

平成 25 年度 修士論文

論文題目

撮影時におけるモバイル映像アノテーションとそれに基づく
映像編集への応用

Mobile Video Annotation at Photographing and its Application
for Video Editing

指導教員

舟橋 健司 准教授

山本 大介 准教授

名古屋工業大学大学院 工学研究科 情報工学専攻

平成 24 年度入学 24417563 番

名前 高柳 亮太

目次

第 1 章	はじめに	1
第 2 章	関連研究	4
2.1	動画撮影時におけるメタデータ付与手法	4
2.1.1	撮影システム	4
2.1.2	メタデータ付与による動画シーン検索	5
2.2	位置・方向情報付き映像の検索方式	6
2.2.1	位置・方向情報の取得	6
2.2.2	注視点からの検索	6
2.3	閲覧者によるオンラインビデオコンテンツへのアノテーション	9
2.3.1	オンラインでのメタデータ共有	9
2.3.2	システム構成	9
2.3.3	アノテーションの応用	9
第 3 章	提案システムの概要	11
3.1	研究目的	11
3.2	システムの全体像	12
第 4 章	視線集中判定モデル	14
4.1	視線メタデータの取得	14
4.2	経緯度から XY 座標への変換	14
4.2.1	平面直角座標系	15
4.2.2	平面直角座標変換式	16
4.3	視線方向ベクトル	17
4.3.1	視線方向直線の範囲	17
4.3.2	視線の始点, 終点	18
4.4	視線メタデータ記述	19
4.5	線分交差判定	20
4.6	交点座標のクラスタ分類	22

4.6.1	k-means 法	22
4.6.2	x-means 法によるクラスタ分類	23
4.7	マハラノビス距離による外れ値除去	28
第 5 章	プロトタイプシステム	30
5.1	撮影システム	32
5.2	メタデータ解析システム	34
第 6 章	実験と考察	36
6.1	実験システムと実験の様子	36
6.2	視線集中判定モデルの確認	36
6.3	主観評価実験	51
第 7 章	むすび	52
	謝辞	54
	参考文献	55
	発表論文リスト	58
	発表論文リスト	58

第1章 はじめに

近年, インターネットの普及により映像・音楽などのメディアコンテンツをインターネット上で閲覧することが可能となった. その中でも, 世界的に有名な YouTube [1] やコメント投稿システムなどの独自の文化を築き人気を博しているニコニコ動画 [2] をはじめとする, 動画投稿サイトの発展と普及は目覚ましいものである. 最近ではパソコンや家庭用テレビ, HDD レコーダだけではなく, 音楽プレーヤーやスマートフォン, タブレット PC といった携帯端末からも動画投稿サイトを閲覧できるようになっている. また, テレビ放送では一般ユーザにより投稿された動画を紹介していく番組が多数放送されている. 動画投稿サイトには, 様々な動画が投稿されており, 誰でも鑑賞することができる. それらの動画は専門家が作成した作品だけではなく, 一般ユーザが撮影・編集した作品も含まれており, 一般ユーザが作成した映像コンテンツの数は爆発的に増加している. その要因として, デジタルカメラやカメラ付き携帯電話などの撮影機器が一般家庭に普及したことで誰でも簡単に撮影が可能となったこと, HDD の大容量化により個人が所有できるデータ量が増加し, 動画編集ソフトの販売・配布により気軽に編集作業ができるようになったことが考えられる. さらに, Youtube, ニコニコ動画をはじめとする動画投稿サイト, facebook や mixi に代表される SNS (ソーシャル・ネットワーキング・サービス) の登場により, 一般ユーザが映像を発信する機会が増加したことが挙げられる [3]. このような背景から, 一般ユーザが撮影, 編集, 投稿することがより身近なものとなっていると言える [4]. それに伴い, 映像の検索・要約に関する研究が積極的に行われている. 代表的な手法として, 映像解析手法とアノテーション手法がある. 映像解析手法の例としては, 画像特徴および音響特徴を解析することで自動的に映像を要約するものがある [5]. 画面内の色配置情報や歓声を検出し, 映像中の冗長なシーンや意味のないシーンを排除, 盛り上がったシーンやストーリーの把握に必要なシーンだけを抽出する技術である. この手法では, 自動で要約映像を作成できるため, 要約映像作成にかかるユーザ負担が少ないというメリットがある. 一方, シーンの設定は機械処理によって行われるため, ユーザの嗜好を反映できないという課題点がある [6]. また, 放送映像のようにブレのないクリアな映像の場合, 映像解析は比較的容易かもしれない. しかし, 一般ユーザの撮影した映像やノイズの多い映像を対象とする場合, 自動処理は極めて困難である [7]. そこで, 映像に対して, 検索や要約に必要な情報 (メタデータ) を付与するアノテーション手法が提案されている. メタデータとは, 「コンテンツに関するデータ」であり, コンテンツに対して様々な観点からの情報が付与される [8]. 例えば, スポーツ映像におけるアノテーションでは, 試合中に起こったイベント情報 (選手名, プレイ, 得点など) を付与する手法や [9], 映像を見ながら視覚的に判

断したフィールド上の位置情報をリアルタイムで付与する手法 [7] などがある。その他にも、感情を表すオノマトペ (ワクワク, がっくりなど) をメタデータとして画像に付与し, それらの画像を対象とした画像検索を行うものもある [10]. このように, メタデータの記述内容によって, 多様な検索・要約を実現可能な点がアノテーション手法の特徴である。つまり, 映像解析のみでは困難であったユーザの嗜好を反映した映像要約が可能である。また映像解析を必要としないため, 要約の質が映像の質に依存しない。このような観点から, 一般ユーザの映像を対象とした検索・要約に適していると考えられる。しかしながら, メタデータの付与作業は人手に頼るところが多く, アノテーションコストが大きな課題となっている。特に映像の内容を記述したアノテーションの付与には人間の高度な判断や解釈が必要となるため多大な人的コストがかかる [11]. つまり, アノテーションの利点を生かしつつ, アノテーションによる人的コストの最小化, 効率化が強く求められている。アノテーションコストの最小化に着目した先行研究として山本ら [12] の閲覧者によるオンラインビデオアノテーションがある。この研究では, 動画の場合, 多くの閲覧者を獲得することが比較的容易であるという点に着目し, その閲覧者からのフィードバックやアノテーションに参加できる環境を整備することで, 閲覧者による簡単かつ負担の少ない手段でより多くのアノテーションを行うシステムを提案している。この方法により, 一人当たりのアノテーションの量が少なくても, 全体として高度なアノテーションとその活用 (検索・要約など) を実現している。さらに, アノテーション情報をユーザ間で共有することにより, 映像コンテンツを中心としたコミュニティ形成にも言及している。このように, オンラインでのアノテーション収集, 共有は SNS が発展・普及している現代社会において有効な手法であると考えられる。ただし, 不特定多数のユーザから集めたアノテーションの中には信頼性の低い情報も数多く含まれている。山本ら [12] は, アノテーション信頼度という指標を導入し, アノテーションの選別を行っているが, 専門家が修正を加えるアノテーション (オフラインアノテーション) と比較すると, データの信頼性という面で課題が残る。このようにアノテーション手法には人的コストやデータの信頼性などの点において長所短所はあるが, 共通して言えることはすでに存在している映像データを対象としている点である。つまり, アノテーション時に対象とする映像は, テレビ局や個人による加工・編集が加えられたものであり, 撮影とアノテーションが非同期な事後処理である。また, 撮影した映像に対して撮影後に検索や要約のためのメタデータを付与することは現実的とは言えない。そこで我々は, 撮影者が撮影中のボタン押下によって映像シーンにメタデータを付与する手法を提案した [13]. この手法では, 撮影と同時に撮影者自身が任意のシーンに対してマークを付けるので撮影後のシーン検索を効率化できる。しかしながら, 「ボタンを押す」という作業はユーザ負担の大きいものであり, また撮影者自身が任意のシーンでメタデータを付与することができる反面, 撮影者のボタン押下のタイムラグやボタンの押し間違えなども検索結果に反映されるという課題点があった。そこで本研究では, アノテーションコストを最小化するため, 撮影中のユーザの自然な行動の中からメタデータを抽出することを検討し, 撮影者の「視線」をメタデータとして活用できないかと考えた。視線情報は, 撮影と同時に GPS・電子コンパスを用

いて取得する。GPSで撮影者の位置情報を、電子コンパスで撮影方向を時刻情報とともにそれぞれ取得し視線メタデータとして活用する。西岡ら [14] は市販のデジタルカメラにGPS・電子方位計を取り付けているが、本研究では近年急速に普及し始めているスマートフォンやタブレットPCなどを代表する携帯情報端末を使用している。携帯情報端末はカメラ機能を搭載しているだけでなく、高精度なGPSで撮影者の位置情報を取得でき、地磁気センサーや加速度センサーなどの各種センサーを利用することで方位情報を取得できる。また、パソコン並みのインターネット通信が可能なので、サーバとの連携によって複数ユーザの視線メタデータを集積する環境を構築しやすい。以上の理由から視線メタデータの取得に適したデバイスであると言える。

ここで、映像コンテンツ作成の手順について説明する。映像コンテンツ作成には、まずはじめに素材となる動画・静止画を撮影し、次にそれらの素材から実際に使用するものを選定する。選定した素材をつなぎ合わせ、タイトルやBGMの挿入、エフェクトを加えるなどの演出を行うことで一つの映像作品として形作っていくというのが基本的なコンテンツ作成の流れである。松本ら [15] はこれらの工程の中で素材選びの良し悪しが映像作品に大きく影響すると考えている。熟練のカメラマンの場合、撮影対象の選定、撮影位置や角度、構図などの撮影スキルを習得しているため、質の高い素材を収集することができると考えられるが、一般ユーザの場合、撮影に関するスキルを習得したうえで撮影している場合は少なく、この違いが素材の質に影響を与えらる。そこで我々は複数の一般ユーザから「いつ、どの位置からどの方向へ向かって撮影している」という視線メタデータを集め、人気の撮影スポットや視線の集中している注目ポイントを導出することで、素材集めに重要な工程である「撮影」の支援ができないかと考えた。撮影支援によって良質な素材を多く集めることができれば、映像コンテンツの質を高めることができるので、映像コンテンツ作成の支援にもつながる。また、副次的利用として、視線メタデータをユーザ間で共有することにより、素材共有を中心としたコミュニティ形成の手段として利用できるのではないかと考えた。具体的には、視線を共有したユーザ同士で素材を共有することで、映像コンテンツ作成に必要な素材を効率良く収集できるのではないかと考えている。例えば、サッカーの試合などでは、座席が固定されており単一方向からしか撮影できないが、スタジアムにいるユーザ同士で撮影素材を共有することができれば、多視点映像の獲得が可能となり素材の充実が期待できる。

これらの考察を踏まえた上で、本研究では動画撮影中に視線メタデータを付与するシステムを提案し、撮影・映像コンテンツ作成を支援することを目指す。具体的には、携帯情報端末を用いた視線メタデータの取得方法、視線メタデータ共有による視線集中判定法を用いた注視点の導出について言及する。そして、本システムを活用することにより人的アノテーションコストの最小化と視線メタデータの有効活用法を提案していく。以下、第2章において関連研究について、第3章では提案システムの概要、第4章では視線メタデータを用いた視線集中判定モデルについて説明する。第5章ではプロトタイプシステムについて、第6章では実験と考察、第7章でまとめと今後の展望について述べる。

第2章 関連研究

本章では, 本研究に関連する3つの先行研究について説明する.

2.1 動画撮影時におけるメタデータ付与手法

2.1.1 撮影システム

はじめに, 以前我々が提案した動画撮影時におけるメタデータ付与手法とシーン検索への応用について説明する [13]. この研究では, 動画編集作業において素材動画から欲しいシーンを抜き出すトリミングという工程に着目し, トリミング作業の効率化を目的に動画撮影中に動画に対してメタデータを付与する手法を提案した. 具体的には, 図 2.1 のようなシステム環境を構築し, 撮影者が撮影中に印象的だと感じたときにボタンを押すことで映像シーンに対してメタデータを付与する.

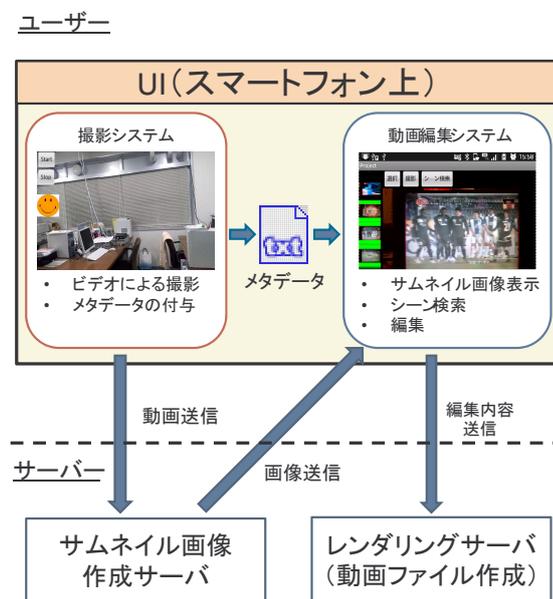


図 2.1: システム構成図

図 2.2 の撮影システムは撮影ボタンと印象ボタンから構成される。印象ボタンとは撮影者が動画撮影中に印象的だと感じたシーンにメタデータを付与するためのボタンインタフェースであり、本システムでは感情カテゴリーに分類した印象ボタンを複数個用意した。感情はエピソード記憶の要素の一つであることが知られており、感情と記憶は密接に関係しているといわれている。その一例として、感情に強く影響を受けた記憶は、長期的に鮮明な記憶として残ることが報告されている。また、撮影した動画からシーン検索する場合、ユーザの記憶がシーン検索に深く関係すると考えられることから感情情報が動画シーン検索に有効であると考えた。そこで、感情を表す表現としてオノマトペをボタン情報として取り入れた印象ボタンを撮影システムに組み込んだ。これらの印象ボタンを自身の感情に該当する場面で押すことでシーンに感情情報を付与することができる。

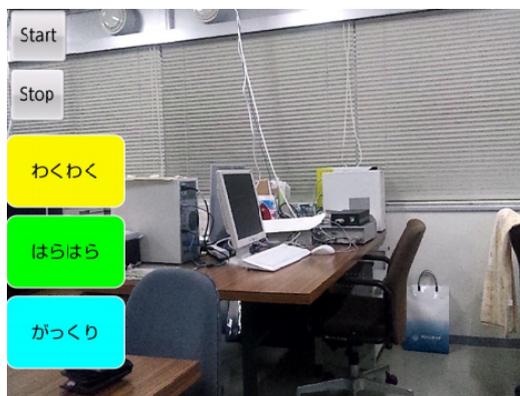


図 2.2: 撮影システム

2.1.2 メタデータ付与による動画シーン検索

撮影システムから取得したメタデータと該当シーンを関連付けることでシーン検索を行う。この研究では、「動画撮影中に印象的だと感じたシーンでボタンを押す」ことに対する評価実験を行った。その結果、撮影者自身が任意のシーンにメタデータを付与でき撮影後に効率良くシーン検索が可能ではあるが、撮影者のボタン押下のタイムラグやボタンの押し間違えなども検索結果に反映されることから主観的性質の強いことが分かった。また、撮影時間の長さに比例して撮影におけるユーザの負担が増加するのではないかという課題点が挙げられた。

2.2 位置・方向情報付き映像の検索方式

2.2.1 位置・方向情報の取得

続いて、位置・方向情報をメタデータとして利用する先行研究 [14] の一部を抜粋し、説明する。西岡ら [14] は収録した映像に対して撮影後に映像データ処理を行って内容検索のためのメタデータを付与することは現実的ではなく、長時間の映像に対して人手によるインデキシングを行うことは、多くの時間を費やすためユーザの負担が大きいと述べている。そこで、様々な環境に応じて簡単なセンサーをビデオカメラに取り付けることで、簡単にデータが入手できる情報を扱うことが望ましいと考え、ビデオカメラの位置と方向が重要な情報としている。映像の収録と同時に GPS で取得するカメラ位置と電子方位計で取得するカメラの向きを位置・方向情報として付与する。

2.2.2 注視点からの検索

位置・方向情報を利用した検索方式として「視点からの検索」と「注視点からの検索」という二つの手法を提案している。視点からの映像検索は、撮影カメラの撮影位置を検索条件として行う。GPS などを用いて撮影カメラの軌跡を保存し、検索者はカメラの撮影位置（緯度、経度）を指定する。検索結果として表示される映像は、指定された場所から撮影されているが、撮影カメラがどちらの方向を撮影しているか分からないため何が映っているか分からない。ウェアラブルカメラを用いて撮影された映像においては、撮影者がどこを見ているかなどを知ることができる。しかし、スポーツイベントなどで撮影された映像に対してはカメラの位置による検索では、見たい映像を探すには不十分である。そこで筆者らは撮影しているカメラの位置を検索条件とするのではなく、カメラに映っている被写体の位置を検索条件とする注視点からの検索手法を提案している。この手法では、“どこから撮影された映像なのか”ということよりも、“どこを撮影しているのか”ということが重要になると考える。視点からの検索では、必ずしも指定した地点からの映像の中に被写体が映っているという保証はなく、被写体が映っている部分映像を見つけるまで何度も地点を変えて検索しなければならない。また、映像の中から被写体が映っている部分映像の全てを探し出すためにはさらに時間がかかる。カメラに映っている被写体をキーとする注視点からの検索であれば、このような煩わしさはない。検索手順を以下に示す。

1. 検索する被写体 (図 2.3 中の A) の位置と被写体からカメラまでの距離 (図 2.3 中の r) を検索者が指定
2. 指定された領域内から撮影した映像を絞り込む
3. 2. で絞り込まれた映像の中から、指定した被写体 (図 2.3 中の A) が撮影カメラのレンズ内に入っている部分映像を抽出する

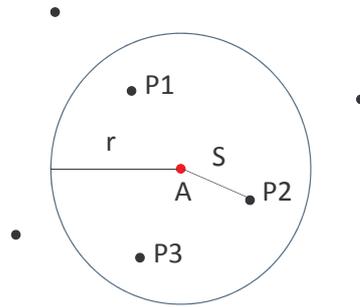


図 2.3: 2次元で表した検索範囲とカメラの軌跡

手順1で検索したい被写体の位置 A と、検索領域の半径 r を指定する (図 2.3 参照). この研究では、焦点距離を考慮しておらず検索項目に距離を加えることで検索時に撮影カメラと注視点まで距離を指定できるようにしている. 手順2でGPSより取得した撮影カメラの位置情報を用いて、検索者が指定した被写体の位置から撮影位置までの距離 S を求め、撮影位置が指定された半径 r の円内に含まれるか調べる. 被写体の位置とカメラの位置は緯度、経度で表現されているため、そのまま距離 S を計算することはできない. そこで、2点の経緯度を平面直角座標に直してから S を求める. この S と利用者が指定した r を比較することで、指定領域内に撮影カメラがあるかが分かり、図 2.3 では、 $P_1 \sim P_2, P_3 \sim P_4$ の部分映像に絞り込むことになる. 手順3では、手順2で絞り込まれた $P_1 \sim P_2$ 間と $P_3 \sim P_4$ 間の部分映像の中から、電子方位計より取得した撮影カメラの方向情報と撮影カメラの画角を用いて、カメラのレンズ内に求めている被写体が映っている部分映像を検索結果として抽出する. 筆者らは実映像に対してプロトタイプを用いて注視点からの検索により検索精度を評価している. 実映像について以下に示す.

- 撮影場所:大学キャンパス内
- 撮影内容:風景
- 撮影時間:6440 秒

プロトタイプにより抽出できた目標映像を表 2.1 に示す. 注視点として選んだものは、大学キャンパス内にある消火栓や道沿いのミラーなど映像の中で目印となりやすいものがある6つ地点を選んだ. 表 2.2, 2.3 に適合率と再現率を示す. 適合率 80%, 再現率 63% となり、適合率は比較的高い値を示したが、再現率は低かった. 再現率の精度が上がれば、GPS、電子方位計を組み込んだウェアラブルカメラやビデオカメラにより日常生活を撮影した映像に対して検索を行えると考えている. さらに、今後の課題としてインタフェースの改良について言及している. 被写体の位置を検索条件

として、緯度・経度の入力を必要としている。しかし、緯度・経度による被写体の位置の指定は慣れ親しんだものではない。よって緯度・経度に代わる新たなインタフェースの提案が必要であると考えている。

注視点	撮影映像数	抽出映像数	適合映像数
ポイント1	21	12	10
ポイント2	10	9	9
ポイント3	14	11	9
ポイント4	14	13	9
ポイント5	6	6	5
ポイント6	5	4	2

表 2.1: 検索結果

抽出映像数	適合映像数	適合率
55	44	80 %

表 2.2: 適合率

撮影映像数	適合映像数	再現率
70	44	63 %

表 2.3: 再現率

2.3 閲覧者によるオンラインビデオコンテンツへのアノテーション

2.3.1 オンラインでのメタデータ共有

最後は, 山本ら [11] の先行研究を通して, オンラインでのメタデータ収集とその有用性について説明する. 山本ら [11] は, 動画コンテンツの意味的な検索や要約をするためには, そのコンテンツへのメタデータの付与が不可欠であると考えている. 動画の場合, 多くの閲覧者を獲得することが比較的容易であるという点に着目し, その閲覧者からのフィードバックやアノテーションに参加してもらえる環境を整備すれば, より多くのアノテーション情報が集まるのではないかという観点から, 一般的な Web ブラウザを用いて, 閲覧者による簡単かつ負担の少ない手段で動画に対してアノテーションを行うシステムを提案している. この手法をオンラインビデオアノテーションという. オンラインビデオアノテーションは, アノテーションに必要な情報をネットワークを通じてリアルタイムに収集し応用に反映させる手段である.

2.3.2 システム構成

この研究では, アノテーションの有用性を示すため閲覧者によるオンラインビデオアノテーションシステムである iVAS(intelligent Video Annotation Server) を構築している. 具体的な処理の流れは図 2.4 で示す. また iVAS の構成図は図 2.5 に示す. ユーザは, ネットワークからアクセス可能な任意のビデオコンテンツに対して, iVAS を通じてアノテーション及び閲覧を行うこととする. iVAS を通じて閲覧したいコンテンツは, 登録サーバを用いて明示的に登録する必要がある. コンテンツを登録すると直ちに, カット検出サーバが連動しカット検出やヒストグラム情報の取得などの自動処理が行われる. ビデオアノテーションサーバによって生成されたページを通してコンテンツを閲覧しつつ, 閲覧者がアノテーションを投稿する仕組みである. 投稿されたアノテーションはアノテーション XML データベースに蓄積され, 各種のアノテーションを利用したサービスなどで利用される.

2.3.3 アノテーションの応用

閲覧者によるビデオアノテーションでは, 比較的容易にアノテーションの量を確保できる可能性は高いが, 信頼性の低い情報や日本語として正しくないものも多く含まれる. そこで, 山本ら [11] は個々のアノテーションに対してアノテーション信頼度という指標を導入し, 情報の選別を行った. アノテーション信頼度は「信頼できる情報を多数入力した人の情報ほど信頼できる」という原則に基づいて計算し, 実験によりその手法の妥当性を確認している. さらに, アノテーション情報のいくつかの応用例を示しており, 本研究ではアノテーションによるコミュニティ支援に注目した. コンテンツの意味的内容に応じたコミュニティ形成支援やコンテンツ推薦を実現するため, 印象アノテ

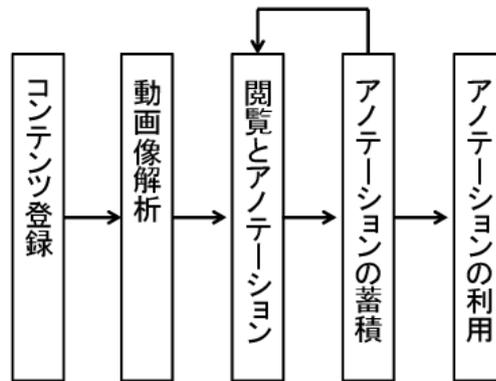


図 2.4: 処理の流れ

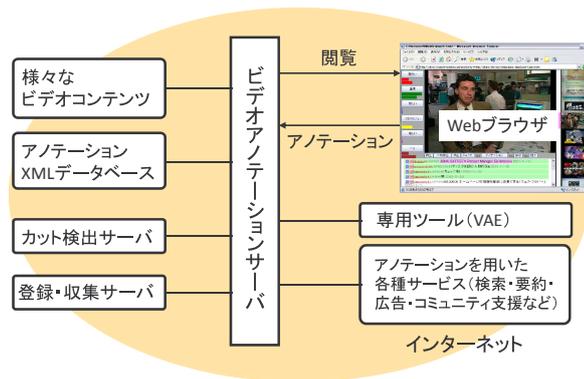


図 2.5: iVAS システム構成

ション情報を用いて、人と人との興味やものの感じ方の近さの一つの指標となる、印象距離を求める方法を提案し、コミュニティ発見やコンテンツ推薦のための一つの指標になればと考えている。

第3章 提案システムの概要

本章では, 第2章で述べた関連研究の課題点を踏まえて, 3.1節で本研究の研究目的を説明する. また, 3.2節で本研究の目的を実現するための提案システムの全体像について説明する.

3.1 研究目的

これまでは映像の検索・要約のためのメタデータ付与作業は人手に頼るところが多く, そのコストが大きな問題となっていた. そのため, 2.1節の先行研究 [13] では, 撮影時にボタンを押下することでメタデータの付与作業を非同期に行われる事後処理としてではなく, 映像収録と同時にメタデータ付与作業を完了させることで映像編集におけるトリミング作業の効率化とアノテーションコストの軽減を試みた. しかしながら, 撮影時のボタン押下によるメタデータ付与は主観的性質が強くメタデータの信頼性の点や, 撮影時間に比例してアノテーションコストが増加するという点に課題があった. また, 映像コンテンツ作成に関するメタデータの活用法についてもトリミング作業の効率化という限定的なものであった. そこで本研究では, アノテーションコストを抑えつつ, 映像コンテンツ作成を支援するための新しい手法を提案する. まずはじめに, 本研究が提案するアノテーションコストの最小化手法について説明する. アノテーションコストの最小化に関して, 2.2節で述べたように撮影後に映像に対して人手によるメタデータを付与していくことは多くの時間を費やすためユーザの負担が大きい. そのため, 撮影時に映像コンテンツ作成支援のためのメタデータを取得することが有効であると考えられる. また, 2.3節で述べたように, ネットワークに繋がっている多数のコンテンツ閲覧者が少ない労力でアノテーションを行うオンラインアノテーションが有効であるが, 不特定多数の閲覧者からアノテーションされるために, データの信頼性や正確さが問題となっていた. このような背景を踏まえ, 我々は「撮影時の人間の自然な行動から映像コンテンツ作成を支援するためのメタデータを取得し, さらにオンライン上でメタデータを共有することで撮影者一人あたりのアノテーションコストを最小化できないか」と考えた. つまり, 撮影者による簡単かつ負担の少ない手段でのアノテーションを行うシステムを提供し, かつ複数の撮影者のアノテーション結果を融合させることで, 全体として高度なアノテーションとその活用法を実現することが狙いである. しかし, デジタルカメラで撮影した静止画が保持するメタデータは, 一般的に時間情報のみである. そこで本提案手法では, 撮影機器として携帯情報端末に着目した. 携帯情報端末であれば時間情報だけでなく, GPS機能により撮影者の位置情報, さらには電子コンパスから撮影方

向をメタデータとして取得可能である。また、携帯情報端末内臓の GPS や電子コンパスの値であれば、データの信頼性という観点からも優れていると考えられる。我々は、時刻・撮影位置・撮影方向といった情報を視線メタデータとして取得し、撮影者同士で共有することでアノテーションコストの最小化と映像コンテンツ作成の手助けとなる情報の提示手法について検討する。

次に前述した視線メタデータによる映像コンテンツ作成支援について説明する。映像コンテンツ作成にはいくつかの工程があるが、中でも素材集めは映像作品の良し悪しに大きく影響するものである [15]。しかしながら映像を撮影する際、撮影対象の選定、撮影位置や角度、構図など考慮すべき事柄が多く、撮影スキルが必要となる。また、視聴者の多様性に応えるためには、より多くの視聴ポイントを実際の視聴者の視点で押さえることが重要である [16]。このような課題に対して、本手法では複数の撮影者からの視線メタデータから他者の撮影位置や視線の集中している視線集中地点を導出することで、素材集めにおいて重要な工程である「撮影」の支援ができないかと考えた。撮影支援によって良質な素材を多く集めることができれば、映像コンテンツの質を高めることができるので、映像コンテンツ作成の支援にもつながる。この考えのもと、第4章で後述する視線集中判定モデルから撮影者から収集した視線メタデータから視線集中地点を導出し、撮影支援情報として有用であるかを検討する。

3.2 システムの全体像

次に、本研究が提案するシステムの全体像について説明する。本システムは撮影システムと視線メタデータ解析サーバの2つのシステムより構成される。撮影システムは携帯情報端末のアプリケーションとして実装している。これは、携帯情報端末に内蔵されているカメラや GPS、電子コンパス、インターネット通信など本システムを実現する上で必要な機能を備えている点や OS やシステムに関する基本的な情報が公開されていて、開発環境が容易に整えられるという点から本システムに適したデバイスであると考えられる。図 3.1 に本システムの全体像を示す。ユーザは、携帯情報端末を用いた撮影システムで静止画を撮影する。撮影と同時に携帯情報端末に内蔵された GPS と電子コンパスを利用して、撮影位置と撮影方向を時刻情報とともに取得する。取得した時刻・位置・方向情報を記述したデータを視線メタデータと呼ぶ。次に、視線メタデータと写真データをサーバに送信する。サーバには各ユーザから送信された視線メタデータを蓄積しておき、集積した視線メタデータに対して視線集中判定モデルを適用し、撮影者の視線が集中している地点を導出する。また、視線メタデータの位置情報から撮影ポイントも導出する。解析結果を再び携帯情報端末へ送信し、アプリケーション内の Map 上に解析結果をユーザに提示する。Map 上に表示された情報を参考にすることでユーザの撮影を支援することができると考えている。さらに、視線を共有したユーザ同士で画像を共有する環境を提供することで、一つの撮影対象に対して様々な角度から撮影された静止画を入手できるようになり、映像コンテンツ作成のための素材の充実が期待できる。

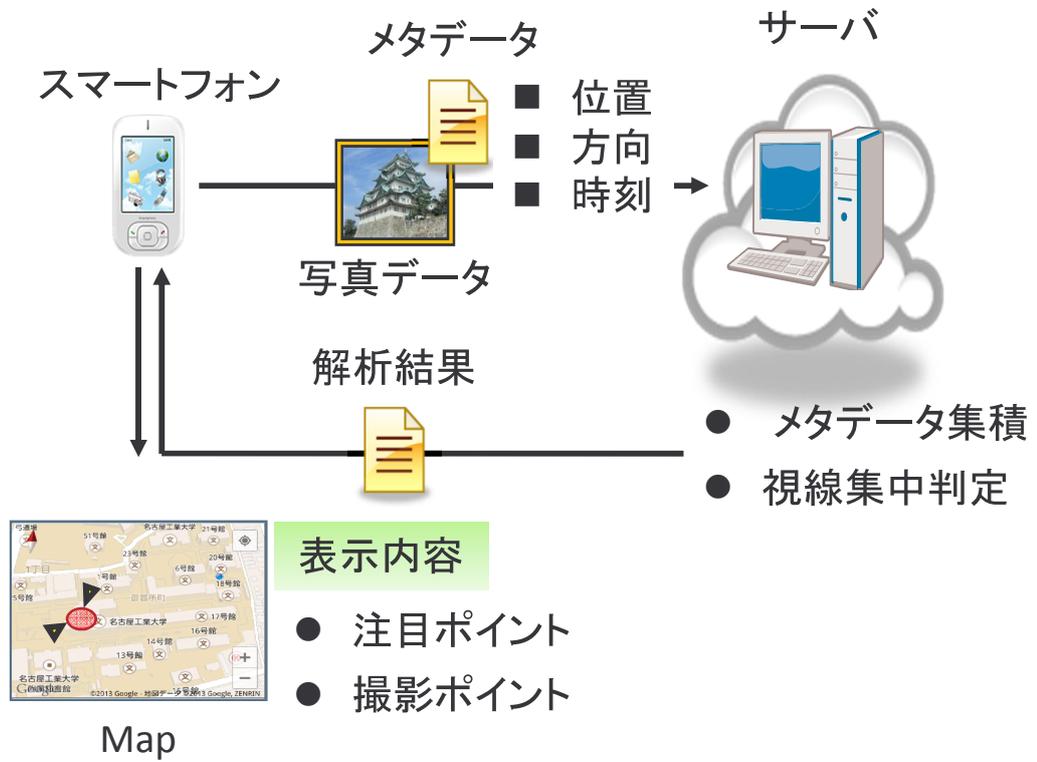


図 3.1: システムの全体像

第4章 視線集中判定モデル

本章では、視線メタデータを用いた視線集中判定法について述べる。視線メタデータは本研究の提案する撮影システムを使用することで取得可能な視線情報である。各ユーザから収集した視線メタデータを利用して視線が集中している地点を導出する。

4.1 視線メタデータの取得

はじめに、視線メタデータの取得方法について述べる。視線メタデータとは撮影時の位置・方向情報を時刻とともに記述したデータである。視線メタデータの取得には携帯情報端末（以下、デバイスと呼ぶ）上に構築した撮影システムを使用する。撮影システムでは静止画の撮影と同時にデバイス内蔵のGPS機能で撮影時の緯度、経度を取得し、テキストデータに記述する。また、撮影と同時にデバイス内蔵のセンサーから方向情報を導出する。デバイスには様々なセンサーが内蔵されているが、視線方向の導出には加速度センサと地磁気センサの二つのセンサを組み合わせることで実現する。加速度センサはデバイスの加速度を検出するセンサで、デバイスの運動速度の変化に伴ってどのくらい加速度が掛かっているか検出する。また同時に重力による地面方向への加速度も検出しており、加速度センサから取得できる加速度はユーザによってデバイスに掛けられている加速度と重力加速度の合力である。加速度センサによって検出された加速度は、図4.2のようなデバイスを中心とした3次元空間のそれぞれの空間軸の成分として取得できる。

一方、地磁気センサは電磁波の強さを測定するセンサである。地磁気センサは加速度センサと同様に、前後・左右・上下の3軸方向の値を取得することができる。前述した加速度センサの3方向の傾き角と地磁気センサの3軸方向の地磁気の値を組み合わせることで、デバイスの姿勢に関係なく正しい方位角を取得することが可能となる。

4.2 経緯度からXY座標への変換

GPSで取得する撮影位置情報は経緯度であるが、経線は赤道を離れるほど間隔が狭くなる。また緯線についても地球を回転楕円体で近似する必要があるため、同じ1度間隔であっても緯度によって距離（子午線弧長）が異なり、平面上の直交座標と比べ扱いが困難となる。視線集中判定において、各ユーザの視線の交点とその位置関係から視線集中地点を決定するため、経緯度の差による距

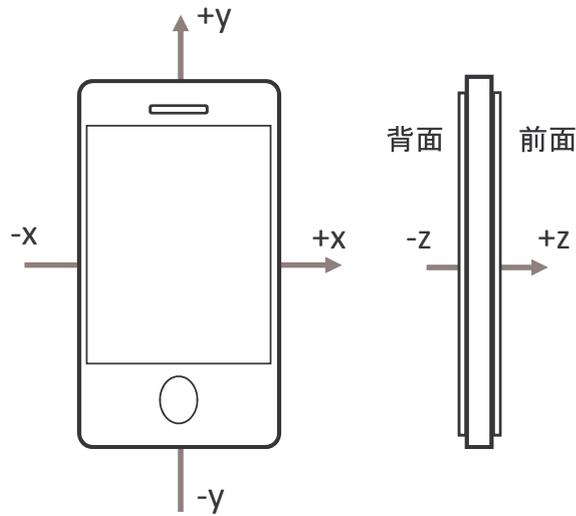


図 4.1: 加速度センサーの軸

離の変化は解析結果に大きく影響すると考えられる。そこで、我々は経緯度を平面直角座標に変換することを考える。

4.2.1 平面直角座標系

本来、どのような地図投影法も、回転楕円体面上の地物を角度や距離の関係を歪めることなく平面に投影することはできない。しかし、角度や歪みを一定限度に収まるように狭い範囲だけ投影を行い、その範囲内だけで用いる平面上の直角座標系を定めることで、直交座標系の様々な利便性を活用することができるようになる。これを平面直角座標系という。平面直角座標系は、公共測量のような測量範囲の狭い場合に共通に利用できるよう定めた座標またはこの座標を基にした位置をいう。日本では、国土交通省国土地理院 [17] により座標原点より東西 130km を適用範囲とする座標系が 19 系設定されている。各系番号ごとの座標系原点の経緯度や適用区域など、具体的な定義は国土交通省告示に記載されている。座標は数学で用いる座標とは異なり、縦座標を X、横座標を Y としている。これにより、地点の位置を (X,Y) の直角座標で表すことができる。座標の原点は、 $X=0\text{m}$ 、 $Y=0\text{m}$ とし、原点から東及び北方向を正、西及び南方向を負の値とする。日本で用いられる平面直角座標は、ガウス・クリューゲルの等角投影法によるもので、座標原点を通る子午線は等長に、図形は等角の相似形に投影される。しかし、距離については、原点から東西に離れるに従って平面距離が増大していくため、投影距離の誤差を相対的に 1/10,000 以内に収めるよう座標原点に縮尺係数 (0.9999) を与え、かつ、座標原点より東西 130km 以内を適用範囲とした座標系を設定している。

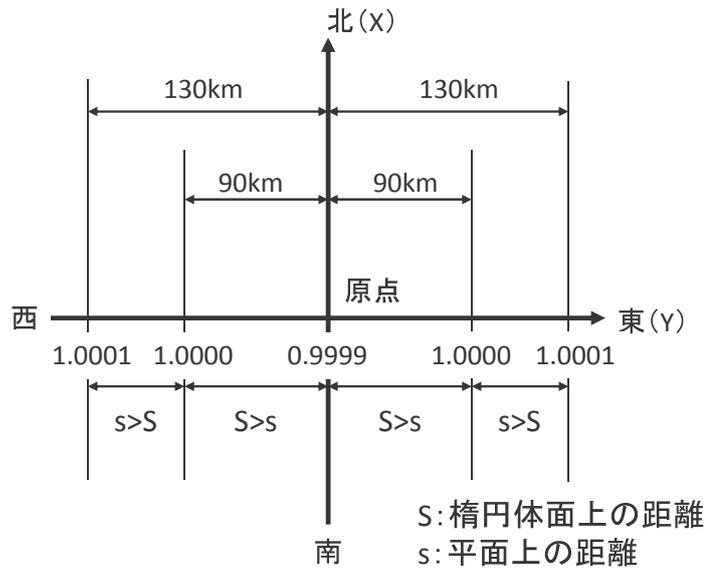


図 4.2: 平面直角座標系

4.2.2 平面直角座標変換式

デバイスのGPSから取得した緯度経度を平面直角座標系のXY座標に変換する計算式を記述する。なお、日本では19系の平面直角座標系が設定されているが、今回の計算には適用範囲に愛知県を含むVII系を選択している。

x座標及びy座標

$$x = \bar{A} \left(\xi' + \sum_{j=1}^5 \alpha_j \sin 2j\xi' \cosh 2j\eta' \right) - \bar{S}_{\psi_0} \tag{4.1}$$

$$y = \bar{A} \left(\eta' + \sum_{j=1}^5 \alpha_j \cos 2j\xi' \cosh 2j\eta' \right) \tag{4.2}$$

ただし、

$$\psi, \lambda : \text{新点の緯度及び経度} \tag{4.3}$$

$$\psi_0, \lambda_0 : \text{平面直角座標系原点の緯度及び経度} \tag{4.4}$$

$$\alpha, F : \text{楕円体の長半径及び逆扁平率} \tag{4.5}$$

$$n = \frac{1}{\sqrt{2F-1}} \tag{4.6}$$

とする。

$$t = \sinh\left(\tanh^{-1} \sin \psi - \frac{2\sqrt{n}}{1+n} \tanh^{-1}\left(\frac{2\sqrt{n}}{1+n} \sin \psi\right)\right), \bar{t} = \sqrt{1+t^2} \tag{4.7}$$

$$\lambda_c = \cos(\lambda - \lambda_0), \lambda_s = \sin(\lambda - \lambda_0) \quad (4.8)$$

$$\eta' = \tan^{-1}\left(\frac{t}{\lambda_c}\right), \xi' = \tanh^{-1}\left(\frac{\lambda_s}{t}\right) \quad (4.9)$$

$$\alpha_1 = \frac{1}{2}n - \frac{2}{3}n^2 + \frac{5}{16}n^3 + \frac{41}{180}n^4 - \frac{127}{288}n^5, \alpha_2 = \frac{13}{48}n^2 - \frac{3}{5}n^3 + \frac{557}{1440}n^4 + \frac{281}{630}n^5 \quad (4.10)$$

$$\alpha_3 = \frac{61}{240}n^3 - \frac{103}{140}n^4 + \frac{15061}{26880}n^5, \alpha_4 = \frac{49561}{161280}n^4 - \frac{179}{168}n^5, \alpha_5 = \frac{34729}{80640}n^5 \quad (4.11)$$

$$\bar{S}_{\psi_0}(VII) = 3985.144116029222 \quad (4.12)$$

$$\bar{A} = 6366.812400856471 \quad (4.13)$$

本研究では、デバイスの GPS から取得した緯度経度を上記の式に従いそれぞれ xy 座標に変換して利用する。計算式の詳細については文献 [18] を参照されたい。

4.3 視線方向ベクトル

次に 4.1 節で取得した撮影位置と方位角からユーザの視線方向直線を生成する。これは各ユーザの視線方向直線の交差判定を行うことで、視線が交差した点を求めるためである。各種センサーで取得した方位角は直線の傾きを求めるときに利用する。方位角は北が 0 度、東が 90 度、南が 180 度、西が 270 度を指す。平面直角座標系は各系の原点緯度経度を原点 (0,0) とし縦軸が x、横軸が y、単位はメートルである。また原点から東及び北方向を正、西及び南方向を負の値となる。

4.3.1 視線方向直線の範囲

ここで、視線を直線として扱う場合の問題点について言及する。直線は両側に無限に延びたものであり、端点を持たない。その場合、図 4.3 左のように視線が交差していないにも関わらず交差していると判定されてしまう。また、図 4.4 左のようにユーザの意図しない無限遠に近い距離で交差した場合にも交差していると判定される可能性がある。そこで我々は、直線ではなく始点、終点を設定し大きさと向きを持つベクトルとして視線を定義する。これを視線方向ベクトルと呼ぶ。視線方向ベクトルを定義することで、図 4.3 右、図 4.4 右のように視線交差における誤判定を防ぐことができる。

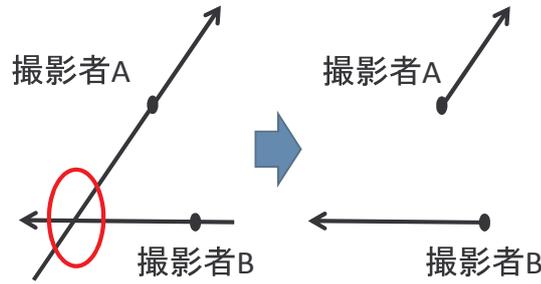


図 4.3: 視線交差の誤判定例 1

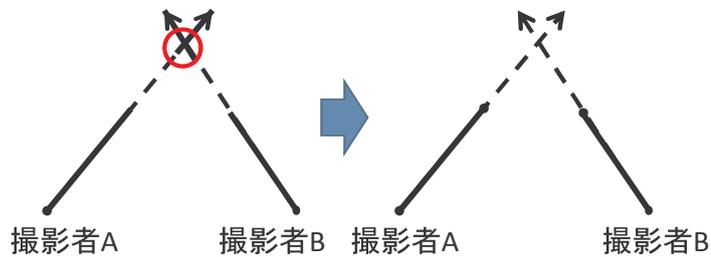


図 4.4: 視線交差の誤判定例 2

4.3.2 視線の始点, 終点

視線方向ベクトルの始点は撮影位置の経緯度を平面直角座標系に変換した座標 (x_0, y_0) とする. 終点は図 4.5 に示すように, 撮影位置座標 (x_0, y_0) と平面直角座標系の x 軸と視線方向ベクトルのなす方位角 θ , 視線方向ベクトルの大きさ α から求める. 終点座標 (x_e, y_e) として式 (4.14), 式 (4.15) により表す.

$$x_e = \alpha \cos \theta + x_0 \tag{4.14}$$

$$y_e = \alpha \sin \theta + y_0 \tag{4.15}$$

視線方向ベクトルの大きさ α は撮影位置から被写体までのおおよその距離にあたる. そこで, α の値を任意で変更可能にすることで, 撮影位置から近い被写体, 遠い被写体の両方への対応が可能となる. これにより, 視線方向ベクトルの交差判定時, 誤判定を減少させることができると考えられる. また, カメラのズームアップと α の値を連動させることでユーザにとってより使い勝手の良いインタフェースとなると期待できる.

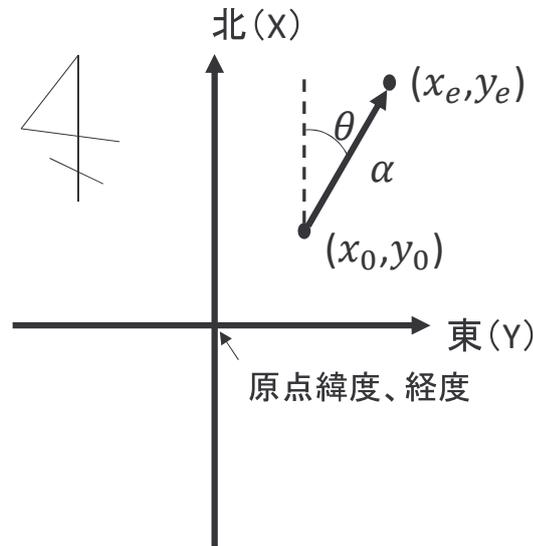


図 4.5: 視線方向ベクトル

4.4 視線メタデータ記述

本システムでは、撮影と同時に視線に関する情報を視線メタデータとしてテキストファイルに記述する。テキストファイルには「年、月、日、時、分、秒、緯度、経度、始点 x_0 、始点 y_0 、終点 x_e 、終点 y_e 、撮影方位角、識別コード」の順に記録される。始点 x_0 、始点 y_0 は 4.3.2 節で述べた撮影位置座標 (x_0, y_0) であり、終点 x_e, y_e は視線方向ベクトルの終点 (x_e, y_e) であり、撮影方位角は平面直角座標系の x 軸と視線方向ベクトルのなす方位角 θ である。このように時刻、撮影緯度経度、始点座標、終点座標、撮影方位角を 1 行に記述した情報を視線レコードと呼ぶ。視線メタデータは視線レコードの集合である。視線メタデータは撮画像とともに、サーバへ送信する。各ユーザから集めた視線メタデータを解析することで視線集中地点を決定する。

4.5 線分交差判定

視線メタデータは撮影システムを使用している各ユーザからメタデータ解析サーバへ送信される。サーバでは各ユーザから集めた視線メタデータより、ユーザの視線が集中していた地点を導出する。本節では視線メタデータを利用した視線集中判定法について説明する。

まずはじめに、各ユーザの視線メタデータの視線レコードを総当たりで判定し視線が交差した点を求める。視線の交差判定には視線レコードの始点座標 (x_0, y_0) と終点座標 (x_e, y_e) を用いる。例えば、ユーザ A のある時刻に撮影したときの始点座標、終点座標をそれぞれ P, Q, ユーザ B のある時刻に撮影したときの始点座標、終点座標をそれぞれ R, S とする。線分 PQ と線分 RS の交差について「点 P, Q を通る直線が線分 RS と交差し、かつ点 R, S を通る直線が線分 PQ と交差している」と考えられる。そこで、線分の交差は「直線を境界線として、線分を構成する 2 つの点が直線に対して両側に存在するとき、直線と線分は交差する」という判定を行っていく (図 4.6)。

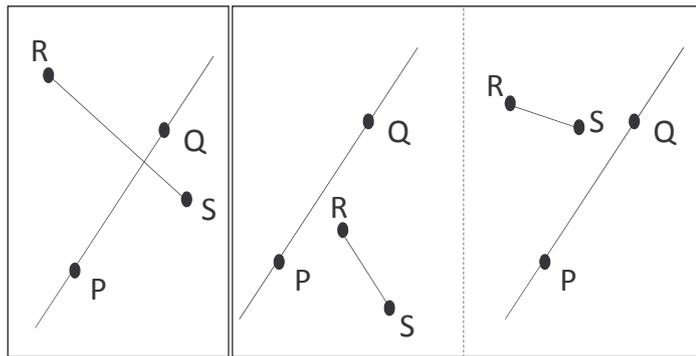


図 4.6: 線分交差判定

ある点が直線に対してどちら側にあるか求めるためには、直線の方程式にその点の座標を代入し、結果得られた値の正負により確認できる (図 4.7)。直線の両側に点があれば交差するので、交差判定は線分を構成する 2 つの点を直線の方程式に代入し、得られた値の符号が異なるかどうかを調べればよい。これを具体的に式で説明するため、点 P, Q の座標を $P(x_1, y_1), Q(x_2, y_2)$ とし、この 2 点を通る直線と点 R, S の座標 $R(x_3, y_3), S(x_4, y_4)$ を結ぶ線分について考える。点 $P(x_1, y_1), Q(x_2, y_2)$ を通る直線 PQ の方程式は、

$$(x_1 - x_2) * (y - y_1) + (y_1 - y_2) * (x_1 - x) = 0 \quad (4.16)$$

と表せる。ここで、この方程式に点 $R(x_3, y_3), S(x_4, y_4)$ を代入し、その符号を調べることによって直

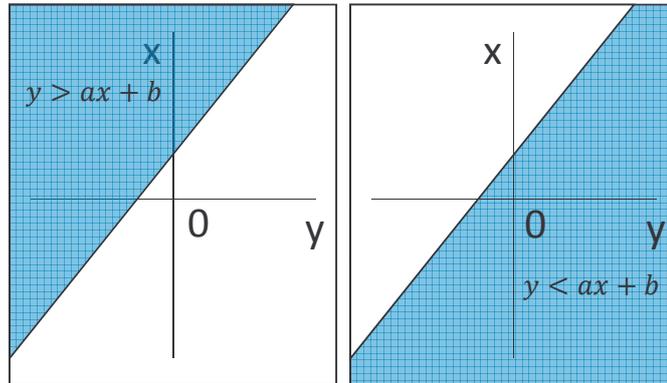


図 4.7: 交点位置の判定

線と線分の交差判定が可能となる.

点 R を代入

$$t_r = (x_1 - x_2) * (y_3 - y_1) + (y_1 - y_2) * (x_1 - x_3) \quad (4.17)$$

点 S を代入

$$t_s = (x_1 - x_2) * (y_4 - y_1) + (y_1 - y_2) * (x_1 - x_4) \quad (4.18)$$

t_r, t_s の値が互いに異なる, つまり

$$t_r * t_s < 0 \quad (4.19)$$

のようになったとき, 交差していると言える. 線分 PQ と直線 CD についても同様の計算を用いて,

$$t_p = (x_3 - x_4) * (y_1 - y_3) + (y_3 - y_4) * (x_3 - x_1) \quad (4.20)$$

$$t_q = (x_3 - x_4) * (y_2 - y_3) + (y_3 - y_4) * (x_3 - x_2) \quad (4.21)$$

式 (4.20), 式 (4.21) より,

$$t_p * t_q < 0 \quad (4.22)$$

式 (4.19), 式 (4.22) の条件を満たせば, 2つの線分は交差していると判定される. 全ての視線レコードを総当たりで判定し, 視線方向ベクトル同士の交点座標を求める. 得られた交点座標に対してクラストリングを適用することで, 視線集中判定を行う.

4.6 交点座標のクラスタ分類

線分交差判定により導出した交点座標に対してクラスタ分析を行う。クラスタ分析とは、2つ以上のデータがあるとき、類似度や距離からデータをいくつかのグループ（これをクラスタと呼ぶ）に分類することである [19]。クラスタ分析の手法は階層的方法と非階層的方法の2種類に大別される。階層的方法は、まず各データをクラスタとし、最も距離の近いクラスタ同士または類似度が最大となるクラスタ同士を順に併合し、新たなクラスタを形成していく方法である。どのようなクラスタが形成されるかはクラスタ生成法に依存するところが大きい。これは途中の段階の結果が後の段階の結果に影響し続けるという性質があるからである。階層的方法の代表的なものとして、最短距離法、最長距離法、群平均法、重心法、メジアン法、ward法などがある。それに対して非階層的方法は、あらかじめクラスタ数を固定し、その数に従って各データを振り分けていくというものである。非階層的方法の代表的なものでは、k-means法がある。この手法は、単純なアルゴリズムで計算できるため、k-means法をベースとした様々な応用手法が提案されている。そこで本研究では、クラスタリング手法としてはじめに k-means 法を検討した。

4.6.1 k-means 法

k-means法は、あらかじめクラスタ数を決めておき、その数に従って各標本点を振り分けていく方法である。各クラスタに基準点を与え、それぞれの標本点を最も近い基準点に割り当てることでクラスタリングを行う。標本点が割り当てられたら、割り当てられた標本点から新たな基準点を決定する。このように基準点の決定と標本点の割り当てを収束するまで繰り返すことで、適切な基準点の推定と標本点の分割が可能となる。k-means法のアルゴリズムを以下に示す（図4.8）。

1. 初期値としてクラスタ数 k を与え、 k 個のクラスタの基準点をランダムで設定する。
2. 基準点と標本点とのユークリッド距離を計算し、基準点が近い方のクラスに分類する。
3. 分類されたクラス内で重心点を求め、基準点を重心点に移動する。
4. 2,3を繰り返し、基準点と重心点が収束し更新できなくなったら、クラス分けを終了する。

本研究における標本点は4.5.1節で求めた視線方向ベクトルの交点座標であり、同じクラスタに分類された交点座標とは、つまり類似度が高いデータであると言える。そこで、同クラスタに分類された交点座標の分布において、交点密度が高い領域が「視線が集中している」場所であると考えることができる。ただし、k-means法を視線集中判定に適用する場合、あらかじめクラスタ数を決定しなければならない。しかしながら、事前にクラスタ数を決定する根拠はないため、交点座標の分布によって正しいクラスタ分類ができない場合がある。このような課題点から、交点座標の分布に応

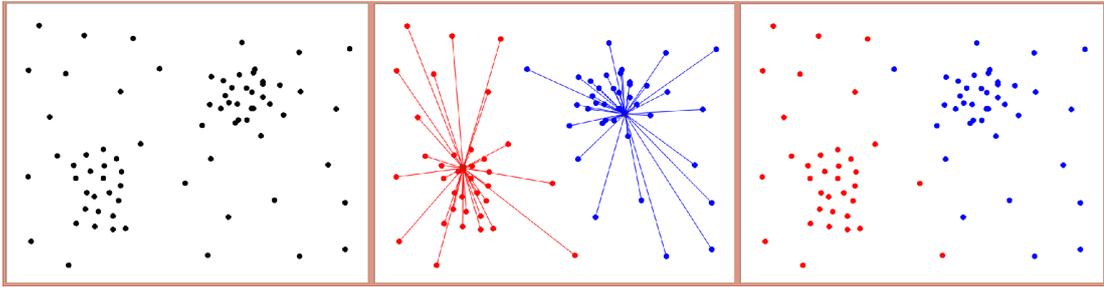


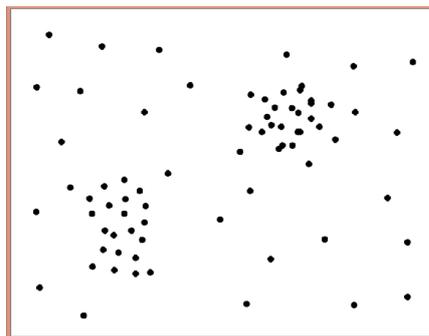
図 4.8: k-means 法

じてクラスター数を自動的に判別することが望ましいと考えた. そこで, データに応じてクラスター数を自動的に決定する x-means 法に着目した. x-means 法は k-means 法の拡張であり, k-means 法の逐次繰り返しと BIC による分割停止基準を用いることで最適なクラスター数を推測するアルゴリズムである [20],[21]. 本研究では, この x-means 法を視線集中判定に適用する.

4.6.2 x-means 法によるクラスタ分類

4.6.2 節では, x-means 法のアルゴリズムについてテストデータに対して x-means 法を適用した様子とともに概説し, 視線集中判定への適用について説明する.

1. n 個の p 次元データを用意し, クラスター数の初期値 $k_0(k=2)$ を定める (図 4.9) .

図 4.9: n 個の 2 次元データ

2. $k=k_0$ として k-means を適用する. 分割後のクラスターを

$$C_1, C_2, \dots, C_{k_0} \quad (4.23)$$

とする (図 4.10) .

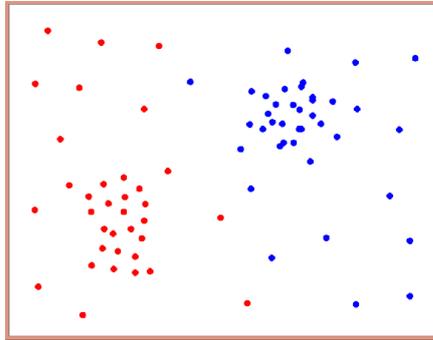


図 4.10: k-means 法による 2 分割

3. $i=1,2,\dots,k_0$ とし, 手順 4~9 を繰り返す.
4. クラスタ C_i に対して $k=2$ として k-means を適用する. 分割後のクラスタを

$$C_i^1, C_i^2 \quad (4.24)$$

とする (図 4.11) .

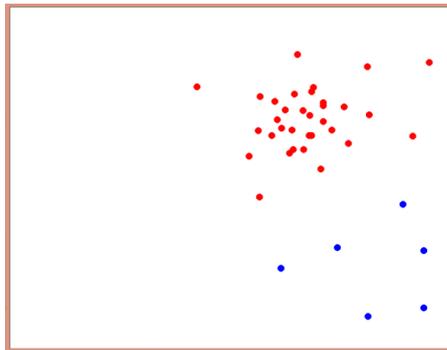


図 4.11: C_i に対して k-means 法を適用

5. C_i に含まれるデータ x_i に p 変量正規分布

$$f(\theta_i; x) = (2\pi)^{-p/2} |V_i|^{-1/2} \exp\left[-\frac{1}{2}(x - \mu_i)^t V_i^{-1} (x - \mu_i)\right] \quad (4.25)$$

を仮定し, そのときの BIC(Baysian In-formation Criterion) を以下により計算する.

$$BIC = -2\log L(\hat{\theta}_i; x_i \in C_i) + q\log n_i \quad (4.26)$$

$\hat{\theta}_i = [\hat{\mu}_i, \hat{V}_i]$ は、 p 変量正規分布の最尤推定値とする。 μ_i は p 次の平均値ベクトル、 V_i は $p \times p$ の分散・共分散行列である。 q はパラメータ空間の次元数で、 V_i の共分散を無視すれば $q=2p$ であり、共分散を無視しなければ $q=p(p+3)/2$ とする。 x_i はクラスター C_i に含まれる p 次元データとし、 n_i は C_i に含まれるデータ数とする。 L は尤度関数で $L(\cdot) = \prod f(\cdot)$ である (図 4.12) .

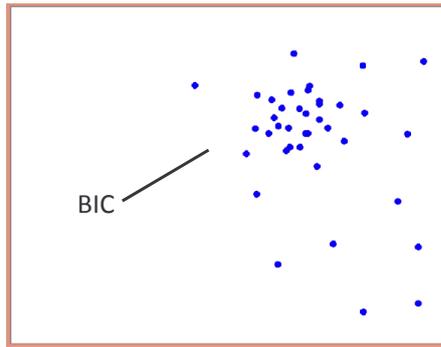


図 4.12: BIC

6. C_i^1, C_i^2 のそれぞれに対して、パラメータ θ_i^1, θ_i^2 をもつ p 変量正規分布を仮定し、この 2 分割モデルにおける BIC' を以下により計算する.

$$BIC' = -2\log L(\hat{\theta}_i'; x_i \in C_i) + q' \log n_i \tag{4.27}$$

$\hat{\theta}_i' = [\hat{\theta}_i^1, \hat{\theta}_i^2]$ は、2 つの変量正規分布の最尤推定値である。共分散を無視すれば、各 p に対し平均と分散の 2 つのパラメータが存在するので、パラメータ空間の次元は $q' = 2 \times 2p = 4p$ となり、共分散を無視しなければ、 $q' = 2q = p(p+3)$ である (図 4.13) .

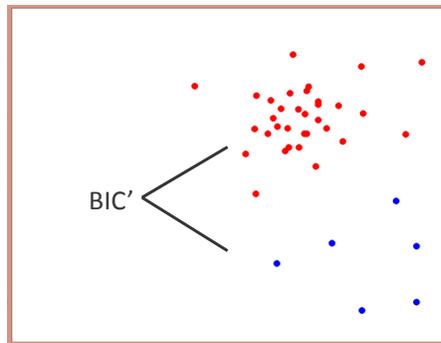


図 4.13: BIC'

7. $BIC > BIC'$ ならば, 2分割モデルがより好ましいと判断し, 2分割を継続するため

$$C_i \leftarrow C_i^1 \quad (4.28)$$

とする. C_i^2 については, p 次元データ, クラスターの重心, 対数尤度と BIC を保持し, これらをスタックに積む. 手順4へ.

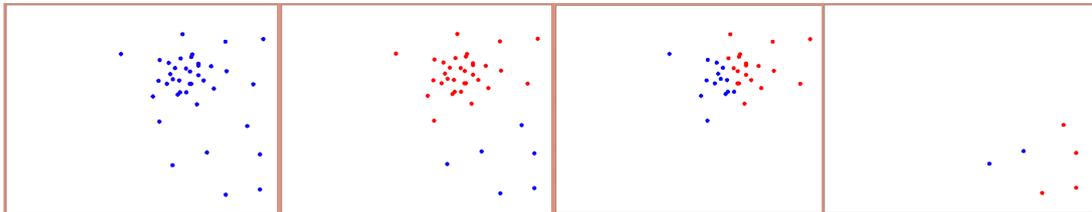
8. $BIC \leq BIC'$ ならば, 2分割しないモデルがより好ましいと判断し, C_i^1 についての2分割を停止する.(手順7で作成された) スタックからデータを取り出し,

$$C_i \leftarrow C_i^2 \quad (4.29)$$

とし, 手順4へ. スタックが空なら次の手順へ進む.

9. C_i における2分割が全て終了. 手順4~8で作成された2分割のクラスターが C_i 内で一意になるようにデータの属するクラスター番号を振りなおす (図4.14).

【 C_1 】



【 C_2 】

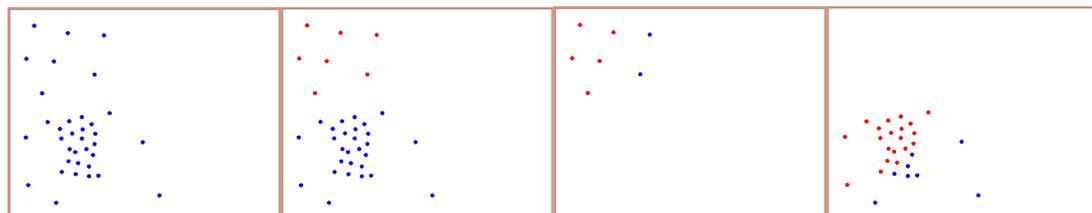


図 4.14: x-means 法による分割の様子

10. はじめに k_0 分割したクラスター全てについて2分割が終了した後, 全データに対してそれらの属するクラスター番号が一意になるように番号を振りなおす (図4.15).

以上のアルゴリズムにより, あらかじめ決められた数ではなく, データに応じたクラスター分類ができるようになる. このアルゴリズムをベースに視線集中判定用のプログラムを実装した. テスト

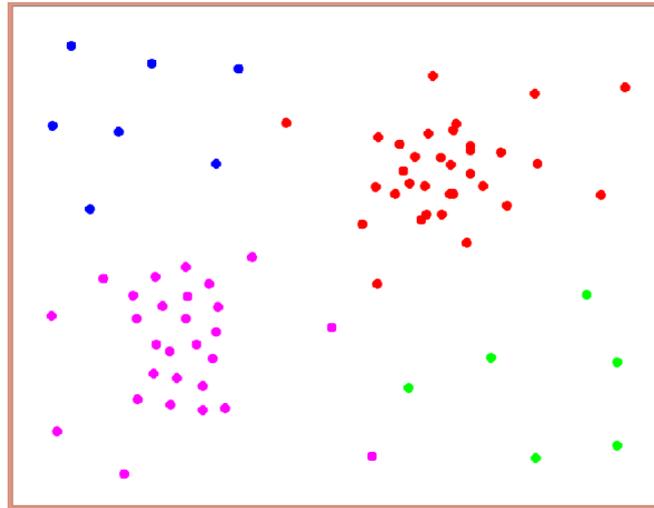


図 4.15: x-means 法によるクラスタ分類結果

データに対して,x-means 法を適用した結果を図 4.16 に示す. 図 4.16 左の場合, 交点座標が密集した集まりは3つであると確認できるので, クラスタ数は3つと判定されると推測できる. 図 4.16 右の x-means 法を実行結果を確認すると, 交点分布に応じて3つのクラスターに分類されていることが確認できる. 4.6.1 節の図 4.17 のような交点座標データに x-means 法を適用すると図 4.17 のように4つのクラスターに分類される. このように交点座標の分布に応じて最適なクラスター数を決定できている. 本章ではテストデータを用いたが, 第6章で実際に撮影システムで取得したデータに対して x-means 法を適用した結果を示す.

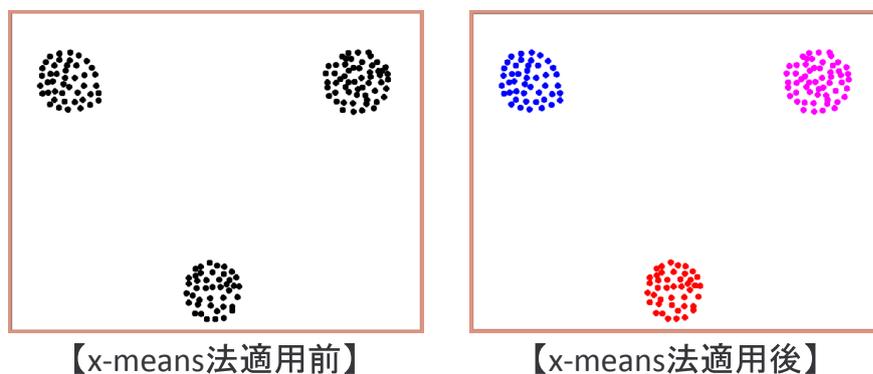


図 4.16: x-means 法適用例

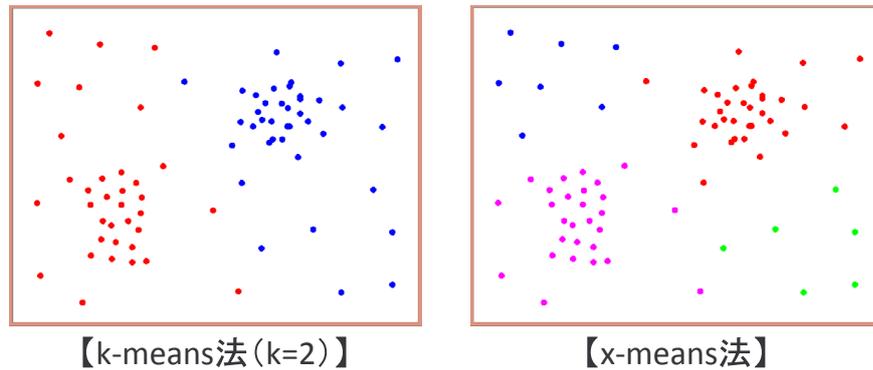


図 4.17: k-means 法との比較

4.7 マハラノビス距離による外れ値除去

視線集中判定において類似度の高い交点座標を同じクラスタに分類し、その中でも交点座標の密集度が高い領域が「視線が集中している」場所である [22]. そこで交点座標の密集領域を特定するためにクラスタの重心に着目した. クラスタの重心座標は交点座標データの平均値から求められる. つまり, 交点座標が密集していればクラスタの重心はその領域付近に位置する. そこで, クラスタの重心と各交点座標の距離から視線集中地点を決定することを考えた. しかしながら, 同クラスタに分類された交点座標データの中には少なからず外れ値が存在しておりクラスタの重心位置に与える影響は大きい. この外れ値が視線集中判定の精度を低下させる原因となる. このため, 外れ値はあらかじめ除去した上で, 解析するデータセットを構築することが望ましい. そこで, クラスタごとに外れ値除去法を適用し視線集中判定の精度向上を図る.

外れ値除去法とは, データから外れ値を検出し, 母集団の中から除去する方法である. 各データの外れた度合いを算出し, ある閾値を超えるものを外れ値とみなす. 本研究では, 距離による外れ値除去法の代表的な手法である Mahalanobis Outlier Analysis(MOA) を適用する. Mahalanobis Outlier Analysis(MOA) とは, マハラノビス距離が閾値 θ_{MOA} を超えるデータを外れ値とみなし除去する, 距離に基づいた手法である [23]. マハラノビス距離とは, 母集団の分散に基づいて算出される距離尺度であり, 母集団の重心と各データとの距離を表す. 図 4.18 にある母集団とマハラノビス距離の例を示す.

図中の点は各データを表し, 楕円はマハラノビス距離の等高線を表す. この図のように楕円状に分布している場合, 重心からの距離が同じデータであっても短軸では外れ値になるが, 長軸では外れ値にならないというケースがある. つまり正規分布とは異なり, 中心からの距離だけでなく, 方向によって外れ値かどうかが決まるということである. \bar{x} を各変数の平均値のベクトル, x_i を標本 i の

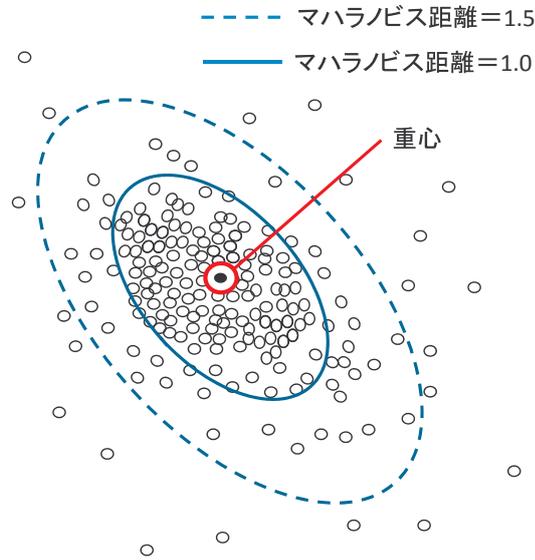


図 4.18: マハラノビス距離の例

持つ変数のベクトル, n を標本の数, T を転置ベクトルとしたとき, 共分散行列 S は

$$S = \frac{1}{n-1} \sum_{i=1}^n 1(x_i - \bar{x})(x_i - \bar{x})^T \quad (4.30)$$

と表され, データ i のマハラノビス距離 MD_i は

$$MD_i = \sqrt{(x_i - \bar{x})^T S^{-1} (x_i - \bar{x})} \quad (4.31)$$

となる. MD_i は平均が 1 になるように正規化され, データ i が重心と同じ位置にある場合 $MD_i = 0$ となり, 重心から離れたデータほど大きな値となる.

このような特徴から本研究の提案する視線集中判定モデルに適していると考えられる. マハラノビス距離による外れ値除去をクラスタ分類した交点データセットに適用することで, マハラノビス距離が閾値を超える交点座標を外れ値とみなしデータセットから除外することができる. これにより, ユーザに視線が集中している地点をより正確に提示できるようになると考えられる.

第5章 プロトタイプシステム

前述の提案モデルに基づいたプロトタイプシステムを開発した。プロトタイプシステムは撮影時に視線メタデータの付与および視線集中判定結果を表示する撮影システムと視線集中判定を行うメタデータ解析システムの二つのサブシステムから構成される。撮影システムとメタデータ解析システムは各種データのやり取りをオンラインで行うことを想定している(図5.1)。撮影者が本研究の提案する撮影システムで撮影した撮画像と視線メタデータをメタデータ解析サーバへ送信する。撮影者から送信された視線メタデータはメタデータ解析サーバのデータベースに集積され、視線集中判定処理が行われる。解析結果を撮影システムに送信し、撮影者に視線集中地点および撮影位置情報を提示する。本章では、二つのサブシステムの機能と処理内容について記述する。

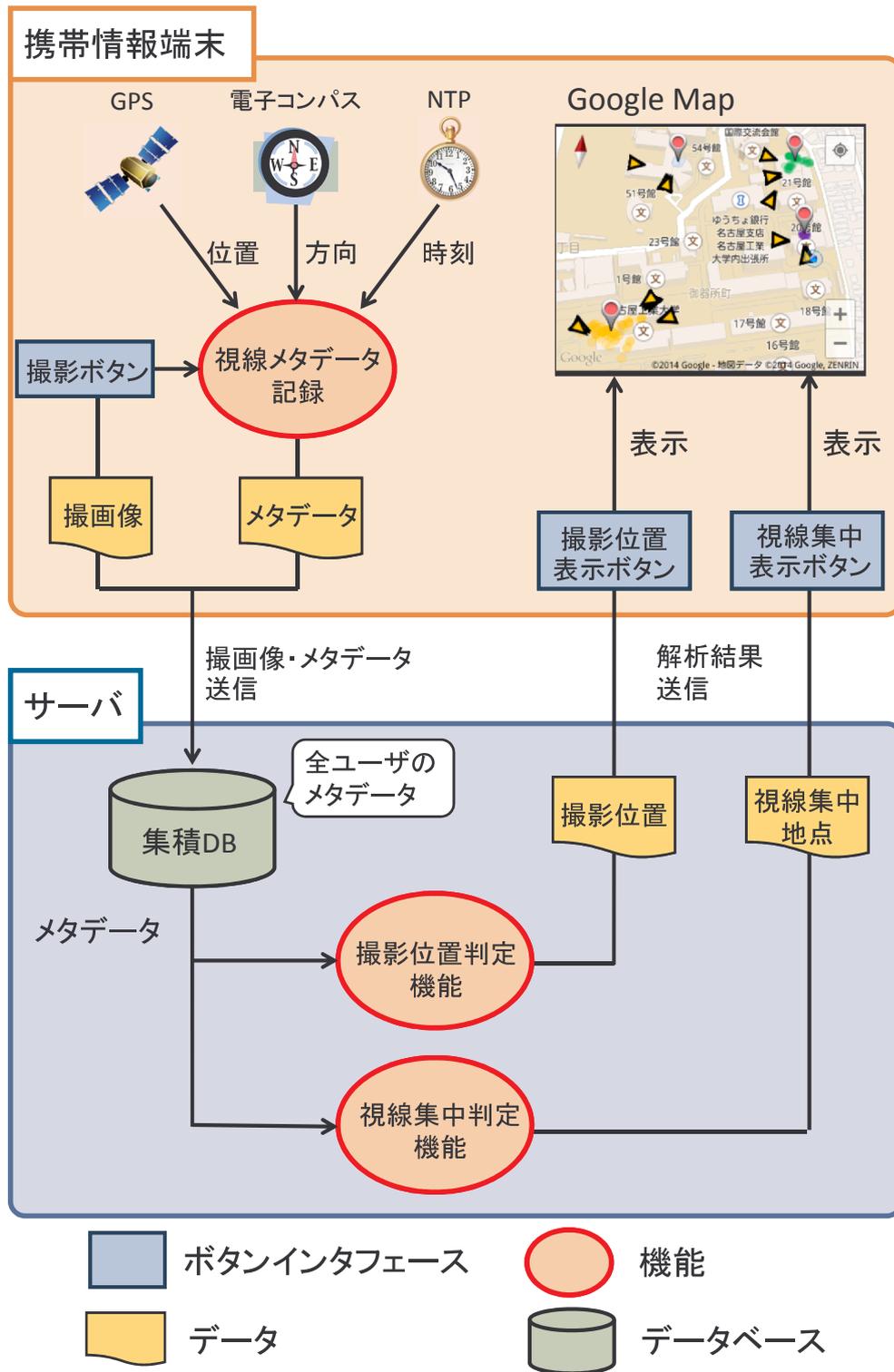


図 5.1: システム構成図

5.1 撮影システム

撮影システムは撮影時に視線メタデータの付与および視線集中判定結果を表示する Android OS 上で動作するアプリケーションである。Android 3.2 搭載の携帯情報端末 (CPU: NVIDIA TegraTM 2 mobile processor with Dual Core ARM CortexTM - A9 CPU) 上で Java 言語により実装した。また、撮影システムに必要なカメラ、GPS および各種センサは携帯情報端末に搭載されているものを使用する。さらに、持ち運び可能な Wifi ルータを活用することで無線ネットワークの位置情報を取得し GPS 精度を向上させている。

次に撮影システムの撮影時のメタデータ付与および視線集中判定結果の表示を実現するための各種機能について説明する。図 5.2 は撮影システムのインタフェース画面である。

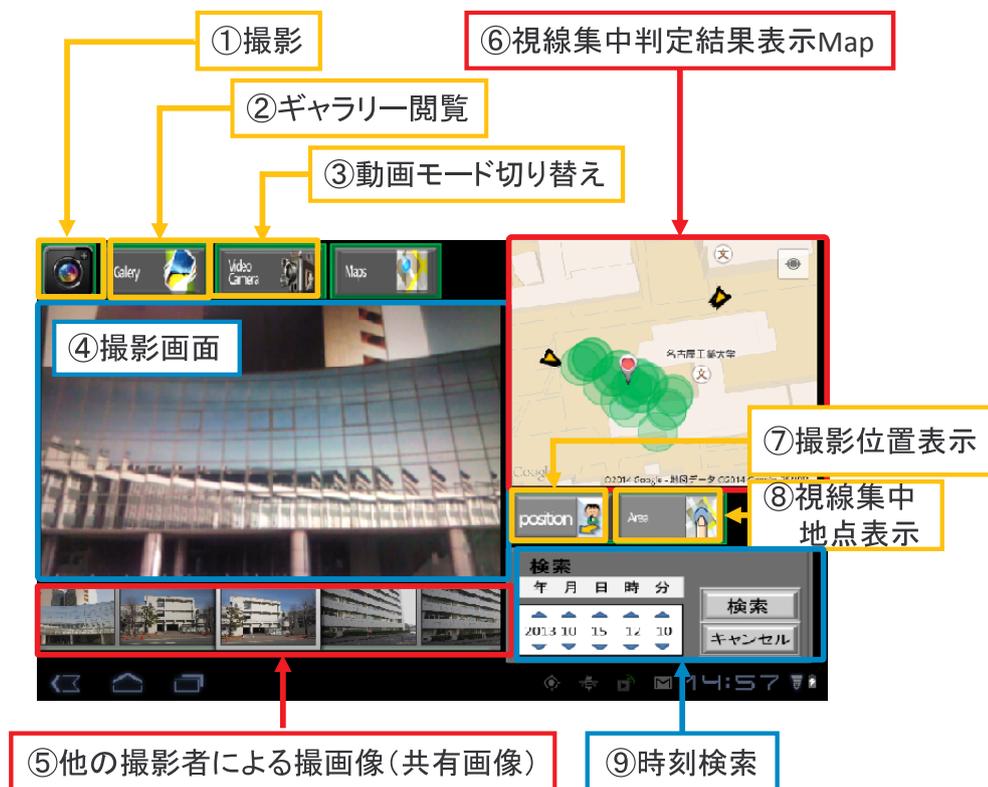


図 5.2: 撮影システム

以下に各種機能の詳細を記述する。

1. 撮影ボタン

静止画を撮影するときに使用するボタンであり、撮影と同時に視線メタデータに撮影時刻、撮影位置、撮影方向などを記録する。

2. ギャラリー閲覧

撮影した映像を閲覧する際に使用する。

3. 動画モード切り替えボタン

現時点では静止画の撮影による視線集中判定モデルを提案しているが、今後の展望として動画撮影時の視線メタデータの活用について検討している。そこで、静止画撮影モードと動画撮影モードを切り替えるためのボタンを用意した。動画撮影時における視線メタデータの活用は今後の課題とする。

4. 撮影画面

撮影対象を表示する画面。今後はタッチ操作によるズーム機能を追加する予定である。

5. 他の撮影者による撮画像（共有画像）

他の撮影者の視点で撮影された画像を一覧表示する。他者の視点を知ることによって撮影を支援することが狙いである。具体的には、一覧表示された撮画像を選択すると、その撮画像が撮影された位置を6.のMap上に表示する。これにより、撮影者はどこからどの方向に撮影すればイメージした写真が撮影できるかを具体的に知ることができる。

また、他の撮影者による撮画像を入手できるようにすれば映像作品に必要な素材を効率良く収集することが可能となり、映像作成支援という点で有用であると考えられる。

6. 視線集中判定結果表示 Map

後述するメタデータ解析システムにより生成された解析データから視線集中地点、撮影位置およびGPSから取得した現在位置をGoogle社よりAndroid開発用に提供されているGoogle Map上に表示する。

7. 撮影位置表示ボタン

他の撮影者の撮影位置を撮影方向とともに表示する。

8. 視線集中地点表示ボタン

視線集中地点を6.の視線集中判定結果表示Map上に表示する。Map上の交点座標位置にあたる地点にクラス番号ごとに色の異なる円を表示することで撮影者に視線集中地点を提示している。また、図5.2のMap上に表示されているマーカーはクラスタの重心位置を表している。

9. 時刻検索インタフェース

時刻によって視線が集中する場所が変化することが考えられる。そこで、時刻を検索条件とするインタフェースを試作した。時刻インタフェースによって様々な場面に対応した情報を提供できる可能性を示す。例えば年単位で検索すれば1年間を通して注目度の高い場所を知ることができる。月単位で検索する場合、季節によって視線が集中する場所が変わると考えられる。例えば、春ならば花見、秋ならば紅葉スポットなどに視線が集中することが予想される。特に、日本は四季がはっきりと分かれているため顕著な傾向が表れることが期待できる。また日単位であれば特定の日に開催されたイベントに視線が集中するであろう。時間単位ならば日の出や日の入り、ライトアップなどのように限られた時間に撮影される状況にも対応できる。このように時刻を検索条件とすることでユーザの嗜好を反映した検索が可能となる。また、ユーザ同士での素材映像の共有が実現すれば、映像作品に必要な素材を効率良く収集することが期待できる。さらに時間あたりの撮影回数などから視線集中度合いを算出し、注目されている場所だけでなく注目度の高さを提示することで有意な影響を与えるのかといった点について評価する必要がある。今後、収集データをさらに増やして検証していきたいと考えている。

5.2 メタデータ解析システム

一方、メタデータ解析システムはCPU: Intel Core i5-3337U, 1.80GHz 上でC++言語により実装した。このシステムでは撮影システムにより生成された視線メタデータに対し視線集中判定処理を行い、解析結果を撮影システムに送信する。本章では、視線メタデータから解析結果を出力するまでのフローと処理の概要について記述する。図 5.3 に視線集中判定処理フローを示す。



図 5.3: メタデータ解析フロー

1. 視線メタデータ読み込み

前述した撮影システムにより生成した視線メタデータの撮影時刻, 撮影位置, 撮影方向などの各種情報を抽出し, ユーザアカウントごとに値を保持する.

2. 線分交差判定

4.5.1 節で記述した始点座標, 終点座標から線分交差判定を行う. 交差判定により導出された交点座標を交点データセットとして出力する.

3. クラスタ分類

2. の線分交差判定により出力された交点データセットに対して 4.6.2 節で記述した x-means 法によるクラスタリングを適用する. 各交点座標をクラスタ分類し, クラス番号を割り振る.

4. 外れ値検出

3. のクラスタ分類により同クラスタに分類されたデータセットから 4.7 節で記述したマハラノビス距離を求める.

5. 解析データ出力

交点座標, クラス番号, マハラノビス距離が記録された視線集中判定データセットを出力し撮影システムへ送信する.

第6章 実験と考察

6.1 実験システムと実験の様子

前述の提案モデルに基づいたプロトタイプシステムを使用して実験を行った。実験システムは撮影時に視線メタデータの付与および視線集中判定結果を表示する撮影システムと視線集中判定を行うメタデータ解析システムの二つのサブシステムから構成される。なお、今回の実験では撮影システムとメタデータ解析システム間でのデータのやり取りはオフラインで行っている。実験の様子を図6.1に示す。



図 6.1: 実験の様子

6.2 視線集中判定モデルの確認

実際に撮影システムを使用して視線メタデータを付与し、メタデータ解析システムにより視線集中判定を行う。解析結果は撮影システムの GoogleMap 上に表示する。視線メタデータから導出した

視線集中地点および撮影位置が実際に注目した被写体の位置及び撮影位置と一致しているかを確認する。そこで、今回の実験では「撮影対象物と撮影位置をあらかじめ指定した上で撮影する」という条件の下で実験を行った。

撮影環境について以下に示す。

- 撮影場所:名古屋工業大学キャンパス内の 10 か所
- 撮影対象物:2号館,20号館,22号館,52号館 (図 6.2 参照)
- 撮影枚数:撮影者 10 人× 5 枚=50 枚
- 撮影範囲:撮影位置から撮影方向へ 30m,60m,100m のいずれか



図 6.2: 名古屋工業大学キャンパス Map

実験は GPS から位置情報を更新・取得しながら、大学キャンパス内の指定した場所まで移動してもらい、撮影対象を撮影する。撮影者一人につき撮影回数は 5 回とする。撮影時に生成された視線メタデータを図 6.3 に示す。

```
2014,1,9,17,33,46,35.156647,136.9238937,-93524.77840626227,-22115.15235223386,-93545.9093193493,-22069.83696288203,115.0
2014,1,9,17,33,49,35.156647,136.9238937,-93524.77840626227,-22115.15235223386,-93551.27436947393,-22072.74994742604,122.0
2014,1,9,17,33,52,35.156647,136.9238937,-93524.77840626227,-22115.15235223386,-93557.5813577118,-22077.416873222723,131.0
2014,1,9,17,33,55,35.156647,136.9238937,-93524.77840626227,-22115.15235223386,-93550.53031000777,-22072.293987198755,121.0
2014,1,9,17,33,59,35.156647,136.9238937,-93524.77840626227,-22115.15235223386,-93546.69696360172,-22070.212649918903,116.0
```

図 6.3: 視線メタデータ

視線メタデータは左から「年, 月, 日, 時, 分, 秒, 緯度, 経度, 始点 x_0 , 始点 y_0 , 終点 x_e , 終点 y_e , 撮影方位角 θ , 識別コード」の順に記録される。緯度, 経度は経緯度表記の現在地, 始点, 始点は平面直角座標系表記の現在地のことで, 始点 x_0 , 始点 y_0 は 3.3.2 節で述べた撮影位置座標 (x_0, y_0) , 終点 x_e, y_e は視線方向ベクトルの終点 (x_e, y_e) であり, 撮影方位角 θ は平面直角座標系の x 軸と視線方向ベクトルのなす方位角である。撮影者 10 人分の視線メタデータは撮画像とともに, サーバへ送信する。収集した視線メタデータから収集した撮影位置の「緯度, 経度」のみを抽出した撮影位置データを生成する (図 6.4)。

```
35.156647000000,136.923893700000,115.0,Data
35.156647000000,136.923893700000,122.0,Data
35.156647000000,136.923893700000,131.0,Data
35.156647000000,136.923893700000,121.0,Data
35.156647000000,136.923893700000,116.0,Data
35.156929900000,136.924866700000,203.0,Data
35.156929900000,136.924866700000,215.0,Data
35.156929900000,136.924866700000,227.0,Data
35.156929900000,136.924866700000,237.0,Data
35.156929900000,136.924866700000,219.0,Data
35.156728800000,136.925113100000,250.0,Data
35.156728800000,136.925113100000,254.0,Data
35.156728800000,136.925113100000,250.0,Data
35.156728800000,136.925113100000,246.0,Data
35.156728800000,136.925113100000,250.0,Data
35.157271000000,136.926766900000,354.0,Data
35.157271000000,136.926766900000,347.0,Data
35.157271000000,136.926766900000,345.0,Data
35.157271000000,136.926766900000,355.0,Data
35.157271000000,136.926766900000,349.0,Data
35.157527500000,136.926323500000,98.0,Data
35.157527500000,136.926323500000,96.0,Data
35.157527500000,136.926323500000,94.0,Data
35.157527500000,136.926323500000,90.0,Data
35.157527500000,136.926323500000,91.0,Data
35.158565200000,136.926240700000,113.0,Data
35.158565200000,136.926240700000,110.0,Data
35.158565200000,136.926240700000,110.0,Data
35.158565200000,136.926240700000,112.0,Data
35.158565200000,136.926240700000,108.0,Data
35.158266200000,136.926273600000,85.0,Data
35.158266200000,136.926273600000,89.0,Data
35.158266200000,136.926273600000,79.0,Data
```

図 6.4: 撮影位置データ

記述内容は左から「撮影位置の緯度, 撮影位置の経度, 方位角, 識別コード」の順である. このファイルを撮影システム側で読み込み, Map 上にプロットすることでユーザに撮影位置情報を提示する. この際, 撮影ファイルの方位角により撮影位置だけでなく撮影の方向も考慮する.

図 6.5, 図 6.6 左から, キャンパス内の 10 か所から撮影されていることが確認できる. また, 方位角から矢印画像を回転させることで撮影方向を表現している. 図 6.6 右の青色の点は撮影者の現在地を示している.



図 6.5: 撮影位置



図 6.6: 撮影位置その2

次に視線集中判定処理結果を以下に示す。各撮影者から収集した視線メタデータを用いて、交差判定処理を実行したところ 200 個の交点座標が算出された。これらの交点座標をデータセットとして 3.6.2 節で記述した x-means 法を適用する。その結果、データセットの各交点座標は 5 個のクラスに分類された。クラス番号ごとの交点数を表 6.1 に示す。クラス番号ごとの交点座標を色付きの円で Map 上に表示する。

クラス番号	色	交点数
1 (52 号館)	赤	15
2 (52 号館)	青	10
3 (2 号館)	黄	75
4 (22 号館)	緑	75
5 (20 号館)	紫	25

表 6.1: クラスタ分類の結果

x-means により分類されたデータセットからクラスタの重心と各交点のマハラノビス距離を求める。これらの値を解析結果として CSV 形式で記録する。解析結果ファイルの一部を図 6.7 に示す。

-93546.264097662104,-22069.076138329568,-93545.356177348280,-22077.295731083974,20,0.697540,Data
-93562.057299019172,-22055.493652951947,-93545.356177348280,-22077.295731083974,20,2.316571,Data
-93553.854072982082,-22068.621558828654,-93545.356177348280,-22077.295731083974,20,1.024274,Data
-93545.466026516617,-22082.045239218485,-93545.356177348280,-22077.295731083974,20,0.400729,Data
-93537.746674572612,-22094.398784676319,-93545.356177348280,-22077.295731083974,20,1.578990,Data
-93551.115278123281,-22073.004546809156,-93545.356177348280,-22077.295731083974,20,0.605805,Data
-93572.708669609681,-22060.014891525727,-93545.356177348280,-22077.295731083974,20,2.729070,Data
-93560.923799042590,-22073.571834309307,-93545.356177348280,-22077.295731083974,20,1.350177,Data
-93549.653983868164,-22086.536273641177,-93545.356177348280,-22077.295731083974,20,0.859621,Data
-93539.915735984265,-22097.738846348504,-93545.356177348280,-22077.295731083974,20,1.784397,Data
-93557.161083461120,-22077.900343439556,-93545.356177348280,-22077.295731083974,20,0.997050,Data
-93560.916046040162,-22055.009219803422,-93545.356177348280,-22077.295731083974,20,2.292707,Data
-93553.067753520983,-22068.070972014557,-93545.356177348280,-22077.295731083974,20,1.014184,Data
-93544.982636315486,-22081.526866692038,-93545.356177348280,-22077.295731083974,20,0.358285,Data
-93537.487867476215,-22094.000256696210,-93545.356177348280,-22077.295731083974,20,1.557514,Data
-93550.434572054161,-22072.453321903089,-93545.356177348280,-22077.295731083974,20,0.591889,Data
-93555.279174560157,-22052.616509817963,-93545.356177348280,-22077.295731083974,20,2.243665,Data
-93549.097445590291,-22065.290932472475,-93545.356177348280,-22077.295731083974,20,1.060641,Data
-93542.484992125406,-22078.848471213827,-93545.356177348280,-22077.295731083974,20,0.275332,Data
-93536.121312898074,-22091.895947180194,-93545.356177348280,-22077.295731083974,20,1.457205,Data
-93554.160728973686,-22052.141757833200,-93545.356177348280,-22077.295731083974,20,2.247962,Data
-93546.971774619335,-22069.649203830184,-93545.356177348280,-22077.295731083974,20,0.659225,Data
-93340.244305913351,-21835.546941103628,-93334.632531271869,-21858.073127529617,21,5.288249,Data
-93344.440216404604,-21825.661995446022,-93334.632531271869,-21858.073127529617,21,7.713810,Data
-93335.671024655807,-21846.320916573790,-93334.632531271869,-21858.073127529617,21,2.687558,Data
-93334.451464814658,-21849.194019510818,-93334.632531271869,-21858.073127529617,21,2.023061,Data
-93331.855500641279,-21855.309727850283,-93334.632531271869,-21858.073127529617,21,0.892440,Data
-93339.478725809880,-21826.796320478996,-93334.632531271869,-21858.073127529617,21,7.209803,Data
-93344.232058652677,-21813.736645826240,-93334.632531271869,-21858.073127529617,21,10.333769,Data
-93334.522648191749,-21840.413031823809,-93334.632531271869,-21858.073127529617,21,4.023007,Data
-93333.238656563277,-21843.940769829860,-93334.632531271869,-21858.073127529617,21,3.234939,Data
-93330.556177169623,-21851.310821392086,-93334.632531271869,-21858.073127529617,21,1.798671,Data

図 6.7: 解析結果

解析結果は左から「交点 x 座標, 交点 y 座標, 重心 x 座標, 重心 y 座標, クラス番号, マハラノビス距離, 識別コード」の順で記録されている. この解析結果ファイルを撮影システム側へ送信する. 撮影システム側では解析結果ファイルを読み込み, クラス番号ごとに交点データと重心データを保持するが, 解析結果ファイルのマハラノビス距離の値が設定した閾値以上であった場合, 外れ値とみなし除去する. 外れ値除去後のデータセットから重心座標とマハラノビス距離を再計算する. 最後に平面直角座標系から経緯度表記に変換し, 解析結果を表示する. 図 6.8, 図 6.9 は解析結果をアプリケーション上の GoogleMap に表示した様子である.

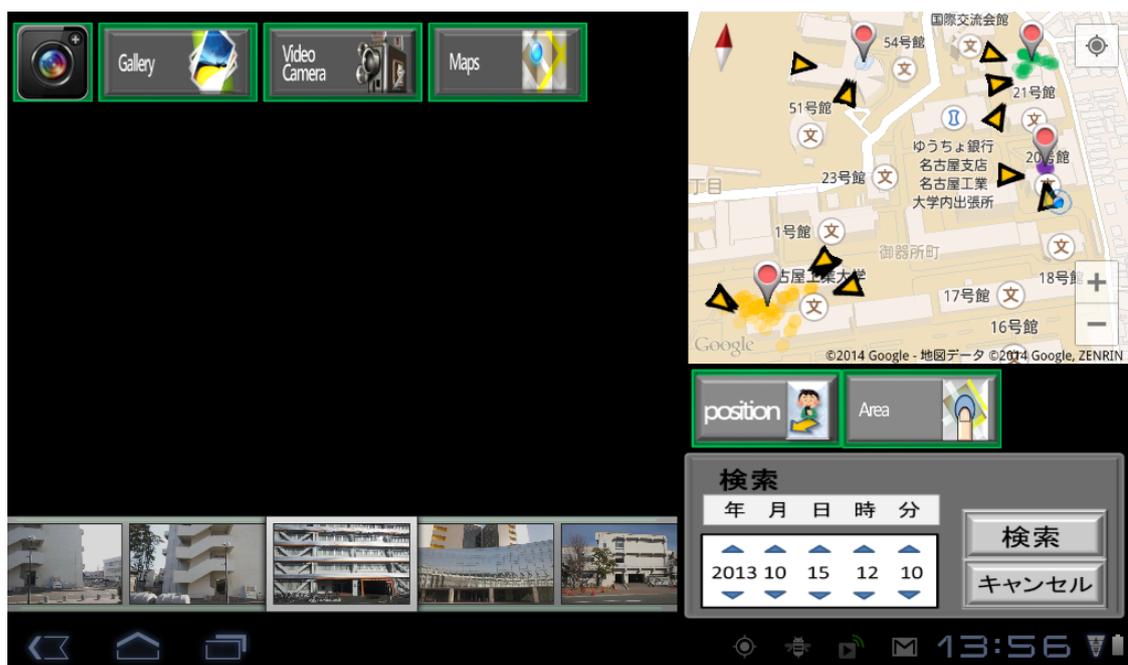


図 6.8: 視線集中地点



図 6.9: 外れ値除去 (左: 閾値なし, 右: マハラノビス距離 4.0)

図 6.8 より, 撮影対象として指定した 2 号館, 20 号館, 22 号館, 52 館付近に視線メタデータから導出した交点座標が分布していることが確認できる. また, 図 6.9 からマハラノビス距離による外れ値除去によって, 撮影対象物から離れた位置に分布していた交点座標が除去されていることが確認できる. この結果から, 実際に視線が集中していた地点と視線集中判定により導出した視線集中地点が一致していると言える. またクラスタリングによる交点データのクラスタ分類によって 4 つの撮影対象物ごとに異なるクラスタであると判定されている. これにより撮影者は視線が集中している地点が何か所あるのかを直観的にだけでなく, データに基づいた根拠によって理解することができる. しかしながら, 図 6.8, 図 6.9 左の 52 号館付近ではクラスタが 2 つ存在している. これは x-means 法によるクラスタリングでは交点分布に応じてクラスタ数が自動的に決定されるため主観的に判断したクラスタ数とは異なるクラスタ数となるケースがあると考えられる. Map 上の交点分布を観察すると撮影対象物付近から極端に離れた場所に交点がほとんど分布していないことが分かる. これはマハラノビス距離による外れ値除去法によってマハラノビス距離が閾値を超える交点を除去できているだけではなく, 視線を直線ではなく, 大きさをもつ視線方向ベクトルとして扱うことにより, 意図しない地点での交差が回避できていることを示している. ただし, 今回のように視線方向ベクトルの大きさ (撮影位置から撮影対象物までの距離) が一定である場合, 撮影者と撮影対象物との距離によっては, 複数の撮影者が同じ撮影対象物を撮影している場合でも視線ベクトルの交差が起きない場合や異なる撮影対象物を撮影しているにもかかわらず, 視線が交差していると判定されてしまうことが考えられる. そこで今後の方針として, 撮影対象物との距離に応じて視線方向ベクトルの大きさを撮影者が任意に変更可能なインタフェースを追加することを検討している. 具体的には, カメラのズーム機能と連動して視線方向ベクトルの大きさを変更することを考えている.



図 6.10: 撮影対象物の領域

次にマハラノビス距離による外れ値除去前後での各クラスタの交点数および適合率を示す. 表 6.2 は外れ値除去前後での各クラスタの交点数の変化を表している. マハラノビス距離の閾値は 8.0, 7.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.0 とし, 表 6.2 は外れ値を除去後の各クラスタの交点数を表している. 表 5.3 ではマハラノビス距離を設定した上で, 各クラスタの交点のうち撮影対象物 (2 号館, 22 号館) の領域内に存在する交点数から適合率を求めた. 判定領域は撮影対象物である 2 号館および 22 号館の中心座標から一定の距離を半径とする円の内部 (図 6.10) とし, その円の領域内に存在する交点の数から適合率を求める. 適合率を表す式は以下のようになる.

$$\text{適合率} = \frac{\text{撮影対象物領域内の交点数}}{\text{各クラスタの交点数}} \times 100 \tag{6.1}$$

ただし, マハラノビス距離による外れ値除去を適用する場合, 各クラスタの交点数は外れ値を除去後の交点数となる.

クラス番号	閾値なし	交点数 (個)							
		閾値 (マハラノビス距離)							
		8.0	7.0	6.0	5.0	4.0	3.0	2.0	1.0
1 (52 号館)	15	2	1	1	1	0	0	0	0
2 (52 号館)	10	10	10	10	10	10	5	0	0
3 (2 号館)	75	74	74	74	74	71	66	55	33
4 (22 号館)	75	70	67	67	61	50	29	12	1
5 (20 号館)	25	25	25	25	25	22	15	4	1

表 6.2: 外れ値除去前後の交点数の比較

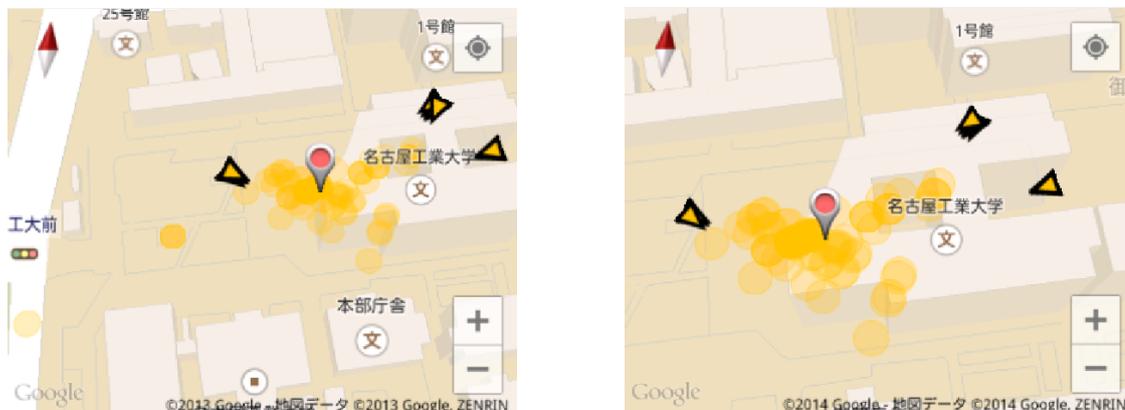


図 6.11: 外れ値除去の様子

の距離を表す。マハラノビス距離を閾値として外れ値を除去した結果、閾値を小さくすると外れ値とみなされる交点数が増加し、各クラスターの交点数が減少していることが確認できる。交点座標のマハラノビス距離が0に近い値ほどクラスターの重心位置に近いことを表しているため、閾値を高く設定すれば広範囲を視線集中地点として撮影システムの Map 上に表示され、閾値を低くすることで視線集中地点の中でも特に視線が集中している地点だけを表示することができる(図 6.11)。ただし、マハラノビス距離の閾値を低く設定することでクラス番号 1,2 のようにすべての交点座標が外れ値とみなされることがある。そのためマハラノビス距離の閾値をどのように設定するかが課題となる。

表 6.3 に注目すると、外れ値除去前の適合率が 2 号館, 22 号館それぞれで 84.00 %, 81.33 % と高い値を示している。平均適合率に関しても、2 号館, 22 号館それぞれで 87.6 %, 93.96 % となっている。これは、視線メタデータから導出した交点座標が視線が集中している地点に多く分布しているためである。また、図 6.12, 図 6.13 が示すように、マハラノビス距離による外れ値除去を適用することで適合率が上昇していることが確認できる。閾値であるマハラノビス距離の値を低く設定することで適合率が上昇していくことから、外れ値除去により視線集中判定の精度が向上したと言える。ただし、マハラノビス距離の値を極端に低くすると、マハラノビス距離 1.0 に設定した際の、22 号館付近の交点数が 1 個となっているようにほとんどの交点が外れ値とみなされる場合が起こりえることが分かった。

マハラノビス距離	適合率 (適合交点数 (全交点数))	
	2号館	22号館
閾値なし	84.00 % (63(75))	81.33 % (61(75))
8.0	85.13 % (63(74))	87.14 % (61(70))
7.0	85.13 % (63(74))	91.04 % (61(67))
6.0	85.13 % (63(74))	91.04 % (61(67))
5.0	85.13 % (63(74))	95.08 % (58(61))
4.0	88.73 % (63(71))	100 % (50(50))
3.0	87.88 % (58(66))	100 % (29(29))
2.0	87.27 % (48(55))	100 % (12(12))
1.0	100 % (33(33))	100 % (1(1))
平均	87.6 %	93.96 %

表 6.3: 適合率

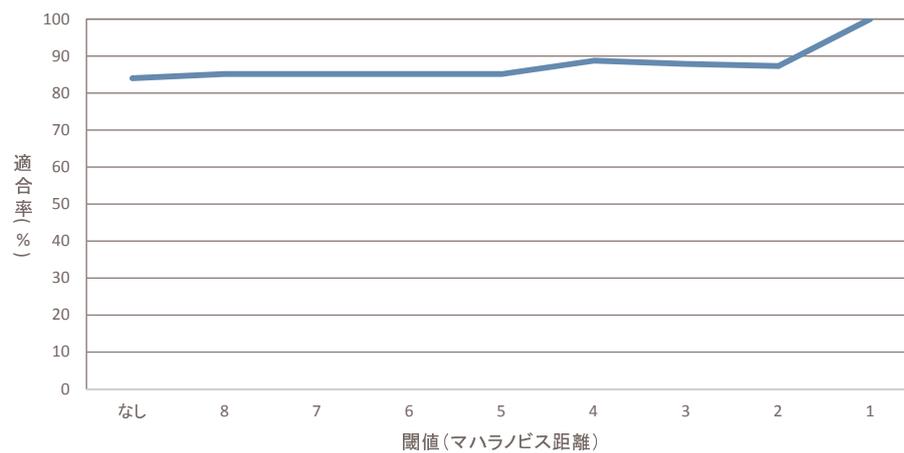


図 6.12: 適合率 (2号館)

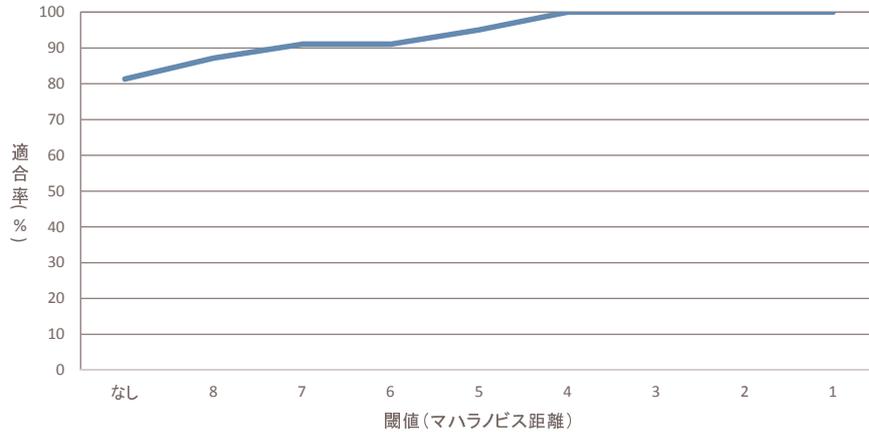


図 6.13: 適合率 (22 号館)

表 6.4~表 6.12 は、外れ値除去前後での各クラスターの重心位置と撮影対象物の中心座標との距離の差を表している。撮影対象物として 22 号館、2 号館を選択し、閾値であるマハラノビス距離の値の変化によって、2 点間の距離の変化を観察した。なお、表中の座標の桁数は小数第 3 位までの表記となっているが、計算時は小数第 12 位まで考慮している。マハラノビス距離 8.0~1.0 まで 1.0 ずつ閾値を下げていった際の、2 点間の距離の変化を図 6.14、図 6.15 に示す。その結果、閾値を下げていくことでクラスターの重心と撮影対象物の中心との距離の差は一時的には減少したが、最終的には外れ値除去前と比べて距離の差が増加した。図 6.14 ではマハラノビス距離 4.0 のときに最も撮影対象物の中心に近づき約 29.514m、マハラノビス距離 2.0 のときに最も撮影対象物の中心から遠ざかり約 35.183m となった。また、図 6.15 ではマハラノビス距離 7.0、6.0 のときに最も撮影対象物の中心に近づき約 14.167m、マハラノビス距離 2.0 のときに最も撮影対象物の中心から遠ざかり約 19.487m となった。このような結果から、極端な外れ値が存在する場合、マハラノビス距離による外れ値除去は視線集中地点の特定に有効であるが、交点座標のばらつきが小さい場合では、必ずしも外れ値除去が視線集中地点の特定に有効であるとは限らないと言える。

クラス番号	平面直角座標 (x,y)		距離 (m)
	撮影対象物	閾値なし	2 点間の距離
3 (2 号館)	(-93543.245,-22043.907)	(-93545.356,-22077.296)	33.455409625259090
4 (22 号館)	(-93347.359,-21866.514)	(-93334.636,-21858.073)	15.271258440262022

表 6.4: 閾値なし

	平面直角座標 (x,y)		距離 (m)
クラス番号	撮影対象物	閾値 8.0 以下	2点間の距離
3 (2号館)	(-93543.245,-22043.907)	(-93544.669,-22075.843)	31.967790093594456
4 (22号館)	(-93347.359,-21866.514)	(-93334.024,-21861.177)	14.363308251861918

表 6.5: マハラノビス距離 8.0

	平面直角座標 (x,y)		距離 (m)
クラス番号	撮影対象物	閾値 7.0 以下	2点間の距離
3 (2号館)	(-93543.245,-22043.907)	(-93544.669,-22075.843)	31.967790093594456
4 (22号館)	(-93347.359,-21866.514)	(-93333.707,-21862.734)	14.167052369250658

表 6.6: マハラノビス距離 7.0

	平面直角座標 (x,y)		距離 (m)
クラス番号	撮影対象物	閾値 6.0 以下	2点間の距離
3 (2号館)	(-93543.245,-22043.907)	(-93544.669,-22075.843)	31.967790093594456
4 (22号館)	(-93347.359,-21866.514)	(-93333.707,-21862.734)	14.167052369250658

表 6.7: マハラノビス距離 6.0

	平面直角座標 (x,y)		距離 (m)
クラス番号	撮影対象物	閾値 5.0 以下	2点間の距離
3 (2号館)	(-93543.245,-22043.907)	(-93544.669,-22075.843)	31.967790093594456
4 (22号館)	(-93347.359,-21866.514)	(-93332.944,-21863.323)	14.763042680288070

表 6.8: マハラノビス距離 5.0

	平面直角座標 (x,y)		距離 (m)
クラス番号	撮影対象物	閾値 4.0 以下	2点間の距離
3 (2号館)	(-93543.245,-22043.907)	(-93543.890,-22073.414)	29.514146573863695
4 (22号館)	(-93347.359,-21866.514)	(-93331.404,-21862.664)	16.412544268944636

表 6.9: マハラノビス距離 4.0

	平面直角座標 (x,y)		距離 (m)
クラス番号	撮影対象物	閾値 3.0 以下	2点間の距離
3 (2号館)	(-93543.245,-22043.907)	(-93544.958,-22075.809)	31.948193000714880
4 (22号館)	(-93347.359,-21866.514)	(-93330.705,-21859.829)	17.945112930030046

表 6.10: マハラノビス距離 3.0

クラス番号	平面直角座標 (x,y)		距離 (m)
	撮影対象物	閾値 2.0 以下	2点間の距離
3 (2号館)	(-93543.245,-22043.907)	(-93544.592,-22079.064)	35.18316681871128
4 (22号館)	(-93347.359,-21866.514)	(-93329.598,-21858.496)	19.48709569670281

表 6.11: マハラノビス距離 2.0

クラス番号	平面直角座標 (x,y)		距離 (m)
	撮影対象物	閾値 1.0 以下	2点間の距離
3 (2号館)	(-93543.245,-22043.907)	(-93545.273,-22078.439)	34.591007237261095
4 (22号館)	(-93347.359,-21866.514)	(-93331.856,-21855.310)	19.12836131953479

表 6.12: マハラノビス距離 1.0

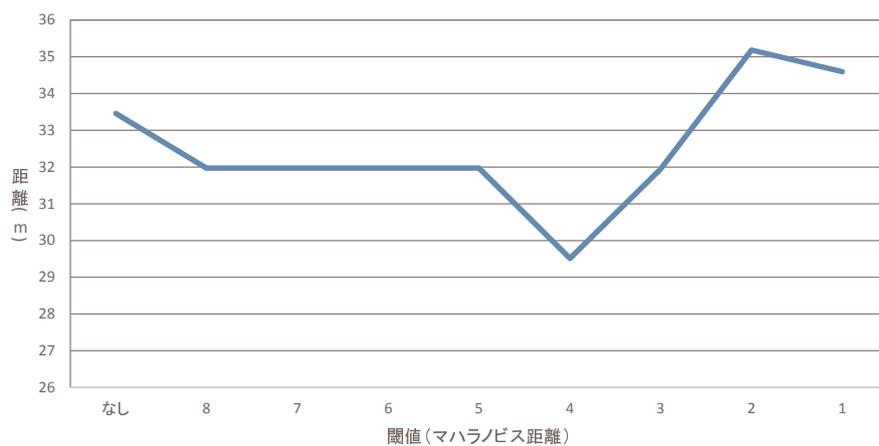


図 6.14: 2号館とクラスタ重心間の距離

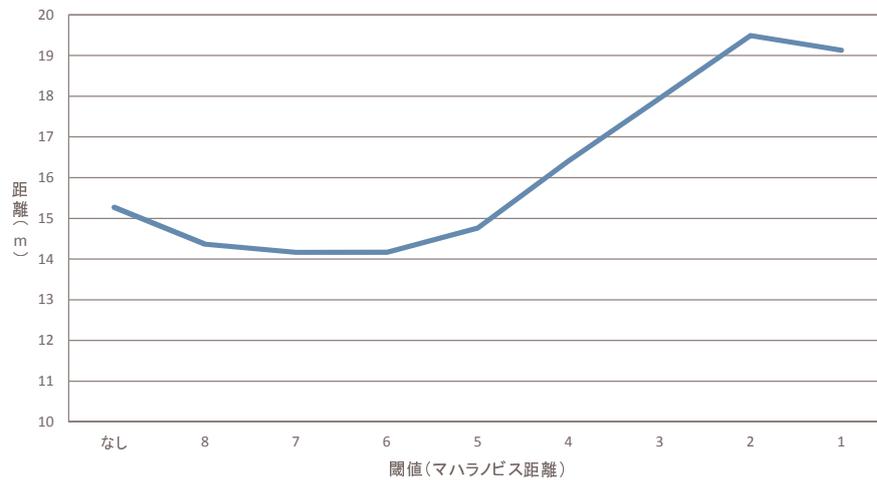


図 6.15: 22号館とクラスタ重心間の距離

6.3 主観評価実験

本研究の目的は、アノテーションコストの最小化と視線集中判定モデルによる撮影者の支援である。そこでアノテーションコストおよび視線集中判定モデルを含むユーザインタフェースの分かりやすさについて被験者数人に対して簡単なアンケート調査を行った。まず肯定的な意見は以下の通りである。

- メタデータ付与が撮影ボタンを押すだけでいいので、アノテーションに対して特に負担を感じなかった。
- 他の撮影者の撮影位置と注目地点が表示されている方が撮影しやすくなった。

次に、本提案システムにおける改善点の指摘は以下の通りであった。

- 撮影時の撮影方向をリアルタイムで表示してあるとよい。
- 視線集中地点の中で特に注目度の高い地点を知りたい。

アンケート調査では、撮影時の位置・方向情報をメタデータとして用いることでアノテーションコストを軽減することができることが確認できた。また、視線メタデータから導出した視線集中地

点および撮影位置の提示による撮影支援が有効であるという一定の評価を得られた。以上より、撮影時の視線をメタデータとして利用することによるアノテーションコストの最小化、視線メタデータを利用した映像コンテンツ作成のための撮影支援の有用性を示したと言える。また、課題点として挙げられた項目は実用性を意識した意見が多く、例えば、視線集中度合いを知りたいという意見には、3次元の棒グラフで視線集中度合いを表現するなどの改良が考えられる、今後の課題としてより多くの被験者に本提案システムを体験してもらい、様々な意見を集めることで更なるインターフェースの向上を追求していきたい。

第7章 むすび

近年、一般ユーザが映像コンテンツを作成することがより身近なものになっている。そのような背景を踏まえ、本研究ではアノテーションコストを抑えつつ、映像コンテンツ作成を支援するための新しい手法を提案した。特に、映像コンテンツ作成に必要な素材集めに焦点を当て撮影を支援するためのシステムを開発した。本提案システムは撮影システムとメタデータ解析システムの二つのサブシステムから構成されている。撮影システムでは撮影時に視線メタデータを付与する手法を提案した。具体的には、撮影デバイスとして携帯情報端末を使用することで、撮影時にGPSと電子コンパスから撮影位置・方向情報を取得し、視線メタデータとして利用した。そのため、撮影という目的を阻害することなく自然にメタデータを付与し、かつ、撮影支援に活用することが可能となった。このことから、本提案手法によりアノテーションコストを最小化できたと言える。また、撮影者から収集した視線メタデータに対して視線集中判定モデルを適用することで視線集中地点を導出し、撮影支援のための情報として提示した。視線メタデータから導出した各クラスターの交点のうち、撮影対象物の領域内に存在する交点数から適合率を求めた結果、実験対象とした2地点において84.00% (2号館)、81.33% (22号館) という高い値を示した。また、外れ値除去を適用することで適合率が上昇することが確認できた。将来的には、視線集中地点をユーザに提示するだけでなくユーザ同士で素材映像を共有することで映像コンテンツ作成のための素材の充実などができるようにしていきたい。

今後さらなるインタフェースの向上を目指す上で必要と考えられる課題を以下に示す。

- 視線共有による素材映像の共有とコミュニティ形成支援

本稿では、視線メタデータによる映像作成支援の一環として、素材収集作業にあたる撮影を支援するシステムを提案した。今後の展望として、撮影の支援だけでなくSNSのようにユーザ間のオンライン上でのコミュニケーションが可能な環境を構築し、撮画像や動画などの映像を共有し、映像作品作成に必要な素材映像を効率良く入手できるようにしたいと考えている。例えば、サッカーなどのスポーツ観戦の場面では、スタジアムの座席が指定されているため撮影者は一視点からの撮影しかできない場合がほとんどである。そこで、スタジアムにいる撮影者でコミュニティを形成し、ユーザ同士で映像を共有することで複数視点からの映像が入手可能となり、より質の高い映像作品の作成が期待できる。また、他者の視点を知ることによって情報の欠落を補完し、より深く試合内容を理解することができると考えられる。このように映像共有による素材映像の充実だけでなく、映像共有を通じた新たなコミュニティ形成を支援する

ことも視野に入れたシステムを検討していく。

- ズーム機能への対応

静止画を撮影するとき、撮影者は被写体に合わせてズームアップを行うと考えられる。今回の研究ではズームアップなどを行わず一定の焦点距離で撮影をしている。提案した撮影システムでは、撮影位置からの距離を指定することで視線方向ベクトルの大きさを設定できるようにしている。つまり、撮影システムの位置と被写体の位置によらず指定された距離内で交差判定条件に当てはまる場合が出てきてしまう。実験では、被写体までの距離に合わせて距離を設定した上で撮影を行ったが、ユーザに距離を入力させることは現実的なインタフェースとは言えない。そこでカメラのズームアップに合わせて被写体までの距離を自動で調整できるようにすることでユーザビリティの向上と視線集中判定精度の向上が期待できる。

- サーバを構築し、オンラインでのデータのやり取り

現時点では、データのやり取りは携帯情報端末とメタデータ解析処理システムとのデータのやり取りはオフラインで行っている。今後は実際にサーバを構築し、メタデータ及び撮画像データのやり取りをオンライン上で実現することを考えている。より実用性を意識したシステムを構築し、そのうえで処理時間やどのようなユーザインタフェースが使いやすいかなどを評価実験を交えながら検証する。さらに、本システムをアプリとして提供することで多くの被験者の獲得が期待できる。

- 動画撮影への対応

静止画だけでなく、動画撮影時にリアルタイムでユーザ同士で撮影位置・方向といった撮影状況を確認できるようにすることで、撮影支援に有用な情報を得られないか検討する。

- 時刻検索インタフェース

時刻による視線集中地点の変化に対応するため、時刻を検索条件とするインタフェースを提案するとともに、その有用性を検証していきたいと考えている。

これらの課題を含め、撮影システムと解析処理システムに関する研究を進めることで多くのユーザに利用される利便性を高めたシステムにしていきたいと考えている。

謝辞

本研究を進めるにあたって、日頃から多大な御尽力を賜った名古屋工業大学 舟橋健司 准教授に心から感謝致します。また、本研究に対して御検討、御協力を頂きました名古屋工業大学 山本大介 准教授、名古屋工業大学 伊藤宏隆 助教、中部大学 岩堀祐之 教授、名古屋工業大学 中村剛士 准教授、愛知教育大学 福井真二 准教授に心から感謝いたします。最後に、本研究を進めるにあたり多大な協力を頂いた舟橋研究室諸氏に心から感謝致します。

参考文献

- [1] YouTube: “YouTube”, <http://www.youtube.com/>
- [2] ニコニコ動画: “ニコニコ動画 (原宿)”, <http://www.nicovideo.jp/>
- [3] 岸本直樹, 若原俊彦, “地図情報と空間情報を用いたコミュニケーション支援”, 電子情報通信学会, IEICE Technical Report MoMuC2010-8 (2010-05)
- [4] 佐野直樹, “XML を利用したインターネット上の動画編集システムに関する研究”, 平成 22 年度名古屋工業大学卒業論文, 2010
- [5] 滝嶋康弘, “Automated Content Summarization Technology”, 映像情報メディア学会誌, Vol.62, No.5, 2008
- [6] 益満健, 越後富夫, “映像重要度を用いたパーソナライズ要約映像作成手法”, 電子情報通信学会論文誌 D-II, Vol.J84-D-II, No.8, pp.1848-1855, Aug.2001
- [7] 大平茂輝, 長尾確: “アノテーションに基づくスポーツ映像要約とナレーション生成”, 情報処理学会全国大会講演論文集, 67th, p55-56, 2005.3
- [8] 林良彦, 松永昭一, 松尾義博, “音声認識・言語処理の適用によるコンテンツ内容記述メタデータの生成”, NTT 技術ジャーナル, 2003.4
- [9] 山本晃司, 大盛善啓, 井田孝, “スポーツ映像を対象としたユーザ適応型映像要約システム”, 電子情報通信学会技術研究報告, IE, 画像工学 103(739), 31-36, 2004-3-12
- [10] 石橋賢, 宮田一乗, “感情を表すオノマトペを用いた感情情報入力手法の提案と画像検索への応用”, The 25th Annual Conference of the Japanese Society for Artificial Intelligence, 2011
- [11] 山本大介, 増田智樹, 大平茂輝, 長尾確, “映像を話題としたコミュニティ活動支援に基づくアノテーションシステム”, 情報処理学会論文誌, Vol.48, No.12, pp.3624-3636, 2007.12.
- [12] 山本大介, 長尾確, “閲覧者によるオンラインビデオコンテンツへのアノテーションとその応用”, 人工知能学会論文誌, Vol.20, No.1, pp67-75, 2005.1

- [13] 高柳亮太, 舟橋健司, 山本大介, “動画撮影時におけるメタデータ付与と動画編集への応用”, 平成 23 年度名古屋工業大学卒業論文, 2011
- [14] 西岡欣亮, 住野優, 大西真昌, 上田真由美, 上島 紳一, “位置・方向情報付き映像の検索方式の提案”, DEWS2005 3A-o2
- [15] 松本剛, “図解デジタルビデオ編集のしくみ”, 2001.2.5
- [16] 大平茂輝, 長尾確: “スポーツ観戦中の情報抽出と映像アノテーションへの応用”, 情報処理学会第 73 回全国大会, 3D-7, 2011.3
- [17] 国土交通省国土地理院, <http://www.gsi.go.jp/>
- [18] 河瀬和重, “Gauss-Kruger 投影における経緯度座標及び平面直角座標相互間の座標換算についてのより簡明な計算方法”, 国土地理院時報, 2011, No.121
- [19] 志津綾香, 松田眞一, “クラスター分析におけるクラスター数自動決定法の比較”, アカデミア情報理工学編 11, 17-34, 2011-03
- [20] 石岡恒憲, “x-means 法改良の一提案-k-means 法の逐次繰り返しとクラスターの再併合-”, 『計量統計学』, 18(1), 3-13, 2006.
- [21] 石岡恒憲, x-means 法のソースコード
<http://www.rd.dnc.ac.jp/tunenori/xmeans.html>(日本語)
- [22] 藤本武司, 砂山渡, 山口智浩, 谷内田正彦, “視線情報の可視化・共有による感性インタラクション支援”, IPSJ SIG Technical Report, 2003-CVIM-140
- [23] 松本真佑, 亀井靖高, 門田暁人, 松本健一, “Fault-Phone モジュール判別モデルに対する外れ値除去法の適用効果”, 情報処理学会研究報告, ソフトウェア工学研究会報告 2007(33), 49-56, 2007-03-22
- [24] 竹村憲太郎, 松本吉央, 小笠原司, “複数人の視線計測に基づく「場の注意」の推定”, IPSJ SIG Technical Report, 2004-HI-110

発表論文リスト

1. 高柳亮太, 舟橋健司, 山本大介, “動画撮影時におけるメタデータ付与と動画編集への応用”, 電気関係学会東海支部連合大会, D1-1, 2012-9-24.

発表論文リスト

1. 高柳亮太, 舟橋健司, 山本大介, “動画撮影時におけるメタデータ付与と動画編集への応用”, 電気関係学会東海支部連合大会, D1-1, 2012-9-24.