

# ニュース記事を利用したサッカー要約映像の生成

吹野直紀<sup>†</sup> 馬強<sup>†</sup> 角谷和俊<sup>†</sup> 田中克己<sup>†</sup>

<sup>†</sup> 京都大学大学院情報学研究科 〒 606-8501 京都府京都市左京区吉田本町

E-mail: †{fukino,qiang}@dl.kuis.kyoto-u.ac.jp, ††{sumiya,tanaka}@i.kyoto-u.ac.jp

あらまし 本稿では、要約映像を生成する一手法を提案する。特に、サッカー映像を対象とする。その際、ビデオ映像・音声とメタデータのみから要約を作るのではなく、人間が作成した高度な要約であり、Web等から入手が容易なニュース記事を利用する。まず、実況コメントは実況と解説が混在しているので、これらを分離し解説を実況に関連付ける。さらに実況からイベント区間を推定しニュース記事中のキーワードを使って映像中のイベント集合に問い合わせを行う。返された複数のクリップをキーワードの種類を考慮した類似度やイベント間の時間距離等を使ってランク付けし、ニュース記事の各部分に適切なビデオクリップを判断する。ニュース記事と映像の対応から判断された記事構造も利用して、最終的な要約映像を提示する。

キーワード ビデオシーン検索, マルチメディアデータベース

## Generating Football Video Summary Using News Article

Naoki FUKINO<sup>†</sup>, Qiang MA<sup>†</sup>, Kazutoshi SUMIYA<sup>†</sup>, and Katsumi TANAKA<sup>†</sup>

<sup>†</sup> Graduate School of Informatics, Kyoto University Yoshidahonmati, Sakyou-ku, Kyoto, 606-8501  
Japan

E-mail: †{fukino,qiang}@dl.kuis.kyoto-u.ac.jp, ††{sumiya,tanaka}@i.kyoto-u.ac.jp

**Abstract** In this paper, we propose a method that summarize video. Especially, we target at football video. In this process, we don't generate summary by video picture, video audio and metadata, but we use news article which is advanced summary human generated and is easily got from website and so on. This process begin by separate play-by-play comments from announcer's comments, and associate explaining comments with play-by-play comments. Next, it estimate event section from play-by-play comments and retrieve these events using keywords in news articles. Next, it rate many video clip using keyword's similarity and time interval between event and event and so on, and decide video clip which is appropriate to each parts of news articles. Final summarized video is presented using article structure which is decided from correspondence between each parts of news article and video clips.

**Key words** Video Scene Retrieval, Multimedia Database

### 1. はじめに

蓄積型TVやストリーミング技術の普及に伴って、我々は大量のランダムアクセス可能なビデオ映像を視聴できるようになる。しかし、ビデオ映像の量が増えても我々がビデオ映像を視聴できる時間はそう変わらない。そのため、長時間のビデオを要約して視聴したいという要求がますます高まると考えられる。特にスポーツ映像では、映像の長さが比較的長く、その中で重要なシーンは限られているので、短時間のダイジェストで見たいという要求はより高まる。

映像を要約するための技術として、カーネギーメロン大学 Informedia プロジェクトの VideoSkimming 技術 [2] やリコー

のダイジェスト生成技術 [3] がある。それらの技術は、映像から抽出されたか、もしくは人手でつけられたメタデータを用いている。しかし、映像とそのメタデータのみから重要シーンを取り出す事はできても、それを繋ぎ合わせたものが要約映像として見やすいかという点、人間が作成したものには遠く及ばないのが現状である。一方で、ドラマ映像・音声・シナリオ文書をDPマッチングを用いて対応付けることで映像を理解するアプローチがある [5]

本研究ではビデオ映像、特にサッカー映像を対象とし、その映像について記述しているニュース記事を映像中の実況コメントを利用して対応付けることで記事に沿った要約映像を得る手法を提案する。ニュース記事は、人間が作成した高度な要約で

あり、Web から入手しやすいという特徴がある。また、同じ映像に対しても複数のニュース記事があるので、複数の視点から要約映像を視聴する事ができる。

ビデオ映像の音声は、音声認識によってある程度得る事が可能であるが、認識ミスも多い。雑音が多いスポーツ映像では特にミスは多くなる。初めの段階で、対象映像をしばらく認識すべき単語をしばらくしておく事も必要になる。また、音声認識と入手を併用する事で生放送に字幕データを付ける試みも行われている[1]。このような字幕データは通常数秒に1回送信されてくる。その数秒のうち、実際にアナウンサーが発言したおおよそのタイミングを推測する事も必要である。

このような要約における課題として、

- ニュース記事の構成が単なる時間軸に沿った要約ではないという事
  - 起こったイベントの説明・まとめたテキスト・解説的なテキストがニュース記事にも、ビデオの音声データにも混在している事
- 等がある。ニュース記事に関しては、その構造を判断しなければならない。また、ビデオの音声データである実況コメントは、目の前で起こっている事を断片的に発言した部分と以前起こったイベントについて説明している部分に分離しなければならない。その上で、説明的なコメントを断片的な実況コメントに関連付ける必要がある。

## 2. 基本的事項

### 2.1 グルー演算

区間同士の演算には、intersection 演算、union 演算等がある。intersection は2つの区間の共通部分を返す。図1では区間Aと区間Bのintersectionとして区間R1を返す事が出来る。union は2つの区間を繋げた区間を返す。図1では区間Aと区間Bのunionとして区間R2を返す事が出来る。しかし、union では2つの区間の間にノイズがあった場合その2区間を繋げる事はしない。図1では、区間R3を返す事が出来ない。ビデオにはノイズが入る事はよくあるので、ノイズがあっても2つの区間を繋げて返す glue 演算が提案されている[4]。区間Aと区間Bの glue として区間R3を返す事が出来る。区間Aの始めの時間を  $A_{start}$ 、区間Aの終わりの時間を  $A_{end}$  とすると、区間Aと区間Bの glue である区間R3は  $R3_{start} = \min\{A_{start}, B_{start}\}$ 、 $R3_{end} = \max\{A_{end}, B_{end}\}$  となる。以下では、intersection 演算を  $\cap$  で表し、glue 演算を  $\bowtie$  で表す。

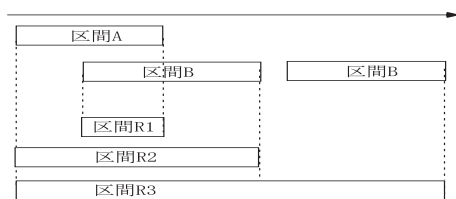


図1 intersection, union, glue 演算

グルー演算は区間と区間に対する演算だが、これを区間集合

と区間集合に拡張したものがペアワイズグルー演算である。2つの区間集合間の任意の区間同士を glue 演算して作った区間の集合を返すものである。図2は、キーワードAに対する区間集合A1, A2, キーワードBに対する区間集合B1, B2, キーワードCに対する区間集合C1に対して、ペアワイズグルー演算  $A \bowtie B \bowtie C$  を行った結果を表している。

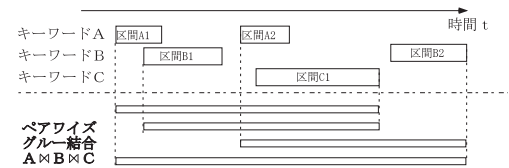


図2 ペアワイズグルー演算

さらに、パワーセットグルー演算  $\bowtie^*$  がある。これは、区間集合A, Bそれぞれから1つ以上の要素を取り出し、それらにグルー演算を行って作り出した区間の集合である。形式的には、次のように定義される。

$$X \bowtie^* Y = \{i | X' \subseteq X, Y' \subseteq Y, X' \neq \emptyset, Y' \neq \emptyset, i = \bowtie(X' \cup Y')\}$$

$$\text{ただし } \bowtie(i_1, \dots, i_n) = i_1 \bowtie \dots \bowtie i_n$$

グルー演算の結果に対するフィルタの例としては、区間の最大長で制限するフィルタや区間中のノイズの最大長を制限するフィルタがある。これらのフィルタはグルー演算の演算途中に挿入する事ができ、演算処理の効率化ができる可能性がある。

## 3. 基本概念

本手法では、サッカー映像の要約の過程でその映像について記述しているニュース記事を利用する。ニュース記事はWeb上から入手する事が可能である。

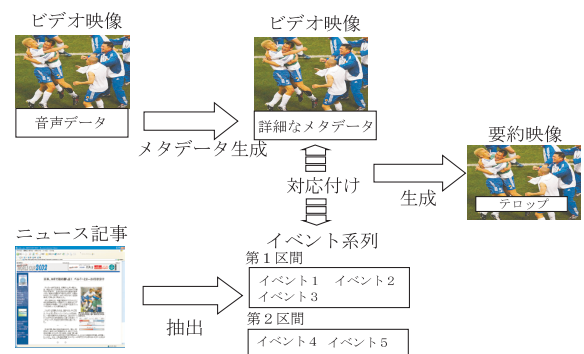


図3 本手法の概要

サッカー映像中に何が映っているかを判断するには、サッカー映像中のアナウンサーの音声を利用する。しかし、アナウンサーは見た事をリアルタイムに話さなければならないため、目の前で起こっている事、特に見れば分かる事の多くを省略して話す。そこで、アナウンサーの音声をもとに、より詳細な映像のメタデータを自動生成する。次に、要約映像のもとになる記事の中の文章を小さな区間に分ける。1つの区間は、例えば記事中の1文となる。本稿では、これを1文と想定する。そして、記事中のそれぞれの文の中にあるイベントを表すキーワー

ドと生成した映像のメタデータをマッチングする事で、記事の各文に対応する映像クリップを取り出す。取り出した映像クリップは記事にそって繋げて再生し、それを要約映像とする。対応する映像クリップを判断する際と要約映像として映像クリップをつなげる際に記事の段落構造を利用する。

#### 4. 実況コメントからのビデオメタデータ生成

実況コメントは、アナウンサーにとって時間が限られているために起こっている事を全て発言している訳ではない。特に、言わなくても視聴者に伝わる部分に関しては基本的に省略される。これら、省略された情報およびリアルタイムに話す必要のため構造化されていない情報を、可能な限り補完する必要がある。また、実況ではイベントの発生を確認し発言するが、その発言をした時刻は分かってもイベントがどの時刻からどの時刻まで継続していたかは分からない。イベントの継続時間を推測する必要もある。補完・推測を行い、ビデオ映像のメタデータを生成する。

##### 4.1 実況と説明の分離

本研究では映像の内容を知るために実況のデータを利用するが、サッカー映像から得る事ができる音・実況には以下のような種類がある。

(1) 流れている情報と同時に反映されるもの

例：笛の音・歓声・アナウンサーのテンション

(2) 映像が流れた直後に反映されるもの

例：試合を見ながらアナウンサーが断片的に発している言葉

(3) 映像が流れたしばらく後に反映されるもの

例：解説者による解説

(1) は重要なイベントの発生と共に起こる現象であり、イベント発生の正確な時刻を知るのに役立つ。(2) も同時刻映像にはどんなシーンが映っているかを推測するのに役立つが、アナウンサーはリアルタイムにそれほど多くの事を発言する事はできないので、情報はそれほど多くない。これを補うのが(3)である。一連のイベントが終わってしばらく経った後、それについての説明を行う場合がある。これは情報が豊富だが、いつのイベントについての話なのか判断できなければ使う事はできない。

(2) を実況、(3) を説明と名づけると、アナウンサーが話している事が実況なのか説明なのか。また、説明だとするとどのシーンに関する話なのか、を判断する事が大切になる。そのためには、重要なイベントの後はそのイベントについて少し話す事が多いという傾向、実況と解説が交互に発言している時は目の前で起こっている事については話していない場合が多いという傾向、リプレイという単語が出てきてからしばらくはリプレイを見ながら話しているので過去の大きいイベントについて話している事が多いという傾向等、様々な傾向から総合的に判断する必要がある。イベントを表す単語及び人名が高い密度で出てきている単語系列・文章は、高確率で実況と判断できる。

##### 4.2 イベントの発生時刻・継続時間の推測

実況の話し方にはイベントの発生時刻との関係で大きく3つに分けることができる

- 宣言 さあ～です、まもなく～です

- 現況 ～します、～です

- 結果 ～しました、～でした

実況のコメントからこれらの語尾等を検出することは、イベントの発生時刻を推測する上で非常に重要である。

また、イベント名も、それがアナウンサーによって発声される時刻とイベントが発生する時刻の関係によって同じように3つに分けることができる。

- 宣言 コーナーキック、フリーキック

- 現況 ドリブル、センタリング

- 結果 ゴール、ファール

コーナーキックという言葉は、多くの場合今からコーナーキックを蹴るという時に発言される。ドリブルは、今まさにドリブルしているという時に発言される事が多い。また、ゴールはその瞬間を確認した後でアナウンサーは発言する。

また、イベントの発生時刻の他にイベントの継続時間もイベントの種類により異なる。例えば、「パス」といったイベントの継続時間は非常に短い、「波状攻撃」といったイベントの継続時間はパス等に比べ非常に長い。そういったイベント毎の特徴を把握するため、図4のような統計をとる。図の横軸はイベント名をアナウンサーが発言してから時間であり、いくつかのケースについてアナウンサーが発言したタイミングの前後どれくらいにイベントが継続しているかを表している。

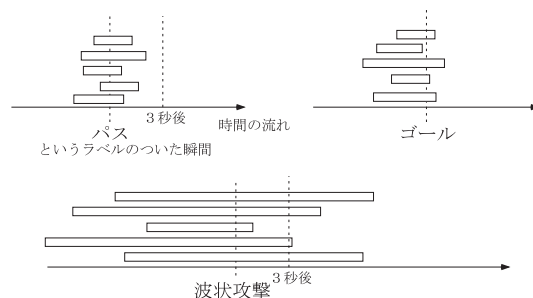


図4 アノテーション時刻を中心としたショットの分布

これらの統計を多くの場合についてとると、図5のようなグラフを作成する事ができる。横軸はラベルの付いた時刻を原点0とした時間軸であり、縦軸はそれぞれの時刻tがラベルに相当するショットに含まれていた確率である。このような評価関数を、本研究ではショット区間評価関数と呼ぶ[6]。

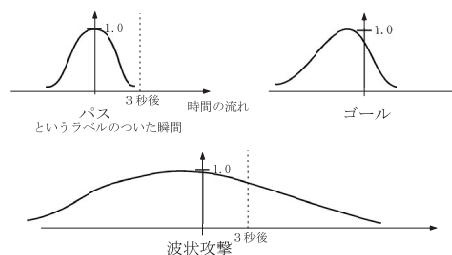


図5 ショット区間の分布を表現する評価関数

このグラフを用いて、映像中ではどこからどこまでがアナウンサーが発言したイベントの該当区間なのかを推測する。実際のイベント区間は、この評価関数に閾値を儲け、確率がある閾

言い回し	ずれ
$x$ 分	80sec, -68sec, 22sec, 40sec
まもなく $x$ 分	-70sec, -118sec, -30sec, -45sec
ちょうど $x$ 分	-13sec, 10sec, 2sec, 28sec

値以上になっている部分として判断する．閾値は，必要に応じて動的に変化する．

### 4.3 時刻の推定

記事上では「23分には がゴールを決め」といったように時刻付きで説明される場合が多い．特に「11分のファール」等，ビデオ区間中何度も発生するイベントを時間のみが修飾しているケースがあり，それらのビデオ区間をビデオ映像全体の中から特定するために，ビデオ中での時刻を推定する必要がある．

例えば，前半の間に時間に関して次のような実況の情報があったとする．

- 「まもなく15分になるところです」
- 「手元の時計でちょうど20分です」
- 「25分の彼のシュートについてなんですが…」
- 「前半の30分で3回給水をしていますね」
- 「残り5分となった前半です」

3つ目，4つ目は現在の時刻を言っているのでは無い．その点を判断して正確な時間を知りたい．

まず，“まもなく”や“ちょうど”等，言い回しによって実際のビデオ中の時刻とのずれの大きさが異なる．

この図の分布をグラフにすると図6のようになる．「まもなく  $x$  秒」などのコメントが発生した時刻を原点にとった時刻  $t$ (秒)である．縦軸はおのこの時間が実際のビデオ内時刻  $x$  分  $\pm 1$  分に含まれている確率密度であり，以下確信度と呼ぶ．



図 6 言い回しごとの確信度

図7で上部にある時間の軸を左右にずらして確信度の総和が最も大きくなったところを正しいビデオ内時間と見なす

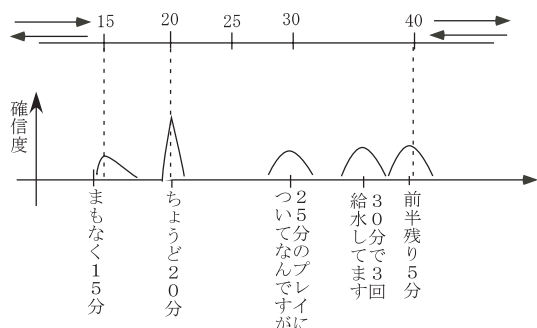


図 7 時間の推定

### 4.4 シーンテンプレート

実況のコメントは断片的である．画面に映る映像を詳細に実況する時間的余裕はない．そのため，それを考慮した質問で問い合わせを行うべきである．

例えば，実況では“稲本のファール”と言って，その後しばらくたってから“イエローカード，26分です”と言ったとする．これを記事中の“稲本がイエローカードを受けた”という文からそのシーンを取り出そうとした場合，“稲本のイエローカード”は実況中には存在しないという不都合が起こる．

そこで，イエローカードが起こるシーンのパターン（ひとまとまりのイベント）を記述したシーンテンプレート（図8）を用いる．このテンプレートが表現しているのは，“ファールの

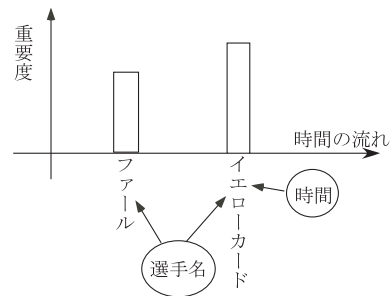


図 8 イエローカードシーンのテンプレート

後にイエローカードが出される．周辺で人名が読み上げられた場合，その人はイエローカードを受けた人であり，同時に直前にファールを犯した人である．周辺で時間が読み上げられた場合，それはイエローカードを出された時間である．”という情報である．この情報を用いて，記事中の“稲本のイエローカード”という条件だけでなく，“イエローカードの直前にある稲本のファール”という条件でも検索を行う．

シーンテンプレートの適用手順は次のようになる．par「ファールがありました 後方からつきます これはイエローカードです ベルヘイエンです 手元の時計で後半16分のイエローカードです」

この場合，始めのファール，イエローカードは実況にあたり，その後の発言は説明にあたる．アナウンサーがイエローカードと言った瞬間このテンプレートが適用され，まずは前方のファールをひとまとまりのシーンとして取り込んだ後，その後のアナウンサーの発言により選手名・時間を埋めていく．これらの情報は全て埋まる必要は無い．

## 5. 記事からのイベント系列取り出し

記事は，タイトルや写真と記事本文に分けられる．記事本文は，1つあるいは複数の段落に分かれており，それぞれの段落はいくつかの文によって構成される．

本手法では，まず記事を1文ごとに区切り，それぞれの文からイベントを表すキーワードを抽出する．例えば「得点した」という言葉からイベント「ゴール」を取り出す．このために，ゴールやパス等サッカーで起こるイベントの様々な表現を記録した辞書を用意しておく．次に各文から複数のイベントが取り



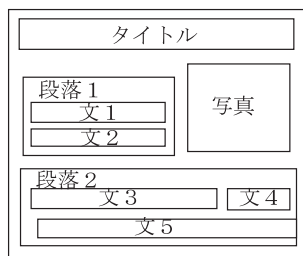


図 9 記事段落の例

出されるので、そのイベントの集合をもとに、その文に対応する映像クリップを見つける。ただし、文に含まれるイベントの情報のみでは、該当する映像クリップが複数見つかったり逆に1つも見つからなかったりする。その際に用いるのが段落構造である。段落構造をもとに、文の前後関係と各文に該当するシーンの包含関係を把握する。

### 5.1 文の前後関係

一般的に同一段落内では、各文の前後関係が映像中のそれと一致している場合が多い。この傾向を用いて、イベント情報の少ない文も前後関係から絞込む条件を増やし、対応する映像クリップを特定する。また、同段落中の文同土程ではないが別の段落同士でも前後関係が映像中の前後関係と一致している事がある。この点も利用して映像クリップを特定する。

### 5.2 段落の包含関係

ニュース記事は、ある段落で概要を述べた後、続く段落でそのシーンの詳細を述べるといった構造をとる。

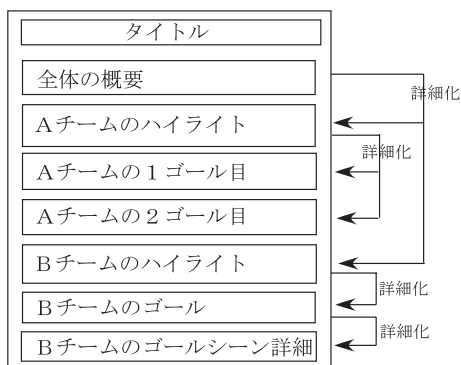


図 10 段落の詳細化関係

この詳細化関係の推定には、各段落中の文から抽出したイベントキーワードの重要度を用いる。重要度は、

$$imp(\text{イベント } e) = \frac{e \text{ の記事中の出現頻度}}{e \text{ の実況中の出現頻度}} \quad (1)$$

で計算される。例えば、重要度の高いイベントが多くある段落に重要度が低いイベントが混ざった段落が続いている場合、続いている段落は重要度の高い段落の詳細化であると考えられる。もとの段落を親段落、詳細化された段落を子段落と呼ぶと、ある段落がある段落の子段落であることが分かった場合、その子段落に相当する映像区間は親段落全体に相当する映像区間の部分区間であると考えられる事ができる。この段落同士の包含関係を利用する事で、段落中の各文に対応するビデオ映像クリップが満たすべき条件を追加して絞り込む事ができる。

## 6. 記事と映像クリップの対応付け

本手法による要約映像は、記事の各部に相当する映像クリップをつなげて作られる。その過程では、記事の各部に相当する映像クリップを定める必要がある。対応する映像クリップは、4章の方法で映像から生成したメタデータ中のイベント系列と5章の方法で記事から抽出したイベント系列を比較する事で推定する。記事の一部に対応する映像クリップを決めるための操作は大きく3段階に分ける事ができる。第1段階で、記事中のある文に対応する映像クリップの候補を列挙する。第2段階では、列挙された候補をそれぞれランク付けする。最後の第3段階で、記事の段落構造を考慮し最終的に対応する映像クリップを決定する。

### 6.1 候補クリップの取り出し

ニュース記事の各文に対応する映像クリップを取り出す際、基準となるのは候補クリップのトータル長さや候補区間中のイベント間の時間距離である。まず、文に含まれるイベントの種類をもとに、その文に相当するクリップの長さがいくらか以下であればよいかを決める。例えば、文に「パス」と「パスカット」というイベントが記述されていたとすると、その文に対応する映像クリップはそれほど長くなるとは考えにくい。一方、文に含まれるイベントが「カウンター」と「ゴール」であった場合、それぞれのイベントがある程度長い時間に渡るイベントなので、その文に相当するクリップはある程度長い可能性がある。このように単語の種類を考慮して最大のクリップの長さを決め、その長さ以下の時間に集まっているビデオ中のイベントの組み合わせを選ぶ。

また、イベント同士で最大どの程度時間的に離れている可能性があるかという事もイベントの種類から判断する。最大のトータル長さや最大のイベント間距離を両方超えていないイベントの組み合わせを全て取り出す。

文に対応する候補クリップは、取り出された組み合わせ中のそれぞれのイベントに対応する映像区間(ショット区間)をグループ結合 [4] する事で取り出す事ができる。

### 6.2 候補クリップのランク付け

候補クリップのランク付けの際に基準となるのは、記事文中のイベントと候補クリップ中のイベントがいかになら一致しているかである。その計算には、イベント毎の重要度とイベント毎の一致確率を用いる。

イベント毎の一致確率とは、例えば記事中に「スルーパス」と書いてあり、映像中に「パス」というイベントが見つかった場合、映像中の「パス」が「スルーパス」である確率である。この確率は、実際に多くのパスシーンのうち何割がスルーパスであったかを統計をとり求める。それを様々なイベントについて求め、図 11 のような辞書を作る事ができる。図中で矢印になっている所は is-a 関係になっているが、そこに記述されている小数が表しているのが「親ノードのイベントが映像中に記述されていた場合そのイベントが子ノードのイベントである」という確率である。辞書中で、同義語は同一ノード内に記述され、一致確率は 1 になる。

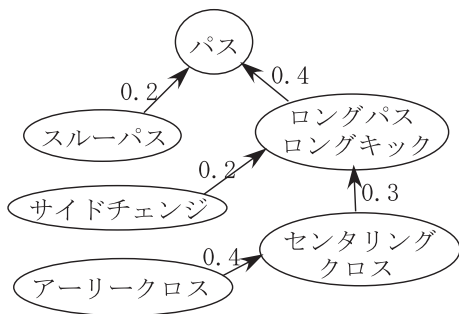


図 11 一致確率付き辞書

記事中のイベント系列を  $(e_1, e_2, e_3, \dots)$  とする。また、候補クリップ中のイベント系列を  $(e'_1, e'_2, e'_3, \dots)$  とする。候補クリップ中のイベント  $e'_1$  がイベント  $e_1$  である確率を  $d(e'_1, e_1)$  とし、イベント  $e_1$  の重要度を  $imp(e_1)$  とすると、イベント系列の一致度は

$$d(e'_1, e_1) \times imp(e_1) + d(e'_2, e_2) \times imp(e_2) + \dots \quad (2)$$

で求める。この式は、重要度の高いイベントが一致している事をより重視する事を表している。

### 6.3 段落構造を利用した対応クリップの決定

候補クリップのうちランク値が最も高いものであっても、各文に対応するクリップ決定し繋げた結果それが記事の文脈にそっていないければ、その候補クリップが間違っているということになる。記事に文脈として判断できるものは、5章で述べた同一段落中の文の前後関係と、段落同士の親子関係による段落の包含関係である。候補クリップのうち最も高いランク値に重み  $w(0 < w < 1)$  を掛けた値以上のランク値を持つ候補クリップを、各文に対して組み合わせる事で記事の文脈を満たした対応クリップの組み合わせを見つけ、それらを最終的な各文に対応する映像クリップとする。

## 7. アウトプットの生成

記事中の各文に対応する映像クリップを特定した後ならば、より詳しく記事の構造を把握する事ができる。単語の重要度からでは段落同士の詳細化関係等しか分からなかったが、各文が記述している映像中の時刻が分かれば図 12 のように文レベルで詳細化等の関係が分かる。

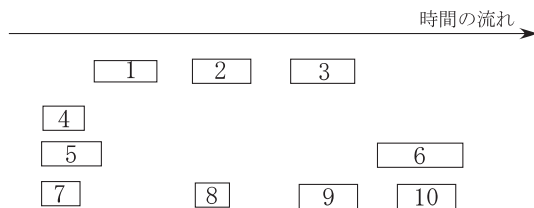


図 12 記事中の文章とビデオ内容の対応

図 12 は、ある記事中の 10 個の文がビデオ内の時間軸でどこに位置していたかを示している。

文 1 ~ 10 に関係を定義する事ができる。まず、文 4 と文 5 はビデオ中の同じ部分についての記述であり、言い換えである

と考えられる。特に、言い換えの中でも文 5 は文 4 よりもイベントが豊富に含まれており、かつ重要度の低いイベントも含まれる。このような場合、文 5 は文 4 の詳細化であると考えられる。

また、各文をさらにグループ化する事ができる。文 1 ~ 3 が共通要素「チーム A」、文 4 ~ 6 が共通要素「チーム B」、文 7 ~ 10 が共通要素「ゴール」を持っているとする。すると、文 1 ~ 3 を「チーム A のハイライト」、文 4 ~ 6 を「チーム B のハイライト」、文 7 ~ 10 を「全体のゴールハイライト」と考えることができる。これらの情報から判断された記事構造を考慮して、要約映像を生成する。具体的に述べると、言い換えている所は同じシーンを 2 度ユーザに見せる事は望ましくないので映像の見せ方を変える。後に出てきた方はリプレイ映像の方を見せたり、スローモーションで見せたりする方法が考えられる。また、後で詳細化して述べられる事が分かっている文は、通常の方法で映像クリップを流すのではなく重要シーンだけを断片的に繋いだ映像を流すだけの方がよいと考えられる。

以上の作業により作成された要約映像をユーザに提示する。提示の方法としては様々なものが考えられるが、一例を以下に示す。



図 13 表示される画面

映像にテロップをオーバーラップさせる。テロップは、ニュース記事から 1 文毎に表示する。背景に流れているのは、テロップに対応する映像クリップである。この場合、対応する映像クリップが非常に短い場合でも、最低ユーザがテロップを最後まで読めるように映像クリップの長さを調節する必要がある。そのほかにも、映像に同期してニュース記事をスクロールさせていく方法等が考えられる。

## 8. おわりに

ニュース記事とサッカー映像の中身を実況コメントをたよりに対応付けることで、ニュース記事に沿った要約映像を生成する手法について提案した。今後の課題として、各段階のアルゴリズムについて検討し、プロトタイプを実装した上でその有効性を検証する。また、本稿では簡単のため記事を 1 文毎に区切ったが、1 シーンに相当する区切りの単位として 1 文が適切とは限らない。1 文では不適切だと判断された場合、その 1 文を分割したり周囲の文と連結してそれぞれを 1 シーンとする手法についても考えなければならない。その他、スポーツのニュース記事にはイベント情報をほとんど含んでいない段落が存在する。場所や結果など、事実を記事の冒頭で 5W1H 的に述べている部分や、記者の感想を述べている段落などである。これらの文

章にどのような映像を割り当てるかを考えなければならない。  
サッカー映像以外のスポーツ映像への拡張についても考える。

## 謝 辞

本研究の一部は、平成 14 年度科研費特定領域研究 (2) 「Web の意味構造に基づく新しい Web 検索 サービス 方式に関する研究」( 課題番号: 14019048, 代表: 田中克己), 平成 14 年度基盤技術研究促進事業 ( 民間基盤技術研究支援制度) 「クロスメディア コンテンツ基盤技術の研究開発」, 及び 21 世紀 COE プログラム 「知識社会基盤構築のための情報学拠点形成」による。ここに記して謝意を表します。

## 文 献

- [1] NHK 生放送字幕のための音声認識システム  
<http://www.nhk.or.jp/str1/open2002/tenji/id03/03.html>
- [2] H.Wactlar: Informedia - Search and Summarization in the Video Medium, Proceedings of Imagina 2000 Conference, January, 2000.
- [3] Takako Hashimoto, Yukari Shirota, Atsushi Iizawa, Hiroyuki Kitagawa: Digest Making Method Based on Turning Point Analysis. WISE (1) 2001: 83-91
- [4] ブラダン スジツ, 田島 敬史, 田中 克己: "ビデオデータ検索のための区間グルー操作と解のフィルタリング", 情報処理学会論文誌・データベース (1999.1)
- [5] 柳沼良知, 坂内正夫: "DP マッチングを用いたドラマ映像・音声・シナリオ文書の対応付け手法の一提案", 電子情報通信学会論文誌, Vol.J79-D-II, No.5, pp.747-755 ( 1996.5)
- [6] 吹野 直紀, 角谷 和俊, 田中 克己: "キーワード毎のショット長分布を用いたビデオ映像シーン検索", 情報処理学会研究報告, Vol.127, No.5(2002.5)