

スライドと映像のメタデータを用いたシーンの意味的關係 に基づくプレゼンテーション管理システム

王 元元[†] 北山 大輔^{††} 角谷 和俊[†]

[†] 兵庫県立大学環境人間学部 〒 670-0092 兵庫県姫路市新在家本町 1 丁目 1-12

^{††} 兵庫県立大学大学院環境人間学研究科 〒 670-0092 兵庫県姫路市新在家本町 1 丁目 1-12

E-mail: †{nf08m104,ne07p001}@stshse.u-hyogo.ac.jp, ††sumiya@shse.u-hyogo.ac.jp

あらまし 近年、スライドや映像などの異種メディアを組み合わせるプレゼンテーションコンテンツが大量にインターネット上に公開されるようになり、E-learning や遠隔会議などの教育の場面で利用されている。しかし、膨大なプレゼンテーションコンテンツが蓄積され、視聴時に検索キーワードを含むシーンを抜き出すだけでは、その内容の理解が困難な場合がある。そのため、プレゼンテーションコンテンツ検索のための異コンテンツ間の視聴区間検索への拡張を行う。その際、単一のプレゼンテーションコンテンツを対象とするのではなく、複数コンテンツを対象としたシーンの関係性の重要度判定方式が必要である。本研究では、蓄積されたプレゼンテーション視聴時に、複数コンテンツから意味的シーン間の関係をスライドと映像のメタデータの変化により判定し、関係となるシーンの組を提示する手法を提案し、評価実験を行った。

キーワード プレゼンテーションコンテンツ, マルチメディア, 区間抽出

An Extracting Method of Semantic Relations using the Metadata of Slide and Video for Presentation Management System

Yuanyuan WANG[†], Daisuke KITAYAMA^{††}, and Kazutoshi SUMIYA[†]

[†] School of Human Science and Environment, University of Hyogo
1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan

^{††} Graduate School of Human Science and Environment, University of Hyogo
1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan

E-mail: †{nf08m104,ne07p001}@stshse.u-hyogo.ac.jp, ††sumiya@shse.u-hyogo.ac.jp

Abstract In the recent years, more and more presentation contents, which have become easier for users to view presentation content consisting of synchronized heterogeneous media such as videos and slides on the Internet. In the case of E-learning or remote conferencing, users simply retrieve the stored content they wish to view. In most cases users are interested in viewing a specific portion of the content. Scenes extracted from a presentation, however, if there is a slide including a keyword which users do not know, it is difficult for them to understand the rest of the presentation contents. In this case, the research object is not single of presentation contents, is multi-contents. A method is needed for extracting viewing intervals from a content which are meaningful to the user. In this paper, we propose a method to extract scenes according to their relation. These relations are decided at the point of change between scenes. Finally, we evaluate our proposed method by evaluation experiment.

Key words Presentation content, Multimedia, Interval extraction

1. はじめに

現在、ブロードバンドネットワークの普及に伴い、テキストのみではなく、高品質の映像や音声コンテンツが多数利用されるようになってきている。このことにより、異種のメディアを

複合させたコンテンツも登場しており、特にビデオやスライドなどの異種メディアを組み合わせるプレゼンテーションコンテンツが手軽にインターネット上に公開されるようになってきている。このようなプレゼンテーションコンテンツ(以降、単にコンテンツと記す)を生成、蓄積するシステムが開発されてい

る [1] . 蓄積されたコンテンツは, E-learning や遠隔会議などの教育の場面で利用される. 例えば, 会議や講義・講演活動の参加できなかった者が後日にプレゼンテーションを視聴する場合や, 発表者や聴講者がプレゼンテーションの内容を知った上で再度視聴する場合に利用することが出来る. しかし, 膨大なプレゼンテーションコンテンツが蓄積された場合, 視聴時に適切なシーンを検索することが困難となる. 例えば, 検索キーワードを含むシーンを抜き出すだけでは, プレゼンテーションのコンテキストが失われるため, その内容の理解が困難な場合がある. そのため, コンテンツのシーン検索のためのシーン間の関係を考慮したコンテンツ管理方式を提案する. その際, 単一のコンテンツを対象とするのではなく, 複数のコンテンツを対象とした関係するシーンの判定方式が必要となる.

コンテンツは, スライドと映像という, 異種メディアによって構成される. スライドの切り替わりによって区切られる単位をシーンと定義する. そして, あるシーンからみて他のシーンが詳しい内容を説明しているというような, シーン間の意味的な関係を定義する. 意味的な関係に当たる他のシーンを共に視聴することで, 効率的な視聴を支援できる. この時, スライドや映像というそれぞれのメディアの変化を分析することによりこのような意味的な関係を求めることができると考えた. 具体的には, シーン間の意味的な関係は, スライドにおける階層構造をもったメタデータと, 映像における構造を持たないメタデータという異なるメディアのメタデータを組み合わせることにより判定する. また, このメタデータを組み合わせることにより, シーン間の関係として, detailed, abstracted, equal, additional という関係を抽出する. さらに, これらの関係を用いることで, detailed, abstracted, equal, additional, compared の5つの種類の意味を持つシーンの検索を行う. そのことにより視聴中のコンテンツを効率的に視聴するためのシーン抽出を行うことが可能となる.

以下, 2 節では本研究のアプローチと関連研究について述べる. 3 節では, シーンの意味的な関係について説明する. 4 節では, シーンの意味的な関係の判定について述べ, 5 節では, 視聴区間の決定について説明する. 6 節で, プロトタイプシステムおよび本手法の評価について述べ, 最後に, 7 節でまとめを述べる.

2. 本研究のアプローチと関連研究

2.1 本研究のアプローチ

ユーザがプレゼンテーションコンテンツを視聴する際, あるシーンに対して, もっとよく知りたいという要求が発生することが考えられる. この時, 指定されたシーンと適切に関係するシーンを提示することで要求に応えることが可能である. しかし, コンテンツの構成は作成者によって異なり, 時間的に前後して出現するシーンや, 指定したシーンに類似したシーンが, 関係するシーンであるとは限らない. また, ユーザの求めるシーンが視聴中のコンテンツに含まれるとは限らないため, 単一のコンテンツを対象とするのではなく, 複数コンテンツを対象として関係するシーンの抽出方式が必要となる. そのため,

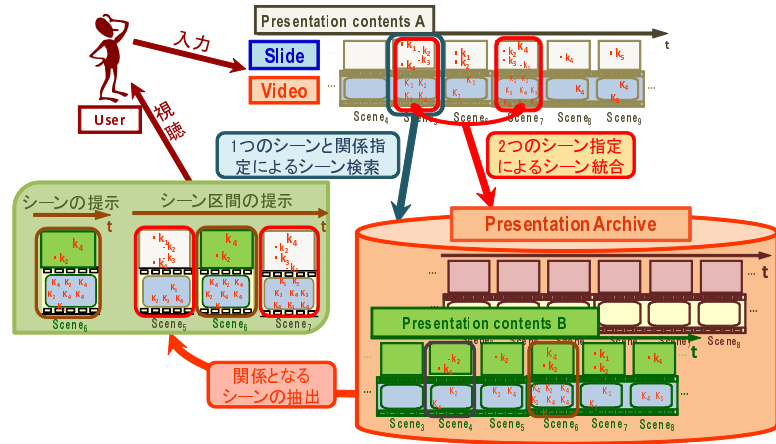


図 1 関係抽出に基づくシーン抽出の概念図

Fig. 1 Concept of extracting relations

コンテンツの構造的な情報を用いて関係するシーンを判定し, コンテンツの時系列的に連続する区間ではなく, 複数コンテンツのシーン間に意味的な関係が存在するシーン群を提示する方式を提案する. 図 1 に本手法の概念図を示す.

複数コンテンツ中のシーンは, 他のシーンと意味的に関係を持つ場合がある. 例えば, あるシーンで説明した内容を, 後のシーンで詳しく説明している場合, これらのシーン間には detailed という関係があると考えた. シーン間の意味的な関係としては, detailed, abstracted, equal, additional の 4 種類を定義する. このように, あるシーンに対し, 意味的に関係するシーンを判定し, 結合することで, 他のコンテンツから適切なシーンを抽出し, ユーザに提示できると考えた. 本手法では, まずコンテンツに対して, シーンが持つ異種メディアのメタデータを直接用いてシーン間の意味的な関係を判定する. 具体的には, 同一キーワードのスライドでの階層構造の変化, ビデオでの発言量の変化を用いる. シーンを指定する入力には, 単一シーンを指定する場合と, 複数シーンを指定する場合がある. 単一シーンを指定した場合に検索可能なシーンタイプとして, detailed, abstracted, equal, additional, compared の 5 種類を定義する. 複数シーンを指定した場合に検索可能なシーンタイプ区間として, detailed 区間, abstracted 区間, equal 区間, additional 区間の 4 種類を定義する. これらのシーンは, 意味的な関係を組み合わせることにより判定する.

2.2 関連研究

異種メディア統合コンテンツのメタデータ特性の利用に関する研究として以下の研究が挙げられる. Smith ら [5] は, CC (Closed Caption), 音声, 画像ストリームの協調処理によりシーンの特徴を検出することで, 映像から区間抽出を行う手法を提案している. レーヒェウハンら [11] は, 動画とスライドからコンテンツを作成するシステムである MPMeister [1] で収録された講義講演の情報から, 重要シーンを抽出しダイジェストを自動的に作成する手法を提案している. 小林らの提案する UPRISE [7] では, メタデータを用いて資料や映像の疎結合を行い, コンテンツを蓄積し, 提供する枠組みを提案している. 従

来の異種メディアコンテンツの特性を用いた研究では、コンテンツ全体から重要なシーンを判定し、要約を生成する手法が主として提案されている。本手法では、シーンの重要性を用いずにシーン間の相対的な関係を抽出し、要約のみならず、様々なアプリケーションへ応用することを目的としている点で異なる。

映像の区間抽出に関する研究として Pradhan らの研究が挙げられる。Pradhan ら [6] は、断片的なビデオ区間集合を組み合わせ、新しい区間を生成するグルー結合を提案している。コンテンツ中のシーン間における部分的な情報から関係するシーンを抽出するという点で本研究と類似しているが、この手法は、映像という単一メディアを対象としたものであり、複数メディアからなるコンテンツのシーン間の関係を抽出する本手法とは対象が異なる。

大谷ら [12] は、コンテンツのシーン間の関係やシーンの役割を用いてコンテンツから視聴区間を抽出する手法を提案している。北山ら [13] は、コンテンツのコンテキストとユーザの視聴時行為に基づく区間検索方式を提案している。これらの研究における入力にはシーン指定であり、同一コンテンツ内でのシーン間の関係しか考慮することができていない。本手法では単一のコンテンツを対象とするのではなく、複数のコンテンツを対象とし、シーン間の関係における、シーンのキーワード全てに対してシーンの関係を抽出することができ、より詳細に関係を判定することが可能である。また、シーンの関係を用いて、複数のコンテンツから、複数のシーン間の関係を判定し、適切な視聴区間を抽出する。

3. シーンの意味的關係

3.1 スライドと映像のメタデータ

本章では提案手法で扱う、コンテンツとそのメタデータについて説明する。コンテンツは、スライドと映像という同期した異種のメディアから構成される。スライドからはスライドに書かれた文字列とインデントによる構造化情報を得ることができ、ビデオからは音声のテキストが得られるものとする。スライドの切り替わりによって区切られる単位をシーンと呼ぶ。図 2 に本稿で扱うプレゼンテーションコンテンツの構造を示す。プレゼンテーションコンテンツのメタデータは、以下の要素で構成される。

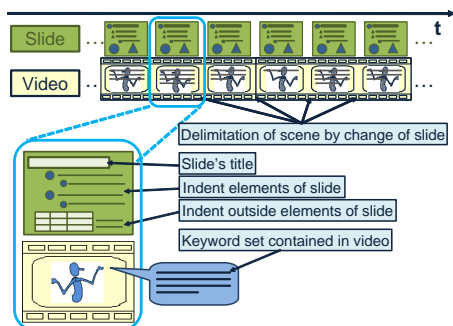


図 2 プレゼンテーションコンテンツの構造
Fig. 2 Metadata of presentation contents

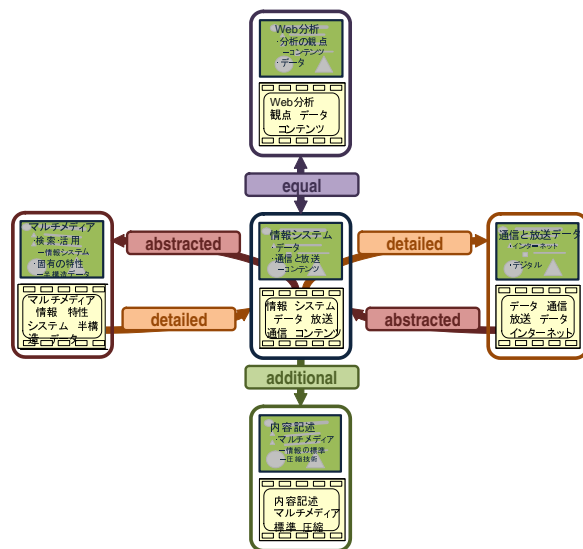


図 3 シーンの意味的關係
Fig. 3 Relation of semantic scene type

- スライドの切り替わりによるシーンの区切り
- スライド中に含まれるキーワード集合
- スライドでのインデントによる構造化情報
- ビデオ中に含まれるキーワード集合

本稿では、スライドのインデントによる階層構造を以下のよう定めた。

- スライドのタイトルに相当する要素を階層の最上位である 1 とする
- インデントの開始要素を 2 として、インデントが深くなるたびに 3, 4 のように 1 階層ずつ深くなるものとする
- 図や表などのインデント外となる要素をスライドの平均階層にあるとして扱う

3.2 意味的關係の定義

本手法では、シーンの意味的關係として以下の 4 種類を定義する。シーンの意味的關係の模式図を図 3 に示す。

detailed 基点シーンに対して、詳しい内容を説明している候補シーンとの関係と定義する。用いているスライドがある内容について特化しており、かつ、その内容に関して講演者が述べる量が多くなるという特徴がある。

abstracted 基点シーンに対して、概要を説明している候補シーンとの関係と定義する。用いているスライドがある内容について一般化した内容となり、かつ、その内容に関して講演者が述べる量が減少するという特徴がある。すなわち、detailed の逆方向の關係が abstracted であり、対となる關係といえる。

equal 基点シーンに対して、対等な内容を説明している候補シーンとの關係と定義する。用いているスライドの内容の詳細さが同じであり、かつ、その内容に関して講演者が述べる量が変化しないという特徴がある。

additional 基点シーンで説明された内容について、関連のないテーマのスライドのシーンにおいて説明している候補シーンとの關係と定義する。用いているスライド自体は關係がないため、基点シーンに出現したキーワードの記述量が減少し、か

つ、口頭での説明はされるため両方のシーンの映像にはそれらのキーワードが出現するという特徴がある。なお、additional は、指定したシーンに対して additional となるシーンを関係づけると効果的であることしか表わさないため、対となる関係は定義しない。

4. シーンの意味的関係の判定

意味的関係の判定手法について述べる。意味的関係は、スライドでの階層構造の変化やビデオでの発言の割合の変化を用いて判定を行う。判定には、各関係の特性を考慮した異なる条件式を用いる。以下、あるコンテンツ A における基点となるシーン番号を a_i とし、候補シーン番号を a_j とした場合の判定条件を説明する。

4.1 detailed・abstracted 判定

detailed 関係と abstracted 関係は対になる関係であり、detailed と判定することができる場合、その逆方向が abstracted 関係となるため、ここでは、detailed 判定に関して説明する。あるキーワードに関して、基点シーン a_i とシーン a_j が以下の条件式を満たし、他のシーンに関しての条件に合致した場合、2シーン間の関係を detailed と判定する。

$$|U(a_i, a_j)| > |S(a_i, a_j)| \quad (1)$$

$$|U(a_i, a_j)| > |D(a_i, a_j)| \quad (2)$$

$$\frac{v_num(U(a_i, a_j), a_i)}{v_count(a_i)} < \frac{v_num(U(a_i, a_j), a_j)}{v_count(a_j)} \quad (3)$$

基点コンテンツ A で、基点スライド a_i よりそれらのキーワードが候補スライド a_j で上位階層に出現するキーワードが多く、かつ、基点ビデオ a_i より候補ビデオ a_j で出現の割合が増加している場合に detailed と判定する。キーワードの出現階層の上位化を判断基準とするのは、スライドの構造で、上位階層に出現するキーワードはタイトルに近い扱いになると考えられ、スライドがそのキーワードに特化した内容になっていると判断できるためである。

$U(a_i, a_j)$ は、シーン a_i と a_j において、出現階層が上位と判定されるキーワードの集合を表す。 $S(a_i, a_j)$ は、シーン a_i と a_j において、出現階層が同じと判定されるキーワードの集合を表す。 $D(a_i, a_j)$ は、シーン a_i と a_j において、出現階層が下位と判定されるキーワードの集合を表す。基点スライド a_i より候補スライド a_j で出現階層が上位のキーワードの数が多いかどうかを条件式 (1), (2) により判定する。

基点ビデオ a_i より上位と判定されるキーワードが候補ビデオ a_j で出現している割合が増加しているかどうかを条件式 (3) により判定する。上位と判定されるキーワードが a_i 番目のビデオにおける出現の割合は、 $v_num(U(a_i, a_j), a_j)/v_count(a_i)$ により算出する。 $v_num(U(a_i, a_j), a_i)$ は、上位と判定されるキーワードの a_i 番目のビデオにおける出現数を返す関数である。 $v_count(a_i)$ は、 a_i 番目のビデオの全キーワードの出現数を返す関数である。

4.2 equal 判定

あるキーワードに関して、基点シーン a_i とシーン a_j が以下

の条件式を満たし、他のシーンに関しての条件に合致した場合、2シーン間の関係を equal と判定する。

$$|S(a_i, a_j)| > |U(a_i, a_j)| \quad (4)$$

$$|S(a_i, a_j)| > |D(a_i, a_j)| \quad (5)$$

$$\left| \frac{v_num(S(a_i, a_j), a_i)}{v_count(a_i)} - \frac{v_num(S(a_i, a_j), a_j)}{v_count(a_j)} \right| < \alpha \quad (6)$$

基点コンテンツ A で、基点スライド a_i よりそれらのキーワードが候補スライド a_j で同じ階層に出現するキーワードが多く、かつ、基点ビデオ a_i より候補ビデオ a_j で出現の割合が類似する場合に equal と判定する。

基点スライド a_i より候補スライド a_j で出現階層が同じしているキーワードの数が多いかどうかを条件式 (4), (5) により判定する。基点ビデオ a_i より候補ビデオ a_j で出現している割合が類似かどうかを条件式 (6) により判定する。出現階層が同じと判定されるキーワードの a_i 番目のビデオにおける出現の割合は、 $v_num(S(a_i, a_j), a_i)/v_count(a_i)$ により算出する。 $v_num(S(a_i, a_j), a_i)$ は、出現階層が同じと判定されるキーワードの a_i 番目のビデオにおける出現数を返す関数である。 $v_count(a_i)$ は、 a_i 番目のビデオの全キーワードの出現数を返す関数である。 α は、出現割合の差異に関する閾値である。

4.3 additional 判定

あるキーワードに関して、基点シーン a_i とシーン a_j が以下の条件式を満たし、他のシーンに関しての条件に合致した場合、2シーン間の関係を additional と判定する。

$$|inter(a_j, a_i)| < |differ(a_j, a_i)| \quad (7)$$

$$\frac{s_num(inter(a_j, a_i), a_i)}{s_count(a_i)} > \frac{s_num(inter(a_j, a_i), a_j)}{s_count(a_j)} \quad (8)$$

$$v_num(inter(a_j, a_i), a_i) > 0 \quad (9)$$

$$v_num(inter(a_j, a_i), a_j) > 0 \quad (10)$$

$$l(inter(a_j, a_i), a_i) < \frac{max_l(a_i)}{2}$$

$$\text{or } l(inter(a_j, a_i), a_j) > \frac{max_l(a_j)}{2} \quad (11)$$

基点コンテンツ A で、基点スライド a_i よりそれらのキーワードが候補スライド a_j で出現しない、または、基点ビデオ a_i より候補ビデオ a_j で基点スライドで出現しないキーワードの割合がある場合に additional と判定する。基点スライド a_i より候補スライド a_j で出現しないキーワード数が多いかを条件式 (7) により判定する。

$inter(a_j, a_i)$ は、シーン a_i と a_j において、共通と判定されるキーワードの集合を表す。 $differ(a_j, a_i)$ は、シーン a_i と a_j において、出現しないと判定されるキーワードの集合を表す。基点スライド a_i より共通と判定されるキーワードが候補スライド a_j で出現している割合が減少しているかどうかを条件式 (8) により判定する。共通と判定されるキーワードの a_i 番目のスライドにおける出現の割合は、 $s_num(inter(a_j, a_i), a_i)/s_count(a_i)$ により算出する。 $s_num(inter(a_j, a_i), a_i)$ は、共通と判定されるキーワードの a_i 番目のスライドにおける出現数を返す関数である。 $s_count(a_i)$ は、 a_i 番目のスライドの全キーワードの

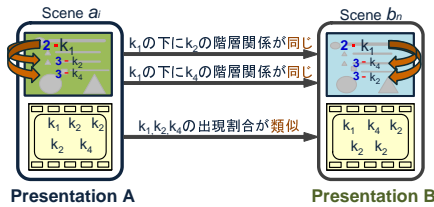


図 4 対応シーン判定の例

Fig. 4 Example of equivalent scene

出現数を返す関数である。また、条件式 (9)(10) により、共通と判定されるキーワードが基点ビデオ a_i と候補ビデオ a_j の両方に出現していることを判定する。条件式 (11) 中の $max.l$ は指定したシーンにおけるスライドの最深階層を返す関数であり、この条件式により、共通と判定されるキーワードが下位の階層から、上位の階層へ上位化していないことを判定する。

5. 視聴区間の決定

5.1 入力方式

本手法では、視聴中のコンテンツにおけるあるシーンの指定を入力とする。シーンを指定する入力には、単一シーンを指定する場合と、複数シーンを指定する場合がある。単一シーンを指定する入力の場合、特定のシーンにおいて、何らかの目的を持って他のシーンを得たいと考えられる。複数シーンを指定する入力の場合、ユーザはシーンの繋がりに関して他のシーンを得たいと考えられる。

単一シーンを指定する入力の場合、ユーザはシーンの入力に加えて、得たいシーンのタイプを選択することができる。このシーンと出力シーンのタイプを入力として、指定したシーンに対して、他のコンテンツから特定の関係に当たるシーンの集合を出力とする。出力を得る手順を以下に示す

- (1) ユーザが単一のシーンとタイプを指定
- (2) 他のコンテンツから単一のシーンに対応するシーンを決定
- (3) シーンの意味的關係を組み合わせたタイプの判定により、適切なシーンを出力

複数シーンを指定した入力の場合、指定したシーン間の関係に対して、他のコンテンツから、指定したシーン間に入るべきシーンとして特定の関係に当たるシーンの集合を出力とする。出力を得る手順を以下に示す

- (1) ユーザが複数のシーンを指定
- (2) 指定したシーン間の関係を判定
- (3) 指定したシーン間内、順序が先のシーンに対応する他のコンテンツのシーンを決定
- (4) 指定したシーン間の意味的關係により、必要なシーンを抽出することで、意味タイプ区間を出力

5.2 他コンテンツにおける対応シーンの決定

他コンテンツにおける対応シーンの決定について説明する。ここでは、ユーザが視聴中のコンテンツ A における指定シーンを a_i 、他のコンテンツ B における候補シーンを b_n とした場合の判定手法を説明する。基点シーン a_i と候補シーン b_n が以

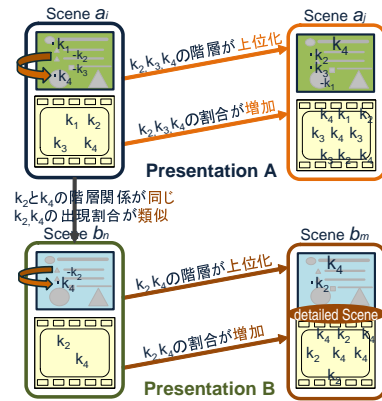


図 5 detailed タイプ検索の例

Fig. 5 Example of detailed relationship

下の条件式を満たす場合、他コンテンツにおける判定用の対応シーンと判定する。図 4 に、基点シーンとコンテンツ B のシーンを対応シーンと判定される例を示す。

$$\left| \frac{K(l, a_i) \cap K(l, b_n)}{\min(K(l, a_i), K(l, b_n))} \right| > \beta \quad (12)$$

$$SVratio(a, i, n) = \frac{s_num(K(l, a_i) \cap K(l, b_n), a_i)}{v_num(K(l, a_i) \cap K(l, b_n), a_i)} \quad (13)$$

$$|SVratio(a, i, n) - SVratio(b, i, n)| < \gamma \quad (14)$$

K は l 基点スライド a_i とスライド b_n における同じ階層のキーワード集合を返す関数であり、 $K(l, a_i) \cap K(l, b_n) / \min(K(l, a_i), K(l, b_n))$ はスライドの階層關係を考慮した基点スライド a_i とスライド b_n の類似度を算出する関数である。基点スライド a_i とスライド b_n の類似度は β により大きい、条件式 (12) により判定する。 β はスライドの相対類似度高いに関する閾値である。 $SVratio(a, i, n)$ はスライドにおける階層が同じキーワードの出現回数とそのキーワードのビデオにおける出現割合を算出する関数である。基点シーン a_i よりシーン b_n で出現している割合は条件式 (13) により算出する。基点シーン a_i よりシーン b_n で出現している割合が類似かどうかを条件式 (14) により判定する。 $s_num(K(l, a_i) \cap K(l, b_n), a_i)$ は、階層關係を維持しているキーワード $K(l, a_i) \cap K(l, b_n)$ の a_i 番目のスライドの出現回数を返す関数である。 $v_num(K(l, a_i) \cap K(l, b_n), a_i)$ は、階層關係を維持しているキーワード $K(l, a_i) \cap K(l, b_n)$ の a_i 番目のビデオの出現回数を返す関数である。 γ は、出現割合の差異に関する閾値である。

5.3 単一シーンと関係指定によるシーン検索

単一シーンを指定した場合のシーン検索を説明する。単一シーン指定の場合、指定されたシーンと指定されたタイプを用いて視聴区間の抽出を行う。意味的關係を組み合わせることで、各タイプの検索が可能である。以下にタイプの作成例を示す。detailed・abstracted タイプ検索 コンテンツ A で detailed の意味的關係があり、コンテンツ B で、detailed となる意味的關係を組み合わせることで detailed シーンタイプを定義する。

基点シーン a_i に対する対応シーン b_n を条件式 (12), (13), (14) により判定する。コンテンツ B で、対応シーン b_n から見た他のシーン b_m が条件式 (1), (2), (3) を満たし、2 シーン間

の関係を detailed と判定し、基点シーン a_i とシーン b_m の関係を detailed というシーンタイプと判定する。図 5 に、基点シーンとコンテンツ B のシーンの関係が detailed と検索される例を示す。

detailed 関係と abstracted 関係は対になる関係であり、detailed と判定することができる場合、その逆方向が abstracted 関係となる。

equal タイプ検索 コンテンツ A で equal の意味的關係があり、コンテンツ B で、equal となる意味的關係を組み合わせることで equal シーンタイプを定義する。

基点シーン a_i に対する対応シーン b_n を条件式 (12), (13), (14) により判定する。コンテンツ B で、対応シーン b_n から見た他のシーン b_m が条件式 (4), (5), (6) を満たす時、2 シーン間の関係を equal と判定し、基点スライド a_i とシーン b_m の関係を equal というシーンタイプと判定する。

additional タイプ検索 コンテンツ A で additional の意味的關係があり、コンテンツ B で、detailed となる意味的關係を組み合わせることで additional シーンタイプを定義する。

基点シーン a_i に対する対応シーン b_n を条件式 (12), (13), (14) により判定する。コンテンツ B で、対応シーン b_n から見た他のシーン b_m が条件式 (1), (2), (3) を満たす時、2 シーン間の関係を detailed と判定し、基点スライド a_i とシーン b_m の関係を additional というシーン意味タイプと判定する。

compared タイプ検索 コンテンツ A で detailed あるいは abstracted 意味的關係があり、コンテンツ B で、abstracted あるいは detailed となる意味的關係を組み合わせることで compared シーンタイプを定義する。

基点コンテンツ A の基点シーン a_i に対して、シーン a_j が detailed 判定となる場合を説明する。

compared の場合、指定シーンではなく、シーン a_j に対する対応シーン b_n を条件式 (12), (13), (14) により判定する。

最後に、対応シーン b_n から見た他のシーン b_m が abstracted の条件を満たす時、2 シーン間の関係を abstracted と判定し、compared というシーンタイプと判定する。

5.4 複数シーン指定によるシーン統合

複数シーン指定の場合における、意味タイプ区間の判定手法について述べる。意味タイプ区間は、基点コンテンツの指定したシーン間の関係を用いて、複数コンテンツから関係に関して意味タイプ区間の判定を行う。指定したシーン間の関係の判定は意味タイプの判定手法の中で、基点コンテンツ A の基点シーンと候補シーンが同様な判定手法を用いて判定する。また、コンテンツ B で指定したシーン間の同じ関係である必要なシーンを抽出し、指定したシーン間によって、適切な位置に入り、意味タイプ区間の生成について説明を行う。本稿では、意味タイプ区間として detailed 区間、abstracted 区間、equal 区間、additional 区間の判定を行う。図 6 に、コンテンツ B から必要なシーンの入る位置に detailed 区間と判定される例を示す。

detailed シーン区間 コンテンツ A で指定したシーン間に対して detailed 意味的關係があり、コンテンツ B で、detailed、abstracted となる意味的關係を組み合わせることで detailed

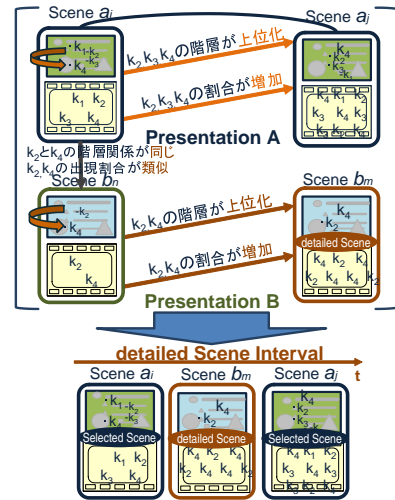


図 6 detailed 区間判定の例

Fig. 6 Example of detailed interval

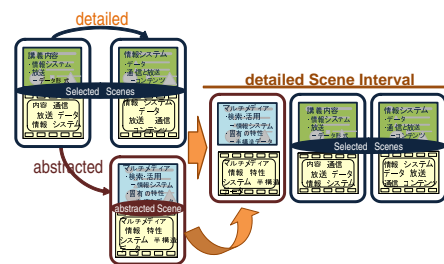


図 7 指定した 2 つのシーンの前に入るシーンの detailed シーン区間

Fig. 7 Detailed interval of semantic scene in front of the appointed scenes

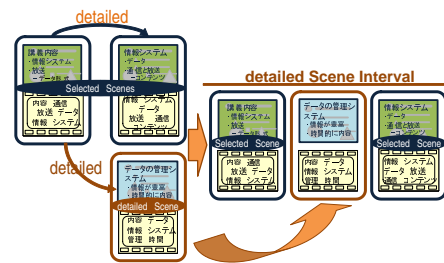


図 8 指定したシーン間に入るシーンの detailed シーン区間

Fig. 8 Detailed interval of semantic scene between the appointed scenes

シーンタイプ区間を定義する。

指定したシーン間に入るシーンとして適切に働くような視聴区間となるように他のコンテンツからシーンを抽出する。指定した 2 つのシーンと、他のコンテンツのシーンを用いることで、順を追って詳しくなる説明となる区間を生成する。他のコンテンツから基点シーンに対応するシーンに対して、詳しい内容、概要を説明している他のシーンを抽出する。このことにより、指定したシーンの内容に関して、詳細さを理解しやすい区間を自動的生成することが可能となる。複数シーンを指定した入力の場合、detailed シーン区間における、それぞれのシーン挿入の例を、図 7、図 8、図 9 に示す。

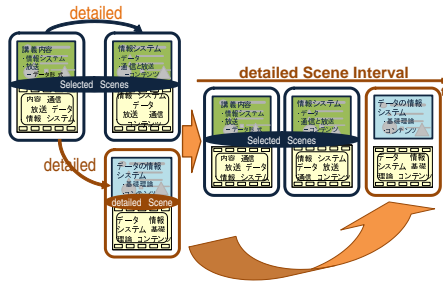


図 9 指定した 2 つのシーンの後に入るシーンの detailed シーン区間
Fig. 9 Detailed interval of semantic scene behind the appointed scenes

abstracted シーン区間 コンテンツ A で指定したシーン間に対して abstracted 意味の関係があり、コンテンツ B で, detailed, abstracted となる意味的关系を組み合わせることで abstracted シーンタイプ区間を定義する。

指定したシーン間に入るシーンとして適切に働くような視聴区間となるように他のコンテンツからシーンを抽出する。指定した 2 つのシーンと、他のコンテンツのシーンをを用いることで、概要をより理解しやすい区間を生成する。他のコンテンツから基点シーンに対応するシーンに対して、詳しい内容、概要を説明している他のシーンを抽出する。このことにより、指定したシーンの内容に関して、概要をより理解しやすい区間を自動的に生成することが可能となる。

equal シーン区間 コンテンツ A で指定したシーン間に対して equal 意味的关系があり、コンテンツ B で, equal となる意味的关系を組み合わせることで equal シーンタイプ区間を定義する。

指定したシーン間に入るシーンとして適切に働くような視聴区間となるように他のコンテンツからシーンを抽出する。指定した 2 つのシーンと、他のコンテンツのシーンをを用いることで、順を追って対等な内容を説明となる区間を生成する。他のコンテンツから基点シーンに対応するシーンに対して、対等な内容を説明している他のシーンを抽出する。このことにより、指定したシーンの内容に関して、対等な説明を追加した区間を自動的に生成することが可能となる。

additional シーン区間 コンテンツ A で指定したシーン間に対して additional 意味的关系があり、コンテンツ B で, detailed となる意味的关系を組み合わせることで additional シーンタイプ区間を定義する。

指定したシーン間に入るシーンとして適切に働くような視聴区間となるように他のコンテンツからシーンを抽出する。指定した 2 つのシーンと、他のコンテンツのシーンをを用いることで、順を追って関連のないテーマのスライドのシーンにおいて説明となる区間を生成する。他のコンテンツから基点シーンに対応するシーンに対して、詳しい内容を説明している他のシーンを抽出する。このことにより、指定したシーンの内容のうち付加された内容に関して詳細に説明することが可能な区間を自動的に生成する。

図 10 に構築したプロトタイプシステムの画面イメージを



図 10 プロトタイプシステムの画面
Fig. 10 Screen image of the prototype system

表 1 実験コンテンツ
Table 1 Experimental contents

組	プレゼンテーションコンテンツ	
1	ニュース映像のシーン順序に基づく Blog 検索 ニュース構成パターンに基づくビデオアーカイブコンテンツ閲覧方式	DEWS2005 2C-i9 DEWS2006 5C-i4
2	地理情報の詳細度を考慮した移動オブジェクト群への情報配信 状況依存型実時間地図合成システムの設計と実現	DEWS2003 1-C-1 DEWS2003 4-P-1
3	注視点を考慮したデジタル写真の検索インタフェース マーケティングマップによる空間認知に基づいた地域情報推薦手法	DEWS2003 6-P-3 DEWS2003 1-C-3
	空間関係を利用した写真と注釈の相補的な共有モデル	DEWS2005 3A-o1

示す。

6. 評価

6.1 プロトタイプシステム

プロトタイプシステムは提案手法に基づき Visual Studio 2008 C# によって開発する予定である。プロトタイプシステムは、ユーザの入力部、シーン間の関係判定部と視聴区間出力部によって構成される。シーン間の関係判定部には、シーン意味タイプ判定部とシーン意味タイプ区間判定部によって構成される。シーン意味タイプ判定部では、指定したシーンに含まれる全てのキーワードについて、シーンの総当りで、各シーン意味タイプを判定する。シーン意味タイプ区間判定部では、シーン意味タイプ判定部により得られた結果を用いて、各シーン意味タイプ区間を判定する。視聴区間出力部は、ユーザの入力に基づいて、ユーザに提示するシーン群、シーン順序を決定する。スライド、映像からの単語抽出は SlothLib [4] [9] より呼び出した茶筌 [2] [3] による形態素解析を用いて行った。

プロトタイプシステムの画面は、入力画面と出力画面からなる。入力画面では、スライド画像の一覧と、画像の横にシーンで出現するキーワードリストを表示する。出力画面では単一シーンと関係を指定した場合、指定したシーンと関係によるシーンのスライド画像を出力する。複数シーン指定した場合、複数シーンによる統合区間のシーンのスライド画像を出力する。Play ボタンがクリックされた場合、別ウィンドウで抽出区間を再生する。

6.2 評価実験 1: 意味的关系判定の妥当性

本手法で定義しているシーンの意味的关系の判定の妥当性を評価するために実験を行った。被験者はシーン間の関係が、detailed, abstracted, equal, additional, others の 5 種のい

表 2 判定結果の分類

Table 2 Experimental result

		システム判定結果				
		detailed	abstracted	equal	additional	others
正解	detailed	52	12	5	3	22
	abstracted	23	81	14	10	42
	equal	13	9	25	3	16
	additional	6	7	13	17	22
	others	27	24	12	8	48

表 3 意味的關係判定の評価

Table 3 Result of detected relation type

	detailed	abstracted	equal	additional	all
Precision	43.0% (52/121)	60.9% (81/133)	36.2% (25/69)	41.5% (17/41)	48.1% (175/364)
Recall	68.4% (52/76)	63.3% (81/128)	50.0% (25/50)	37.8% (17/45)	58.5% (175/299)
F-measure	0.53	0.62	0.42	0.40	0.53

ずれに該当するかを判定した。コンテンツ内の 2 シーンと、2 シーンに出現するキーワードを被験者に提示することで、自由記述による、あるキーワードに関するシーンの関係を得た。被験者は 5 名であり、3 組の 7 つコンテンツ^(注1)からランダムにサンプリングした 115 個のシーン間の関係を判定した。表 1 に実験に用いたコンテンツのタイトルを示す。3 名以上が自由記述により何らかの関係があると記述したシーンを正解とし、システムがいずれかの意味的關係と判定した結果を用い、適合率・再現率により評価を行った。判定結果の分類を表 2 に示す。表 2 では、縦にシステム判定結果の数、横に被験者判定結果による正解の数を示した。表中の下線が引かれた数値は、システムの解と正解が一致した解の数である。また、システム解における others とは、判定不能であったデータである。結果を表 3 に示し、考察を以下に述べる。

- detailed の判定では、適合率が低く、再現率が高い結果を得た。実験で detailed のシステム解は、判定結果の分類において他のタイプの正解を多数含む傾向が見られた。detailed の定義は妥当であると考えられる。例えば、「まとめ」というシーンで見たときに、「ニュース構成パターンの抽出」というシーンでは「構成パターンの抽出」という点で detailed であり、システムも正しく判定していた。一方、「本研究のアプローチ」というシーンで見たときに、「ニュース間の関係の決定」というシーンでは「ニュース間の関係の決定」という点で detailed であるが、システムは abstracted と判定しており、シーン中のあるキーワード出現数を比較する際に、あるキーワードと同階層のキーワードに範囲を広げるなど、キーワードの判定方式を改良する必要があると考えられる。

- abstracted の判定では、適合率が低く、再現率が高い結果を得た。実験で abstracted のシステム解は、判定結果の分類において他のタイプの正解を多数含む傾向が見られた。例えば、「順序による重要度付け」というシーンで見たときに、「まとめ」というシーンでは「出現順序による重要度」という点で abstracted であり、システムも正しく判定していた。一方、「以前の研究」というシーンで見たときに、「発表の流れ」という

シーンでは「以前の研究」という点で abstracted であるが、システムは equal と判定しており、シーン中のあるキーワード出現数を比較する際に、あるキーワードと同階層のキーワードに範囲を広げるなど、キーワードの判定方式を改良する必要があると考えられる。一方、実験において比較的 abstracted 判定に関する F 値が高いことから、abstracted の定義は妥当であると考えられる。

- equal の判定では、適合率が低く、再現率が高い結果を得た。実験で equal のシステム解は、判定結果の分類において他のタイプ特に abstracted の正解を多数含む傾向が見られた。equal の定義は妥当であると考えられる。例えば、「以前の研究 (3)」というシーンで見たときに、「以前の研究 (2)」というシーンでは「地理情報」と「グループ化」という点で equal であり、システムも正しく判定していた。一方、「従来手法による合成結果」というシーンで見たときに、「単一メタファによる合成結果」というシーンでは「合成結果」という点で equal であるが、システムは abstracted と判定しており、シーン中のあるキーワード出現数を比較する際に、あるキーワードと同階層のキーワードに範囲を広げるなど、キーワードの判定方式を改良する必要があると考えられる。

- additional の判定では、適合率が高い結果を得ることができた。よって、additional の定義が適切であると考えられる。一方で、再現率は低い結果を得た。また、多くの正解がシステムにおいて、いずれのタイプにも判定されず、無関係となる傾向が見られた。例えば、「配信データのスケジューリング」というシーンで見たときに、「以前の研究 (4)」というシーンでは「移動オブジェクト」と「グループ」という点で additional であり、システムも正しく判定していた。一方、「以前の研究 (3)」というシーンで見たときに、「問題点」というシーンでは「位置情報」という点で equal であり、「移動方向」という点で abstracted であるが、システムは additional と判定していた。これは、説明の度合いとキーワードの出現割合が一致しないことが一因と考えられる。そのため、スライド中のあるキーワード出現数を比較する際に、あるキーワードと同階層のキーワードに範囲を広げるなど、キーワードの判定方式を改良する必要があると考えられる。

以上のことより、シーンの意味的關係のうち、detailed, abstracted, equal, additional は適切に定義が行えていることを確認した。ただし、additional に関しては、定義の追加を検討する必要があると考えられる。

6.3 評価実験 2: シーンタイプ区間判定の妥当性

本手法で定義しているシーンタイプ区間の判定の妥当性を評価するために実験を行った。被験者はシーン間の関係が、detailed タイプ区間、abstracted タイプ区間、equal タイプ区間、additional タイプ区間の 4 種のいずれに該当するかを判定した。被験者は 5 名であり、実験データは、評価実験 1 と同様のデータを用いて、計 50 個のシーンタイプ区間の妥当性を判定した。その結果を正解として、システムの解を用いた適合率・再現率により評価を行った。結果を表 4 に示し、考察を以下に述べる。

(注1): DBSJ Archives:

<http://www.dbsj.org/Japanese/Archives/archivesIndex.html>

表 4 シーンタイプ区間判定の評価
Table 4 Result of detected relation interval type

	detailed	abstracted	equal	additional	all
Precision	43.8% (14/32)	34.8% (8/23)	55.6% (10/18)	25.0% (1/4)	42.9% (33/77)
Recall	73.7% (14/19)	88.9% (8/9)	83.3% (10/12)	11.1% (1/9)	67.3% (33/49)
F-measure	0.55	0.50	0.67	0.15	0.52

- detailed タイプ区間の判定では、適合率が低く、再現率が高い結果を得た。実験で detailed タイプ区間のシステム解は、挿入したシーンが他のタイプの正解を多数含む傾向が見られた。detailed タイプ区間の定義は妥当であると考えられる。例えば、「注視点を考慮したデジタル写真の検索インタフェース」というコンテンツの中で、「結論」というシーンと「空間スキーマ付写真への空間ラベル取得および写真上位置算出」というシーンでは detailed であり、と「空間関係を利用した写真と注釈の相補的な共有モデル」というコンテンツの中で、「対象データ」というシーンではその 2 つのシーンの後に入り、「空間スキーマ」と「空間ラベル」という点で detailed であり、その生成区間が detailed であり、システムも正しく判定していた。一方、「ニュース映像のシーン順序に基づく Blog 検索」というコンテンツの中で、「まとめ」というシーンと「順序による重要度付け」というシーンでは detailed であり、と「ニュース構成パターンに基づくビデオアーカイブコンテンツ閲覧方式」というコンテンツの中で、「ニュース間の関係の決定」というシーンではその 2 つのシーンの後に入り、「出現順序」と「対象」という点で abstracted であり、その生成区間が detailed ではないが、システムは detailed と判定していた。そのシーンが 2 つのシーンの前に入ったら、その生成区間が detailed であるが、シーンの挿入位置の判定を改良により、detailed タイプ区間判定の適合率も改善できると考える。

- abstracted タイプ区間の判定では、適合率が低く、再現率が高い結果を得た。実験で abstracted タイプ区間のシステム解は、挿入したシーンが他のタイプ特に detailed の正解を多数含む傾向が見られた。例えば、「空間関係を利用した写真と注釈の相補的な共有モデル」というコンテンツの中で、「空間スキーマ付写真への空間ラベル取得および写真上位置算出」というシーンと「結論」というシーンでは abstracted であり、と「注視点を考慮したデジタル写真の検索インタフェース」というコンテンツの中で、「撮影ベクトル」というシーンではその 2 つのシーンの前に入り、「写真の視点」という点で abstracted であり、その生成区間が abstracted であり、システムも正しく判定していた。一方、「空間関係を利用した写真と注釈の相補的な共有モデル」というコンテンツの中で、「ユーザーにとっての視点と注視点の重要度」というシーンと「人間による写真の読み取り」というシーンでは abstracted であり、と「空間スキーマ付写真への空間ラベル取得および写真上位置算出」というコンテンツの中で、「地理的位置・写真上位置が既知の 3 組以上の特

徴点を利用した写真の空間スキーマ算出」というシーンではその 2 つのシーンの前に入り、「空間写真」と「位置」という点で detailed であり、その生成区間が detailed ではあるが、システムは abstracted と判定しており、意味的關係判定アルゴリズムの改良により、abstracted タイプ区間判定の適合率も改善できると考える。

- equal タイプ区間の判定では、適合率が低く、再現率が高い結果を得た。実験で equal タイプ区間のシステム解は、挿入したシーンが他のタイプ特に others の正解を多数含む傾向が見られた。例えば、「注視点を考慮したデジタル写真の検索インタフェース」というコンテンツの中で、「動機」というシーンと「位置情報付き写真に着目する社会的・技術的背景」というシーンでは equal であり、「マーキングマップによる空間認知に基づいた地域情報推薦手法」というコンテンツの中で、「マーキングマップの提案 (1/2)」というシーンではその 2 つのシーンの後に入り、「空間認知」という点で equal であり、その生成区間が equal であり、システムも正しく判定していた。一方、「注視点を考慮したデジタル写真の検索インタフェース」というコンテンツの中で、「動機」というシーンと「位置情報付き写真に着目する社会的・技術的背景」というシーンでは equal であり、と「空間関係を利用した写真と注釈の相補的な共有モデル」というコンテンツの中で、「結論」というシーンではその 2 つのシーンの前に入り、「空間ラベル」と「空間スキーマ」という点で abstracted であり、その生成区間が abstracted ではあるが、システムは abstracted と判定しており、意味的關係判定アルゴリズムの改良により、equal タイプ区間判定の適合率も改善できると考える。一方、実験において比較的 equal タイプ区間判定に関する F 値が高いことから、equal タイプ区間の定義は妥当であると考えられる。

- additional タイプ区間の判定では、適合率が高い結果を得ることができた。よって、additional タイプ区間の定義が適切であると考えられる。一方で、再現率は低い結果を得た。また、多くの正解がシステムにおいて、いずれのタイプ区間にも判定されず、無効となる傾向が見られた。例えば、「注視点を考慮したデジタル写真の検索インタフェース」というコンテンツの中で、「写真と空間とを関連付ける幾何モデル」というシーンと「同一の視点で注視点異なる」というシーンでは additional であり、「空間関係を利用した写真と注釈の相補的な共有モデル」というコンテンツの中で、「結論」というシーンではその 2 つのシーンの後に入り、「空間ラベル」と「位置」という点で additional であり、その生成区間が additional であり、システムも正しく判定していた。一方、「ニュース構成パターンに基づくビデオアーカイブコンテンツ閲覧方式」というコンテンツの中で、「まとめ」というシーンと「本研究のアプローチ」というシーンでは detailed であり、と「ニュース映像のシーン順序に基づく Blog 検索」というコンテンツの中で、「まとめ」というシーンではその 2 つのシーン間に入り、と「ニュース構成パターンに基づくビデオアーカイブコンテンツ閲覧方式」というコンテンツの「まとめ」というシーンが「出現順序」という点で additional であり、その生成区間が additional ではない

いが、システムは additional と判定しており、additional タイプ区間判定アルゴリズムの改良する必要があると考えられる。また、実験用データは適切なデータが少ないと考えられる。以上のことより、シーンタイプ区間のうち、detailed タイプ区間、abstracted タイプ区間、equal タイプ区間、additional タイプ区間は適切に定義が行えていることを確認した。ただし、additional タイプ区間に関しては、定義の追加を検討する必要があると考えられる。また、適切なデータを用いて評価を行う必要がある。

7. おわりに

本稿では、複数コンテンツにおけるシーン間の関係やシーン意味タイプ区間を定義、解析し、視聴区間を抽出する方式について提案を行った。その中で、他コンテンツのシーンを抽出するための意味的關係を定義し、それらの関係を組み合わせたシーン検索手法について説明した。このことにより、意味的につながりのあるシーンや区間の抽出が可能である。

評価実験の結果、意味的關係の判定による他コンテンツのシーンを抽出として、適応できるコンテンツに関して考察を行った。また、各判定に関しては十分な精度が得られていないため、意味的關係の定義についての改良が必要であることを確認した。スライドにおける階層変化やビデオにおける出現の割合だけでなく、スライドにおける字の色や配置、ビデオにおける音量の変化などを考慮し、シーン間の変化を解析することで、判定を行う必要があると考える。また、本稿ではシーン中のあるキーワードに対して判定を行っているが、複数コンテンツの内容から意味的關係の組み合わせ判定を行う必要がある。今後の課題としては、意味的關係やシーンタイプ区間判定アルゴリズムの改良が必要である。また、シーンタイプやシーンタイプ区間を提示することの有効性も評価を行う必要がある。さらに、ユーザの視聴操作に基づく区間判定方式への拡張を行う必要がある。

謝 辞

本研究の一部は、平成 20 年度科研費基盤研究 (B)(2) 「ユーザの潜在的意図を用いたレス・コンシャス情報検索基盤の構築」(課題番号: 20300039) によるものです。ここに記して謝意を表すものとします。

文 献

- [1] Ricoh Corporation: MPmeister, <http://www.ricoh.co.jp/mpmeister/>.
- [2] Japanese Morphological Analysis System ChaSen: <http://chasen.naist.jp/hiki/ChaSen/>.
- [3] Asahara, M., Matsumoto, Y. Extended Models and Tools for High-performance Part-of-Speech Tagger. In Proc. of The 18th International Conference on Computational Linguistics(COLING 2000), pp. 21-27, 2000.

- [4] SlothLib. <http://www.dl.kuis.kyoto-u.ac.jp/slothlib/>.
- [5] Smith, M.A. and Kanade, T.: Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques, Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR 07), 1997.
- [6] Pradhan, S., Tajima, K. and Tanaka, K.: A Query Model for Retrieving Relevant Intervals within a Video Stream, Proc. of the 6th IEEE Int'l Conference on Multimedia Computing and Systems (ICMCS 99), Vol.2, pp.788-792, 1999.
- [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] 戈 指夷, 角谷和俊. 遠隔会議システムにおける資料操作ログに基づくアーカイブコンテンツ作成支援方式. 電子情報通信学会第 17 回データ工学ワークショップ (DEWS2006) 論文集, 6C-i5, 2006.
- [9] 大島裕明, 中村聡史, 田中克己. SlothLib:Web サーチ研究のためのプログラミングライブラリ. 日本データベース学会 Letters, Vol.6, No.1, pp. 113-116, 2007.
- [10] 北山大輔, 大谷亜希子, 平元綾子, 角谷和俊. プレゼンテーションコンテンツのための意味的シーンタイプに基づく視聴区間検索. データベースと Web 情報システムに関するシンポジウム (DBWeb2007), 2007.
- [11] レーヒェウハン, ルートラットデーチャークン, ティティポーン, 渡部 徹太郎, 横田 治夫. 講義講演ビデオからダイジェスト自動作成のための重要シーン抽出手法の評価. 電子情報通信学会第 19 回データ工学ワークショップ (DEWS2008) 論文集, E4-1, 2008.
- [12] 大谷亜希子, 北山大輔, 平元綾子, 角谷和俊. プレゼンテーションコンテンツのためのシーンの意味的關係を用いた視聴区間抽出方式. 電子情報通信学会 第 19 回データ工学ワークショップ (DEWS2008) 論文集, E4-5, 2008.
- [13] 北山大輔, 角谷和俊. プレゼンテーションコンテンツのためのコンテキストと視聴時行為に基づく区間検索方式. 電子情報通信学会 第 19 回データ工学ワークショップ (DEWS2008) 論文集, E4-6, 2008.