I-70

文脈の流れと特徴的なイベントに基づく適応的な映像要約手法の検討

Study on Adaptive Summarization of Video based on Context Flow and Noticeable Events

宮森　恒†
Hisashi Miyamori

1．まえがき

大容量のデジタル映像を手軽に記録・蓄積できる環境が整うのに伴い，自分の興味のあるハイライトだけを短時間に視聴するための映像要約技術が重要になると予想される．

従来，要約の生成法としては，必要なデータを手作業で入力することを大幅に許容し，内容に直接関連した索引を生成・利用する方法が報告されている[1][2].

本稿では，可能な限り自動解析で得られる索引を用いて映像要約するアプローチをとる．索引としては，テニス映像を例として選手基本動作を自動識別した結果を利用する[4][5]．これらの索引を利用することで，オリジナル映像の意味的内容を保持した要約を，利用者の嗜好に応じて適応的に生成する手法について検討する．

2．要約の生成

図1に，要約生成の処理概要を示す．

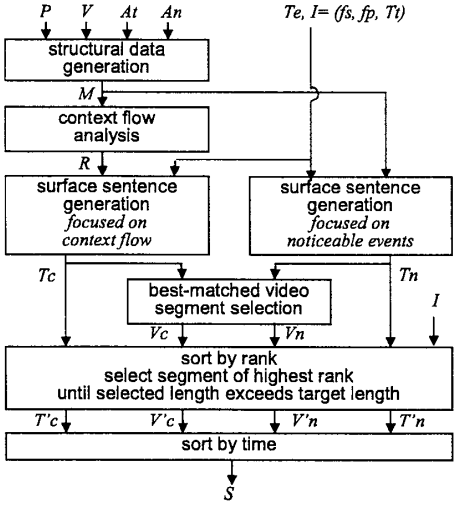

図1．要約生成の処理手順

入力データは，スコアデータ P，映像情報 V，時間順選手動作 A_t，顕著な選手動作 A_n，要約用テキスト要素 T_e，利用者入力 I の6種類のデータから構成される．

スコアデータ P については，映像中のテロップ領域の意味を特定し，他のメディアと関連付ける解析手法[3]等があるが，本稿記述時における実験映像はカメラから直接出力した未編集の映像であるため，暫定的にスコアとその開始・終了時刻を手入力で用意することとした．

A_t, A_n は，選手基本動作を自動識別した結果[4]から得られる索引で，動作ID，動作主ID，開始・終了時刻等から構成される[5]．A_t, A_n は，それぞれ，各選手の動作を時間順に表すための動作イベント，および，各選手のスーパープレイや顕著な動作に該当する動作イベントを表す．表1,2は本稿で用いた動作種別を表す．

表1．選手動作の時間順を表現するイベント

ID	動作名	ID	動作名
0	フォアハンドストローク	3	バックハンドボレー
1	バックハンドストローク	4	スマッシュ
2	フォアハンドボレー	5	サービス

表2．選手の顕著な動作に相当するイベント

ID	動作名	ID	動作名
0	サービスエース	4	スマッシュ成功
1	ダブルフォルト	5	スマッシュ失敗
2	サーブ＆ボレー	6	パッシング成功
3	ストロークエース	7	パッシング失敗

利用者入力 I は，利用者の興味・嗜好を反映させるための項目で，内容構成 f_s，着目選手 f_p，要約時間 T_t から成る．ここで，f_s は試合全体の流れを重視した要約を見たいのか，スーパーショットなど個々の特徴的イベントを重視した要約を見たいのか，といった内容構成に関する要求で，f_p はどちらの選手を重視するか，あるいは，均等な配分で要約を生成するかといった点を指定する項目である．T_t は，生成した要約の再生に許容できる時間を表す．

まず，試合全体のセット，ゲーム，ポイント，サーバ選手，ポイント獲得選手，各選手の時間順動作等を階層的に記述した構造データ M を生成する．これにより，試合中の任意時刻における選手の動作や試合の状態を把握する．

次に，M を用いて試合展開の概略を記述した内部表現 R を生成する．ここでは，優勢度 s とその推移から試合展開を分析した．あるセット（あるいはゲーム）の優勢度 s とは，各選手がどの程度有利にそのセット（あるいはゲーム）を展開したかを示す指標で，以下のように計算する．

$$s = d_i / d_{\max}$$

ここで，

d_i =(そのセット/ゲームの最後の1ゲーム/ポイントを取るまでにつけることのできたゲーム/ポイント差)

$d_{\max}$ =(そのセット/ゲームの最後の1ゲーム/ポイントをとるまでにつけることのできる最大のゲーム/ポイント差)

例えば，6-2 で決まったセットにおける優勢度は，$s_s = 3/5 = 0.6$ となる．また，2 回目の deuce で決まったゲームの優勢度は，40-A(得点に直すと 4-5)なので，$s_g = -1/5 = -0.2$ となる（マイナスは他方の選手が優勢であることを示す）．$|s_s| \le s_{s\_th1}$ ならば接戦，$|s_s| \ge s_{s\_th2}$ ならば一方的，$s_{s\_th1} < |s_s| < s_{s\_th2}$ ならば順調な展開であったと判断する．

これら優勢度 s_s, s_g の値が各セットにおいて，一方的であったか，接戦であったか，いずれでもなく順調な展開であったか，また，そのセット全体を通して持続したか，

†通信総合研究所, CRL

前半・後半で変化したか，といった分類を行い，試合展開を表した内部表現Rを獲得する．

次に，R, T_e, Iを用いて，出力される要約テキストに相当する表層文Tを生成する．

ここで，要約用テキスト要素T_eとは，試合展開や選手のさまざまな動作等を表現するのに十分な名詞・動詞・形容詞・副詞・助詞等の集合で，現在は ID とテキスト内容からなる簡単なテーブルで構成されている．

例えば，第 1 セットの試合展開がほぼ一方的であったとすると，予め定めてあるルールを参照することにより，表層文として次のような出力例を得ることができる．

「第 1 セットは，選手 Oka が 6-2 で危なげなくとった．」

ここでは各セットについて，Rで表現される試合展開に関連した表層文T_cと，顕著な選手動作に関連した表層文T_nの 2 種類を生成する．

次に，生成した表層文T_c, T_nにそれぞれ対応する典型的な映像情報V_c, V_nを，該当する範囲内から取得する．

ここで，試合展開に関連した映像としては，セットおよびゲームが終了する直前にポイントした選手の最後の打撃イベントを含むシーンや，接戦あるいは一方的と判断される全てのゲーム中で，最も頻繁に起きた動作イベントを含むシーンを各セットから選択することとした．

また，顕著な選手動作に関連した映像としては，パッシングやサービスエースといったスーパープレイに類するものに予め順位付けしておき，試合展開と着目選手f_pに対応する選手の動作イベントを順位の高いものから選択した．

最後に，利用者の指定時間T_t内に収まるようにランクの高い順にT_c', T_n'を選択する．同時に，対応するV_c', V_n'も選択する．選択した映像セグメントの合計時間がT_tより大きくなったら処理を終了する．最後に，時間順にソートすることで要約Sが得られる．

ここで，ランクとは，表層文Tに付随した数値で，その文の要素が文中あるいは文脈中でどれだけ重要であるかを示す指標である．一般に，表層文Tは要素$\{T_i\}$から成り立ち，$\{T_i\}$はその内容によって重要度が異なる．例えば，文の主語・述語にあたる部分は文として不可欠な要素としてランクが高く，修飾語句になると徐々にランクが下がるように設定している．

3．実験結果

要約の内容構成および着目選手を適宜変化させ，いくつかの指定時間で要約文および要約映像を生成した．

以下に，内容構成＝試合展開重視，着目選手＝*Oka*，要約時間＝30(sec)の設定で生成された要約文の例を示す．

「第 1 セットは，」「Oka のほぼ一方的な展開となり，」「選手 Oka が 6-2 で危なげなくとった．」「第 2 セットは，Hinomura が一時有利に試合を進めるものの」「ほぼ互角の接戦となり」「選手 Oka が 6-4 で逃げ切った．」

また，内容構成＝顕著なプレイ重視，着目選手＝*Oka*，要約時間＝30(sec)の設定で生成された要約文の例を示す．

「第 1 セットは，パッシングや」「サーブ&ボレーを決めるなどしてリードした」「選手 Oka が 6-2 で危なげなくとった．」「第 2 セットは，Oka がサービスエースや」「ストロークエースを決めるなどしたが，」「選手 Oka が 6-4 で逃げ切った．」

また，内容構成＝顕著なプレイ重視，着目選手＝*Hinomura*，要約時間＝30(sec)の設定で生成された要約文の例を示す．

「第 1 セットは，Hinomura がサーブ&ボレーや」「サービスエースを決めるなどして好プレーを見せたが」「選手 Oka に 6-2 で危なげなくとられた．」「第 2 セットは，Hinomura がパッシングや」「ストロークエースを決めるなどして好プレーを見せたが」「選手 Oka に 6-4 で逃げ切られた．」

内容構成や着目選手を変化させることにより，文脈の流れを保持しつつ，利用者の嗜好に応じた要約内容を構成できていることが確認できる．

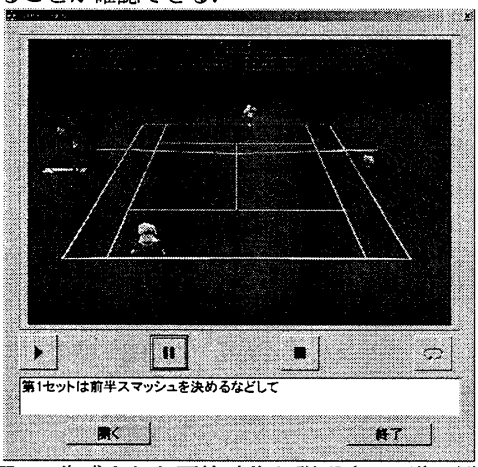

図 2．生成された要約映像と説明文の同期再生

要約再生時には，要約テキスト（上記で示した「」で囲まれた各部分に相当）とそれに対応した映像シーンが図 2 のように同期して再生され，ダイジェストを視聴できる．

実験映像を初めて視聴する被験者に対して要約映像を提示した予備実験の結果，オリジナル映像の代わりとなりうる意味内容を良好に伝達できていることを確認した．

今後は，要約生成部分のチューニングと，得られた要約の評価方法の改善を行い，より多くの被験者に対して本手法によるダイジェスト視聴の効果を検証する必要がある．

4．まとめ

映像中の文脈の流れと個々の特徴的なイベントに着目することにより，オリジナル映像の意味的内容を含む要約映像を，利用者の嗜好に応じて適応的に生成し，オリジナル映像の代わりとなりうる意味内容が良好に伝達できていることを確認した．

文献

[1] 植田和憲, 鎌原淳三, 下條真司, 宮原秀夫: “シナリオテンプレートによるストーリー性を重視したダイジェスト生成機構”, 情処 DBS-119-24, pp.139-144, 1999.

[2] 橋本隆子, 白田由香利, 真野博子, 飯沢篤志: “TV 受信端末におけるダイジェスト視聴システム”, 情処論文誌, Vol.41, No.SIG3(TOD6), pp.71-84, May 2000.

[3] 渡辺靖彦, 長尾 真: “画像の内容を説明するテキストを利用した画像解析”, 人工知能学会誌, Vol.13, No.1, pp.66-74, 1998.

[4] H. Miyamori: “Improving Accuracy in Behaviour Identification for Content-based Retrieval by Using Audio and Video Information”, ICPR, (to appear), 2002.

[5] 宮森 恒: “動作インデックスを用いた映像の自動注釈付けとその柔軟な内容検索への応用”, 情処 DBS127-2, FI67-2, pp.9-16, 2002.