社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2006－DBS－140（Ｉ）（2）
2006／7／12

講義・講演シーン検索におけるスライドおよび音声中の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

仲野　亘[†]　小林隆志[††]　勝山　裕[†††]
直井　聡[†††]　横田治夫[††,†]

我々はこれまで，講義・講演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高度なシーン検索を提供する UPRISE を提案してきた．本稿では，UPRISE の検索機能の精度をより向上させるために，これまで検索に利用してきた講師が用いるレーザーポインタ情報から，スライド中の文字列を強調する目的以外で行われたレーザーポインタ照射を除去することを試みる．スライド中および音声中の検索語出現状況に基づいてレーザーポインタ情報をフィルタリングすることで，検索に有用なレーザーポインタ情報のみを利用する手法を提案する．また，実際の講義コンテンツに対して検索実験を行い，提案手法の有用性を示す．

Filtering the Laser Pointer Information based on Keyword Occurrence in Slides and Speech in Presentation Scene Retrieval

Wataru Nakano,[†] Takashi Kobayashi,[††] Yutaka Katsuyama,[†††] Satoshi Naoi[†††]
and Haruo Yokota[††,†]

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 Slide Retrieval by Impression Search Engine). In this paper, we try to get rid of irrelevant laser pointer information that is used in unified contents retrieval. We propose a method to filter laser pointer information based on keyword occurrence in slides and speech. We evaluate our approach by applying it to actual presentation contents.

1. はじめに

近年，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，検索するシステムが数多く研究，および提案されており[1)~4)]，e-Learning など，様々な用途に用いられている．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，コンテンツのどの箇所から視聴するべき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，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である UPRISE(Unified Presentation Slide Retrieval by Impression Search Engine) を提案してきた[5)]．

UPRISE では，動画ストリームを資料スライドの切り替え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付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，スライドの切り替えタイミングによってシーンを分割し，それら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，巻き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．

我々はこの UPRISE の検索手法において，講師が用いるレーザーポインタなどのポインティング情報に着

† 東京工業大学 大学院 情報理工学研究科 計算工学専攻
Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology
†† 東京工業大学 学術国際情報センター
Global Scientific Information & Computing Center, Tokyo Institute of Technology
††† 株式会社 富士通研究所
FUJITSU LABORATORIES LTD.

目し，ポインティング情報を統合することで検索精度を向上させる検索手法を提案してきた[6]．しかし，レーザーポインタの照射には講師の手のぶれや，複数概念間の関連を軌跡で示すもの，指示対象を明確に意図しないあいまいなものなど，スライド中の文字列を強調する目的で行われるもの以外の照射も多く見られるが，従来の手法では，これらの検索に考慮するべきではない照射も一様にレーザーポインタ情報として扱い，シーン検索に利用してきた．

そこで本稿では，そのような検索に考慮するべきではないレーザーポインタ情報を他の情報を用いて除去し，スライド中の文章を強調する目的で使用されたレーザーポインタだけを選び取ることで，検索の精度向上を実現する手法を提案する．提案手法では，複数キーワードのスライド中の出現の有無や，講師の発話した音声中の出現の有無を考慮することで，各レーザーポインタ照射のフィルタリングを行う．また，それらのフィルタリングしたレーザーポインタ情報を従来の検索手法に統合した手法を提案する．

以下では，まず 2 節で UPRISE の概要を示し，3 節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従来手法について述べる．4 節では，レーザーポインタ情報をフィルタリングする 2 手法について提案し，従来の適合度との統合を行う．また，5 節では，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．最後に 6 節においてまとめと今後の課題を述べる．

2. UPRISE の概要

本節では，UPRISE の概要について述べる．まず，UPRISE によるコンテンツ統合とその検索の概要を示し，次に検索に用いる基本的な適合度について述べる．そして，音声情報を統合した適合度について説明する．

2.1 UPRISE のシステム

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に対するインデクスを含める．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．UPRISE のシステムの詳細については[7] 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．

2.2 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，上位のシーンから表示する．この適合度のうち，最も基本的なものは適合度 I_c である．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 種類の情報を元に算出される．

以下では，まず I_c を構成する適合度 I_p および I_d について説明した後，I_c について述べる．

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．

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

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．

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．

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

ここで，$T(s)$ はシーン s の時間であり，θ は時間の影響の強弱を定めるパラメタである．これによって，説明を長く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．

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

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

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

適合度 I_c では，シーンの適合度はその前後 δ の範囲から影響を受け，ε 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．

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タ群であるため，本稿では以降これを Φ として簡略化表現する．

2.3 音声情報を考慮した適合度

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．さらに，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．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[8]．

[8] では，まず音声認識によって講義中の音声情報を抽出する．そして，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) とした．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．

3. レーザーポインタ情報を考慮した適合度

講師は，主に資料スライド中のある部分を強調する目的でレーザーポインタを用いる．そこで，我々は検索キーワードに対してレーザーポインタが多く当たっているシーンはそのキーワードに対してより適切であると考え，それらのレーザーポインタ情報を考慮した適合度を提案してきた[6]．本節では，レーザーポインタ情報の抽出手法と，レーザーポインタ情報を統合した適合度について説明する．

3.1 レーザーポインタ情報の抽出

撮影動画より画像認識によって抽出したポインタの光点座標[9] 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，レーザーポインタ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く，単語単位で取得するのではうまくキーワードと関連付けられないと考えたためである．

次に，同じ行の文字列を取得した連続の光点を 1 回のレーザーポインタ照射と定義し，一つのレーザーポインタ情報として統合する．一つのレーザー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．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と考えたため，1 回の照射としての統合は同一シーンの中でのみ行う．

ここで，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ないため，レーザー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レーザーポインタ情報として抽出してしまうことがある．そこで，一回のレーザーポインタに対し，近傍の数行を次候補として取得しておく．最も近い行とその付近のいくつかの行を組にしてレーザー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．

また，この 1 回のレーザーポインタに相当する部分をサブシーンと定義する．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り，各サブシーンはある特定のシーンに属することになる．このように定義することで，各シーンで講師がいくつの話題について説明しているのかを知ることができる．

3.2 レーザーポインタ情報と I_c の統合

3.1 節において抽出したレーザーポインタ情報を，適合度 I_c に統合する．まず，レーザー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ということはすでに述べた．そこで，レーザーポインタの照射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 $H(l,q)$ として数値化する．$H(l,q)$ はサブシーン q のレーザーポインタが，行 l に照射された回数の期待値であり，全ての行の $H(l,q)$ を合計すると 1 となる．

こうして得られたレーザーポインタごとの照射回数期待値に対し，各レーザーポインタが当たっていた時間を掛け合わせることで，レーザーポインタ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．ただし，s はシーン，k はキーワードである．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，$T(q_i)$ 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タの照射時間を表す．

この phd を適合度 I_c と統合することにより，レー

ザーポインタ情報を利用した適合度を提案してきた．これまでの報告，実験では，シーンごとの時間情報である $T(s)$ に対し，レーザーポインタの時間の期待値である $phd(s,k)$ を足し合わせ，レーザーポインタが当たっていたときにそのシーンの時間に加点するという統合手法が最も有効であるということがわかっている．この適合度を $I_{c[d+phd]}$ とし，以下の式で定義する．

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

$$I_{d[d+phd]}(s,k,\theta,\omega_d) = \{T(s)+\omega_d\cdot phd(s,k)\}^{\theta}\cdot I_p(s,k)$$

ここで，ω_d は phd の適合度計算における影響度合いを調節するパラメタである．例えば，$\omega_d = 10$ の場合，レーザーポインタが 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．

4. レーザーポインタ情報のフィルタリング

3 節で述べたように，従来の手法では，レーザーポインタ情報は抽出した情報全てを同じ条件で利用していた．そのため，スライド中文字列の直接的な強調という目的以外で行われた照射も適合度の計算に用いていた．そこで，抽出したレーザーポインタ照射の中から，そのような強調目的以外の照射を排除するために，phd を別の情報を用いてフィルタリングする．以下では，複数キーワードのスライド中の出現条件によるフィルタリングと，講義の音声情報を利用したフィルタリングの 2 種類の手法を提案する．また，それら 2 種類の条件を同時に考慮したフィルタリングについても提案を行う．

4.1 複数キーワードのスライド中の出現条件による フィルタリング

UPRISE では，検索語を形態素解析して形態素ごとに適合度を算出し，その和を求めることで検索語の適合度を求める．この形態素解析によって複数に分割される検索語を用いた検索や，複数のキーワードによる AND 検索など，複数の単語を用いた検索を行う機会は多い．このとき，従来の適合度算出手法では，分割された単語それぞれへのレーザーポインタの影響を考慮していたため，スライド中に複数の単語全てが出現していなくても，レーザーポインタにより適合度が加算されていた．

しかし，キーワード全てを含まない行に向けてのレーザーポインタ照射は，その検索キーワードに対しての強調の意味を持たないと考える．そこで，複数の単語による検索の際に，その単語全てがスライド上になければそのシーンにおけるキーワードへのレーザーポインタ照射は無視する，という手法を提案する．この条件でフィルタリングを行った $phd(s,k)$ を，$phd_{/p}(s,k)$ として以下のように定義する．

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

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．

また，この $phd_{/p}(s,k)$ を，3.2 節で述べた適合度 $I_{c[d+phd]}$ に統合する．この適合度を $I_{c[d+phd_{/p}]}$ とし，以下のように定義する．

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

$$I_{d[d+phd_{/p}]}(s,k,\theta,\omega_d) = \{T(s)+\omega_d\cdot phd_{/p}(s,k)\}^{\theta}\cdot I_p(s,k)$$

4.2 音声情報を用いたフィルタリング

検索キーワードに対してレーザーポインタが当たっているにもかかわらず，講師がそのシーンの中でキーワードを発話していない場合，そのレーザーポインタ照射は検索キーワードを説明する補助として用いられたのではないと考える．そこで，あるシーン中でキーワードが発話されていない場合，つまりそのシーンの音声中出現回数である $skc(s,k) = 0$ の場合には，そのシーンにおけるキーワードへのレーザーポインタ照射は無視するという手法を提案する．この条件でフィルタリングを行った $phd(s,k)$ を，$phd_{/s}(s,k)$ として以下のように定義する．

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

この $phd_{/s}$ を適合度 $I_{c[d+phd]}$ に統合する．さらに，[8] で行ったように，skc による適合度の加算分も考慮する．この適合度を $I_{c[d+phd_{/s},p+skc_{/p}]}$ とし，以下の式で表す．

$$I_{c[d+phd_{/s},p+skc_{/p}]}(\Phi,\omega_d,\psi) = \sum_{\gamma=s-\delta}^{s+\delta} E(\gamma-s,\varepsilon_1,\varepsilon_2)\cdot I_{d[d+phd_{/s},p+skc_{/p}]}(\gamma,k,\theta,\omega_d,\psi)$$

$$I_{d[d+phd_{/s},p+skc_{/p}]}(s,k,\theta,\omega_d,\psi) = \{T(s)+\omega_d\cdot phd_{/s}(s,k)\}^{\theta}\cdot\{I_p(s,k)+\psi\cdot skc_{/p}(s,k)\}$$

ここで，ψ は skc が適合度に与える影響度合いを調節するパラメタである．また，$skc_{/p}(s,k)$ はシーン s で用いられるスライドの文字列中にキーワード k が出現

する場合のみ $skc(s,k)$ を計算するというもので，以下の式で表される．

$$skc_{/p}(s,k) = \begin{cases} skc(s,k) & I_p(s,k) \neq 0 \\ 0 & I_p(s,k) = 0 \end{cases}$$

4.3 2 条件を同時に考慮したフィルタリング

4.1，4.2 節で用いた条件はそれぞれが独立である．そこで，両者の条件を満たす場合にのみレーザーポインタ情報を考慮することを考える．この 2 つの条件でフィルタリングを行った $phd(s,k)$ を，$phd_{/ps}(s,k)$ として以下のように定義する．

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

$I_{c[d+phd_{/s},p+skc_{/p}]}$ における $phd_{/s}(s,k)$ を $phd_{/ps}(s,k)$ に置き換えたものを，適合度 $I_{c[d+phd_{/ps},p+skc_{/p}]}$ とする．

5. 実　　験

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．以下ではその実験に関して説明し，実験結果に対して考察を行う．

5.1 実験に用いたデータ

実験では，データベースについての講義 (全 11 回)，計算機アーキテクチャについての講義 (全 12 回) をコンテンツ化し，検索対象とした．

講義の音声情報は，連続音声認識ソフトウェア Julius ☆を用い，言語モデルと音響モデルとして，山崎が[10] で作成したものを用いた．[10] では，日本語話し言葉コーパス (CSJ)[11),12)] の，学会講演 953 講演と模擬講演 1543 講演を学習データとして音響モデルを作成し，CSJ の学会講演 967 講演 (約 300 万単語) の書き起こしを学習データとして，バイグラムおよび逆向きトライグラムを言語モデルとして作成している．言語モデル作成時の形態素解析には，茶筌☆☆，形態素解析用の辞書として，ipadic ☆☆☆を用いている．また，認識用の辞書として，学習データ (約 300 万語) での出現頻度が高い順から選んだ 22,860 語を登録している．

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．そこで，山崎が[10] において作成した辞書に，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．

☆ http://julius.sourceforge.jp/
☆☆ http://chasen.naist.jp/
☆☆☆ http://chasen.naist.jp/

5.2 実　　験

提案手法の評価を行うため，5.1 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以下の条件の下で行った．

- 基本となるパラメタ Φ は $\theta = 0.4$，$\delta = 4$，$\varepsilon_1 = 5.0$，$\varepsilon_2 = 0.5$ とし，skc の影響の強弱を表すパラメタ ψ は 1 に固定した．
- レーザーポインタの光点に対し 5 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 という値に設定した．
- 各適合度ごとに 124 種類のキーワードを検索した．なお，計算機アーキテクチャ関連のキーワードが 78 種類，データベース関連のキーワードが 46 種類である．
- 各適合度に対して，phd の影響の強弱を表すパラメタ ω_d を 1 から 30 まで 5 刻みに変更し，計 7 回の計測を行った．
- 検索対象範囲はキーワードの正解シーンの含まれる講義ごととした．
- 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，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．

ψ を 1 に固定することで，skc の強弱の変化による影響を排除し，phd のフィルタリングの効果を検証しやすくした．

また，今回の実験では講義ごとに辞書を作成し，音声認識を行ったため，検索範囲を各講義ごとに限定した．しかし，検索キーワードとして用いた単語には講義間で共通のものはほとんどなく，評価にはほとんど影響を与えないと考える．

なお，評価に際して指標となる再現率と適合率[13] について述べる．今回の実験では，正解シーンを各キーワードに対して 1 つとしており，また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．

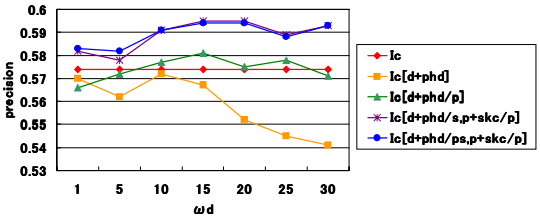

図 1 ω_d を変化させたときの適合度ごとの適合率の推移

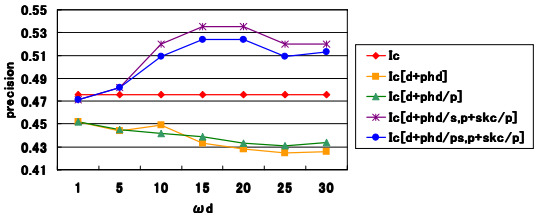

図 2 データベースに関する講義における適合率の推移

$$precision = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\text{the rank of the best scene in the i-th test}}$$

5.3 実験結果と考察

提案手法である $I_{[d+phd_{/p}]}$，$I_{c[d+phd_{/s},p+skc_{/p}]}$，$I_{c[d+phd_{/ps},p+skc_{/p}]}$ の 3 種類の適合度と，従来のフィルタリングを行わない適合度である I_c，$I_{c[d+phd]}$ の比較を行った．

図 1 は，2 講義の検索キーワードを用いた検索による，各適合度の適合率の変化を示したグラフである．グラフより，提案手法である *phd* をフィルタリングして用いている適合度は全体的に検索精度が向上していることがわかる．特に，従来の $I_{c[d+phd]}$ ではパラメタ ω_d を大きくした場合に検索精度が下がるという傾向があったが，フィルタリングした *phd* を用いた適合度では，少なくとも今回の ω_d の範囲では十分な精度を保っている．このことは，これまで ω_d を大きくすると不要なレーザーポインタ照射による適合度計算への影響が顕著になり，精度を下げていたが，それらをフィルタリングしたために，レーザーポインタ情報の影響度を高めても精度を維持できている，ということであると考える．

図 2，3 はそれぞれ，計算機アーキテクチャに関する講義とデータベースに関する講義のそれぞれで検索を行い，各適合度ごとに適合率を算出したグラフである．図 2 が示すように，従来の手法では，データベースに関する講義はレーザーポインタ情報を利用することで検索精度が下がってしまっていた．これは，基本適合度である I_c の値が計算機アーキテクチャに関する講義と比較して低めであることからも考えて，この講義が，スライド文字列情報に基づいたランク付けにはあまり適していないスライド構造，および講義のトピック構造をしていたためであると考える．しかし，レーザーポインタ情報を音声情報でフィルタリングすることにより，ω_d の値を大きくする，つまりレーザーポインタ情報の影響度を高めると，検索精度が大幅に向上するという結果を得た．このことから，この講義では特に，検索キーワードの強調以外の目的で行われたレーザーポインタ照射を，音声情報によって非常によく除去出来ているということがわかる．

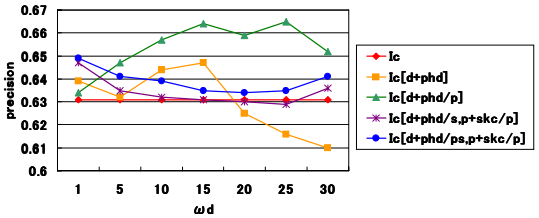

図 3 計算機アーキテクチャに関する講義における適合率の推移

また，計算機アーキテクチャに関する講義では，図 3 が示す通り，従来の手法でもレーザーポインタ情報は検索に良い影響を与えており，それに複数キーワードのスライド上での出現の有無によるフィルタリングを行うことで，全適合度の中で最も良い精度を達成している．このことは，この講義がデータベースに関する講義とは講義の特性が異なっており，スライド構造を基本として適合度を算出する手法に十分に適した講義だということである．

このように，2 つの講義だけを取り上げてみても，その講義ごとに特性は様々である．この 2 つの講義は講義形態も若干異なっている．例えば，データベースに関する講義においては講義 1，2 回ごとに 1 回の頻度で演習を行う形式であるが，計算機アーキテクチャに関する講義では演習を全く行っていない．このような講義形態による特性の分析や，それに基づいて，検索に利用する適合度の選択やパラメタの調節などを行うことが必要であり，それらは今後の課題である．

6. まとめと今後の課題

本稿では UPRISE の検索精度を向上させるために，レーザーポインタ情報を他の情報でフィルタリングし，有用なレーザーポインタ情報だけを検索に利用する手法を提案した．提案手法では複数キーワードのスライド中での出現の有無と，キーワードの音声中での出現の有無によって *phd* をフィルタリングしている．さら

に，フィルタリングした *phd* を従来の適合度に統合し，実際の講義をコンテンツ化して UPRISE に登録したものを利用して，検索実験を行った．

評価実験の結果，フィルタリングした *phd* を用いた適合度は従来の適合度 $I_{c[d+phd]}$ などに比べて精度が向上していることを確認した．また，講義ごとに適合率の推移を算出し，フィルタリングにより不要なレーザーポインタ情報が除去されていることを確認し，講義ごとの各情報に対する特性の違いを考察した．講義ごとの特性をより詳細に調べるためには，今後さらに異なる講義や話者のコンテンツを蓄積し，様々な講義に対して実験を行うことが必要であることがわかった．

本研究のその他の今後の課題としては，まず，今回の提案では音声情報をレーザーポインタ情報のフィルタリングに利用する形で統合したが，音声情報自体の検索への有用性も報告を行ってきた．このため，レーザーポインタ情報と音声情報をより関連付け，一つの適合度計算手法としてまとめることでより高度な検索が可能になると考えている．

次に，今回の実験では音声情報として，音声認識によって抽出した情報を用いた．音声認識による情報においては，実際にはその場面で発話されていないにもかかわらず認識されてしまうという認識誤りや，音声認識に用いた辞書とスライド上の単語間での表記ゆれ，辞書中の英単語の読み付加ができていないことなどを解決することで，より精度の高いレーザーポインタ情報のフィルタリングが可能になると考える．また，音声認識精度が 100%になった条件における実験として，講義音声を人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．

さらに，提案した適合度では音声中，およ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は有効であり[8),14)]，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さら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ご協力頂いた，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熙教授，岩野公司助手，山崎裕紀氏に感謝致します．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，および東京工業大学 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．

参 考 文 献

1) Müller, R. and Ottmann, T.: 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) The Informedia Project, C. M. U.: Informedia II Digital Video Library. http://www.informedia.cs.cmu.edu/.
3) Abowd, G.D.: Classroom 2000: an experiment with the instrumentation of a living edu cational environment, *IBM Syst. J.*, Vol. 38, No. 4, pp. 508–530 (1999).
4) Fujii, A., Itou, K. and Ishikawa, T.: LODEM: A system for on-demand video lectures, *Speech Communication*, Vol.48, No.5, pp.516–531 (2006).
5) Yokota, H., Kobayashi, T., Muraki, T. and Naoi, S.: UPRISE: Unified Presentation Slide Retrieval by Impression Search Engine, *IEICE Trans. on Info. and Syst.*, Vol.E87-D, No.2, pp.397–406 (2004).
6) Nakano, W., Ochi, Y., Kobayashi, T., Katsuyama, Y., Naoi, S. and Yokota, H.: Unified Presentation Contents Retrieval Using Laser Pointer Information, *Proc. of SWOD*, pp.170–173 (2005).
7) 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， Vol.J88-D-I, No.3, pp.715–726 (2005).
8) 岡本拓明，仲野　亘，小林隆志，直井　聡，横田治夫，岩野公司，古井貞熙 : プレゼンテーション蓄積検索システムにおける講義・講演音声情報を利用した適合度の改善，データ工学ワークショップ，pp.DEWS2006–6C–o1 (2006).
9) Katsuyama, Y., Ozawa, N., Sun, J., Takebe, H., Kobayashi, T., Yokota, H. and Naoi, S.: A New Solution for Extracting Laser Pointer Information from Lecture Videos, *Proc. of E-learn2004*, pp. 2713–2718 (2004).
10) 山崎裕紀 : 講義音声認識の高精度化に関する研究，東京工業大学 工学部 卒業論文 (2005).
11) 国立国語研究所 : 日本語話し言葉コーパス. `http://www2.kokken. go.jp/˜csj/public/`.
12) Maekawa, K., Koiso, H., Furui, S. and Isahara, H.: Spontaneous speech corpus of Japanese, *Proc. LREC2000*, Vol. 2, Athens, Greece, pp. 947–952 (2000).
13) Grossman, D.A. and Frieder, O.: *Information Retrieval Algorithm and Heuristics.*, Kluwer (1998).
14) Yokota, H., Kobayashi, T., Okamoto, H. and Nakano, W.: Unified Contents Retrieval from an Academic Repository, *Proc. of International Symposium on Large-scale Knowledge Resources LKR2006*, pp.41–46 (2006).