

Title	視認性とプレゼンス性を両立させる遠隔講義映像の送受信方法
Author(s)	小原, 理
Citation	
Issue Date	2003-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1687
Rights	
Description	Supervisor:篠田 陽一, 情報科学研究科, 修士

修 士 論 文

視認性とプレゼンス性を両立させる
遠隔講義の送受信方法

北陸先端科学技術大学院大学
情報科学研究科情報システム学専攻

小 原 理

2003年3月

修 士 論 文

視認性とプレゼンス性を両立させる
遠隔講義の送受信方法

指導教官 篠田 陽一教授

審査委員主査 篠田 陽一 教授
審査委員 日比野 靖 教授
審査委員 丹 康雄 助教授

北陸先端科学技術大学院大学
情報科学研究科情報システム学専攻

110033 小原 理

提出年月: 2003 年 2 月

目次

第1章	はじめに	1
1.1	背景	1
1.2	本研究の目的	1
1.3	本論分の構成	2
第2章	遠隔講義システム	4
2.1	既存のシステム	4
2.1.1	WIDE University School of Internet	4
2.1.2	Web Collaboration	4
2.1.3	TIDE Project	5
2.2	遠隔講義の分類	6
2.2.1	本研究での対象	8
第3章	本研究の概要	9
3.1	本研究の提案	9
3.2	実現手法の調査	9
3.2.1	赤外線を用いる手法	9
3.2.2	可視光を用いる手法	12
3.2.3	点光源を用いる手法	14
3.2.4	スクリーン位置の取得	14
3.3	手法の検討	15
第4章	システムの設計/実装	19
4.1	システムの設計	19
4.1.1	環境	21
4.2	実装に伴う諸問題と解決法	21
4.2.1	PowerPoint の同期	21
4.2.2	PowerPoint からのイメージの取得	22
第5章	映像からの領域抽出	28
5.1	スクリーンの領域抽出	28

5.1.1	特徴点の抽出	29
5.1.2	Hough 変換による直線検出	30
5.1.3	評価関数による選別	32
5.2	講師の領域抽出	36
5.2.1	背景差分法	36
5.2.2	背景の抽出	37
第 6 章	評価	39
6.1	評価実験	39
6.1.1	輝度による検出精度	39
6.1.2	PowerPoint テンプレートの検出	40
6.1.3	圧縮に対する耐性	40
6.2	考察	41
第 7 章	おわりに	43
7.1	おわりに	43
7.2	今後の課題	43

目 次

1.1	はめ込み合成のイメージ図	2
2.1	SOIの講義映像	5
3.1	赤外線 LED アレイの配置図	10
3.2	撮影画像及びヒストグラム	11
3.3	光源をスクリーン直近に配置した例	16
3.4	図 3.3 を二値化したもの	16
3.5	可視光の輝度情報を二値化した例	17
3.6	色相の差をあらわした例	17
3.7	ピクセルの色情報の距離を二値化した例	18
4.1	全体の構成	19
4.2	Mixer 部の内部構造	20
4.3	RPT	23
4.4	画面のプロパティ	25
4.5	処理の流れ	26
5.1	サンプルの初期状態	28
5.2	二値化及び特徴点の例	29
5.3	Hough 変換 (特徴点)	30
5.4	Hough 変換 ($\rho - \theta$ 平面に写像)	31
5.5	Hough 変換 (点 P_1, P_2 を通る直線)	31
5.6	Hough 変換の結果を $\rho - \theta$ 平面に描画した例	33
5.7	移動体が横切った際の背景抽出の例	38
6.1	劣化処理した画像	42
6.2	図 6.1 のスクリーン領域検出	42

第1章 はじめに

1.1 背景

ネットワークの急速な普及に伴い、遠隔地にいる人とのコミュニケーション手段も多様化している。また、広帯域伝送も同様に普及を見せ、従来難しかった映像を用いたコミュニケーション手段、携帯電話間での映像通信や遠隔教育システム、遠隔会議システムなども一般的になりつつある。

遠隔教育や遠隔会議システムはすでに多く実用化されており、場所を問わないメリットを生かし、英会話教育や会議などに幅広く利用されている。そのなかで、講義映像を遠隔地にネットワークを通して伝送する遠隔講義も行われるようになった。すでに遠隔講義システムの環境構築は各方面から進んでおり、専用 ATM 網を利用した試みや、衛星通信を用いた試みなど多種の試みが盛んに行われている。

遠隔講義において、講義者は、講義映像を提供すると共に資料を提示する。講義の受講者は講義社と同一の資料を共有しながら講義映像を踏まえ受講する。この遠隔講義の資料提示に目を向けると、対面講義で使われる黒板等をそのまま撮影し、講義映像と共に伝送するスタイルの資料提示法では高解像度での映像伝送が必要となり、設備の面で利用提供に制限が生まれ更なる普及の妨げとなる。

そこで、電子媒体教材による遠隔講義が注目されている。共有ボード wb[1] や、Web ページ等で教材を提示し、講師と受講者の間で同期して動作させるこの手法は、これらのツールによる鮮明な画像提供や、講師と受講者間での資料の共有の面で有効である。

しかし、電子媒体教材と講義映像を分離することは、遠隔講義システムの良否である、一つの講義室で講師と受講者が直接対話する現在の講義形態にいかにか近づけるかという問題の中の一つの要素であるプレゼンス性が損なわれる。

このように、遠隔講義においては映像や音声の伝送と共に資料提示の面が極めて重要な要素となるが、遠隔講義は一般の講義と比べ資料提示が視覚的もしくは構成的に制限を受け、受講者の講義に対する集中の妨げとなる場合がある。

1.2 本研究の目的

本研究では、単一画面による遠隔講義スタイルを前提に、受講者の講義映像に対する視認性の向上と単一画面による集中やプレゼンス性の高さを両立させることを目的とする。撮影された映像中の講義者自身の画像領域と資料の領域を抽出した後、受講者側では

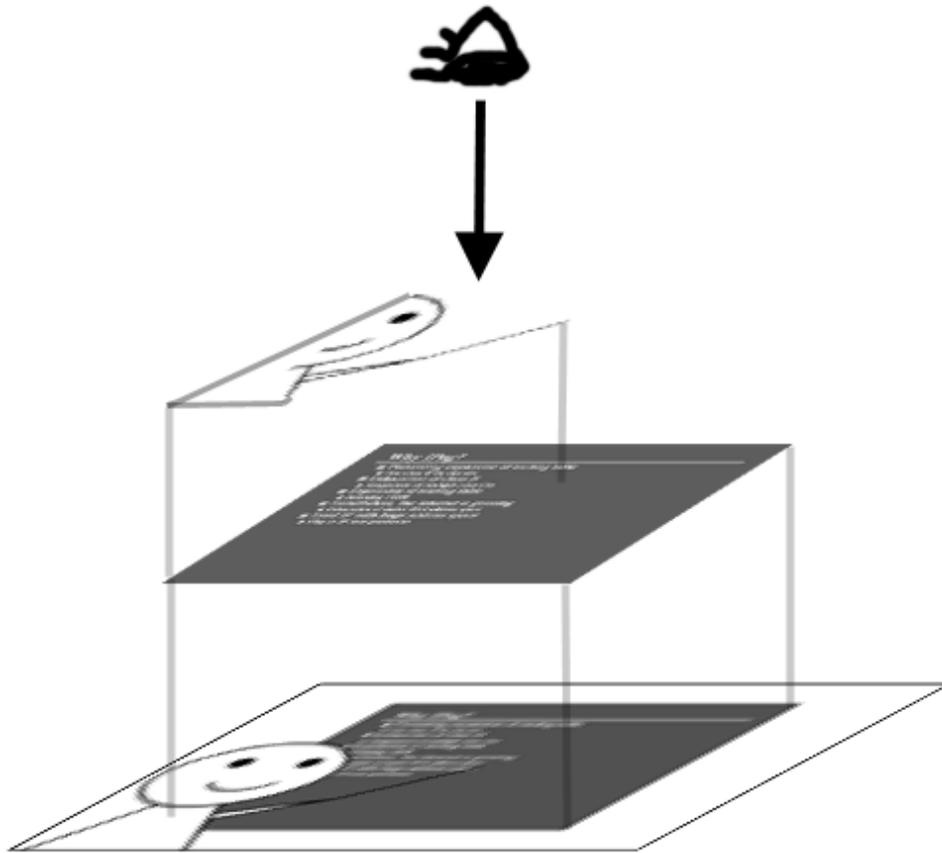


図 1.1: はめ込み合成のイメージ図

資料画像は高品位な画像に、講義者の領域は半透明に加工し、それらの画像を再びはめ込み合成する(図 1.1).

講義映像からの資料や講師の領域抽出を行い、再加工を施し受講者に提示するこの形態は、講義資料や講師の存在など基本的な要素に対して構成変更を行うことなく、受講者の前に提示することを可能とし、プレゼンス性の確保にもつながる。

これにより、事前共有可能な情動的な情報源である講義資料をより詳細に受講者に示す事を可能にすると共に、テレプレゼンスをもたらす動的な情報源である講義映像も単一画面で提供することができる。

また、本研究は提案に検証、実装も含め、実際の遠隔講義に使用できるシステムの構築を目指す。その上で実際の使用に耐えうるか検証することも含めこれを目的に定める。

1.3 本論分の構成

本論分の以降の構成を以下に示す。

- 2章: 既存の遠隔講義システムの例を挙げ、資料提示手法の現状を示すと共に、遠隔講義形態の分類を行い、本研究の位置を明確にする。
- 3章: 本提案を実現するための領域抽出手法を検討し、決定する。
- 4章: システムの設計/実装を行うにあたっての問題点を明らかにし、解決策を検討する。
- 5章: 3章において決定した実現手法である、輝度による領域抽出手法について詳しく述べる。
- 6章: 輝度による領域抽出の適応範囲を探るために、複数の観点から本システムを評価/考察する。
- 7章: 本研究のまとめと、今後の課題について述べる。

第2章 遠隔講義システム

この章では、遠隔講義システムについての既存の例を挙げる。また、それらのシステムを分類することにより、本提案システムの位置づけや適用範囲を明確にする。

2.1 既存のシステム

遠隔講義は既に多方面で実践的な試みがなされており、システムとしても既に多数のものが存在する。以下では本研究において重要となる資料提示の面を中心に既存のシステムの例を挙げる。

2.1.1 WIDE University School of Internet

WIDE プロジェクトの School on the Internet (SOI) workgroup[2] が提供するインターネット基盤を利用した研究活動。

大学における教育資源をデジタル化し、インターネット上で受講者に対して公開する。質疑応答や、課題提出もインターネット上で行われ、インターネット基盤上で教育資源を共有する実証実験が行われている。

また、後述する RPT などのツールの提供や、本学も参加する SOI-Asia による衛星を利用したインターネットによるアジア諸国への高等教育など幅広く活動している。

SOI の講義映像は、インタラクティブコンテンツを提供するためのマークアップ言語 SMIL(Synchronized Multimedia Integration Language) によって記述されており、講義映像と同期して資料を提示することができるよう設計されている(図 2.1)。

2.1.2 Web Collaboration

IBM の T.J ワトソン研究所と東京基礎研究所が開発した Web ブラウザを同期させてそれに注釈付けや指し示しのための機能を提供する技術 [3]。

同期型遠隔教育システムで、インターネットを通し資料として用いられる HTML ページを多地点で同期して表示することを可能にする。

- ハイパーリンクを使ったページジャンプ



図 2.1: SOIの講義映像

- フォーム入力
- ウィンドウスクロール

などの，web ページを閲覧する際の基本的なアクションの同期を備えるほか，

- 注釈付け
- マウスポインタの同期

等の特別なアクションも共有することができる．

2.1.3 TIDE Project

TIDE(Trans-pacific Interactive Distance Education) Project は，京都大学と UCLA 間で NTT GEMNet 海外実験用ネットワークを介して相互に遠隔講義を行うプロジェクト．TIDE Project では，資料提示に 2 つのツールを用いている．

- IMED

UCLA CDI が開発した WWW ベースの教育用システム．資料提示のほかに資料作成ツールや講義評価ツール，オンラインでのオフィスアワーを実現する機能を持っている [4]

- Meeting Perfe

NTT-AT 社製のグループウェア。資料提示として、HTML ファイルと PowerPoint を資料として用いることができ、資料提示を同期させられるほか、マウス操作やウィンドウのスクロール、ポインティング等の操作も同期させることができる。

2.2 遠隔講義の分類

同期/非同期

遠隔講義はニーズに合わせて様々な形態をとっている。まず大きな分類として、時間の観点で分類すると2つに大別できる。

- リアルタイム遠隔講義
- アーカイブ遠隔講義

前者は、離れた地点にいる講師と受講者がネットワークを通してリアルタイムで主に映像と音声を送信して講義を行う同期型の講義形態である。利点として、意見交換、質疑応答のしやすさがあるが、時間が束縛される、双方向で受講者間を接続した場合の設備が高くなるなどの問題がある。

後者は、講師が行った遠隔映像や資料、講義録を保存し、受講者は好きなときにそれにアクセスして受講する、非同期型の講義形態である。利点として、好きな時間に受講できることや、繰り返し受講可能、場所の制限が無いことなどがあるが、質問や意見交換、論議がしにくい、学習のペースが掴み難いなどの問題がある。

資料提示

大半の遠隔講義では講義の際、資料提示が行われる。資料提示の観点では、2つに大別することができる。

- 講義映像とに含めて提示する
- 講義映像に含めず、別に提示する

前者は、講義映像と講義資料が単一画面に取り込まれ配信される場合のほか、黒板などの講義資料を特別に記録するカメラなど複数のカメラを用い視認性を上げるなど様々な手法があるが、映像として資料を撮影し、伝送する方法をこの分類とする。

この場合、単一画面であれば、一つの画面に集中することが出来る。また、講師が直接資料を指し示した場合にその場所が明確に判別することができる。しかし、映像中に含まれる資料の領域が比較的小さくなり、文字や図形の視認を妨げる恐れがある。また、複数

画面で伝送する場合は、視認性の問題は単一画面に比べ改善されるが、基本的に視認性は伝送媒体やコーデックなどによる制限を受けるため、視認性を高めるには高品位の映像を伝送する必要がある。

後者は、講師の映像を伝送し、資料は別の手段で受講者に提示するものの分類とする。この場合は、講義資料があらかじめ印刷されて受講者に配布される場合や、2.1.1 節 SOI の例のように資料と講義映像を 2 分割し同時に伝送する場合などがある。

この場合、資料の視認性は前者に比べ比較的高い。しかし、講師が直接資料を指し示す場合、別提示である講義資料を直接指し示すことが出来ない。この問題を改善するには、何らかの手段で講師の指し示す場所を、別に提示する資料の場所にマップする必要がある。

また、受講者は講義資料と講義映像の両方に目を遣らねばならず、集中の低下やプレゼンス性の低下などの現象をもたらす。

映像伝送

資料提示の観点からも、遠隔講義映像の品質が問題となる。遠隔講義において主に使われる映像伝送機構には以下のものがある。

- DVTS
- Real Video
- vic/vat
- Windows Media Video

DVTS は WIDE Project が開発した DV 伝送システムである。DV のデータは DVTS によって IP パケットにラッピングされ、インターネットを通して遠隔地に伝送することができる。DVTS で DV データを伝送するには概ね 30Mbps の帯域を必要とする。

Real Video は Real Networks が開発したビデオプレイヤーで Real Video 形式のビデオを再生出来るほか、リアルタイムで映像を伝送することが出来る。また、インタラクティブコンテンツを提供するためのマークアップ言語 SMIL にも対応しており、単にビデオ映像を提供することにとどまらず、様々な利用することが出来る。

vic/vat はマルチキャストバックボーンである MBONE の研究の中で開発されたツールで、vic はマルチキャストで映像の配信/受信を可能にするツールで、H.261 ビデオコーデックでの圧縮をサポートしている。ユニキャストも使用することが出来る。

vat は同様にマルチキャストで音声の配信/受信を可能にするツールである。

Windows Media Video は Microsoft が開発したビデオエンコーダ/デコーダで Windows に標準で添付されている、Microsoft Media Player で再生することが出来、Windows のシェアから最近普及している。ライブで配信することもオンデマンドで配信することも可能である。

他に、DV over ATM で伝送する方法、ISDN 網を用いる方法などがある。

2.2.1 本研究での対象

本研究での対象として、同期すなわちリアルタイムで遠隔講義が行われるタイプの遠隔講義を対象とする。その上で、資料提示の観点本研究の目的である、視認性とプレゼンス性の両立を図るべく、本提案手法により両者のメリットを生かす。すなわち、単一画面での講義映像の伝送により、プレゼンス性の向上を図り、その上で提示資料をあらかじめ講師/受講者間で保持しておき、受講者側すなわち受信側で資料のはめ込み合成を行い、表示する。

送受信媒体は特に問わないものとするが、本研究の性質上低帯域の伝送媒体でも使用可能なものでなくてはならない点を考慮する。

また、本研究においては、講師から受講者側への送信にのみ注目し、逆向きの送信や受講者からのフィードバック等は研究の対象外とする。

本研究では、資料提示の媒体として Microsoft PowerPoint を対象とする。PowerPoint は、現在プレゼンテーション及び講義において広く利用されているソフトウェアであり、昨年度の商用プレゼンテーションソフト販売の 99.3% のシェアを持つソフトウェアである [5]。

第3章 本研究の概要

この章では、本研究の提案について述べ、その長所を明らかにする。また、提案の実現法について論議する。

3.1 本研究の提案

本研究の提案は、単一画面での映像提供を行い、提示資料はその画面にはめ込み合成で合成を行い、その上で講師の領域を半透明に合成する方式である。

本提案方式では、受信側で生成した資料を用いる。これにより高精細の資料提示が実現でき、視認性の面で有効である。また、映像提供の観点では、資料提示は別の手段で行われることになり、資料の視認性を無視することが出来るようになる。これにより高精細な映像提供を必要としなくなり、伝送する映像の帯域を低く抑えることが可能である。これは、帯域の面の他、コストの面で有効である。

3.2 実現手法の調査

本提案手法において、はめ込み合成を行うための合成位置、すなわちスクリーンの位置の取得は重要な要素である。

スクリーンの位置の取得をするにあたっての条件として、送信側、すなわち講師側でも同時に対面講義が行われていることを想定した。この条件付けは、スクリーンの位置の取得には講師側においてはスクリーンを良好に視認することが出来なくてはならないことを意味する。従って、位置取得に際してスクリーンに何らかの機能を追加する場合においても、その追加は最小限度に留めなくてはならない。

以上のことを踏まえ、スクリーンの取得手法について調査検討を行った。

3.2.1 赤外線を用いる手法

この節の手法は、赤外線の照射装置を用いてスクリーンを照らし、赤外カメラを用いて赤外領域の輝度情報を取得し、その情報を用いてスクリーンの位置を検出する試みである。

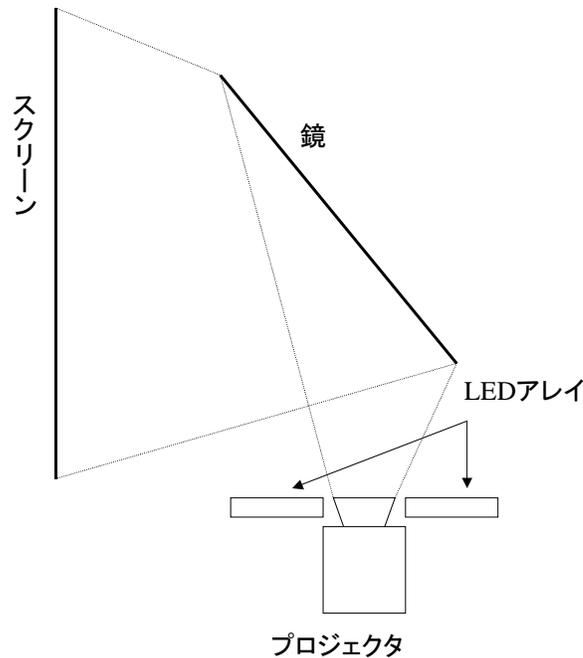


図 3.1: 赤外線 LED アレイの配置図

スクリーン全体を背後から照射する手法

スクリーンに対し、背後から全体を赤外線で照射し、赤外線カメラを用いて位置の検出を試みた。

この場合、機材として赤外線照射装置と、一般のカメラに加え、赤外線カメラが必要となる。また、エッジを検出する際良好に検出するためには、スクリーンのみを照らす必要がある。そのため、この構成を採用する場合、対象は背面投射型のみプロジェクタに限定される。

この試みは、赤外 LED アレイ (500mW) を 10 個背面投射型プロジェクタの投影部近くに設置し (図 3.1)、実際に輝度を測定した。

結果、実際の使用を模し、自然光及び蛍光灯の光る状況で撮影した画像 (図 3.2) では、スクリーン以外からの赤外線が入り込み、スクリーン内の最低輝度をスクリーン外の最高輝度が上回り、この手法単独でのスクリーンの取得は困難だという結論に至った。

この手法での取得を試みる場合、赤外線に発光パターンを持たせスクリーンとの他の領域との識別を明確にするなどの手法の併用や、赤外領域でのスクリーンの輝度を上げるために更なる赤外線 LED の設置、別の赤外線照明の使用を検討するなどの対策が必要である。

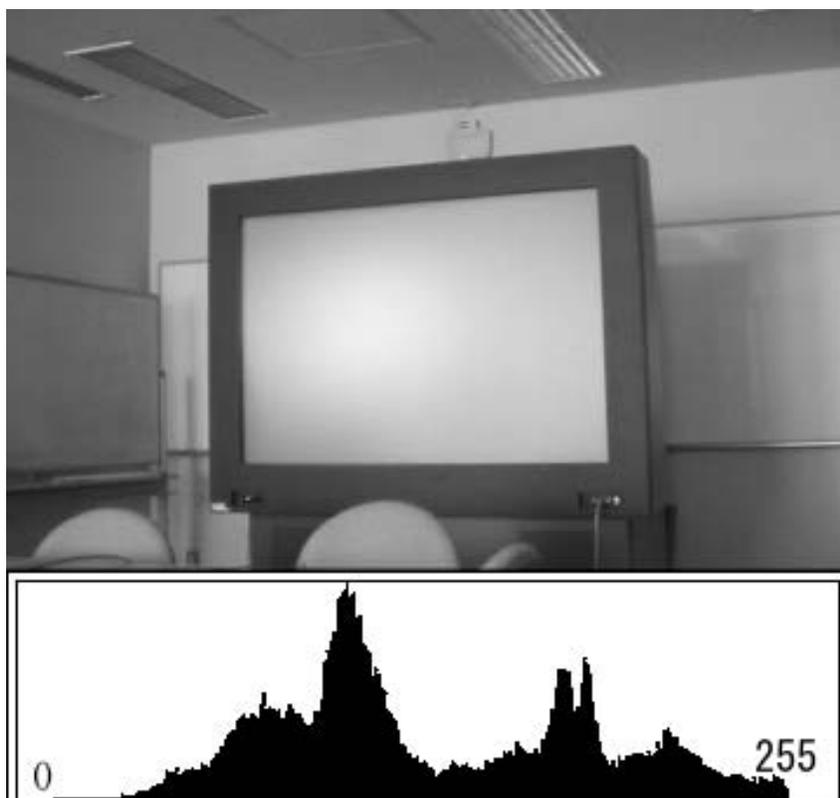


図 3.2: 撮影画像及びヒストグラム

スクリーンの一部を照射する手法

スクリーン全体を赤外線で照射し、それを取得する手法は赤外領域の輝度の面で問題があった。そこで、全体を照らさず、スクリーンの直近から一部のみを照らすことを試した。

スクリーンの裏面に直接赤外 LED を張り、前面から撮影し、輝度情報を調べた。可視光領域を撮影する一般のカメラは赤外線を遮断するために、ローパスフィルタが内蔵されている。しかし、一部の赤外光が漏れて検出されることがある。この試みでも一般のカメラでの検出が確認された。しかし、微量であり、これを用いることは難しいと考えられる。

赤外線カメラでは、良好に検出することができた(図 3.3)。この撮影映像を閾値 220/256 で二値化した画像が図 3.4 である。

この画像から、スクリーン直近に赤外光源を設置した場合、良好に位置情報を取得できることが確認できた。この手法は、背面投射型プロジェクタ以外にも適用することが出来る。欠点としては、図 3.3 から分かる通り、スクリーン直近に光源を配置しているため、スクリーンに影をつくることとなる。この問題避けるには、スクリーンの投影面以外に光源を配置する必要がある。

3.2.2 可視光を用いる手法

この節の手法は，スクリーンのに投影される画像自体の情報を用いて，位置情報を取得する試みである．

スクリーンの輝度情報を用いる手法

スクリーン自体が発する光を取得し，その輝度情報を元にスクリーンの位置を検出できるか試みた．

この場合，特に特別な機材を必要としない．また，赤外線を用いた手法の場合，送信側で検出を行い，領域情報を送信しなければならないため，送信側及び受信側で2台のコンピュータを必要とするがこの手法の場合は，受信側で閉じているため使用するコンピュータは1台で済む．しかし，スクリーンに映される画像の種類，特徴により輝度情報が変化する恐れがあり，その点を考慮しなければならない．

スクリーンの輝度をRGB要素ごとに $b_r(x, y)$, $b_g(x, y)$, $b_b(x, y)$ とし，閾値を D とすると二値情報 $f(x, y)$ は以下の式で求めることができる．

$$d = b_r(x, y) + b_g(x, y) + b_b(x, y)$$
$$f(x, y) = \begin{cases} 0 & (d < D) \\ 1 & \textit{otherwise} \end{cases}$$

実際にスクリーンに資料を模した白い矩形を投影した状態で，カメラでスクリーンを撮影し，輝度情報を閾値 $66/256$ で二値化した画像が図 3.5 である．

この図から，投射した資料部分の輝度はスクリーンの周辺より輝度が高くなっており，資料の領域が矩形に浮き出ていることが分かる．しかし，スクリーンのみが浮き出ているのではなく，背景の照明に照らされている部分も浮き出ているのが分かる．投影した資料の背景にある程度の輝度があれば，スクリーンの周囲より輝度は増す．この増分に注目すれば，スクリーン自体の検出は可能である．しかし，スクリーン以外の領域の輝度が蛍光灯などの照明の影響でそれより高いことが十分考えられるため，スクリーンの発光かそれ以外の発光かを選別する手段が必要となる．

スクリーンの色相情報を用いる手法

資料の背景色の色相情報に注目して，スクリーン領域かその他の領域かの選別ができるか試した．

資料の背景色を青に設定してスクリーンに投影し，それを撮影したものを用意した．各RGB要素ごとのピクセルの輝度情報 RGB と色相 H の関係は式 3.1 で表すことができる．

$$r = \frac{R}{R + G + B}$$

$$\begin{aligned}
g &= \frac{G}{R + G + B} \\
b &= \frac{B}{R + G + B} \\
H &= \tan^{-1} \left(\frac{g - b}{2r - g - b} \right)
\end{aligned} \tag{3.1}$$

映像の色相 $H(x, y)$ を導き出し，元の背景色の色相 H_b からの距離 d を算出し，閾値 D で二値化し二値情報 $f(x, y)$ を算出した (式 3.2).

$$\begin{aligned}
d &= |H(x, y) - H_b| \\
f(x, y) &= \begin{cases} 1 & (d < D) \\ 0 & \text{otherwise} \end{cases}
\end{aligned} \tag{3.2}$$

スクリーンの背景色を青に設定し，上記式を適用し，閾値 128/256 で二値化した例を図 3.6 に示す。

図 3.6 は，ピクセルが暗いほど距離が近いことを表している．スクリーンは一様に暗くなっており，検出可能なことが分かる．しかし，黒に近い色 (図ではスクリーンのフレーム付近) では，色相の値が安定せず，検出の誤りが起こっているのが分かる．従って，暗い部分マスクを行わねばならない．

この場合も，輝度を用いる手法と同様に特別な機材は必要としないメリットがある．しかし，事前にスクリーンの背景色を設定しなければことや，グラデーションのかかった背景など，色相が一樣でないものの対処などの問題が挙げられ，それに対する対処策が必要である．

色情報から距離を求め，用いる手法

背景色の色情報に注目し，映像の各ピクセルと背景色との距離をとり，それからスクリーン領域の選別ができるか試した．距離 d を RGB 要素ごとの映像の色情報 $f_r(x, y)$ $f_g(x, y)$ $f_b(x, y)$ ，資料の背景色を $C = (r, g, b)$ から算出し，閾値 D で二値化し二値情報 $g(x, y)$ を算出した (式 3.3)．

$$\begin{aligned}
d &= \sqrt{\{f_r(x, y) - r\}^2 + \{f_g(x, y) - g\}^2 + \{f_b(x, y) - b\}^2} \\
g(x, y) &= \begin{cases} 1 & (d < D) \\ 0 & \text{otherwise} \end{cases}
\end{aligned} \tag{3.3}$$

資料の背景色を青に設定してスクリーンに投影し，それを撮影したものを用意した．その画像かに上記式を適用し距離 d を求め，閾値 70/256 で二値化したものが図 3.7 である．

この例では，良好にスクリーンとその他の領域を分離することが出来た．この手法は，色相情報を用いる手法と同様に，機材の面などで有効だが，一樣でない背景の対処などへの検討が必要である．

3.2.3 点光源を用いる手法

スクリーンの四隅に点光源を配置し，その光源を元に位置取得できるか試みた．

機材として点光源を4個必要とする．またこの場合も可視光の輝度を利用する場合と同様に，受信側で閉じているため，使用するコンピュータは1台で済む．

この場合，点光源の取得が問題である．カメラでの撮影の際，点光源以外の場所の輝度情報が最高値に達してしまう場合があり，単純な輝度情報での選別は不可能であると思われる．従って，点光源を定期的に点滅させ，その点滅パターンを取得する手法が考えられる．点滅パターンの点灯と消灯の2つの画像をそれぞれ二値化し，その2枚の画像の排他的論理和を取ると，輝度に変化のある部分すなわち2枚の二値画像で違いのある部分のみを取得することができる．一回の処理ではノイズが含まれるが，処理結果を重ね合わせることで点光源を正確に取得することが可能であると考えられる．

この手法では，コンピュータ1台での取得を前提としているため，点滅を行う際，送信側と受信側で同期を取ることが出来ない．したがって，定期的な点滅を受信側で認識する必要がある．

3.2.4 スクリーン位置の取得

点光源を用いる場合を除いて，これらの手法は，スクリーンとそれ以外の領域を二値化して示すだけに留まっている．従って，その情報からスクリーンの四隅の座標を求めなければならない．その手法について検討した．

snake モデル

snake モデルは，kass らによって提案された手法で，画像から直接輪郭を取得する手法である．snake は輪郭線をモデル化した関数を定義して，目的の画像に対してフィッティングすることによって抽出を行う．モデル化した関数のエネルギー項画像上でその輪郭線を動的に移動させ目的の領域を抽出することから動的輪郭モデルと呼ばれる．

snake は滑らかな輪郭の抽出には適しているが，鋭角物体のエッジへのフィッティングが難しいことや，エネルギーに関するパラメタの選定が難しいなどの問題がある．

Hough 変換による直線検出

Hough 変換 [6] を用いると，画像中から特徴点から，それらの特徴点を通る直線を求めることができる．それらの直線の交点を求めることでスクリーンのエッジが求められる．この手法については，別項にて詳しく述べる．

3.3 手法の検討

赤外線を用いる手法の場合、赤外線カメラと可視光のカメラ、2つのカメラの同期が問題となる。2つのカメラが全く同じ映像を撮影していない限り、この手法は実現できない。2つのカメラで同じ映像を撮影するには、カメラ毎に異なる位置を補正する必要があることになる。この視差の補正を行えば、カメラの位置のずれはある程度許容される。しかし、ズームを行う場合は、何らかの手段でズームが行われたことを検出し、2つのカメラの拡大値を同期する必要がある、これは困難が予想される。輝度及び、点光源を用いる手法は、このような補正を行う必要が無く、機器の構成も比較的シンプルになるため有利である。

点光源を用いる手法は、点光源をスクリーンに配置するという物理的な作業を行う必要がある。輝度を用いる手法は、このような制約が無く構成もシンプルであり、輝度を用いる手法で実際の領域検出が行えるのであれば、この手法が適していると考えられる。

また、輝度情報から実際のスクリーン四隅の座標への変換は Hough 変換を用いることとした。これは、snake モデルの動作が低速であることと、エッジのあるオブジェクトの抽出には比較的不向きであることが理由である。

以下の章では、この輝度による抽出手法を採用することとし、システムを設計/実装し、輝度情報を用いた領域抽出が実際に可能であることを明らかにする。



図 3.3: 光源をスクリーン直近に配置した例



図 3.4: 図 3.3 を二値化したもの

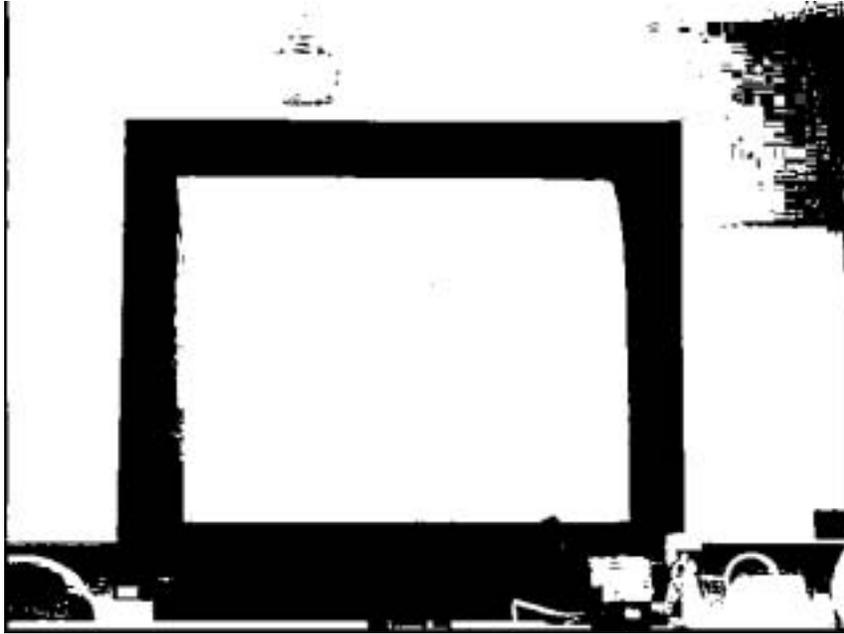


図 3.5: 可視光の輝度情報を二値化した例



図 3.6: 色相の差をあらわした例

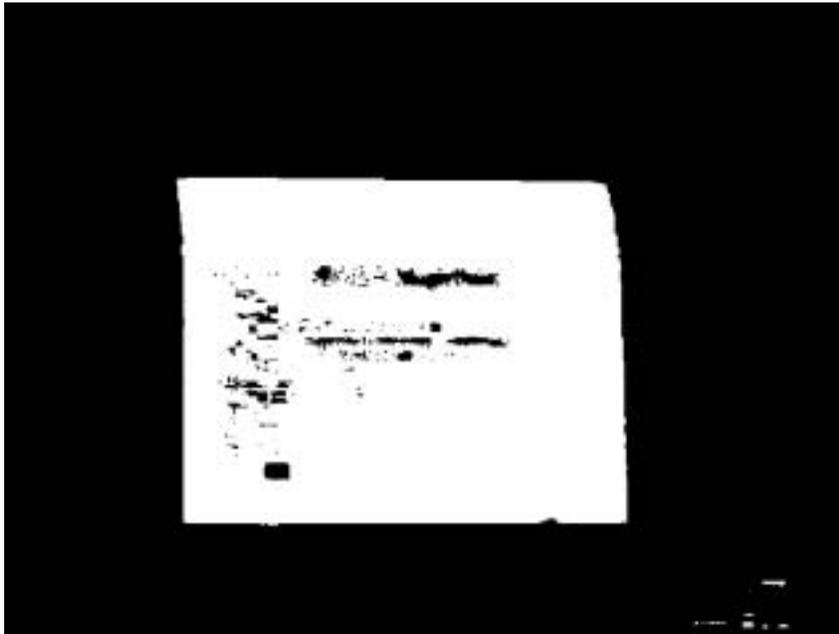


図 3.7: ピクセルの色情報の距離を二値化した例

第4章 システムの設計/実装

本章では、今まで述べてきた提案及び手法を実現するべく、設計/実装を行いその詳細について述べる。また、設計にあたり問題となった点について、解決策について述べる。

4.1 システムの設計

本提案に基づくシステムの設計にあたり、用いる手法は3.3節で述べた輝度を用いる手法とした。これに基づきシステムの構成を定めた(図4.1)。

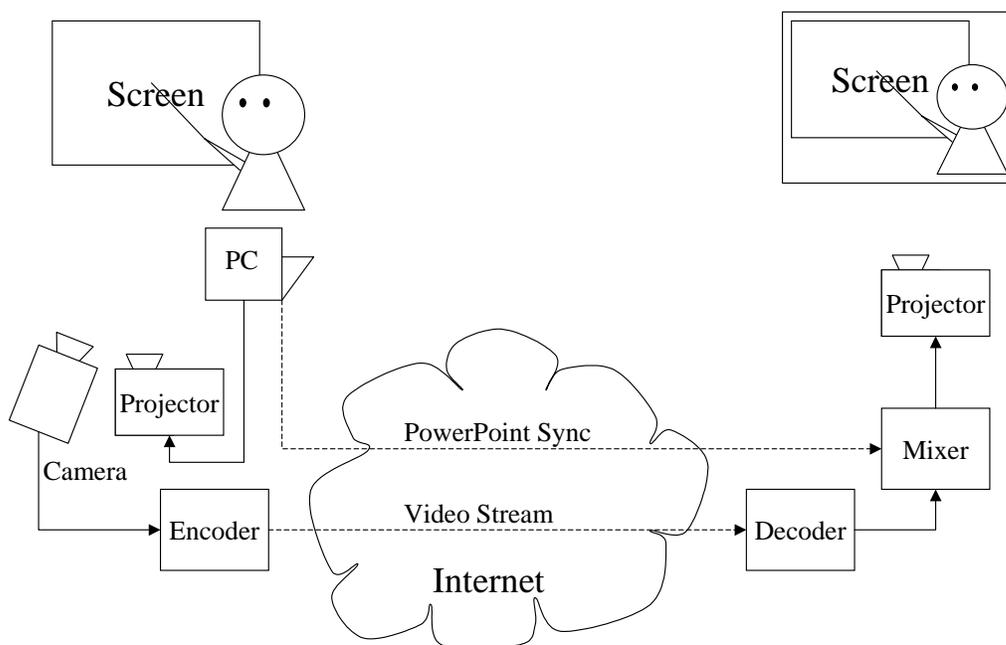


図 4.1: 全体の構成

図 4.1 中央に示すインターネットを挟んで左側を講師側の機器構成，右側を受講者側の

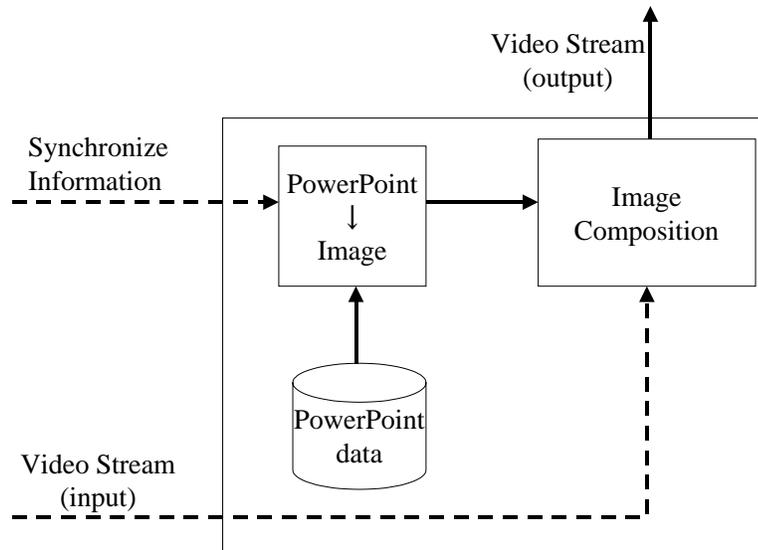


図 4.2: Mixer 部の内部構造

機器構成を示すものである。本提案システムにおいて実装される箇所は、受講者側の構成要素である、Mixer 部である。

講師側の PC と示してある機器は PowerPoint が動作しているパーソナルコンピュータであり、プロジェクタに画面を出力し、資料の提示を行っている。この PC はネットワークを通して、同期情報を出力する。また、講師側のカメラは講義映像を撮影し、エンコーダを通して圧縮処理を行い、ネットワークを通して受講者側へ送信される。

受講者側では、講義映像がデコーダで伸張処理され、mixer にデコードされた講義映像が入力される。

Mixer 部の内部構造を図 4.2 に示す。Mixer 部は、内部で PowerPoint 形式の資料データを保持している。このデータは、講師側の PC 内に保持されている資料データと同一のものでなければならないため、事前に何らかの手段で受講者側に伝送し、Mixer 内に保持させる必要がある。同時に Mixer 内部では PowerPoint が動作し、レンダリングを行い資料データをイメージに変換する。

作成した資料イメージは入力される講義映像と合成される。同時に講師の領域を抽出し半透明処理を施し、合成され出力される。実際の出力は Mixer が動作しているコンピュータの画面からの出力となる。

4.1.1 環境

本提案システムの Mixer 部は，Microsoft Windows で動作する．これは，Microsoft PowerPoint が同時に動作する構成から必至であると考えられる．他の OS において資料の変換を行うとなると，データの変換を行わなければならない，互換性の面から問題となることが考えられる．

また，映像の取り込みは，Windows で動作する DirectX コンポーネントの機能の一部である，DirectShow を利用した．これは，多種のソースからの映像の入力をサポートしており，それらを統一されたインターフェイスとして，平易に扱うことが出来る．また，対応するデバイスの種類も多いという利点もある．

DirectShow の扱うことができるデバイス/ソースは具体的に以下の通りである．

- DV (IEEE1394)
- ビデオキャプチャカード
 - － Composite 入力
 - － S 端子入力
- ファイル
- ストリーム

本システムでは，DV 及びビデオキャプチャカードの Composite 及び，S 端子からの入力を想定している．これは，DV 形式は，DVTS や DV over ATM 等を間接的に用いれば，ネットワークとの親和性が高いためである．Composite 及び S 端子からの入力は現在広く用いられているソースであることや，PC に付くビデオカードの出力として，Composite 端子を用いているものがあり，それを Decoder とすることで，多種のフォーマットを扱うことが可能となる点で有利であると判断し採用した．

また，2.2.1 節で述べたとおり，資料提示には，Microsoft PowerPoint を用いる．

4.2 実装に伴う諸問題と解決法

上記設計を実現するにあたり，主に問題となった個所に対する解決策を述べる．

4.2.1 PowerPoint の同期

PowerPoint を同期して動作させるには，下記の処理をする必要がある．

- 同期シグナルの送信，受信

- PowerPoint のページ更新の検知
- 受信したシグナルに対する PowerPoint の反映

最後の要素，PowerPoint の反映を実現する手法として，

- PowerPoint 自体を制御する
- PowerPoint のデータを別形式にして，それを制御する

などの手法が考えられる．PowerPoint 自体を制御する場合，それを実現するには Microsoft の提供するインターフェイス，具体的には ActiveX コントロールを操作すれば，この手法での操作は可能となる．しかし，ActiveX コントロールを用いる場合，MFC(Microsoft Foundation Class) を用いる必要があり，これは DirectX との親和性が低い．

後者の手法を用いる場合，別形式自体にすることによる再現性が問題となる．PowerPoint には，アニメーションなどの機能が内包されており，それらを忠実に再現することは，今現在出回っている形式では難しい．

したがって，PowerPoint を用いつつ，ActiveX コントロールを用いずに PowerPoint 自体を操作する手法が必要となる．

本手法では，これらの問題を解決するために，RPT を用いることとした(図 4.3)．RPT は SOI の開発した PowerPoint の同期を取るためのツールである．RPT は起動すると，RPT のコントロール下にある PowerPoint を起動する．起動後の操作は全て RPT が監視し，アクションの情報を IRC プロトコルを用いてネットワークに送出する．受信側では同様に RPT が IRC からアクションの情報を取得し，RPT の管理下にある PowerPoint へアクションを伝える．これにより PowerPoint の同期を実現する．

4.2.2 PowerPoint からのイメージの取得

PowerPoint から，イメージを取得する最も簡潔な手法は，画面データを直接 VRAM からメモリにコピーすることである．この手法であれば，アニメーションやムービーの再生など，PowerPoint の複雑な動作があっても様に VRAM からイメージを取得することができる．プレゼンテーション画面を表示しておき，そのデータ逐次コピーするこの手法は，簡潔ではあるが下記のような問題を含んでいる．

- VRAM アクセスのウェイト
- Windows の制約からくる画面の問題

前者は，コピーする画面は VRAM 上にある問題からくるものである．現在の標準的な PC において VRAM はシステムメモリ上には存在せず，AGP バスや PCI バスを通してア

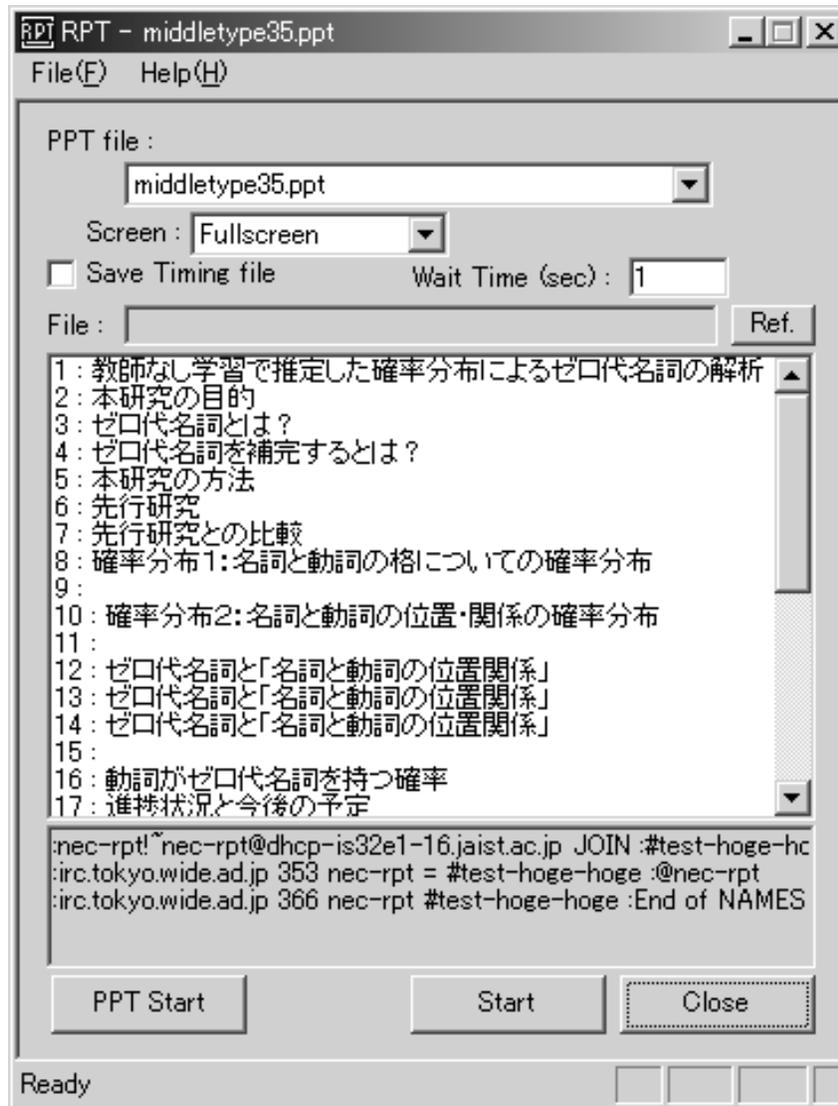


図 4.3: RPT

クセスされる。この時、バスのバンド幅がボトルネックとなり、アクセス時間がシステムメモリに比べ多くかかる。

表 4.1 は、テストマシン (DELL Dimension 4500 CPU 1.7GHz) でのメモリ間転送帯域実験の結果の表である。System System 間転送の実効値の計測は memcpy を使って行っている。また理論値は搭載しているメモリ (DDR SDRAM) の理論値である。System VRAM 間転送の実効値の計測は Win32 GDI インターフェイスの BitBlt コールを実際に発行し、それにかかった時間から逆算したものである。また、理論値は AGP バス (AGP 4x モード) のバンド幅である。

転送		実効値	理論値
System	System	450Mbytes/sec	2.1Gbytes/sec
System	VRAM	170Mbytes/sec	1.0Gbytes/sec

表 4.1: メモリ間での転送帯域

これらの結果から，System VRAM 間転送は，System System 転送に比べ，約 2.5 倍のアクセス時間を要することがわかる．画面のデータは VRAM 上以外には存在しておらず，かならず VRAM にアクセスすることとなる．PowerPoint の表示領域を，1024x768 ピクセルと仮定すると，VRAM 上のデータをシステムメモリ上に一回転送するのに要する時間は，18.5msec で 1 秒間に 30 回更新すると仮定すると，255msec を要し，プロセス全体の 1/4 の時間を必要とする．

したがって，VRAM 上のデータに対しアクセスする方式は，処理時間の面から好ましくない．

また，後者の問題 Windows の制約からくる画面の問題とは，Windows は現在表示されているアプリケーションの画面情報しか取得できないことからくる制約に起因する問題である．PowerPoint の画面情報を取得するには，PowerPoint が動作していて，かつ，画面に表示されている状態でなければ，画面情報を取得できない．しかし，画面は本システムの出力に使われており，画面にシステムの出力以外のウィンドウを配置することは出来ない．したがって，PowerPoint の画面を取得することが出来ないことになる．

仮想画面出力ドライバ

これらの問題を解決するために，本システムでは画面を 2 個設け，一方に本システムの出力を，もう一方に PowerPoint の画面を出力する．しかし，物理的に 2 枚のビデオカードを本システムが動作する PC に挿すことは，使用する機器がふえることとなり好ましくない．また，この場合でも，VRAM アクセスが発生することには変わりはない．それらの問題から，機器の増設を必要とせず，また，出力をシステムメモリに行う機構の検討を行った結果，仮想の画面を生成し，そこに出力を行うことの出来るドライバの実装を行うこととした．

ドライバの概要

このドライバ (以降 `vmidrv` と呼ぶ) は，Windows のシステムに対し，実体の無いビデオカードをあたかも存在しているように見せる働きをする．Windows システムから発行されるメモリ領域の取得命令に対し，本来であれば VRAM を取得しなければならないが，代わりにシステムメモリを取得し，システムにその位置を通知する．

本来ビデオカードが持つグラフィックスアクセラレーション機能全てを，全てのビデオ

カードが備えているわけではない．一部の機能のみアクセラレーションを用いる場合は，その他の描画を全て Windows システムが代わりに代行する．すなわち，グラフィックスドライバとシステム間でのメッセージ交換に用いられるインターフェイスである DDI に対し全てのコールバック関数を実装する必要がない．

vmidrv では，描画命令に対するコールバック関数は一切備えておらず，描画全てに対し Windows システムが代行して描画を行う構造になっている．

vmidrv をインストールし，Windows の画面のプロパティで確認すると，2 台のモニタが接続されていることが確認できる (図 4.4) ．



図 4.4: 画面のプロパティ

図 4.4 の 2 番のモニタは，vmidrv が生成した画面で，Windows のアプリケーションは，この画面に自由に移動することが出来る．また，Microsoft PowerPoint は，2 つの画面を検出した場合，1 番の画面 (プライマリモニタ) に，ノートや操作画面を，2 番の画面 (セカンダリモニタ) にプレゼンテーションの画面を表示するようになっている．すなわち，PowerPoint を用いる場合，特別な操作なしに，vmidrv の生成した画面にプレゼンテーション画面を表示することが出来る．

また，生成した画面へのアクセスすなわちドライバとユーザーモードプロセスとの通信は，Windows GDI(Graphics Driver Interface) の持つ API の一つで，GDI 経由ではアクセスできない特定のデバイスの機能にアクセスするためのインターフェイスを用いる．これにより，カーネルモードにある画面バッファのアドレスをユーザーモードにマップしたアドレスをユーザーモードのプロセスに渡す．

PowerPoint からイメージを取得するプロセス Grabber までの処理の流れを図 4.5 に示し、詳しく解説する。

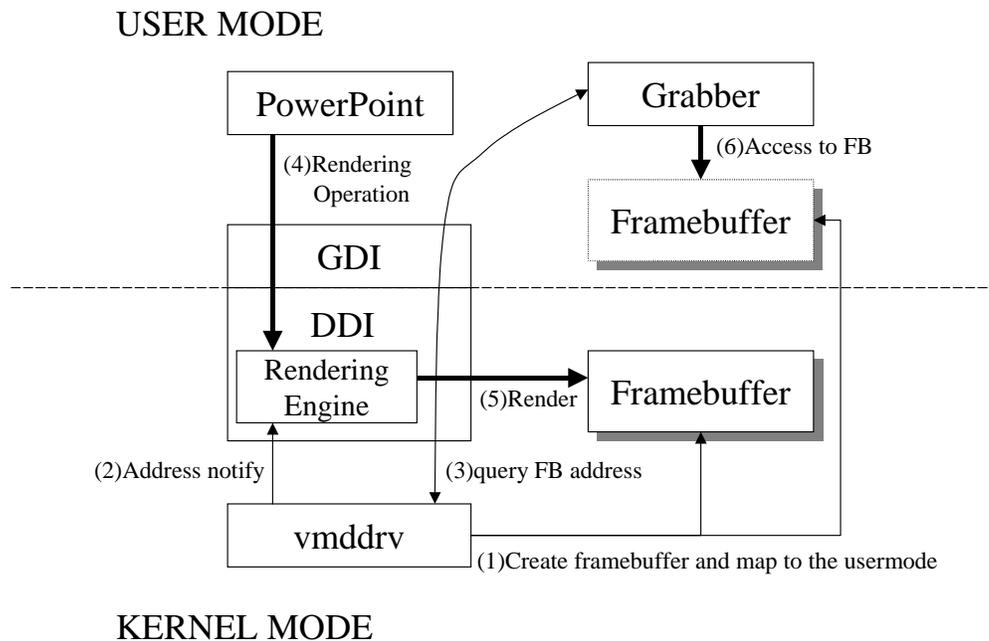


図 4.5: 処理の流れ

- (1) vmidrv がインストールされると、カーネル空間にメモリアドレスを確保し、フレームバッファとする。ユーザー空間にフレームバッファの写像を作る。
- (2) vmidrv は、フレームバッファの存在をレンダリングエンジンに通知する。これ以降レンダリングエンジンはフレームバッファに対し、描画処理を行う。
- (3) Grabber はマップされたフレームバッファのアドレスを知るために、GDI を通して vmidrv にクエリーを発行してアドレスを得る。
- (4) PowerPoint は資料の描画を行うために、GDI インターフェイスを通して描画命令を発行する。
- (5) レンダリングエンジンは、命令を受けフレームバッファに対して描画処理を行う。
- (6) Grabber は、(3) で得たアドレスに対してアクセスすることで、描画されたイメージを得る。

以上の処理の流れにより、アプリケーション”Grabber”は PowerPoint によって描画された資料イメージを得ることが可能になる。

第5章 映像からの領域抽出

この章では、映像中からのスクリーンの領域及び、講師の領域の抽出手法について述べる。

5.1 スクリーンの領域抽出

映像中のスクリーンの領域に、生成した PowerPoint のイメージを貼り付けるために、スクリーンの四隅を検出する必要がある。本手法において、映像から四隅の位置情報を得るまでの一連の流れを以下の節で詳しく述べる。

また、以下で処理途中の映像の例を挙げるが、その初期状態は図 5.1 である。

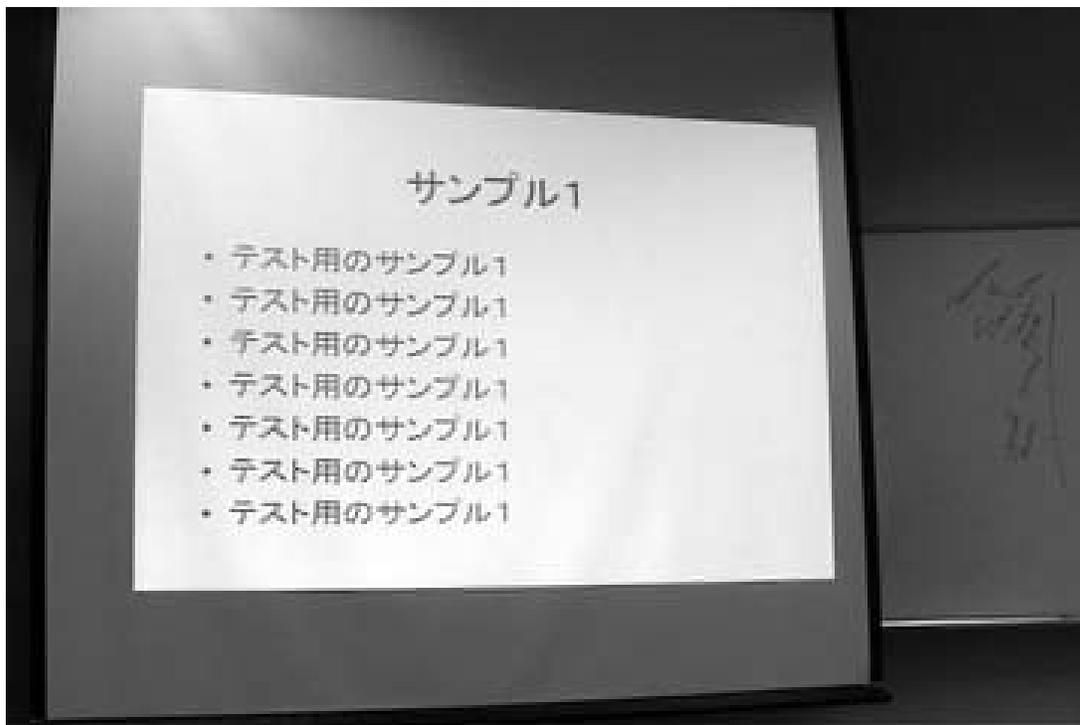


図 5.1: サンプルの初期状態

5.1.1 特徴点の抽出

ここで言う特徴点とは、輝度情報が大きく変化する場所のことである。特徴点は、エッジの抽出を行うことにより取得することができる。

本手法では、典型的なエッジ検出の手法である、平滑化フィルタを施し元画像との差分を取る手法は取らずに、閾値を用いた二値化画像からのエッジ抽出により、特徴点を検出している。なお、閾値はユーザーによって適切な値に設定されることを前提としており、実運用の段階では、二値化画像を実際に見ながら、適切な値を設定できるように設計した。

画像を二値化した後、画像内の全てのピクセルに注目し、各ピクセルの4近傍の値を得る。それらの値のうち、1つが低輝度・残りの3つが高輝度の場合、そのピクセルはエッジにあると判断され、特徴点となる。

図 5.2 は、サンプル画像である図 5.1 の処理結果の例である。二値化に伴い、高輝度の場所をグレイで、低輝度の場所を黒で表し、特徴点を白い点で表している。

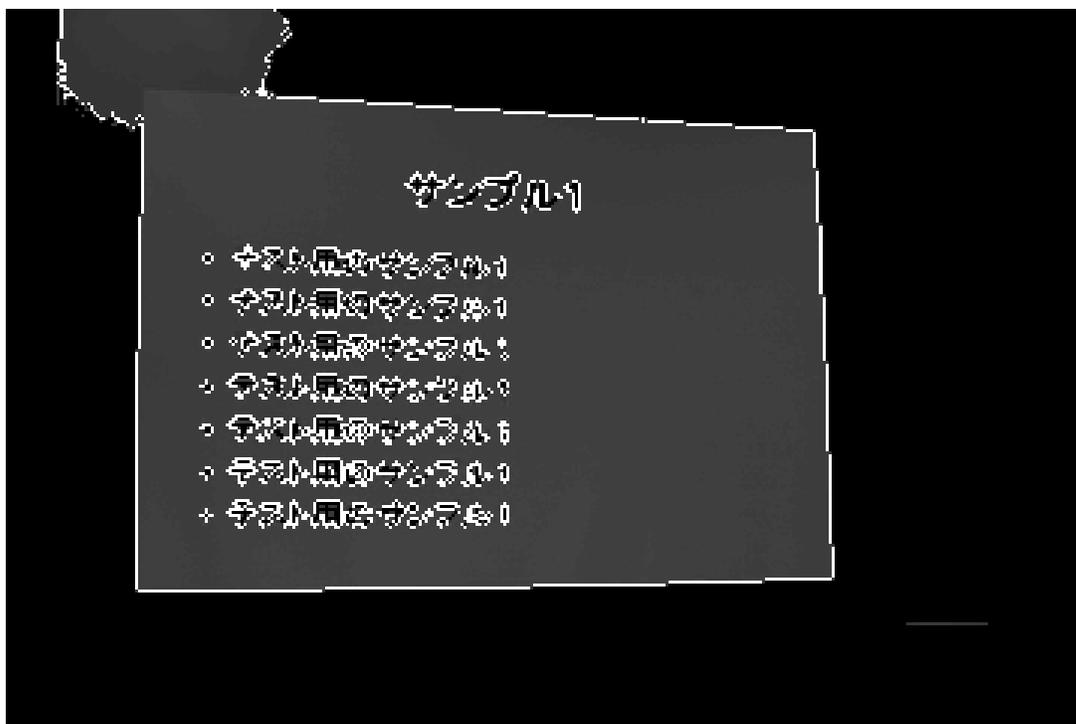


図 5.2: 二値化及び特徴点の例

図 5.2 の例では、スクリーンの周囲はほぼ特徴点として検出できているが、照明の影響で左上部の検出に不都合ができています。また、資料内の文字が閾値を跨いでおり、文字周辺に特徴点が見れているのが分かります。これらのノイズは次節 Hough 変換の際、除去される。

5.1.2 Hough 変換による直線検出

本システムでは直線検出の手法として Hough 変換を用いる。以下では、一般的な Hough 変換の手法及び、本システムでの適用について述べる。

Hough 変換

Hough 変換とは、複数の特徴点から直線や円などの線を表す方程式のパラメタを抽出するための手段として有効な手段である。以下では、直線の方程式に対する Hough 変換について説明する。

ある点 (x, y) とその点を通る直線 $l(y = ax + b)$ がある。この直線 l は無限に存在するが、これらの直線 l と原点との距離を ρ 、原点を通り直線 l と垂直に交わる直線と y 軸とのなす角を θ とすると点 (x, y) と ρ, θ の間には式 5.1 の関係が成立する。

$$\rho = x \cos \theta + y \sin \theta \quad (5.1)$$

$x-y$ 平面の点 (x, y) を $\rho-\theta$ 平面に写像する式 5.1 を Hough 変換という。逆に $\rho-\theta$ 平面の点 (ρ, θ) を $x-y$ 平面に写像することを逆 Hough 変換と言い、式 5.2 で表すことができる。

$$x = -\left(\frac{x}{\tan \theta}\right) + \left(\frac{\rho}{\sin \theta}\right) \quad (5.2)$$

ある特徴点 $P_1(x_1, y_1), P_2(x_2, y_2)$ があり、この 2 点を通る直線を求めたいとする (図 5.3)。

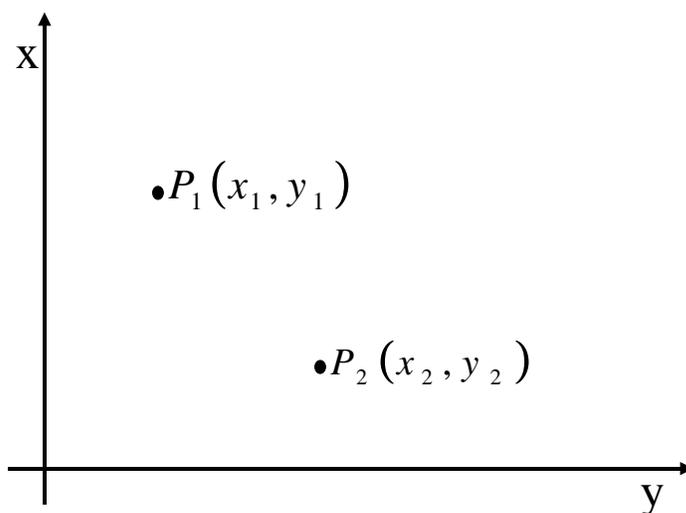


図 5.3: Hough 変換 (特徴点)

この2点をそれぞれ式 5.1 に代入し, $\rho - \theta$ 平面に写像する. そのときの交点を (ρ, θ) とする (図 5.4).

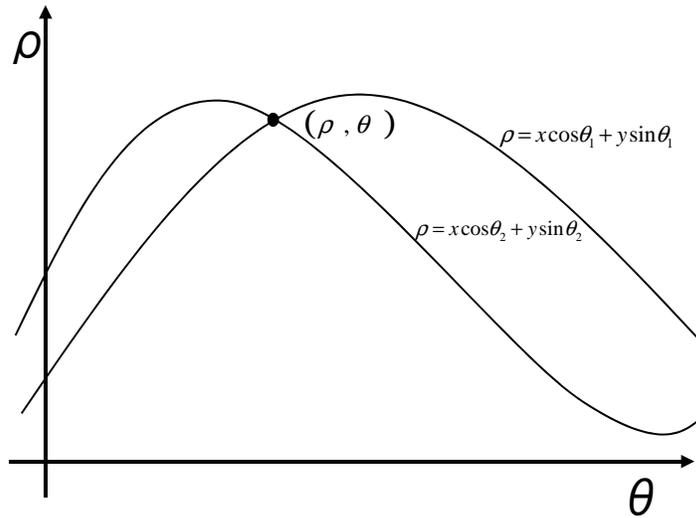


図 5.4: Hough 変換 ($\rho - \theta$ 平面に写像)

それぞれの点 P_1, P_2 を通る直線は無数に存在するが, 唯一点 (ρ, θ) を逆 Hough 変換して得られる直線だけが同じ傾きと y 切片を持っている. すなわち, 点 P_1, P_2 を通る直線である (図 5.5).

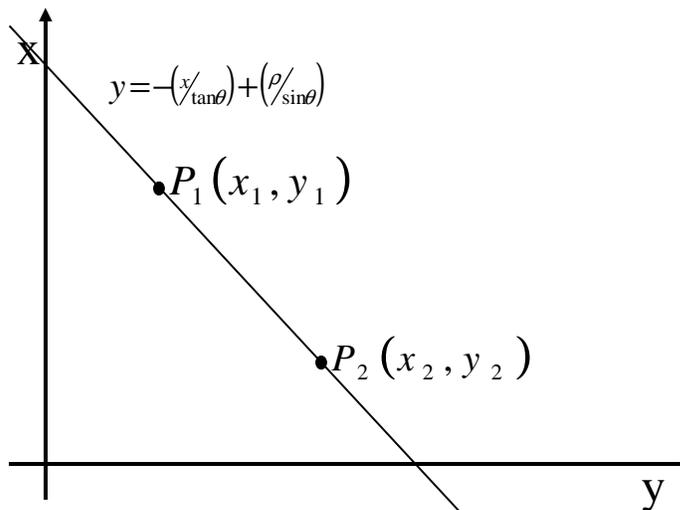


図 5.5: Hough 変換 (点 P_1, P_2 を通る直線)

なお, Hough 変換の後に $\rho - \theta$ 平面での交点を発見する手法として, Hough 変換によっ

て得られた曲線がある区域を通った場合，その区域を表すカウンタを1回加算し，全ての特徴点においてその操作を繰り返した後，カウンタを調べ多くカウントされた場所を交点とする手法が一般的である．

従って，その区域を小さく刻めば検出制度が向上するが，処理に時間がかかる．

本システムにおける適用

本システムにおいて Hough 変換に用いる特徴点は，5.1.1 節で求めた特徴点を用いる．各々の特徴点に対し Hough 変換を施し， $\rho - \theta$ 平面に描画する． $\rho - \theta$ 平面の θ 軸の変域は 0-360 で，これは六十分法の一周期からであり θ 軸の 1 ピクセルに 1 度が対応している． ρ 軸の変域は， $x - y$ 平面の大きさを元に，Hough 変換後にとり得る最小値と最大値を変域とする．具体的には， $x - y$ 平面の最大値を x_{max}, y_{max} とすると，0 から $\frac{2}{3}x_{max} + \frac{2}{3}y_{max}$ の間である．同様に 1 ピクセル毎に 1 が対応している．

図 5.2 における特徴点を元に Hough 変換を施し， $\rho - \theta$ 平面に描画した例を図 5.6 に示す．

描画の後，ピクセルの値が設定した閾値より高く，かつそのピクセルの八近傍の値よりそのピクセルの値が高い，すなわちそのピクセルが頂点となっている部分を抜き出す．図 5.6 では，薄い十字と濃い十字で表した点がそれを表している．

これらの点は，描画した $\rho - \theta$ 平面の解像度やノイズの影響で誤認識した点が多く含まれる．従って，本来一点で示されるはずの頂点が，解像度の低さから複数点で示されてしまっている点があり，それを一点に戻す処理を施す必要がある．この処理は，重心法によるクラスター分析により行った．

初期サンプルとしてそれぞれの点を割り当て，最も近いサンプルを新たに一つのクラスターとしてそれぞれのユークリッド距離がある値以上になるまで処理を繰り返す．その結果いくつかのクラスターが生成され，その中に含まれる点を同一のグループとした．それらのグループの中で最も値の高い点を選び出し，それ以外の点は破棄することにより選別を行った．

図 5.6 では，濃い十字が選ばれた点を示し，薄い十字が破棄された点を示す．これにより，画面内に含まれる直線を表す点の内，独立して存在する点と大きな値を持つ点を同時に選択することができる．

5.1.3 評価関数による選別

Hough 変換及び，クラスター分析による選別によって選ばれた点のうち，スクリーンの 4 辺を表す点を選び出す手法について述べる．

スクリーンを表す直線と x 軸とのなす角度は，スクリーンの上下の辺でほぼ 180 度，スクリーン左右の辺でほぼ 0 度であると言える．従って，選ばれた点の θ 軸の値に注目し， θ が $(-10 < \theta < 10)$ の条件を満たす点を左右の辺の候補， $(170 < \theta < 190)$ の条件を満たす点を上下の辺の候補として選び出し，その他の点は破棄する．

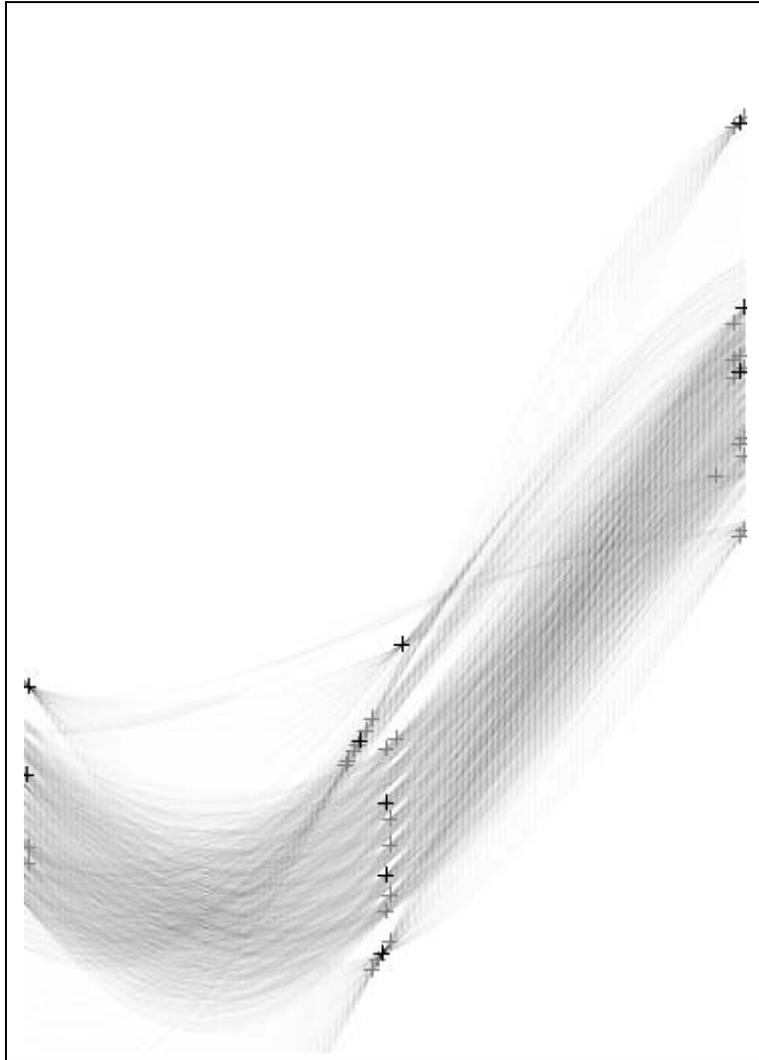


図 5.6: Hough 変換の結果を $\rho - \theta$ 平面に描画した例

これによって選ばれた上下の辺の候補及び、左右の辺の候補について、出来得る全ての組み合わせについて評価関数にかけ評価し、最も評価の高かったもので、なおかつ評価値が設定した値より高いものをスクリーンを構成する 4 辺とみなすこととした。

評価関数

この手法で用いた評価関数は、以下の 4 関数である。なお、これらの関数の戻り値の変域は 0.0 ~ 1.0 で値が低いほど適していることを表す。また、以下では、直線の組み合わせによってできる矩形の四隅の座標を左上 $P_{lt}(x_{lt}, y_{lt})$ 、右上 $P_{rt}(x_{rt}, y_{rt})$ 、左下 $P_{lb}(x_{lb}, y_{lb})$ 、右下 $P_{rb}(x_{rb}, y_{rb})$ として表す。

- アスペクト比による評価

現在用いられているプロジェクタ解像度のアスペクト比は特殊なものを除いてほぼ4:3(縦:横)である。実際にプロジェクタによって映された映像から計測すると、アスペクト比は約4.5:3であることが分かった。よって、評価される矩形のアスペクト比を調べ4.5:3に近いほど適しているとして評価した。

評価 $EVAL$ は以下の式で求めることができる。

$$h_1 = \sqrt{(x_{rt} - x_{lt})^2 + (y_{rt} - y_{lt})^2}$$

$$h_2 = \sqrt{(x_{rb} - x_{lb})^2 + (y_{rb} - y_{lb})^2}$$

$$v_1 = \sqrt{(x_{lb} - x_{lt})^2 + (y_{lb} - y_{lt})^2}$$

$$v_2 = \sqrt{(x_{rb} - x_{rt})^2 + (y_{rb} - y_{rt})^2}$$

$$e = \left| 1 - \frac{\frac{|h_1 + h_2|}{v_1 + v_2}}{\frac{4.5}{3}} \right|$$

$$EVAL = \begin{cases} e & (0 < e < 1) \\ 1 & (e \geq 1) \end{cases}$$

- 矩形のなす角による評価

スクリーンと同じ高さから撮影された場合、スクリーン右辺の両端の角度は等しい。同様に左辺の両端の角度も等しいといえる。従って、長方形のスクリーンが垂直に立っていて、スクリーンと同じ高さから撮影されたと仮定すると、それらの角度から評価を行うことができる。

評価 $EVAL$ は以下の式から求めることができる。

$$\vec{h}_1 = (x_{rt} - x_{lt}, y_{rt} - y_{lt})$$

$$\vec{h}_2 = (x_{rb} - x_{lb}, y_{rb} - y_{lb})$$

$$\vec{v}_1 = (x_{lb} - x_{lt}, y_{lb} - y_{lt})$$

$$\vec{v}_2 = (x_{rb} - x_{rt}, y_{rb} - y_{rt})$$

$$EVAL = \frac{\left| \frac{\vec{h}_1 \cdot \vec{v}_1}{|\vec{h}_1| |\vec{v}_1|} - \frac{\vec{h}_2 \cdot -\vec{v}_1}{|\vec{h}_2| |\vec{v}_1|} \right| + \left| \frac{-\vec{h}_1 \cdot \vec{v}_2}{|\vec{h}_1| |\vec{v}_2|} - \frac{-\vec{h}_2 \cdot -\vec{v}_2}{|\vec{h}_2| |\vec{v}_2|} \right|}{4}$$

- 大きさによる評価

この評価は、4本の線からなる矩形の面積が大きいほど適していると評価される。この評価を導入する理由は、スクリーンに映された資料の中に描かれた矩形が適した矩形と認識されないようにするためである。すなわち他の評価がほぼ同じ矩形が2つあった場合、大きい矩形が選択されるようバイアスをかける意味がある。

この評価 $EVAL$ は以下の式から求めることができる。なお w, h は、それぞれ撮影された映像の縦解像度と横解像度を表す。

$$EVAL = 1 - \frac{\{(x_{rt} - x_{lt}) + (x_{rb} - x_{lb})\} \{(y_{lb} - y_{lt}) + (y_{rb} - y_{rt})\}}{2wh}$$

- 閾値による評価

特徴点を求める際、入力画像を明と暗に2値化した。その二値化情報を元に矩形の内側と外側の明るさを比較評価する。

このとき、矩形の内側と外側全ての二値情報を調べるのではなく、矩形を構成する線分の直近の情報のみを用いる。すなわち、矩形を構成する線分のうち、縦の線分の場合右側と左側、横の線分の場合上側と下側の二値情報を用いる。

評価は、矩形を構成する4つの線分で独立して行われる。線分の各ピクセルごとに縦の線分の場合右側と左側、横の線分の場合上側と下側の二値情報を得る。その2つの二値情報を矩形の内側と外側に分け、内側のピクセルが明だった場合+1、暗だった場合-1する。外側は明だった場合-1、暗だった場合+1する。すなわち1ピクセルごとに適する場合は+2、不適だった場合-2の値をとる。この処理を辺を構成する全てのピクセルに対して行い評価を足し合わせる。その後、ピクセルの数×4で割り、0.5を足すことにより0.0~1.0の間に収まるようスケーリングする。

この処理により、各4辺毎の評価値を得ることができる。実際の評価は4辺のうち一番評価値の低いものをその矩形の評価とする。平均を取らずに最も評価の低い値を採用する理由は、この手法は輝度を用いた領域検出法であり、スクリーンの領域がその周辺より輝度が高く二値化によって明確に区分けされていることが前提となっており、一辺でも評価が低くなっていれば、それは区分けした領域から外れている可能性が高いためである。評価という手段でそのような外れた領域を効率的に除外するには、平均して値を鈍らす手法は適さない。

これらの評価の値の和がその矩形の評価値となる。また実験の段階で、大きさによる評価が厳しすぎ、適切な矩形も認識されない不都合があったため、大きさに対する評価は20%にスケーリングを施してある。

本提案システムの実装にあたって、特徴点を求めるために行われる映像の二値化の際の閾値の設定は、ユーザーが処理結果を見ながら適切な値へ変更する。その際、これまで述べてきた手法により選別された最適な矩形が画面に表示され、その矩形がスクリーンと一致した場合にユーザーの手によって、設定画面を終了させ、運用画面に変更される。設定終了時の矩形はスクリーンの位置と一致していると見なし、アスペクト比の比率及び、矩形のなす角を変更し、それ以降評価にはこの時変更した数値が用いられる。

なお、このスクリーン領域の検出処理は、CPUの空き時間、すなわちアイドル時に表示合成処理とは別のプロセスと優先度で動作している。

スクリーン領域の更新

スクリーン領域の検出処理は常に動作しており、映像が変化したり、カメラが動いた際にスクリーン領域を表す矩形も追従できるよう設計されている。しかし、検出処理が常に動作しているため、映像の少しの変化により矩形領域が変動し安定しない動作が確認された。そこで安定した領域検出が行えるよう、1回前の検出時に最も適していると評価した矩形領域を繰り越し、今回の評価にもその矩形領域を加える手法を採った。これにより、スクリーンの矩形領域の検出が安定して適した位置を保つことができるようになった。

また、評価値がある一定以下になった場合は、スクリーンの前に講師が立つなどして評価が下がったと考えられるため、以前取得したスクリーン領域を保持し、評価値が一定以上になるまで更新は行わない。

5.2 講師の領域抽出

スクリーンの前面に立った講師が講義資料を基に説明を行うとき、講師の影に隠れて講義資料が見えないことがある。この時講義資料が見えるよう、講師の領域を半透明に加工する処理を行う。そのために講師の領域を検出する必要がある。

5.2.1 背景差分法

講師の領域を抽出する手法として背景差分法による移動物体抽出法を用いた。ここで言う移動物体とは、講師の領域をさす。背景差分法とは、あらかじめ設定しておいた背景画像と、入力画像で差分をとり、設定した閾値により2値化を行う手法である。入力画像を $g_i(x, y, t)$ 、背景画像を $b_i(x, y, t)$ 、閾値を D 、時間軸を t とし、 $i = r, g, b$ は画像の各RGB要素を表すものとする、出力 $f(x, y, t)$ は次の式で表すことができる。

$$g_r g b(x, y, t) = g_r(x, y, t) + g_g(x, y, t) + g_b(x, y, t)$$

$$b_r g b(x, y, t) = b_r(x, y, t) + g_g(x, y, t) + g_b(x, y, t)$$

$$f(x, y, t) = \begin{cases} 1 & (|g_{rgb}(x, y, t) - b_{rgb}(x, y, t)| > D) \\ 0 & otherwise \end{cases} \quad (5.3)$$

閾値 D を、大きくとると入力画像中の移動物体が検出できない。小さくとると逆に移動物体中の背景と似た色のピクセルが検出できない。したがって適切な閾値を設定する必要がある。

5.2.2 背景の抽出

背景差分法は、あらかじめ背景画像を取得しておくことが前提となっている。従って初期に背景が設定されていない場合は、未知の背景を推定する手法が必要となる。本システムでは積分法を用いて背景の推定を行った。

積分法は、現在の画像と1フレーム前の画像の加重平均値によって、背景を推定する手法である。加重平均のパラメタを α とし、入力画像を $g(x, y, t)$ 、背景画像を $b(x, y, t)$ とし、 t は時間軸を表すものとする、背景は以下の式で求めることができる。

$$b(x, y, t) = \alpha g(x, y, t) + (1 - \alpha)b(x, y, t - 1)$$

なおこの処理は、RGB 各要素毎に行う必要がある。加重平均のパラメタ α は、大きくとると適用性が高くなり、小さくとると適用性が低くなるが誤検出が抑えられる。

移動物体を考慮した背景抽出

上記の手法では、生成する背景に移動物体があった場合でも背景の更新が行われ、背景中に移動物体の映像が含まれてしまう恐れがある。このため背景に移動物体がある時はその領域の更新をとめ、背景の抽出精度を上げる試みがなされている [7]。

時間軸を t とし、 α は加重平均のパラメタ、 $f(x, y, t)$ は式 5.3 と同じものとする。背景画像を $b(x, y, t)$ 入力画像を $g(x, y, t)$ 、時間軸を t とすると、この手法では以下の式に則り背景画像を更新する。

$$b(x, y, t) = \begin{cases} \alpha g(x, y, t) + (1 - \alpha)b(x, y, t - 1) & (f(x, y, t) = 0) \\ b(x, y, t - 1) & (f(x, y, t) = 1) \end{cases}$$

図 5.7 は、移動体が横切った際の背景と入力画像の関係のグラフである。また、この映像中の輝度は背景と入力画像ともに R 要素のもののみ抜き出して表示してある。

図 5.7 では、時間軸が 150 ~ 250 の間で輝度が大きく変化しており、前面に移動体が通過したことを表している。しかし、抽出を試みている背景画像の輝度は移動体による影響をあまり受けずに背景の輝度とほぼ同じ輝度を表している。しかし、220 ~ 250 フレーム付近で、背景画像の輝度が移動体の影響を受け、輝度値 5 ほど上昇している。これは、移動体の輝度値の変化が閾値を下回ったために移動体が背景として認識されてしまったことを表している。

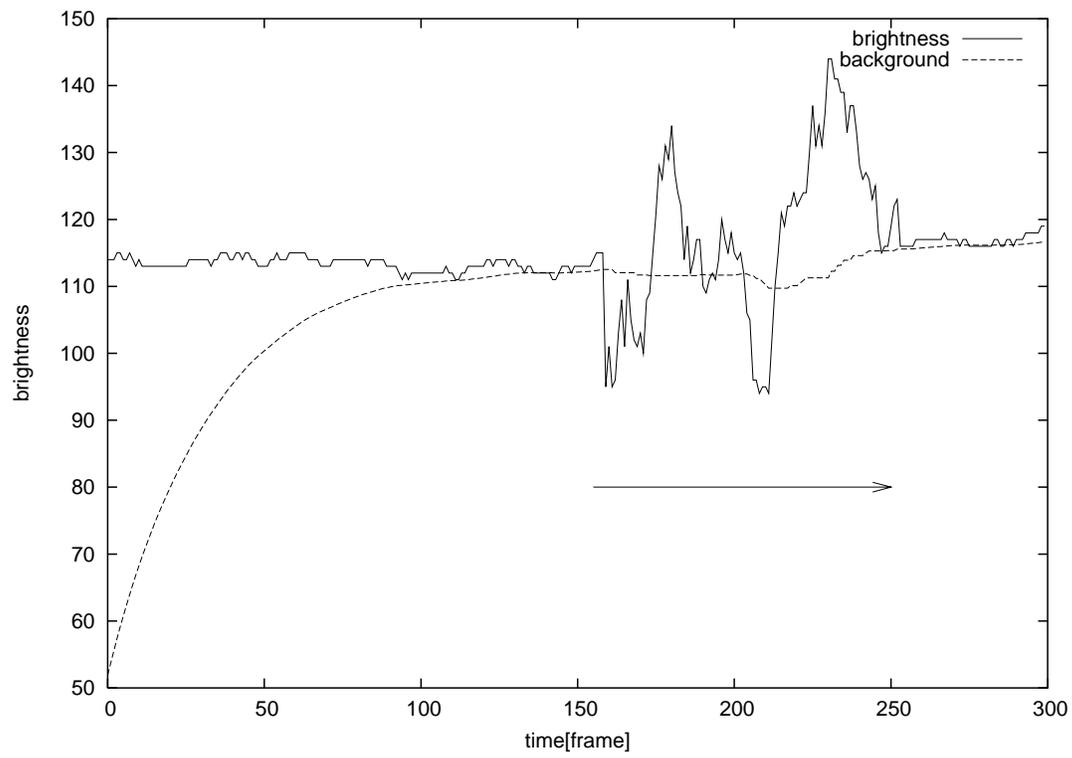


図 5.7: 移動体が横切った際の背景抽出の例

第6章 評価

この章では、本システムを評価し、目標である運用に耐えうるかを調べる。また、評価に基づき考察を行う。

6.1 評価実験

本システムでは、輝度を用いた領域検出手法を用いている。したがって、映し出す資料の輝度が検出精度に影響を及ぼす。そこで、資料画像の輝度の偏りなどの要因による、検出精度の差異を明らかにすべく評価実験を行った。

6.1.1 輝度による検出精度

資料映像の輝度と検出精度の関係を明らかにすべく、輝度分布が一様で異なる輝度値を持つ実験用資料を作成し、輝度と検出精度の関係を明らかにした。

実験は、室内の明るさに左右されるため、比較的実運用に近いと思われる、

- 全ての照明が点いている状態
- スクリーン上の照明のみ消えている状態

の2種類で実験を行った。なお実験は昼間に自然光が入る状態で行った。また、実験は実際の資料提示で使われるであろう機器を想定し、

- 背面投射型プロジェクタ (NEC NX-8061BK)
- 前面投射型プロジェクタ (TOSHIBA TLP-X21:2500 ANSI lumens)

の2台のプロジェクタで行った。

実験は、プロジェクタに実験用資料を投影し、本システムにより検出できるか測定した。

以下はその実験結果の表である。表中の数値は、検出できた最低輝度を表しており、値は最も暗い0から最も明るい256までの値をとり、実験では16刻みで測定した。

実験では、両プロジェクタ共に照明が全て点いた状態でも、半分の輝度値以下の値で検出することができた。また、スクリーン上の照明を消した状態では、さらに検出精度が上がり1/4以下の明るさでも検出することができた。

	全照明点灯	スクリーン上照明のみ消灯
背面投射型	96	48
前面投射型	112	48

表 6.1: 輝度に対する評価実験の結果

6.1.2 PowerPoint テンプレートの検出

PowerPoint では、テンプレートと呼ばれる背景画像を選択することができる。テンプレートは、Microsoft PowerPoint 2000 の場合 61 種類含まれており、これらのテンプレートを含む資料が本システムにおいて認識できるかは、運用の段階では大きな問題となる。

この実験では、これら 61 種類のテンプレートを 6.1.1 節と同じ部屋、機材を用いて投影し、本システムが認識可能か調べた。なお、照明はスクリーン上の照明のみ消灯した状態で実験した。

以下はその実験結果の表である。

	識別できたテンプレート
背面投射型	49/61
前面投射型	50/61

表 6.2: テンプレートに対する評価実験の結果

両環境とも約 8 割のテンプレートを認識することができた。認識できなかったテンプレートは、背景が黒ベースのテンプレートと黒から何色かへのグラデーションのかかったテンプレートであった。

6.1.3 圧縮に対する耐性

本システムは、提示資料の視認性が低下している映像での視認性の向上を目的としている。したがって、低いビットレートで伝送された映像においてスクリーン領域が検出できる必要がある。この実験では、低ビットレートで伝送された映像での領域検出を試みた。

実験は、21kbps で伝送される MPEG4 映像に対して領域検出を試み成功した。なお、21kbps 以下以下の帯域はエンコーダーの制限で試みることが出来なかったため、代わりに画像を意図的に劣化させた画像を複数用意し、その画像に対して領域検出を試みた。

劣化させた画像は 360x240pixel の大きさのテスト画像 (図 5.1) を、縮小処理を施した後に一旦 JPEG に保存し、それを元の大きさに復元することによって生成した。

実験では、7%(25x17pixel) にまで縮小した画像 (図 6.1) までで、検出させることができた。

6.2 考察

これらの実験で資料提示を行うスクリーンの位置検出に関して以下のことが確認できた。

- ある程度の輝度を持つ資料が必要である。
- グラデーションなどで輝度が場所により変化する場合も、一定の輝度がないといけない。
- 良好に検出させるには、スクリーン周囲の明るさを落とす必要がある。

以上の条件さえ満たされていれば、輝度によるスクリーン位置の検出ができることと言える。運用の段階においては、資料の製作者に対してこの制限について説明し、条件にあった資料を作る必要がある。しかし、通常複数枚存在する資料全てがこの条件を満たして必要は無い。なぜなら、一度スクリーン領域を検出させれば、条件を満たしていない検出不能な資料が現れた場合に評価段階で閾値を下回り、その結果スクリーンの位置は以前の位置が保たれる。その間カメラの視点を移動させることは出来ないが、固定している状態ではスクリーン位置が変化することは無い。



図 6.1: 劣化処理した画像

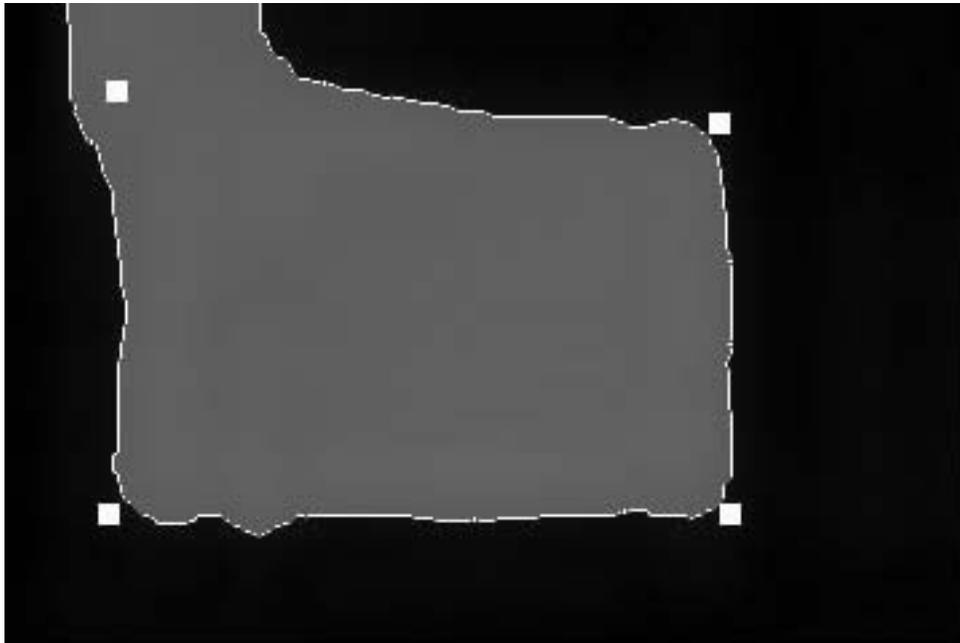


図 6.2: 図 6.1 のスクリーン領域検出

第7章 おわりに

7.1 おわりに

本研究では、視認性とプレゼンス性の向上を目的として、講義映像中に資料をはめ込み合成し、講師の領域を半透明に合成する方式を提案した。

提案を実現させるために、輝度を用いてはめ込み合成に必要なスクリーンの領域を検出する方式及び、背景差分法を用いた講師の領域抽出の手法を採用し、設計/実装を行った。その上で、輝度を用いた領域検出の耐性を調べるべく、実運用を模した環境で実験を行い、輝度を用いて領域検出が可能なことを実証した。

7.2 今後の課題

検出に伴う評価の問題

本手法では、スクリーン検出が正常に行われているかの尺度として4種類の評価を施し、その和を尺度とした。これにより、正常/異常かの判断は可能であるが、なぜ異常な状態なのかは判断できない。これは、スクリーンが画面外に移動したことと、講師などの障害物が検出を妨げているかの判断ができないことを表している。これら二つの状態に対してシステムがとるべき行動は異なる。前者は、追跡不能であるため、はめ込み合成を中止する必要があるが、後者の場合は保持する必要がある。

したがって、これらの状態の違いを判断する機構が必要である。

提示資料の制限の問題

本手法である、輝度を用いた領域抽出は、提示資料にある程度の輝度が必要である。そのことは資料の作成に制限を設けることとなり、好ましくない。したがって、この制限を払拭すべく、他の手法との併用が考えられる。

様々な場面を想定し、負担が少なく正確に領域が抽出できるよう、複数の検出法を実装しユーザーが選択可能であれば好ましい。

謝辞

本研究を行うにあたり，常日頃より，数多くの有益なご指導，ご助言を賜りました篠田陽一教授に心より感謝いたします．

また，貴重なご意見を多数いただきました篠田研究室所属の皆様にも厚く御礼申し上げます．

参考文献

- [1] Information and Computer Science Division at Lawrence Berkeley National Laboratory, <http://www-nrg.ee.lbl.gov/>
- [2] WIDE project, School of Internet workgroup <http://www.soi.wide.ad.jp/>
- [3] IBM 東京基礎研究所, <http://www.trl.ibm.com/projects/learning/webshare.htm>
- [4] Center for Digital Innovation, University of California at Los Angeles
<http://www.cdi.ucla.edu/>
- [5] BCN 総研, <http://www.computernews.com/>
- [6] P.V.C.Hough, "Method and means for recognizing complex patterns." U.S.Patent 3069654,1962
- [7] 川端敦, 谷藤真也, 諸岡泰男, 移動物体像の抽出技術 情報処理学会論文誌 vol28,no4 pp395-402 april 1987