

ユーザの視聴時行為を利用したビデオ映像編集支援システム

田中 博子[†] 木村 敏文^{††} 角谷 和俊^{††}

[†] 兵庫県立大学大学院応用情報科学研究科 〒 650-0044 兵庫県神戸市中央区東川崎町 1-3-3

^{††} 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 1-1-12

E-mail: [†]aa04e102@ai.u-hyogo.ac.jp, ^{††}{kimura,sumiya}@shse.u-hyogo.ac.jp

あらまし ビデオ映像の作成が容易になり、多くのビデオ映像を利用できる環境になってきている。しかしながら、大量のビデオ映像素材を効率良く編集し、視聴する環境は十分整っているとは言えない。本研究では、素材ビデオを視聴する行為からユーザの意図を抽出し、素材映像の編集のための支援作業に反映する方式を提案する。また、ビデオ映像に付与された視聴時のユーザの意図をビデオ映像とともに格納し、他のユーザが新たに視聴をした場合にも視聴を支援する機構について検討する。本稿では、視聴時のユーザの行為から意図を抽出する方式について述べ、提案する方式に基づくプロトタイプシステムの実装と評価について述べる。

キーワード ビデオ映像、編集、視聴、意図、インデックス、支援システム

A Video Editing Support System using Users' Viewing Action

Hiroko TANAKA[†], Toshifumi KIMURA^{††}, and Kazutoshi SUMIYA^{††}

[†] Graduate School of Applied Informatics, University of Hyogo

1-3-3 higashikawasaki-cho chuou-ku, Kobe, Hyogo, 650-0044 Japan

^{††} School of Human Science and Environment, University of Hyogo

1-1-12 Shinzaike-honcho, Himeji, Hyogo, 670-0092 Japan

E-mail: [†]aa04e102@ai.u-hyogo.ac.jp, ^{††}{kimura,sumiya}@shse.u-hyogo.ac.jp

Abstract Recently we can edit video easily and view a lot of video contents. However there are not enough environment to view and edit them efficiently. In this paper, we propose a method of editing video contents using users' intention from viewing behavior. Then we also discuss a system that supports to view that video for other users using their intentions. We describe a method to recognize users' intention and a prototype based on our proposed method.

Key words video stream, editing, viewing, intension, index, support system

1. はじめに

近年、質の向上や価格の低下などにより、デジタルビデオカメラが一家に一台所有されていることも珍しくなくなった。携帯電話やデジタルカメラでもビデオカメラ（動画撮影）機能は標準装備されている。このようにビデオ映像を個人が撮影する機会が益々増えている。しかしながら、それらのビデオ映像は一度か二度視聴されるだけで、あとはそのままの状態でストックされてしまうことが多い。素人の撮影者は、プロの撮影者とは違い、撮影時に後の視聴や編集作業を念頭において撮影することは稀である。構図や構成など関係なく撮影することがほとんどであり、必要以上にズームやパンを繰り返すカメラワークは逆効果で、視聴が苦痛になることも少なくない。そこで何らかの編集を行いたいが、効率的な編集作業が容易に行える環境は未だ整っていない。特にアノテーションすることのできない

一般ユーザのホームビデオやプライベートビデオの編集は困難である。

一方、構図や構成などが整っていないビデオ映像ほど、視聴時に画面の中であえて一部に注目したいという時がある。また複数人で視聴する場合、誰かが推奨するポイントを指示しながら、全員が映像の一部に注目して視聴する場合がある。本研究では、このような場合に注視という視聴行為を行うことを想定している。注視とは、視聴範囲を絞りその範囲を一定時間見ることとする。自分の興味などによって、あえて映像の一部に注目したい場合などに注視を行う。例えば、音楽会などのイベントで親が撮影したビデオ映像には、同じ服装をした子どもたちが大勢映っている。舞台全体を視聴したい場合もあるが、数人に焦点を絞って視聴したい場合や、一人の子どもに注目して視聴したい場合もある。その時に映像自体を変えてしまうのではなく、画面上を一部拡大し視聴することで、自由に見方が

変えられる。こうした複数の注視情報を集めることで編集を行うことができれば、容易な編集が可能になる。このような視聴形態を導入することで、撮影者自身も撮影時にカメラを固定して、応援や鑑賞に専念できるというメリットがある。

また注視を行う例に、普段何気なく撮影するプライベートショートビデオがある。実際視聴してみると、撮影時の意図とは違った興味深い対象が、思いがけず映ることがある。例えば、「携帯電話のビデオカメラモードで友人を撮影しているにもかかわらず、友人の後ろにかわいい犬が偶然映っていた」といった場合である。このように日常だからこそ何が映るかはわからない。しかし、手間をかけて編集するのでは、手軽に撮影して楽しめるショートビデオの利点が薄れてしまう。

他にも、例えば修学旅行などの行事で撮影されたビデオ映像を生徒がそれぞれ視聴し、注視した情報をもとにビデオを編集することで、生徒達自身にとって興味深い場面を抽出することができる。

本研究では、視聴行為から注視情報を抽出し、ビデオ映像の編集支援を行うシステムについて検討する。具体的には、注視行為から得られる注視時間、注視領域、および視聴順序の関係を使うことで、編集情報の抽出を行うフレームワークを提案する。

以降、2節で関連研究、3節でユーザの視聴行為とその抽出について、4節で編集支援について述べる。5節で実装について、6節では今後の課題について述べる。

2. 関連研究

Lynn Wilcox らが、Hitchcock システム [1], [2] の中で、カメラワークの動きを利用することでキーフレームを自動抽出し再構成する編集方法を提案している。このシステムではユーザインタフェースを工夫し、キーフレームの選択などでユーザの意思を反映することも考えられている。Juan Casares らによる Silver システム [3] は、ニュース映像とオーディオそれぞれによるシーン検出の特性の差をメタデータとしてビデオ編集支援を行っている。また、天野らによる映像文法に基づく映像編集支援システム [4] は、ビデオ映像の編集支援の方法を提案しているが、映像中の情報を利用したものである。これらは、映像の編集の自動化により、ユーザの補助を行うことを目的とした研究であり、ユーザの意図をより強く反映することはできない。また、視聴行為には特に触れていない。

視聴行為を利用した研究としては、福村らが、視聴履歴に基づくデジタルコンテンツの個人化された提示手法 [5] を提案している。これは Web ページ閲覧者の視聴履歴に基づき、個人化した Web ページを作成するというものである。また、清光らは、閲覧履歴に依存した Web コンテンツの動的生成方式 [6] を提案している。この研究では Web ページのコンテンツ付加にルールを設けることで、閲覧履歴に基づき Web ページ内容を変化させることができる。また、J. Freire による WebVCR [7] は、ユーザが Web に対して行う行為を記録し、後で再生することができる。しかし、いずれの研究も Web ページのみを対象としている。映像の視聴行為を利用した研究では、ユーザの視聴

形態を考慮した映像品質制御手法の提案 [8] が、加藤らから提案されている。映像の視聴形態（ザッピングながら見/視聴に専念）によって映像の配信品質を制御するものである。映像の視聴時における注視情報を利用する点は、本研究と類似しているが、他の作業をしながらの視聴スタイルを想定している点が大きく異なる。また、映像配信サービスを対象とした品質制御に関する研究であり、映像の編集支援のための研究ではない。

吹野らは、ニュース記事を利用したサッカー要約映像の生成 [9] を提案している。ニュース記事とサッカー映像の内容を実況コメントで対応づけることによって、ビデオメタデータを生成し、要約映像を生成する手法である。これは映像のメタデータから重要なシーンを抽出し、さらにそれをつなぎ合わせるだけではなく、時間を考慮した上で要約を行っている。また、同じ映像に対して複数のニュース記事があることから、複数の視点での要約映像を生成することができる。しかし、これはサッカー映像の特性を活かしたものであるため、その他の映像にはそのまま適用することができず、一般のビデオ映像には向いていない。

映像作家の宇田は、実写を使ったインタラクティブムービーを制作している [10]。宇田の表現は、インタラクティブムービーの中でも、ユーザがマウスを画面上へ自由に置くことで、そのマウスの位置により次の映像を展開するマルチストーリービデオである。本研究との関連は、ユーザの注視点によって、表示される映像が変化する点である。一般の映像の編集作業とは大きく異なり、ユーザに編集作業を意識させることはない。しかし、リアルタイムに表示を変化させることが目的であるため、その視聴行為を後の編集に利用することはできない。また、個人の注視行為のみをその対象としており、効果的な編集として他人が利用することができない点も本研究と異なっている。

時区間同士の演算では、スジットらによって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 [11] の中で glue 演算が提案されている。これは、ビデオ映像からキーワード検索をする際に、ノイズが存在してもまとまりのある区間を抽出することができる演算である。しかし、複数の注視行為から編集を行う際には、許容して範囲を拡張するだけではなく、注視行為として意味のある区間を抽出することが必要である。

3. ユーザの視聴時行為

3.1 視聴時行為

映像の視聴時行為には、受動的視聴と能動的視聴がある。ここで扱うのは、ユーザ（視聴者）自身が視聴する際、ビデオ映像に対し何らかの意図を持って行う行為、すなわち能動的視聴を対象とする。

能動的視聴の中には、何かに注目した視聴があげられる。時区間を注目する視聴時行為として、例えば、再生、早送り、シーンマーク、シーンスキップ、シーンリピート、巻き戻しや時間指定巻き戻しがある。特に音量を扱ったものには、音量通常操作（アップ、ダウン）、ミュートなどがある。また、時点を注目する視聴行為には、例えば、一時停止、イメージマーク、プリントアウトがある。また、画面領域への注目として、例えば、

画面のズーム、拡大縮小、レーザーポインター等の操作があげられる。

視聴時行為とは、いずれの行為もビデオ映像を視聴することが目的であり、編集することを目的とした行為ではない。

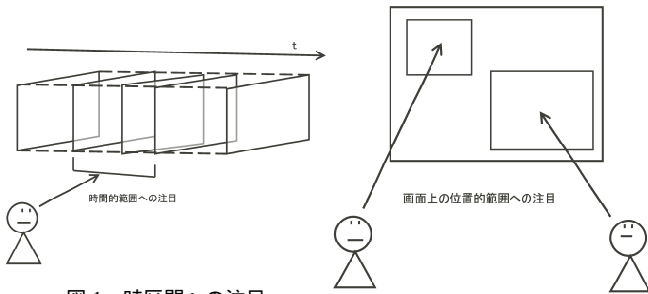


図1 時区間への注目

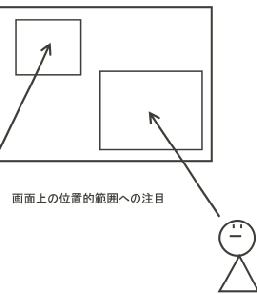


図2 領域への注目

一方、視聴情報を利用して素材ビデオを編集する際は、完全な編集を行うのではない。本システムにおいて視聴とは、半編集された映像をさらに視聴することであり、初めての視聴を行ったユーザとは異なるユーザが新たに視聴する場合や、同じユーザ自身が繰り返し視聴する場合がある。これを本研究では半編集 (semi-edit) と呼んでいる。なお、二回目以降の視聴で新たに情報を加えていくことが可能であるため、視聴を繰り返すことで、形は変化し続ける。

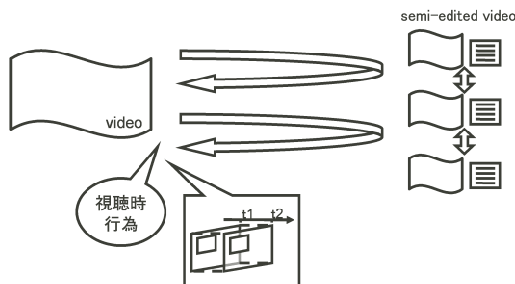


図3 半編集

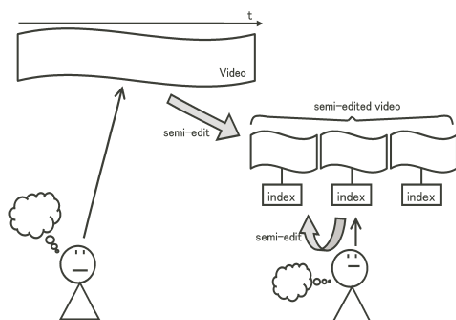


図4 視聴時行為による半編集

3.2 視聴情報

3.2.1 視聴情報からの編集情報の抽出

画面上の一部に映っているある対象に視点を合わせ、その範囲を一定時間視聴する行為を視聴とよぶ。視聴とは、ただ何気なく目に入ったという受動的な行為ではなく、能動的に見る行為である。

視聴者は気になる対象が出てきた時、注視行為を行いながら視聴する。複数人での視聴の場合は、まず一人目が視聴する際、ユーザ自身が興味のある範囲を選び注視を行いながら視聴する。その後二人目のユーザは一人目のユーザの注視行為を提示されながら視聴し、自分が興味のある範囲を選び注視する。このように複数ユーザの注視行為の組み合わせから、編集情報を抽出し、ビデオ映像の編集を支援する。

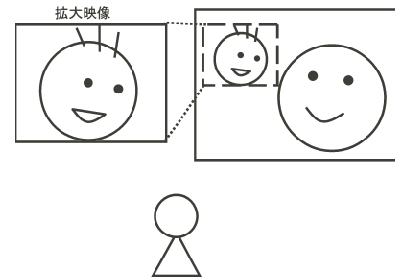


図5 視聴イメージ

3.2.2 注視情報の記述

注視を記述する内容は以下の通りである。

- 注視実行区間 (時区間)
 - 注視実行開始時間 映像中の時間 t_s
 - 注視実行終了時間 映像中の時間 t_e
- 注視実行位置 (画面領域)
 - 注視実行左上座標 画面上の位置 (x_1, y_1)
 - 注視実行右下座標 画面上の位置 (x_2, y_2)
- 視聴者識別子

4. 映像編集支援

4.1 編集と半編集

編集 (edit) とは、素材ビデオ映像そのままの状態ではなく、素材ビデオ映像とは異なる新しいビデオ映像を作成することである。素材ビデオ映像とは、何も加工していない撮影した状態のビデオ映像を指す。

半編集ビデオ (semi-edited video) とは、複数人のユーザの注視行為の組み合わせから、編集に適した出力情報を抽出し、それらをメタデータとして付与したビデオ映像である。メタデータは新たに視聴を行うたびに追加・更新される。

4.2 編集情報によるビデオ映像の編集方法

視聴者の注視行為から抽出するのは、注視時区間、注視領域、および視聴順序であり、これらは互いに独立である。

4.3 注視時区間

4.3.1 注視時区間の抽出

注視時区間の関係において、ある視聴者が注視行為を行った時区間は、他の視聴者の注視行為にとってどのような意味をもつかという関係を考え、編集の際に必要な時区間を判断する。つまり、注視していない時区間の重要度を、複数ユーザの注視データから算出する。

一方のユーザが注視した時区間を、二人のユーザの注視時区

間の重なりによって分割し、もう一方のユーザの視聴行為をもとに重要度を求める。

- 両方のユーザが注視を行った時区間
あるユーザが注視をし、他のユーザも注視をした時区間は、基本的に編集結果として出力する。これは、どちらも注視を行った時区間であることから、その時区間は重要であると考えられるためである。
- 一方のユーザだけが注視を行った時区間
あるユーザが注視をし、他のユーザは注視をしていない時区間は、その時区間自身を含めた注視されていない空白区間のうち、その時区間が占めている割合により判断する。割合が閾値 α 以上であった場合は出力せず、閾値 α 未満であった場合は出力する。これは、一方のユーザだけが注視を行った時区間が、もう一方のユーザの注視の空白区間の多くを占めている場合、その時区間の重要度は低いと考えられるからである。言い換えると、一方のユーザがその近くで注視を行っていないときに、もう一方のユーザがあえて注視をしている時区間には価値があり、出力すべきと考えるためである。
- 両方のユーザが注視を行わなかった時区間
両方のユーザが注視を行わなかった時区間は、どのユーザにも注視されていないという理由から基本的には出力しない。ただし、二つの注視時区間の間にある区間、すなわち注視の空白区間は、注視時区間が互いに接近しているという条件を満たす場合のみ出力する。二つの注視時区間が重なっていない場合、一方の注視時区間は、時系列で見たときに重なる他のユーザの空白区間に、完全に含まれる形となる。よって、その空白区間に占めている割合を見るだけでなく、空白区間の間で前後のどちらの注視に影響しているかを考慮する必要がある。これは、前後の空白区間の割合を閾値 β によって判断する。

同一のビデオ映像に対して、ユーザ A、ユーザ B の順で視聴を行ったとする。ユーザは視聴時にいくつかの時区間で注視を行うとする。ユーザ A の注視時区間を $A_i \{A_1, A_2, A_3, \dots, A_m\}$ ($i = 1, 2, 3, \dots, m$)、ユーザ B の注視時区間を $B_j \{B_1, B_2, B_3, \dots, B_n\}$ とする ($j = 1, 2, 3, \dots, n$)。また、 $t_s(A_m) < t_e(A_m)$ の時、 A_m の注視開始時間を $t_s(A_m)$ 、 A_m の注視終了時間を $t_e(A_m)$ とし、 $[t_s(A_m), t_e(A_m)]$ とあらわす。同様に $t_s(B_n) < t_e(B_n)$ の時、B の注視開始時間 $t_s(B_n)$ 、 B_n の注視終了時間を $t_e(B_n)$ とし、 $[t_s(B_n), t_e(B_n)]$ とあらわす。

注視の間の時区間、すなわち注視の空白区間 (blank) を $blank(A_i, A_{i+1})$ とあらわす。 $blank(A_i, A_{i+1}) = [t_e(A_i), t_s(A_{i+1})]$ である。ただし、最初の注視 A_1 の前の空白区間は、ビデオの開始時間から $t_s(A_1)$ までを指し、最後の注視 A_m の後の空白区間は、 $t_e(A_m)$ からビデオの終了時間までとする。また、 $t_e(A_i) < t_s(B_j)$ のときは、 A_i と B_j の間にできた空白区間を $blank(A_i, B_j)$ と表し、 $blank(A_i, B_j) = [t_e(A_i), t_s(B_j)]$

とする。

4.3.2 注視時区間の抽出例

(1) 注視時区間に重なりがある場合

二人の注視時区間に重なりがある場合の例として、 A_i と B_j の二つの注視時区間が一部で重なっている場合について述べる (図 6)。

$[t_s(B_j), t_e(A_i)]$
 A_i と B_j の注視と重なっているため、出力する。
 $[t_s(A_i), t_s(B_j)]$
 $t_s(B_j) - t_s(A_i) / blank(B_{j-1}, B_j) < \alpha$ より出力する。
 $[t_e(A_i), t_e(B_j)]$
 $t_e(B_j) - t_e(A_i) / blank(A_i, A_{i+1}) \geq \alpha$ より出力しない。

よって図 6 の場合は、 $[t_s(A_i), t_e(A_i)]$ が出力となる。通常は、誰かが注視をした時区間として $[t_s(A_i), t_e(B_j)]$ が出力されるか、もしくは二人が注視をした時区間として $[t_s(B_j), t_e(A_i)]$ が出力されると考えられるが、本方式では、注視時区間と空白区間から、その重要度を考慮することができる。

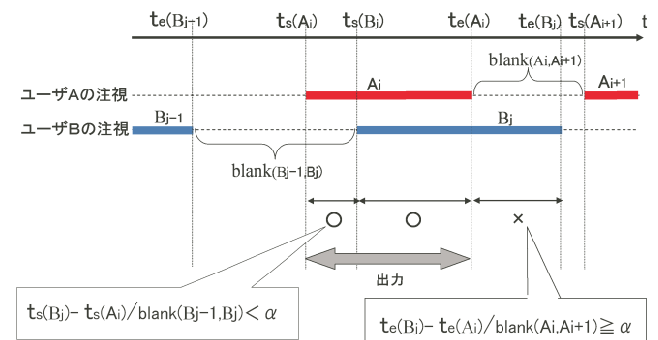


図 6 注視時区間に重なりがある場合

(2) 注視時区間に重なりがない場合 (I)

注視時区間に重なりがない場合について述べる。

$A_i [t_s(A_i), t_e(A_i)]$
 $t_e(A_i) - t_s(A_i) / blank(B_{j-1}, B_j) \geq \alpha$ より出力しない。
 $B_j [t_s(B_j), t_e(B_j)]$
 $t_e(B_j) - t_s(B_j) / blank(A_i, A_{i+1}) < \alpha$ より出力する。
 $blank(A_i, B_j) [t_e(A_i), t_s(B_j)]$
 $blank(A_i, B_j) / blank(B_{j-1}, A_i) > \beta$
 $blank(A_i, B_j) / blank(B_j, A_{i+1}) < \beta$ より出力しない。

図 7 の場合の出力は、 $[t_s(B_j), t_e(B_j)]$ となる。通常、この場合の出力は、 $[t_s(A_i), t_e(A_i)]$ と $[t_s(B_j), t_e(B_j)]$ 、もしくは間の $blank(A_i, B_j) = [t_e(A_i), t_s(B_j)]$ を許容して $[t_s(A_i), t_e(B_j)]$ が抽出されると考えられる。しかし、本方式では時区間の前後の間隔を考慮し、一方のユーザだけが注視している時区間を出力する。

(3) 注視時区間に重なりがない場合 (II)

注視時区間に重なりがない場合 (I) とは異なり、 $blank(A_i, B_j)$ を出力する場合の例 (図 8) である。

$A_i [t_s(A_i), t_e(A_i)]$
 $t_e(A_i) - t_s(A_i) / blank(B_{j-1}, B_j) < \alpha$ より出力する。

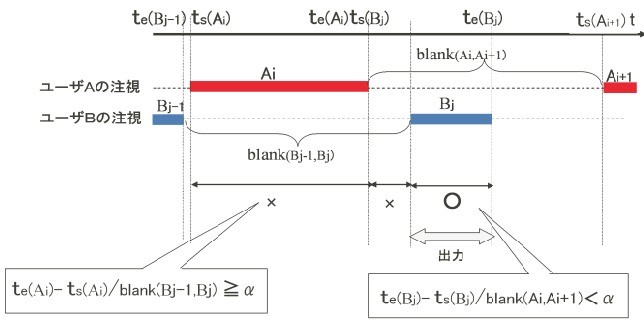


図7 注視時間に重なりがない場合 (I)

$B_j[t_s(B_j), t_e(B_j)]$
 $t_e(B_j) - t_s(B_j) / \text{blank}(A_i, A_{i+1}) < \alpha$ より出力する。
 $\text{blank}(A_i, B_j) [t_e(A_i), t_s(B_j)]$
 $\text{blank}(A_i, B_j) / \text{blank}(B_{j-1}, A_i) < \beta$
 $\text{blank}(A_i, B_j) / \text{blank}(B_j, A_{i+1}) < \beta$ より出力する。

$\text{blank}(A_i, B_j)$ について説明する。注視時区間 A_i は $\text{blank}(B_{j-1}, A_i)$ の時区間に完全に含まれている。よって、 $\text{blank}(B_{j-1}, A_i)$ と $\text{blank}(A_i, B_j)$ の長さを比較することで、 $\text{blank}(B_{j-1}, A_i)$ の中でも A_i が B_j 側で注視されているかを考える。つまり、 $\text{blank}(A_i, B_j) / \text{blank}(B_{j-1}, A_i) < \beta$ (閾値) であれば、 A_i は B_j に近い時区間で注視されたと考えることができる。同様に $\text{blank}(A_i, B_j) / \text{blank}(B_j, A_{i+1}) < \beta$ を比較することで $\text{blank}(A_i, A_{i+1})$ の中で B_j は A_i 側で注視されているのかを考える。この例の場合、どちらも閾値 β 未満であるため、 $\text{blank}(B_{j-1}, A_i) > \text{blank}(A_i, B_j)$ かつ $\text{blank}(A_i, B_j) < \text{blank}(B_j, A_{i+1})$ を満たす。よって、 A_i 、 B_j ともに両者に近い位置で注視していることから、間の $\text{blank}(A_i, B_j)$ は単なる空白区間ではなく、意味がある時区間である可能性がある。したがって、 $\text{blank}(A_i, B_j)$ を出力する。

通常、離れた時区間は関係がないものとして、その間の空白区間は出力されないと考えられるが、本方式では、注視の前後との間隔を比較することによって、注視されていない時区間の重要度を判断し、出力する場合がある。

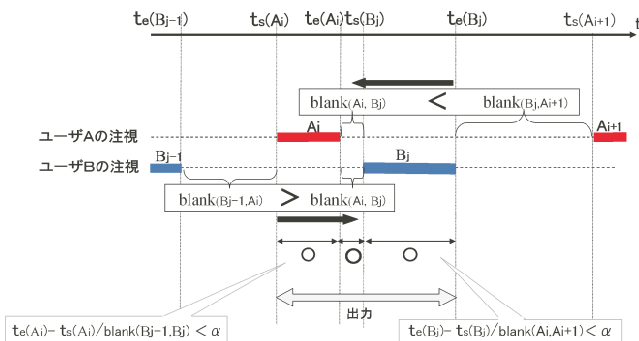


図8 注視時間に重なりがない場合 (II)

4.4 注視領域

注視領域の関係は、ユーザ同士の注視領域による重なりの有無、領域の中心点の位置ともう一方の領域の包含関係、領域面

積の大小関係を使用する。例えば、中心点の位置が同じ場合、二つの注視の対象が同じである可能性を考えることができる。

領域面積の大小関係では、注視としての価値を考える。基本的には面積が小さい方に価値がある。なぜならば、注視という行為は範囲を絞って視聴することであると考えられるためである。また、中心点の位置関係を合わせて考えることにより、同じ対象であっても領域をあえて広げて注視したという注視の意図を反映することができる。

同一のビデオ映像に対して、ユーザA、ユーザBの順で視聴を行った場合について述べる。ユーザBはユーザAの注視を提示されていると仮定する。

時区間の関係が図9のように、 $t_s(A_i) < t_s(B_j)$ である場合を考える。時区間の出力は、 $B_j[t_s(B_j), t_e(B_j)]$ である。領域には重なりがあり、二つの領域の中心点が互いの領域に含まれている場合、かつユーザBの注視面積がユーザAの注視面積より小さい場合である。対象の中心は同じであることから、ユーザBがさらに注視領域範囲を絞ったことになる。この場合、 $[t_s(A_1), t_e(B_1)]$ の時区間においてユーザBの領域を出力する。通常、編集時には $[t_s(A_1), t_s(B_1)]$ でユーザAの注視領域を出力し、 $[t_s(B_1), t_e(B_1)]$ でユーザAの注視領域とユーザBの注視領域を含む最小矩形領域、つまりユーザAの注視領域を出力すると考えられる。しかし本方式では、注視領域の重なりと中心点の位置からその対象を考慮した上で、領域範囲をさらに絞ったユーザAの注視領域を出力する。

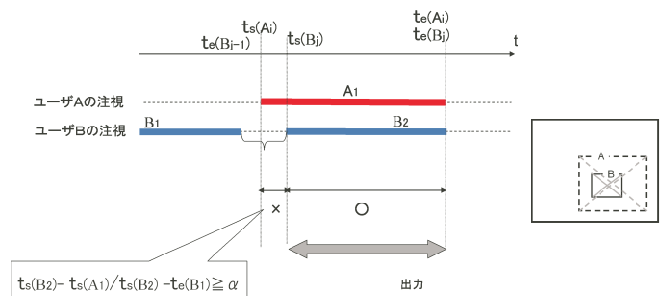


図9 領域例

4.5 視聴順序

一般的に複数ユーザの視聴では、ユーザの視聴順序により、提示される注視データが異なるために、出力する時区間、領域が異なる。例えば、ユーザAとBの注視データで演算した後にユーザCの注視データを演算した場合と、ユーザBとCの注視データを演算した後、ユーザAの注視データを演算した場合では、出力結果が異なる。

5. 実装

4節で述べた提案手法を基にプロトタイプシステムを作成した。システムの構成、実装環境、ユーザインタフェースについて説明を行い、使用例を示す。

5.1 システム構成

プロトタイプシステムの構成を図10に示す。システムは大きく分けて、表示部、制御部、提示注視データ生成部、編集映

像生成部，データベースで構成する。

表示部では，ユーザからの要求により，制御部に映像の再生要求を行う。また，再生要求によって，制御部から送られてくる映像のコンテンツデータおよび提示注視データを基にして，映像を再構成し，ユーザに提示する。視聴しているユーザから注視情報を取得して，制御部に渡す役目も果たす。

制御部は，要求やデータの制御を行う役目を果たす。表示部より受け取った映像再生要求に基づき，再生するために必要なデータを表示部に送る。このとき，制御部は再生する映像のコンテンツデータを取得し，さらにその映像の提示注視データが存在するかどうかをデータベースに問い合わせ，存在する場合は提示注視データを取得する。これらのデータを表示部に送る。また，ユーザが注視した情報を制御部を通じて，注視データデータベースに保存する。

提示注視データ生成部では，次にユーザが視聴する際に提示するデータを生成する役割を果たす。視聴することによって取得される注視情報を基に提案手法によって，次に視聴するユーザに提示する注視データを生成する。生成したデータは一度提示注視データデータベースに保存し，ユーザの要求で制御部へと提供する。

編集映像生成部では，ユーザの編集要求を制御部から受け，その時点での提示注視データおよび映像データを用いて，映像の編集を行う。生成された映像は制御部を通して，ユーザに送る。

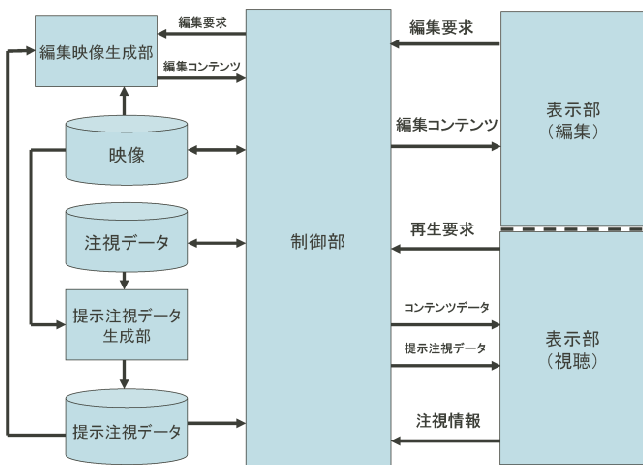


図 10 プロトタイプのシステム構成図

5.2 ユーザインタフェース

プロトタイプはテレビの画面を想定し，PC 上で構築した。プロトタイプのユーザインタフェースを図 11 に示す。上部には映像再生画面を配置した。この画面上では映像を再生する他に，ユーザの注視情報を提示・抽出する役割がある。再生画面の右側では，ラジオボタンにより，あらかじめ決めてある注視領域の大きさをユーザが指定することができる。注視領域の大きさの変更を行うことができる。注視情報（注視領域の中心座標や注視開始時間，終了時間）の表示も行う。また，別に注視領域のみを表示する画面がある。

実際にユーザが注視行為を行う際の手順を以下に示す。

1. システムを起動する。

2. システムによって，再生される映像を視聴する。その際に，以前に視聴したユーザの注視情報も同時に表示される。
3. ユーザはマウスを操作し，システムに対してリクエストを送る。ユーザができるリクエストを次に示す。
 - (1) 注視領域の大きさの変更
 - (2) 注視する時区間と領域の指示
 - (3) 映像再生の終了
4. 映像再生が終了するまで，手順 2,3 を繰り返す



図 11 TVを想定したユーザインタフェース例

他の利用環境としては，携帯電話でのショートビデオの交換も想定される。例えば，携帯電話の画面にビデオ映像を表示させ，十字キーで領域を指定し，拡大領域の大きさをボタンで指定するインタフェースが考えられる。図 12 は携帯電話でビデオ映像の領域を選択する場合のユーザインタフェースである。



図 12 携帯でのユーザインタフェース例

5.3 実装環境

プロトタイプシステムは，映像再生機能を除いて，Microsoft Visual C# .NET によって実装している。ユーザインタフェースは映像画面上の任意の点や制御ボタンを選択することにより制御を行う。映像再生機能は，Windows Media Player SDK(Software Development Kit) で実装している。Windows Media Player で再生できる映像ファイルすべてが素材映像として利用可能である。また，提示注視データ生成部は映像が終了した際に外部で起動する。このモジュールを取り替えることだけで，修正，

拡張が可能である．プロトタイプシステムの構築に用いたマシンは，OS:WindowsXP HOME Edition/CPU: Celeron 2.53GHz / Memory: 512MB である．

5.4 処理の流れ

プロトタイプではユーザがシステムを起動することにより映像の再生が始まり，ユーザ自身が注視したい領域をマウスを用いて指定する仕組みになっている．プロトタイプシステムの処理の流れを以下に示す．

1. ユーザによってプロトタイプシステムが起動される．
2. あらかじめ選択されている映像ファイルをシステムに読み込む．
3. 選択された映像ファイルに対して，提示すべき注視情報の有無をシステムが確認する．
4. 提示すべき注視情報がデータベースに存在する場合は，該当する注視情報をシステムが読み込む．
5. システムは映像を再生しながら，注視情報に記されている時間・領域を再生映像に加えてユーザに提示する．
6. 映像再生中，システムはユーザからの以下のリクエストをマウスによって随時受け付ける．
 - (1) 注視領域の大きさの変更
 - (2) 注視する時区間と領域の指示
 - (3) 映像再生の終了
7. システムはユーザからのリクエストに対して，リクエストにあった処理を行う．
 - (1) 指定された注視領域の大きさへ変更する．
 - (2) 注視領域の大きさ情報と注視した時区間の開始時間と終了時間の取得し，メモリ上に記憶しておく．
 - (3) 映像再生終了要求により，映像再生終了と同時にシステムで取得した注視情報をデータベースに一括して保存し，提示注視データ生成部を起動する．
8. 映像再生が終了するまで，4 に戻り繰り返す．

次に再生終了時に実行される提示注視データ生成部（以下，データ生成部）の処理の流れを以下に述べる．

1. データ生成部はユーザの再生終了リクエストによって，起動する．
2. データ生成部は映像が再生されていた際に提示されていた注視情報とシステムが取得した注視情報を読み込む．
3. 読み込まれた注視情報から，注視時区間の重なりや途切れを計算し，時区間領域の組み合わせを算出する．

ユーザが視聴するたびに，上記の処理が繰り返される．

5.5 適応例

音楽会で子どもたちが合奏しているビデオ映像を視聴する例を挙げる．ただしカメラの視点は舞台上に固定された状態で撮影したものとする．視聴者 A，B が注視を行いながら視聴する．

途中まで視聴者 A，B とともに，数十秒ずつの間隔で注視していた．B は全体での演奏になったのであえて注視をやめたが，A はその間に指揮者を長めに注視した (A_i)．次に指揮者がリコーダーのソロパートに振ったので視聴者 A は注視をやめた．

その時 B は，A の注視の影響でリコーダーの子どもを注視した (B_j)．その後はソロパートもなく，落ち着いた調子になったため，視聴者 A，B はしばらく注視をしなかった．この場合，本方式では $[t_e(A_i), t_s(B_j)]$ で視聴者 A 視聴者 B の注視領域を含む最小矩形領域， $[t_s(B_j), t_e(B_j)]$ で視聴者 B の注視領域を出力する．

領域の関係から，視聴者 A と B による注視対象は異なっていることがわかる．しかし，本システムでは二人とも注視していなかった視聴者 A，B の注視時区間の間の区間 $[t_e(A_i), t_s(B_j)]$ を， B_j と共にまとめた区間として出力する．これは視聴者の複数の注視の中で A_i, B_j 二つの注視時区間が近いことから，大きく影響を受けていると考えられるためである．この場合，影響を受けた時区間は映像中の盛り上がり場面でもあり，内容とも合っている．このように，複数人の注視データから編集を支援することは実現可能であると考えられる．

6. 終わりに

本稿では，視聴時の注視行為から得られる注視時区間情報，注視領域情報，および注視順序の関係により，編集情報を抽出可能にし編集支援を行う方式を提案した．今後の課題として，

- 注視領域面積，注視領域同士の距離が出力にどう影響するか
- 複数人が一緒に視聴する場合のユーザインタフェースについて

- 影響を与える視聴行為の提示方法

などが挙げられる．また，視聴回数やコメントによる半編集への影響，注視以外の視聴行為の利用（例えば，連続的なズーム，早送り，巻き戻し等），さらに，編集結果としての効果音やエフェクトの利用，異なる複数本のビデオ映像への適応についても検討を行う必要がある．

謝 辞

本研究の一部は，平成 16 年度科研費基盤研究 (B)(2)「Web アーカイブと映像アーカイブを融合した次世代デジタル・ライブラリに関する研究」(課題番号：16300028) 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] Lynn Wilcox, John Doherty, and Andreas Girgensohn, "A Hitchcock Assisted Video Edited Night at the Opera," ACM Multimedia 2002, Dec.2002
- [2] Andreas Girgensohn, Sara Bly, John Boreczky, and Lynn Wilcox, "Home Video Editing Made Easy - Balancing Automation and User Control," In Proceedings of Human-Computer Interaction (INTER-ACT '01), IOS Press, pp. 464-471, Jul.2001
- [3] Juan Casares, Brad A. Myers, A. Chris Long, Rishi Bhatnagar, Scott M. Stevens, Laura Dabbish, Dan Yocum, and Albert Corbett, "Simplifying Video Editing Using Metadata." In Proceedings of Designing Interactive Systems (DIS 2002), Jun.2002
- [4] 天野美紀, 上原邦昭, 熊野雅仁, 有木康雄, 下條真司, 春藤憲司, 塚田清志, "映像文法に基づく映像編集支援システム", 情報処理学会論文誌, Vol.44, No.03, pp.915-924, 2003
- [5] 福村真哉, 中野 賢, 春本 要, 下條真司, 西尾章治郎, "視聴履歴に基づくデジタルコンテンツの個人化された提示手法", 第 13 回 データ工学ワークショップ (DEWS 2002), A4-5
- [6] 清光英成, 竹内淳記, 田中克己, "閲覧履歴に依存した Web コンテ

- ントの動的生成法”, 第 11 回データ工学ワークショップ (DEWS 2000), 2A-5
- [7] J. Freire, V. Anupam, B. Kumar, D. Lieuwen, “Automating Web Navigation with the WebVCR”, Proceedings of the 9th International World Wide Web Conference, Amsterdam, Netherlands, May.2000
 - [8] 加藤由花, 箱崎勝也, “ユーザの視聴形態を考慮した映像品質制御手法の提案”, 電子情報通信学会技術研究報告, CQ2003-37, pp.25-28, 2003.
 - [9] 吹野 直紀, 馬 強, 角谷 和俊, 田中 克己, “ニュース記事を利用したサッカー要約映像の生成”, 第 14 回データ工学ワークショップ (DEWS 2003), 8-P-03
 - [10] Uda Atsuko's Site.
<http://wwwold.iamas.ac.jp/makura/index.html>
 - [11] プラダン 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, Vol.40, No.SIG3, (TOD1) Jan.1999