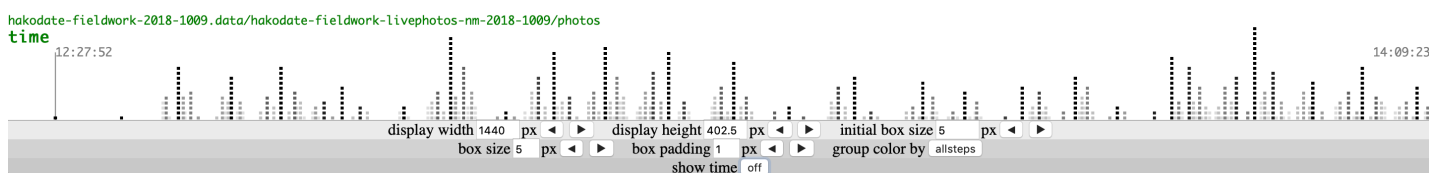
allstepsはグループ化の現在の状態から最初のグループ化してない状態まで全てを黒から白までグラデーションで表します。

distanceCyclic10はグループ化の段階を用意した10色を順に適用して表します。グループの範囲を楕円形で表します。

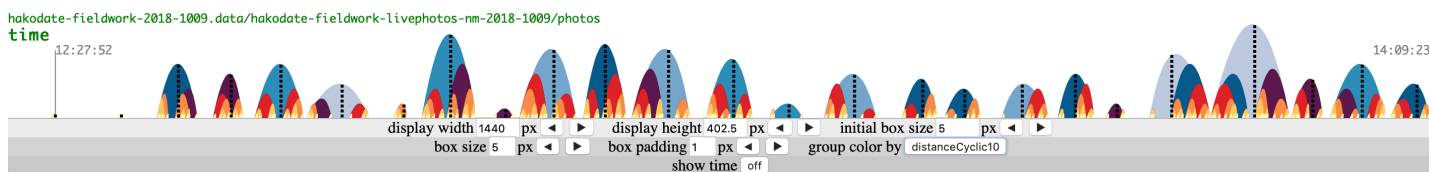
itemCyclic10はグループ化前のデータに、時間順に10色を順繰りに適用して表します。グループの色はメンバーの先頭の色になります。

show time はグループの開始時刻と終了時刻と真ん中の時刻を表示します。

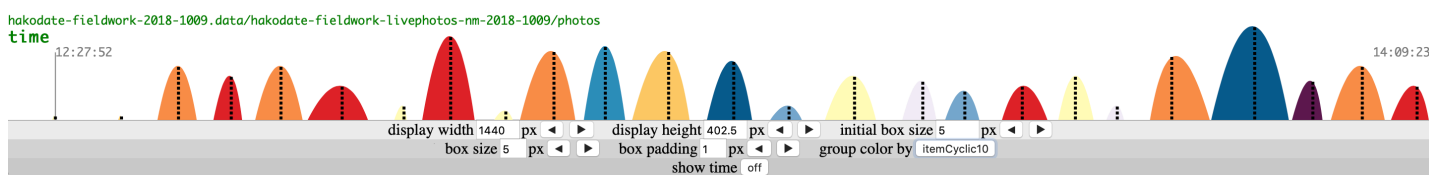
「データをグループにする」と、言葉にすると1つですが、人によって頭の中のイメージは違ってくると思います。頭の中のイメージと実装との印象も違うと思います。データをグループにして表示して見ると、グループの範囲が分かり、データが重なる範囲も表現できるようになりました。



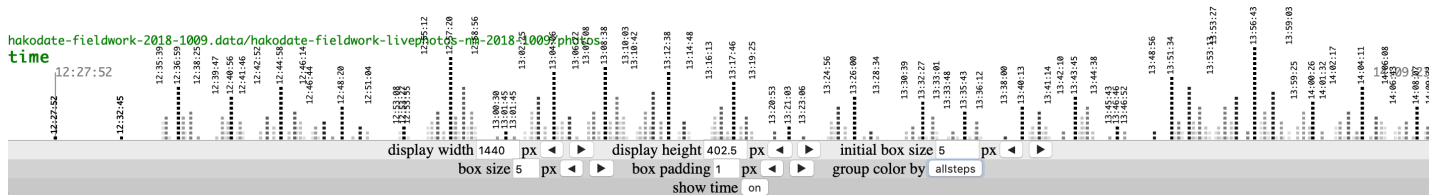
allsteps



distanceCyclic10



itemCyclic10



show time

GSLetterNeo vol.126

発行日 2019年1月20日

発行者 株式会社 S R A 先端技術研究所

編集者 土屋 正人

バックナンバー <https://www.sra.co.jp/gslletter/>

お問い合わせ

gsneo@sra.co.jp

〒171-8513 東京都豊島区南池袋2-32-8

