

論文 / 著書情報
Article / Book Information

論題(和文)	講演コンテンツにおける視覚的效果に基づくダイジェストの自動生成
Title(English)	Automatic Digesting of Presentation Contents based on the Visual Effects
著者(和文)	呉怡, 渡辺陽介, 横田治夫
Authors(English)	Yi Wu, Yousuke Watanabe, Haruo Yokota
出典(和文)	DEIM Forum 2010 D4-3, , ,
Citation(English)	DEIM Forum 2010 D4-3, , ,
発行日 / Pub. date	2010, 3

講演コンテンツにおける視覚的効果に基づくダイジェストの自動生成 (O)

呉 怡[†] 渡辺 陽介^{††} 横田 治夫^{†,††}

[†] 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

^{††} 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

E-mail: [†]{goi,watanabe}@de.cs.titech.ac.jp, ^{††}yokota@cs.titech.ac.jp

あらまし 近年、講演・講義の様子を収録したプレゼンテーションコンテンツが大量に蓄積されてきた。しかし、大量の講演動画を視聴し、その中から欲しい情報を見つけ出すことは多くの時間を要する。そのような状況に対して、我々はキーワードに合致する複数の講演コンテンツから、重要なシーンを抽出し、講演の概要となるダイジェストを自動生成するシステムを構築している。本稿では、利用者の興味を引き出すダイジェストの提供を目的とし、講演動画における視覚的効果の高いシーンを重要視する。提案手法では、スライドにおける色分布やアニメーションの使用回数などの情報を用いて、シーンの重要度を算出する。さらに被験者実験で、手法の有効性を確認する。
キーワード eラーニング, 映像要約, プレゼンテーションコンテンツ

Automatic Digesting of Presentation Contents based on the Visual Effects

Yi WU[†], Yousuke WATANABE^{††}, and Haruo YOKOTA^{†,††}

[†] Department of Computer Science, Graduate School of Information Science and Engineering
2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552 JAPAN

^{††} Global Scientific Information and Computing Center, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552 JAPAN

E-mail: [†]{goi,watanabe}@de.cs.titech.ac.jp, ^{††}yokota@cs.titech.ac.jp

Abstract A large amount of presentation contents which record presentation slides and videos have been accumulated in archives to be provided to the Internet. However, it takes users a long time to find what they really want to get from a lot of presentation videos. Thus, we have been building a system to summarize multiple presentation contents that match given keywords. In this paper, we intend to obtain an interesting digest including visually attractive scenes instead of generating a sketch of presentation contents. We propose methods to calculate the importance of scenes by considering color distribution and animation occurrence. Finally, we evaluate the effectiveness of the proposed methods.

Key words E-learning, Video Digest, Presentation Contents

1. はじめに

近年、プレゼンテーションコンテンツの作成を支援するツールが多く開発されてきた。それにより、マルチメディア情報を含むコンテンツが容易に作成できるようになって、講演・講義の様子を収録した動画アーカイブが大量に蓄積されるようになってきた。例えば、奈良先端科学技術大学院大学では平成16年度から現在までの約6年間にわたる授業アーカイブを蓄積している[1]。また、日本データベース学会[2]では過去に行なわれた研究シンポジウムにおける発表の様子を収録した1,200件以上の講演コンテンツを会員向けに公開している。しかし、長

時間にわたる動画の中から、欲しい情報を見つけ出すことは多くの時間を要する。

そのような状況を鑑み、我々の研究グループはまず、統合プレゼンテーションコンテンツを蓄積し、その中からキーワードに関連する動画シーンを検索するシステム UPRISE (Unified Presentation Slides Retrieval by Impression Search Engine)[3]を提案した。UPRISEでは、スライドの切り替わるタイミングを利用し、動画をシーンに分割したあと、シーンのメタデータを利用して、キーワードに対するシーンの適合度を算出している。次に、ユーザが短時間で講義講演の内容を把握できるように、1件の講義講演コンテンツから、そのトピックをよく表す

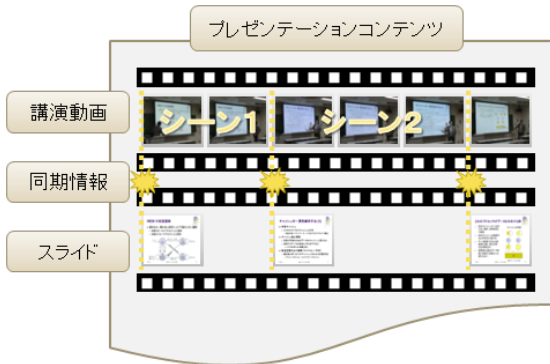


図1 プレゼンテーションコンテンツの構成

シーンを抽出する手法を提案した [4] .

一方、講演コンテンツに対しては、あるトピックに関連する複数の講演の内容を簡潔にまとめるように、講演動画アーカイブスから、キーワードに合致する複数の講演コンテンツの概要となるダイジェストを自動生成するシステムを構築している [5] . しかしながら、複数の講演のダイジェストとして、要点がよく説明されるシーンが求められている一方、限られた時間の中、講演の魅力を視聴者に伝えることもその役割のひとつだと考える . それを実現するには、文字の羅列からなるシーンよりも、図やアニメーションを使って説明するような映像ならではの伝え方を生かしたシーンのほうが適している . だが、これまでの手法 [5] では、主に単語の出現状況をベースとしていたため、視覚的効果の高くても、文字数の少ないシーンへの評価が常に低いという問題点があった .

そこで本稿では、スライド内の色分布やアニメーションの使用回数などの情報を用いて、視聴者の興味を引くようなダイジェストを生成する手法を提案する . 提案手法では、まずスライドのサムネイル内の色分布を調べ、講演ごとに背景の検出を行なって、その影響を除去したあと、シーンの重要度算出に利用している . また、アニメーション使用回数に基づくシーン重要度算出式を提案し、さらに、色分布情報とアニメーション使用回数との組み合わせについて考える . 実験では、被験者によるシーンの点数評価に基づく正解セットを用いて、評価を行なう .

以下に本稿の構成を述べる . 2. 節で、プレゼンテーションコンテンツに対して、我々のこれまでの研究内容について紹介し、3. 節において、講演コンテンツから視覚的効果の高いシーンを抽出する手法について述べて、4. 節で提案手法で抽出したダイジェストシーンの一例を示す . 5. 節では、被験者実験による手法の評価を行う . そして 6. 節で、関連システムについて議論し、最後に 7. 節で、まとめと今後の課題を述べる .

2. これまでの取り組み

我々はこれまでに、講義などで使用されるプレゼンテーション資料と動画とを統合したプレゼンテーションコンテンツに対し、高度な検索機能を提供するシステム UPRISE [3] を提案し、[6] において実装を行なって、有効性を示した . UPRISE では、スライドの切り替えタイミングからシーンを定義し、シー

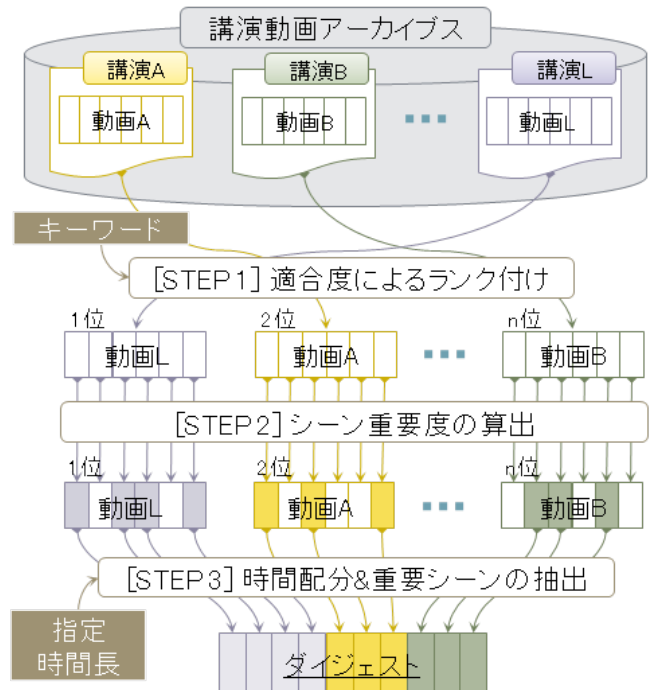


図2 ダイジェスト自動生成するまでの流れ

ンが対応するスライド内のテキスト情報から、動画シーンの検索を実現した . 検索を行なう際に、単語の出現回数と出現位置、スライドの提示時間及びシーンの前後関係といったプレゼンテーションの特性を考慮した適合度指標を用いている .

また、十分な時間の余裕がない利用者に対し、1 件のプレゼンテーションコンテンツのダイジェストを自動的に作成する手法を提案し、実際の講義コンテンツを用いて評価を行なった [4] .

そして、利用者が短時間で、あるトピックに関連する講演の概要を把握できるように、キーワードに合致する複数の講演コンテンツを含むダイジェストを自動生成するシステムを構築している . ダイジェストを自動生成するまでの流れ (図 2) として、まず既存手法の tf.idf [7] を用いて、キーワードに関連するコンテンツをランク付けする (ステップ 1) . 次に、講演コンテンツの特性を考慮し、各講演に含むシーンの重要度を算出する (ステップ 2) . 最後に、利用者が指定する時間長に納まるように時間配分を行い、その結果に基づき、重要なシーンを抽出し、ダイジェストを生成する (ステップ 3) .

しかし、以前提案した重要度指標 [5] では主にシーンに含まれる単語の出現頻度に基づいたもののため、重要な単語を多く含むシーンへの評価が高く、図やグラフを用いた視覚的効果の高いシーンに対しては常に重要度が低い結果となってしまうため、利用者の興味を引き出すダイジェストの生成には不向きだと考える . そこで本稿では、図やグラフを使って説明するシーンを重要とし、色分布やアニメーションの使用回数などの情報を考慮した重要度を利用して、ダイジェストの自動生成を試みる . なお、本稿でダイジェストを自動生成する際の処理のステップ 1 およびステップ 3 についてはこれまでと同様であるため、論文 [5] を参照されたい .

```

- <AudioVisualSegment id="audioVisualSegment-47">
  <Title>Slide 110</Title>
  <TextAnnotation type="headline" xml:lang="en">パイプラインの性能向上 (2)
  </TextAnnotation>
  <TextAnnotation type="commentary" xml:lang="en">スーパースカラ
  (Superscalar) をステージで同時に実行できる命令を増やす同時に2命令実行す
  ると2倍にスーパーパイプラインvs. スーパースカラ実装の容易さ、多重度向上の容易さ
  組み合わせ</TextAnnotation>
  <MediaRelTimePoint>P0DT0H3M33S500N1000F</MediaRelTimePoint>
  <MediaIncrDuration>81200</MediaIncrDuration>
- <AudioVisualSegment id="audioVisualSegment-51">
  <Title>Animation 0</Title>
  <MediaRelTimePoint>P0DT0H3M33S500N1000F</MediaRelTimePoint>
  <MediaIncrDuration>8100</MediaIncrDuration>
</AudioVisualSegment>
...
</AudioVisualSegment>

```

図3 MPMeisterによるメタ情報の記述例

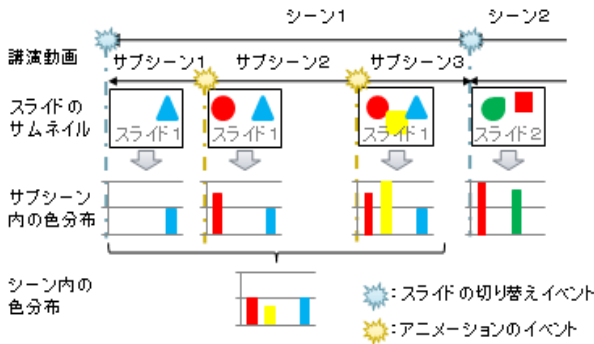


図4 イベントによるシーンの分割

3. 視覚的効果の高いシーンの抽出手法

我々は利用者の興味を引き出すには、動画ならではの説明部分をダイジェストに含む必要があると考え、スライド内の色分布やアニメーションの使用回数を利用して、視覚的効果の高いシーンの抽出手法を提案する。

手法の適用対象として、MPMeister [8] を使用して記録されたコンテンツを用いる。MPMeister はスライドの切り替えやアニメーションのイベントが発生した時のスライドの様子をサムネイルに記録する。また、そのイベントに対応するスライド資料にあるテキスト情報、スライドの提示時間およびイベントの開始時間などのメタ情報を MPEG-7 ファイルで記述している (図3)。

以下では、アニメーションのイベントによって分割されたシーンの断片をサブシーンという。サブシーンごとにスライドの表示画面のサムネイル (図4) がある。

3.1 色分布を用いた重要度

講演コンテンツにある視覚的効果の高いシーンには、図やグラフを用いて説明するスライドが多く、また色遣いがカラフルである特徴があるため、ここでは、スライド内の色分布を用いて、視覚的効果の高いシーンの抽出手法を述べる。まず前処理において、各講演コンテンツのスライドのサムネイルに対し、RGB 色空間における R ビンの色ヒストグラムを求める。スライドのテンプレートの影響をなくするため、背景の検出を行ない、背景色を除去した画素値を用いている。

N 個のサブシーンからなるシーン c に対して、色 i ($1 \leq i \leq R$) のサブシーン j ($1 \leq j \leq N$) のサムネイルにおける画素値を $r_{i,j}$ とした時に、シーン c における色 i の画素値 r_i を

表1 色分布を用いた重要度

	色平均	色の二乗平均	色種類
黒・グレーを考慮する	I_{avg_color}	I_{ms_color}	I_{var_color}
黒・グレーを考慮しない	I'_{avg_color}	I'_{ms_color}	I'_{var_color}

$r_i = \frac{1}{N} \sum_{j=1}^N r_{i,j}$ で求め、スライド内の色分布の相加平均をシーン重要度とする I_{avg_color} と、各色の二乗平均を重要度とする I_{ms_color} と、色の種類数を重要度とする I_{var_color} を定義する。

$$I_{avg_color} = \frac{1}{R} \sum_i r_i \quad (1)$$

$$I_{ms_color} = \frac{1}{R} \sum_i r_i^2 \quad (2)$$

$$I_{var_color} = \frac{1}{R} \sum_i (r_i > 0 ? 1 : 0) \quad (3)$$

ただし、視覚的効果の低い黒 (#000000) とグレー (#808080) の2色を区別するため、以下では、その2色を重要度算出に用いない場合の重要度を I'_{avg_color} , I'_{ms_color} , I'_{var_color} で表す。

3.2 アニメーションの使用回数を考慮した $I_{animation}$

ここでは、あるシーン c におけるアニメーションの使用回数を考慮して、 c に含むサブシーンの枚数を利用し、 $I_{animation}$ を以下のように求める。

$$I_{animation} = \log(N + 1) \quad (4)$$

ただし、 N はシーン c にあるサブシーンのサムネイルの枚数とする。

3.3 色分布とアニメーション使用回数との組み合わせ

スライド内の色分布情報をアニメーションの使用回数と組み合わせるために、色比率パラメータ γ を用いて、式 (5) を提案する。

$$I_{c+a} = I_{color}^\gamma * I_{animation}^{1-\gamma} \quad (5)$$

ただし、 I_{color} は 3.1 節で提案した色分布を用いた重要度のいずれかである。

4. 生成したダイジェスト

図6で本稿の提案手法を用いて生成したダイジェストを示している。今までの手法 (図5) に比べると、全体的にカラフルなイメージになり、単語の出現頻度をベースとした手法では取れなかった視覚的効果の高いシーンがダイジェストシーンとして選ばれているため、興味を引き出すダイジェストが得られたと言える。

5. 実験

本実験では、MPMeister で生成した34件の講演コンテンツ (表2) から、3種類のキーワード (「ストリーム」、「信頼性」、



図 5 テキストが中心となるダイジェスト



図 6 視覚的効果の高いシーンが中心となるダイジェスト

表 2 実験用データセット

講演コンテンツ	34 件
動画の長さ	7 分 3 秒 ~ 24 分 20 秒
シーンの枚数	12 ~ 49 枚
サブシーンの平均枚数	5.26 枚 / シーン

「スライド」) それぞれに関連する講演に対して、提案した手法を適用し、3 分間ダイジェストの自動生成を行なう。色分布に関しては、RGB 色空間の各成分を 2 等分し、合計 8 ビンの色ヒストグラムに解析をした結果を利用した。

5.1 評価方法

提案手法の有用性を調べるために、被験者の協力を得て、各講演コンテンツのシーンに点数を付け、それに基づき、正解セットを作成した。生成したダイジェストと正解セットとを比較することで、評価を行なった。

まず正解セットを作成するため、キーワードに関連する講演コンテンツに含まれるタイトルシーンを除いた C 枚のシーンに対して、興味を引く・印象に残るシーンであることなどを基準に、 $\lceil \log_2 C \rceil$ 段階で点数を与えることにした。

被験者らが付けた点数の和と、各講演コンテンツからダイジェストとして抽出されたシーンの枚数に基づき、評価に用いる正解セットを定める。表 3 にて、今回実験に用いられた正解セットの一例を示す。例では、被験者により、ある講演コンテンツがもつ 11 枚のシーンに対して、4 段階 (0 点を 6 枚、1 点を 3 枚、3 点を 1 枚、5 点を 1 枚) の点数が付けられる。評価対象となるダイジェストがその講演から 3 枚のシーンを抽出した時、被験者が付けた点数の和が上位 3 枚までのシーンを正解シーンとする。

また、ダイジェストを自動生成する際に、各シーンがダイジェスト内で表示される時間長の割り当てを行なうが、本実験では、

表 3 正解セットの一例 (ダイジェストシーンが 3 枚のとき)

シーン	被験者 A	被験者 B	被験者 C	被験者 D	sum	正解
1	0	0	0	0	0	
2	1	3	5	0	9	
3	0	0	1	1	2	
4	1	0	0	0	1	
5	3	5	1	3	12	
6	0	0	0	0	0	
7	1	0	1	5	7	
8	0	1	0	1	2	
9	0	1	0	1	2	
10	5	0	0	0	5	
11	0	1	3	0	4	

時間配分の結果の影響を考慮しない。作成した正解セットに基づき、以下の式 (6) で定義した *Accuracy* で評価を行なう。

$$Accuracy = \frac{|\text{正解セット} \cap \text{ダイジェスト}|}{|\text{ダイジェスト}|} \quad (6)$$

5.2 実験結果と考察

以下において、提案手法の有効性を調べるために、作成した正解セットを用いた評価実験の結果と考察を述べる。

5.2.1 実験 1

実験 1 では、3. 節で提案した色分布を考慮した I_{avg_color} 、 I_{ms_color} 、 I_{var_color} とアニメーションの使用回数を考慮した $I_{animation}$ を用いて、実験を行なった。色分布情報については、黒とグレーの 2 色を考慮するとき (I_{avg_color} 、 I_{ms_color} 、 I_{var_color}) と、黒とグレーを考慮しないとき (I'_{avg_color} 、 I'_{ms_color} 、 I'_{var_color}) の比較を行なった。最後に、テキストを中心とした重要度算出式とを比較した。

3 種類のキーワードを用いた実験の結果を図 7、図 8、図 9 で示すように、キーワード「ストリーム」を用いた実験では、全体的にキーワード「信頼性」に関する実験よりよい結果が得られている。また、キーワード「ストリーム」を用いた実験では、黒とグレーの 2 色を考慮した色の二乗平均が最も良い結果を示しているのに対して、キーワード「信頼性」と「スライド」を用いた実験ともに、色の種類数を考慮した抽出手法がよりよい有効性を示している。

その理由を調べると、キーワード「ストリーム」に関連する講演コンテンツでは、キーワード「信頼性」と「スライド」に関連する講演コンテンツよりも、図を多く使用していることがわかった。また、抽出結果のよい講演コンテンツでは、手法やシステム全体の説明などに図が多く用いられているが、抽出結果が望ましくない講演コンテンツでは、図を用いて説明する場面が少ない傾向が見られる。そして、アニメーション使用回数を考慮した抽出手法は色分布ほどの有効性を示さなかった一因として、本実験で使用した講演コンテンツの中に、アニメーションを用いて説明するシーンが少ないことが挙げられる。

スライド内の色分布を用いて、3 種類のキーワードに対して生成したダイジェストの *Accuracy* の平均値を表 4 にまとめた。実験結果からわかるよう、色の種類数の多いスライドに、利用

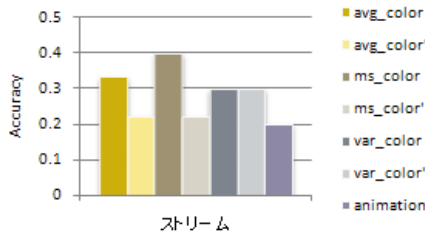


図 7 キーワード「ストリーム」に関するダイジェストの評価結果

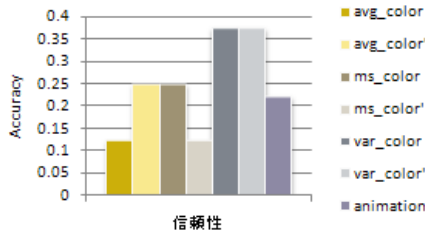


図 8 キーワード「信頼性」に関するダイジェストの評価結果

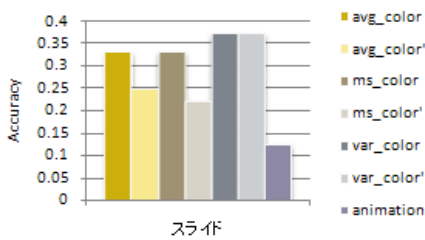


図 9 キーワード「スライド」に関するダイジェストの評価結果

表 4 生成したダイジェストの平均 Accuracy

	色平均	色の二乗平均	色の種類
黒・グレーを考慮する	0.264	0.328	0.350
黒・グレーを考慮しない	0.241	0.190	0.350

者が興味をもつ可能性が大きい。色の平均を用いてダイジェストシーンを抽出する場合と色の二乗平均を用いた場合ともに、黒とグレーの二色を考慮したほうがより有効である。また、色の種類を用いた重要度算出式で、結果に差が出なかった理由は、黒とグレーの二色がほとんどのシーンに含まれているためだと考えられる。

各手法の違いを調べると、色平均を用いた重要度では、色の種類に関係なく、より図またはテキストに埋まっているスライドが重要となる。それに対して、色の二乗平均を用いた手法では、異なる種類の色が程よく使用されたスライドのほうが高い重要度となる。そして、色種類を考慮した重要度では、色の付いているピクセルの数に関係なく、より種類多くの色を使っているシーンがダイジェストシーンに選ばれる。

実験 1 で平均 Accuracy の最も良かった I_{var_color} をこれまでダイジェスト生成に用いられたテキストをベースとした重要度と比較するために、3種類のキーワードに対する平均 Accuracy を算出した。図 10 にて、3種類のキーワードに対する I_{var_color} とテキストベースの重要度の平均 Accuracy を示す。グラフが

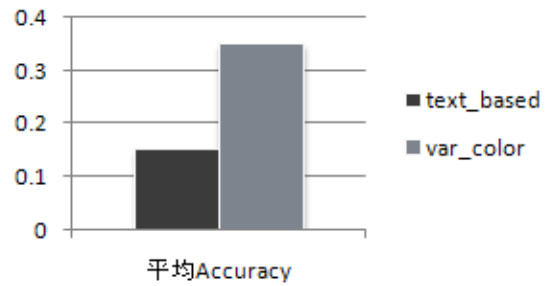


図 10 I_{text_based} と I_{var_color} を用いた時の実験結果

表 5 実験 2 の結果のまとめ

	色平均	色の二乗平均	色の種類
黒・グレーを考慮する	0.383	0.369	0.361
黒・グレーを考慮しない	0.398	0.319	0.361

らわかるように、色の種類数を考慮した重要度がテキストをベースとした重要度よりも、利用者の興味を引きだすダイジェストを生成する際、有効性を示している。

5.2.2 実験 2

実験 2 では、色分布情報とアニメーション使用回数を組み合わせた時に、利用すべき色分布情報の種類および最適なパラメータについて調べた。

実験方法として、 I_{c+a} に色分布情報を用いた 6 つの式をそれぞれ適用したとき、色比率パラメータ γ を 0.0 から 1.0 まで 0.1 刻みで変化させ、3種類のキーワードを使って生成したダイジェストの Accuracy の平均値で実験を行なった。 I_{c+a} の定義より、 $\gamma = 0$ のときに $I_{c+a} = I_{animation}$ となっており、 $\gamma = 1.0$ のとき、 $I_{c+a} = I_{color}$ となる。

I_{avg_color} と I'_{avg_color} の実験結果を図 11 で、 I_{ms_color} と I'_{ms_color} の実験結果を図 12 で、 I_{var_color} と I'_{var_color} の実験結果を図 13 で示す。

実験結果からわかるように、3手法について、色比率パラメータ $\gamma = 0.1 \sim 0.2$ までの区間で、アニメーションの使用回数と組み合わせることにより、Accuracy が大きく改善した。その後、色比率パラメータ γ が大きくなるにつれて、結果が下がる傾向があるため、 γ を小さい値に設定することが適切であるとされる。

表 5 にて、各式を用いた時の最大 Accuracy を示す。実験 1 における色分布情報のみ考慮したときの結果 (表 4) と、表 5 で示す実験 2 の結果とを比較すると、全ての項目において、結果の改善があったことから、色の分布情報とアニメーションを組み合わせる手法の有効性を示した。

今回の実験では、黒とグレーの 2 色を考慮しない色の二乗平均とアニメーションを組み合わせる重要度算出手法が $\gamma = 0.1$ のときに、Accuracy が最大となった。

6. 関連システム

ニュース番組の要約システム VideoQA [9] では、映像の特性を分析し、音声認識と文字認識によって得られたテキスト情報を

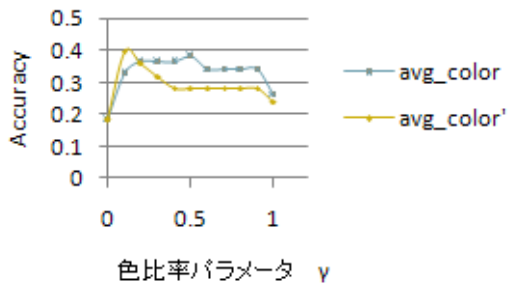


図 11 I_{avg_color} と I'_{avg_color} を用いた時の実験結果

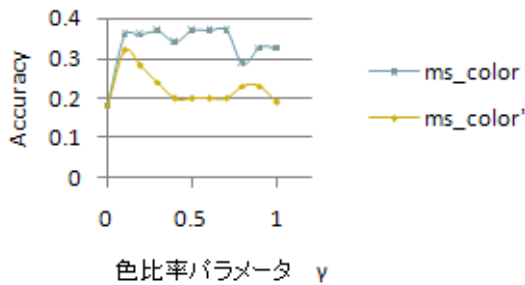


図 12 I_{ms_color} と I'_{ms_color} を用いた時の実験結果

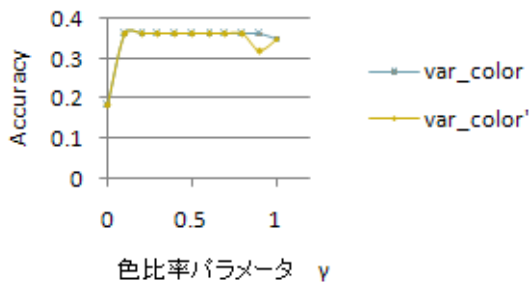


図 13 I_{var_color} と I'_{var_color} を用いた時の実験結果

を使って、さらにインターネット上にある関連新聞記事を探す。最後にそれらの情報を利用して、ユーザの質問に対し、ニュース動画アーカイブスの中から、回答となるニュースの要約映像を生成する。しかし、講義・講演動画では、クローズキャプションがなく、またその発話には高い自発性を持っているため、ニュース番組で用いられる手法はそのまま適用できない。

また、スポーツ動画に着目したシステムとして、Bezerra が提案したサッカーの動画を要約するシステム [10] があるが、講演コンテンツではスポーツ動画のように、映像が常に大きく変化したり、止まったりすることがないため、スポーツ動画に用いられる要約手法は講演コンテンツに適していないと考えられる。

そして、プレゼンテーションコンテンツを対象とした王らの研究 [11] では、シーン間の関係を考慮することで、単一シーンと関係指定によるシーンの検索や複数シーンの統合を実現したが、本研究ではシーン間の関係を利用していない。そのほかに、講演要約に着目した藤井らの研究 [12] では、講演に対応する

予稿論文の文書構造と表層情報を利用して、利用者が指定した論文の一部に対応する講演音声の要約を提供する手法を提案している。また、堀ら [13] では、講演音声の発話文から重要な単語を抽出し、それらを接合することで、講演に対する音声の自動要約を実現したが、いずれも講演内容の概要となるダイジェストで、本研究の目的と異なる。また、本稿でダイジェストを生成する際にあたって、講演に対応する論文や音声認識の情報を考慮していないが、今後の課題として考えていきたい。

7. まとめと今後の課題

本稿では、利用者の興味を引くようなダイジェストの自動生成を実現するため、スライド内の色分布を考慮した 3 つの I_{color} 、アニメーションの使用回数を考慮した $I_{animation}$ および両者を組み合わせた I_{c+a} といったシーン重要度指標を提案した。提案手法の有効性を確認するために、各講演コンテンツに対して、被験者によるシーンの点数評価に基づき、作成した正解セットを使用した。

評価実験において、これまでのテキストを中心とした手法よりも、提案手法のほうが視覚的効果の高いシーンの抽出に適していることがわかった。色分布情報を利用した 3 種類の式については、黒とグレーを重要度算出に利用したときと、利用しないときの比較を行なって、色の種類数を用いた算出式が最もよい結果となった。また、アニメーション使用回数を考慮した $I_{animation}$ を用いて、ダイジェストを自動生成した。さらに、色分布情報とアニメーション使用回数を組み合わせた時の利用すべき色分布情報と最適なパラメータについて調べ、有効性を確認した。

今後の課題として、まずキーワードの種類を増やして実験を行いたい。また、今まで提案した手法の統合や、音声情報の利用が挙げられる。

謝 辞

本研究を進めるにあたって、貴重な実験用データをご提供くださった筑波大学北川博之先生、森嶋厚行先生、並びに北川研究室の皆様へ深く感謝致します。

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 (#21013017) の助成により行われました。

文 献

- [1] 奈良先端科学技術大学院大学附属図書館. NAIST Digital Library. http://library.naist.jp/library/archive_top/index.html.
- [2] 日本データベース学会. DBSJ Archives. <http://www.dbsj.org/Japanese/Archives/archivesIndex.html>.
- [3] Haruo Yokota, Takashi Kobayashi, Taichi Muraki, and Satoshi Naoi. UPRISE: Unified Presentation Slide Retrieval

- by Impression Search Engine. *IEICE Trans Inf Syst (Inst Electron Inf Commun Eng)*, Vol. E87-D, No. 2, pp. 397–406, 2004.
- [4] 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. 第 19 回電気通信学会データ工学ワークショップ (DEWS2008) 論文集, pp. E4–1, 2008.
 - [5] 呉怡, 渡辺陽介, 横田治夫. 講演動画アーカイブスからキーワードに合致する講演ダイジェストの自動作成. DEIM フォーラム 2009 論文集, pp. E9–2, 2009.
 - [6] 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, 3 2005.
 - [7] Gerald Salton, editor. *Automatic text processing*. Addison-Wesley Longman Publishing Co., Inc., 1988.
 - [8] Ricoh Japan. MPMeister II. <http://www.ricoh.co.jp/mpmeister/>.
 - [9] Hui Yang, Lakha Chaisorn, Yunlong Zhao, Shi-Yong Neo, and Tat-Seng Chua. VideoQA: question answering on news video. In *the 11th ACM international conference on Multimedia*, pp. 632–641. ACM, 2003.
 - [10] F.N. Bezerra and E.Lima. Low cost soccer video summaries based on visual rhythm. In *the 8th ACM international workshop on Multimedia information retrieval*, pp. 71–78. ACM, 2006.
 - [11] 王元元, 北山大輔, 角谷和俊. 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. DEIM フォーラム 2009 論文集, pp. E9–4, 2009.
 - [12] 藤井敦, 伊藤克亘, 秋葉友良, 石川徹也. 音声言語データの構造化に基づく講演発表の自動要約. 話し言葉の科学と工学ワークショップ講演予稿集, pp. 173–177, 2001.
 - [13] 堀智織, 古井貞照. 講演音声の自動要約の試み. 話し言葉の科学と工学ワークショップ講演予稿集, pp. 165–171, 2001.