

筑波大学大学院博士課程

システム情報工学研究科修士論文

大規模データの特徴把握が可能な  
折れ線グラフの描画手法の開発

石川 凜太郎

修士（工学）

（コンピュータサイエンス専攻）

指導教員 三末 和男

2016年3月

## 概要

時系列データ等の可視化手法として折れ線グラフが広く用いられている。折れ線グラフは傾向や外れ値といったデータの特徴の把握を助けてくれるが、系列が多くなると大量の線で描画領域が埋め尽くされ、特徴の把握が困難になってしまう。そこで大規模なデータに対しても適用可能な折れ線グラフの描画手法の開発を目的とした研究を行った。我々はまず描画領域の各部分(画素)をその集積量に基づいて着色することを考えた。そして着色を制御するために、集積量の分布の利用に加えトーンマッピングやガンマ変換といった画像処理の技術を取り入れた。さらに我々は開発した描画手法に基づき、各部分の集積量と部分に割り当てる色の関係性を手動で制御できるツールを開発し、実在する大規模なデータに対して適用した。実在するデータに対して手法を適用した結果、提案手法の有効性が示された。また、「集積量の分布」と「集積量と色の関係性を表す変換関数のトーンカーブ」の間に形状の類似性が見られることが分かった。最後に我々はこの知見に基づき、提案手法に基づいた最適な着色の制御方法の開発に向けた考察を行った。

# 目次

<b>第1章</b>	<b>序論</b>	<b>1</b>
1.1	系列の多いデータに対する折れ線グラフの適用	1
1.2	本研究の目的とアプローチ	1
1.3	貢献	2
<b>第2章</b>	<b>提案手法の適用範囲と解決すべき問題の定義</b>	<b>4</b>
2.1	提案手法が対象とするデータ形式と手法の適用範囲	4
2.2	問題となる状態の定義	4
<b>第3章</b>	<b>関連研究</b>	<b>6</b>
3.1	系列ごとに異なる領域を用意する	6
3.2	1つの描画領域に全ての線を描く	6
3.2.1	線の削減	7
3.2.2	線の強調	7
<b>第4章</b>	<b>線の集積量に基づく折れ線グラフ描画手法の考案</b>	<b>9</b>
4.1	表現の説明に利用するデータ	9
4.2	試作1: 線の不透明度を下げる	9
4.3	試作2: 不透明度を変えて組み合わせる	10
4.3.1	試作2における課題	12
4.4	提案手法	13
4.4.1	集積量の算出	13
4.4.2	変換関数の作成	14
4.4.3	ガンマ変換	15
4.4.4	ガンマ変換を元にした変換関数の改良	18
4.4.5	画素値が上限を超えた微細領域集団に割り当てる画素値の決定	19
<b>第5章</b>	<b>実装</b>	<b>23</b>
5.1	折れ線グラフを描画するプログラムの開発	23
5.1.1	線が描画される画素の算出	23
5.2	提案手法を用いた折れ線グラフ描画ツールの開発	25
5.2.1	トーンカーブとヒストグラムを表示	25
5.2.2	インタフェース部分	25

<b>第 6 章</b>	<b>ユースケース</b>	<b>30</b>
6.1	時系列データ . . . . .	30
6.1.1	降水量データ . . . . .	30
6.1.2	気温データ . . . . .	31
6.2	PCP への適用 . . . . .	35
6.2.1	シカゴの犯罪履歴データ . . . . .	35
6.2.2	大学のスコアカードのデータ . . . . .	39
<b>第 7 章</b>	<b>考察</b>	<b>50</b>
7.1	提案手法の有効性 . . . . .	50
7.1.1	変換関数のパラメータに関する議論 . . . . .	50
7.2	今後の課題 . . . . .	51
7.2.1	着色の自動制御 . . . . .	51
7.2.2	より柔軟なトーンカーブの操作 . . . . .	52
<b>第 8 章</b>	<b>結論</b>	<b>54</b>
	謝辞	55
	参考文献	56

# 目次

1.1	全国の降水量の月単位の変化を表す折れ線グラフ. 計 1227 本の折れ線からなり, ひとつの折れ線はひとつの観測地点を表す. グラフの横軸は観測月と, 縦軸は降水量とそれぞれ対応している. . . . .	2
2.1	千葉県の子年齢別の人口の推移を表す折れ線グラフ. 計 101 本の線からなり, ひとつの折れ線はひとつの年齢を表す. また, グラフの縦軸は人口と, 横軸は観測年とそれぞれ対応している. . . . .	5
4.1	アプローチの説明に用いる擬似データ. 計 1150 本の線からなり, グラフの横軸は時刻と, 縦軸は擬似的に作成した系列の値とそれぞれ対応している. . .	10
4.2	不透明な線で描いた擬似データの折れ線グラフ. . . . .	11
4.3	試作 1 の折れ線グラフ. 半透明な線で描いた擬似データの折れ線グラフ. . .	11
4.4	試作 2 の折れ線グラフ. 小領域ごとに透明度を変えて組み合わせている. . .	12
4.5	試作 2 において不透明度の代わりに色相を割り当てたグラフ. . . . .	13
4.6	提案手法を用いて描いた折れ線グラフ. . . . .	14
4.7	提案手法において線形な変換関数を適用したときの折れ線グラフ. . . . .	15
4.8	擬似データの集積量の分布を表すヒストグラム. 横軸が集積量, 縦軸が微細領域数とそれぞれ対応している. . . . .	16
4.9	ガンマ変換の変換関数における $\Gamma$ の値とトーンカーブの関係性. トーンカーブの線の色は図中右下の $\Gamma$ の値を示す文字の色と対応している. 横軸は入力画像の画素値, 縦軸は出力画像の画素値とそれぞれ対応している. . . . .	17
4.10	$x_{border}$ の値を操作した時のトーンカーブの変化を示す図. 図の横軸は線の集積量に, 縦軸は微細領域に割り当てられる画素値にそれぞれ対応している. $x_{border}$ はトーンカーブが頭打ちになる点を変更するためのパラメータである. . . . .	19
4.11	$x_{trans}$ の値を操作した時のトーンカーブの変化を示す図. 図の横軸は線の集積量に, 縦軸は微細領域に割り当てられる画素値にそれぞれ対応している. $x_{trans}$ は $x$ 軸方向の平行移動量を表すパラメータである. . . . .	20
4.12	$y_{trans}$ の値を操作した時のトーンカーブの変化を示す図. 図の横軸は線の集積量に, 縦軸は微細領域に割り当てられる画素値にそれぞれ対応している. $y_{trans}$ はトーンカーブの $y$ 軸方向の平行移動量を表すパラメータである. . . . .	21
4.13	上図: 割り当てられる画素値が 255 を超えてしまった領域のみで構成される描画領域. 下図: 上図にガウシアンフィルタをかけた結果. . . . .	22

5.1	開発したツール. . . . .	26
5.2	ツール上に表示されるトーンカーブとヒストグラム. トーンカーブ, ヒストグラム共に横軸は集積量と対応している. また, トーンカーブにおける縦軸は微細領域に割り当てられる画素値と, ヒストグラムにおける縦軸は微細領域の数とそれぞれ対応している. . . . .	27
5.3	ツール起動時のインタフェース部とトーンカーブ及びヒストグラム表示部. . .	29
6.1	3種類の手法を用いて描いた降水量データの月ごとの変化を表す折れ線グラフ. 各グラフは計 1227 本の線からなり, ひとつの折れ線はひとつの観測地点を表す. また, グラフの横軸は観測月と, 縦軸は降水量とそれぞれ対応している. . . . .	32
6.2	降水量データの折れ線グラフにおける変換関数のパラメータ調整過程. . . . .	33
6.3	提案手法を適用した降水量データの折れ線グラフ. $\frac{1}{F} = 0.3, x_{border} = 0.4, x_{trans} = 0, y_{trans} = 0$ . . . . .	34
6.4	提案手法を適用した降水量データの折れ線グラフに対してトーンマッピングを行った結果. . . . .	35
6.5	3種類の手法を用いて描いた全国の気温の毎月の変化を表す折れ線グラフ. 計 823 本の折れ線からなり, ひとつの折れ線はひとつの観測地点を表す. また, グラフの縦軸は気温と, 横軸は観測月とそれぞれ対応している. . . . .	36
6.6	気温データの折れ線グラフにおける変換関数のパラメータ調整過程. . . . .	37
6.7	提案手法を適用した気温データの折れ線グラフ. $\frac{1}{F} = 0.3, x_{border} = 1.0, x_{trans} = 0, y_{trans} = 0$ . . . . .	38
6.8	3種類の手法を用いて描いたシカゴの犯罪履歴データの PCP. 計 291976 本の折れ線からなり, ひとつの折れ線はひとつの 1 件の犯罪を表している. また, PCP の各軸は犯罪に関する情報ひとつひとつと対応している. . . . .	40
6.9	提案手法の変換関数において, $\frac{1}{F}$ と $x_{border}$ の値のみを変更し作成した PCP. 右側の橙色のアノテーションを付けた箇所において, 密な領域の色のコントラストが乏しい. . . . .	41
6.10	シカゴの犯罪履歴データの PCP における変換関数のパラメータ調整過程. . .	42
6.11	提案手法を適用したシカゴの犯罪履歴データの PCP. $\frac{1}{F} = 0.2, x_{border} = 0.15, x_{trans} = 0, y_{trans} = 0.3$ . . . . .	43
6.12	提案手法を適用した犯罪履歴データの PCP に対してトーンマッピングを行った結果. . . . .	44
6.13	3種類の手法を用いて描いた大学のスコアカードのデータの PCP. 計 7804 本の折れ線からなり, ひとつの折れ線はひとつの大学を表している. PCP の各軸は学問領域と, 軸の値は学問領域毎の賞の受賞割合と対応している. . . . .	45
6.14	提案手法の変換関数において, $\frac{1}{F}$ と $x_{border}$ の値のみを変更し作成した PCP. 橙色のアノテーションを付けた箇所において, 密な領域の色のコントラストが乏しい. . . . .	46
6.15	大学のスコアカードのデータの PCP における変換関数のパラメータ調整過程. . .	47

6.16	提案手法を適用した大学のスコアカードのデータのPCP. $\frac{1}{T} = 0.3, x_{border} = 1.0,$ $x_{trans} = 0.05, y_{trans} = 0.2$ . . . . .	48
6.17	提案手法を適用した大学のスコアカードのデータのPCPに対してトーンマッ ピングを行った結果. . . . .	49
7.1	ユースケースで用いたデータの集積量の分布とトーンカーブ. . . . .	53

# 第1章 序論

## 1.1 系列の多いデータに対する折れ線グラフの適用

折れ線グラフは時系列データ等の特徴を把握するために広く用いられている可視化手法である [13]. データを折れ線グラフの形に可視化することによって系列の変化の様子の把握が容易になる [12]. また, 複数の折れ線を重ねることによって類似した傾向を持つ系列集団や特異な系列の発見が可能になる. しかし, 折れ線グラフは大規模なデータに対する適用が困難という課題を抱えている. 図 1.1 は全国の降水量の月単位の変化を表す折れ線グラフである [20]. 黒い背景に白い線を引いており, グラフの横軸は観測月と, 縦軸は降水量とそれぞれ対応している. 図 1.1 を見るとグラフの下部が真っ白に塗りつぶされており, 特徴の把握が困難になっていることが分かる. この描画領域が大量の線で埋め尽くされてしまう問題の解決を目的としてこれまでに数多くの手法が提案されている.

1 つの方針は線のバンドリングや線の集約, 削減によって画面上の視覚的混雑を減らす方針である. この手法は類似した傾向を持つ系列集団の発見を目的としてデータの分析を行う場合, 有効な手法である. その一方でデータから取得したい情報が明確でない仮説発見を目的とした分析においては重要な情報を喪失してしまう可能性を秘めている.

もう 1 つの方針は線の削減や集約は行わず, 色付けや線の透明度の調整によって特定の情報を目立たせたる方針である. 前の方針と比較してデータ全体の様子を正確に俯瞰しつつ, ある傾向を持った系列集団の発見もできるという利点がある. 一方でこの色付けや透明度の調整はデータや線分に対して排他的なクラスタリングを行った結果に基づいて行われる場合が多い. 排他的なクラスタリングを行った基づいた色付けや透明度の調整は視覚的には傾向の異なる系列が同一のクラスタに入ったり, 逆に視覚的には類似した傾向を持つ系列同士が同じクラスタに入らず目立たなくなってしまうといった人間の直感との不一致を起す可能性がある.

## 1.2 本研究の目的とアプローチ

本研究では大規模なデータの特徴把握が可能な折れ線グラフの描画手法を開発することを目的とする. 我々はデータから取得したい情報が明確でない仮説発見を目的とした分析において, データ全体の様子をまず俯瞰するために折れ線グラフを用いることを想定し, 画面上の線の削減による視覚的混雑の解消は行わない. また, 本研究ではデータや線分に対する排他的なクラスタリングを行わず, 描画領域の各微細領域 (画素) における線の集積量に着目し



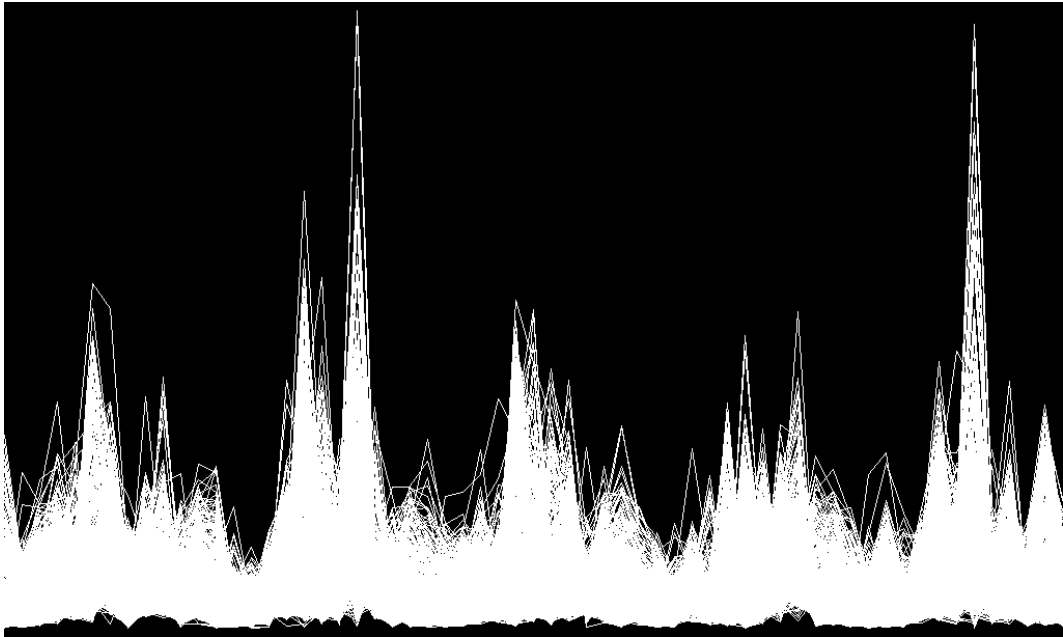


図 1.1: 全国の降水量の月単位の変化を表す折れ線グラフ. 計 1227 本の折れ線からなり, ひとつの折れ線はひとつの観測地点を表す. グラフの横軸は観測月と, 縦軸は降水量とそれぞれ対応している.

同じ集積量を持つ微細領域を同一の色で着色するアプローチをとる. 1 これは人間の直感との不一致が起きる可能性を回避し, データそのものの情報をできるだけ正確に表現するねらいがある. 微細領域への着色を制御するための手段として, 我々は画像処理において用いられている技術を折れ線グラフの描画に適用することを考える. 我々は画素値の分布に基づいた画像の濃淡変換の手法を参考に, 集積量の分布に基づいて微細領域毎の集積量と微細領域に割り当てる画素値の関係性を定義する変換関数を作成した. 変換関数はガンマ変換に用いられる関数を基に, 集積量の分布に応じた微細領域への柔軟な色付けを可能にする 3 種類のパラメータを付与した. さらにこの変換関数に加えてトーンマッピングの技術を導入することによって, データの規模が大きく取りうる集積量の値の範囲が広い場合にも微細領域に対する可能な限りデータの特徴を俯瞰できるような色の割り当てを可能にする.

開発した提案手法の有効性を示すために, 我々は提案手法を適用したデータ分析ツールを開発する. そして開発したツールによって実在するデータを用いたユースケースを作成し, 提案手法の有効性を示すとともに集積量に基づく最適な着色の制御方法の開発に向けた考察を行う.

### 1.3 貢献

本研究の貢献は以下の通りである.

1. データそのものの持つ情報をできるだけ損なわずにその特徴を俯瞰できるような折れ線グラフの描画手法を開発した点。データに関する知見が少ない状況におけるデータの特徴の俯瞰を目的とするため、開発した描画手法には線の集約や削除、及びデータや線分に対する排他的なクラスタリングを行わない。
2. 折れ線グラフの描画に分布に基づくトーンカーブの操作やトーンマッピング、ガンマ変換といった画像処理の技術を導入した点。開発した描画手法において描画領域に対する着色を制御するに辺り、これらの技術を取り入れている。そしてユースケースを元にその有効性に関する考察を行った。

## 第2章 提案手法の適用範囲と解決すべき問題の定義

本研究の目的は大規模データに対して適用してもデータの特徴が俯瞰できるような折れ線グラフを開発することである。本章ではまず提案手法が対象とするデータ形式と、提案手法を適用することができる折れ線グラフ以外の可視化表現について述べる。続いてデータの特徴を俯瞰するために解決すべき問題について、実例を示しながら解説する。

### 2.1 提案手法が対象とするデータ形式と手法の適用範囲

本研究が対象とするデータ  $V$  において  $v \in V$  をデータの系列とする。また、 $v_i$  は系列  $v$  の  $i$  番目の値を表すものとする ( $1 \leq i \leq m$ )。我々が提案する描画手法は折れ線グラフを描画する際に適用することを想定しているが、線を座標領域に複数描画することによってデータのもつ特徴を明らかにする表現は折れ線グラフ以外にも存在する。代表的なものが多変量データの特徴の把握を目的として開発された PCP(Parallel Coordinate Plots)[14][15] である。PCP は多変量データの 1 つ 1 つの変量に対して座標軸を用意して並列にかつ等間隔に並べ、隣り合う座標軸における点を結んでいくことによって 1 つのデータレコードを 1 本の線で表現する可視化手法である。1 つのデータレコードを表す折れ線が単一の座標領域上に複数重なり合うことになるため、その見た目は折れ線グラフに近いものとなる。我々が提案する手法はこの PCP にも適用することが可能である。本論文のユースケースでは実際に多変量データを用いて作成した PCP に対しても提案手法を適用する。

### 2.2 問題となる状態の定義

本研究では複数の線が重なりあうことで結果的に描画領域のある領域が単一の色で塗りつぶされ線を認識することが困難になったとき、その領域の状態を本研究が問題として取り上げる「特徴把握が困難になった状態」として定義する。ここで、特徴把握が困難になったグラフとそうでないグラフについて具体的な例を挙げて示す。

まずは特徴把握が可能なグラフの例を示す。図 2.1 は千葉県の子供の年齢別の人口の推移が記されたデータを折れ線グラフの形に可視化したものである [19]。データは 101 の系列と 27 の項目からなる。データの系列は年齢を、項目は観測年を表し、グラフの縦軸は人口と、横軸は観

測年とそれぞれ対応している。グラフを見ると複数の線が重なりあってはいるが、線の視認が困難な領域は確認できない。

続いて特徴把握が困難になったグラフの例が序論においても示した全国の降水量の月単位の変化を表す折れ線グラフ(図 1.1)である [20]。グラフの横軸は観測月、縦軸は降水量を表しており、観測地点を表すデータ系列の数は 1227、観測月を表す項目の数は 61(2010 年 1 月～2015 年 1 月)である。図 1.1 を見るとグラフの下部が真っ白に塗りつぶされてしまっている。この状態では個々の線を視認することができず、塗りつぶされているように見える領域において各系列がどのように交わり合っているのかを読み取ることは難しい。本研究ではこの状態になっている領域を可能な限り減らすことを研究の目的とする。

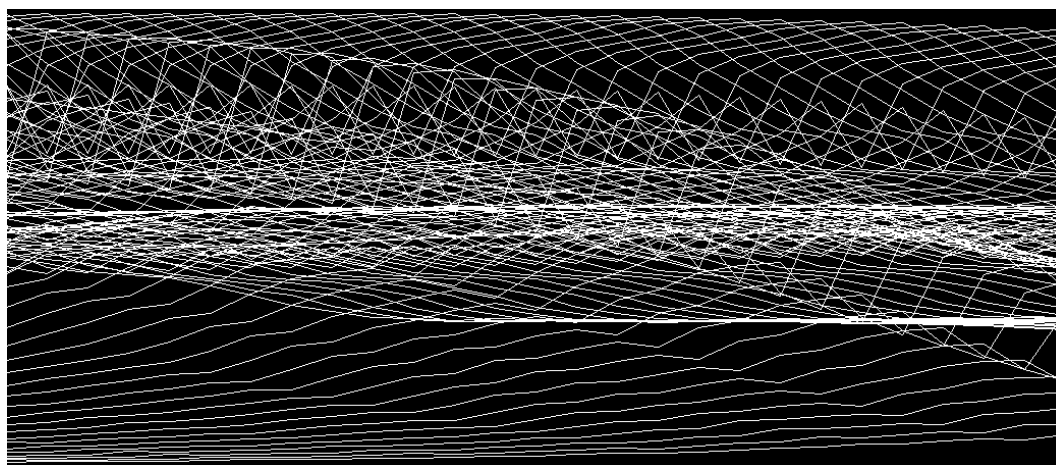


図 2.1: 千葉県の年齢別の人口の推移を表す折れ線グラフ。計 101 本の線からなり、ひとつの折れ線はひとつの年齢を表す。また、グラフの縦軸は人口と、横軸は観測年とそれぞれ対応している。

## 第3章 関連研究

本章では本研究の関連研究として、折れ線グラフや PCP を用いてデータの特徴を把握することを目的とした研究について述べる。本研究では単一の描画領域に大量の線を引くことを前提としているが、系列毎に異なる描画領域を用意する研究についても言及する。

### 3.1 系列ごとに異なる領域を用意する

折れ線グラフや PCP を用いてデータの特徴を把握することを目的としたアプローチとして本研究で用いている単一の座標領域に線を描いていくアプローチ以外に、系列ごとに異なる座標領域を用意し、それを並べて比較する手法が考えられる。Munzner[1] は1つ1つの系列の特徴を詳細に観察していくようなローカルなタスクを行う際にはこちらの手法の方が適していると述べている。既存研究ではなるべく多くの系列を同時比較できるよう、小領域に対してより多くの情報を詰め込むことを目的とした研究が行われている。Saito ら [2] は折れ線グラフを横にスライスし、スライスしたグラフにスライスする前の階級に応じて区別のし易い色を割り当てて重ねることによって小領域に表示してもその系列の値の変化を容易に読み取れるような表現を開発した。Javed ら [3] は“Simple line graph”, “Braided graph”, “Small multiples”, “Horizon graphs”の4種類の折れ線グラフをベースとした手法について、小領域上に表示したタスクをこなす際にどのようなタスクに対してどの手法がより適しているか、その検証を行った。4種類の手法の内“Simple line graph”と“Braided graph”は1つの描画領域に全ての線を描くアプローチ、“Small multiples”, “Horizon graphs”は系列ごとに異なる領域を用意するアプローチである。検証の結果、それぞれの系列の値の分散具合を比較するタスクを行う際には系列ごとに異なる領域を用意するアプローチの方が適している事が明らかになった。Javed らの研究では比較的小規模なデータを対象としており大規模データを用いたタスクにおいてどのような手法が適用しているかを検証することは今後の課題であるとしている。この系列頃に異なる描画領域を用意する手法は系列数が1000を越えるような大規模なデータに対して適用することは難しいとされており、Munzner[1]はこのアプローチは系列の数が数ダース~100程度を越えた際には適用が難しいと述べている。

### 3.2 1つの描画領域に全ての線を描く

本研究では規模の大きなデータを扱うことを目的とし単一の座標領域に全ての線を描くアプローチをとる。前の章でも説明したようにこのアプローチを採用する上で問題となるのが

大量の線が単一の描画領域上に描かれることで視覚的混雑が生まれ、特徴把握が難しくなることである。そしてこの視覚的混雑を解消することを目的としてこれまでに様々な手法が開発されている。

### 3.2.1 線の削減

問題解決のための1つの方針として線のある規則に従って集約したり削減したりすることによって描画領域上の視覚的混雑を減らす手法が考えられる。Ellisら[4]は大規模データのPCP(Parallel Coordinate Plots)による分析を可能にするためにPCP中の注目したい箇所を分析者に指定させ、その箇所を線のサンプリング数を削減した上で拡大表示する”Sampling Lens”を提案した。Trimmら[5]は学生たちの選択した講義とその成績との関係性を探るために、全ての生徒の選択したコースの変遷を表す線の平均のみを表示し残りの線の情報は背景色を用いて表すことを試みている。

Uchidaら[12]は大規模な時系列データの全体の変動の特徴を保持しつつその特徴を把握するために一定間隔ごとにグラフ形状に基づくクラスタリングを行い、類似した動きをする線分を要約することでグラフの詳細度を制御する手法を開発した。系列の数は減らさず、線をバンドリングすることによって描画領域上の視覚的混雑度の解消を試みるアプローチも存在する。Zhouら[6]は距離が近い線分をバンドリングし、線の密度に応じて色と透明度を決定することに寄って視覚的混雑度の解消を実現した。McDonnellら[7]は線のバンドリングと主要なクラスタの可視化、クラスタの密度を表す不透明度に基づいたヒント、そしてローカルな線の密度を表す影の4種類の技法と表現を組み合わせた手法によって専門家でない人にも見やすいようなPCPを開発している。

これらの手法はデータが持つ傾向を見ることを主要な目的とした分析を行う場合非常に有効な手法であると言える。一方で分析者データに関する知識が十分になく、データに関する仮説発見を目的とした分析を行う場合、集約や線の削除によって重要な情報が失われてしまう可能性がある。

### 3.2.2 線の強調

もう1つの方針は特定の情報を持つデータ系列のみを強調して表示する手法である。Hauserら[8]はある項目において、線分の傾きが近い集団を強調して表示する手法を提案している。特定の線を強調するアプローチにおいて、多く行われているのはデータや線分に対して排他的なクラスタリングを行い、クラスタに属する系列や線分を強調表示する手法である。Fuaら[9]が開発した手法はまずデータをクラスタリングし、クラスタ毎に色付けをおこない、更にクラスタの中心を強調して表示している。Arteroら[10]はデータセットの密度と頻度の情報に着目して傾向を持ったデータ群のみを高い精度で取り出し、それを強調表示する可視化手法を考案した。本研究ではデータや線分に対して排他的なクラスタリングは行わず描画領域の各微細領域(画素)における集積量に着目し、同じ集積量を持つ微細領域を同一の色で着色

する。これは視覚的には傾向の異なる系列や線分が同じクラスタに入っていたり、逆に視覚的には同一傾向を持つ系列や線分がクラスタに入らず目立たなくなるといった直感との不一致が起きる可能性を回避し、データそのものの情報をできるだけ正確に表現するねらいがある。本研究と同じく線の集積量に着目した色付けを行っている研究に Wegman らの研究が挙げられる。Wegman ら [11] も折れ線グラフを描画する際に線を構成する各画素の色をその画素の線の集積量に基づいて決定する手法を提案している。しかし、彼らは線の密な領域の様子を見せることを目的として集積量に応じて線形に色を対応させており、着色の制御についての考察は行っていない。本研究では画像処理の技術を応用した着色制御の仕組みを導入することで密な領域だけでなく、疎な領域も同時に見ることができるような折れ線グラフの描画手法の開発を行った。

## 第4章 線の集積量に基づく折れ線グラフ描画手法の考案

問題解決のためのアプローチとして、我々は折れ線グラフを描画した際の描画領域の微細領域における線の集積量及びその分布に着目する。本章ではまず提案手法の考案に至るまでに開発した2つの問題解決アプローチについて述べ、2つのアプローチが抱える問題を明らかにした後で提案手法で用いたアプローチについて説明する。

### 4.1 表現の説明に利用するデータ

提案手法の説明をするに辺り擬似データを用いる。擬似データは62の項目を持つ計1150系列のデータである。項目は時刻を表し、各時刻において各系列が取る値は全てプログラムを用いて作成した擬似的な値である(値の範囲は1~100)。データの詳細とそれぞれを不透明な折れ線グラフに可視化したものを図4.1に示す。黒い背景に白い線を引いており横軸は時刻と、縦軸は擬似的に作成した系列の値とそれぞれ対応している。1150系列のデータには3種類の傾向、“Pattern1”、“Pattern2”、“Pattern3”を持つ系列集団と傾向を持たないランダムな系列の集団で構成される。3種類の傾向を持つ系列集団の内、Pattern1とPattern2は300系列の比較的大規模な系列集団でPattern3は50系列の小規模な系列集団である。図4.2が示す通り、全ての線をそのまま単一の座標領域に描くと絵が潰れてしまいPatternの発見や線の観察が不可能になる。

### 4.2 試作1：線の不透明度を下げる

問題解決のための素朴なアイデアとして、線の不透明度を下げることによって線1本1本の視認性を高めることを考える。図4.3は全ての線に均一な不透明度が割り当てられた折れ線グラフである。グラフを見ると不透明な線のみを用いて描いた折れ線グラフと比較して密な領域のデータの特徴が掴みやすくなっていることがわかる。一方で特にグラフの上端、下端において多く見られる線が疎な領域の線が殆ど見えなくなってしまうという課題が生じてしまう。この課題を解決すべく疎な領域の線がある程度視認できるまで不透明度を上げると、逆に密な領域が真っ白に塗りつぶされてしまい傾向が見えなくなってしまう。



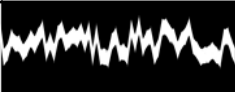
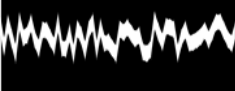



	Quantity	Outline
Pattern1	300	
Pattern2	300	
Pattern3	50	
Random Pattern	500	
Total	1150	

図 4.1: アプローチの説明に用いる擬似データ. 計 1150 本の線からなり, グラフの横軸は時刻と, 縦軸は擬似的に作成した系列の値とそれぞれ対応している.

### 4.3 試作 2 : 不透明度を変えて組み合わせる

試作 1 において, 線の不透明度を一律に設定するだけでは疎な領域と密な領域の線を同時に見せることが難しいことが分かった. そこで試作 1 における問題を解決するために今度は描画領域全体を正方形の小領域に区切り, 小領域毎に適した不透明度の線を引くことを考える. 各小領域に引かれる線の不透明度を決定するための情報として小領域内の線の密度に着目する. ここで言う線の密度とは, 小領域を更に細かい正方形の微細領域に区切ったとき, 各小領域における線が引かれている微細領域の割合のことを指す(式 4.1).

$$congestion_{x,y} = \frac{S_{x,y}}{l * l} \quad (4.1)$$

$congestion_{x,y}$  : 注目小領域内に引かれる線の不透明度

$x$  : 注目している小領域の行番号

$y$  : 注目している小領域の列番号

$S_{x,y}$  : 注目している小領域における線が引かれている微細領域の数

$l$  : 小領域 1 辺辺りの微細領域の数

式 4.1 に示されるように我々は線の密度が高い領域には低い不透明度の線が, 低い領域には高い不透明度の線が描かれるように不透明度を設定することによって全ての領域において線の視認が可能になると考えた. プログラム上では各小領域に引かれる線の不透明度は以下に示す手順で算出される.

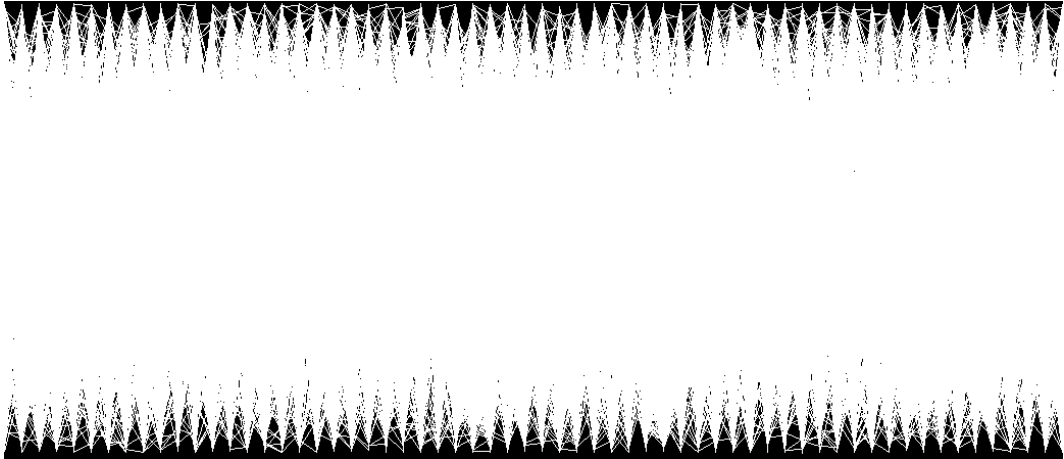


図 4.2: 不透明な線で描いた擬似データの折れ線グラフ.

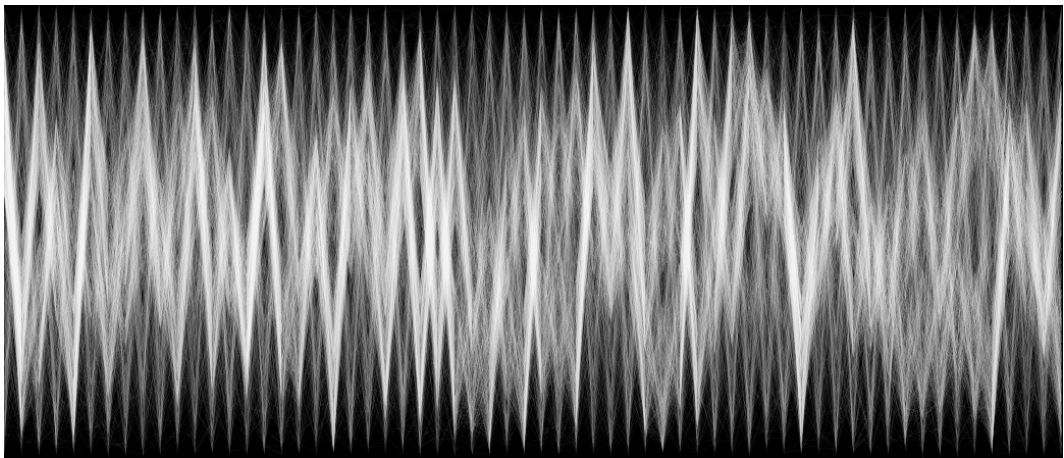


図 4.3: 試作 1 の折れ線グラフ. 半透明な線で描いた擬似データの折れ線グラフ.

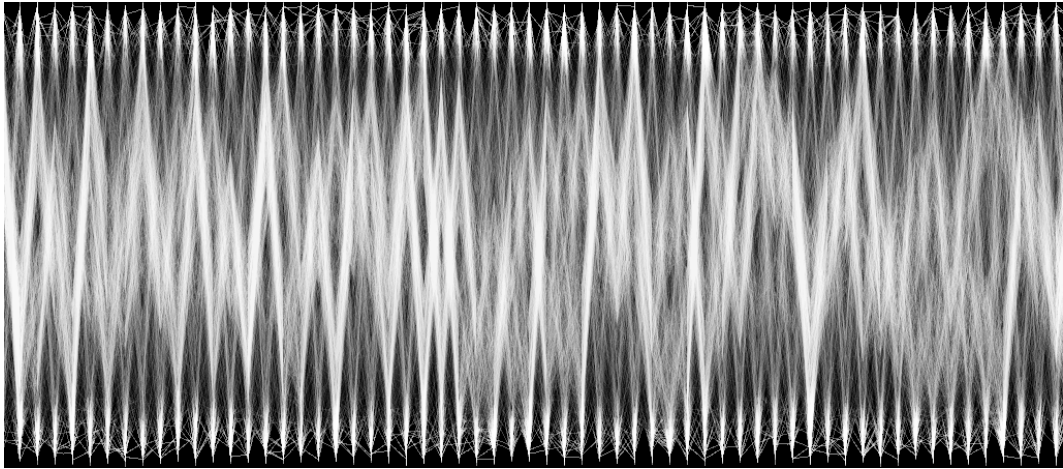


図 4.4: 試作 2 の折れ線グラフ. 小領域ごとに透明度を変えて組み合わせている.

1. 不透明な線を用いて折れ線グラフを描画する
2. 描画領域全体を正方形の小領域に区切る
3. 描画領域を走査し, 小領域毎に色が塗られている微細領域 (画素) の割合 (密度) を算出する.
4. 算出した密度を元に小領域内に引かれる線の不透明度を計算する

試作 2 のアイデアに従って作成した折れ線グラフを図 4.4 に示す. 全ての線に一律な不透明度が割り当てられたグラフと比較して線が疎な領域の線の視認を可能にしつつ密な領域の傾向が読み取れるようになっていることがわかる.

#### 4.3.1 試作 2 における課題

図 4.4 を見ると確かにグラフの特徴は掴みやすくなっているが, 一方で上下端の線が疎な領域が極端に目立ってしまっていることが分かる. 原因は上端の領域とその下の領域, 下端の領域とその上の領域の線の密度の差が極端に大きくなったことにある. この密度の差についてより詳しく見るために不透明度の代わりに色相を割り当てた折れ線グラフを作成し線の密度の分布を観察した (図 4.5).

図 4.5 のグラフの上下端に注目すると, 線の密度がかなり極端に偏っていることが読み取れる. 我々は手法を実データに対しても適用し, 密度差が極端な領域は実データに対して手法を適用した際にも生まれてしまうことを確認した. このような領域が生まれてしまう原因は線の密度という 2 次元的な情報を用いたことにあると考えられる.

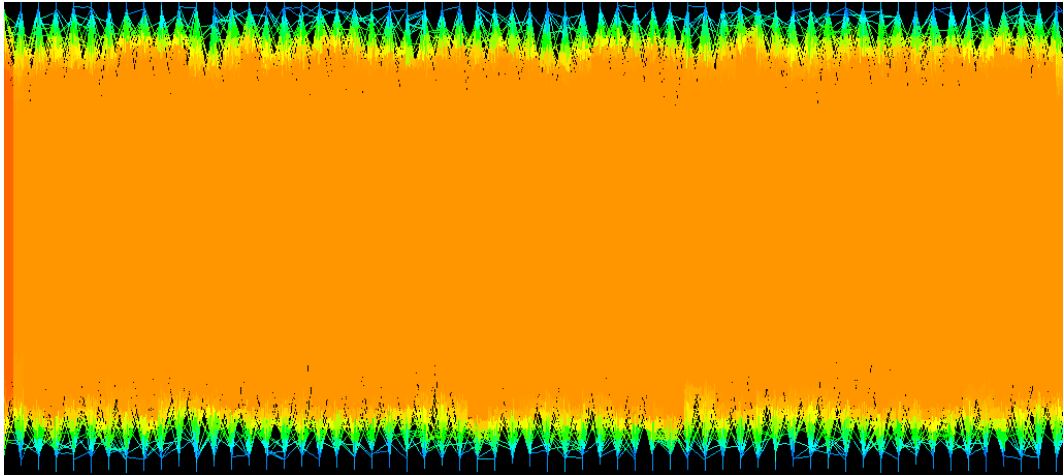


図 4.5: 試作 2 において不透明度の代わりに色相を割り当てたグラフ.

## 4.4 提案手法

試作 2 では線の密度という 2 次元的な情報に着目したが隣接する極端な密度差のある領域を生み出してしまい、それが視覚的な違和感を生みだした。提案手法では、試作 2 で着目した小領域よりも細かい微細な領域ごとの線の集積量という 3 次元的な情報に着目する。なお、「線の集積量」とは各微細領域における線の交差数のことを指す。最初に提案手法を用いて描画した擬似データの折れ線グラフを図 4.6 に示す。線が密な領域に着目すると擬似データを持つ類似した傾向を持った系列集団の存在が確認できる。また、線が疎な領域に着目すると上下端の疎な領域の線も視認できることが分かる。試作 2 における課題であった上下端の疎な領域が極端に目立ってしまう現象も回避できている。続いて本手法を用いた折れ線グラフの描画手順を以下に示す。

1. 描画領域全体を均等な大きさの微細な領域に区切る
2. 描画領域に収まるように折れ線グラフを描画したときの各微細領域における線の交差数を算出し、これをその微細領域における線の集積量とする
3. 2 で求めた各微細領域における集積量を変換関数を用いて画素値に変換し、描画領域に対して色付けを行う

以降の節では描画手順の 2 と 3 について詳しく解説を行う。

### 4.4.1 集積量の算出

まず、ある系列  $v$  の折れ線グラフを描画したときの描画領域上の各微細領域における線の有無を  $P_{x,y}(v)$  で表す。 $x, y$  はそれぞれ微細領域の  $x$  座標、及び  $y$  座標を表し ( $x \in N, y \in N$ ),

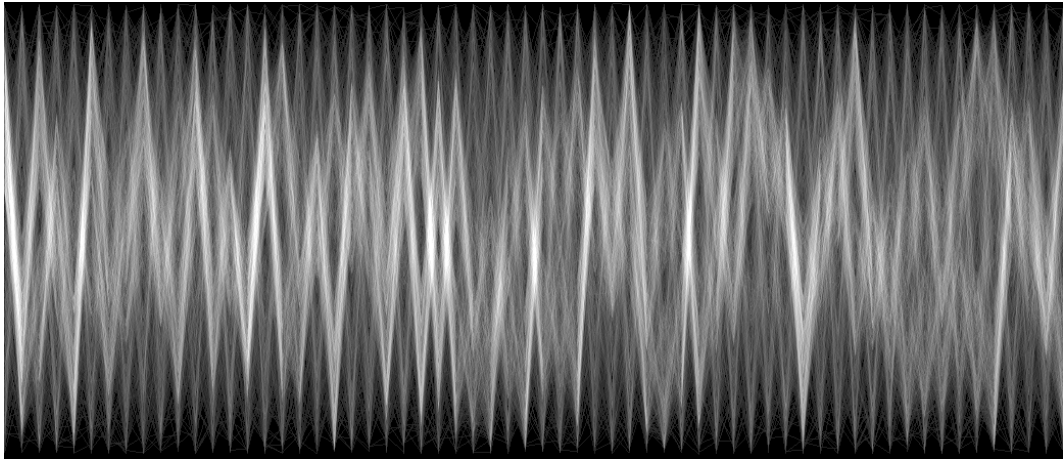


図 4.6: 提案手法を用いて描いた折れ線グラフ.

$P_{x,y}(v) = 1$  のとき座標  $(x, y)$  の微細領域上に系列  $v$  の折れ線が存在するものとする (式 4.2). 線の集積量  $L_{x,y}$  はこの  $P_{x,y}(v)$  の値を全ての系列について足し合わせることによって求められる (式 4.3).

$$P_{x,y}(v) = \begin{cases} 1 \\ 0 \end{cases} \quad (4.2)$$

$$L_{x,y} = \sum_{v \in V} P_{x,y}(v) \quad (4.3)$$

提案手法ではこの集積量を元に各微細領域に割り当てる画素値を決定する.

#### 4.4.2 変換関数の作成

線の集積量を描画領域の微細領域に割り当てる画素値に変換する際, 集積量と画素値の関係性を定義する必要がある. 最初に, 素朴なアイデアとして, 集積量と割り当てられる画素値とを線形に対応させることを考える. この変換関数を用いて色付けを行った折れ線グラフが図 4.7 である. グラフを見ると, また試作 1 で作成したグラフと同様に疎な領域の線が見えなくなってしまっていることが分かる. 原因を探るために描画領域全体における集積量の分布のヒストグラムを観察すると集積量の少ない微細領域集団が描画領域の大半を占めていることが分かる (図 4.8). 線形な変換関数を使用すると, この微細領域集団の殆どに背景色に近い画素値が割り当てられたことによって図 4.7 のような特徴の読み取りづらいグラフになったのだと考えられる.

そこで, 描画領域の大半を占める集積量の少ない微細領域集団を高いコントラスト, かつ視認可能な色で塗り分けられるような変換関数を作成する. ここで参考にしたのがデジタル画像処理における濃淡変換の技術である. デジタル画像処理における濃淡変換とは, デイ

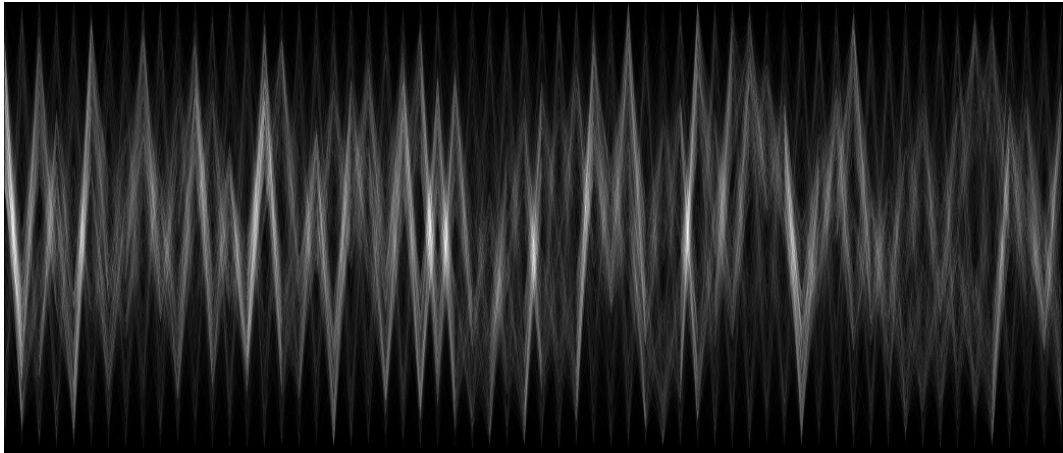


図 4.7: 提案手法において線形な変換関数を適用したときの折れ線グラフ。

デジタル画像の色のコントラストを変更する事によって画像の特定の箇所を目立たせたり、全体をぼかす処理である。デジタル画像の濃淡変換を行うためには入力画像の画素値と出力画像の画素値とを対応付ける階調変換関数を設定する必要がある [17]。この階調変換関数を決定するに辺り画像作成者は画像全体の画素値の分布に基づき、強調したい箇所のコントラストが強くなるようなトーンカーブを描く階調変換関数を設定する。この濃淡変換における階調変換関数として、トーンカーブが曲線を描くような関数を使われることがある。我々はこの曲線状の関数を折れ線グラフに利用することでコントラストを強くして見たい集積量区間には固有の色を割り当て、そうでない区間に割り当てる色の数を少なくすることを考えた。

#### 4.4.3 ガンマ変換

トーンカーブが曲線の形をとるような階調変化関数としてデジタル画像処理における濃淡変換において用いられているのがガンマ変換である [17]。式 4.4 がガンマ変換の変換式である。ガンマ変換の変換関数において、入力の値は 0~1.0 の範囲に収める必要がある。そのため、入力画像の画素値を 0~1.0 の範囲に正規化してから変換関数を適用し、その後 255 を乗算することで画素値に戻す計算処理を行う。 $\Gamma$  の値は任意の値をとるパラメータで、この値を変更することによってトーンカーブの曲率を変化させ、入力値と出力値の関係性を調整する。図 4.9 は  $\Gamma$  の値とトーンカーブの形状の関係性を示したものである。図からも読み取れるように  $\Gamma$  の値を大きくとると画素値の小さい画素間のコントラストが強調され、逆に  $\Gamma$  の値を小さくとると画素値の大きい画素間のコントラストが強調される。

$$y = 255 * \left(\frac{x}{255}\right)^{\frac{1}{\Gamma}} \quad (4.4)$$

$y$ : 出力画像の画素値

$x$ : 入力画像の画素値

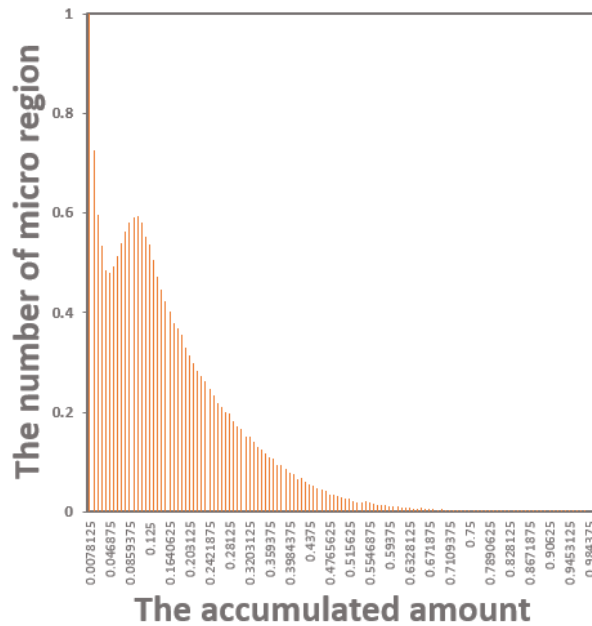


図 4.8: 擬似データの集積量の分布を表すヒストグラム。横軸が集積量，縦軸が微細領域数とそれぞれ対応している。

$\Gamma$  : 変換関数の曲率を制御するパラメータ

このガンマ変換の変換関数における「入力画像の画素値  $x$ 」と「出力画像の画素値  $y$ 」を「ある微細領域における線の集積量」と「ある微細領域に割り当てる画素値」に読み替え，変換関数として用いる (式 4.5)。

$$\begin{aligned}
 H_{x,y} &= 0 \\
 S_{x,y} &= 0 \\
 V_{x,y} &= 255 * \left( \frac{L_{x,y}}{L_{max}} \right)^{\frac{1}{\Gamma}}
 \end{aligned} \tag{4.5}$$

$H_{x,y}$  : 微細領域に割り当てる画素値の色相

$S_{x,y}$  : 微細領域に割り当てる画素値の彩度

$V_{x,y}$  : 微細領域に割り当てる画素値の明度

$L_{x,y}$  : 微細領域における線の集積量

$L_{max}$  : 線の集積量の最大値 (式 4.6)

$\Gamma$  : 変換関数の曲率を制御するパラメータ

$$L_{max} = \max_{x,y} L_{x,y} \quad (4.6)$$

式 4.5 からわかる通り算出された集積量の値が 0 のとき，割り当てられる画素値も 0 になる．そのため完成したグラフは図 4.6 に示したような黒い背景にグレーの線を描画したようなグラフとなる．先にも述べた通りガンマ変換において入力の値を 0~1.0 の範囲に収める必要があるため，式 4.5 に示されるように「ある領域における線の集積量」を「描画領域内の全微細領域中の集積量の最大値」で割ることによって正規化し用いる． $\Gamma$  の値を変更することによってトーンカーブの曲率が変化し，各微細領域の集積量と微細領域に割り当てられる画素値との関係性が決まる．

1.  $\Gamma$  の値を 1.0 より大きくすると変換関数のトーンカーブは曲率が負の曲線を描く．この変換関数を用いると線の集積量が少ない微細領域間の色のコントラストが強くなり，逆に線の集積量が多い微細領域間の色のコントラストは弱まる．
2.  $\Gamma$  の値を 1.0 より小さくするとトーンカーブの曲率が正の曲線を描く変換関数になる．この変換関数を用いると線の集積量が多い微細領域間の色のコントラストが強まり，線の集積量が少ない微細領域間の色のコントラストが弱くなる．

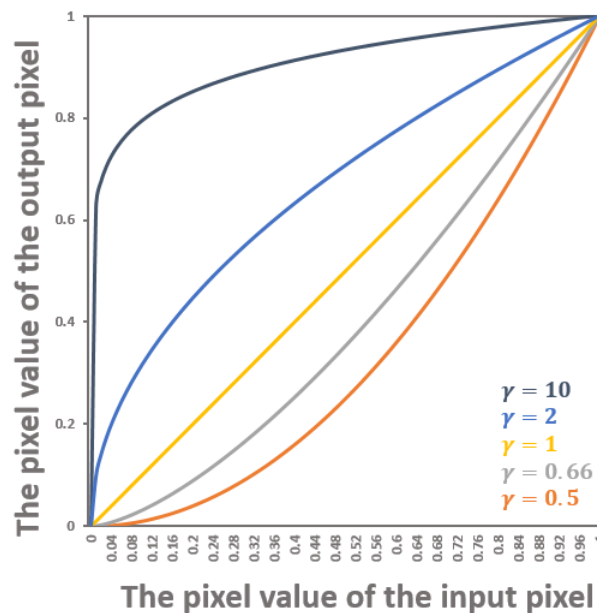


図 4.9: ガンマ変換の変換関数における  $\Gamma$  の値とトーンカーブの関係性．トーンカーブの線の色は図中右下の  $\Gamma$  の値を示す文字の色と対応している．横軸は入力画像の画素値，縦軸は出力画像の画素値とそれぞれ対応している．



#### 4.4.4 ガンマ変換を元にした変換関数の改良

前節ではガンマ変換を基に作成した変換関数を示した。ここでガンマ変換と我々が作成した変換関数の入力として用いる値の範囲の違いについて考える。ガンマ変換において、入力として用いることのできる値は0~255の256種類である。それに対して、線の集積量は全ての系列が完全に重なっていた場合には0と1の2種類、全ての系列が交わらない場合にはデータ系列の数だけ存在する。そのため規模の大きなデータを扱う際には多くの微細領域集団に対して同じ色が割り当てられることとなる。ところが、集積量の分布が図4.8のように集積量の分布が偏っている場合、使用されない色が大量に出来てしまうこととなる。この問題を解決するべく、我々は式4.5を改良し更にいくつかのパラメータを設定することで、ヒストグラムの形状に応じたより柔軟な色付けを可能にする。

最初にパラメータ設定後の式を提示する(式4.7)。

$$\begin{aligned} H_{x,y} &= 0 \\ S_{x,y} &= 0 \\ V_{x,y} &= 255 * \left( \frac{L_{x,y}}{L_{max} * x_{border}} - x_{trans} \right)^{\frac{1}{\Gamma}} + y_{trans} \end{aligned} \quad (4.7)$$

$H_{x,y}$  : 微細領域  $i$  に割り当てる画素値の色相

$S_{x,y}$  : 微細領域  $i$  に割り当てる画素値の彩度

$V_{x,y}$  : 微細領域  $i$  に割り当てる画素値の明度

$L_{x,y}$  : 微細領域  $i$  における線の集積量

$L_{max}$  : 線の集積量の最大値

$\Gamma$  : 変換関数の曲率を制御するパラメータ

続いて新たに設定した各パラメータについて解説する。

1.  $x_{border}$  は固有の色を割り当てる集積量の上限を設定するパラメータである(図4.10)。このパラメータは主に集積量の分布が左側に固まるようなデータを扱う際に用いることを想定している。集積量の多い微細領域集団にはユニークな色を割り当てず、分布が固まっている集積量の少ない領域集団に多く色を割り当ててコントラストを重点的に上げることが可能である。
2.  $x_{trans}$  はトーンカーブの  $x$  軸方向の平行移動量を表すパラメータである(図4.11)。このパラメータは主に集積量の少ない微細領域が非常に少なく、逆に集積量の多い微細領域が多くなるようなデータを扱う際に用いることを想定している。集積量の少ない微細領域集団にはユニークな色を割り当てず、集積量の多い領域集団のコントラストを上げることが可能になる。

3.  $y_{trans}$  はトーンカーブの  $y$  軸方向の平行移動量を表すパラメータである (図 4.12). このパラメータは系列数が多く集積量の範囲も大きくなるようなデータを扱う際に用いることを想定している.  $\Gamma$  の値を補正しても集積量の少ない微細領域の視認が困難なときにこのパラメータの値を調整し, 視認可能な状態にすることが可能である.

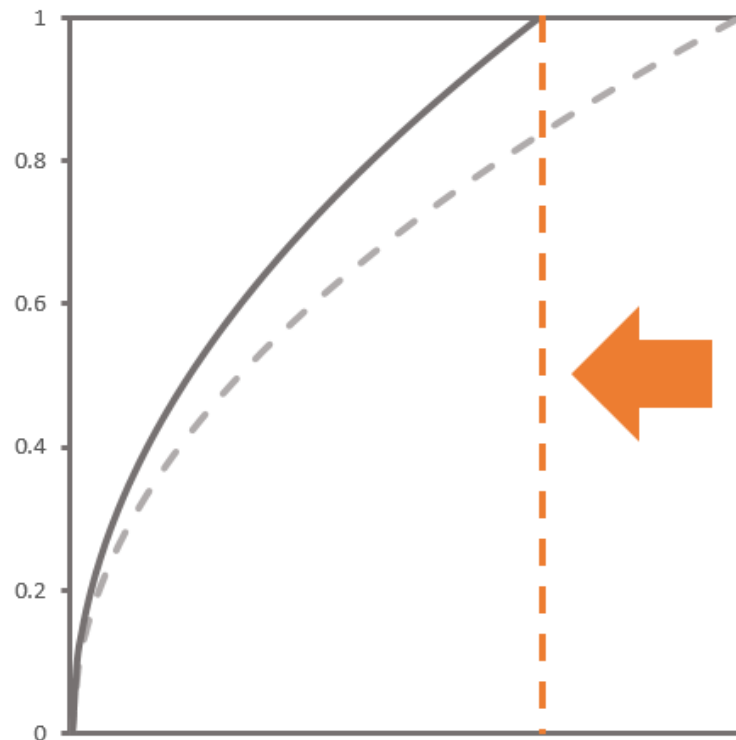


図 4.10:  $x_{border}$  の値を操作した時のトーンカーブの変化を示す図. 図の横軸は線の集積量に, 縦軸は微細領域に割り当てられる画素値にそれぞれ対応している.  $x_{border}$  はトーンカーブが頭打ちになる点を変更するためのパラメータである.

#### 4.4.5 画素値が上限を超えた微細領域集団に割り当てる画素値の決定

新たに追加したパラメータの値を調整することで, 割り当てられる画素値が割り当て可能な上限である 255 を超えてしまう微細領域集団が現れる. これらの領域集団に対して割り当てる色について, 本研究では 2 通りの決定方法を考える. 1 つ目は上限を越えた微細領域は全てグレースケールの明度の最大値 (255) で着色する手法である. シンプルな色の決定方法であるが, 上限を越えた微細領域集団のみが際立って目立ってしまう可能性がある. 2 つ目は上限を越えた微細領域の周囲の領域の色をぼかすことで上限を越えた領域を上手くその他の領域になじませる手法である. この決定方法は写真の加工などに用いられている技術である

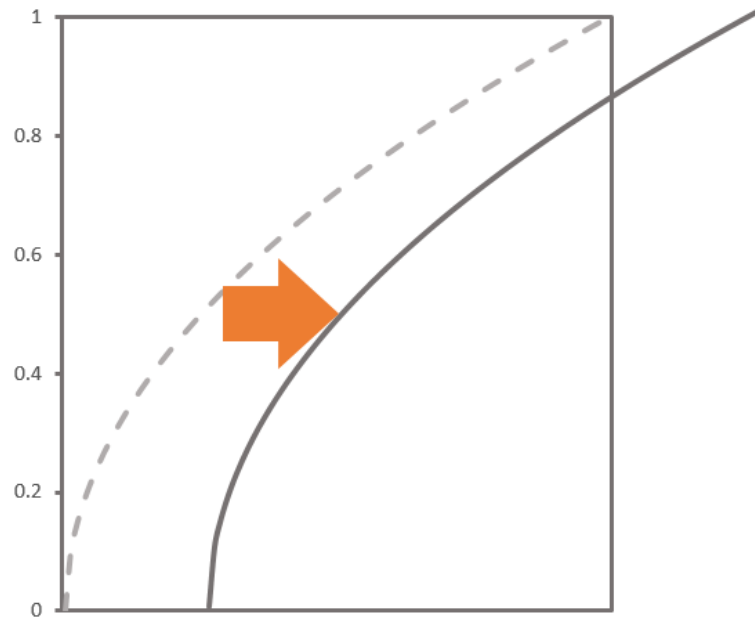


図 4.11:  $x_{trans}$  の値を操作した時のトーンカーブの変化を示す図. 図の横軸は線の集積量に, 縦軸は微細領域に割り当てられる画素値にそれぞれ対応している.  $x_{trans}$  は  $x$  軸方向の平行移動量を表すパラメータである.

トーンマッピングの手法を利用している [18]. まず, 図 4.13 の上の図のように割り当てられる画素値が 255 を超えてしまった微細領域のみで構成される描画領域を作成する. 描画領域のサイズはグラフの描画領域と同じでその他の微細領域の色は黒色である. 続いて, この画像に対してガウシアンフィルタをかける (図 4.13). ガウシアンフィルタをかけることで割り当てられる画素値が 255 を超えてしまった微細領域集団の周囲の領域の色が黒色から白に近づく. 最後にこの画像を元の画像に対して加算合成する. この手順をふむことで割り当てられる画素値が 255 を超えてしまった微細領域集団の周囲の領域をぼかし, 一部の領域が極端に目立ってしまうことを防ぐ狙いがある.

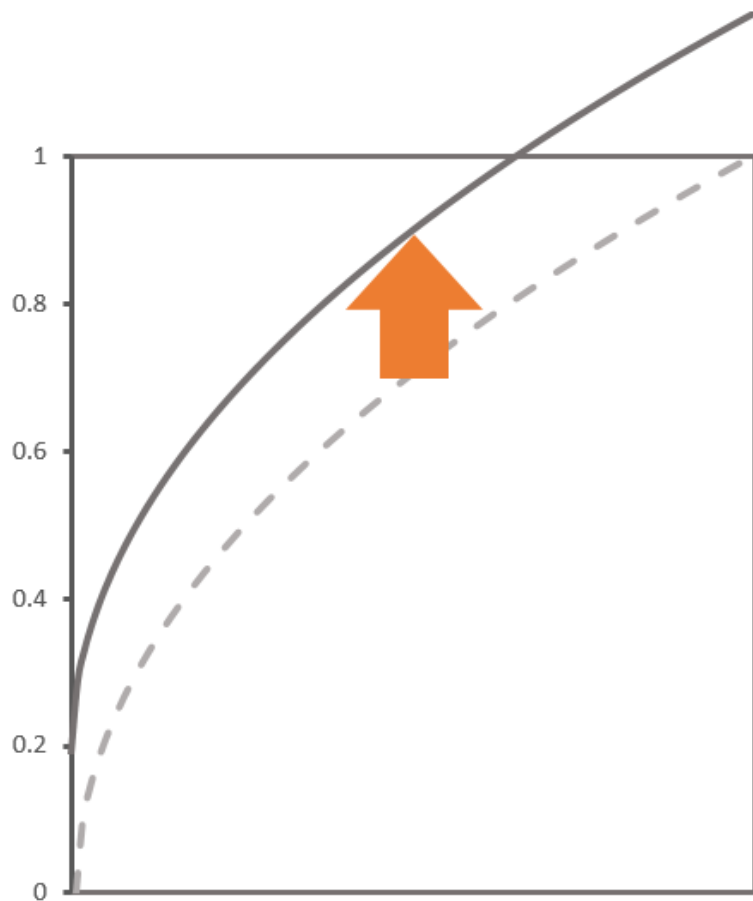
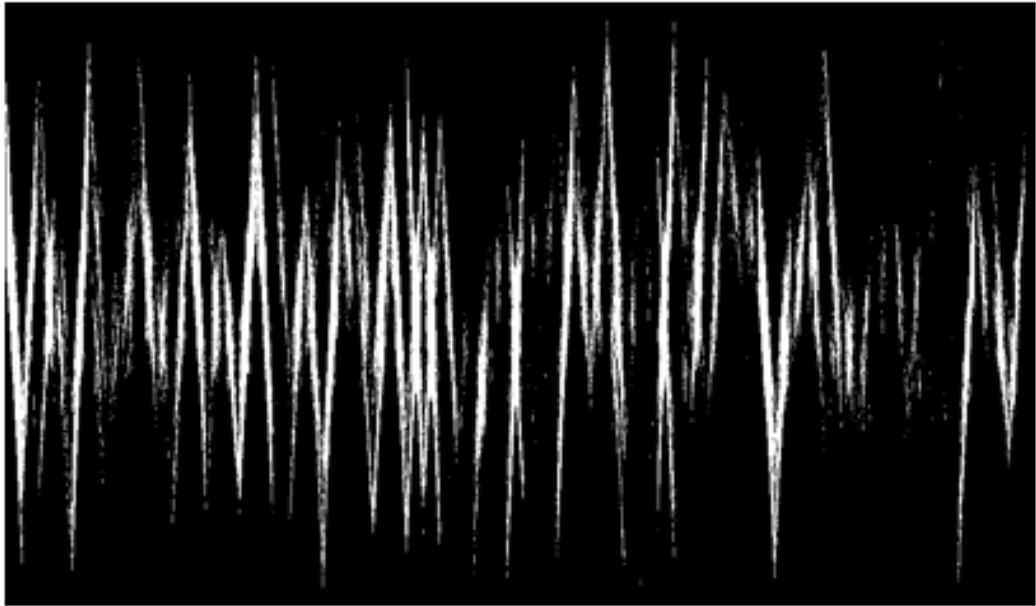


図 4.12:  $y_{trans}$  の値を操作した時のトーンカーブの変化を示す図。図の横軸は線の集積量に、縦軸は微細領域に割り当てられる画素値にそれぞれ対応している。 $y_{trans}$  はトーンカーブの  $y$  軸方向の平行移動量を表すパラメータである。



↓ Gaussian filter

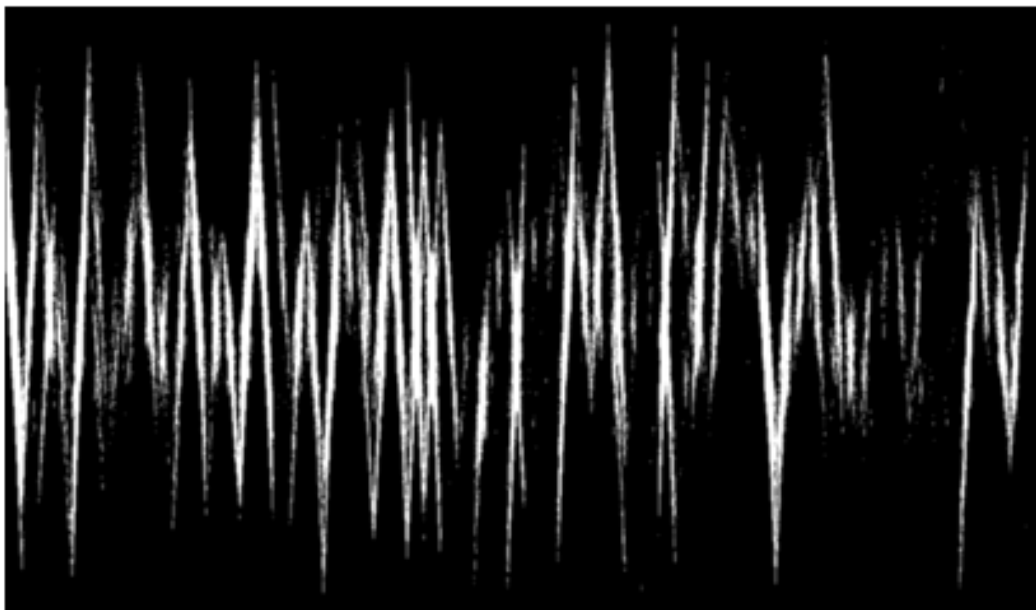


図 4.13: 上図：割り当てられる画素値が 255 を超えてしまった領域のみで構成される描画領域。下図：上図にガウシアンフィルタをかけた結果。

## 第5章 実装

我々は提案手法の有効性を確かめるために折れ線グラフ描画プログラムを開発した。さらに提案手法において、データに応じた最適なトーンカーブを分析者が集積量のヒストグラムと折れ線グラフを見比べられながら設定できる分析ツールを開発した。

### 5.1 折れ線グラフを描画するプログラムの開発

最初に開発したプログラムにおける折れ線グラフの作成手順を示す。

1. 描画領域の画素数と同じサイズの2次元配列を用意する
2. ある系列を描画領域に描いた際に色が塗られる画素を算出する
3. 算出された画素に対応する2次元配列の要素に+1する
4. 2, 3を全てのデータ系列に対して行う
5. 2次元配列に格納された集積量の値を変換関数を用いて画素値に変換し、対応する画素に色付けを行う

開発したプログラムではディスプレイの1つの画素を提案手法における1つの微細領域とみなす。画素ごとの線の集積量を算出するためにまず、描画領域の画素数と同じサイズの2次元配列を用意する。そしてすべての  $v \in V$  に対して系列  $v$  の線が引かれている画素に対応した2次元配列の要素に1を加算する処理を行う。最後に求めた集積量の値に基づいて画素に色付けを施す。

#### 5.1.1 線が描画される画素の算出

本研究では、線が描かれる画素の導出に最もポピュラーな直線描画アルゴリズムであるブレゼンハムのアルゴリズム [16][23] を採用している。ブレゼンハムのアルゴリズムは始点と終点の座標を入力とし2点間を結ぶ線分が通る画素の座標を算出するアルゴリズムである。ここではブレゼンハムのアルゴリズムを用いた系列  $v$  の隣り合う項目  $v_i, v_{i+1}$  間を結ぶ線分が通る画素の座標の求め方を説明する。前提条件としてブレゼンハムのアルゴリズムでは左上端を原点  $(0, 0)$  とし、画素の座標は右下方向に進むにつれて大きくなる。また、各画素の中心が

画素の整数座標と正確に対応しているとする。この前提条件の元で、項目  $v_i, v_{i+1}$  間を直線で繋いだときに線が通る画素を求める。

まず、 $v_i$  を描画領域上にプロットしたときの座標  $(x_i, y_i)$  は式 5.1, 及び式 5.2 によって表される。

$$x_i = e * i + 1 \quad (5.1)$$

$$y_i = h - margin - (v_i - v_{min}) * s \quad (5.2)$$

$$e = \left\lceil \frac{w}{m - 1} \right\rceil$$

$$s = \frac{h - margin * 2}{v_{max} - v_{min}}$$

$e$  : 描画領域上における項目間の距離

$s$  : 各項目の値のスケーリングを行うための定数

$v_{max}$  :  $v \in V$  における  $v_i$  の最大値

$v_{min}$  :  $v \in V$  における  $v_i$  の最小値

$h$  : 描画領域の縦の画素数

$w$  : 描画領域の横の画素数

$m$  :  $v \in V$  の項目数

$margin$  : 描画領域の上下のマージン

ブレゼンハムのアルゴリズムは整数座標を入力としたアルゴリズムである。そのため、2つの式を用いて項目  $v_i, v_{i+1}$  間を結ぶ線分の始点の座標  $(x_i, y_i)$  及び終点の座標  $(x_{i+1}, y_{i+1})$  を算出した後、値を四捨五入し整数座標に変換した始点の座標  $(X_i, Y_i)$  及び終点の座標  $(X_{i+1}, Y_{i+1})$  を入力として用いる。出力となる線分が通る画素の座標は  $x \in N$  の値を  $x = X_i$  を初期値として  $x = X_{i+1}$  まで1ずつ増やしながらか関数に代入し、それぞれの  $x$  に対応する  $y$  の値を算出することで求まっていく。このとき線分の傾きにあたる  $\frac{Y_{i+1} - Y_i}{X_{i+1} - X_i}$  の値に応じて線分が通る画素の座標が以下のように変わる [23].

1.  $\left| \frac{Y_{i+1} - Y_i}{X_{i+1} - X_i} \right| \leq 1$  のとき  
 $(x, y)$  を線分が通る画素の座標とみなし、対応する2次元配列の要素に1を加算する(1つの  $x$  座標の整数値毎に1つの画素が塗られる [17])
2.  $\left| \frac{Y_{i+1} - Y_i}{X_{i+1} - X_i} \right| > 1$  のとき  
 $(y, x)$  を線分が通る画素の座標とみなし、対応する2次元配列の要素に1を加算する(1つの  $y$  座標の整数値ごとに1つの画素が塗られる [17])

以上の処理を  $1 \leq i \leq m - 1$  を満たすすべての  $i \in N$  に対して行うことである系列  $v$  の折れ線グラフを描いた時に線が引かれる画素の座標がすべて求まる。さらにこの処理をすべての  $v \in V$  に対して行い、最終的な2次元配列の要素を対応する画素の集積量  $L_{x,y}$  とする。

## 5.2 提案手法を用いた折れ線グラフ描画ツールの開発

開発した折れ線グラフ描画プログラム，ガンマ変換の関数における  $1/\Gamma$  の値等のパラメータは予め指定しておく必要がある。そのため，思い通りの折れ線グラフが描かれるまでコードを書き換えて折れ線グラフを描画する作業を繰り返さなければならない。本研究では分析者がトーンカーブを操作しながら特徴把握により適した折れ線グラフを見つけ出すことができるような折れ線グラフ描画ツールを開発した(図 5.1)。

ツールは3つの部分から構成される。グラフ描画部分，変換関数を調整するインタフェース部分，変換関数のトーンカーブ及び集積量の分布のヒストグラムを表示する部分である。グラフ描画部分には前節で解説した描画手順に従って折れ線グラフが描かれる。残りの2つの部分についてこれから解説する。

### 5.2.1 トーンカーブとヒストグラムの表示

ツールの左下には変換関数のトーンカーブに集積量の分布のヒストグラムを重ねて表示する(図 5.2)。2つを重ねて表示することによって分析者は現在の変換関数を用いることでグラフ中のどの位の量の微細領域にどの色が割り当てられるのかを直感的に把握することが可能になる。

横軸は集積量を0~1.0の範囲に正規化したものである。軸の左端は1，軸の右端は集積量の最大値である。縦軸はトーンカーブとヒストグラムでそれぞれ異なる意味を持っている。トーンカーブの縦軸は微細領域に対して割り当てられる画素値である。画素値の範囲は0~255だが，それを0~1.0の範囲に正規化し，軸に用いている。ヒストグラムの縦軸は画素数を表している。これもヒストグラムの高さ(画素数)の最大値を用いて0~1.0の範囲に正規化し，軸の値として用いる。

### 5.2.2 インタフェース部分

開発したツールにおいて，分析者は変換関数のパラメータを変更し各画素に対する色の割り当てられ方を変更する。こうしてより特徴が見えやすい形に折れ線グラフの見目を調整することが可能になる。この変換関数のパラメータの変更に利用するのがインタフェース部分である。インタフェース部分は4種類のスライダと1種類のラジオボタンによって構成される。4種類のスライダは変換関数におけるパラメータの値を変更するためのものである。インタフェースを操作することによって変更できるのは式 4.7 中の  $\frac{1}{\Gamma}$ ,  $x_{border}$ ,  $x_{trans}$ ,  $y_{trans}$  のパラメータである。それぞれのパラメータの値を操作することによって分析者は折れ線グラフ



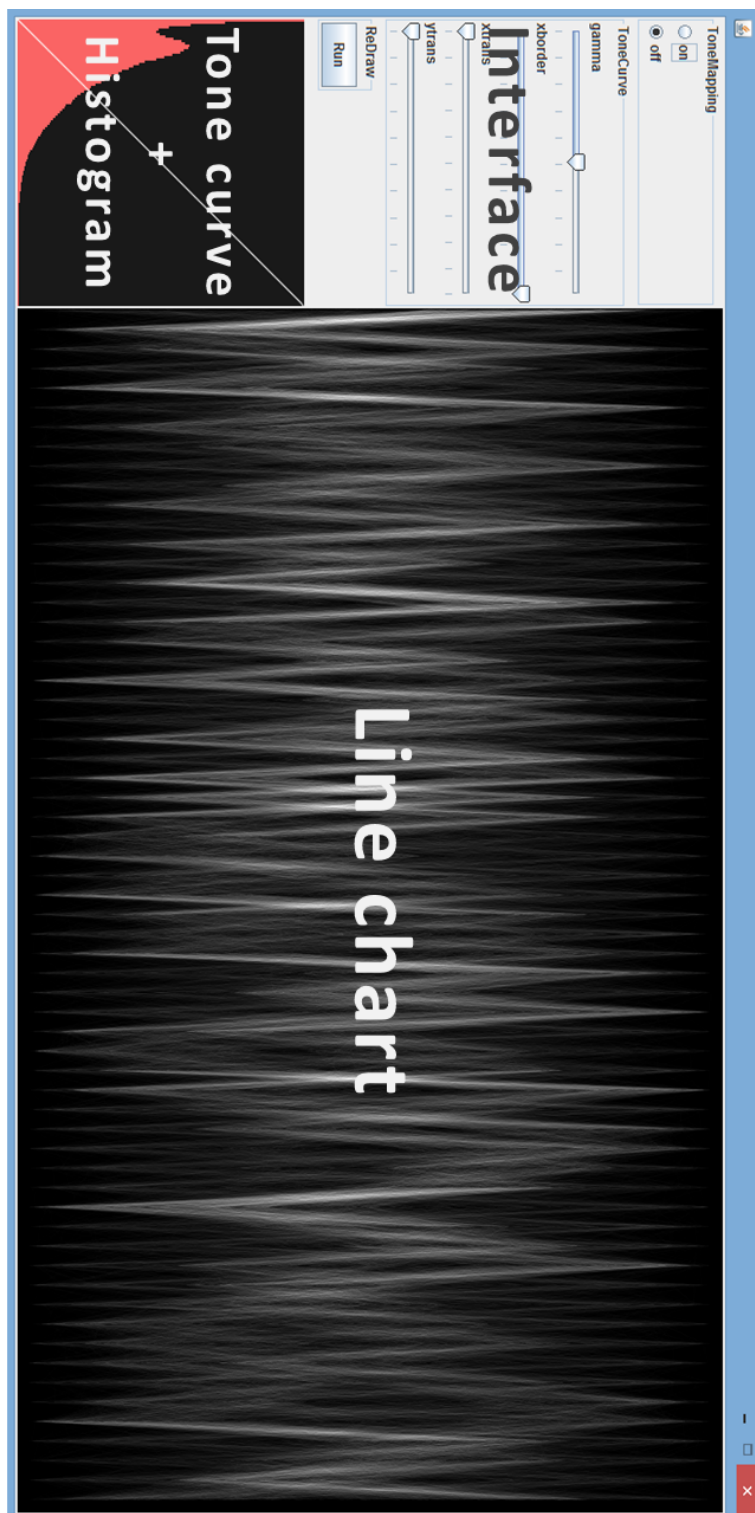


図 5.1: 開発したツール.

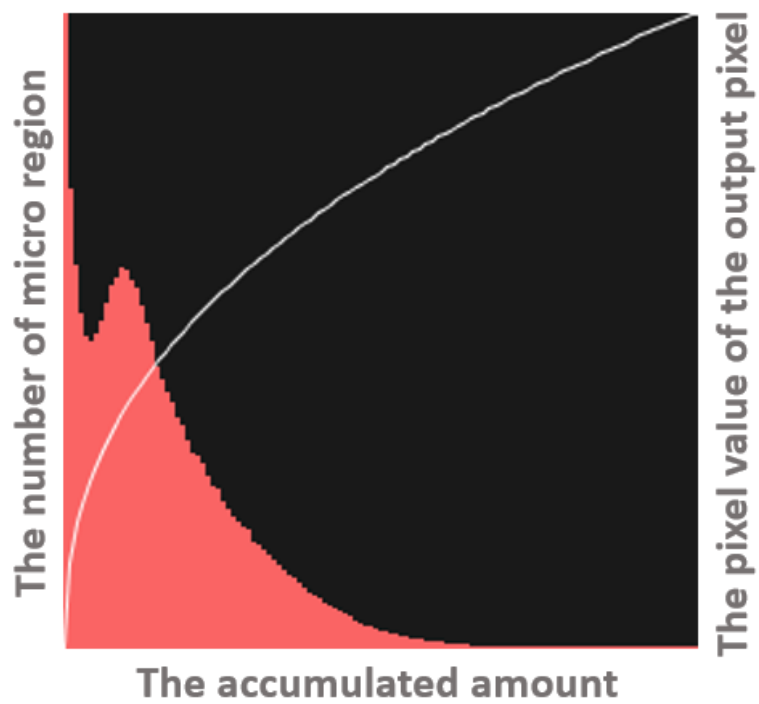


図 5.2: ツール上に表示されるトーンカーブとヒストグラム。トーンカーブ、ヒストグラム共に横軸は集積量と対応している。また、トーンカーブにおける縦軸は微細領域に割り当てられる画素値と、ヒストグラムにおける縦軸は微細領域の数とそれぞれ対応している。

の見た目を調整する。 $\frac{1}{\Gamma}$ の値は0.01から2.0の範囲を0.01刻みで、 $x_{border}$ ,  $x_{trans}$ ,  $y_{trans}$ の値は0.01から1.0の範囲を0.01刻みで調整することができる。スライダを動かして変換関数を変更するとそれに連動してトーンカーブの形状が変わり、折れ線グラフが新たに設定された変換関数を用いて再描画される。分析者は折れ線グラフとトーンカーブ、そしてヒストグラムの3つを見比べながら4種類の項目をスライダを用いて調整する。

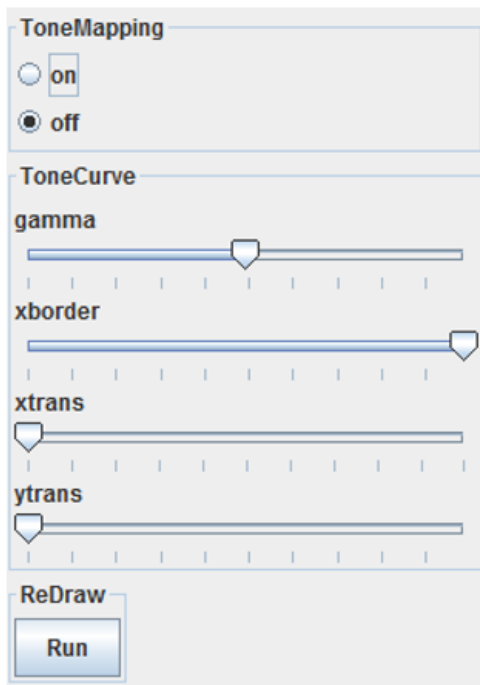
ラジオボタンは割り当てられる画素値が255を超えてしまった微細領域集団に割り当てる色を決めるに辺り、トーンマッピングの技術を適用するか否かを設定するためのラジオボタンである。ラジオボタンが“OFF”の状態のときは割り当てられる画素値が255を超えてしまった微細領域全てに対して白色を割り当てる。ラジオボタンが“ON”の状態のときには領域に対してトーンマッピングを行う。

描画ツール起動時、4種類のパラメータ及びラジオボタンは以下のように設定されている。

- $\frac{1}{\Gamma} = 1$
- $x_{border} = 1.0$
- $x_{trans} = 0$
- $y_{trans} = 0$
- ラジオボタン：“off”

この変換関数をトーンカーブで表すと図5.3のような線形のトーンカーブになる。ユースケースではこの状態を初期状態として各パラメータを調整していく。

## Interface



## Tone curve + Histogram

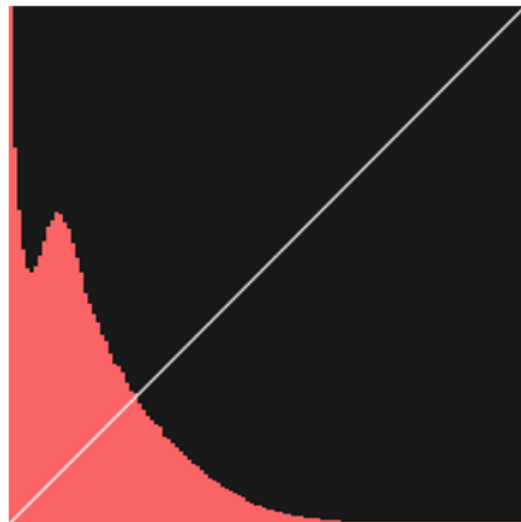


図 5.3: ツール起動時のインターフェース部とトーンカーブ及びヒストグラム表示部.

## 第6章 ユースケース

我々は開発した分析ツールを用いて数種類の実在するデータを可視化した。ユースケースを作成する目的は2つある。1つ目の目的は基礎的な手法を用いて作成した折れ線グラフと提案手法を用いて作成した折れ線グラフを比較し提案手法を適用することによって大規模データの特徴把握が容易になっていることを示すことである。本研究では不透明な線と半透明な線を用いて作成した折れ線グラフ、2種類のグラフと提案手法を用いて作成したグラフを比較し提案手法の有意性を示す。2つ目の目的はトーンカーブとヒストグラムの関係性を観測することである。両者の形状比較から最適な着色の制御方法の決定に向けた知見が得られることを期待する。

### 6.1 時系列データ

最初に、折れ線グラフが適用されることの多い時系列データの折れ線グラフを提案手法を用いて可視化する。ユースケースでは、気象庁が公開している2種類の気候に関する時系列データを用いる。

#### 6.1.1 降水量データ

気象庁が公開している全国各地の観測地点における降水量のデータ [20] に対して提案手法を適用した。各データ系列は観測地点を、各項目は観測月を表している。また、系列の数は1227、項目の数は61(2010年1月~2015年1月)である。図6.1の上の図はこのデータを不透明な折れ線で描いたグラフである。グラフの縦軸は降水量と、横軸は観測月とそれぞれ対応しており、全国の降水量の月毎の変化を表している。図6.1の上の図を見ると線が複数重なり合っている箇所が潰れ、特徴が把握できなくなっている。続いて図6.1の中央の図は半透明な線を用いて描いた折れ線グラフである。密な領域の線の様子が視認可能な程度まで線の不透明度を下げると上の図では見えた疎な領域の線の視認ができなくなっている。

今度はこのデータを提案手法を用いて可視化する。まず集積量と色とを線形に対応させると図6.2の上の図のようになる。密な領域の線の重なりあう様子がはっきりと確認できる。しかしその一方で半透明な線の不透明度を低くしたとき以上に密な領域以外の線の視認が困難になってしまっている。ツール左下のヒストグラムに着目すると集積量の分布が少ない領域に固まっており、そこに背景色に限りなく近い色が割り当てられてしまっている事がわかる。

ヒストグラム及びトーンカーブから得た情報から、まず  $\frac{1}{F}a$  の値を補正することによってこの疎な領域の線を視認可能にする。ヒストグラムを見ながらスライダを操作し、分布の密集している領域に背景色との区別がつく色が割り当てられるよう  $\frac{1}{F}$  の値を大きくしていくと見えなかった線が見えてくる (図 6.2 の中央の図)。ここで  $\frac{1}{F}$  の値を大きくしすぎてしまうと今度は不透明な線を用いて作成したグラフのように密な領域の様子が読み取れないグラフになってしまうため、疎な領域の線がはっきりと視認可能になったところ ( $\frac{1}{F}=0.3$ ) で  $\frac{1}{F}$  の値の調整を終える。続いて、密な部分の様子を更に詳細に見るため、集積量のスケールを小さくすることで分布が固まっている画素集合のコントラストを上げる ( $x_{border} = 0.4$ )。すると密な部分の様子がより詳細に見えてくる (図 6.2 の下の図)。こうして完成したグラフが図 6.3 である。図 6.1 を見ると上の 2 種類のグラフでは同時に見るができなかった密な領域と疎な領域の特徴両方が提案手法を適用した下のグラフでは視認可能になっていることが分かる。続いてこのグラフに対してトーンマッピングを適用したグラフが図 6.4 である。あまり目立った効果は現れていないが図 6.4 のアノテーションがついた箇所を見るとトーンマッピングを行うことで特に集積量の多い微細領域集団が目立って見えている。

### 6.1.2 気温データ

同じく気象庁が公開している全国の気温のデータ [20] に提案手法を適用する。各系列は観測地点を、項目は観測月を表しており、系列の数は 823、項目の数は降水量のデータと同じく 61 (2010 年 1 月～2015 年 1 月) である。図 6.5 の上の図は不透明な線を用いて描いた折れ線グラフである。グラフの縦軸は気温と、横軸は観測月とそれぞれ対応しており、全国の気温の月毎の変化を表している。図 6.5 の上の図を見ると中心に大規模な系列集団があり、さらに極端に気温が高い値を取る系列と逆に低い値をとる系列が数本あることが確認できる。一方で大規模な系列集団の中で線ががどのように重なり合っているのかはこの図からは確認することができない。続いて図 6.5 の中央の図は半透明な線を用いて作成した折れ線グラフである。系列集団の中の線の重なりあいの様子が視認できるようになり、特に気温の高い月、低い月における系列のばらつき具合を観察することができる。一方で、上の図では視認することが可能だった極端な値をとる数本の線が見えなくなった。

このデータに対し、提案手法を適用する。ツール起動時の状態が図 6.6 の上の図である。降水量のデータに対して手法を適用したときと同じく線形の変換関数では疎な領域の線を視認することが難しい。ヒストグラムを見ると降水量のデータと同じくこのデータも分布に偏りがあり集積量の少ない領域が多い。そこで疎な領域の線が視認可能になる程度まで  $\frac{1}{F}$  の値を補正する ( $\frac{1}{F} = 0.3$ )。トーンカーブから、全ての微細領域に対して十分判別可能な色が割り当てられていると判断しパラメータの操作を中止する (図 6.6 の下の図)。完成したグラフが図 6.7 である。グラフの上下にある特異な動きをするレコードが見え、かつ中央部の密集した線集団の様子も観察することができる。全ての微細領域の画素値が 255 以内に収まっているためこのグラフに対するトーンマッピングは行わない。

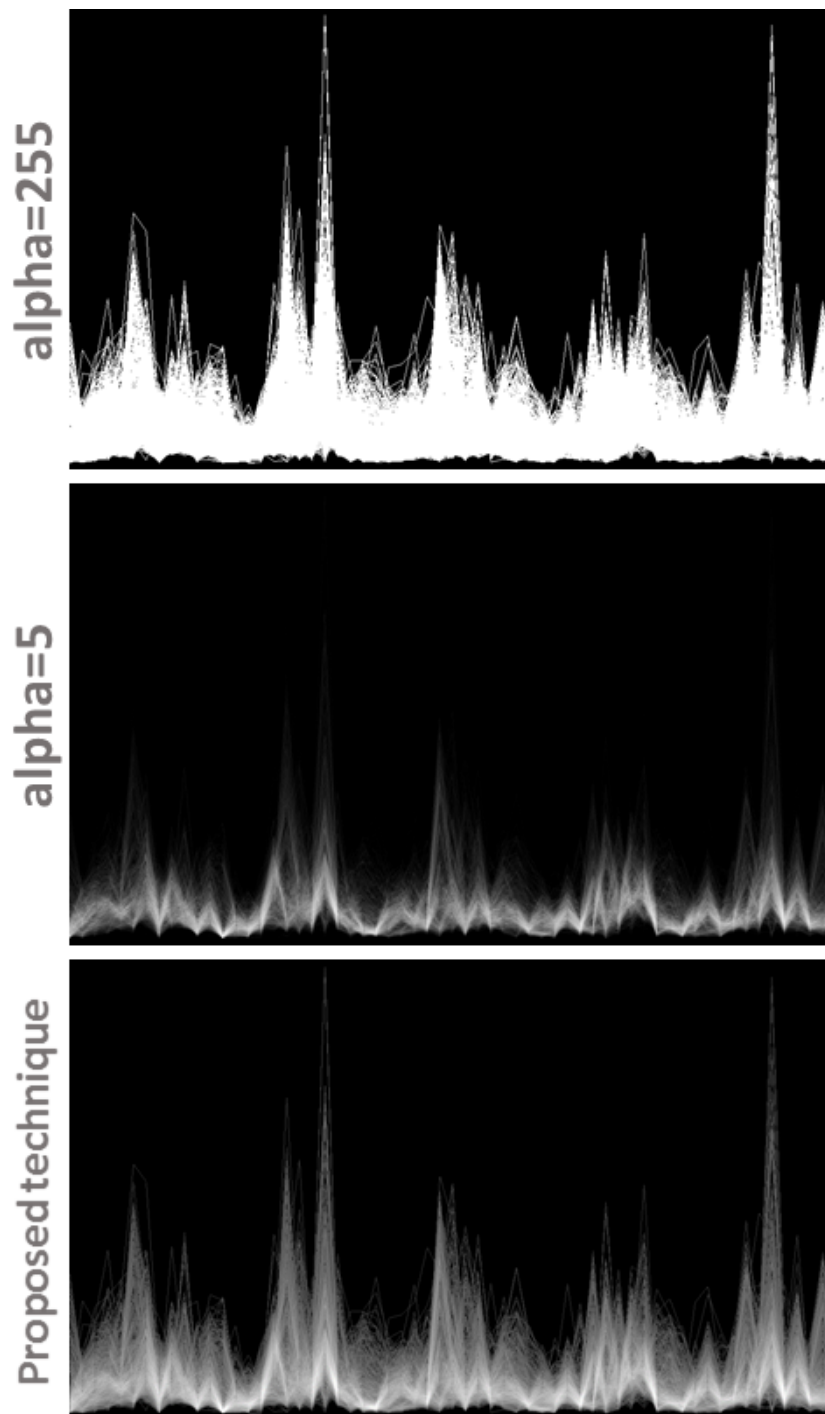


図 6.1: 3 種類の手法を用いて描いた降水量データの月ごとの変化を表す折れ線グラフ. 各グラフは計 1227 本の線からなり, ひとつの折れ線はひとつの観測地点を表す. また, グラフの横軸は観測月と, 縦軸は降水量とそれぞれ対応している.

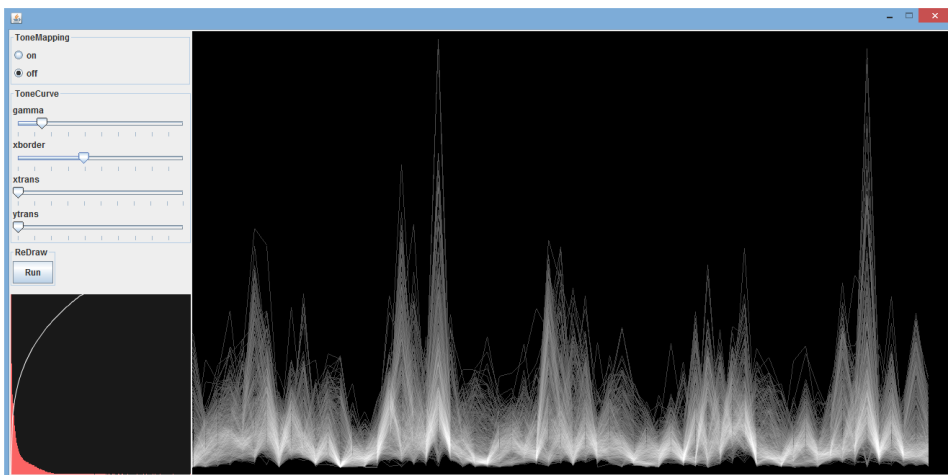
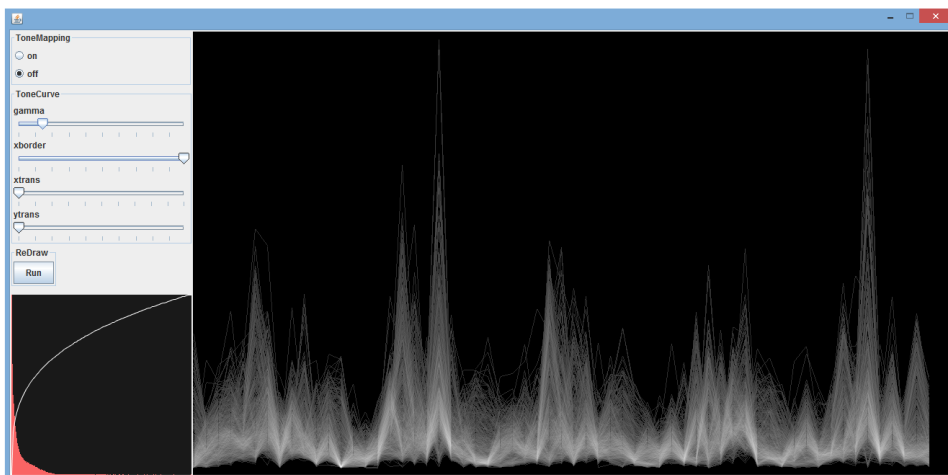
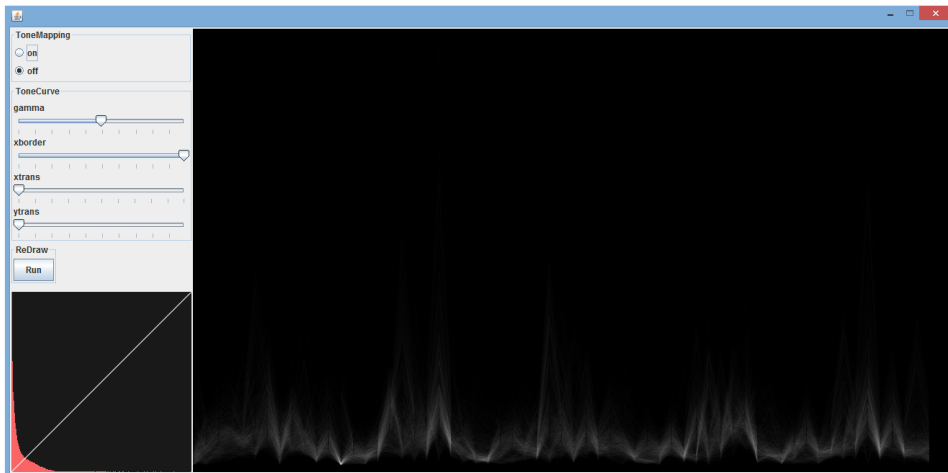


図 6.2: 降水量データの折れ線グラフにおける変換関数のパラメータ調整過程.



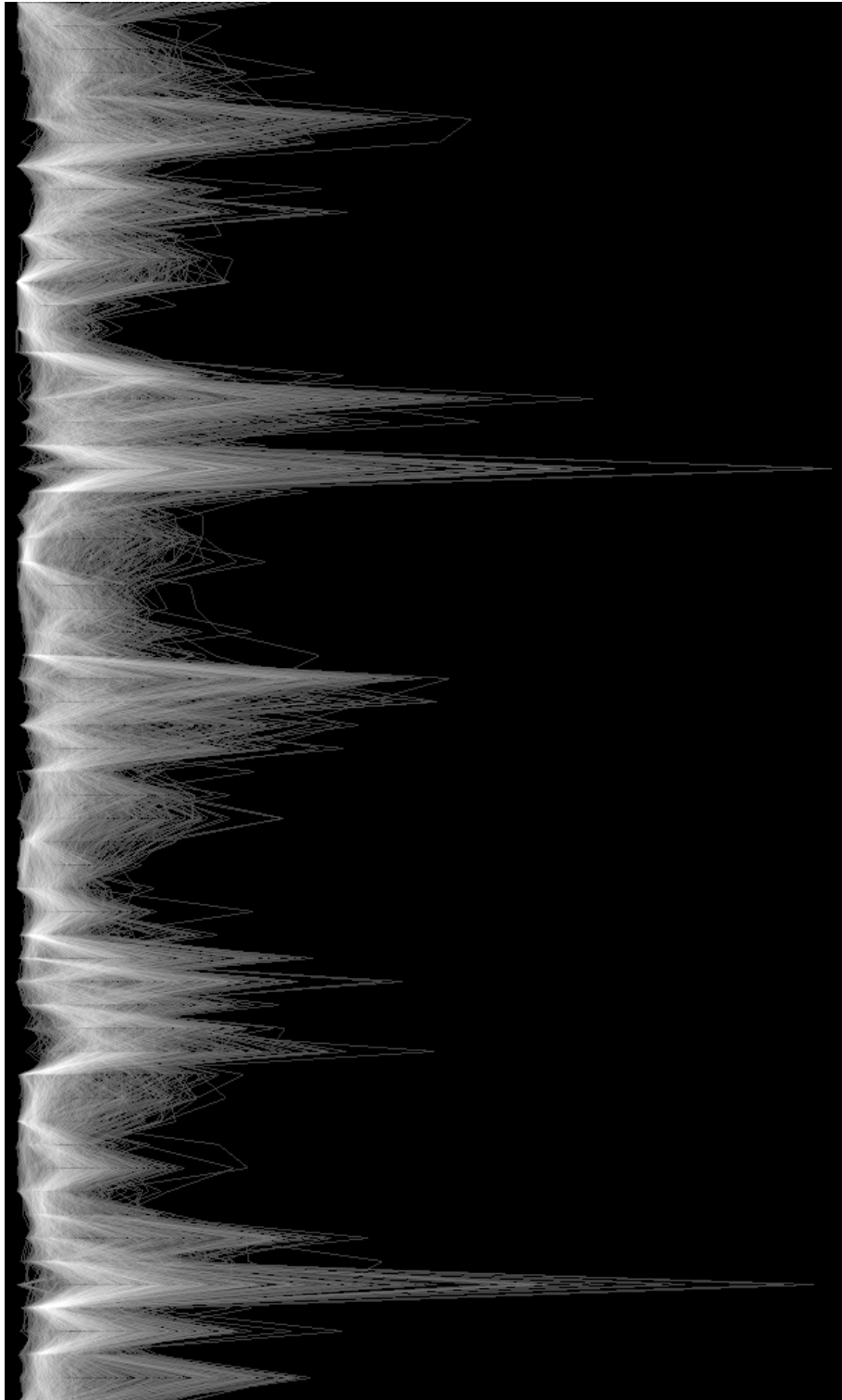


図 6.3: 提案手法を適用した降水量データの折れ線グラフ.  
 $\frac{1}{T} = 0.3, x_{border} = 0.4, x_{trans} = 0, y_{trans} = 0$

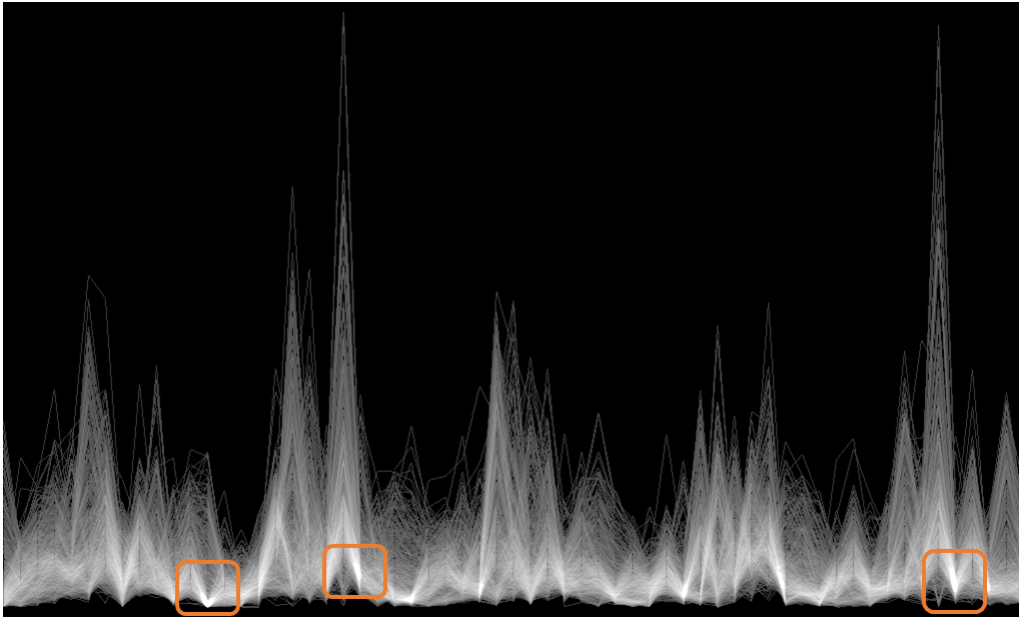


図 6.4: 提案手法を適用した降水量データの折れ線グラフに対してトーンマッピングを行った結果.

## 6.2 PCP への適用

我々が提案する折れ線グラフ描画手法は多変量データの可視化手法である PCP(Parallel Coordinates plot) に対しても適用することが可能である. そこで PCP に対して手法を適用した際の有効性を確かめるべく, 2 種類の時系列データではない多変量データを 2 種類用意し, 提案する手法を用いて可視化した.

### 6.2.1 シカゴの犯罪履歴データ

1 つ目の多変量データはシカゴの犯罪履歴データ [21] である. 1 つの系列が 1 つの犯罪を表しており, 各項目はその犯罪に関する情報を表している (表 6.1). 系列数は 291976 と前の 2 つのデータと比較して大きく, 8 つの項目からなる. 図 6.8 の上の図は不透明な線を用いて描いたこのデータの PCP である. PCP の各軸は表 6.1 に示した犯罪に関する情報ひとつひとつと対応している. 特に図の左側において密な領域の構造が見えづらくなっていることが分かる. このデータを半透明な線を用いて PCP にしたものが図 6.8 の中央の図である. 系列の数が多いため, 不透明度を設定可能な最低の値まで下げても密な領域の線の重なり合う様子を読み取ることができない. 加えて不透明度を下げた影響で疎な領域の線も見えなくなっている.

そこで今度は, 提案手法を適用した PCP を作成する. まず, 線形な変換関数を用いると殆どの線が視認不可能になってしまっている. ヒストグラムを確認すると分布にこれまでのデー

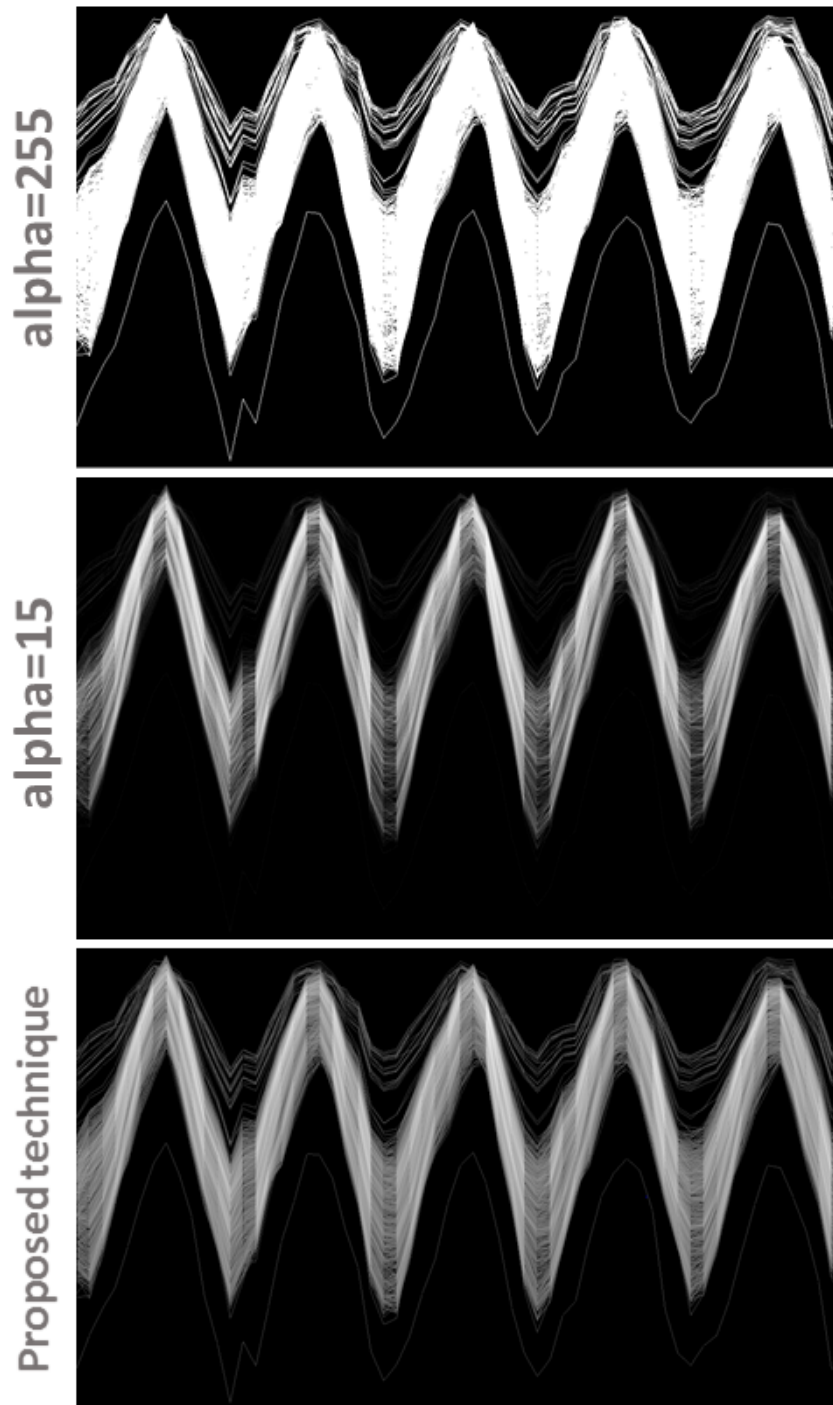


図 6.5: 3 種類の手法を用いて描いた全国の気温の月毎の変化を表す折れ線グラフ. 計 823 本の折れ線からなり, ひとつの折れ線はひとつの観測地点を表す. また, グラフの縦軸は気温と, 横軸は観測月とそれぞれ対応している.

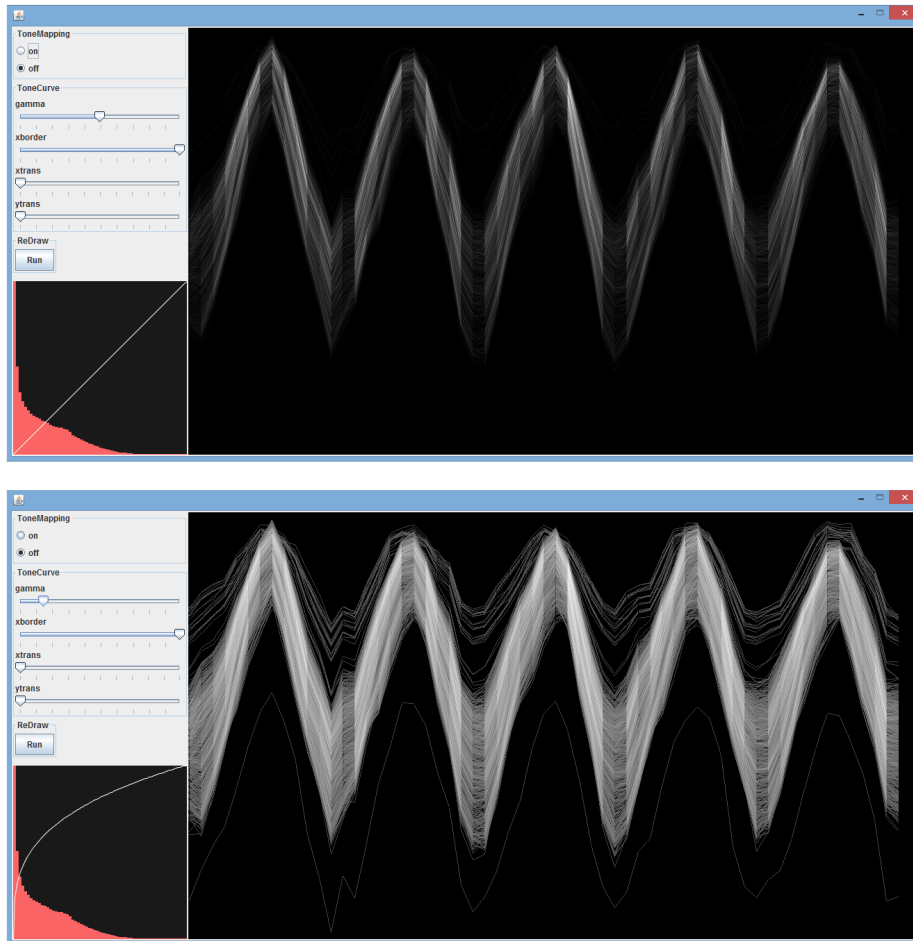


図 6.6: 気温データの折れ線グラフにおける変換関数のパラメータ調整過程.

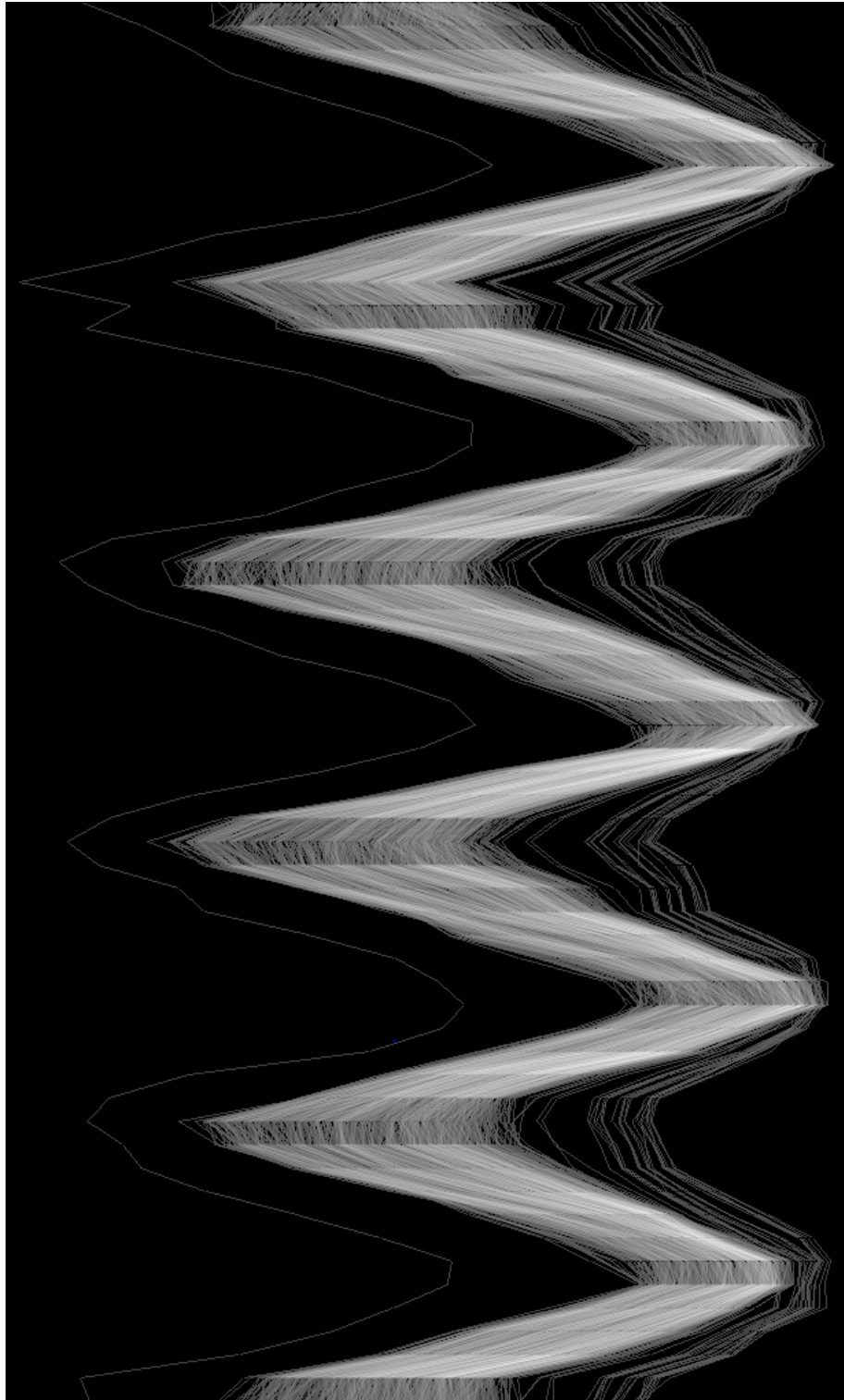


図 6.7: 提案手法を適用した気温データの折れ線グラフ.  
 $\frac{1}{T} = 0.3, x_{border} = 1.0, x_{trans} = 0, y_{trans} = 0$

タよりも更に極端な偏りがあることが分かる 6.9. そこで、これまでの2種類のデータと同じく最初に  $\frac{1}{F}$  の値を設定し ( $\frac{1}{F} = 0.1$ ), その後  $x_{border}$  の値を調整して ( $x_{border} = 0.1$ ) 集積量が少ない微細領域集団の色のコントラストを上げる (図 6.9). 図 6.9 を見ると不透明な線, 及び半透明な線のみを用いて描画した PCP と比較して疎な領域と密な領域, 双方の線が視認可能になっている. しかし, 図 6.9 右側の橙色のアノテーションを付けた箇所において, 密な領域の色のコントラストが乏しく, 線の識別が難しい. 原因は  $\frac{1}{F}$  の値を小さく設定し過ぎたことで集積量の多い微細領域集団に十分な数の色が割り当てられなかったことによる. しかし,  $\frac{1}{F}$  の値をこれ以上大きく設定してしまうと図 6.9 の緑色のアノテーションを付けた領域の線等の視認が困難になってしまう. そこで  $y_{trans}$  の値を変更することを考える. 最初に  $y_{trans}$  の値を図 6.9 の緑色のアノテーションを付けた領域が視認可能になるまで ( $y_{trans} = 0.3$ ) 上げる (図 6.10 の上図). この状態で  $\frac{1}{F}$  の値を調整する ( $\frac{1}{F} = 0.2$ , 図 6.10 の中央の図). 最後に  $x_{border}$  の値を調整し集積量が少ない微細領域集団の色のコントラストを上げる ( $x_{border} = 0.15$ , 図 6.10 の下図). こうして完成した PCP が図 6.11 である. 図 6.8 を見ると疎な領域と密な領域, 双方の線の様子が視認できるようになっていることに加え, 提案手法を用いた事によって不透明な線, 及び半透明な線のみを用いて描画した PCP では見えなかった密な領域における線の重なり合う様子が読み取れるようになっている. 続いて, 作成した PCP に対してトーンマッピングを行った PCP が図 6.12 である. 降水量データに対して適用した際にはあまり効果が現れなかったが, 今回はトーンマッピングを行うことで同じ傾向を持った系列集団が目立って見えるようになっている.

表 6.1: シカゴの犯罪履歴データの各項目の内容

項目名	項目の内容
Primary Type	犯罪の種類
Location Description	犯罪の起きた場所の区分
Arrest	逮捕の成否
Domestic	家庭内で起きた犯罪か否か
Beat	警察の巡回区域
District	管区
Ward	区
Community	地域

## 6.2.2 大学のスコアカードのデータ

2つ目の多変量データはアメリカの大学のスコアカードのデータから学問領域ごとの賞の受賞割合のデータを用いる [22]. 各系列は大学を表しており, 項目は学問領域を表している. また, 系列の数は 7804, 項目の数は 37 である. これまでの3つのデータと同じく, 最初に不透明な線を用いて作成した PCP (図 6.13 の上図) と半透明な線を用いて作成した PCP (図 6.13 の

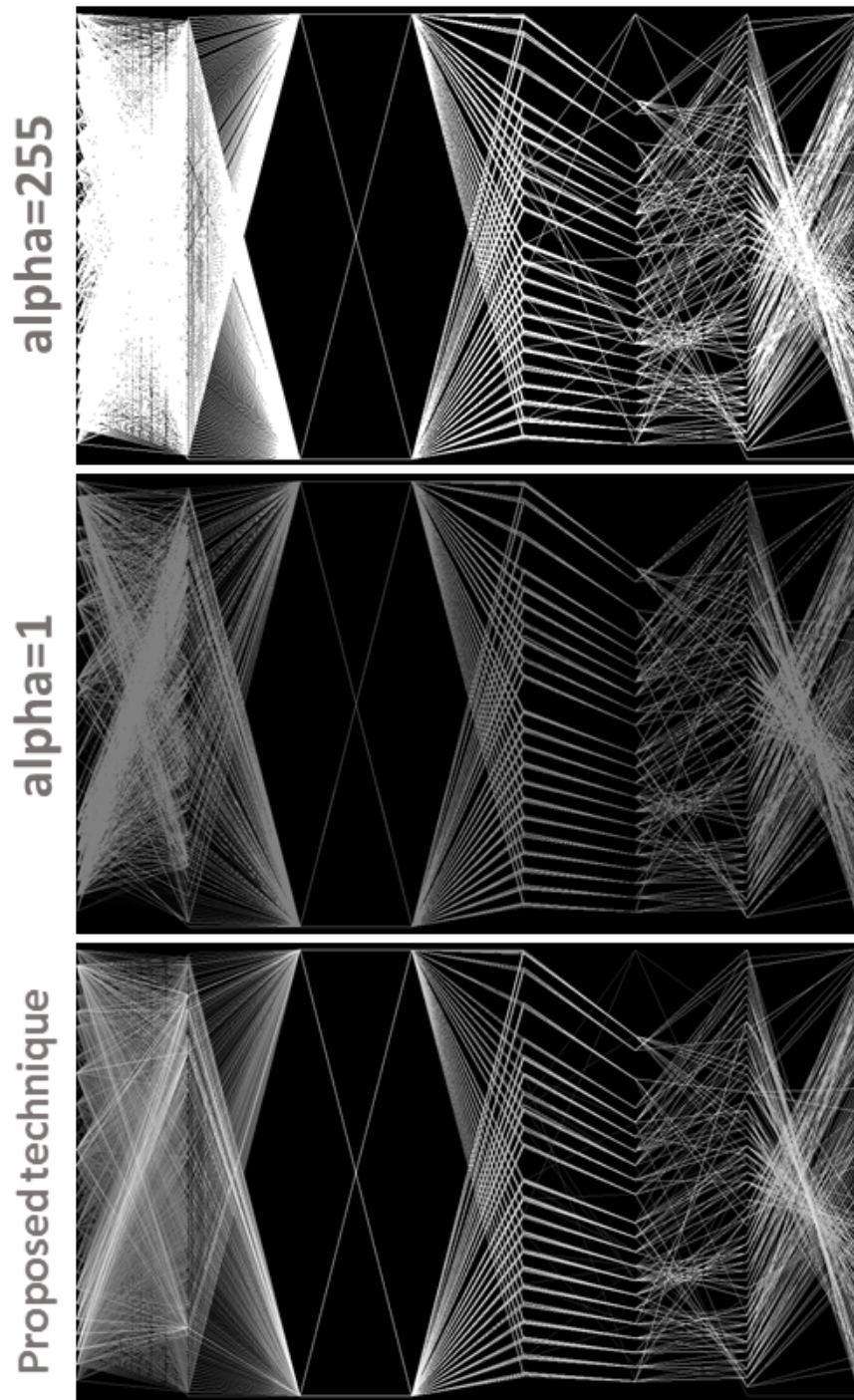


図 6.8: 3 種類の手法を用いて描いたシカゴの犯罪履歴データの PCP. 計 291976 本の折れ線からなり、ひとつの折れ線はひとつの 1 件の犯罪を表している. また, PCP の各軸は犯罪に関する情報ひとつひとつと対応している.

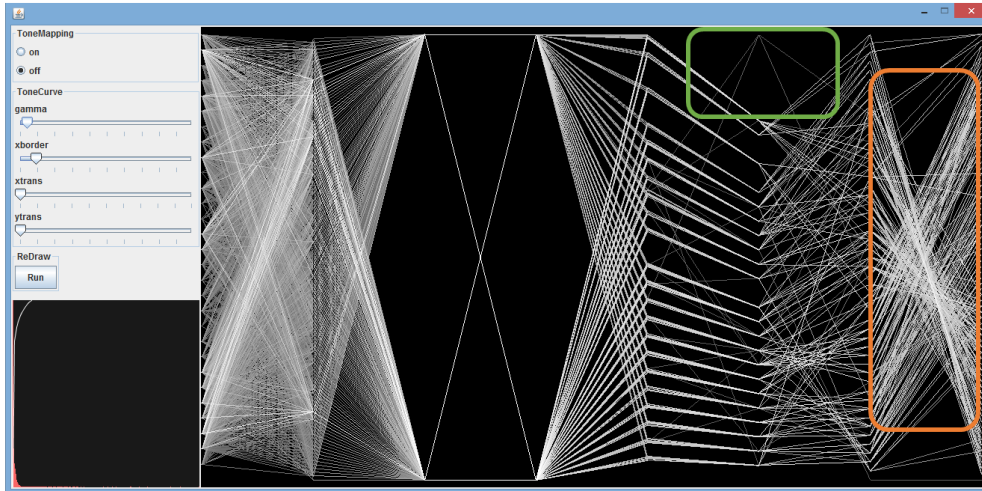


図 6.9: 提案手法の変換関数において、 $\frac{1}{\Gamma}$  と  $x_{border}$  の値のみを変更し作成した PCP. 右側の橙色のアノテーションを付けた箇所において、密な領域の色のコントラストが乏しい。

中央の図)を示す。PCPの各軸は学問領域と、軸の値は学問領域毎の賞の受賞割合と対応している。疎な領域な線が微かに見える程度まで不透明度を下げた状態においても右端の密な領域が真っ白に塗りつぶされてしまっている。これまで扱ってきたデータと同じようにまず  $\frac{1}{\Gamma}$  の値を変更し、疎な領域の線が見えるようにした後に  $x_{border}$  の値を変更する。完成した PCP が図 6.14 である。PCP 下部の線が密な領域の色のコントラストが少なく、ぼやけたように見えてしまっている。集積量のヒストグラムに着目すると分布の偏りが非常に極端なことがよく分かる。シカゴの犯罪履歴データと同じく疎な領域の線が見えるまで  $\frac{1}{\Gamma}$  の値を小さくすると密な領域の色のコントラストが弱まってしまうのである。そこで今回も最初に  $y_{trans}$  の値を疎な領域の線が視認可能になるまで増やし ( $y_{trans} = 0.2$ , 図 6.15 の上図), その上で  $\frac{1}{\Gamma}$  と  $x_{border}$  の値を調整する ( $\frac{1}{\Gamma} = 0.3$ ,  $x_{border} = 0.05$ , 図 6.15 の中央, 及び下図)。こうして図 6.16 に示されるような PCP を作成した。続いてこの PCP に対してトーンマッピングを行ったものを図 6.17 に示す。同じ傾向を持った系列集団が目立って見えるようになっているが、一方で線が必要以上に太く見えてしまっている。



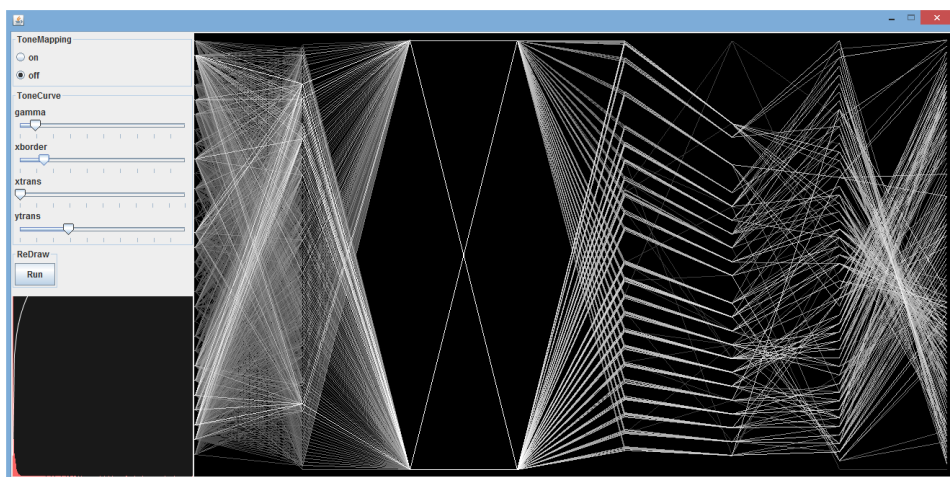
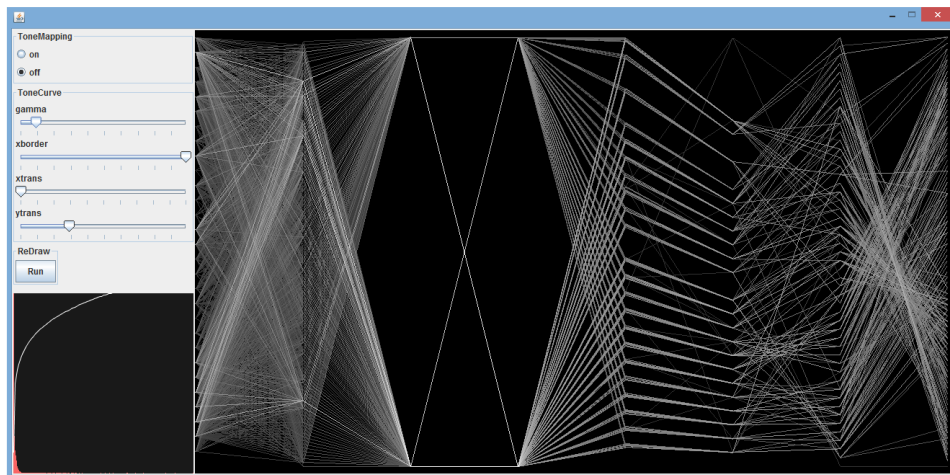
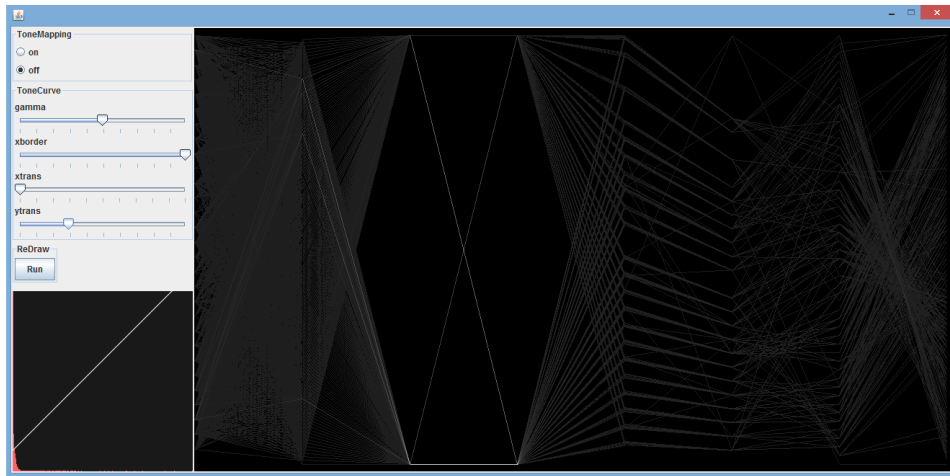


図 6.10: シカゴの犯罪履歴データの PCP における変換関数のパラメータ調整過程.

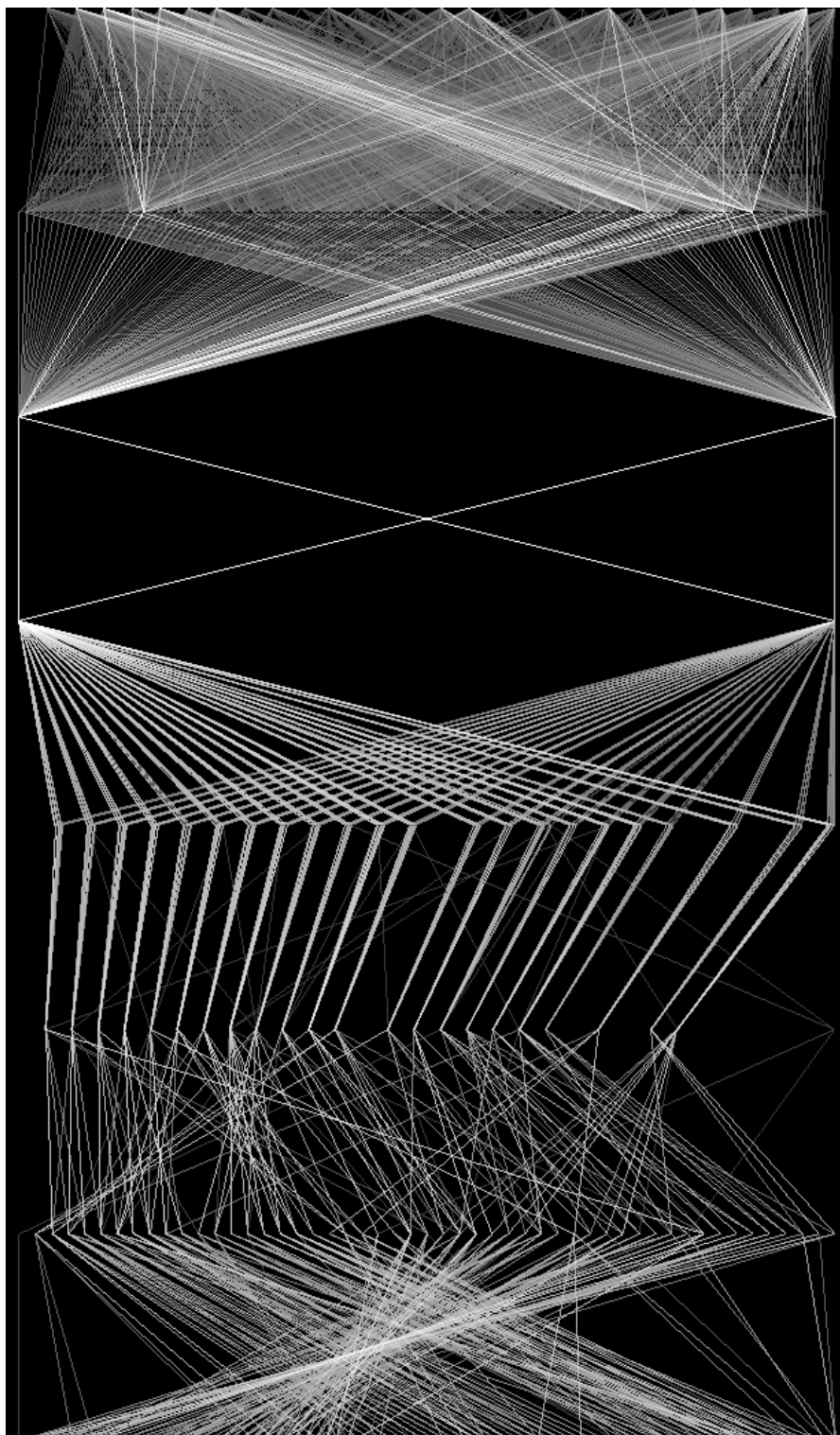


図 6.11: 提案手法を適用したシカゴの犯罪履歴データの PCP.  $\frac{1}{\Gamma} = 0.2$ ,  $x_{border} = 0.15$ ,  $x_{trans} = 0$ ,  $y_{trans} = 0.3$

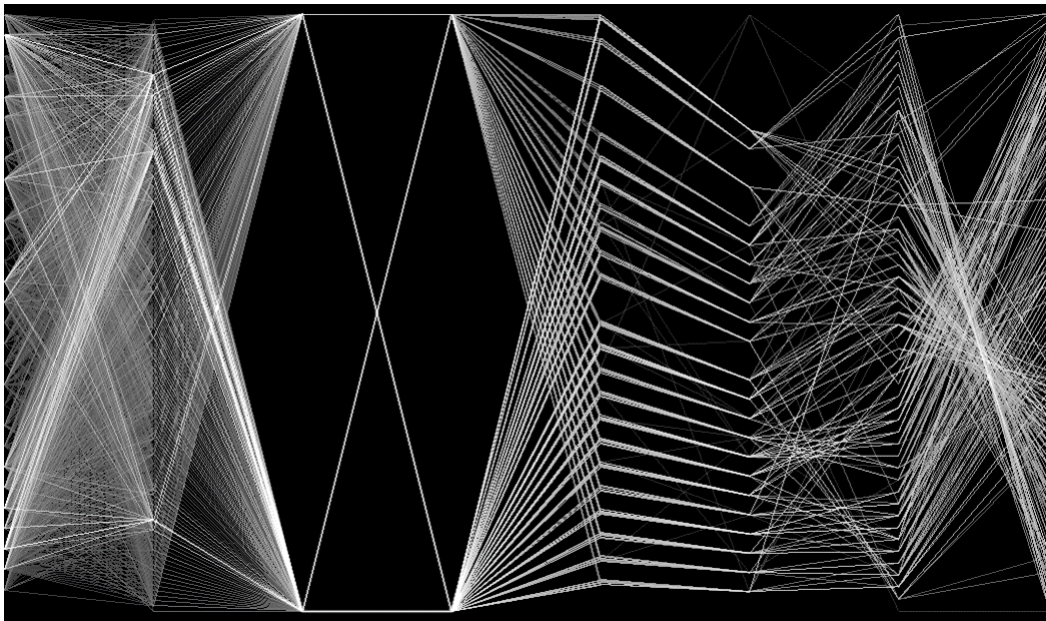


図 6.12: 提案手法を適用した犯罪履歴データの PCP に対してトーンマッピングを行った結果.

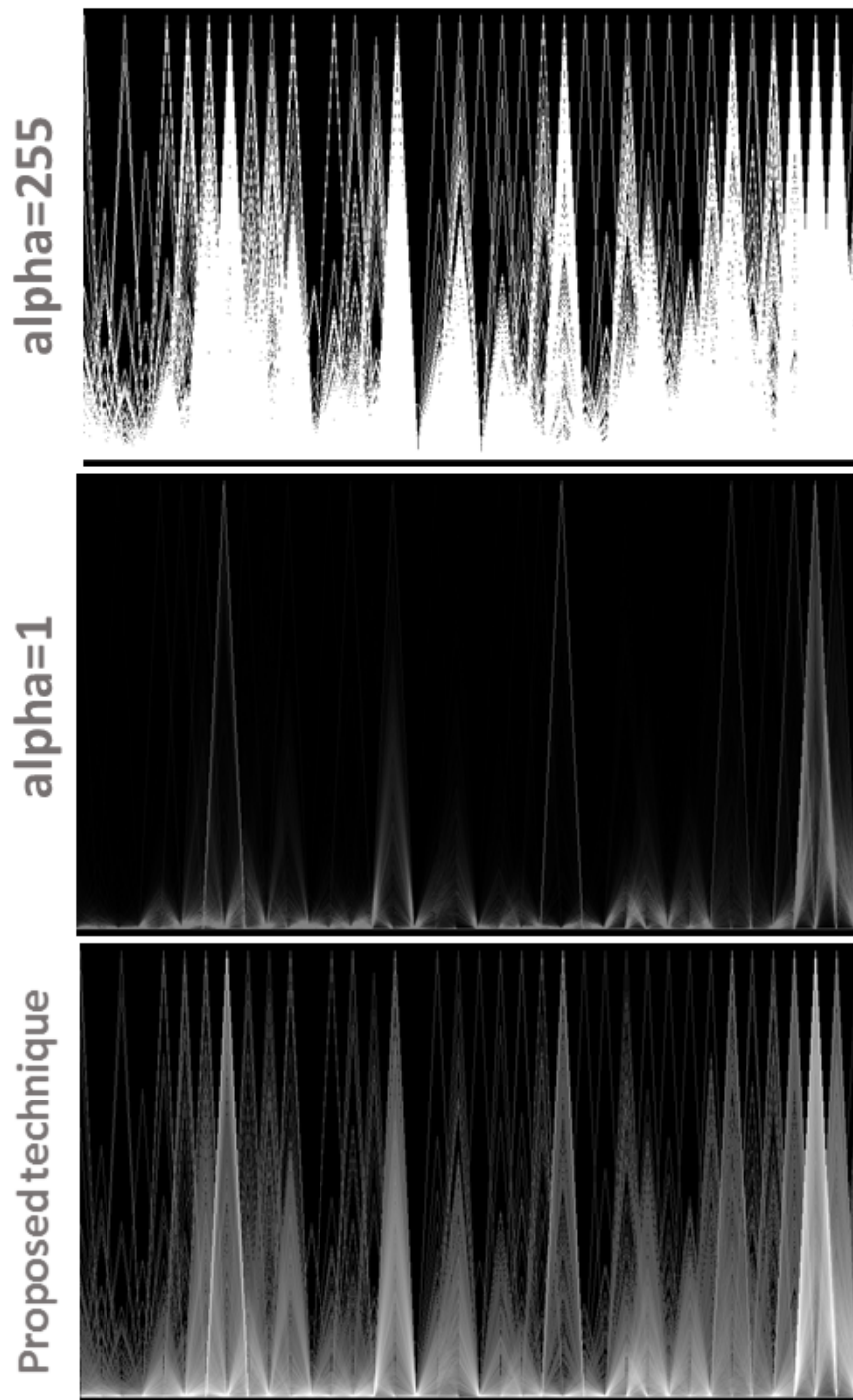


図 6.13: 3 種類の手法を用いて描いた大学のスコアカードのデータの PCP. 計 7804 本の折れ線からなり, ひとつの折れ線はひとつの大学を表している. PCP の各軸は学問領域と, 軸の値は学問領域毎の賞の受賞割合と対応している.

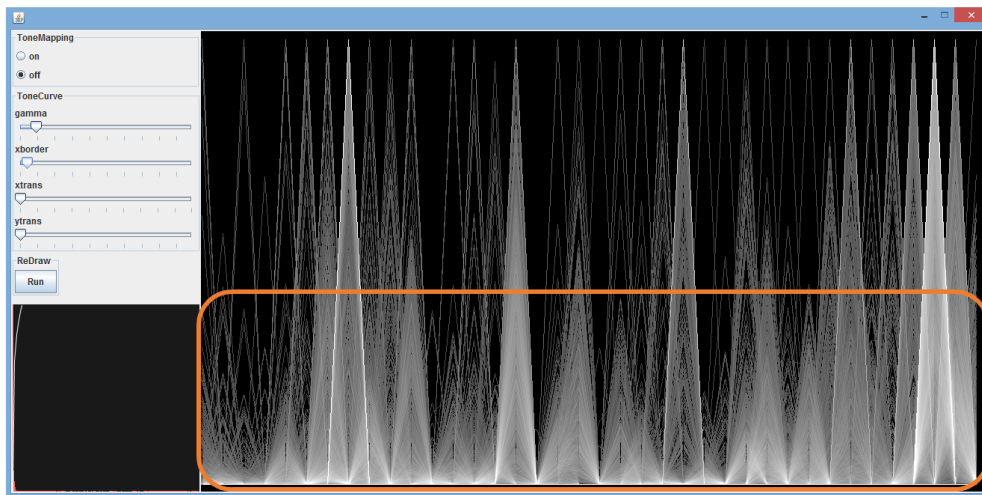


図 6.14: 提案手法の変換関数において,  $\frac{1}{\Gamma}$  と  $x_{border}$  の値のみを変更し作成した PCP. 橙色のアノテーションを付けた箇所において, 密な領域の色のコントラストが乏しい.

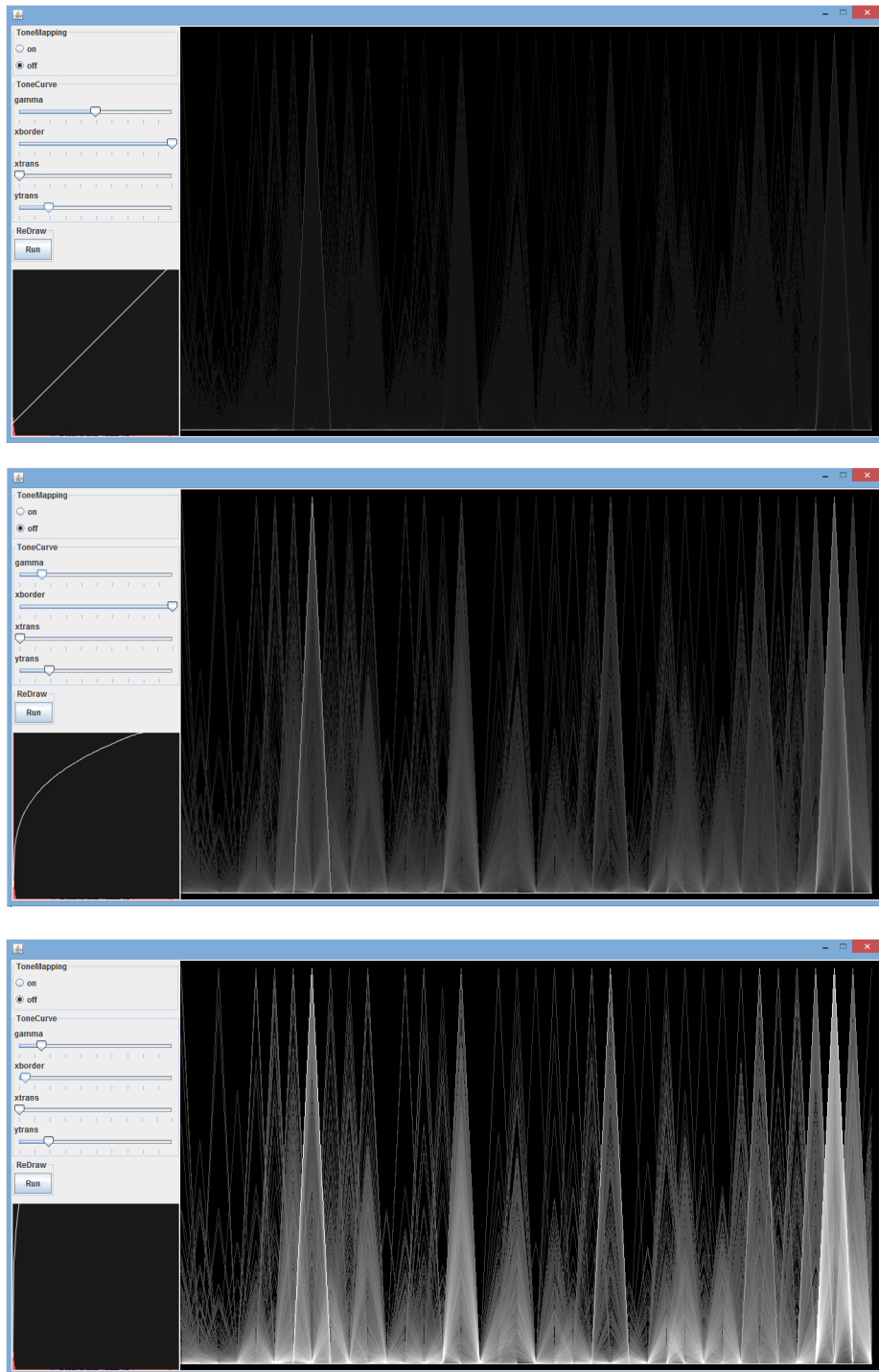


図 6.15: 大学のスコアカードのデータの PCP における変換関数のパラメータ調整過程.

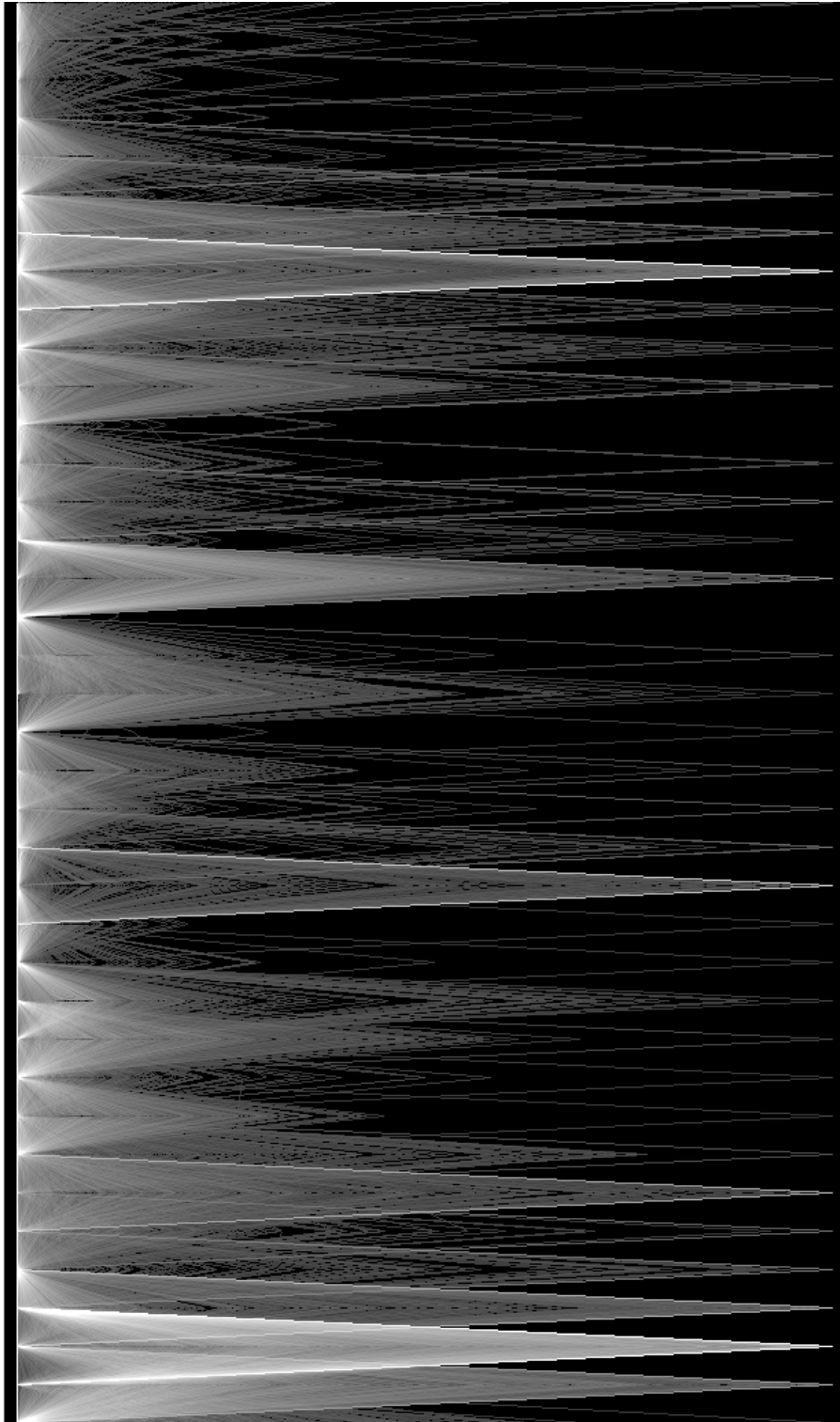


図 6.16: 提案手法を適用した大学のスコアカードのデータの PCP.  $\frac{1}{\Gamma} = 0.3$ ,  $x_{border} = 1.0$ ,  $x_{trans} = 0.05$ ,  $y_{trans} = 0.2$

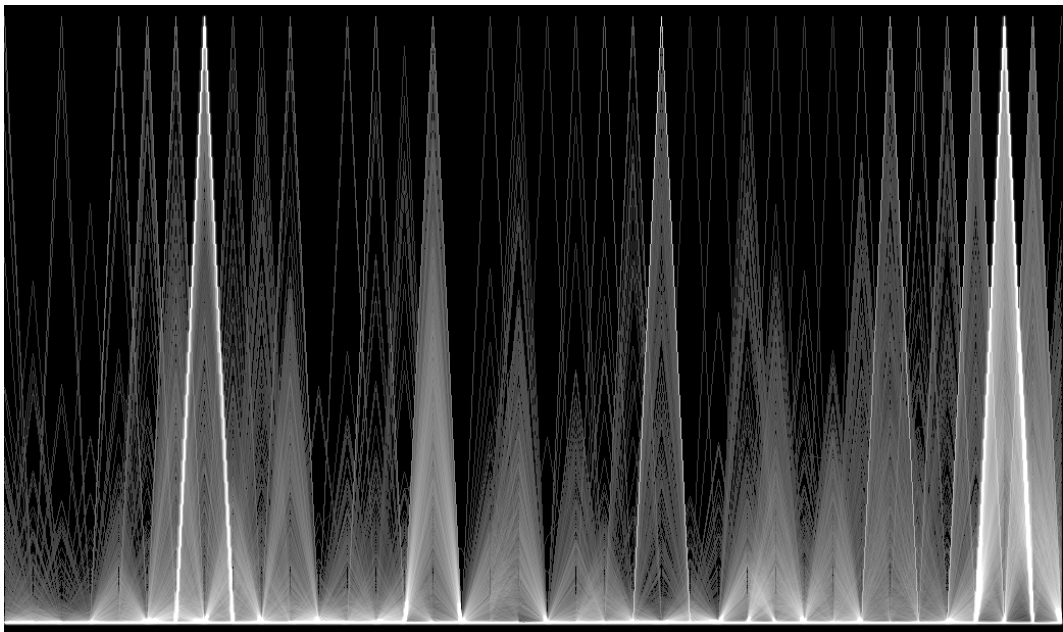


図 6.17: 提案手法を適用した大学のスコアカードのデータの PCP に対してトーンマッピングを行った結果.



## 第7章 考察

本章ではユースケースを元に手法の有効性に関する検証を行い、検証を基にし提案手法の今後の課題について考える。

### 7.1 提案手法の有効性

我々は作成したユースケースにおいて、時系列データと時系列でない多変量のデータ、2種類のデータに対して提案手法を適用し、その有効性を示そうとした。結果、いずれのデータに対して提案手法を適用した場合にも疎な領域の線が見えるようにしつつ、密な領域における線の重なり合いの様子も見えるような折れ線グラフ、及びPCPを作成することができた。2種類のPCPについては不透明度を限界まで下げても見ることが出来なかった情報が提案手法を用いることによって明らかになった。また、ツールを用いて折れ線グラフ、及びPCPを作成する仮定においてツール上の集積量の分布とトーンカーブは有効に機能し、どの位の集積量を持つ領域にどの程度の明るさの色が割り当てられているかを確認し、その情報を元にパラメータの調整を行うことができた。トーンマッピングは集積量の多い領域をその他の領域に上手く馴染ませる目的で導入したが逆に集積量の多い系列群が目立つ結果となった。

#### 7.1.1 変換関数のパラメータに関する議論

我々がガンマ変換にパラメータを追加し作成した変換関数の有効性に関してユースケースから得られた知見を記す。 $\frac{1}{\Gamma}$ の値の操作はユースケースで用いた全てのデータに対して行った。 $\frac{1}{\Gamma}$ の値を操作したのは線形な変換関数ではこの集積量の少ない領域に背景との識別が困難な色が割り当てられたためである。 $x_{border}$ の値の操作は降水量のデータと犯罪履歴データ及び大学のスコアデータに対して手法を適用した際に行った。3つのデータはいずれも集積量の分布が極端に左に寄っており、そのため $x_{border}$ の値を変更しヒストグラムにおける集積量が密集している領域のコントラストを上げる事でより多くの線の視認を可能にした。 $y_{trans}$ の値の操作は犯罪履歴データ、及び大学のスコアカードのデータに対して手法を適用した際に行った。この2種類の多変量データは $\frac{1}{\Gamma}$ の値を集積量の少ない疎な領域の線が見える程度まで小さくすると、密な領域の色のコントラストが弱くなり過ぎてしまう。そこで $y_{trans}$ の値を変更し、集積量が最も少ない領域の線を視認可能にした上で $\frac{1}{\Gamma}$ の値を変更することによって密な領域のコントラストを保ちつつ疎な領域の線を視認できるようなPCPを作成することができた。両者のデータに共通しているのは集積量の分布が極端に左隅に寄っていたことであ

る。このようなデータに対して  $\frac{1}{F}$  の値の変更のみによって集積量の少ない疎な領域の線を見ようとする  $\frac{1}{F}$  の値を非常に小さく取る必要があり、 $\frac{1}{F}$  の値を小さく取りすぎたために密な領域の色のコントラストが弱まったと推測される。このようなデータに対してはユースケースで示したように  $y_{trans}$  の値から変更していくことによって  $\frac{1}{F}$  の値を小さく取り過ぎる必要がなくなり、密な領域の色のコントラストをある程度強めることが可能になる。 $x_{trans}$  の値の変更は今回作成したユースケースにおいて1度も行わなかった。理由はユースケースでも示した通り、利用したデータはいずれも、集積量の少ない領域に集積量の分布が偏っていたことによる。30万本近い系列数をもっているシカゴの犯罪履歴データも集積量の少ない微細領域に分布が極端に偏っていた。原因はデータ内に外れ値が存在する事による。例えば、特定の項目において外れ値が1つ存在するとその線上の全ての領域の集積量が1となる。複数の項目で外れ値が存在するようなデータはそれだけでつまり、各項目が数種類の値しか取らず、かつ外れ値が存在しないようなデータ(例えば、アンケートの集計結果)に対して手法を適用するとき以外は  $x_{trans}$  の値は使用しづらいという事が分かる。割り当てられる画素値が255を超えてしまった微細領域集団に対するトーンマッピングの適用はそうした集団をその他の領域に馴染ませることを目的として導入したが、集積量の多い領域を目立たせる結果となった。犯罪履歴データに対して適用した際には同じ傾向を持つ系列集団が上手く目立ったが、大学のスコアカードのデータに対して適用した際には系列集団が過度に目立つ結果となった。また、降水量データに対して適用した際には余り有効に機能しなかった。トーンマッピングがどのようなデータに対して有効に機能するのかを明らかにするべく、今後更にいくつかのデータに対してトーンマッピングを適用してみる必要がある。

## 7.2 今後の課題

ユースケースから、提案手法の有効性が示された。次の目標は作成した変換関数のパラメータ設定手順、及びヒストグラムとトーンカーブの形状を参考に、集積量に基づく最適な着色の制御方法を考案することである。

### 7.2.1 着色の自動制御

ユースケースで使用したデータの集積量の分布とグラフを作成した際のトーンカーブとを図7.1に示す。4種類のデータについて作成したヒストグラム及びトーンカーブを見比べるとy軸の midpoint を横切る線を中心としてトーンカーブを反転させるとヒストグラムの頂点間を結んでいった線に近い形状を取っている。つまり、ヒストグラムの近似曲線とトーンカーブの曲率が類似している。ここから着色を自動制御するに辺り、最終的なトーンカーブの曲率がヒストグラムの近似曲線の曲率に近づくようパラメータ調整を行うことによって疎な領域と密な領域の両方が視認可能な折れ線グラフが生成されるのではないかという仮説を立てる。また、仮説を元にユースケースで用いた気象庁のデータのような集積量の範囲があまり広くな

いデータに対しては以下の手順でパラメータ設定を行うことで目的に合った折れ線グラフが自動生成されるのではないかと我々は推測する。

- 集積量が少ない領域の線が視認可能になるまで (=集積量が最も少ない微細領域の色が一定値以上になるまで)  $\frac{1}{F}$  の値を増加させる
- ヒストグラムの近似曲線の曲率を計算し、この値に限りなく近い値を取るまで  $x_{border}$  の値を小さくする。

一方で犯罪履歴データや大学のスコアカードのデータのような集積量の範囲が広いデータに対して、上記の手順でパラメータ決定を行うとユースケースのように密な領域のコントラストが弱まってしまふことが予想される。そこでユースケースで踏んだ手順と同じように最初に  $y_{trans}$  の値を増加させ集積量が少ない領域の線を視認可能にすることを考えられる。ここで、 $\frac{1}{F}$  と  $x_{border}$ 、2つの値を用いてどのようにヒストグラムの近似曲線の曲率にトーンカーブの曲率を近づけるかを考える必要がある。ユースケースでは用いられなかった  $x_{trans}$  の利用が求められるようなデータに対する対応も考えなければならない。集積量の範囲、及びヒストグラムの形状に応じたパラメータの自動決定方法を完成させ、本節で立てた仮説の正しさについて検証することは今後の課題である。

## 7.2.2 より柔軟なトーンカーブの操作

前節では自動で最適な着色を行うための制御方法について考察した。本節では手動によってより最適な着色を行うために変換関数自体をより柔軟に決定することを考える。我々は微細領域毎の集積量に基づきそこに割り当てる色を決定するための変換関数としてガンマ変換の式にいくつかのパラメータを加えた式を作成した。ユースケースでは、集積量の分布の近似曲線がいずれも2次の曲線に近くなるような分布であり、そのため作成した変換関数で十分に対応することができた。しかし、全てのデータの集積量の分布の近似曲線が2次の曲線に近づくとはいえない。例えば分布の密集地帯が数か所あるようなデータがあったときにその数カ所の密集地帯のコントラストを上げるような変換関数は作成変換関数のパラメータを操作するだけでは作ることが難しい。この問題の解決手法として、Adobe Photoshop 等で用いられているようなパスを用いてトーンカーブの線を自在にコントロールする手法が考えられる。Adobe Photoshop ではトーンカーブの線にパスを追加し、直接操作することによってユーザが自由にトーンカーブの形を決めることができる。1つ目の課題に上げた自動決定、パラメータ操作による変換関数の決定に加えて直接トーンカーブの形状操作も行えるようにすることで分布が特殊な形をとるようなデータに対しても柔軟な適応が可能になるのではと我々は考える。

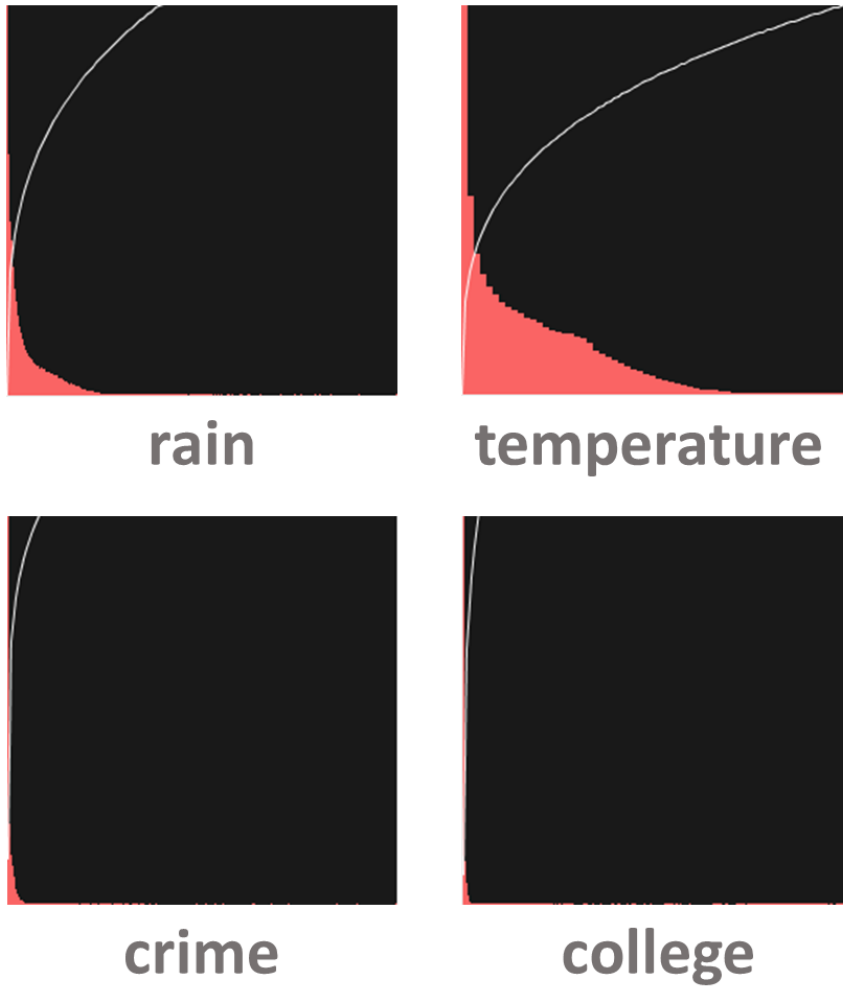


図 7.1: ユースケースで用いたデータの集積量の分布とトーンカーブ.

## 第8章 結論

本研究では大規模データに適用しても，その特徴を俯瞰することができるような折れ線グラフの描画手法を開発した．系列の多いデータを折れ線グラフにして単一の座標領域に描画しようとした際，大量の線によって描画領域が埋め尽くされデータの特徴の把握が困難になってしまうという問題が生じる．この問題を解決するために我々はまず，小領域毎の線の見かけの密度という2次元的な情報に着目し，この密度の値によって線の不透明度を調節する手法を考案した．しかし実際にデータに対して適用すると隣接する小領域間で極端な密度差のある領域が生まれてしまい，描画領域の視覚的混雑はある程度解消されたものの見た人に誤解を与える可能性のある違和感のあるグラフが生成されてしまった．そこで提案手法では描画領域の各微細領域(画素)における集積量という3次元的な情報に着目し，この値に基づいて微細領域へ着色を施す手法を考えた．着色を制御するために導入したのがトーンマッピングやガンマ変換といった画像処理において用いられている技術である．我々は画素値の分布に基づいた画像の濃淡変換の手法を参考に，集積量の分布に基づいて微細領域毎の集積量と微細領域に割り当てる画素値の関係性を定義する変換関数を作成した．変換関数はガンマ変換に用いられる関数を基に，集積量の分布に応じた微細領域への柔軟な色付けを可能にする3種類のパラメータを付与した．さらに我々は開発した描画手法に基づいて，各部分の集積量と部分に割り当てる色の関係性を手動で制御できるツールを開発し，実在する大規模なデータに対して適用した．実在するデータに対して適用した結果，提案手法の有効性，及び変換関数に付与したパラメータの有用性を示すことができた．加えて，集積量の分布と作成した変換関数のトーンカーブの形状に類似性が見られることが分かった．最後に我々はこの知見に基づき，提案手法に基づいた最適な着色の制御方法の開発に向けた考察を行った．

## 謝辞

本研究を進めるにあたり，テーマ選びから発表資料の作成に至るまで終始適切な助言を賜り，また丁寧に指導して下さった三末 和男教授に感謝致します。また，日常の議論を通じて多くの知識や示唆を頂いたビジュアライゼーションとインタラクティブシステム研究室の皆様  
に感謝します。

## 参考文献

- [1] Tamara Munzner. *Visualization Analysis & Design*. Chicago IL, 2015.
- [2] T. Saito, H.N. Miyamura, M. Yamamoto, H. Saito, Y. Hoshiya, T. Kaseda. Two-Tone Pseudo Coloring: Compact Visualization for One-Dimensional Data. *IEEE Symposium on Information Visualization 2005*, pp. 173–180, 2005.
- [3] W. Javed, B. McDonnel, N. Elmqvist. Graphical Perception of Multiple Time Series *IEEE Transactions on Visualization and Computer Graphics*, Vol. 16, Issue. 6, pp. 927–934, 2010.
- [4] G. Ellis, A. Dix. Enabling Automatic Clutter Reduction in Parallel Coordinate Plots. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 12, No. 5, pp. 717–723, 2006.
- [5] D. Trimm, P. Rheingans, M. desJardins. Visualizing Student Histories Using Clustering and Composition. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 18, Issue. 5, pp. 2809–2818, 2012.
- [6] H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen. Visual Clustering in Parallel Coordinates. *Computer Graphics Forum*, Vol. 27, Issue. 3, pp. 1047–1054, 2008.
- [7] K. T. McDonnell, and K. Mueller. Illustrative Parallel Coordinates. *Computer Graphics Forum*, Vol. 27, Issue. 3, pp. 1031–1038, 2008.
- [8] H. Hauser, F. Ledermann, and H. Doleisch. Angular Brushing of Extended Parallel Coordinates. *IEEE Symposium on Information Visualization 2002*, pp. 127–130, 2002.
- [9] Ying-Huey Fua, Matthew O. Ward, and Elke A. Rundensteiner, E.A.. Hierarchical Parallel Coordinates for Exploration of Large Datasets. *In Proceedings of IEEE Visualization '99*, pp. 43–50, 1999.
- [10] AO. Artero, MCF. de Oliveira, and H. Levkowitz. Uncovering Clusters in Crowded Parallel Coordinates Visualizations. *IEEE Symposium on Information Visualization 2004*, pp. 81–88, 2004.
- [11] E. J. Wegman, and Q. Luo. High Dimensional Clustering Using Parallel Coordinates and the Grand Tour. *Classification and Knowledge Organization*, Section. 1, pp. 93–101, 1997.

- [12] Y. Udhida, and T. Ito. 大規模時系列データの可視化と対話的な詳細度制御の一手法. 芸術科学会論文誌, Vol. 8, No. 2, pp. 108–119, 2009.
- [13] T. Ito. マッピング技術: 特に時系列データ可視化技術の体系化に向けて. 日本計算幸楽会講演会, Vol. 17, Issue. 4, 2012.
- [14] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, Vol. 1, Issue 2, pp. 69–91, 1985.
- [15] A. Inselberg, and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. *Springer*, 1987.
- [16] J. E. Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems Journal*, Vol. 4, Issue 1, pp. 25–30, 1965.
- [17] コンピュータグラフィックス編集委員会. コンピュータグラフィックス. CG-ARTS 協会, 2011.
- [18] マイナビニュース. <http://news.mynavi.jp/column/graphics/050/>, accessed December 25, 2015.
- [19] 千葉県, 人口・世帯. <https://www.pref.chiba.lg.jp/cate/kt/toukeidata/bunya/jinkou-setai.html>, accessed December 15, 2015.
- [20] 国土交通省, 気象庁, 各種データ・資料. <http://www.jma.go.jp/jma/menu/menureport.html>, accessed December 20, 2015.
- [21] City of Chicago, Data Portal. <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>, accessed December 20, 2015.
- [22] U.S. DEPARTMENT OF EDUCATION, College Scorecard. <https://collegescorecard.ed.gov/>, accessed December 20, 2015.
- [23] Wikipedia contributors, “ブレゼンハムのアルゴリズム”, Wikipedia. <https://ja.wikipedia.org/wiki/ブレゼンハムのアルゴリズム>, accessed January 5, 2016.