

# 統合プレゼンテーションコンテンツ検索における レーザーポインタ情報の利用

仲野 亘<sup>†</sup> 越智 悠太<sup>††</sup> 小林 隆志<sup>†††</sup> 勝山 裕<sup>††††</sup> 直井 聡<sup>††††,††††</sup>  
横田 治夫<sup>††,††††</sup>

<sup>†</sup> 東京工業大学 工学部 情報工学科 〒 152-8552 東京都目黒区大岡山 2-12-1

<sup>††</sup> 東京工業大学 大学院 情報理工学研究科 計算工学専攻 〒 152-8552 東京都目黒区大岡山 2-12-1

<sup>†††</sup> 東京工業大学 学術国際情報センター 〒 152-8552 東京都目黒区大岡山 2-12-1

<sup>††††</sup> 株式会社 富士通研究所 〒 211-8588 神奈川県川崎市中原区上小田中 4-1-1

E-mail: <sup>†</sup>{wnakano,yutaochi}@de.cs.titech.ac.jp, <sup>††</sup>tkobaya@gsic.titech.ac.jp,

<sup>††††</sup>{katsuyama,naoi.satoshi}@jp.fujitsu.com, <sup>††††</sup>yokota@cs.titech.ac.jp

あらまし 我々はこれまで、プレゼンテーションにおける資料とその撮影動画をメタデータによる統合コンテンツとして蓄積し、その特性を利用して検索・提供する UPRISE を提供してきた。本稿では、この UPRISE の検索機能に対し、講義、講演中で講師が用いるレーザーポインタの情報を利用する方法を提案する。提案手法では撮影動画から画像認識によりレーザーポインタの指す位置情報を抽出し、その付近の文字列情報を利用した検索関数を用いる。さらに、実際の講義コンテンツに適用することで提案手法を評価する。

キーワード 情報統合, 情報検索, e-learning

## Application of Laser Pointer Information for Retrieving Unified Presentation Contents

Wataru NAKANO<sup>†</sup>, Yuta OCHI<sup>††</sup>, Takashi KOBAYASHI<sup>†††</sup>, Yutaka KATSUYAMA<sup>††††</sup>, Satoshi  
NAOI<sup>††††,††††</sup>, and Haruo YOKOTA<sup>††,††††</sup>

<sup>†</sup> Department of Computer Science, Tokyo Institute of Technology  
Ookayama 2-12-1, Meguro-ku, Tokyo, 152-8552 Japan

<sup>††</sup> Department of Computer Science, Graduate School of Information Science and Engineering,  
Tokyo Institute of Technology Ookayama 2-12-1, Meguro-ku, Tokyo, 152-8552 Japan

<sup>†††</sup> Global Scientific Information and Computing Center, Tokyo Institute of Technology  
Ookayama 2-12-1, Meguro-ku, Tokyo, 152-8552 Japan

<sup>††††</sup> FUJITSU LABORATORIES LTD.

Kamikodanaka 4-1-1, Nakahara-ku, Kanagawa, 211-8588 Japan

E-mail: <sup>†</sup>{wnakano,yutaochi}@de.cs.titech.ac.jp, <sup>††</sup>tkobaya@gsic.titech.ac.jp,

<sup>††††</sup>{katsuyama,naoi.satoshi}@jp.fujitsu.com, <sup>††††</sup>yokota@cs.titech.ac.jp

**Abstract** We have proposed unifying presentation contents, such as lecture video and presentation slides used in lectures, using metadata. For the unified contents, we have also proposed a search mechanism named UPRISE (Unified Presentation contents Retrieval by Impression Search Engine). In this paper, we pay attention to information of laser pointer and propose a method of using it for retrieval. we extract position of laser pointer from movie, and use slide's sentences that are near from laser pointer position, for retrieval. We evaluate to measure the effectiveness of our approach.

**Key words** information integration, information retrieval, e-learning

## 1. はじめに

近年、動画や文書、音声ストリームなど複数のメディアコンテンツを統合し、それらを蓄積、検索するシステムが数多く研究および提案されており、e-Learning を始めとする様々な用途に用いられている [1]~[4]。このようなシステムの中で、講義や講演の内容を対象としたシステムでは、それらで講師や講演者が用いるレーザーポインタの情報を強調表示に利用しているものも提案されている [5], [6]。レーザーポインタによるスライド内の文字や図の指示は、講師や講演者が指示箇所に対する注目を引くために使われる場合が多いため、この情報を検索に利用することは効果的であると考えるが、そのようなシステムはまだ提案されていない。そこで本稿では、レーザーポインタの情報を統合プレゼンテーションコンテンツの検索に利用することを目的とする。

教育コンテンツの統合機構、および統合コンテンツに対する高度な検索機能を実現するシステムとして、我々は UPRISE(Unified Presentation contents Retrieval by Impression Search Engine)[7]~[12] を提案してきた。UPRISE は、プレゼンテーションを撮影した動画と、用いた資料等の教育コンテンツをメタデータによって統合することで、それらの教育コンテンツの同期表示を実現する。加えて、統合された教育コンテンツに対する高度な検索機能や、その検索結果を効率的に提供するユーザインターフェースを備えている。UPRISE では、メタデータによるコンテンツの統合のために動画ストリームをシーンの連続であると抽象化し、各シーンとそこで使用された資料とを対応付けている。そしてシーンと資料との対応情報、および各シーン、資料の検索用インデックスをデータベースに格納し、検索に利用している。

本稿では、レーザーポインタの情報を UPRISE の検索機能に利用し、その精度を向上させる手法を提案する。UPRISE で提案しているコンテンツ検索機能は、検索キーワードに対する適合度という、コンテンツの特徴を活かした検索指標を利用している。適合度とは、あるキーワードに対してプレゼンテーション中の各シーンごとに算出されるポイントであり、現在はシーン中で使用している資料スライドの情報や、そのスライドの説明に要した時間等を利用して計算されている。UPRISE はこの適合度を用いて、ユーザに動画のシーン中で用いられたスライドのサムネイルを適合度順に表示する。これにより従来のビデオやマルチメディアプレゼンテーションに対する検索手法 [3], [13], [14] では実現することのできない、重要なシーンの効率の良い検索を可能にしている。

提案手法では、レーザーポインタの情報を画像認識を用いて動画から抽出し、その情報をスライド上で光点付近の行文字列と、ポインタが当たっていた時間の組として取得する。取得したポインタ情報は従来の UPRISE のモデルには統合できないため、従来のシーンに対してより細かい概念であるサブシーンという抽象モデルを導入し、ポインタ情報を UPRISE のシステムに組み込む。こうして統合・蓄積したポインタ情報を指示回数と指示時間という形で数値化し、検索語に対してのシーン適合

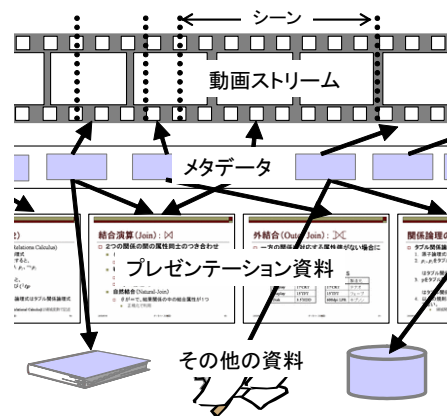


図1 プレゼンテーション資料と動画の統合

度に組み込むことが可能になるようにする。そして従来の適合度に対し、ポインタ情報を統合する箇所を 1). スライド文書構造情報、2). シーン時間情報、3). シーン適合度全体、の3通りにすることで、大きく分けて3つの新しい適合度を提案する。さらに、実際の講義コンテンツを UPRISE に登録し、提案手法によってポインタ情報を抽出したのち、提案した適合度を従来の適合度と比較する検索実験を行い、提案手法が UPRISE の検索精度を向上させることを確認する。また、提案した適合度の持つ、ポインタ情報と従来適合度との影響度合の比率を指定するパラメータを変化させて検索実験を行い、パラメータが検索結果に及ぼす影響についても考察を行う。

## 2. UPRISE の概要

### 2.1 UPRISE のシステム

UPRISE のコンテンツ統合の概念図を図1に示す。メタデータには、動画のどの時刻にスライドの切り替えが起こったかというシーン情報と、その際にどのスライドを用いていたかという同期情報、スライドに含まれる文字列情報に対するインデックスを含める。これらの情報を保持するメタデータによってコンテンツを緩く結合することにより、個々のコンテンツが持つ情報に修正を加えることなくコンテンツの同期表示を実現し、柔軟な統合を可能にしている。また、このメタデータから得られるスライドの使用順序やスライドごとの説明に要した時間という情報を用いることによって各シーンの特性が具体化され、シーンの特性に基づいた検索が可能になる。UPRISE のシステムの詳細についてはこれまでの報告 [11] を参照されたい。

UPRISE では、動画中に同じスライドが複数回出現する場合にそれらを異なるシーンとして区別し、個別に適合度を算出する。これにより、それぞれのプレゼンテーションは対応する動画のシーンの集合として抽象化され、プレゼンテーション中の任意のシーンが検索可能になる。

以下では、UPRISE の検索において用いる、従来の適合度算出手法について述べる。

### 2.2 従来の適合度算出手法

現在、UPRISE は以下の適合度を検索に使用している。

- $I_p$ : スライドの文書構造 (インデント, 文字のサイズ等)

を考慮した適合度

- $I_d$  :  $I_p$  にシーンの時間情報を考慮した適合度
- $I_c$  :  $I_d$  に前後のシーンを考慮した適合度
- $SFP$  : キーワードの特定性を考慮した適合度

なお、複数キーワードにおける適合度の積和計算等に関しては、[10]を参照されたい。

### 2.2.1 適合度 $I_p$

適合度  $I_p$  はスライドの文書構造を考慮した適合度であり、以下の式によって定義される。

$$I_p(s, k) = \sum_{l=1}^L P(s, l) \cdot C(s, k, l)$$

ここで、 $s$  はシーン、 $k$  はキーワード、 $l$  は行数であり、 $P(s, l)$  はシーン  $s$  で用いられたスライドの行  $l$  に与えられるポイント、 $C(s, k, l)$  はシーン  $s$  で用いられたスライドの行  $l$  にキーワード  $k$  が含まれる個数を表している。さらに  $P(s, l)$  において行のインデントや文字の大きさに応じて重み付けをし、キーワードの出現回数だけではなく出現位置も考慮することができる。

### 2.2.2 適合度 $I_d$

適合度  $I_d$  は  $I_p$  にシーンの時間情報を付加した適合度であり、以下の式によって定義される。

$$I_d(s, k, \theta, u) = \left( \frac{T(s)}{u} \right)^\theta \cdot I_p(s, k)$$

ここで、 $T(s)$  はシーン  $s$  の時間であり、 $\theta$  は時間の影響の強弱を定めるパラメタ、 $u$  は単位時間を定めるパラメタである。これによって、長い説明を行っているシーンを重要視することができる。

### 2.2.3 適合度 $I_c$

適合度  $I_c$  は  $I_d$  にシーンの前後関係を付加した適合度であり、以下の式によって定義される。

$$I_c(s, k, \theta, u, \delta, \varepsilon_1, \varepsilon_2) = \sum_{\gamma=s-\delta}^{s+\delta} E(\gamma - s, \varepsilon_1, \varepsilon_2) \cdot I_d(s, k, \theta, u)$$

ここで、 $\delta$  は考慮する前後シーンの範囲を定めるパラメタであり、 $E(\gamma - s, \varepsilon_1, \varepsilon_2)$  は前後関係の強弱を定める関数である。 $E(\gamma - s, \varepsilon_1, \varepsilon_2)$  は以下のように定義される。

$$E(x, \varepsilon_1, \varepsilon_2) = \begin{cases} \exp(\varepsilon_1 x) & (x < 0) \\ \exp(-\varepsilon_2 x) & (x \geq 0) \end{cases}$$

この適合度によって、適合度はそのシーンの前後  $\delta$  だけの範囲の影響を受け、 $\varepsilon$  が小さいほど影響を受けやすくなる。例えば  $\delta = 4, \varepsilon_1 = 5.0, \varepsilon_2 = 0.5$  の時、そのシーンの適合度は前後 4 シーンの適合度に影響を受け、後に続くシーンのほうにより強い影響を受ける。

### 2.2.4 適合度 $SFP$

$I_p$  において、キーワードの出現回数の多いシーンに多くの適合度を与えている。これはキーワードの網羅性 (exhaustivity) を考慮した手法で、 $I_p$  を拡張した  $I_d, I_c$  も同様に網羅性を考慮している。

しかし、網羅性への考慮だけでは、キーワードが複数の際に

適切な結果を得ることはできない。多くのシーンに出現するキーワードと、いくつかのシーンのみ出現するキーワードでは、後者の方がシーンをより特定しているのである。多くのシーンに出現するようなキーワードは目的のシーンを特定する性質 (特定性: specificity) が低い。したがって、シーン検索においては、網羅性だけではなくキーワードの特定性も考慮して適合度を算出しなければならない。

適合度  $SFP$  は、キーワードのプレゼンテーション中での出現頻度を表す適合度である。シーン  $s$  が出現するプレゼンテーションを  $P(s)$ 、 $P(s)$  に含まれるシーン数を  $N(s)$ 、プレゼンテーション  $P(s)$  でキーワード  $k$  が出現するシーンの数を  $pf(k, s)$  とし、適合度  $SFP$  を次のように定義している。[12]

$$SFP(s, k) = \log \frac{pf(s, k)}{N(s)}$$

この適合度  $SFP(s, k)$  の逆数を掛け合わせることによって、多くのシーンに頻出するキーワードの適合度への影響を抑え、限られたシーンに出現するキーワードの影響を多く反映することができる。

## 3. ポインタ情報利用の手法

以下では、動画中のレーザーポインタ情報を UPRISE の検索に利用する手法について提案する。まず、動画コンテンツからポインタ情報を抽出する手法について説明し、抽出したポインタ情報をどのようなメタデータとして格納するかについて示す。そのあとで、ポインタ情報をシーンの適合度として扱う手法について述べ、従来の適合度に対してポインタ情報を統合した新しい適合度を提案する。

### 3.1 ポインタ情報の抽出

画像認識によって、動画コンテンツからポインタの光点情報を抽出する [15], [16]。画像認識は 1 秒ごとに行い、光点のスライド中での座標とその時刻との組として光点情報を抽出する。これにより、プレゼンテーション中の時間軸の上のある地点にポインタが出現したことを情報として取得できる。我々の手法では、動画中の光点からの座標抽出は 90 % 以上の精度をもつて行うことができる [16]。

次に、画像認識によって抽出したポインタの光点座標に対してスライド上で最も近い行の文字列を取得する。これは、座標ではなく文字列としてポインタ情報を持つほうが検索の際に利用しやすいためである。また、ポインタを当てるという行為の対象がスライド中の一行であることが多いことも理由の一つである。

同じ文字列を取得した連続の光点を一回のポインタを指す行為と定義し、一つのポインタ情報として統合する。一つのポインタ情報は行文字列の他に、当たった時間の長さの情報を持つ。

また、レーザーポインタはある 1 行に対して正確に当て続けることが容易ではないため、ポインタの光点は対象行から外れてしまうことが多い。そのため、1 秒ごとに光点に最も近い行を取得しているだけでは講師の意図と異なる行をポインタ情報として抽出してしまうことがある。そこで、一回のポインタに対し、近傍の数行をポインタが当たった行の次候補として取得

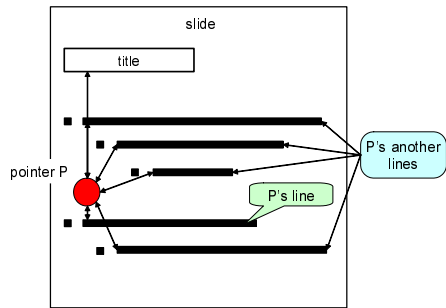


図2 動画認識によるポインタの抽出

```
<?xml version="1.0" encoding="Shift_JIS" ?>
<presentation id="" title="">
<movie filename="" />
<slides>
<slide matId="" page="1" menuTitle="計 算 機 ア ー キ テ ャ 第 二 ( エ ク ラ ス ) 対 象 : 6 字 期 学 籍 番 号 の 右 か ら 二 目 の 数 字 が 直 前 - <pointer>
<dot from="32.000000" x="0.031944" y="0.981481" />
<dot from="36.000000" x="0.004167" y="0.994444" />
<dot from="41.000000" x="0.059722" y="0.972222" />
<rectangle from="32.000000" left="0.098611" right="0.772222" top="0.525926" bottom="0.587037" />
<rectangle from="36.000000" left="0.098611" right="0.772222" top="0.525926" bottom="0.587037" />
<rectangle from="41.000000" left="0.098611" right="0.772222" top="0.525926" bottom="0.587037" />
<line from="32" consumption="10" linePos="0.587037" string="(兼 情報理工学研究所 計 算 工 学 専 攻)">
<line_near rank="1" from="32" consumption="10" linePos="0.735185" string="yokota@cs.titech.ac.jp" />
<line_near rank="2" from="32" consumption="10" linePos="0.370370" string="対 象 : 6 字 期 学 籍 番 号 の 右 か ら 二 目 の 数 字 が 直 前 - <pointer>
<dot from="153.000000" x="0.102778" y="0.940741" />
<dot from="173.000000" x="0.038889" y="0.964815" />
<dot from="191.000000" x="0.444444" y="0.257407" />
<dot from="192.000000" x="0.737500" y="0.242593" />
<dot from="221.000000" x="0.094444" y="0.088889" />
</line>
</pointer>
</slide>
<slide matId="" page="2" menuTitle="講 義 要 要" from="116" consumption="260">
<pointer>
<dot from="153.000000" x="0.102778" y="0.940741" />
<dot from="173.000000" x="0.038889" y="0.964815" />
<dot from="191.000000" x="0.444444" y="0.257407" />
<dot from="192.000000" x="0.737500" y="0.242593" />
<dot from="221.000000" x="0.094444" y="0.088889" />
</pointer>
</slide>
</presentation>
```

図3 ポインタ情報を含んだ同期情報 xml

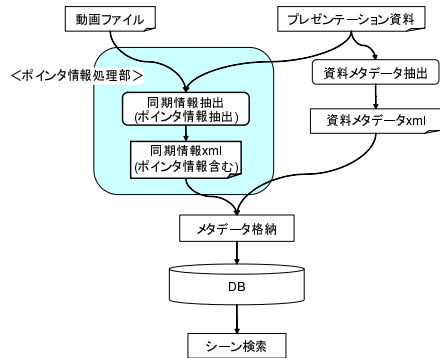


図4 ポインタ情報処理付近の処理フロー

しておく。最も近い行とその付近のいくつかの行を組にしてポインタ情報とすることで、講師の意図と光点とのぶれをある程度解消することができる。光点による行の取得方法を図2に示す。

ポインタ情報は図3のようなxmlとして抽出され、保管される。UPRISEはこのxmlを入力として、ポインタ情報を格納する。この節で述べた処理をUPRISEの処理フローの中から抜き出すと、図4のようになる。

### 3.2 ポインタ情報の抽象化

あるポインタが当たっている間、講師はポインタを当てている行に対して説明をしていると考える。このことから、あるシーンの中で1回のポインタに対応する部分をより細かい概念として抽象化すると、その一つ一つはある事柄に対して説明を

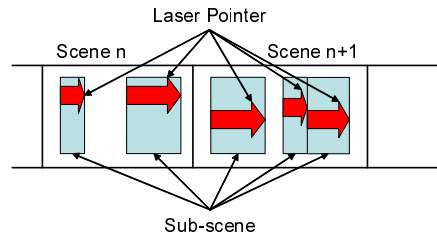


図5 シーン・サブシーンモデル

行っている場面がまとまったものとなる。我々はこの1回のポインタに相当する部分をサブシーンと定義した。したがって、各シーンはより詳細な内容ごとにまとまったサブシーンを複数持つことになる。このように定義することで、各シーンで講師がいくつかの話題について説明しているのを知ることができる。

UPRISEにおけるシーンとサブシーンの関係について図5に示す。ここで、複数のシーンにまたがるポインタはないと考えるため、各サブシーンはある特定のシーンに属することになる。

### 3.3 ポインタ情報の数値化

3.1, 3.2節において、ポインタ情報を時間情報と付近行の文字列という2つのメタデータとして抽出する手法について提案した。抽象化したポインタ情報をシーンごとの適合度として数値化する方法について以下に述べる。

まず、キーワードを含む行に対してポインタが当てられた回数を考慮した適合度を考える。ポインタは確実に講師の意図どおりに当たるわけではなく、抽出した情報も誤差を含む。この問題を解消するため、対象行以外に候補となる行を持つことはすでに述べた。このため、ポインタの回数はその情報の信頼度を考慮し、キーワードが全候補行に含まれていたときに1とするような、回数の期待値として数値化する。こうして得られるポインタごとの回数の期待値をシーンごとに合計し、各シーンの適合度加算分とする。シーン  $s$ 、キーワード  $k$  に対するこの加算分を  $poc(s, k)$  (Pointer Of Count) とし、以下の式で表す。

$$poc(s, k) = \sum_{t=1}^T W(s, t, k)$$

ここで、 $t$  はサブシーンを表し、 $T$  はシーン  $s$  が持つサブシーン数を表す。 $W(s, t, k)$  はシーン  $s$  のサブシーン  $t$  におけるキーワード  $k$  の回数の期待値である。

次に、ポインタの時間情報を考慮した適合度を考える。上で算出したポインタごとの回数の期待値に対して、各ポインタが当たっていた時間を掛け合わせることで、ポインタの時間の期待値が得られる。この時間の期待値をシーンごとに合計し、適合度加算分とする。この加算分を  $pod(s, k)$  (Pointer Of Duration) とする。ここで  $s, k$  は  $poc(s, k)$  と同様である。 $pod(s, k)$  の式を以下のように定義する。

$$pod(s, k) = \sum_{t=1}^T W(s, t, k) \cdot T(s, t)$$

ここで、 $T(s, t)$  はシーン  $s$  のサブシーン  $t$  の時間を表す。すな

わち、サブシーン  $t$  に対応するポイントの当たっている時間を表す。

### 3.4 適合度の提案

ポイント情報による適合度を従来の適合度のどの箇所と統合するかによって、いくつかの適合度が得られる。以下ではポイント情報を統合した新適合度について述べる。

#### • 適合度 $I_{p+poc}$

まず、適合度  $I_p$  の段階、つまりスライドの文書構造情報に統合する手法を提案する。適合度  $I_p$  によるキーワードの網羅性に対し、ポイントの出現回数の期待値である  $poc(s, k)$  を足し合わせ、ポイントが当たっていたときにそのシーンにおけるキーワード出現回数が増えたとみなす。この値を  $I_p$  として  $I_c$  を求めたものを適合度  $I_{p+poc}$  とする。 $I_{p+poc}$  の式を以下のように定義する。

$$\begin{aligned} I_{p+poc}(s, k, \delta, \theta, \varepsilon_1, \varepsilon_2, d) &= \sum_{\gamma=s-\delta}^{s+\delta} E(\gamma - s, \varepsilon_1, \varepsilon_2) \cdot I'_d(s, k, \theta, u, d) \\ I'_d(s, k, \theta, u, d) &= \left( \frac{T(s)}{u} \right)^\theta \sum_{l=1}^L \{P(s, l) \cdot C(s, k, l) + poc(s, k) \cdot d\} \end{aligned}$$

$d$  は  $I_p$  の  $P(s, l)$  と対応するパラメタで、ポイント回数のスライド文書情報に対する影響度を与える。例えば、 $P(s, l)$  においてタイトル行に 10 ポイント与えると定義されていて、 $d = 10$  とした場合、キーワードを含む行にポイントが 1 回当たることは、そのスライドのタイトルにキーワードが 1 つ含まれていることと同等の意味を持つことになる。

#### • 適合度 $I_{d+pod}$

次に、適合度  $I_d$  の段階、シーンの時間情報に統合する手法を考える。シーンごとの時間情報である  $\{T(s)/u\}^\theta$  に対し、ポイントの時間の期待値である  $pod(s, k)$  を足し合わせ、ポイントが当たっていたときにそのシーンの時間に加点する。こうして求めた時間情報を用いて  $I_c$  を求める。この適合度を  $I_{d+pod}$  とし、以下の式で定義する。

$$\begin{aligned} I_{d+pod}(s, k, \delta, \theta, u, \varepsilon_1, \varepsilon_2, \omega_1, \omega_2) &= \sum_{\gamma=s-\delta}^{s+\delta} E(\gamma - s, \varepsilon_1, \varepsilon_2) \cdot I''_d(s, k, \theta, u, \omega_1, \omega_2) \\ I''_d(s, k, \theta, u, \omega_1, \omega_2) &= \left( \frac{T(s) \cdot \omega_1 + pod(s, k) \cdot \omega_2}{u} \right)^\theta \cdot I_p(s, k) \end{aligned}$$

ここで、 $\omega_1$  および  $\omega_2$  はポイントの時間とシーンの時間との影響度を調節するパラメタである。例えば、 $\omega_1 = 1$ 、 $\omega_2 = 10$  の場合、ポイントが 1 秒当たることはそのシーンが 10 秒伸びることに相当する。

#### • 適合度 $I_{c+poc}$ 、 $I_{c+pod}$

最後に、 $I_c$  で与えられた各シーンごとの適合度に対して、ポイント情報による適合度の全ポイント情報に対する比率を掛け合わせる。この手法では、該当シーンのポイント情報のみを考慮している。この手法を  $poc(s, k)$  および  $pod(s, k)$  に対して適

用し、それぞれを  $I_{c+poc}$ 、 $I_{c+pod}$  と表すことにする。 $I_{c+poc}$  および  $I_{c+pod}$  を以下のように定義する。

$$\begin{aligned} I_{c+poc}(s, k, \delta, \theta, u, \varepsilon_1, \varepsilon_2, \alpha) &= I_c(s, k, \delta, \theta, u, \varepsilon_1, \varepsilon_2) \cdot \left( 1 + \frac{poc(s, k) \cdot \alpha}{\sum_{s=1}^S poc(s)} \right) \\ I_{c+pod}(s, k, \delta, \theta, u, \varepsilon_1, \varepsilon_2, \beta) &= I_c(s, k, \delta, \theta, u, \varepsilon_1, \varepsilon_2) \cdot \left( 1 + \frac{pod(s, k) \cdot \beta}{\sum_{s=1}^S pod(s)} \right) \end{aligned}$$

ここで、 $\alpha$ 、 $\beta$  はポイント情報の影響度を表すパラメタである。また、 $poc(s)$  と  $pod(s)$  はシーン  $s$  に属する全サブシーンにおける回数期待値と時間期待値の総和を表し、 $S$  はプレゼンテーションに対するシーンの数である。

## 4. システムの実装と実験

3. 節で提案した手法を用いて、実際の講義のコンテンツを提案手法によって UPRISE に登録し、登録したコンテンツに対して各適合度ごとの検索実験を行う。その後、実験結果に対して考察を行う。

### 4.1 実験に用いたデータ

3.1 節で提案した手法を用いて、実際の講義のコンテンツをポイント情報も含めて UPRISE に登録する。今回は、全 12 回のそれぞれの講義を 1 つのコンテンツとして登録した。

画像認識を行って各コンテンツのメタデータを抽出した結果、得られたことを以下に示す。

- 全シーンの 538 個に対し、サブシーンを含むシーンは 368 個であり、これは全シーンの 68 % であった。
- 全コンテンツにおいて、サブシーンは 3833 個抽出された。
- サブシーンを含むシーンにおいて、サブシーンの個数の平均は 10 個であった。
- サブシーンの平均時間、つまり 1 回のポイントの平均時間は 3.6 秒であった。なお、シーンの平均時間は 93 秒であった。

### 4.2 実験

提案手法の評価を行うため、4.1 節で登録したコンテンツに対し、キーワードについて説明しているシーンを実際に検索する実験を行った。実験は前節で提案した適合度  $I_{c+poc}$ 、 $I_{c+pod}$ 、 $I_{p+poc}$ 、 $I_{d+pod}$  および従来の適合度である  $I_c$  を用い、以下の条件の下で行った。

- パラメタは  $\theta = 0.5$ 、 $u = 60$ 、 $\delta = 4$ 、 $\varepsilon_1 = 5.0$ 、 $\varepsilon_2 = 0.5$  に固定した。
- 1 回のポイントに対し 5 つの候補行を取得し、光点からの距離の近い順に 0.4、0.3、0.15、0.10、0.05 という、回数の期待値を与えた。
- 各適合度ごとに 78 種類のキーワードを検索した。
- 格納された全てのシーンに対して適合度を算出し、順序付けを行った。
- キーワードに対し、事前にプレゼンテーション中で最もよく解説していると判断したシーンをそのキーワードの正解シーンとした。
- 適合度の種類ごとに、正解シーンが何番目に順序付けされたかを記録した。

$u$  を 60 とすることで、シーン時間の基本単位を分としている。これは、4.1 節で述べたように、シーンの平均時間が約 1 分半であるため、シーン中のスライド文書情報とシーン時間の影響比率をある程度揃えるためである。

$\varepsilon_1$  と  $\varepsilon_2$  をこのように定めた理由は、正解シーンとされるシーンの傾向として、検索キーワードの定義や詳細説明を行ったシーンが多いこと、およびそのようなシーンの後ろにはそのシーンの内容を補足したり、具体的な利用方法、より詳細な説明やその発展形などを説明したりするようなシーンが続くことが多い、と考えるためである。つまり、検索キーワードに関するシーンはある程度連続して続く傾向があり、そのシーン群の先頭もしくは先頭付近のシーンが検索結果としてよりよい場合が多いためである。

$\delta$  および  $\theta$  での上記の値は、これまでの研究において用いられた値をそのまま使用しており、その値は経験的に定めたものである。これらのパラメタに関してより適切な値を設定することに関しては、今後の課題とする。

また、回数期待値を上記の様に定めた理由は、レーザーポインタ指示時の手のぶれを考慮すると、第一候補および第二候補行を指す意図がある可能性が高いと考えたため、そのために第一、二候補行の重みを比較的重く設定した。

なお、評価に際して指標となる再現率 (recall) と適合率 (precision) [17] について述べる。再現率は検索結果に含まれていた正解が全正解の中で占める割合、適合率は検索結果に対して正解が占める割合であり、今回の実験においては以下の式で表される。

$$\text{再現率} = \frac{\text{検索で得られた正解シーン数}}{\text{全正解シーン数}}$$

$$\text{適合率} = \frac{\text{検索で得られた正解シーン数}}{\text{検索で得られたシーン数}}$$

今回の実験では、正解シーンを各キーワードに対して 1 つしか与えていないため、再現率は常に 0 か 1 をとる。しかし、今回全ての試行において検索結果には正解シーンが含まれているため、以下の結果における再現率は常に 1 である。

また、検索結果の範囲は正解シーンの順位までとする。これは、UPRISE の検索インターフェースでは適合度が 0 でないシーンが全て表示される仕様となっていること、および検索を行うユーザは検索結果の上位からシーンを見ると考えることができることから、検索精度の評価において正解シーン以下の順位にあるシーン群は無視できると考えるためである。

以上の前提のもとでは、適合率が以下の式で求まる。ただし、 $N$  は検索回数である。

$$\text{適合率} = \frac{1}{N} \sum_{i=1}^N \frac{1}{i \text{ 番目の検索での表示順位}}$$

適合率が大きいほど、正解シーンが上位に表示される場合が多いことを示し、システムがより高性能な検索機能を有していることを示す。

#### 4.3 実験結果と考察

表 1 は各適合度における検索結果を順位別に表したものであ

る。なお、表示順位は 10 位までに限っており、評価の指標として適合率、正解シーンが 3 位までに入った検索の件数、および全検索中最も悪かった順位を併記した。

表 1 での各適合度における、ポインタ情報の影響比率を定めるパラメタは、 $d = 10$ 、 $\omega_1 = 1$ 、 $\omega_2 = 1$ 、 $\alpha = 10$ 、 $\beta = 10$  とした。これは、4. 節で得られたことなどを参考に、ポインタ情報とそれを統合した情報とのバランスを適切にとるように定めたものである。

表 1 各適合度による検索結果

表示順位	$I_c$	$I_{p+poc}$	$I_{d+poc}$	$I_{c+poc}$	$I_{c+pod}$
1	32	27	34	34	34
2	17	15	16	15	16
3	8	12	9	9	7
4	8	6	6	6	7
5	3	6	4	4	3
6	3	0	3	3	4
7	1	4	1	1	1
8	1	2	1	1	1
9	0	0	0	0	0
10	0	0	0	0	0
適合率	0.601	0.544	0.608	0.614	0.615
3 位までの件数	57	54	58	58	57
最も悪い順位	30	32	30	26	27

表 1 より、適合度  $I_{p+poc}$  以外の適合度では従来の  $I_c$  よりも適合率が上回っていることがわかる。したがって、ポインタ情報による適合度加算が正解シーンの順位を引き上げる効果を持つと言える。つまり、利用者がより求めているシーンに対し、多くのポインタが用いられる傾向があることが導かれる。

なお、正解シーンの中でサブシーンを持たないシーンの数は、78 シーン中 20 個であった。利用者が求めるシーンにはおおむねポインタが使われていると言えるが、20 キーワードについてはポインタ情報を考慮することによる精度の改善は起こらないということでもある。また、全シーンにおいてポインタが一度も当たらなかった検索キーワードは 78 キーワード中 7 個であった。これらのキーワードを用いた検索においては、実験で用いたどの適合度においても検索結果に差異は生じない。

適合度  $I_{p+poc}$  の結果が良くないため、次に  $I_{p+poc}$  について追加実験を行った。 $I_{p+poc}$  におけるパラメタとして  $d$ 、そして  $poc(s, k)$  内での回数の期待値の与え方がある。そこで、回数の期待値を変更して検索実験を行った。その結果を表 2 に示す。

表 2 では、大きな精度の向上は見られなかった。そのため、 $I_{p+poc}$  の精度低下の原因は回数期待値の重み付けではないと考えられる。また、複数候補をとっていない場合により精度が低下していることから、この適合度に関しては、第 2, 3 候補を取ることが光点のぶれを吸収する目的を果たしていること、および、第 4, 5 候補行は検索結果にほぼ影響を与えていないことが確認できた。

$I_{p+poc}$  の精度が良くない原因について、ポインタの回数情報が正確に取得できていないことを考える。今回、画像認識からポインタを統合する際に、第一候補行が同一であるかどうかをポインタ統合の指標とした。このため、例えば 10 秒の行  $l$  を指すポインタにおいて、ポインタ光点がぶれ、取得した第一候補行が行  $l, l+1, l, l-1, l$  のようにぶれてしまった場合では、本来は 1 つであるポインタが 5 つに分割されてしまい、行  $l$  に当たった回数も 1 回から 3 回に誤認識されてしまう。このよう

表2  $I_{p+poc}$  におけるポイント回数の期待値の影響

表示順位	0.4, 0.3, 0.15, 0.1, 0.05	0.5, 0.4, 0.1, 0, 0	0.7, 0.3, 0, 0, 0	0.9, 0.1, 0, 0, 0	1.0, 0, 0, 0, 0
1	27	28	28	26	24
2	15	15	16	15	17
3	12	10	8	10	10
4	6	8	9	8	8
5	6	3	2	6	6
6	0	2	2	0	1
7	4	4	5	4	3
8	2	2	2	3	2
9	0	0	1	1	2
10	0	0	0	0	0
適合率	0.544	0.551	0.552	0.531	0.518
3位までの件数	54	53	52	51	51
最も悪い順位	32	29	29	29	28

表3  $I_{p+poc}$  におけるパラメタ  $d$  の影響

表示順位	$d = 1$	$d = 10$	$d = 20$	$d = 30$
1	30	27	27	25
2	19	15	10	9
3	6	12	13	13
4	10	6	7	7
5	3	6	4	2
6	3	0	4	7
7	2	4	4	4
8	0	2	0	0
9	0	0	1	1
10	0	0	1	0
適合率	0.586	0.544	0.523	0.517
3位までの件数	55	54	50	49
最も悪い順位	31	32	29	28

にポイントの回数情報を誤って取得した結果、たまたま回数の多い余分なシーンが上位にきてしまったのではないかと考える。

表3は期待値を0.4, 0.3, 0.15, 0.1, 0.05に固定した上で、パラメタ  $d$  を変化させた結果を示す。なお、以下の全ての実験では回数の期待値はこの0.4, 0.3, 0.15, 0.1, 0.05という値に固定した。

表3より、回数の期待値を変化させたときよりも精度への影響は大きいことがわかる。さらに、 $d$  を大きくするほど精度が下がってしまっている。このことから、ポイント情報による適合度への影響は、従来のスライド文書構造、シーン時間、文脈構造などと適切な比率をもって考慮すると良い結果が得られるのではないかと推察する。この推察を確かめるために、他の適合度  $I_{d+pod}$ ,  $I_{c+poc}$ ,  $I_{c+pod}$  についても各パラメタを変化させて検索実験を行った。

表4, 5, 6はそれぞれ、 $I_{d+pod}$ ,  $I_{c+poc}$ ,  $I_{c+pod}$  のパラメタ  $\omega_2$ ,  $\alpha$ ,  $\beta$  を変化させた結果を表している。 $\alpha$ ,  $\beta$  は10付近、 $\omega_2$  は1付近に向かって精度が向上することが期待できる。パラメタの最適値がばらつくのは、パラメタと適合度ごとに影響度が異なるためであると考え。今後、その比率を与える適切な指針を導き出すことが課題となる。

また、表4, 5, 6に見られる特徴として、ポイントの影響度を大きくするにつれて、全キーワードの順位の中で最も悪い順位が改善されていることがある。これは、ポイント情報による有効なシーンの順位上昇が低い順位のシーン間においても起きていて、かつ、それが有効に働いていることを示している。適合率には現れないが、これもポイントの有効性を示す一つの結果と考えてよい。音声データなどの他メディアと連携を行えば、実質的な有効性となる可能性が十分あると考える。

## 5. おわりに

本稿では、講義中に講師が用いるレーザーポイントの情報をUPRISEの検索に利用することを提案した。まずポイント光点座標を画像認識により抽出し、座標からポイント情報を行文字列と指示時間として抽出した。次にポイント情報をUPRISEのモデルに統合するため、サブシーンという抽象モデルを導入した。さらに、ポイントの情報を適合度に統合するために、ポイントの指示回数および指示時間の期待値という形式でポイント情報を数値化し、それらを用いてポイント情報を従来のUPRISEの適合度に考慮する手法を提案した。また、実際の講義のコンテンツをUPRISEに登録し、検索キーワードを用いて評価実験

を行い、提案した適合度の検索精度に対する有効性を確認した。

評価実験の結果、提案した適合度のうち  $I_{p+poc}$  以外のものであることは、従来の  $I_c$  よりも検索精度が向上することを確認した。このことから、レーザーポイントの情報はシーンの重要性を示す性質を持ち、検索の指標の一部として用いることが妥当であることを示した。また、提案した適合度において、ポイント情報と従来適合度との影響比率を定めるパラメタを変化させて検索実験を行い、パラメタが検索結果に与える影響について考察を行った。その結果、パラメタごとにある特定の範囲において精度が極大値を示し、その範囲を外れると徐々に精度が低下していくことを確認した。これにより、適切なパラメタの設定をすることが検索結果に対してよい影響を与えることを確認した。

最後に、本研究の今後の課題を以下に述べる。

- ポイント情報の正確な取得

4.3節でも述べたように、現在のポイントの使用回数の認識方法では、実際の使用回数と異なる結果が出る場合がある。このことは、適合度  $I_{p+poc}$  の結果が  $I_c$  よりも良くないことの大なる原因であるため、使用回数の認識方法を改善する必要がある。

また、今回光点から最も近い行を取得し、連続して同じ行を取得した光点を一つのポイントとしてまとめるという手法をとった。しかし、ある光点と次の光点との間で時間の間隔が空き、且つ2つの光点がたまたま同じ行を取得してしまった場合、動画上ではこの2つの光点の間ではスライドにはポイントが当たっていないにもかかわらず、2点間を統合して一つのポイントとして認識してしまう。このような場合があるため、実際のポイント区分とは異なる結果を認識してしまうことがある。そのため、時間情報の取得手法についても改善を行う必要がある。

- 各パラメタの最適化

今回適合度のパラメタは任意に指定したいくつかの値を用いたが、4.3節でも述べたようにパラメタを最適化することでより精度を向上することができる。上で述べた正確なポイント情報取得のための閾値も、最適化の対象になると考える。

- 単語単位ポイント情報への対応

今回、ポイント情報として、ポイント光点付近の行文字列を単位として抽出した。しかしポイントを指す行為には、行ではなく単語に対して焦点を当てるという場合も多く存在する。そこで、画像認識で得られた座標付近の単語を抽出し、検索に用いることでよりポイントの影響を特定することができる。

さらに、今回の提案手法との比較も行い、より適切なポイン

表4  $I_{d+pod}$  におけるパラメタ  $\omega_2$  の影響

表示順位	$\omega_2 = 1$	$\omega_2 = 10$	$\omega_2 = 20$	$\omega_2 = 30$
1	33	31	31	29
2	16	18	18	19
3	9	6	8	8
4	6	6	3	4
5	4	4	5	4
6	3	5	4	3
7	1	2	3	5
8	1	0	1	1
9	0	1	0	0
10	0	1	0	0
適合率	0.608	0.588	0.589	0.572
3位までの件数	58	55	57	56
最も悪い順位	30	26	21	24

表5  $I_{c+poc}$  におけるパラメタ  $\alpha$  の影響

表示順位	$\alpha = 1$	$\alpha = 10$	$\alpha = 20$	$\alpha = 30$
1	33	34	33	32
2	16	15	17	20
3	9	9	8	7
4	7	6	6	5
5	3	4	4	4
6	3	3	3	3
7	1	1	1	2
8	1	1	1	0
9	0	0	0	0
10	0	0	0	0
適合率	0.608	0.614	0.610	0.610
3位までの件数	58	58	58	59
最も悪い順位	30	26	21	18

表6  $I_{c+pod}$  におけるパラメタ  $\beta$  の影響

表示順位	$\beta = 1$	$\beta = 10$	$\beta = 20$	$\beta = 30$
1	33	34	31	33
2	16	16	16	16
3	8	7	9	6
4	7	7	8	9
5	3	3	3	3
6	4	4	4	4
7	1	1	2	2
8	1	1	0	0
9	0	0	0	0
10	0	0	0	0
適合率	0.606	0.615	0.588	0.604
3位までの件数	57	57	56	55
最も悪い順位	30	27	25	22

タ情報の抽出方法を検討する必要があると考えている。

#### ● ポインタの使用傾向の考慮

ポインタの使用傾向は講師によって異なる部分が多い。使用頻度によって、ポインタを指すことの価値も異なる。全部のスライドに対して網羅的にポインタを使うのであれば、ポインタ一つ一つの価値は低い。逆に極めてポインタが限られた回数しか用いられない場合でも、ポインタを指すことを忘れていたという場合もある。

また、スライドの種別によっても、ポインタを指すことの意味が変わることもあると考える。例えば目次のスライドでは、項目が多いためにポインタを多用することが多いが、そのスライドを用いたシーンの重要度が大きいとは言えないことが多い。また、図を含むスライドでは図中に余分な行要素が散在することによって、ポインタが多く使われているにも関わらずキーワードを持たないために重要度が上がらないという問題もある。これらを分析し、有用なポインタとそうでないポインタを区別しなければならない。

#### ● 音声情報の利用

シーン中で説明している内容をより細かく取得できる方法として、動画中の音声情報の利用を検討している。音声認識もしくは文字起こしを行ったテキストをポインタ情報の文字列と連携させることにより、よりシーンの特性を正確に抽出することができると考える。

また、ポインタの誤認識および指示部の揺れに対して本稿では複数候補をとるという信頼度での対応を行ったが、音声情報を用いればポインタ情報の精度を向上させることができると考える。

## 謝 辞

本研究の一部は、文部科学省科学研究費補助金特定領域研究(15017233,16016232)、独立行政法人科学技術振興機構 CREST、および21世紀COEプログラム「大規模知識資源の体系化と活用基盤構築」の助成により行われた。

## 文 献

- [1] 株式会社富士ゼロックス: MediaDEPO. <http://ubiquitous-media.com/mediadepo.htm>.
- [2] 株式会社リコー: MPMeister. <http://www.ricoh.co.jp/mpmeister/>.
- [3] R. Müller and T. Ottmann. The "Authoring on the Fly" system for automated recording and replay of (tele)presentations. *Multimedia Syst.*, 8(3):158-176, 2000.
- [4] 株式会社アイ・ピー・イー: EZ プレゼンター. [http://www.mpeg.co.jp/products/ez\\_presentater/index.html](http://www.mpeg.co.jp/products/ez_presentater/index.html).

- [5] 株式会社アーネット: 遠隔講義システム P4Web. [http://www.earnnet.co.jp/rims\\_main.html](http://www.earnnet.co.jp/rims_main.html).
- [6] 株式会社富士通インフォソフトテクノロジー: Internet Navigware. <http://www.navigware.com/>.
- [7] 横田 治夫. 東工大学術国際センターの情報蓄積・活用 教育コンテンツの統合とその手法 - 研究会報告 DBS-125-58, 情報処理学会, 2001.
- [8] 村木 太一, 吉田 誠, 小林 隆志, 直井 聡, 横田 治夫. メタデータによる講演資料と動画の統合と検索. In *Proc. of DBWeb2002*, pages 97-104. 情報処理学会, 2002.
- [9] 村木 太一, 吉田 誠, 小林 隆志, 直井 聡, 横田 治夫. メタデータによる教育資料の統合における検索絞り込み指標の評価. Issn 1347-4413, DEWS2003, 5-c, 電子情報通信学会データ工学ワークショップ, 3 2003.
- [10] Haruo Yokota, Takashi Kobayashi, Taichi Muraki, and Satoshi Naoi. UPRISE: Unified Presentation Slide Retrieval by Impression Search Engine. *IEICE Transactions on Information and Systems*, E87-D(2), February 2004.
- [11] 小林 隆志, 村木 太一, 直井 聡, 横田 治夫. 統合プレゼンテーションコンテンツ蓄積検索システムの試作. 電子情報通信学会論文誌 *DI*. (採録決定, ショートバージョン: Proc. of DBWeb2003, pp.61-68, 11 2003).
- [12] 岡本 拓明, 小林 隆志, 横田 治夫. プレゼンテーション蓄積検索システムにおける適合度計算の改善. Issn 1347-4413, DEWS2004; 1-b-3, 電子情報通信学会データ工学ワークショップ, 3 2004.
- [13] O. Marques and B. Furht. *Content-Based Image and Video Retrieval*. Kluwer, 2000.
- [14] Y. Kambayashi, K. Katayama, Y. Kamiya, and O. Kagawa. Index generation and advanced search functions for multimedia presentation material. In *Proc. of ER97 Workshop on Conceptual Modeling in Multimedia Information Seeking*, 1997.
- [15] 勝山 裕, 小澤 憲秋, 武部 浩明, 小原 敦子, 直井 聡, 横田 治夫. 講義動画中のスライド同定を利用したレーザーポインタの同期表示. D-12-69, 電子情報通信学会 総合大会, 3 2003.
- [16] Yutaka Katsuyama, Noriaki Ozawa, Jun Sun, Hiroaki Takebe, Takashi Kobayashi, Haruo Yokota, and Satoshi Naoi. A New Solution for Extracting Laser Pointer Information from Lecture Videos. In *Proc. of E-learn2004*, pages 2713-2718, 10 2004.
- [17] D. A. Grossman and O. Frieder. *Information Retrieval Algorithm and Heuristics*. Kluwer, 1998.