

平成23年度卒業論文

論文題目
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

指導教員
舟橋 健司 准教授
山本 大介 助教

名古屋工業大学 工学部 情報工学科
平成20年度入学 20115093番

名前 高柳 亮太

目次

第1章	はじめに	1
第2章	関連研究	3
2.1	アノテーションを用いたシーン検索	3
2.1.1	オンラインビデオアノテーションシステム	3
2.1.2	システム構成	4
2.1.3	アノテーション編集ページ	5
2.1.4	印象アノテーション	5
2.2	感情を表すオノマトペを用いた感情情報入力手法と画像検索への応用	6
2.2.1	オノマトペを用いた撮影システムと画像検索	6
2.2.2	システム構成	6
第3章	提案システムの概要	7
3.1	システムの狙い	7
3.2	システム構成	7
第4章	印象ボタン押下による情報付与手法を用いた撮影システム	10
4.1	印象ボタン	10
4.2	印象ボタンの感情カテゴリー分類	11
4.3	印象ボタン押下による情報付加手法	12
第5章	動画シーン検索の実現法	13
5.1	ユーザインタフェース	13
5.2	動画からサムネイル画像を生成	14
5.3	撮影システムからの情報取得とサムネイルへの関連付け	15
5.4	編集機能	15
5.4.1	シーンの並び替え	16
5.4.2	不要シーンの削除	17
第6章	実験と考察	18
6.1	実験システム	18
6.2	実験1:提案システムの動作確認	19
6.2.1	撮影システムによる撮影と情報付加	19
6.2.2	Cappuccinoによるサムネイル画像生成	22
6.2.3	シーン検索	23
6.3	シーン検索効率に関する評価実験	25
6.4	考察	28

第7章 むすび	29
7.1 まとめ	29
7.2 今後の課題	29
謝辞	31
参考文献	32

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コンテンツの数は爆発的に増加している。また、デジタルビデオやカメラ付き携帯電話の普及により、一般ユーザが日常生活やスポーツイベントなどを気軽に撮影する機会が増え、また動画編集ソフトも数多く販売・配布され、一般ユーザが動画を撮影、編集することがより身近なものとなっている。

ここで、動画編集の作業工程を確認する。はじめに素材となる動画を撮影し、次にそれらの素材動画から欲しいシーンを抜き出しひとつの動画としてつなぎ合わせ、タイトルやBGMの挿入、エフェクトを加える等の演出を行うことで一つの動画作品として形作っていくと考えられる。素材動画を撮影する工程はデジタルカメラやカメラ付き携帯電話などの普及により、比較的容易に行えると想像できる。また、動画をつなぎ合わせる工程やタイトルやBGMの挿入、エフェクトを加える等の演出を行う工程は編集する内容や個人の趣向によるところが大きい。しかしながら、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では、動画再生時間内のどこに欲しいシーンがあるか分からない場合が多く、動画をひと通り確認しなければならない。そのため、素材動画の再生時間や数に比例してトリミングの作業量が増加してしまうという課題がある。そこで、動画の中にいつ印象的なシーンがあるのかを検索できるようにすることで、動画をひと通り確認することなく、効率良くトリミング作業を行うことができるのではないかと考えた。

動画のシーン検索に関する先行研究としてアノテーションを用いた映像シーン検索手法が報告されている [3][4]。アノテーションとは、あるデータに対して関連する情報（メタデータ）を付与することであり、映像に関するアノテーション手法として動画閲覧者が映像シーンに対してコメントを付けたり、評価ボタンによる映像の評価を行うものがある。これらの手法は映像に対してマークを付与する役割があり、このマークをもとに映像シーンの検索を行う仕組みである。しかし、本研究では動画を再生する前に必要なシーンが動画再生時間内のどこにあるのか分かるようにすることで、トリミング作業を効率化することを目標としている。そのため、動画撮影と同時に検索に利用できるメタデータを記録し、検索に用いることが望ましい。さらに、撮影者によって印象的だと感じるシーンは異なるので、撮影者が任意のシーンに対してメタデータを能動的に付与できることが重要であると考えられる。これら視点からの研究として、撮影時に感情を表すオノマトベにより感情情報を画像に付与する手法 [5] や撮影時に位置・方向情報を付与する手法 [6] などがある。

そこで本研究では、撮影者が撮影中に印象的なシーンだと感じたときにボタンを押すことで映像シーンにメタデータを付与する手法を提案する。この手法では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、的確にシーン検索が行えると期待できる。さらにボタンを複数個用意し、状況に合わせてボタンを押し分けることでより詳細なシーン検索を行うことができると考えられる。例えば、nice と boo という2種類のボタンを用意した場合、映像シーンに対して良い印象を受けたら nice ボタンを押し、悪い印象を受けたら boo ボタンを押すといった具合である。こうすることで、シーン検索の際にどのように印象的なシーンであったかを具体的に知ることが可能になる。本稿ではこの手法を実現するために、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などを代表とする携帯情報端末の利用に着目する。携帯情報端末はビデオカメラ機能を搭載していることやパソコン並みのインターネット通信が可能だけでなく、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研究の提案手法を実現するのに適したデバイスであると考えられる。

これらの考察を踏まえた上で、本研究では携帯情報端末を用いたボタン押下による情報付与手法を用いた動画撮影システムと、メタデータが付与された映像シーンの検索を行う動画編集システムの構築を目指し、動画編集支援を行うことを目的とする。以下、第2章において関連研究について説明する。また、第3章では提案するシステムの概要について説明し、第4章ではボタン押下による情報付与手法を用いた撮影システムについて、第5章ではメタデータが付与された映像シーンの検索方法について説明する。第6章ではシステムの実験と考察、第7章ではまとめと今後の課題について述べる。

第2章 関連研究

本章では、本研究で提案するシステムに関連する2つの研究の概要を説明する。

2.1 アノテーションを用いたシーン検索

2.1.1 オンラインビデオアノテーションシステム

近年のインターネットの発達、デジタルカメラ・動画編集ソフトの普及、さらには動画共有サイトの発展などの背景から映像コンテンツが爆発的に増加している。それに伴い、これらのコンテンツの意味的な検索や要約などが重要になっており、これらを実現するためにはアノテーションの取得が必要であると考えた。アノテーションとは、あるデータに対して関連する情報(メタデータ)を付与することであり、映像シーンに対してアノテーションを取得することでシーンの重要度の推定やシーンの検索が可能となる。映像シーンに関連したアノテーションの取得に関する手法として、以下の3種類があると考えている。

- 自動ビデオアノテーション
- 半自動ビデオアノテーション
- オンラインビデオアノテーション

映像認識や音声認識などの自動解析技術を利用する自動ビデオアノテーション方式は人間の手が介在しないためにアノテーションコストが低いという利点がある一方、コンテンツの種類によって解析手法や解析精度が異なるため一般的なコンテンツに対して適用することが困難だという欠点がある。一方、専任の作業者が専用のツールを用いてアノテーションを作成する半自動ビデオアノテーションは人間の手が介在する分アノテーションコストがかかるものの、意味的な情報を付与することができるという大きな利点があるが、詳細なアノテーションを行うためには、膨大な時間がかかるという欠点がある。

最後に、オンラインビデオアノテーションとは、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。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ので、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、閲覧者よりオンラインで情報を収集しアノテーションとして反映させる、閲覧者によるビデオアノテーションシステムを提案している。

2.1.2 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 2.1 で示す。また iVAS の構成図は図 2.2 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。

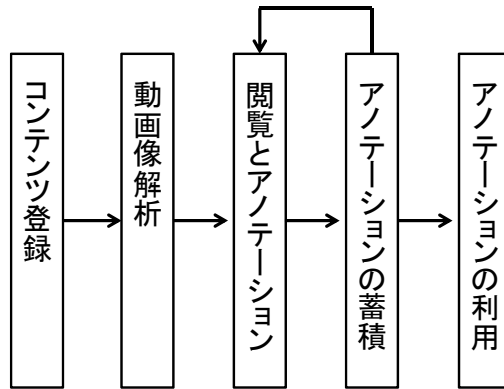


図 2.1: 処理の流れ

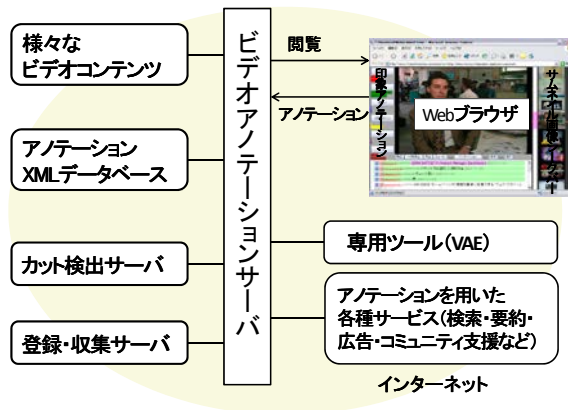


図 2.2: iVAS システム構成

2.1.3 アノテーション編集ページ

閲覧者は、iVASのアノテーション編集ページを用いてアノテーションを行うことができる。アノテーション編集ページのブラウザは図2.3のようになる。画面左に印象アノテーションインタフェース、中央部上部に動画閲覧画面、画面中央下部にテキストアノテーションの一覧、右側にサムネイル画像を用いたスクロール可能なシークバーが配置してある。今回、これらのインタフェースの中から、特に本研究が提案するシステムに関連が深い印象アノテーションインタフェースについて説明する。



図 2.3: アノテーション編集ページ

2.1.4 印象アノテーション

印象アノテーションとは、ビデオコンテンツの雰囲気や閲覧者の主観的印象、例えば、面白い・緊迫・悲しいなどをマウスクリックで入力できる仕組みである。より印象深いシーンではボタンの連打度合いによって印象の強弱を表現できる。印象アノテーションが多く入力されたシーンは重要度の高いシーンとしてユーザーに提示され、シーン検索の手がかりとなる。また、この研究の筆者が以後の研究 [3] で提案した、映像シーンへのボタンアノテーションという手法では閲覧者がボタンを押すことで動画の任意のシーンに対してチェックを行い、映像シーンの引用の手がかりとしている。しかしながら、本研究では動画を再生する前に印象的だと感じたシーンが動画再生時間内のどこにあるのかわかるようにすることでトリミング作業を効率化することを目標としている。そのため、動画撮影と同時に検索に利用できるメタデータを記録し、検索に用いることが望ましい。さらに、撮影者によって印象的だと感じるシーンは異なるので、撮影者が任意のシーンに対してメタデータを能動的に入力できることが重要であると考えられる。そこで、動画撮影時にメタデータを付与することを考える。

2.2 感情を表すオノマトペを用いた感情情報入力手法と画像検索への応用

2.2.1 オノマトペを用いた撮影システムと画像検索

この研究 [5] では、感情を表すオノマトペにより感情情報を画像に付与し、それらの画像を対象とした画像検索を行う。さらに感情が記憶と密接に関係しており、かつ、オノマトペは感情を表現する言葉として適切であると考えており、オノマトペの検索キーが画像検索に効果的であると期待している。すなわち、オノマトペによる感情情報入力手法を用いた撮影システムと、オノマトペが付与された画像データ群から画像検索が行えるシステムの実現がこの研究の目的である。

2.2.2 システム構成

このシステムは、撮影システムと画像検索システムの二つのサブシステムで構成される。図 2.4 にシステムの概念図を示す。図 2.4 のユーザ A が写真を撮影するユーザであり、ユーザ B が画像検

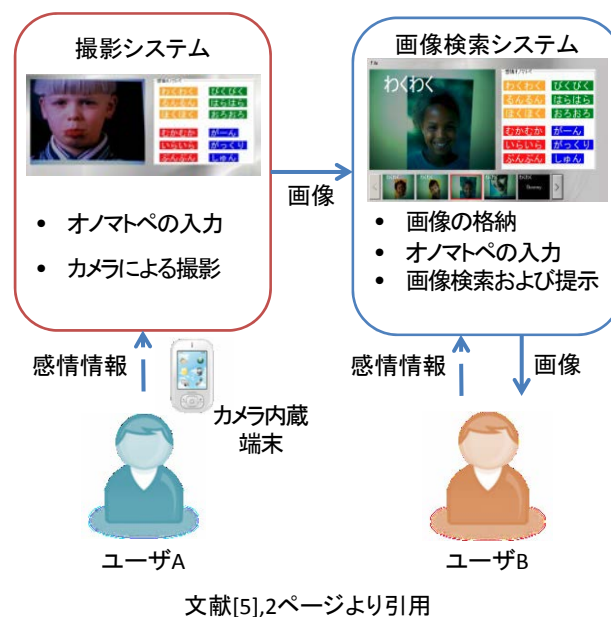


図 2.4: システム概念図

索システムを使用するユーザである。ユーザ A は、被写体の感情を推測し、カメラ内蔵のスマートフォンや PDA などを通して、被写体の感情情報を入力する。感情情報はオノマトペを用いて、入力し、同時に撮影する。撮影システムでは、オノマトペが付与された画像が生成される。ユーザ B は、画像検索の検索キーとして感情情報を入力する。画像検索システムの感情情報の入力も、オノマトペを用いる。画像検索システムでは、入力したオノマトペに基づいた検索結果が提示される。このシステムの注目すべき点として、撮影時に感情情報を付与するところにある。撮影時に情報を付与することで、一枚ずつ画像を確認する手間を省くことができる。さらに感情情報の入力とカメラのシャッターボタンを同期させたシステムとなっているため、ユーザの負荷が抑えられるという特徴がある。

これら 2 つの関連研究を踏まえて上で、次章で本研究が提案するシステムの概要について説明する。

第3章 提案システムの概要

本章では、提案するシステムの概要について説明する。

3.1 システムの狙い

システムの構成を説明する前に、本システムの狙いについて説明する。動画編集にはいくつかの工程があるが、本研究では素材となる動画から欲しいシーンを抜き出すトリミングという作業に動画編集作業の課題点があると考えた。それは、動画再生時間内のどこに欲しいシーンがあるか分からないので動画を一通り確認しなければならず、素材動画の再生時間や数に比例してトリミングの作業量が増加してしまうという点である。この課題点を解消するために、撮影時に印象的だと感じたシーンに情報を付加し、動画確認時にその情報をもとに印象的なシーンを検索する手法を提案した。こうすることで、ユーザは動画を一通り確認しなくても、自身が印象的だと感じたシーンを確認することができ、動画編集作業の効率化を期待できる。以上が本システムの狙いである。

3.2 システム構成

本システムは撮影システム、動画編集システム、サムネイル画像作成サーバ、レンダリングサーバの4つのサブシステムより構成される。撮影システムと動画編集システムはスマートフォン上で実装する。今回使用するスマートフォンをはじめとする携帯情報端末は、ビデオカメラ機能を搭載していることやパソコン並みのインターネット通信が可能だけでなく、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を実現するのに適したデバイスであると考えられる。図3.1に本システムの構成図を示す。ユーザは、スマートフォンを用いた撮影システムで動画を撮影すると同時に印象的だと感じたシーンに対してボタンを押すことでメタデータを付与する。メタデータはテキストファイルに書き込まれ、SDカードに保存される。動画撮影後、動画をサムネイル画像サーバに送信し、一定時間間隔ごとのサムネイル画像を生成する。生成したサムネイル画像をスマートフォンに送信し、動画シーン検索システム内で時系列にリスト表示する。このサムネイル画像は時間ごとの動画のシーンの内容を表す。このサムネイル画像とテキストファイルに書き込まれたメタデータを関連づけることでシーン検索を行う。さらに、シーンの並べ替えや不要シーンの削除などの簡単な編集を行い、その編集内容をレンダリングサーバに送信することで編集内容を反映した動画ファイルを作成する。以上が本研究が目指すシステムの流れである。

本稿では、このうち、撮影システムと動画編集システムの構築、シーンの検索までを行う。本稿で扱うシステムの構成図を図3.2に示す。サムネイル画像生成には山本ら [3][4] が開発したシステム (Cappuccino) を使用し、データのやり取りはオフラインで行うものとする。なお、サムネイル画像作成サーバと動画編集サーバの構築は今後の課題とする。第4章で撮影システムについて、第5章で動画編集システムの詳細について説明する。

ユーザー

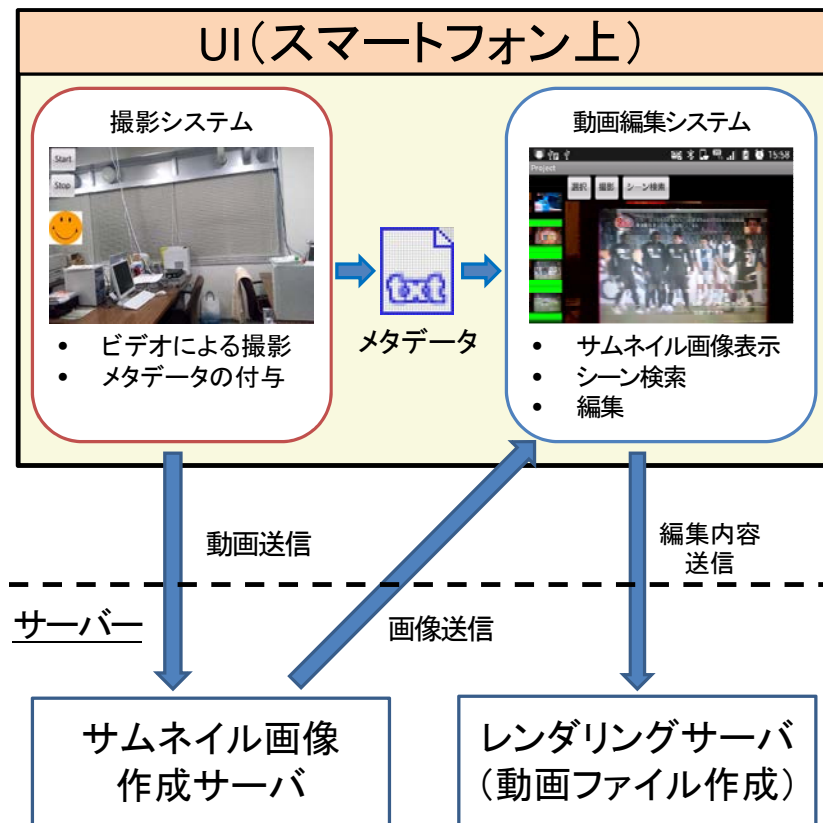


図 3.1: 最終的なシステム構成

ユーザー



図 3.2: 本稿のシステム構成

第4章 印象ボタン押下による情報付与手法を用いた撮影システム

本システムのかなめとなる機能は、撮影者が動画撮影中に印象的だと感じたシーンに対してメタデータを付与することである。撮影中にメタデータを付与することで、動画を確認する段階で印象的だと感じたシーンを容易に検索することが可能となり、トリミングの効率化を行えると期待できる。そこで、本章では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を提案し、このボタンインタフェースを用いた撮影システムについて説明する。なお、本研究では撮影システムを実現するデバイスとしてスマートフォンを使用する。

4.1 印象ボタン

動画にメタデータを付与するには、映像の内容の特徴に着目して自動的に付与することが理想的である。しかしながら、印象的と感じるシーンは人によって異なるので、現段階では、撮影者自身が任意のシーンに対してメタデータを付与する方法が、的確にシーン検索を行えると考えられる。この情報付与手法を実現するために、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を作成した。本研究では、このボタンインタフェースのことを印象ボタンと呼ぶ。図 4.1 に印象ボタンを取り入れた撮影システムの撮影画面を示す。

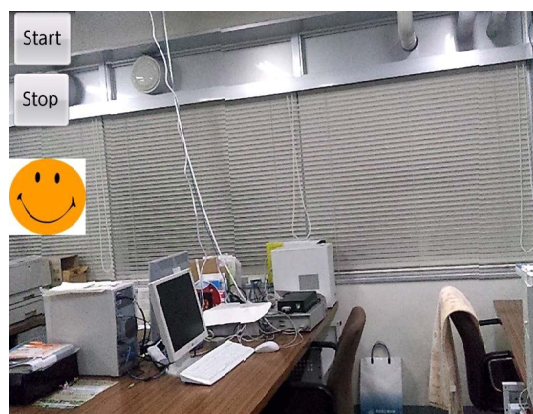


図 4.1: 印象ボタンを取り入れた撮影システム

図 4.1 の start が撮影開始ボタン、stop が撮影終了ボタン、顔マークが印象ボタンである。ユーザは start ボタンで撮影を開始し、撮影中印象的だと感じたときに印象ボタンを押すことでメタデータを付与し、stop ボタンで撮影を終了する。図 4.1 のインタフェースは、印象ボタンがひとつ、かつ、印象的だと感じたときに押すだけというシンプルなものであるためユーザへの負荷が抑えられ、誰にでも容易に扱えるという利点を持つ。しかしながら、印象ボタンがひとつの場合、シーン検索時どのように印象的だったのかまでは判断することができない。

そこで、印象ボタンを複数個用意しそれぞれのボタンに意味を持たせ、撮影者が状況に合わせてボタンを押し分けることで、より詳細なシーン検索が行えるのではないかと考えた。

4.2 印象ボタンの感情カテゴリー分類

本システムでは、単一の印象ボタンだけでなくカテゴリー別に分類した印象ボタンについて検討する。複数のカテゴリー別に分けた印象ボタンを用意することで、シーン検索時、撮影者がどのように印象的だと感じたのかを詳細に知ることができると期待できる。そこで本研究では、撮影者が印象的だと感じたとき何らかの感情が働いていると推測し、感情カテゴリー別に分類した印象ボタンについて検討した。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言われている。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。さらに、感情を表す表現としてオノマトペをボタンの文字情報とすることを考える。オノマトペとは、擬声語(擬音語、擬態語)を意味する語であり、「わくわく」、「はらはら」、などのような表現を包括する概念である。オノマトペは、スポーツの指導において、言葉で表現できない運動の理解を促す効果や、養護教育において、物事を分かりやすく児童に伝える効果があり、実社会での有効性が報告されている [7],[8]。オノマトペの特徴は、覚えやすく直観的、かつ、複雑な動作やニュアンスを簡易化できることである。したがって、単純な形容詞(嬉しい、悲しいなど)をボタンの文字情報とするより、オノマトペを用いる方が感情情報の曖昧さを直観的に伝達することができると考えられる。これらの考察をもとに、図 4.2 に感情カテゴリーで分類した印象ボタンを取り入れた撮影システムの撮影画面を示す。

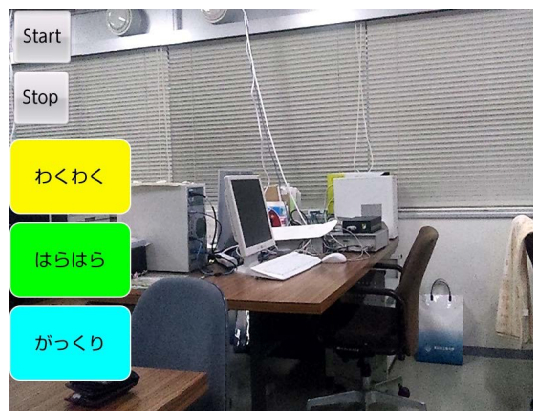


図 4.2: 感情カテゴリーで分類した印象ボタンを取り入れた撮影システム

図 4.2 の「わくわく」、「はらはら」、「がっかり」と記述されたボタンが新たに感情カテゴリーで分類した印象ボタンである。撮影者は撮影中に印象的だと感じたとき、自身の感情に該当する印象ボタンを押すことでシーンに感情情報を付与することができる。こうすることで、印象的だと感じたシーンをさらに分類できるので、より詳細なシーン検索を行うことが可能になる。なお、「わくわく」、「はらはら」、「がっかり」の3つの印象ボタンは一例であり、どのようなオノマトペがボタンとして有効であるか、また何個必要かといった検証は今後の課題とする。以後、感情カテゴリーで分類したものを印象ボタンとして取り扱うものとする。

4.3 印象ボタン押下による情報付加手法

本システムでは、撮影中印象的だと感じたとき印象ボタンを押すことでシーンにメタデータを付与し、次章で述べるシーン検索の手がかりとして用いることを目標としている。この手法を実現するため、印象ボタンの押下情報をテキストファイルに記録する手法を考えた。テキストファイルには、ボタンが押された時刻とその種類を記録する。図 4.3 に情報が記録されたテキストファイルを示す。



図 4.3: ボタン情報の記録

ボタンが押された時刻と種類はコンマで区切られており、左から年、月、日、時、分、秒、種類の順に記録される。本システムでは、印象ボタンを押す場合、撮影者が印象的だと感じている間ボタンを長押しする仕様となっている。理由としては、動画は静止画とは異なり時間という長さの単位があることから、シーン検索時に動画再生時間内のどこからどこまでが印象的なシーンであったのかを反映させるためである。記録するボタンの種類は、撮影開始を表す「Start」、わくわくしたと感じたシーンの開始を表す「WakuStart」、わくわくしたと感じたシーンの終了を表す「WakuStop」、はらはらしたと感じたシーンの開始を表す「HaraStart」、はらはらしたと感じたシーンの終了を表す「HaraStop」、がっかりしたと感じたシーンの開始を表す「GakuStart」、がっかりしたと感じたシーンの終了を表す「GakuStop」の全7種類となる。これらの情報を記録したテキストファイルは携帯情報端末内のSDカードに保存される。次章において情報が記録されたテキストファイルを利用した動画シーン検索の実現法について説明する。

第5章 動画シーン検索の実現法

本章では、第4章の撮影システムから取得したメタデータから動画のシーン検索の実現法について説明する。さらに、スマートフォンの特徴であるタッチインターフェースを用いた動画編集インターフェースについて説明する。はじめに、動画編集システムのインターフェースについて説明する。

5.1 ユーザインターフェース

図5.1に動画編集システムのインターフェースを示す。



図 5.1: 動画編集検索システム

画面上部左から、「選択」ボタン、「撮影」ボタン、「シーン検索」ボタンを配置した。また画面中央部には再生動画画面、画面左に再生動画を一定時間間隔ごとに分割したサムネイル画像を用いたスクロール可能なシークバーを配置した。このサムネイル画像は動画の時間毎のシーンの内容を表しており、上から順に時系列で並んでいる。さらにサムネイル画像には先頭から順に0、1、2…などのようにID番号が割り振られている。このサムネイル画像と撮影システムから取得したメタデータを関連付けることでシーンの検索を行う。

5.2 動画からサムネイル画像を生成

シーン検索時、ユーザが動画再生時間内のどこからどこまでが印象的なシーンであるか視覚的に分かりやすくするため、撮影動画を一定時間間隔ごとにサムネイル画像で表示することを考える。サムネイル画像とは、多数の画像を一覧表示するために縮小された画像のことである。本研究では、山本ら [3][4] が開発したシステム (Cappuccino) を利用して一定時間間隔ごとのサムネイル画像を生成する。図 5.2、図 5.3 に Cappuccino のインタフェースを示す。

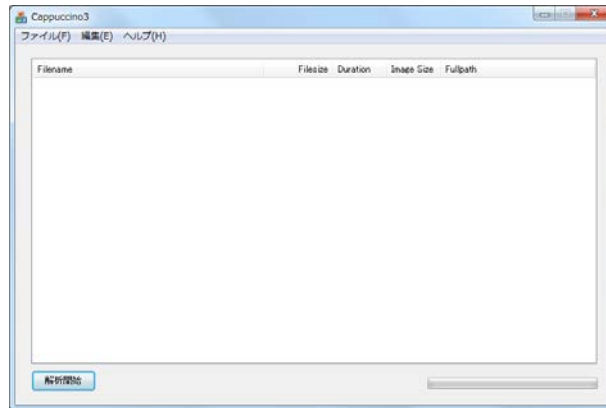


図 5.2: Cappuccino

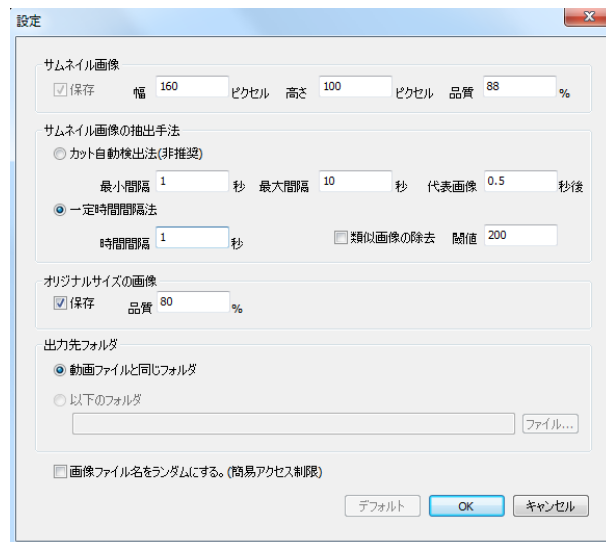


図 5.3: Cappuccino の設定画面

図 5.2 の中央のスペースに動画をドラッグ & ドロップし、解析開始ボタンをクリックすることでサムネイル画像が生成される。サムネイル画像生成の時間間隔や画像サイズの変更は図 reffig12で行う。撮影システムで撮影した動画をこの Cappuccino に取り込むことで一定時間間隔ごとのサムネイル画像を生成し、動画編集システム内で使用する。

5.3 撮影システムからの情報取得とサムネイルへの関連付け

撮影システムでは印象ボタンの押された時刻と種類をテキストファイル (図 4.3) に記録した。ボタンが押された時刻と種類はコンマで区切られており、左から年、月、日、時、分、秒、種類の順に記録されている。動画編集システムでは、「シーン検索」ボタンによりテキストファイルの情報を読み込む。テキストファイルの情報は上から順番に一つずつ読み込んでいき、ボタンの種類から時刻を格納する変数を決定し、それぞれの変数に時刻を格納する。時刻が格納された変数を用いて撮影開始から何秒後に印象ボタンによる記録がされたのかを導出する。図 4.3 の場合、例えば撮影の開始を表す Start の日、時、分、秒をそれぞれ D_{st} 、 h_{st} 、 m_{st} 、 s_{st} とし、わくわくしたと感じたシーンの開始を表す WakuStart の日、時、分、秒をそれぞれ D_{wk} 、 h_{wk} 、 m_{wk} 、 s_{wk} としたとき、撮影開始からわくわくボタンの押下が開始されるまでの時間 T_{wk} は、

$$T_{wk} = (h_{wk} - h_{st}) * 3600 + (m_{wk} - m_{st}) * 60 + (s_{wk} - s_{st}) \quad (5.1)$$

となる。(5.1) 式は時刻の差を秒に変換した式である。ただし、例外として動画撮影中に日付が変わる場合がある。撮影中に日付が変わった場合は、

$$T_{wk} = (h_{wk} + 24 - h_{st}) * 3600 + (m_{wk} - m_{st}) * 60 + (s_{wk} - s_{st}) \quad (5.2)$$

となる。日付が変わったかどうかは撮影開始時の日付と印象ボタンが押されたときの日付によって判定する。記録されたボタン押下情報全てに対して (5.1)、(5.2) と同様の計算を行い、撮影開始から何秒後に印象ボタンによる記録がされたのかをそれぞれ導出する。その後、導出した値とサムネイル画像の ID 番号を照合し、サムネイル画像に目印を付ける。サムネイル画像は動画の時間毎のシーンの内容を表しているため、サムネイル画像に目印を付けることは撮影中に印象的だと感じたシーンを検索することに帰着される。ユーザは、動画編集システムによって、動画をひと通り再生し確認しなくても、自身が印象的だと感じたシーンを即座に発見することが可能になり、トリミング作業を効率良く行うことができるようになる。

5.4 編集機能

本システムでは、トリミング作業を効率化するため撮影システムで動画撮影中に印象的だと感じたシーンに対して印象ボタン押下によってメタデータを付与し、動画編集システムで印象ボタン押下によって得られたメタデータから印象的だと感じたシーンに目印を付けることでシーンの検索を行った。さらに本稿では取り扱わないが、最終的なシステム構成 (図 3.1) として、動画シーン検索システム内で簡単な動画編集を行ったのち、編集内容を動画編集サーバに送信することで動画ファイルを作成することを目指している。そこで、本稿では今後、動画編集サーバを構築することを想定し、動画編集インタフェースを作成した。以下に動画編集インタフェースの機能を示す。

- シーンの並べ替え
- 不要シーンの削除

これらの編集方法については順を追って説明する。

5.4.1 シーンの並び替え

はじめに「シーンの並び替え」について説明する。本研究では、システムを実現するためのデバイスとしてスマートフォンを利用している。スマートフォンはタッチインターフェースによる直感的な操作が特徴であり、本システムでもタッチインターフェースによるシーンの並べ替えを行う。シーンの並び替えには動画のシーンの内容を表すサムネイル画像を用いる。移動したいサムネイル画像をロングタッチで選択し、ドラッグ&ドロップすることでシーンを任意の場所へ配置することができる。シーンの並び替えの様子を図5.4に示す。図5.4では、×のマークが描かれたシーンを表すサムネイル画像をドラッグ&ドロップ操作で○のマークが描かれたシーンを表すサムネイル画像と並べ替えを行っている。この機能によって、ユーザは動画の構成を直感的に変更することが可能となる。

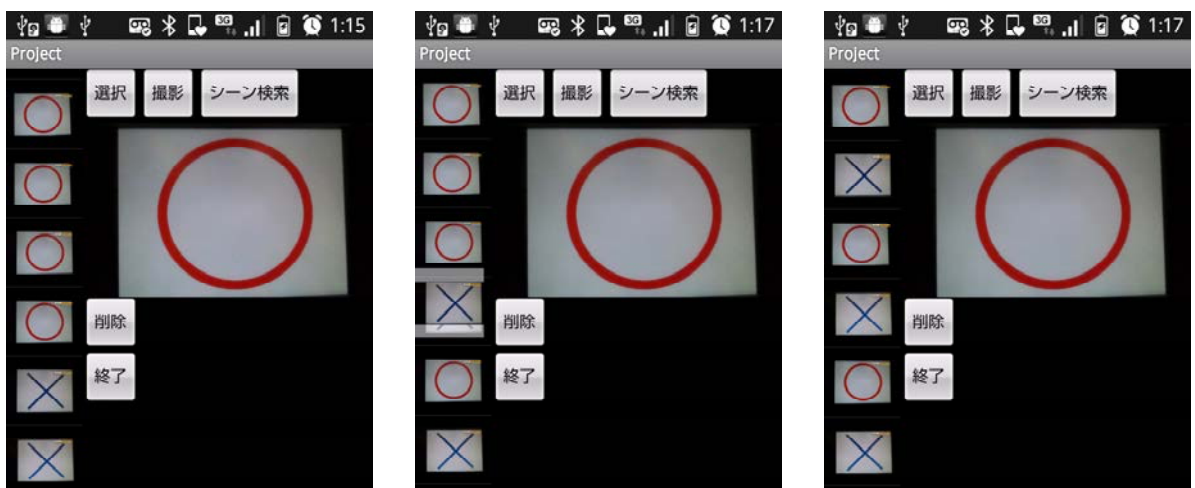


図 5.4: シーンの並べ替えの様子

5.4.2 不要シーンの削除

次に「不要シーンの削除」について説明する。動画シーン検索システムによって印象的なシーンを検索することで不要なシーンも同時に浮き彫りになることが予想される。そこで、不要なシーンを削除するためのボタンインタフェースを作成した。図5.5の「削除」ボタンを押すと削除モードに切り替わる。削除モードの状態では不要なシーンに該当するサムネイル画像をタッチすることで不要なシーンの削除を行うことができる。図5.5では、×のマークが描かれたシーンを表すサムネイル画像と *bigcirc* のマークが描かれたシーンを表すサムネイル画像の間の不要なシーンを表すサムネイル画像を削除している。なお、削除モードの終了は「終了」ボタンで行う。

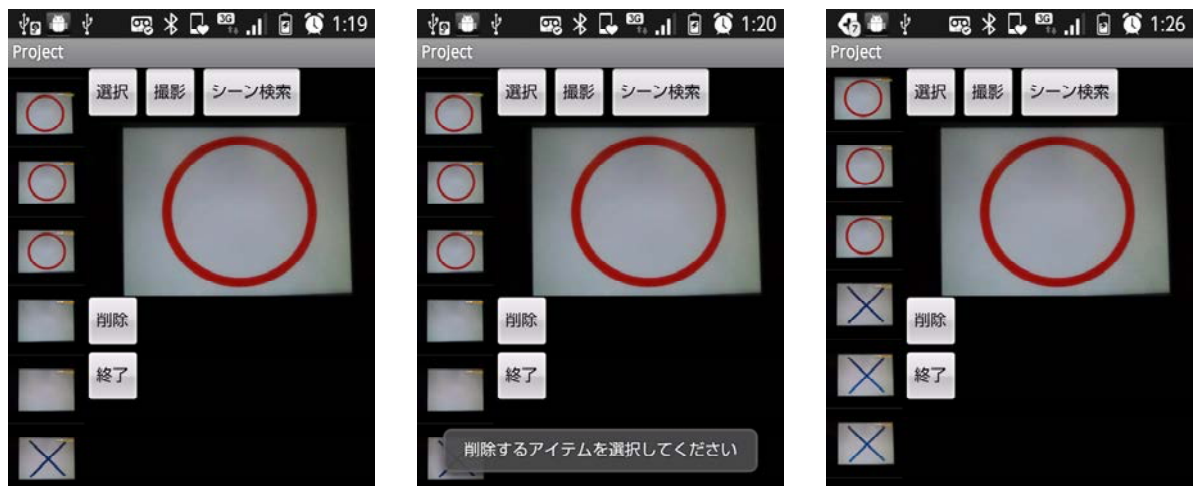


図 5.5: 不要シーン削除の様子

第6章 実験と考察

本章では、前述の提案システムを構築し、シーン検索を行う実験を行った。システムの動作を確認することで、その有用性を確認する。

6.1 実験システム

本システムは撮影システムと動画シーン検索システムの二つのサブシステムから構成されている。今回、システムを実装するデバイスとして Android[10] 搭載のスマートフォンを使用する。

Android とは、アメリカ合衆国のソフトウェア会社 Google[11] が発表したスマートフォンやタブレット PC などの携帯情報端末を主なターゲットとして開発されたプラットフォームである。Android はオープンソースで開発環境が無償で提供されており、誰でもアプリを自由に開発ができるという特徴がある。また、アプリ配布に関して Google による審査は必要最低限しか行われないので、将来的にアプリとしてシステムを配布し、システムの評価実験を行うことを考えてる本研究にとっては大きなメリットである。さらに、現在、Android 搭載のスマートフォンは急速に普及しはじめており、国内スマートフォン OS シェアではトップに立っており、今後も普及し続けると予想される。このような背景から、評価実験を行う場合、より多くの被験者を獲得できることが期待できる。以上が Android 搭載のスマートフォンをデバイスとして使用する理由である。

また、動画編集システム内で使用するサムネイル画像生成には第5章で説明した Cappuccino を用いる。

6.2 実験 1:提案システムの動作確認

実際に撮影システムを用いて動画を撮影し、動画編集システムを用いてシーンの検索を行う。撮影システムでは撮影中印象的だと感じたシーンに対して印象ボタンを押すことでメタデータを付与する。しかしながら、日常行事やスポーツイベントなどで撮影した場合、印象的だと感じるシーンは人によって異なるので、シーン検索が行えているかどうか客観的な評価が行いにくい。そこで、今回の実験では撮影時いくつかの条件を設定することで客観的にシステムの動作を確認する。

- 条件 1:被写体を用意し、撮影システムで撮影する
- 条件 2:撮影中特定のシーンが現れたら、特定のシーンに対応する印象ボタンを押す

上記のような二つの条件下で実験を行い、シーン検索時、特定のシーンを表すサムネイル画像に目印が付与されているかを確認することでシステムの動作を確認する。実験の流れを順に説明する。

6.2.1 撮影システムによる撮影と情報付加

はじめに、今回の実験で条件として設定した被写体と特定のシーンについて説明する。今回の実験では、無地のスライドと○、△、×というマークが描かれたスライドによるスライドショーを被写体(図 6.1)として使用する。

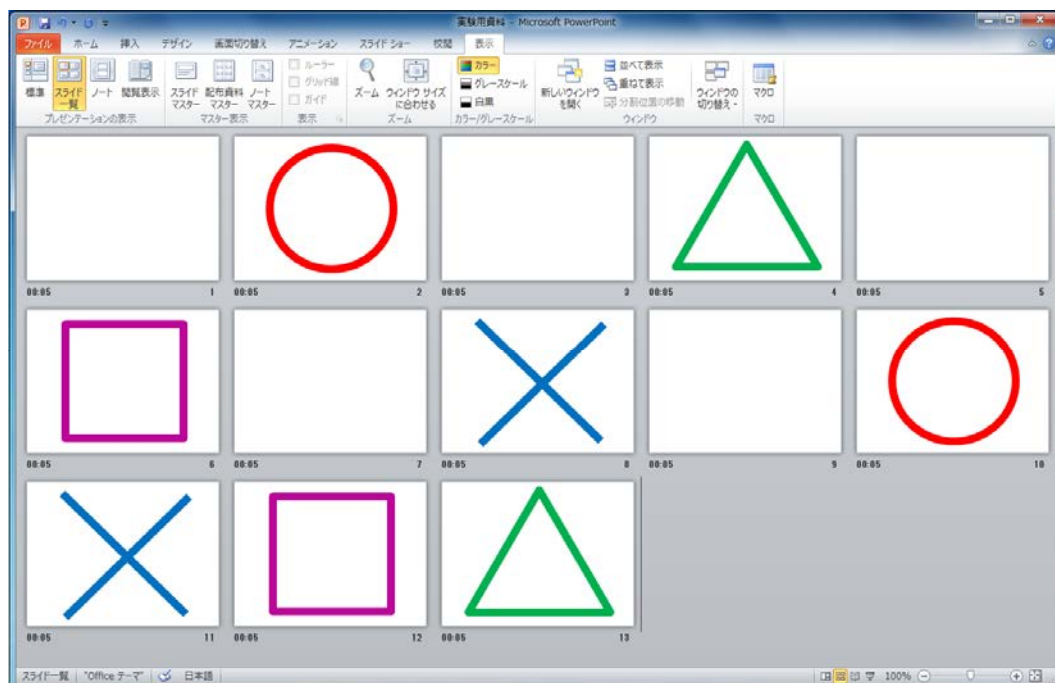


図 6.1: 被写体として使用するスライド

本実験では、無地のスライドを 5 枚、○、△、×のスライドを 2 枚ずつ計 13 枚用意した。なお、13 枚のスライドはランダムに配置され、5 秒間隔で切り替わる仕様となっている。このスライドショーを被写体として扱い、スマートフォンを用いた撮影システムで撮影する。5 種類のスライドのうち、○、△、×の 4 種類を特定のシーンに設定する。

撮影中特定のシーンが表示されている間、印象ボタンを長押しする。第4章では、感情カテゴリーに分類した印象ボタンを定義したが、今回の実験では、特定のシーンに対応する○、△、□、×の4つの実験用印象ボタンを作成した。図6.2に実験用印象ボタンを示す。

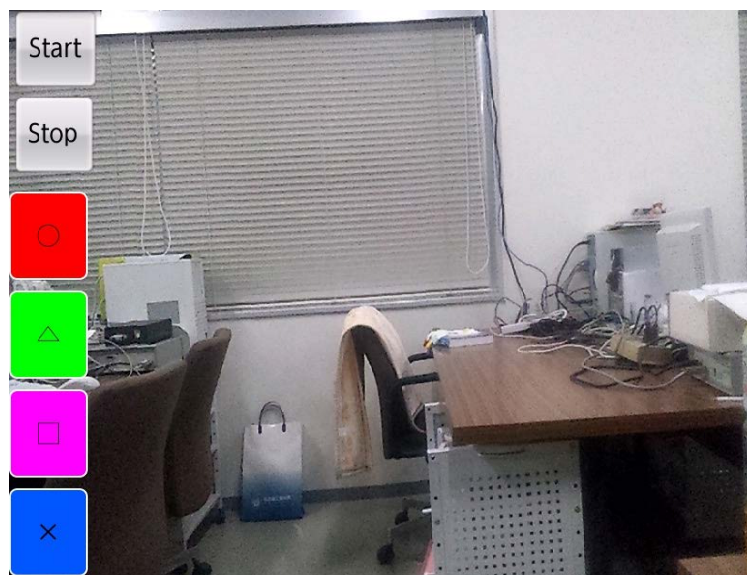


図 6.2: 実験用印象ボタン

特定のシーンが表示されたとき、対応する印象ボタンを押すことでメタデータを付与する。図6.3に実験の様子を示す。

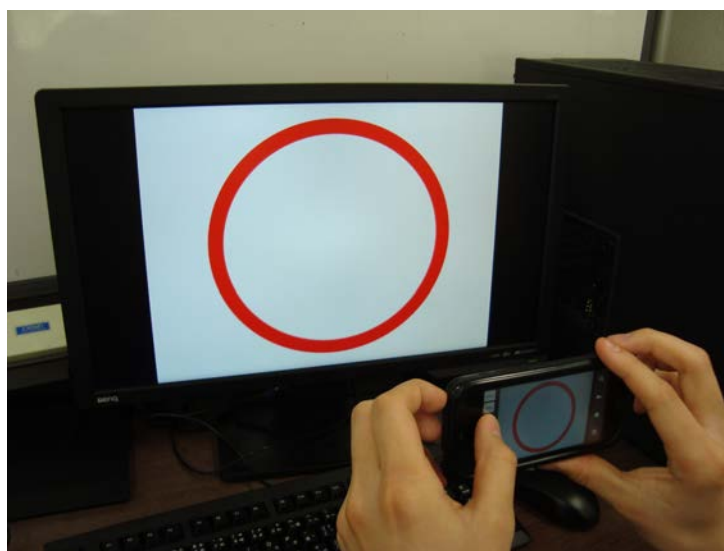


図 6.3: 実験の様子

撮影中に押した印象ボタンの情報はテキストファイルに書き込まれる。書き込まれる情報は印象ボタンの押された時刻とその種類である。実験によって図 6.4 のような押下情報を取得した。

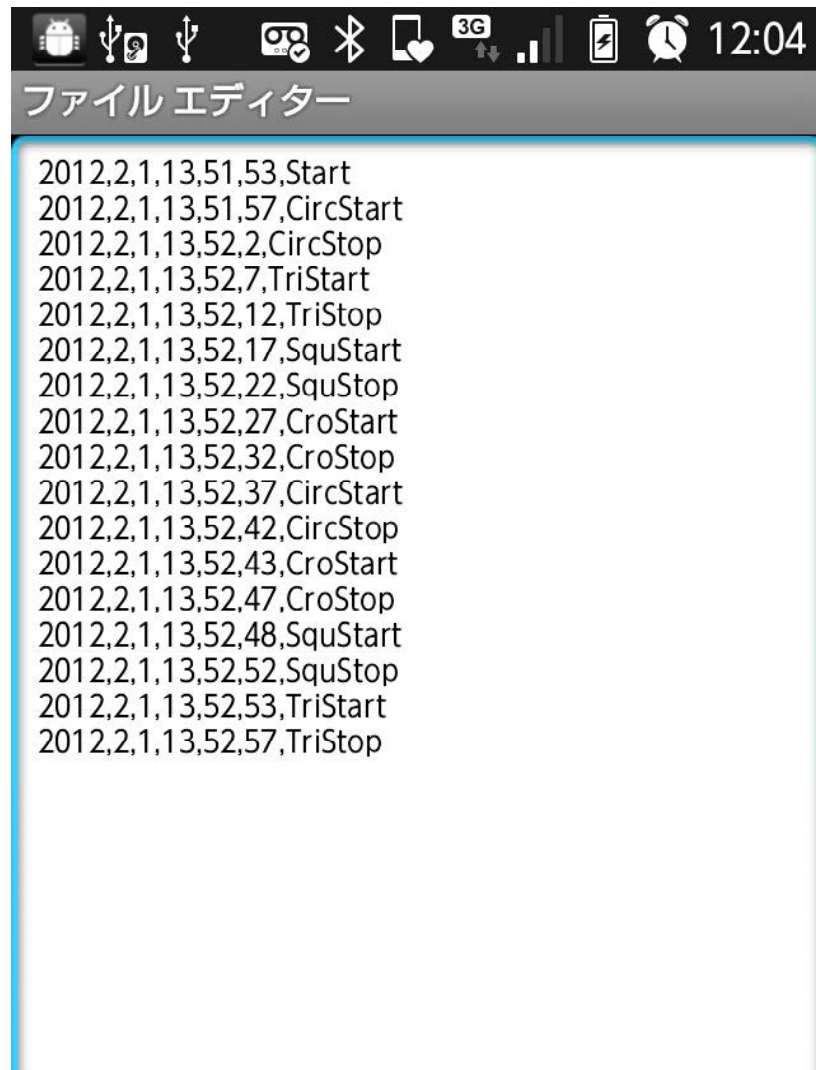


図 6.4: 押下情報を記録したテキストファイル

記録したボタンの種類は、撮影開始を表す「Start」、○ボタンの押下開始を表す「CircStart」、○ボタンの押下終了を表す「CircStop」、△ボタンの押下開始を表す「TriStart」、△ボタンの押下終了を表す「TriStop」、ボタンの押下開始を表す「SquStart」、ボタンの押下終了を表す「SquStop」、×ボタンの押下開始を表す「CroStart」、×ボタンの押下終了を表す「CroStop」の全9種類である。押下情報の書き込まれたテキストファイルはスマートフォン内のSDカードに保存される。

6.2.2 Cappuccino によるサムネイル画像生成

次に、撮影システムで撮影した動画から一定時間間隔ごとのサムネイル画像を生成する。サムネイル画像生成には第5章で説明した Cappuccino を使用する。本実験では、サムネイル画像生成間隔を1秒とした。スマートフォンで撮影した動画を Cappuccino へ取り込み、1秒間隔ごとのサムネイル画像を生成した後、スマートフォンへサムネイル画像を返信する。なお、スマートフォンと Cappuccino とのデータのやり取りはオフラインで行うこととする。Cappuccino で生成したサムネイル画像は動画編集システム内で使用する。

6.2.3 シーン検索

動画シーン検索システムのインターフェースを図 6.5 に示す。

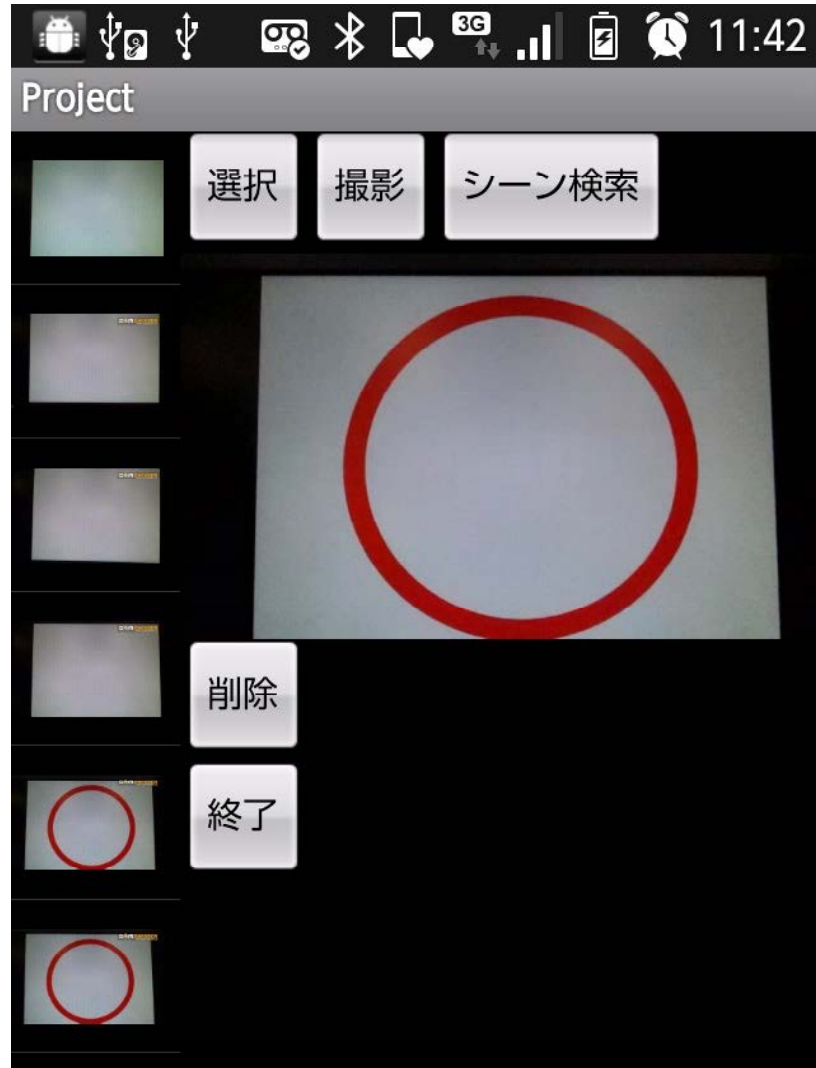


図 6.5: インターフェース

画面中央が撮影した動画の再生画面、画面左が Cappuccino により生成した 1 秒間隔ごとのサムネイル画像を用いたスクロール可能なシークバーである。サムネイル画像は動画の時間毎のシーンの内容を表している。なお、サムネイル画像は上から時系列に並んでおり、先頭から順に ID 番号が割り振られている。画面上部の「シーン検索」ボタンを押下し、撮影システムにより記録した印象ボタンの押下情報 (図 6.4) を読み込む。読み込んだ押下情報より、撮影開始から何秒後に印象ボタンが押されたのかを導出する。導出は 5.3 節で説明した方法で行う。導出された値とサムネイル画像の ID 番号を関連付け、シーンの内容を表すサムネイル画像に目印を付与することで撮影者が印象的だと感じたシーンをユーザに提示する。今回の実験では、印象ボタンが押されたシーンと判定されたサムネイル画像に対して色付けを行うことで、ユーザに結果を提示する。サムネイル画像へ付与される色は実験用印象ボタンの色と連動している。今回の実験では、印象ボタンの ○ ボタンが赤、△ ボタンが緑、□ ボタンが紫、× ボタンが青となっている。実験結果を図 6.6、図 6.7 に示す。

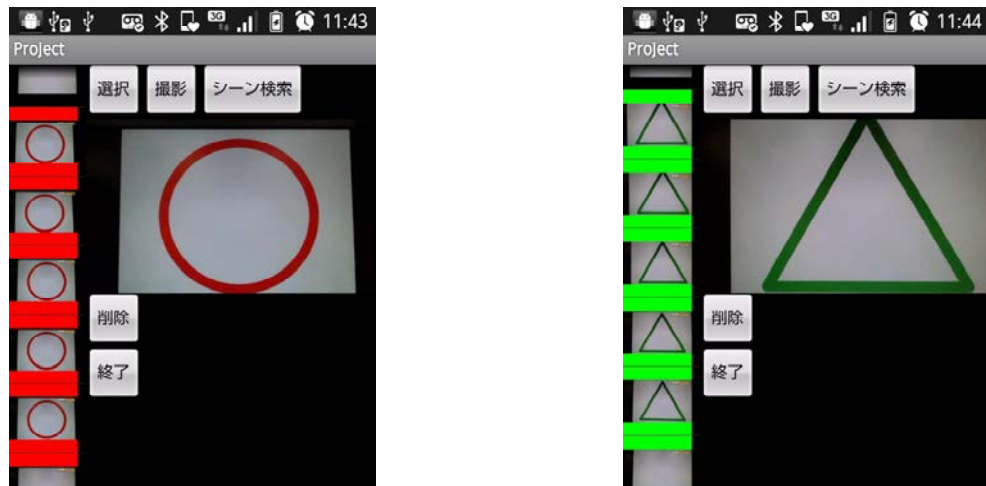


図 6.6: シーン検索結果



図 6.7: シーン検索結果

図 6.6、図 6.7 より、特定のシーンと設定した○、△、□、×のサムネイル画像に対して印象ボタンに対応する色付けがされていることが確認できる。この結果から撮影システムで付与したメタデータを用いてシーンの検索が行えていることが確認できた。

6.3 シーン検索効率に関する評価実験

本研究の目的は、撮影中印象的だと感じたシーンに対して印象ボタンを押下することでメタデータを付与し、さらにそのメタデータを利用することで効率良く欲しいシーンを見つけ出すことである。そこで、本システムのシーン検索の有効性を評価するため、被験者8人に実験システムを体験してもらい、簡単なアンケートを行った。実験では事前に撮影システムを用いて動画を撮影する。撮影中特定のシーンに対して印象ボタンを押下することで、特定のシーンに目印を付与する。撮影した動画を使用して、被験者に以下の3つの手法で特定のシーンを探してもらおう。

- 手法1:被験者に既存の動画再生ソフトで動画を閲覧してもらい、特定のシーンを探してもらおう
- 手法2:被験者に動画を一定時間間隔(1秒間隔)ごとに分割したサムネイル画像のリストを閲覧してもらい、特定のシーンを探してもらおう
- 手法3:被験者に撮影システムで取得したメタデータを用いてシーンの検索を行った動画編集システムで動画を閲覧してもらい、特定のシーンを探してもらおう
- 被験者に手法3、4を踏まえた上で、手法5の提案手法によるシーン検索の分かりやすさを5段階評価でアンケートに答えてもらおう

なお、動画の撮影時間を2分とし、特定のシーンは撮影中複数回現れる。5段階評価の目安は

- 1: 手法1、手法2と比べて特定のシーンを探す手間がかかった
- ⋮
- 3: 手法1、手法2と比べて特定のシーンを探す手間は変わらなかった
- ⋮
- 5: 手法1、手法2と比べて効率良く特定のシーンを探すことができた

とした。アンケートの結果を図6.8に示す。

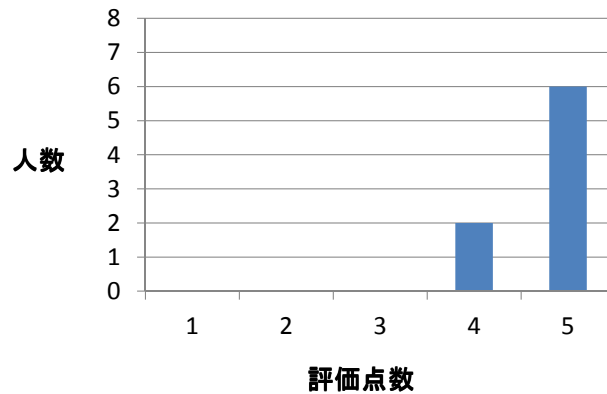


図 6.8: シーン検索の分かりやすさの評価結果

アンケートの結果、動画編集システムによるシーン検索の分かりやすさに対して、被験者 8 人中 6 人から評価点 5 という高い評価を得られた。また、動画から全ての特定のシーンを見つけるまでにかかった時間を計測したところ、表 6.1 のような結果が得られた。

	手法 1	手法 2	手法 3
被験者 1	90 秒	90 秒	60 秒
被験者 2	119 秒	43 秒	25 秒
被験者 3	120 秒	60 秒	32 秒
被験者 4	80 秒	51 秒	30 秒
被験者 5	63 秒	48 秒	35 秒
被験者 6	125 秒	110 秒	50 秒
被験者 7	55 秒	118 秒	37 秒
被験者 8	61 秒	48 秒	27 秒
平均時間	89.1 秒	71 秒	37 秒

表 6.1: シーン検索までにかかった時間

表 6.1 の結果より、全ての被験者においてシーン検索を行った動画編集システムを用いた方が短い時間で特定のシーンを見つけ出すことができたことが確認できる。これらの結果から、本システムにより効率良く欲しいシーンを見つけ出すことができることが確認できた。また、本実験では再生時間を比較的短い 2 分と設定したが、再生時間がより長い動画の場合、シーン検索手法によって更なる時間短縮が期待できる。しかし、高い評価点を付けた被験者から「所々で特定のシーンとは関係のないシーンにも目印が付いていて戸惑った」という意見を頂いた。この原因として、撮影者の印象ボタンの押し間違いが検索結果に反映されてしまったと考えられる。

本実験では、タッチインターフェースを用いた動画撮影システムのインターフェースの使いやすさについても5段階評価アンケートに答えてもらった。5段階評価の目安は

1:使いにくかった

:

3:実用に耐えうる使いやすさ

:

5:快適な使いやすさ

とした。アンケートの結果を図6.9に示す。

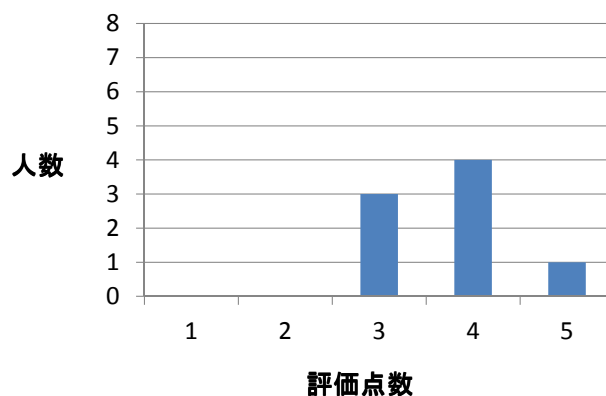


図 6.9: インターフェースの使いやすさの評価結果

アンケートの結果、8人中5人から評価点4以上の評価を得ることができた。特に動画の時間ごとのシーンの内容をサムネイル画像で表現している点や、タッチインターフェースによる直感的な操作に関して評価を得られた。しかしながら、高い評価点を付けた被験者から「シーンの内容を表すサムネイル画像をタッチすることで該当する動画のシーンへジャンプできたら更に良い」という意見や、「サムネイル画像のサイズが小さく若干見づらかった」などの意見を頂いた。これらの意見を参考に、今後更なるインターフェースの改良を目指していく。

6.4 考察

今回の実験では、システムの動作を確認することで、その有用性を確認することが目的であった。実験の結果、特定のシーンを表すサムネイル画像に対して色付けがされ、特定のシーンの検索ができていることが確認できた。また、評価実験により本システムを利用することで効率良く欲しいシーンを見つけ出すことができることが確認できた。ここで、シーン検索結果を分析し、以下の2つの事柄について考察する。

- 特定のシーンを表すサムネイル画像以外のサムネイル画像に対して目印が付与されていることがある
- 特定のシーンと印象ボタンの対応関係が一致していないことがある

これらは全て、シーンの切り替わり直後に起きていることが確認できた。このような事が起こる原因として、印象ボタンの性質が関係していると考えられる。印象ボタンは撮影者が印象的だと感じたシーンに対してメタデータを付与するボタンインタフェースであり、撮影者が印象的だと感じている間、ボタンを長押しする仕様となっている。本実験では、特定のシーンが表示されている間、印象ボタンを長押しし、シーンが切り替わったら印象ボタンから指を離す、または別の印象ボタンを押すという操作を行うことになるが、撮影者がシーンの切り替わりを認識してから実際に操作を行うまでにタイムラグがあると推察される。また、ボタンを複数個用意したことによってボタンの押し間違いが起こったと推察される。これらのことが検索結果に反映されたことで、上述した2つの事柄が起きたと推察される。このように印象ボタンは撮影者自身が任意のシーンにメタデータを付与することができる半面、撮影者のボタン押下のタイムラグやボタンの押し間違い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ボタンインタフェースであると言える。このような特徴から、印象ボタンを用いたメタデータ付与によるシーン検索の質は撮影者に依存すると考えられる。つまり、ユーザにとって最適な撮影システムの構築が、動画編集システムでのシーン検索の質の向上につながると捉えることができる。

次に、一般ユーザが日常生活やスポーツイベントなどの種々のイベントにおいて、本システムを利用した場合について考察する。一般ユーザが動画を撮影する場合、日常生活、イベント、スポーツ観戦など様々な場面が考えられる。そのような場面において印象的だと感じるシーンは撮影者によって異なる。また、シーン検索時、印象ボタンを押したシーンが具体的にどのようなシーンであったかをユーザに提示することが重要であると考えられる。そこで、本研究では、撮影者が撮影中、印象的だと感じたとき何らかの感情が働いていると推測し、感情カテゴリーで分類した印象ボタンを作成した。感情カテゴリーであれば、撮影者の感情に則して印象ボタンを押すので、どのような場面であっても適用可能であり、高い汎用性が期待できる。今後の研究において、感情カテゴリーに分類した印象ボタンに対して主観評価を行い、最適なボタンインタフェースを構築することを目指す。

第7章 むすび

7.1 まとめ

ビデオデバイスや動画編集ソフトの普及、動画投稿サイトの発展などの背景から動画編集が一般ユーザにとってより身近なものになっているといえる。そこで本研究では、動画編集を効率良く行うことが重要であると考え、スマートフォン上で撮影システムと動画編集システムの二つのシステムを構築した。撮影システムでは印象ボタンを作成し、撮影中に印象的だと感じたシーンに対してメタデータを付与する手法を提案した。動画編集システムでは撮影システムによって付与したメタデータを用いてシーンの検索を行った。これらのシステムにより、動画の中から効率良くシーンを見つけ出すことが可能となった。効率良く欲しいシーンを見つけ出すことができるようになることで動画編集の負担軽減が期待できる。将来的には、動画編集インタフェースで編集した内容をレンダリングサーバに送ることで動画ファイルを作成することを考えている。

7.2 今後の課題

以下に本研究の今後の課題をあげる。

- サムネイル画像作成サーバを構築し、オンラインでのデータのやり取り

現時点では、スマートフォンとサムネイル画像作成システム Cappuccino とのデータのやり取りはオフラインで行っている。そこで、サムネイル画像作成サーバを構築し、オンラインでのデータのやり取りを実現することで、より使い勝手の良いシステムに改良する。

- 印象ボタンの改良

本研究では、撮影者が印象的だと感じたとき、何らかの感情が働いていると推測し、感情カテゴリで分類した印象ボタンを採用している。さらにボタンの文字情報としてオノマトペを用いることで感情情報の曖昧さを直感的に伝達できると考えている。そこで、どのようなオノマトペが文字情報として有効であるか、またユーザにとって最適な個数はいくつなのかを検証する。

- サムネイル画像表示間隔の変更

動画シーン検索システム内において1秒間隔ごとのサムネイル画像を表示していた。しかしながら、長時間の動画を扱う場合、サムネイル画像の枚数が膨大になってしまう。そこで、スマートフォン上でサムネイル画像の表示間隔を変更可能にすることでこの問題を解決する。

- 動画編集インタフェースの拡張

シーンの並べ替えや不要シーンの削除だけでなく、指定した範囲を抜き出すトリミング機能などを追加する。

- レンダリングサーバの構築と動画ファイル作成

スマートフォン上で編集した内容をレンダリングサーバに送信することで、編集内容を反映した動画ファイルを作成する。

これらの課題を含め、動画撮影システムと動画編集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う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いただき、ご指導を賜った名古屋工業大学 舟橋健司 准教授、名古屋工業大学 山本大介助教に心から感謝いたします。また、本研究に対してご検討、ご協力いただきました名古屋工業大学 伊藤宏隆助教に心から感謝致します。最後に、本研究を進めるにあたり多大な協力をいただいた舟橋研究室諸氏に心から感謝いたします。

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] 山本 大介, 増田 智樹, 大平 繁輝, 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.
- [4] 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
- [5] 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
- [6] 西岡 欣亮, 住野 優, 大西 真晶, 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
- [7] 有働 眞里子, 高野美由紀:養護学校小学部の授業に見られるオノマトペ的発話-対話活性化の言語学的要因-, 学校教育学研究, Vol.19, pp.17-26(2007)
- [8] 藤野 良孝, 井上 康生, 吉川 政夫, 仁科 エミ, 山田 恒夫:運動学者のためのスポーツオノマトペ電子辞典の開発と評価, 日本教育工学会論文誌, Vo.29, No.4, pp.515-525(2006)
- [9] 石橋 賢, 宮田 一乗:”色彩効果による画像への感情記録の可能性に関する検討”, FIT2010 (第 9 回情報科学技術フォーラム)
- [10] Android:”Android”,<http://www.android.com/>
- [11] Google:”Google”,<http://www.google.co.jp/>