

論文 / 著書情報  
Article / Book Information

論題(和文)	講義講演シーン検索におけるレーザーポインタ情報の活用法
Title(English)	Method of Exploiting Pointing Information for Presentation Scene Retrieval
著者(和文)	仲野 亘, 小林 隆志, 直井 聡, 横田 治夫
Authors(English)	Wataru Nakano, Takashi Kobayashi, Satoshi Naoi, Haruo Yokota
出典(和文)	電子情報通信学会和文論文誌 (D), J91-D, 3, 654-666
Citation(English)	IEICE TRANS.INF.& SYST, J91-D, 3, 654-666
発行日 / Pub. date	2008, 3
URL	<a href="http://search.ieice.org/">http://search.ieice.org/</a>
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 2008 Institute of Electronics, Information and Communication Engineers.

## 講義講演シーン検索におけるレーザーポインタ情報の活用法

仲野 亘<sup>†\*a)</sup> 小林 隆志<sup>††</sup> 直井 聡<sup>†††,††</sup> 横田 治夫<sup>††,†</sup>

Method of Exploiting Pointing Information for Presentation Scene Retrieval

Wataru NAKANO<sup>†\*a)</sup>, Takashi KOBAYASHI<sup>††</sup>, Satoshi NAOI<sup>†††,††</sup>,  
and Haruo YOKOTA<sup>††,†</sup>

あらまし 我々はこれまで、講義・講演における撮影動画と発表資料をメタデータにより疎結合した統合コンテンツとして蓄積し、その特性を利用した高度なシーン検索を提供する UPRISE (Unified Presentation slide Retrieval by Impression Search Engine) を提案してきた。更に講師が発話した音声情報を利用することで、シーン検索をより効果的に行う手法も提案してきた。本論文では、シーン検索精度を改善するために、講師が用いたレーザーポインタの情報を活用し、レーザーポインタの照射情報を統合する手法を提案する。基本的な統合方法に加え、実際の講義動画における各レーザーポインタ照射の解析を行うことで、提案手法の特性を分析し、それに基づいた改良手法の提案も行う。提案手法及びその改良手法を実際の講義コンテンツに適用して検索実験を行い、手法の有効性を確認する。

キーワード 情報統合, 情報検索, e-learning, レーザポインタ, 講演シーン検索

## 1. ま え が き

近年のネットワーク帯域や計算機性能の向上により、動画像の取扱が容易になったことから、文書や動画等の複数のメディアをコンテンツとして統合し、それらを蓄積、検索するシステムが数多く研究、及び提案されている [1]~[5]。

そのようなコンテンツは e-Learning や講演のアーカイブ化など、様々な用途に用いられており、それらに対する高度な検索の要求は高い。特に e-Learning 用のコンテンツに対しては、利用者が必要とするコンテンツを検索できるだけでなく、動画コンテンツ中での箇所から視聴を始めればよいかを効果的に発見できることが重要である。

そのような検索を実現するために、我々は教育コンテンツの統合、蓄積、及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である UPRISE (Unified Presentation Slide Retrieval by Impression Search Engine) を提案してきた [6]。

UPRISE では、動画ストリームを資料スライドの切替タイミングによってシーンという単位に分割し、各シーンとそこで使用された資料スライドを対応づけることでそれらを統合する。また、各シーンに対して、対応する資料スライドの文字・構造情報、シーンの時間長の情報などから検索用インデックスを作成することで、高度な検索を可能としている。従来用いられてきたスライド中の文書検索ではなく、スライドの切替タイミングによって分割したシーンを検索の単位とすることで、動画中で講師がバックトラックをしたり、巻戻りがあったりすることなどにより複数のシーンで同一のスライドが使用されている場合でも、それらを異なるシーンとして区別することができるという利点がある。

本論文では、UPRISE のシーン検索精度を向上させるために、講師が用いたレーザーポインタの照射情報に着目し、その情報を統合した、より高度な検索を実現する手法を提案する。

講義講演コンテンツを対象としたシステムにおいて、講師が用いたレーザーポインタに着目した研究としては、マルチメディアプレゼンテーションコンテンツの配信

<sup>†</sup> 東京工業大学大学院情報理工学専攻, 東京都  
Graduate School of Information Science and Engineering,  
Tokyo Institute of Technology, Tokyo, 152-8552 Japan

<sup>††</sup> 東京工業大学学術国際情報センター, 東京都  
Global Scientific Information & Computing Center, Tokyo  
Institute of Technology, Tokyo, 152-8552 Japan

<sup>†††</sup> (株)富士通研究所, 川崎市  
Fujitsu Laboratories Ltd., Kawasaki-shi, 211-8588 Japan

\* 現在 (株) 東芝

a) E-mail: wnakano@de.cs.titech.ac.jp

と検索手法についての報告 [7] や、レーザーポインタ指示情報を会議コンテンツの提示に利用する手法 [8] があるが、どちらもレーザーポインタ情報をコンテンツの視聴の補助にのみ利用している。

提案手法では、レーザーポインタが長時間照射されたスライド中の文字列について、その重要度を高くすることにより、レーザーポインタ照射情報をシーン検索に利用する。まず、撮影動画から画像認識によって抽出したレーザーポインタ情報を利用し、講師の手のぶれなどによるレーザーポインタの不正確さを考慮して、レーザーポインタ情報を従来の検索手法に統合する手法 [9] を提案する。

次に、撮影動画から抽出できるレーザーポインタ情報のうち、検索に考慮すべきではないレーザーポインタ情報の影響を排除するために、レーザーポインタ情報をフィルタリングする手法に関して説明する。

更に、我々がこれまでに提案してきた撮影動画や資料スライドから得られる 2 種類の情報を利用するフィルタリング手法 [10] の適用結果を分析し、実際の講義コンテンツ間の特性の差異や、実際に抽出したレーザーポインタ照射の分析を行った結果 [11] に基づき、レーザーポインタ情報のフィルタリング手法を改良する手法 [12] の提案を行う。

以下では、まず 2. において本研究の関連研究を述べる。次に、3. で UPRISE の概要を示し、4. で動画中のレーザーポインタ情報を検索に利用するための手法について述べる。5. では、レーザーポインタ情報をフィルタリングする手法について提案し、従来の適合度との統合を行う。また、6. では、実際の講義を用い、検索実験を行って提案手法の評価を行う。その後、7. においてそれぞれのレーザーポインタ照射を分析し、その結果に基づいて提案手法の問題点を解決する手法の提案を行う。この手法についても同様に実験を行い、その有効性を確認する。最後に 8. においてまとめと今後の課題を述べる。

## 2. 関連研究

動画や文書、音声ストリームなどの複数のメディアによるコンテンツを統合し、それらを蓄積、検索するシステムとしては、Authoring on the Fly [1], [13] や Informedia [2], eClass (Classroom 2000) [3], [14] などがある。また、遠隔講義のための商用システムとしては、MPMeister [15], P4Web [16], Internet Navigator [17] など多くのシステムが提供されているが、

コンテンツ作成支援や配信、蓄積のみを提供しており、各コンテンツの詳細なシーン検索を実現してはいない。

ビデオ中の検索を行う研究として、湯本ら [18] は、Web ページに対してビデオストリームの動的統合を行う Video-Augmented Web を提案している。この手法はビデオから効率良く部分映像を探すという側面をもっているが、部分文章に対して複数の部分動画を対応づけることについては考慮していないため、資料スライドと撮影動画を統合したコンテンツへの適用は困難であると考えられる。

オンデマンド講演システム LODEM [5] は、資料テキストの情報と音声認識による講義音声を用いて、パッケージと呼ばれる発話のまとまり単位での検索を行う手法を提供している。ユーザは教科書などの資料の部分文書をクエリとして入力することで、入力した内容に関連するビデオの場面を検索することができる。しかし、資料スライドのバックトラックや巻戻りへの対応や、音声認識における誤認識への対応は行っていない。本研究の手法では、音声認識結果のみを検索に用いるのではなく、他の様々なメタデータと合わせて総合的に各シーンの適合度を算出するため、音声認識の誤認識の影響を直接受けることはない。

講師が用いたレーザーポインタに着目した研究としては、マルチメディアプレゼンテーションコンテンツの配信と検索手法について述べた報告 [7] や、レーザーポインタ指示情報を会議コンテンツの提示に利用する手法 [8] があるが、いずれもレーザーポインタ情報をコンテンツの視聴の補助にのみ利用しており、コンテンツやシーンの検索に利用した研究はほとんど行われていない。

片山ら [19] は、遠隔講義システムにおける講義検索手法として、検索キーワードを含むスライド検索などのほかに、教師がマウスポインタを用いてスライドのある位置を指している部分を検索する手法を提案している。しかし、これらの検索手法はそれぞれが別個のものであり、ポインティングデバイスを用いた情報を利用して各シーンの重要度を総合的に算出するような検索は行っていない。

## 3. UPRISE

### 3.1 UPRISE の概要

UPRISE では、メタデータを用いてコンテンツを管理、統合する。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とその構造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も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。UPRISEのシステムの詳細についてはこれまでの報告[20]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。

スライドの切替によるシーン分割においては、バックトラックや、スライドの再使用などの理由により、同一のスライドが複数のシーンで使用されるということが発生する。例えば、図1のスライドAのように、複数の概念を説明したスライドがあり、その後にそれぞれの概念についての説明や具体例、応用などのスライドがある場合を考える。このとき、図1のような順番でスライドAを2回使用するということが起こる。そのため、従来の文書検索技術をスライド文書内容に対して適用するだけでは、これらのシーンを異なるシーンとして順位付けすることができない。

この問題を解決するため、また、各シーンの内容を考慮し、より適切に順位付けを行うために、UPRISEでは、スライド文書内容以外の様々な情報を用い、検索キーワードに対し、総合的に各シーンの適合度を算出する。これにより、検索キーワードに対してより適切なシーン検索を実現する。

### 3.2 基本的な適合度

UPRISEの検索機能は、検索キーワードに対する適合度をシーンごとに算出し、全シーンの順位付けを行った後、上位のシーンから表示する。この適合度の

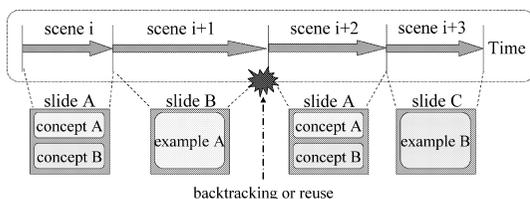


図1 シーン検索における問題

Fig.1 A problem in scene retrievals.

うち、最も基本的なものは適合度  $I_c$  である[6]。 $I_c$ はスライドの文書構造、シーンの時間の長さ、前後シーンの文脈の3種類の情報をもとに算出される。

まず、スライドの文書構造を考慮した適合度  $I_p$  を以下のように定義する。

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

ここで、 $s$ はシーン、 $k$ はキーワード、 $l$ は行、 $L(s)$ はシーン  $s$ のスライドに含まれる文字列の行数であり、 $P(s, l)$ 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。更に  $P(s, l)$ 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することにより、キーワードの出現回数だけでなく出現位置も考慮することができる。

次に、シーンの時間情報を付加した適合度  $I_d$  を以下のように定義する。

$$I_d(s, k, \theta) = T(s)^\theta \cdot I_p(s, k)$$

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ータである。これによって、シーンの長さが長いほどそのシーンは適合度が高くなり、検索されやすくなる。

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。

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

ここで、 $\delta$ は考慮する前後シーンの範囲を定めるパラメータであり、 $E(x, \varepsilon_1, \varepsilon_2)$ は前後関係の強弱を定める関数である。

なお、この  $I_c$  のパラメータ群  $(s, k, \theta, \delta, \varepsilon_1, \varepsilon_2)$  はUPRISEにおける適合度関数の基本パラメータ群であるため、本論文では以降これを  $\Phi$  として簡略化表現する。

### 3.3 音声情報の統合手法

講義における講師の発話内容は、スライド中の文字列情報と同様に直接的にそのシーンの内容を表している。そこで我々は音声情報を利用した適合度の提案を行ってきた[21]。

あるシーン  $s$  中でキーワード  $k$  が発話された回数を  $skc(s, k)$  (Spoken Keyword Count) とし、この  $skc(s, k)$  を適合度  $I_c$  と統合することにより、音声情報を利用した適合度を提案した。更に、実験によって、同一スライドを用いたシーンにおいても、発話内容によってそのシーンの差別化を行うことができることを示した。

#### 4. レーザポイント情報の統合

##### 4.1 レーザポイント情報の抽出

講義講演等で講師がレーザーポイントを使用する場合、講師の手のぶれなどの理由により、ある単語や行に対して正確に当て続けることが容易でない。更に、たとえ目的の単語や行に対して正確にレーザーポイントを当て続けなくても、レーザーポイント照射の意図は聴衆には伝わるため、レーザーポイントの照射にはそれほど正確さは必要とされない。また、実際のレーザーポイント照射の事例においては、目的の行を囲むように円を描いて、中心付近の内容を強調する当て方などもよく行われている。

本研究では、レーザーポイントの照射が必ずしも目標の単語が最も近くなるような座標に対して行われるわけではないことを考慮し、レーザーポイントの照射対象を行単位で扱い、行に対するレーザーポイント照射の回数と時間という観点でレーザーポイント情報を抽出する方法を提案する。

提案手法では、まず、撮影動画から高精度でレーザーポイントの光点座標を抽出する手法 [22] を用いて、1秒ごとにスライド中の光点座標を抽出する。更に、常に光点に最も近い行が必ずしも講師の目的とする行ではないことを考慮し、スライド上で最も近い行の文字列の他に、その近傍の数行を次候補として取得しておく。

連続した光点のうち、同一の行文字列が最も近い行として関連づけられた光点の集合を1回のレーザーポイント照射として扱う。この1回のレーザーポイント照射をシーン内のある区間として扱うために、サブシーンと定義する。複数のシーンにまたがって同一のキーワードを強調することはないため、1回の照射としての統合は同一シーンの中でのみ行う。

このように抽出したレーザーポイント情報の数値化を行う。このとき、照射の不正確さに対応するため、レーザーポイントの照射回数を、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

$H(l, q)$  として数値化する。 $H(l, q)$  はサブシーン  $q$  のレーザーポイントが、行  $l$  に照射された回数の期待値であり、すべての行の  $H(l, q)$  を合計すると1となる。

このレーザーポイントごとの照射回数期待値に対し、各レーザーポイントが当たっていた時間を掛け合わせることで、レーザーポイントの照射時間の期待値が得られる。この照射時間の期待値をシーンごとに合計したものを、 $phd(s, k)$  (Pointer Hit Duration) とする。シーン  $s$ 、キーワード  $k$  における  $phd(s, k)$  の式を以下のように定義する。

$$phd(s, k) = \sum_{q_i \in s} \sum_{l=1}^{L(s)} H(l, q_i) \cdot T(q_i)$$

ここで、 $L(s)$  はスライド中の行数、 $T(q_i)$  はサブシーン  $q_i$  の時間を表す。すなわち、サブシーン  $q_i$  に対応するレーザーポイントの照射時間を表す。

このように、レーザーポイント情報をキーワードへの照射時間期待値で扱い、期待値が高ければ高いほどそのキーワードを目標としていた可能性が高いという扱い方をすることで、レーザーポイント照射が常に目標に最も近い位置へ行われるわけではないという問題を解決することができる。例えば、目的とする行を囲むように円を描くようなレーザーポイント照射に対しても、期待値の総和をとることで、目的とする行に対する期待値が周囲の他の行と比較してより多く与えられることになる。

##### 4.2 レーザポイント情報と $I_c$ の統合

本研究では、4.1 で得られた  $phd$  を検索に利用できるように、適合度  $I_c$  と統合する。統合の際、 $I_c$  のどの箇所と統合するかにより、いくつかの候補が存在するが、予備実験の結果、シーンの時間情報に統合する方法が有効であることが分かったため、統合方法として以下の適合度  $I_{c[d+phd]}$  を提案する。

$I_{c[d+phd]}$  では、以下の式で定義されるように、シーンごとの時間情報である  $T(s)$  に対し、レーザーポイントの時間の期待値である  $phd(s, k)$  を足し合わせ、レーザー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を行う。

$$I_{c[d+phd]}(\Phi, \omega_d) = \sum_{\gamma=s-\delta}^{s+\delta} E(\gamma - s, \varepsilon_1, \varepsilon_2) \cdot \{T(\gamma) + \omega_d \cdot phd(\gamma, k)\}^{\theta} \cdot I_p(\gamma, k)$$

ここで、 $\omega_d$  は  $phd$  の適合度計算における影響度合を

調節するパラメータである．例えば， $\omega_d = 10$  の場合，レーザポイントが対象キーワードに 1 秒間当たるとはそのシーンが仮想的に 10 秒伸びたとみなされる．

3.2 において述べたように，適合度  $I_c$  はシーンの時間が長いほど値が大きくなる．そのため，本手法によって，レーザポイント照射が多くあったシーンではその適合度が増加することになる．

## 5. レーザポイント情報のフィルタリング

講義や講演で行われるレーザポイント照射は，あるキーワードやトピックを強調する目的以外にも，様々な目的で行われる．例えば，図や表を説明する際に補助的に用いるものや，複数概念間の関連を示す軌跡を描くもの，または照射意図がはっきりしないあいまいなものなどがある．

そこで，このような様々な照射の中から，検索キーワードに対して関連する照射だけをシーン検索に利用するために，我々はレーザポイント情報をフィルタリングする手法の提案を行ってきた [10] ．

しかしながら，これまで提案してきたフィルタリング手法では，レーザポイント情報はサブシーンという単位で扱われているが，音声情報はシーンごとに利用されており，情報統合が異なった粒度において行われていた．本研究では，統合する粒度を統一し，二つの情報を高精度に対応させることで，よりいっそうの効果が望めると考え，これまでのフィルタリング手法を改良する．

以下では，まず，我々がこれまで提案してきたフィルタリング手法を簡単に説明し，その後，従来手法を改良した手法を提案する．

### 5.1 従来のフィルタリング方法

[10] では，2 種類の条件を用いてレーザポイント照射のフィルタリングを提案してきた．まず，スライドテキスト中でのキーワードの出現の有無に着目し，検索クエリが複数キーワードの場合に，そのすべてがスライド中に出現しない場合はそのシーンでの検索キーワードへのレーザポイント照射を無視するという手法を提案した．このフィルタリング手法を用いて計算した  $phd(s, k)$  を  $phd_{/p}(s, k)$  とし，以下の式で定義している．

$$phd_{/p}(s, k) = \begin{cases} phd(s, k) & \prod_{k \in K} I_p(s, k) \neq 0 \\ 0 & \prod_{k \in K} I_p(s, k) = 0 \end{cases}$$

ここで， $K$  は検索語を形態素解析して得られた単語の

集合である．

次に，講師の発話した音声でのキーワードの出現の有無に着目し，検索キーワードがそのシーンで 1 度も発話されていない場合はそのシーンでの検索キーワードへのレーザポイント照射を無視するという手法を提案した．このフィルタリング手法を用いて計算した  $phd(s, k)$  を  $phd_{/s}(s, k)$  とし，以下の式で定義している．

$$phd_{/s}(s, k) = \begin{cases} phd(s, k) & skc(s, k) \neq 0 \\ 0 & skc(s, k) = 0 \end{cases}$$

以下，本論文では，これらのフィルタリング手法をそれぞれ P-フィルタリング，及び S-フィルタリングと呼称する．

なお，仲野ら [10] は，P-フィルタリングと S-フィルタリングの両方を同時に用いて計算した  $phd(s, k)$  を  $phd_{/ps}(s, k)$  とし，これらのフィルタリング手法を適合度  $I_{c[d+phd]}$  に統合して評価した結果， $phd_{/ps}(s, k)$  を統合した適合度  $I_{c[d+phd_{/ps}, p+skc/p]}$  が最も優れているという結果を得ている．

### 5.2 NS-フィルタリング

5.1 で説明した S-フィルタリング手法では，各照射をフィルタリングするかどうかはそのシーンの発話中におけるキーワードの有無により決定されていた．しかしこの方法では，例えば，レーザポイント照射の時刻とキーワードの発話の時刻が数分離れていたとしても，それらが同じシーンに属するものであれば，音声中出現条件を満たし，有効な照射としてシーン検索に利用されてしまうことになる．これは，レーザポイント情報をサブシーンという粒度で扱うことに対し，音声情報はシーンという粒度で扱っていたことが原因である．

そこで，本論文では，音声情報をサブシーンの粒度で扱うことで，二つの情報の粒度を統一し，より高精度に対応させることを考える．そのために，レーザポイント照射中，及びその前後  $r_1, r_2$  秒以内においてキーワードが発話されていない場合，その照射を無視する，というフィルタリング手法を提案する．この手法により，実際に発話と同時に行われたレーザポイント照射と，それ以外の照射を区別することができ，前者のレーザポイント照射のみをシーン検索に利用することができる．図 2 はこの手法の処理を示した図である．

このフィルタリング手法を NS-フィルタリングと

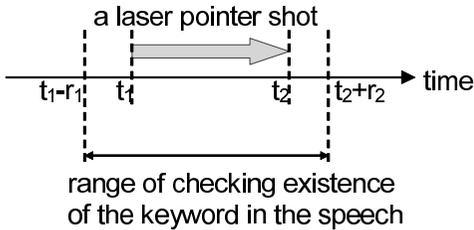


図 2 改良した S-フィルタリング手法  
Fig. 2 Improved S-filtering.

し、NS-フィルタリングを用いて計算した  $phd(s, k)$  を、 $phd_{/ns}(s, k)$  とする。  $phd_{/ns}(s, k)$  を以下の式で定義する。

$$phd_{/ns}(s, k) = \sum_{q_i \in s} \sum_{l=1}^{L(s)} H(l, q_i) \cdot T(q_i) \cdot exist(q_i, k, r_1, r_2)$$

ただし、 $exist(q_i, k, r_1, r_2)$  はサブシーン  $q_i$  中、及びその前後  $r_1, r_2$  秒においてキーワード  $k$  が発話されていれば 1、されていない場合は 0 の値をとる関数である。

また、NS-フィルタリングと同時に、5.1 で述べた P-フィルタリングを用いた  $phd(s, k)$  を、 $phd_{/pns}(s, k)$  とする。  $phd_{/pns}(s, k)$  は、 $K$  を検索語を形態素解析して得られた単語の集合とすると、以下の式で定義される。

$$phd_{/pns}(s, k) = \begin{cases} phd_{/ns}(s, k) & \prod_{k \in K} I_p(s, k) \neq 0 \\ 0 & \prod_{k \in K} I_p(s, k) = 0 \end{cases}$$

このように定義した  $phd_{/ns}$ 、 $phd_{/pns}$  をそれぞれ適合度  $I_{c[d+phd]}$  に統合する。更に、5.1 で行ったように、音声情報  $skc$  による適合度の加算分も考慮する。この適合度をそれぞれ  $I_{c[d+phd_{/ns}, p+skc/p]}$ 、 $I_{c[d+phd_{/pns}, p+skc/p]}$  とする。このように統合することで、二つの情報の粒度を考慮した統合が可能となる。

## 6. 評価実験

5.2 で提案した NS-フィルタリング手法の評価を行うため、評価実験を行った。以下では、まず 6.1 で本実験に用いたデータの詳細について説明し、次に、6.2 で本実験における各パラメータの設定や評価方法について述べ、その後、6.3 において実験結果を述べ、その考察を行う。

### 6.1 実験に用いたデータ

実際に大学において行われたデータベースについて

の講義（全 11 回）と、計算機アーキテクチャについての講義（全 12 回）の動画と資料を UPRISE に格納し、スライド同期情報やレーザーポインタ情報などのメタデータ抽出を行って講義コンテンツを作成した。講義は 2 種類ともに 1 回が 90 分であり、資料スライドの枚数はデータベースが 260 枚、計算機アーキテクチャが 282 枚である。1 回の講義の平均シーン数はデータベースの講義が 46.4 シーン、計算機アーキテクチャの講義が 44.2 シーンであり、2 講義の総計シーン数は 1040 シーンである。なお、撮影時の録音に問題があり、部分的に音声録音できていなかった講義が存在したため、その回は除外している。

ここで、作成した講義コンテンツにおけるレーザーポインタの使用状況について述べる。2 種類の講義はどちらも、すべての回でレーザーポインタが用いられている。1 回の講義において、レーザーポインタが使用されたシーン数が全シーンに占める割合の平均は、データベースの講義が 81.0%、計算機アーキテクチャの講義が 86.9% である。また、レーザーポインタが使用されたシーンにおいて、1 回のシーンにおける平均使用回数は 10.0 回、照射時間がシーン時間に占める割合の平均は 19.1% である。

講義の音声情報は、連続音声認識ソフトウェア Julius<sup>(注1)</sup> [23] を用い、言語モデルと音響モデルとして、山崎らが作成したもの [24] のうち、話者適応を行っていないものを用いた。

また、単語辞書に登録されていない用語は音声認識の結果に出現しないため、山崎らが作成した辞書 [24] に、資料スライド中から辞書に含まれていない単語を追加したものを講義ごとに作成し、音声認識に使用した。データベースの講義用の辞書には、資料スライドから抽出した 1099 語のうち、辞書に含まれていない名詞 123 語を、計算機アーキテクチャの講義には、同様に資料スライドから抽出した 1109 語のうち、辞書に含まれていない名詞 176 語を、それぞれ追加した。なお、本実験では簡単化のため、英語表記の専門語などへの読み付与は行わず、検索クエリにも含めていない。

音声認識用辞書と資料スライド間の表記揺れや、音声認識と UPRISE のコンテンツ登録時に異なる形態素解析エンジンを用いていることによる影響を緩和するため、以下の処理を行った。

- 英語表記された検索キーワードを除外

(注1): <http://julius.sourceforge.jp/>

- 全角英数字を半角に置換
- 検索キーワードとして用いた専門語の表記を資料スライドのものに統一

## 6.2 実験設定

上述の講義コンテンツ群に対し、以下の設定のもとで実験を行った。

- 基本となるパラメータ  $\Phi$  は  $\theta = 0.4, \delta = 4, \varepsilon_1 = 5.0, \varepsilon_2 = 0.5$  に固定した。
- レーザポイントの光点に対し五つの候補行を取得し、照射回数期待値  $H(l, q)$  を第 1 候補から順に 0.4, 0.3, 0.15, 0.10, 0.05 という値に固定した。
- 計算機アーキテクチャ関連の検索クエリ 59 種類と、データベース関連の検索クエリ 33 種類の、合計 92 種類の検索クエリを用いて検索実験を行った。なお、検索クエリは資料スライド中の専門語などを、形態素解析した複数のキーワードから構成される。実験に使用したすべての検索クエリを表 3 に示す。
- キーワードに対して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。

なお、正解シーンでは、94.6%のシーンにおいてレーザーポイントが使用されている。正解シーン中のレーザーポイントの平均使用回数は 13.4 回、照射時間がシーン時間に占める割合の平均は 16.3%である。

評価に際しては、正解シーンを各キーワードに対して一つとしていることから、平均逆数順位 (Mean reciprocal rank : MRR) を用いた [25]。MRR は質問応答システムの評価に用いられることが多く、質問ごとに最初に出現した正解の順位の逆数を求め、それらを全質問にわたって平均することで定義される。

本実験の MRR は、 $N$  を検索回数とすると以下の式で求めることができる。

$$MRR = \frac{1}{N} \sum_{i=1}^N \frac{1}{i \text{ 番目の検索における正解の検索順位}}$$

## 6.3 NS-フィルタリング手法の評価

5.2 で提案した NS-フィルタリング手法の効果を検証するため、実際の講義のコンテンツを 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った。

まず、提案手法である  $I_{c[d+phd/p_{ns}, p+skc/p]}$ 、 $I_{c[d+phd/p_{ps}, p+skc/p]}$  の適合度と、従来のフィルタ

リング手法を用いた適合度  $I_{c[d+phd/p_s, p+skc/p]}$ 、 $I_{c[d+phd/p_{ps}, p+skc/p]}$ 、及びレーザーポイント情報と音声情報を考慮せず、スライド構造、シーン時間長、シーンの前後関係のみから算出する適合度  $I_c$  の比較を行った。なお、NS-フィルタリングのパラメータ  $r_1, r_2$  の値は  $\{r_1, r_2\} = \{5, 5\}$  とした。

図 3 は、2 講義の検索キーワードを用いた検索による、各適合度の MRR の変化を示したグラフである。なお、グラフでは適合度名の表記を簡略化している。まず、従来手法で S-フィルタリングを適用した  $I_{c[d+phd/p_s, p+skc/p]}$  を TYPE1、S-フィルタリングと P-フィルタリングを適用した  $I_{c[d+phd/p_{ps}, p+skc/p]}$  を TYPE2 と表記する。また、提案手法で NS-フィルタリングを適用した  $I_{c[d+phd/p_{ns}, p+skc/p]}$  を TYPE3、P-フィルタリングと NS-フィルタリングを適用した  $I_{c[d+phd/p_{ps}, p+skc/p]}$  を TYPE4 とそれぞれ表記する。グラフより、本節の提案手法である NS-フィルタリングを適用した適合度 TYPE3, TYPE4 は、パラメータ  $\omega_d$  の値によっては従来のフィルタリングを行った適合度よりも良い結果を得るが、全体として従来のフィルタリング手法 TYPE1, TYPE2 とほぼ変わらない検索精度であるということが分かる。また、S-フィルタリング、NS-フィルタリングどちらにおいても、P-フィルタリングと組み合わせる方が MRR が向上することが分かる。

次に、適合度  $I_{c[d+phd/p_{ns}, p+skc/p]}$  (TYPE4) において、NS-フィルタリングのパラメータである  $r_1, r_2$  の値を  $\{r_1, r_2\} = \{10, 0\}$  に変更し、再度検索実験を行った。図 4 は、 $\{r_1, r_2\} = \{5, 5\}, \{10, 0\}, \{0, 10\}$  の 3 種類の設定による MRR の変化を示したグラフである。

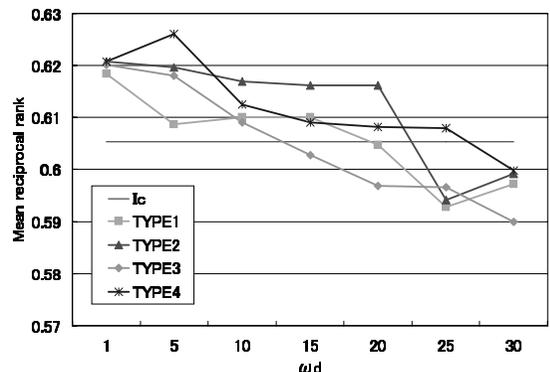


図 3  $\omega_d$  を変化させたときの適合度ごとの MRR の推移  
Fig. 3 MRRs for the two lectures varying  $\omega_d$ .

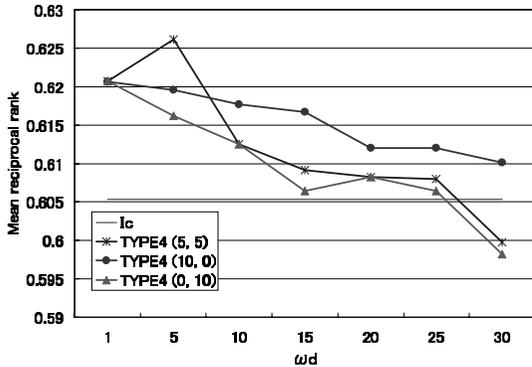


図 4  $I_{c[d+phd/pns,p+skc/p]}$  における  $r_1, r_2$  の MRR への影響

Fig.4 Effect of  $r_1, r_2$  in  $I_{c[d+phd/pns,p+skc/p]}$ .

図 4 より,  $\{r_1, r_2\} = \{0, 10\}$  では他の 2 種類の設定よりも MRR が低下している。 $\{r_1, r_2\} = \{0, 10\}$  という設定は、レーザーポインタ照射と同時に、照射後の 10 秒間に検索キーワードが発話されているような照射のみをシーン検索に利用する設定である。このことから、今回の実験に用いた講義においては、キーワードが発話の前に行われたレーザーポインタ照射よりも、発話の後に行われたレーザーポインタ照射の方が、より多く検索キーワードに関連している照射であったということが分かる。

### 7. フィルタリング結果の解析

6.3 で示した実験結果では、提案手法である NS-フィルタリングの十分な効果を確認することができなかった。そこで、S-フィルタリングと NS-フィルタリングによって実際のレーザーポインタ照射情報がどのように除去されているのかを分析することで、NS-フィルタリングの特性を調査する。

以下では、まず、レーザーポインタ照射情報をその意図する行を分析することで分類し、どれだけの照射が検索キーワードに関連しているのかを調査する。その後、その分析結果を元に、NS-フィルタリングを改良する手法を提案する。

#### 7.1 レーザーポインタ照射情報の分類

まず、P-フィルタリングと S-フィルタリングを同時に用いた  $phd/ps$  において、二つのフィルタリングで除去されずに  $phd/ps$  の計算に利用されたレーザーポインタ照射情報の分類を行い、どれだけのレーザーポインタ照射が検索キーワードに関連しているのかを調査し

表 1 レーザーポインタ照射情報の分類

Table 1 Classification of laser pointer information.

DB	ps	pns (5,5)	pns (10,0)
	51	16	17
	186	9	13
x	97	6	8
x x	30	2	1
	17	1	4

ARCH	ps	pns (5,5)	pns (10,0)
	80	48	50
	269	58	51
x	206	50	39
x x	16	1	0
	60	19	22

た。分類では、6.3 の実験に用いた検索クエリのうち、各講義 10 件ずつの合計 20 クエリに対し、フィルタリングで除去されなかった照射すべてを講義動画で確認し、講師が照射しようとしている行を発話や軌跡から判断した。その結果を用い、各レーザーポインタ照射情報を以下の 5 種類に分類した。

検索キーワードを含む行に当てようとした照射  
 検索キーワードを含まない行に当てようとしているが、検索キーワードと関連があると判断した照射  
 x 検索キーワードとは関連がないと判断した照射  
 x x 実際の動画では照射は行われていない、誤認識による照射情報

上記 4 種類の分類をするのに判断が難しい照射  
 検索キーワードと関連がないと判断した照射 (x) には、複数のトピックをもつようなスライドにおいて、検索キーワードとは別のトピックに属する行を意図した照射や、テキストがアニメーションによって順次出現するようなスライドにおいて、まだキーワードに関する行が出現していない状態で行われた照射などが含まれる。

また、判断が難しいとした照射 ( ) には、あいまいに行われたため、動画からは意図する行が分からない照射や、関連の有無の判断が難しい行や図への照射などが含まれる。

更に、S-フィルタリングの代わりに、本節の提案手法である NS-フィルタリングを行うとこれらの照射情報がどれだけ除去されるのかを調査した。なお、NS-フィルタリングのパラメータ  $r_1, r_2$  の値として、 $\{r_1, r_2\} = \{5, 5\}, \{10, 0\}$  の 2 種類の設定について調査を行った。

表 1 は各講義ごとの P-フィルタリング+ S-フィルタリングで残ったレーザーポインタ照射の分類結果、及

び、それらが P-フィルタリング+ NS-フィルタリングではいくつ残ったかを示す表である。

まず、P-フィルタリングと S-フィルタリングを行っただけでは、検索キーワードと関連のない多くの照射情報が除去されないままであることが分かる。また、NS-フィルタリングは S-フィルタリングと比較して、 $\times$ 、 $\times \times$ に分類される検索キーワードと関連のない照射情報の大部分を除去できているが、同時に  $\times$ 及び  $\times \times$ に分類される検索キーワードに関連した照射情報も多くのものが除去してしまっていることが分かる。特に  $\times \times$ については、NS-フィルタリングにすることで非常に多くの照射が除去されてしまっている。

この、NS-フィルタリングが非常に多くのキーワードに関連する照射を除去しているという問題の原因を調べるために、撮影動画を実際に視聴した。その結果、NS-フィルタリングの結果において、 $\times$ に比べて  $\times \times$ の除去率が高い理由として、レーザーポインタ照射は対象となる行の内容、またはそれに類似する内容を発話しながら行うことが多いことが分かった。これは、 $\times$ に分類される照射では、対象となる行の文章を読みながら照射を行った場合、レーザーポインタ照射の近傍の音声にキーワードが出現することになるが、 $\times \times$ に分類される照射では、対象となる行に検索キーワードが含まれていないために近傍の音声にキーワードが出現しないことが多いと考える。

#### 7.1.1 NS-フィルタリングの改良

7.1の分析により、NS-フィルタリングはレーザーポインタ照射情報を過剰に除去していたことが分かった。そこで、検索キーワードに関連する照射をできるだけ除去しないために、特に  $\times \times$ に分類される照射について考える。

に分類された照射の中で多く見られた例として、各照射を分類する過程において以下の3種類の事例を確認した。

- 各照射において、講師が意図した行はインデント構造において下位であり、その上位レベルの行にはキーワードが含まれる。意図した行は、キーワードが含まれる行を補足説明している。
- 各照射において、講師が意図した行と同じ段落にキーワードを含む行があり、段落全体の主題がそのキーワードを含む行になっている。
- スライドタイトルには検索キーワードが含まれるが、本文にはキーワードは含まれない。ただし、本文全体がタイトルに含まれている検索キーワードにつ

いて説明を行っている。

これらのうち、スライドタイトルにのみ検索キーワードが含まれる事例に着目し、タイトルに検索キーワードが含まれないシーンは P-フィルタリング + NS-フィルタリングを行い、検索キーワードが含まれるシーンでは例外として P-フィルタリング + S-フィルタリングを行うという手法を提案する。この、例外処理を含めたフィルタリング手法を、S/NS-フィルタリングとする。

表2は、例外処理を含めた手法である S/NS-フィルタリングを行った後の照射情報の分類結果を表1に追加したものである。スライドタイトルの例外処理を含めることによって、 $\times \times$ に分類されたレーザーポインタ照射を多く残すことに成功していることが分かる。特に計算機アーキテクチャの講義においては、NS-フィルタリングによって約80%の  $\times \times$ に分類される照射が除去されてしまっていたのが、S/NS-フィルタリングでは約25%の除去に抑えることができています。

#### 7.2 再実験とその結果

7.1.1において提案した、スライドタイトルにおける検索キーワードの出現の有無に基づいた例外処理がシーン検索に与える効果を検証するため、再度検索実験を行った。なお、実験に用いたデータ及び、各パラメータの設定は6.3で行った実験と同一のものを用いた。

図5は適合度  $I_{c[d+phd/pns,p+skc/p]}(\text{TYPE4})$  に対してスライドタイトルによる例外処理を適用したときの MRR の推移を示したグラフである。なお、グラフ中では例外処理を適用した適合度を TYPE5 と表記する。グラフより、適合度  $I_{c[d+phd/pns,p+skc/p]}$  にお

表2 レーザーポインタ照射情報の分類 (例外処理含む)  
Table 2 Classification of laser pointer information.  
(including the exception processing)

DB	ps	pns (5,5)	pns (10,0)	pns (5,5) +title	pns (10,0) +title
	51	16	17	22	23
	186	9	13	153	157
$\times$	97	6	8	12	14
$\times \times$	30	2	1	19	19
	17	1	4	1	4

ARCH	ps	pns (5,5)	pns (10,0)	pns (5,5) +title	pns (10,0) +title
	80	48	50	54	55
	269	58	51	202	204
$\times$	206	50	39	58	47
$\times \times$	16	1	0	5	5
	60	19	22	32	39

表 3 検索クエリー一覧  
Table 3 List of search queries.

アボート	スキーマ	内部, スキーマ	キー
結合, 演算	コミット	トランザクション	時刻, 印
射影	選択	割り算	集合, 和
集合, 積	障害	ハッシュ	拡張, ハッシュ
ハッシュ, 結合	属性	ダブル	直積
デッドロック	問い合わせ, 木	ログ	関係, データ, モデル
ロック	チェックポイント	定義, 域	関係, 代数
ページ	レコード	索引	原子, 性
一貫, 性	コンピュータ, 性能	スルーブット	性能, 向上, 率
性能, 向上, 比	アムダール, 法則	合成, ベンチマーク	パイプライン
時, 空間, ダイアグラム	命令, パイプライン	インター, ロック	ハザード
構造, ハザード	ハーバード, アーキテクチャ	インターリーブ	制御, ハザード
分岐, 予測	静的, 分岐, 予測	動的, 分岐, 予測	分岐, 履歴
分岐, 履歴, テーブル	プリフェッチ, バッファ	遅延, 分岐	データ, ハザード
フォーワーディング	内部, フォワーディング	スーパー, パイプライン	スコアボード
共通, データ, パス	乱, 命令, 実行	レジスタ, リ, ネーミング	リ, オーダ, バッファ
算術, パイプライン	パイプライン, 予約, 表	衝突, ベクトル	最小, 平均, レイテンシ
縮小, 化, 状態, 図	ベクトル, プロセッサ	ロード, ストア, パイプライン	可変, 遅延
相互, 結合, 網	静的, 結合, 網	動的, 結合, 網	クロスバー, 結合
多段, 結合, 網	ホップ	直径	ルーティング
ブロッキング	二, 重, 反射, 網	共有, メモリ, 型, マルチプロセッサ	分散, メモリ, 型, マルチプロセッサ
ディジーチェーン	イネーブル, ディジーチェーン	許可, ディジーチェーン	要求, ディジーチェーン
キャッシュ, 一貫, 性	パス, スヌープ	データ, フロー	講義, 配信

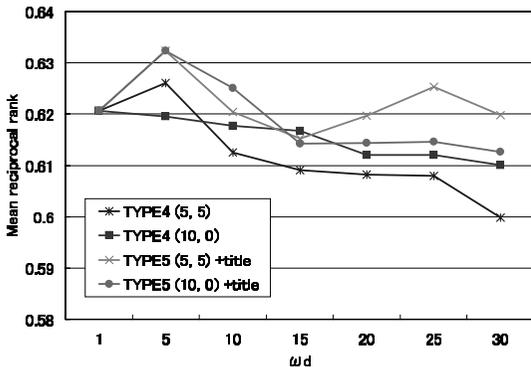


図 5 例外処理が MRR に与える影響  
Fig. 5 Effect of the exception processing.

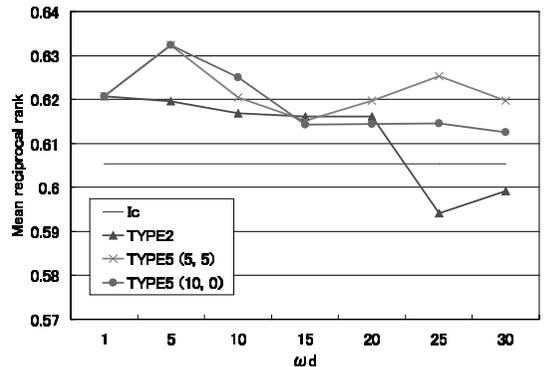


図 6 従来の適合度との比較  
Fig. 6 Comparison for original weighting schemata.

いて、スライドタイトルによる例外処理を適用することで全体的に MRR が向上している。このことから、7.1.1 の分析結果で示した、S/NS-フィルタリングによって多くの に分類されるレーザポイント照射が残されるということが、実際のシーン検索に対しても良い影響を与えていることが分かる。

また、図 6 は、S/NS-フィルタリングを適用した適合度  $I_{c[d+phd/pn.s,p+skc/p]}$  (TYPE5) と、従来の適合度である  $I_c$  や  $I_{c[d+phd/p.s,p+skc/p]}$  (TYPE2) を比較したグラフである。図 6 より、従来の適合度と比較しても、MRR が改善している。これにより、本節の提案手法である NS-フィルタリングは、その特性を補う処理を

行うことによって、シーン検索精度の改善に有益な効果を与えることが分かる。

したがって、検索キーワードに関連するレーザポイント照射の情報のみを選別し、それ以外の照射情報を排除する精度を向上させていくことが、実際の講義コンテンツにおけるシーン検索の検索精度を向上させることにつながったと考える。

## 8. む す び

### 8.1 ま と め

本論文では、講義や講演を統合コンテンツとして蓄積、検索、配信するシステム UPRISE において、講

師が用いるレーザーポインタの照射情報を利用することで、シーン検索機能の精度を向上させるための手法を提案した。

UPRISE のシーン検索ではスライド文書への文書検索では不十分であるため、撮影動画や資料スライドから得られる様々な情報を利用した検索を実現しなければならないことを述べた。

本論文の提案として、まず、講師が用いたレーザーポインタ情報の抽出方法を提案し、適合度計算手法に統合した。講師の手のぶれなど、レーザーポインタ照射特有の不正確さに対応するため、1回のレーザーポインタ照射に対し複数の行文字列を対応させることで、レーザーポインタ照射を照射期待値として扱い、シーン検索に利用した。

抽出したレーザーポインタ情報のうち、検索キーワードに関連する、キーワードを強調する目的で行われたレーザーポインタ照射の情報だけを利用し、その他の照射の影響を緩和するために、レーザーポインタ情報のフィルタリングを行う手法を説明した。フィルタリング手法として、我々がこれまでに提案してきた複数キーワードの際のスライド中キーワード出現条件によってフィルタリングを行う P-フィルタリングと、音声中キーワード出現条件によってフィルタリングを行う S-フィルタリングの 2 種類の手法を説明した。

更に、S-フィルタリングにおけるレーザーポインタ情報と音声情報の粒度を統一することで、S-フィルタリングを改良した NS-フィルタリングを提案した。NS-フィルタリングは各レーザーポインタ照射の近傍でキーワードの発話が行われない場合はそのレーザーポインタ照射を考慮しないという処理を行う。

また、S-フィルタリングと NS-フィルタリングの結果を分析し、各レーザーポインタ照射の分類を行うことで、二つのフィルタリング手法の特性を解析した。その解析に基づき、NS-フィルタリングに対しその特性を補うための例外処理を加えた、S/NS-フィルタリングを提案した。

これらの提案手法の評価を、実際の講義コンテンツに適用することで行った。検索キーワードに関連したレーザーポインタ照射と、それ以外の照射を選別し、キーワードに関連するレーザーポインタ照射のみをシーン検索に利用することが、より適切にシーン検索の精度向上に結び付くということを確認した。したがって、レーザーポインタ情報の選別における精度を改善していくことが、シーン検索精度のよりいっそうの向上をも

たらすということが分かった。更に、音声情報とレーザーポインタ情報を本提案手法のように統合し、同時に、利用することで、効果的なシーン検索が実現できることを示した。

## 8.2 今後の課題

本研究の今後の課題として、まず、レーザーポインタ照射情報の分析を本実験に用いたすべての検索クエリで行うことが必要である。更に、より高精度に各レーザーポインタが検索キーワードに関連しているか否かを選別する手法を考案することが必要である。特に、7.1 においてと分類されたレーザーポインタ照射を除去されないようにすることが重要であり、7.1.1 で示したようなタイプの特徴を利用することで良い結果が得られると考える。

また、NS-フィルタリングのパラメータである  $\{r_1, r_2\}$  の適切な値は、講義の種類、特に異なる講師の講義において変化する可能性がある。そのため、異なる講師の講義コンテンツを増やして実験を行うことは非常に有効であると考えられる。

その他の課題としては、音声認識精度がレーザーポインタ情報のフィルタリングに与える影響について調べる必要があると考える。音声認識精度が 100% になった条件における実験として、講義音声を手により書き起こしたテキストを用いての実験は有効である。

更に、提案した適合度では音声中、及びスライド文字列中での複数の検索キーワードにおける特定性を考慮していない。これらの特定性をシーン検索において考慮することの有効性はこれまでの報告 [21] において既に示しており、今回の提案手法においてもこれらの特定性を考慮することで更に精度が向上すると考える。

謝辞 本研究で用いた Julius と音響、言語モデルの使用にあたり御協力頂いた、東京工業大学大学院情報理工学研究科計算工学専攻の古井貞照教授、篠田浩一准教授、岩野公司助教、山崎裕紀氏に感謝致します。また、本研究で用いたスライド同定技術及びレーザーポインタ照射の情報抽出技術について御協力頂いた (株) 富士通研究所の勝山裕氏、小澤憲秋氏に感謝致します。

なお、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, 18049026)、独立行政法人科学技術振興機構 CREST、及び東京工業大学 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] R. Müller and T. Ottmann, "The "Authoring on the

- Fly” system for automated recording and replay of (tele)presentations,” *Multimedia Syst.*, vol.8, no.3, pp.158–176, 2000.
- [2] A.G. Hauptmann and M.J. Witbrock, “Informedia: News-on-demand multimedia information acquisition and retrieval,” in *Intelligent Multimedia Information Retrieval*, ed. M.T. Maybury, pp.215–239, MIT Press, 1997.
- [3] M. da G. Pimentel, Y. Ishiguro, G.D. Abowd, B. Kerimbaev, and M. Guzdial, “Supporting educational activities through dynamic web interfaces,” *Interacting with Computers*, vol.13, no.3, pp.353–374, Feb. 2001.
- [4] 森本容介, 室田真男, 清水康敬, “教育用動画検索システムと時間情報同期方法の開発,” *信学論 (D-I)*, vol.J88-D-I, no.10, pp.1515–1524, Oct. 2005.
- [5] A. Fujii, K. Itou, and T. Ishikawa, “Lodem: A system for on-demand video lectures,” *Speech Commun.*, vol.48, no.5, pp.516–531, 2006.
- [6] H. Yokota, T. Kobayashi, T. Muraki, and S. Naoi, “UPRISE: Unified presentation slide retrieval by impression search engine,” *IEICE Trans. Inf. & Syst.*, vol.E87-D, no.2, pp.397–406, Feb. 2004.
- [7] Y. Kambayashi, K. Katayama, Y. Kamiya, and O. Kagawa, “Index generation and advanced search functions for multimedia presentation material,” *Proc. ER97 Workshop on Conceptual Modeling in Multimedia Information Seeking*, 1997.
- [8] 大平茂輝, 長尾 確, “レーザーポインタ指示情報に基づく会議コンテンツの協調的提示手法の提案,” *情報処理学会第 68 回全国大会*, 4G-2, 2006.
- [9] W. Nakano, Y. Ochi, T. Kobayashi, Y. Katsuyama, S. Naoi, and H. Yokota, “Unified presentation contents retrieval using laser pointer information,” *Proc. SWOD2005*, pp.170–173, April 2005.
- [10] 仲野 亘, 小林隆志, 勝山 裕, 直井 聡, 横田治夫, “講演シーン検索における検索語出現状況に基づくレーザーポインタ情報のフィルタリング,” *日本データベース学会 DBSJ Letters*, vol.5, no.2, pp.1–4, Sept. 2006.
- [11] W. Nakano, T. Kobayashi, Y. Katsuyama, S. Naoi, and H. Yokota, “Treatment of laser pointer and speech information in lecture scene retrieval,” *Proc. Eighth IEEE Intl. Symp. on Multimedia*, pp.927–932, Dec. 2006.
- [12] 仲野 亘, 小林隆志, 直井 聡, 横田治夫, 古井貞照, “講義講演シーン検索手法におけるレーザーポインタ情報と音声情報の粒度を考慮した統合,” *DEWS2007 論文集*, no.E1-3, March 2007.
- [13] W. Hürst and R. Müller, “A synchronization model for recorded presentations and its relevance for information retrieval,” *Proc. Seventh ACM International Conference on Multimedia (Part 1)*, pp.333–342, ACM Press, 1999.
- [14] G.D. Abowd, “Classroom 2000: An experiment with the instrumentation of a living educational environment,” *IBM Syst. J.*, vol.38, no.4, pp.508–530, 1999.
- [15] (株)リコー: MPMeister.  
<http://www.ricoh.co.jp/mpmeister/>
- [16] (株)アーネット: 遠隔講義システム P4Web.  
<http://www.earnnet.co.jp/jp/>
- [17] (株)富士通ソフトウェアテクノロジー: Internet Navigware. <http://www.navigware.com/>
- [18] 湯本高行, 吹野直紀, 馬 強, 角谷和俊, 田中克己, “Video-augmented web: ビデオストリームの web ページへの動的統合,” *日本データベース学会 Letters*, vol.2, no.2, pp.41–44, 2003.
- [19] 片山 薫, 香川修見, 神谷泰宏, 對馬英樹, 吉廣卓哉, 上林彌彦, “遠隔教育のための柔軟な講義検索手法,” *情処学論*, vol.39, no.10, pp.2837–2845, Oct. 1998.
- [20] 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,” *信学論 (D-I)*, vol.J88-D-I, no.3, pp.715–726, March 2005.
- [21] 岡本拓明, 仲野 亘, 小林隆志, 直井 聡, 横田治夫, 岩野公司, 古井貞照, “音声情報を統合したプレゼンテーションコンテンツ検索,” *信学論 (D)*, vol.J90-D, no.2, pp.209–222, Feb. 2007.
- [22] Y. Katsuyama, N. Ozawa, J. Sun, H. Takebe, T. Kobayashi, H. Yokota, and S. Naoi, “A new solution for extracting laser pointer information from lecture videos,” *Proc. E-learn2004*, pp.2713–2718, Oct. 2004.
- [23] 河原達也, 李 見伸, “連続音声認識ソフトウェア Julius,” *人工知能誌*, vol.20, no.1, pp.41–49, 2005.
- [24] 山崎裕紀, 岩野公司, 篠田浩一, 古井貞照, 横田治夫, “講義音声認識における講義スライド情報の利用,” *情処学研報*, 2006-SLP-64-39, pp.221–226, Dec. 2006.
- [25] 酒井哲也, “よりよい検索システム実現のために,” *情報処理*, vol.47, no.2, pp.147–158, Feb. 2006.

(平成 19 年 5 月 29 日受付, 9 月 28 日再受付)



仲野 亘

平 17 東工大・工・情報工卒・平 19 同大大学院・情報理工・計算工・修士課程了。現在 (株) 東芝研究開発センター。複合メディアコンテンツの管理・検索・推薦に関する研究に従事。日本データベース学会会員。



小林 隆志 (正員)

平 9 東工大・工・情報工卒．平 16 同大大学院・情報理工・計算工・博士課程了．平 14 同大学術国際情報センター助手．平 19 より 名大大学院・情報科学・特任准教授，現在に至る．工博．ソフトウェア開発方法論，ソフトウェア再利用技術，複合メディアコンテンツの管理・検索，Web サービス連携などの研究に従事．情報処理学会，日本ソフトウェア科学会，日本データベース学会，ACM 各会員．



直井 聡 (正員)

昭 58 慶大・工・電工卒．昭 60 同大大学院・工・電気工学・修士課程了．同年より(株)富士通研究所，現在，部長．平 13 東工大・学術国際情報センター・客員助教授．平 17 より同センター・客員教授．工博．文字パターン処理，画像処理，文字認識や e ラーニングの研究に従事．情報処理学会会員．



横田 治夫 (正員：フェロー)

昭 55 東工大・工・電物卒．昭 57 同大大学院・情報・修士課程了．同年富士通(株)．同年 6 月(財)新世代コンピュータ技術開発機構研究所(ICOT)．昭 61(株)富士通研究所．平 4 北陸先端大・情報・助教授．平 10 東工大・大学院情報理工・助教授．平 13 より同大学術国際情報センター教授，現在に至る．工博．主として分散インデキシング，データ工学向けアーキテクチャ，高機能ストレージシステム，ディペンダブルシステム等に関する研究に従事．日本データベース学会理事．ACM SIGMOD 日本支部評議委員．情報処理学会，人工知能学会，IEEE，ACM 各会員．