

## リアルタイムオーサリングのための音声データによる構造化内容記述

4U-9

飯嶋 亨<sup>†</sup> 赤迫 貴行<sup>‡</sup> 角谷 和俊<sup>‡</sup> 田中 克己<sup>‡</sup><sup>†</sup>通信・放送機構 神戸リサーチセンター<sup>‡</sup>神戸大学大学院自然科学研究科

## 1 まえがき

本稿では、ストーリーがなく、かつリアルタイムに状況が変化する野球などのスポーツ実況中継を対象として、リアルタイムに試合を内容記述、索引付け、編集するリアルタイムオーサリングシステムについて述べる。

## 2 本研究の動機

CSデジタル放送の多チャンネル化や将来のBSデジタル化、地上波デジタル化により映像と共に付加情報を送信することが可能となってきている。しかし、リアルタイムに付加情報を自動作成し送信する技術は充分とは言えない。そのため、これらの作業を支援し作業量を軽減することが求められている。

関連研究として以下のものがあげられる。ライブ映像ストリームに対するリアルタイムな内容記述ツールに関する研究 [1]、ライブ映像ストリームに対するリアルタイム編集に関する研究 [2] が行われている。また、音声認識を用いた研究として、大相撲番組での決まり手を音声認識を用いて字幕スーパーを行う研究 [3] がある。さらに、ニュース音声を対象として、キーワードスポッティング機能により音声認識をし、その結果得られた単語から話題と同定する研究 [4]、さらに、スポーツの試合に画像処理を用いた研究として、野球中継を放送している画面上のテロップを画像処理してイベントの候補を推定し提示する研究 [5] がある。

しかし、これらの研究は、内容記述に有効時間が含まれないために映像検索に利用できない場合やリアルタイムに内容記述を行うことができない場合がある。本研究では、リアルタイム内容記述によるビデオインデックス作成に必要な要素として音声認識、インデックス表現、区間推定について検討を行った。

筆者らは、これまでにリアルタイムオーサリングを支援するデータベースに関する研究 [6]、映像に対するリアルタイム内容記述方式と実装の検討に関する研究 [7] を行っている。

本稿では、インデックス表現と区間推定及びプロトタイプシステムの実装について述べる。

## 3 インデックス表現と区間推定

1つの映像データに対する内容記述の集合をインデックスと呼ぶ。

動画像  $V$  に対するインデックス  $I_V$  は、以下のように表される。

$$I_V = \{r | r = (d, dt, tt, v), v \subseteq V\}$$

$r$  はインデックスを構成する要素で、インデックスレコードと呼ぶ。インデックスレコードは映像データの部分的な表現を担う。 $r$  は  $d$  (記述を表し、原子イベント型、状況型、複合イベント型の3種類から成る)、 $dt$  (記述型)、 $tt$  (トランザクション時間)、 $v$  (有効区間) の4項から成っている。

ここでは、内容記述をインデックスとして表す。また、各インデックスの有効範囲を求めるために区間推定を用いる。

## 3.1 インデックス表現

インデックス表現を構成する記述内容とイベント内容により、次のように状況型、原子イベント型、複合イベント型の3種類に分類する。

## 1. 状況型

状況 (situation) の値 (value) を記述する。

$$d_s = (\text{situation}, \text{value})$$

## 2. 原子イベント型

オブジェクト (object) がどういう役割 (role) でどういうイベント (瞬間的な動作を表すイベント) を行ったかを記述する。

$$d_e = (\text{object}, \text{role}, \text{event})$$

## 3. 複合イベント型

オブジェクト (object) がどういう役割 (role) でどういうイベント (一連の動作結果を表すイベント) を行ったかを記述する。

$$d_e = (\text{object}, \text{role}, \text{event})$$

## 3.2 区間推定

3種類の型毎に、記述内容の有効範囲を区間推定する。

## 1. 原子イベント型

インデックスレコード系列において、直前の、原子イベント型の記述を持つインデックスレコードのトランザクション時間を参照する。 $l$  は考慮すべき時間的なずれ。

$$r_i.v = [r_{i-k}.tt + l, r_i.tt + l]$$

## 2. 状況型

インデックスレコード系列において、直後の、situationの値が同じで、valueの値が異なる状況型の記述を持つインデックスレコードにおけるトランザクション時間を参照する。

$$r_i.v = [r_i.tt, r_{i+k}.tt]$$

## 3. 複合イベント型

複合イベント型の記述を持つインデックスレコードの有効区間を決定するために、決定性有限オートマトンの利用を考える。

本研究では、有効区間導出オートマトン  $M$  を以下のように定義した。 $M$  はミーリー型の決定性有限オートマトンであり、インデックスレコード  $r$  において、記

“Structured Description of Content by Audio-Data for Real-Time Authoring”

Toru Iijima<sup>†</sup>, Takayuki Akasako<sup>†</sup>, Kazutoshi Sumiya<sup>†</sup>, and Katsumi Tanaka<sup>†</sup>.

<sup>†</sup> Telecommunications Advancement Organization of Japan.

<sup>‡</sup> Graduate School of Science and Technology, Kobe University.

述を  $d$ , 有効区間を  $v$  とすると図 1 のようになる。

$$M = (Q, \Sigma, \Delta, \delta, \lambda, q_0, F)$$

- (a) 状態の有限集合  $Q$
- (b) 入力記号の有限集合  $\Sigma$
- (c) 出力記号の有限集合  $\Delta$
- (d) 状態推移関数  $\delta$
- (e) 出力関数  $\lambda$
- (f) 初期状態  $q_0 (\in Q)$
- (g) 最終状態の集合  $F (\subseteq Q)$

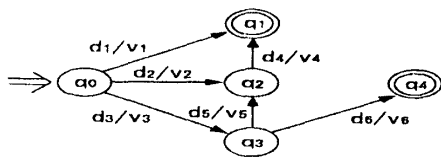


図 1: 有効区間導出オートマトン

複合イベント型の記述が生成されると、その記述に対応する有効区間導出オートマトンが起動し、過去の記述系列を入力記号列として遡っていく。記述系列が受理されたとき得られる出力系列は、各記述に対する有効区間の集合となっている。

## 4 システムの実装

### 4.1 システム構成

システムは、図 2 のように音声認識部、内容記述部 (インデックスレコード生成部、有効区間決定部、有効区間設定部)、インデックスデータベースから構成される。このうち、有効区間決定部と有効区間設定部を除いた部分を実装している。(図 3)

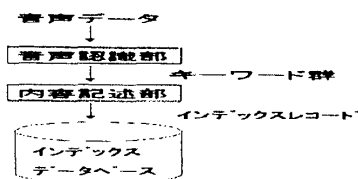


図 2: システム構成図

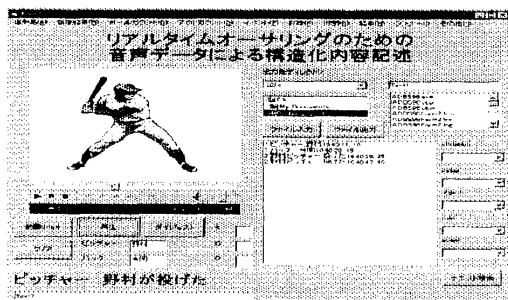


図 3: プロトタイプの画面例

## 4.2 処理の流れ

音声データの対象として、ラジオアナウンサーの実況中継を用いる。この理由は、ラジオの場合はテレビの実況中継に比べて、記述対象の実況が多いからである。逆に、テレビの場合は、解説の割合の方が多い。(テレビの実況と解説はそれぞれおよそ 8%, 92%) これに対して、ラジオの場合は、実況が主である。(ラジオの実況と解説はそれぞれおよそ 45%, 55%)

以下に処理の流れの概要を示す。

- 音声認識  
音声データを取り込み、キーワード群を生成する。
- 内容記述部  
キーワード群から、インデックスレコードを生成しタイムコードを付加する。
- 映像検索  
キーワード検索を用いて、対象となる映像を検索し再生を行う。

## 4.3 実験

1998.5.22 プロ野球 阪神対横浜 (甲子園) の毎日放送 (ラジオ) での実況内容 (1 回裏) を用いて、音声認識と内容記述についての評価実験を行った。この実験で用いたキーワード数は 108 個である。また、生成されたインデックスは 59 個であった。

## 5 終わりに

本稿では、音声データによるリアルタイム映像データの内容記述のための、構造化内容記述方式を提案した。また、プロトタイプを作成し、その評価実験についても述べた。今後は、蓄積型映像に対しても、この方式を適用する予定である。

**謝辞** この研究は、一部、日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高度処理の研究」および文部省科学研究費重点領域研究「高度データベース (No.275)」(課題番号 08244103) による。ここに記して謝意を表す。

## 参考文献

- [1] 川口 知昭, 土居 明弘, 飯嶋 亨, 田中 克己: "内容記述付き映像ストリームの配信とフィルタリング方式", 電子情報通信学会 第 9 回データ工学ワークショップ, DEWS'98, 1998
- [2] 川口 知昭, 土居 明弘, 角谷 和俊, 田中 克己: "放送型ネットワークにおけるライブ映像ストリームの編集と配信-ScoopCast-", 情報処理学会研究報告, Vol.98, No.57, 98-DBS-116-15, pp.111-118, Jul.1998
- [3] 今井 亨: "音声認識技術の放送への応用", NHK 技研 R&D No.46, Aug.1997
- [4] 恒川俊克, 山下洋一, 溝口理一郎: "キーワードスポットティングに基づくニュース音声の話題分類", 情報処理学会 音声言語情報処理研究報告, 98-SLP-20, Feb.1998
- [5] 広部 一弥, 牛尼 剛聡, 酒井 宏治, 孫 魯英, 渡邊 豊英: "イベントと状況変化の依存関係に基づいた野球中継のインデキシング支援", 第 115 回データベースシステム研究会 98-DBS-115-11, pp.87-94, 1998
- [6] 赤迫 貴行, 飯嶋 亨, 角谷 和俊, 田中 克己: "映像のリアルタイムオーサリングを支援するデータベース機能について", 第 57 回情報処理学会全国大会講演論文集 (3), pp.33-34, 1998
- [7] 赤迫 貴行, 飯嶋 亨, 角谷 和俊, 田中 克己: "映像データのリアルタイム内容記述方式とその実装", 電子情報通信学会 第 10 回データ工学ワークショップ, DEWS'99, 1999