

MPEG-7を利用したウェアラブルカメラ映像の索引付け手法

兵 清弘[†] 天笠 俊之[†] 吉川 正俊^{†,‡} 植村 俊亮[†]

[†] 奈良先端科学技術大学院大学 情報科学研究科

[‡] 国立情報学研究所 ソフトウェア研究系

[†] 〒 630-0101 奈良県生駒市高山町 8916-5

[‡] 〒 101-8430 東京都千代田区一ツ橋 2-1-2

{kiyohi-h, amagasa, yosikawa, uemura}@is.aist-nara.ac.jp

あらまし 本研究では、マルチメディアコンテンツに対し内容を記述する規格である MPEG-7 を用いて、ウェアラブルカメラ映像への索引付けを行い、映像検索を可能にする枠組みを提案する。これによりウェアラブルカメラで日常生活を撮影した大量の映像からでも、利用者は所望の場面を検索することが可能となる。具体的には、ウェアラブルカメラでの映像撮影と同時に時間情報、位置情報 (GPS による緯度・経度情報)、アノテーション情報を取得、これらの情報を用いて MPEG-7 ファイルを自動生成する。さらに利用者の検索要求に合わせた MPEG-7 ファイルを動的に生成することにより、利用者との対話的な検索システムを提案する。同時にダイジェスト作成の手法についても議論する。

キーワード 映像検索, MPEG-7, ウェアラブルカメラ, 地理情報, GPS

An Indexing Method for Video Data recorded by Wearable Cameras using MPEG-7

Kiyohiro Hyo[†], Toshiyuki Amagasa[†], Masatoshi Yoshikawa^{†,‡}, and Shunsuke Uemura[†]

[†] Graduate School of Information Science, Nara Institute of Science and Technology

[‡] Software Research Division, National Institute of Informatics

[†] 8916-5 Takayama, Ikoma, Nara 630-0101, Japan

[‡] 2-1-2 Hitotsubashi, Chiyoda, Tokyo 104-8430, Japan

{kiyohi-h, amagasa, yosikawa, uemura}@is.aist-nara.ac.jp

Abstract In this paper, we propose an indexing method for video data recorded by wearable cameras using MPEG-7. This enable us to retrieve the desired scene(s) from such as enormous video data recorded by wearable cameras. In detail, a user gets temporal and geographical information such as latitude and longitude, as well as video data. Using these information, we create MPEG-7 files automatically. Moreover we propose user interactive retrieval system which generate MPEG-7 files with matching retrieval demand of user, and discuss a method for creating digests.

Key words Video Data Retrieval, MPEG-7, Wearable Camera, Geographical Information, GPS

1 はじめに

ここ数年、身体に装着して利用するタイプの計算機、いわゆるウェアラブルコンピュータが注目されている。最近ではその開発が急ピッチで進み、我々が実際に購入できるような製品も見られるようになってきた。近い将来、多くの人がウェアラブルコンピュータを身に付け日々の生活を送るようになることが期待される。

ウェアラブルコンピュータには各種の機材を接続し、多種多様な情報をリアルタイムで取得することができる。例えば、GPS (Global Positioning System) を利用すれば、利用者の位置情報を衛星からの電波を捕捉することによって得ることができる。また、小型のビデオカメラ、いわゆるウェアラブルカメラを利用すれば、利用者の視点の映像を取得することができる。これらの機材を利用することによって、利用者が意識することなく常に様々な情報を取得できるようになることができる。

一般に映像データは容量が膨大になることから、計算機を利用した長時間の撮影データの作成、保存は現実的ではなかった。ところが近年、ハードディスクドライブ等の記憶装置の大容量化、低廉化が急速に進んでいる。例えば、数 10GB クラスの小型ディスク装置は 1 万円程度で入手できる。仮に、映像をビデオ CD 相当の画質の MPEG-1 フォーマットで記録するとしよう。この場合、1 時間の録画のために 650MB 程度の容量を消費すると言われるが、70GB のディスクであれば 110 時間 ($= 70(GB) \times 1024(MB) / 650(MB)$) の映像データを格納することが可能である。このことから、近い将来ウェアラブルカメラ映像を全て計算機上に保管することも十分現実的であると言える。

このようにして蓄積された映像データは、利用者の生活の全ての記録であると言える。その応用の一例としては、ビデオ日記 [1]、利用者の記憶補助等が挙げられる。このようなアプリケーションを実現するためには、大量の映像データから所望の映像を素早く検索する技術が重要である。

これまでも、映像検索に関する研究は数多く行われてきた。その研究の多くは、ニュースやスポーツなどの映像に対してある種の索引付けを行い、その索引に対して検索処理を行うことによって目標を達成していた [2, 3]。しかしながら、ウェアラブルカメラ映像の場合、これらの手法は有効でない場合が多い。これはウェアラブルカメラ映像には撮影時間が長い、切れ目がない、利用者の移動に応じて画像がぶれる等の特徴があり、音声検出やショット切替検出等の手法がうまく適用できないためである。

これに加えて、ウェアラブルカメラ映像では、利用者の位置や状況等が重要な意味を持っている。例えば、過去に訪れた観光先の映像を検索したい場合、映像の内容だけではなく、いつ、どこで、どのように撮影されたか

と言った日時に関する情報や、だれと行ったか、どのような出来事があったかといった、利用者の状況に関する情報を元に映像を検索できる必要がある。これを実現するためには、上で挙げたような既存の枠組だけでは不十分である。

このため上田等は、地理オブジェクトに基づいたウェアラブルカメラ映像の検索、要約手法を提案している [4]。この手法において、利用者はウェアラブルカメラと GPS レシーバが接続されたウェアラブルコンピュータを常時携帯し、視点映像、位置情報などを常に取得し続ける。さらにカメラにはジャイロセンサが接続されており、カメラの向きの情報も取得する。これらの情報と地理情報を利用し、(1) 利用者の近くにあった地理オブジェクトを特定する。(2) それぞれの地理オブジェクトと利用者の距離、角度をもとに、その地理オブジェクトが利用者にとってどの程度重要であったかを数値化した指標 (重要度) を算出する。このように地理オブジェクトの重要度を利用することによって、利用者自身に身近な地理情報を利用した映像検索、要約の作成を可能にしている。

上田等の研究においては、取得したデータや算出した各地理オブジェクトの重要度などのメタデータが重要な意味を持っている。すなわち、予めそれらのメタデータが用意されていることにより、利用者は地理オブジェクト名などを用いて検索を行うことが可能となる。しかしながら、メタデータの格納には独自のシステムを用いているため、他のシステムやアプリケーションとデータのやり取りを行うには、新たにプログラムを作成する必要がある。

そうした問題の解決のために、MPEG-7 という規格の策定作業が行われてきた。これはマルチメディアコンテンツのメタデータの記述形式を定めたものであり、国際標準規格である。この規格にそって映像のメタデータを記述することにより、様々なアプリケーションやオープンなデータベースにおいても同じデータを利用することが可能となり、上で述べたようなデータのやり取りに関する問題は解決する。

しかし MPEG-7 ファイルの作成には多くの時間と手間がかかるため、ファイル作成の自動化が課題となっている。MPEG-7 の標準化の対象となっているのはメタデータの記述形式だけであり、どのように記述するか、どのように利用するかなどはアプリケーション次第であるためだ。既に MPEG-7 をサポートする幾つかのアプリケーションの開発が進められているが、そのなかでも課題として挙げられる点でもある。

そこで本研究では上田等の手法を利用し、(1) ウェアラブルカメラ映像に対して自動で索引付けを行い、それを記述した MPEG-7 ファイルを自動で生成する、(2) 利用者は (1) のファイルを基に検索を行うが、検索内容に合った MPEG-7 ファイルを動的に生成することで利用者

との対話的な検索システムを実現する。(1)の目的の実現のため、我々は時間、位置、アノテーションといったメタデータに着目し、MPEG-7 ファイルを自動で生成する手法を提案する。また(2)で提案する手法により、ウェアラブルカメラ映像のような膨大な映像データからでも所望の場面を効率良く検索することが可能となる。

本研究の構成は以下の通りである。まず2章で MPEG-7 の概要を述べる。3章では MPEG-7 ファイルを自動で生成する手法、4章で自動生成された MPEG-7 ファイルを用いた検索手法について述べる。5章で関連研究について述べ、6章でまとめと今後の課題を述べる。

2 MPEG-7 の概要

MPEG-7 はマルチメディアの内容を記述するためのフォーマットであり、ISO/IEC JTC1 SC29/WG11 により国際標準規格として策定作業が進められ、現在では規格発行の最終段階である。国際標準規格であることから、今後対応するアプリケーションが数多く開発されることが予想され、MPEG-7 形式のファイルのやり取りも頻繁になると考えられる。また記述方法を定義するためのベース言語として XML Schema を用いており、XML 関連の多くのアプリケーションが利用できるという利点も挙げられる。

数多くの利点がある MPEG-7 であるが、MPEG-7 を利用するにあたって、MPEG-7 ファイルを作成する時間や手間について議論する必要がある。既に開発されてきているようなオーサリングツールは、ニュースやスポーツなどの従来の映像を対象にしたものがほとんどであり、カット点の自動検出やあるイベントに着目した映像の分割を行っている。またそれ以上の内容記述は人手で行う必要があり、それにかかる時間は寺口等 [5] の研究で報告されているもので、映像の約 1.6 倍である。

対応するシステムの開発、また MPEG-7 ファイルをいかに効率よく作成するか、という点が MPEG-7 普及の鍵となる。

3 MPEG-7 ファイルの自動生成

本研究での一つの目的として、ウェアラブルカメラ映像に対して自動で索引付けを行うことが挙げられるが、以下そのために必要な情報、処理手順について述べる。

図 1 に索引付けの流れを示す。

3.1 要求分析

ウェアラブルカメラにより常時撮影した映像に対する検索を考えた場合、代表的なものとして次の問合せが挙げられる。

1. 「東大寺の映像が見たい」
2. 「 月 日の昼頃の映像は？」
3. 「桜が写っている映像」

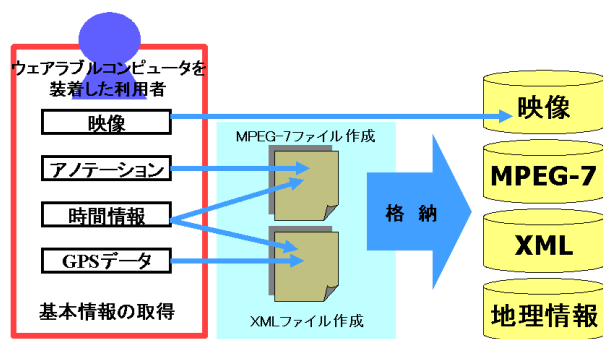


図 1: 映像への索引付けの流れ

4. 「友人と共に何かをしている場面」
5. 「感動した場面」

これらの検索を実現するために必要な情報について考えると、1 は映像中の地理オブジェクト名、2 は日付および時間、3 はフリーキーワードが索引として付与されている必要がある。また 4 の場合には、映像中の人物に対し友人の誰であるか、また何をしている場面であるかについての情報が索引として必要であり、5 の場合も同様に手作業で索引付けを行う必要がある。

ウェアラブルカメラ映像に対し、これらの情報を人手で索引付けすることは非現実的な作業であり、本研究の第一段階として、ウェアラブルカメラ映像に対して自動で索引付けを行うことを試みる。しかし全ての索引を自動で付与することは現在の技術では不可能であり、自動化が可能でありかつウェアラブルコンピュータにより取得可能な情報に関して、索引付けの自動化を行う。

先の検索要求において索引付けの自動化について考えると、1 では GPS データにより地理オブジェクトを基にした映像の分割を行うことで解決を目指す。2 の場合、データ取得時の日付や時間を参照することで容易に解決できる。3 の場合、映像に対する索引付けが手作業で完全に行われていれば、フリーキーワードによる検索も簡単に実現できるが、ここでは索引付けの自動化を目的としており、音声認識ソフトによりテキスト変換されたアノテーションを用いることで解決を図る。

4 は画像処理による人物の認識技術がさらに進んだとしても、自動化はかなり困難な問題であり、手作業による索引付けが必要な例である。5 の場合も 4 と同様に手作業による索引付けが必要である。発汗作用や脳波といった利用者の生体情報を用いた研究 [6, 7] の利用により実現できる可能性もあるが、現段階ではまだそこまで研究が進んでいない。

3.2 地理オブジェクトに基づいた映像の索引付け手法

先に述べた情報を利用し、地理オブジェクト名・アノテーションといったある地点に関する位置情報を基にし

で映像の分割を行い、索引として付与、自動的に MPEG-7 ファイルを生成する。

本研究では上田等 [4] の研究で提案された地理オブジェクトの重要度を計算する手法を用いる。

3.2.1 地理オブジェクトの重要度の計算方法

まず最初にある時刻における地理オブジェクトの重要度の算出方法について簡単に述べる。映像に対して索引として付与される地理オブジェクトは、取得した GPS データと地理情報データベースに格納されている各地理オブジェクトの緯度・経度を比較し、そのユークリッド距離がある一定の値以下であるものが選ばれる (図 2)。具体的には、ある時刻 t における利用者の座標を $P_t(x_t, y_t)$ 、ある地理オブジェクトの座標を $O_k(X_k, Y_k)$ としたとき、2 点間のユークリッド距離 $d(P_t, O_k)$ は次の式により算出される。

$$d(P_t, O_k) = \sqrt{(X_k - x_t)^2 + (Y_k - y_t)^2}$$

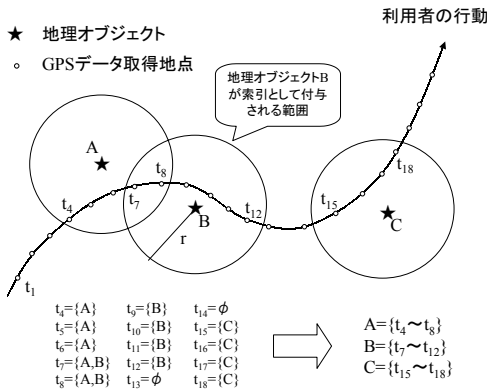


図 2: 索引として選択される地理オブジェクト

この式により求められた $d(P_t, O_k)$ がある一定の値 r より小さい場合、 $d(P_t, O_k)$ から計算される重みを時刻 t における地理オブジェクト O_k の距離に基づく重み $w_{distance}(P_t, O_k)$ とする。重みの計算方法には距離の逆数を用いる。

カメラの向きに基づく重みの計算方法を記す。ある時刻 t におけるカメラの向きを α_t 、利用者からみた地理オブジェクト O_k の角度を $\theta(P_t, O_k)$ とすると $\theta(P_t, O_k)$ は次の式により求まる。

$$\theta(P_t, O_k) = \tan^{-1} \frac{Y_k - y_t}{X_k - x_t}$$

重みを付ける範囲を $\theta(P_t, O_k)$ を中心に 2β とし、 α_t が $\theta(P_t, O_k)$ に近い程大きな重みが付くようにする。ここでは、 $\alpha_t - \theta(P_t, O_k)$ の値が大きくてもある程度の重みが付与されるよう正規分布を用いた重み付けを行う (図 3)。

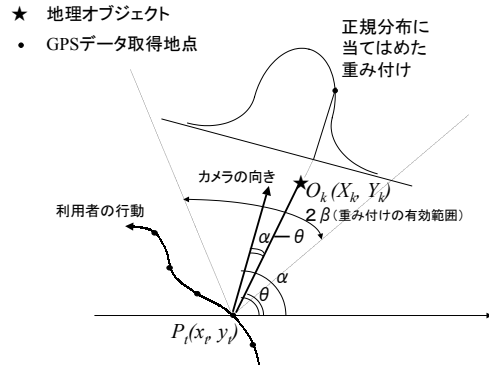


図 3: カメラの向きと地理オブジェクトの関係

地理オブジェクト O_k のカメラの向きに関する重み $w_{direction}(P_t, O_k)$ は、 $-\beta < \alpha_t - \theta(P_t, O_k) < \beta$ を満たすとき、正規分布 $N(0, \sigma^2)$ に従って次の式で求められる。

$$w_{direction}(P_t, O_k) = \frac{1}{1-C} \times \exp\left(-\frac{(\alpha_t - \theta(P_t, O_k))^2}{2\sigma^2}\right) - \frac{C}{1-C} \quad (\text{ただし } C = \exp(-\frac{\beta^2}{2\sigma^2}))$$

$|\alpha_t - \theta(P_t, O_k)| \geq \beta$ の場合には、 $w_{direction}(P_t, O_k) = 0$ とする。

以上の、地理オブジェクトと利用者の距離に基づく重み $w_{distance}$ と、カメラの向きと地理オブジェクトの方向に基づく重み $w_{direction}$ の 2 つを用いて、ある地理オブジェクト O_k の、ある時刻 t における重要度 $w(P_t, O_k)$ を次の式により計算する。

$$w(P_t, O_k) = zw_{distance}(P_t, O_k) + (1-z)w_{direction}(P_t, O_k)$$

どちらの重みを重視するかは、利用者がパラメータ $z(0 \leq z \leq 1)$ を指定することにより決定する。

上田等 [4] の論文における重要度とはある連続した時間の地理オブジェクトの重みを統合したものであるが、本研究での重要度とは、ある時刻における地理オブジェクトの距離に基づく重みと、カメラの向きによる重みを統合したものである。

3.2.2 地理オブジェクトを基にした映像分割

前節の手法により計算された重要度は地理オブジェクト名と、時間、緯度、経度、角度などの情報と共に、効率を考え独自に定義した DTD により XML 形式で記述する。Point タグにより 1 秒毎のデータを記述、GeographicPoint タグの属性に緯度、経度、角度を記述し、Place タグによ

り地理オブジェクト名, その属性として value を定義し重要度を記す.

このファイルは MPEG-7 ファイルを作成するための基礎データとして使用する. 図 4 に記述例を示す.

```
<?xml version="1.0" encoding="Shift_JIS"?>
<Point time="2002-02-08T14:35:04">
  <GeographicPoint longitude="135.442086" latitude="34.438085" direction="90"/>
  <Place value="100">奈良先端科学技術大学院大学</Place>
  <Place value="90">NEC</Place>
</Point>
:
<Point time="2002-02-08T14:35:24">
  <GeographicPoint longitude="135.442099" latitude="34.438214" direction="120"/>
  <Place value="90">奈良先端科学技術大学院大学</Place>
  <Place value="70">高山サイエンスプラザ</Place>
  <Place value="65">NEC</Place>
</Point>
```

図 4: 基礎データの記述例

図 4 の Point タグは 1 秒毎 (GPS のデータ取得の周期) の記述となるため, ある地理オブジェクトに注目すると, 当然連続して同じ地理オブジェクト名が出現する区間が存在する. MPEG-7 ファイルに記述する際には, そういった区間をまとめて記述する. 同じ地理オブジェクトで複数の区間が存在する場合には, 別々の区間として記述する.

図 5 に MPEG-7 ファイルへの記述例を示す. 簡単に説明を行うと, AudioVisualSegment 記述子により地理オブジェクトに関する情報を記述する. 地理オブジェクトに関する情報であることは, id 属性を見ると place* となっていることから分かる. GeographicPointType 記述子により経度・緯度, PointOfView 記述子の viewpoint 属性で地理オブジェクト名を記述, その重要度を Value の値に記す. MediaTime 記述子内では, この映像区間の始まりの時間 (MediaTimePoint) と継続時間 (MediaIncrDuration) を記す.

```
:
<AudioVisualSegment id="place1">
  <GeographicPointType longitude="135.442099" latitude="34.438214" />
  <PointOfView viewpoint="奈良先端科学技術大学院大学">
    <Importance><Value>重要度</Value></Importance>
  </PointOfView>
  <MediaTime>
    <MediaTimePoint>
      2002-02-08T14:35:24:0F30
    </MediaTimePoint>
    <MediaIncrDuration timeUnit="PT1S">
      374
    </MediaIncrDuration>
  </MediaTime>
</AudioVisualSegment>
:
```

図 5: 地理オブジェクトを基にした映像区間の MPEG-7 ファイル記述例

以上により地理オブジェクトを基にした映像の分割を実現する.

3.2.3 アノテーション地点を基にした映像の分割

利用者は映像撮影時に, 後々検索したいと考える場面で音声によるアノテーション付けを行う. すなわち「桜

が写っている映像」といった質問に応えるためには, 桜を見ている場面で「桜」というキーワードを発し, アノテーション付けを行っておく必要がある. アノテーション付けを行った地点 (アノテーション地点) は, 後の処理プログラムにより, その時刻の位置情報 (緯度・経度) が仮想の地理オブジェクトとして登録される. 地理オブジェクト名にはアノテーションそのものをあてる. この仮想の地理オブジェクトは地理情報データベースに登録され, その後の処理は先述の地理オブジェクトの場合と同様である. これによりアノテーション地点を基にした映像区間が作成される.

図 6 に MPEG-7 ファイルへの記述例を示す. 基本的な記述方法は前節の地理オブジェクトの記述方法と同じであるが, アノテーションであることを明示するため AudioVisualSegment 記述子の id 属性を annotation* とする.

```
:
<AudioVisualSegment id="annotation1">
  <GeographicPointType longitude="135.442099" latitude="34.438214" />
  <PointOfView viewpoint="桜">
    <Importance><Value>重要度</Value></Importance>
  </PointOfView>
  <MediaTime>
    <MediaTimePoint>
      2002-02-08T14:35:24:0F30
    </MediaTimePoint>
    <MediaIncrDuration timeUnit="PT1S">
      180
    </MediaIncrDuration>
  </MediaTime>
</AudioVisualSegment>
:
```

図 6: アノテーション地点を基にした映像区間の MPEG-7 ファイル記述例

3.3 MPEG-7 ファイルの自動生成の処理手順

これまで述べた方法を用いて, MPEG-7 ファイルの自動生成を行う手順を以下に記す.

1. ウェアラブルコンピュータ (カメラ) により, 映像, 時刻, GPS データ, 向き (角度), アノテーションを取得.
2. アノテーションの時刻と GPS データの時刻が一致する地点 (アノテーション地点) を, 仮想の地理オブジェクトとして地理情報データベースに登録.
3. 取得したデータと地理情報データベースの情報の照合により, ある時刻における地理オブジェクトおよびアノテーション地点の重要度を算出.
4. 時刻, GPS データ (緯度・経度), 向き (角度), 地理オブジェクト名 + 重要度, のデータを独自に定義した DTD により記述 (基礎データ).
5. 地理オブジェクトを基にした映像区間を特定.
6. 上記の情報を基に, MPEG-7 ファイルを作成.

MPEG-7 ファイルには, 地理オブジェクトとアノテーション地点の 2 つの位置情報を基にした映像区間が記述

される。これは、AudioVisualSegment 記述子の id 属性を参照することでどちらを記述したものであるかが分かる。この MPEG-7 ファイルの作成にあたり人の手による索引付けは行っておらず、処理自体はウェアラブルコンピュータを取り外し、夜寝ている間などに自動で行うことができる。

この MPEG-7 ファイルを基にすることで、大量のウェアラブルカメラ映像から、日付、地理オブジェクト名、キーワードを用いた検索が可能となる。検索手続きについては、次節以降で述べる。

4 MPEG-7 ファイルを用いた検索

自動生成された MPEG-7 ファイルを用いて、日付・地理オブジェクト名・キーワードによる映像検索を行なうシステムの概要について述べる。本研究では 1 日の映像に対し、1 つの MPEG-7 ファイルが生成される。そのため日付による検索の場合には MPEG-7 ファイルが特定され、その後の処理としては希望する映像区間の提示もしくはダイジェスト作成を行う。

しかし地理オブジェクト名やキーワードによる検索の場合には、該当する映像区間が複数の日にまたがっている可能性があり、MPEG-7 ファイルが特定されとは限らない。従ってこの場合には、検索要求に合う映像区間のみを日付が異なる MPEG-7 ファイルより抽出し、新たに MPEG-7 ファイルを生成することを考える。さらにこのファイルを基に、検索、映像区間の提示、ダイジェスト作成を行うことができるようにし、利用者との対話的な検索システムにする。

ここでは図 7 の MPEG-7 ファイルを例に、各処理の具体例を考える。このファイルは地理オブジェクトとアノテーション地点についての索引を 2 つずつ記述したものである。

4.1 日付による検索

日付による検索は、MPEG-7 ファイル中の AudioVisual 記述子の id 属性を参照することで容易にファイルが特定される。ファイルが特定された後は、映像区間の提示・ダイジェスト作成の処理に移る。

4.2 地理オブジェクト名による検索

地理オブジェクト名による検索処理の流れを述べる。

1. 検索する地理オブジェクト名を入力
2. AudioVisualSegment 記述子の id 属性が Place*の中から、PointOfView 記述子の viewpoint 属性に検索語が含まれる AudioVisualSegment を抽出。
3. 2 で抽出された AudioVisualSegment と、映像に関する情報が記述されている部分 (AudioVisual 記述子以下)、AudioVisualSegment 記述子の id 属性が annotation*のもの全てを用いて、新たな MPEG-7 ファイル

```
<?xml version="1.0" encoding="Shift_JIS"?>
<Mpeg7 type="complete" xmlns="http://www.mpeg7.org/2001/MPEG-7_Schema" xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance">
  <ContentDescription xsi:type="ContentEntityType">
    <MultimediaContent xsi:type="AudioVisualType" id="multimediaContent-1">
      <AudioVisual id="20020217">
        <MediaTime> ... </MediaTime>
        <MediaLocator xsi:type="TemporalSegmentLocatorType"> ... </MediaLocator>
        <TemporalDecomposition>
          :
        <AudioVisualSegment id="place1">
          <GeographicPointType longitude="135.442099" latitude="34.438214" />
          <PointOfView viewpoint="奈良先端科学技術大学院大学">
            <Importance><Value>重要度</Value></Importance>
          </PointOfView>
          <MediaTime>
            <MediaTimePoint>
              2002-02-08T14:35:24:0F30
            </MediaTimePoint>
            <MediaIncrDuration timeUnit="PT1S">
              374
            </MediaIncrDuration>
          </MediaTime>
        </AudioVisualSegment>
        :
      <AudioVisualSegment id="place2">
        <GeographicPointType longitude="135.442125" latitude="34.438639" />
        <PointOfView viewpoint="高山サイエンスプラザ">
          <Importance><Value>重要度</Value></Importance>
        </PointOfView>
        <MediaTime>
          <MediaTimePoint>
            2002-02-08T14:38:00:0F30
          </MediaTimePoint>
          <MediaIncrDuration timeUnit="PT1S">
            420
          </MediaIncrDuration>
        </MediaTime>
      </AudioVisualSegment>
      :
    <AudioVisualSegment id="annotation1">
      <GeographicPointType longitude="135.442099" latitude="34.438000" />
      <PointOfView viewpoint="校">
        <Importance><Value>重要度</Value></Importance>
      </PointOfView>
      <MediaTime>
        <MediaTimePoint>
          2002-02-08T14:34:00:0F30
        </MediaTimePoint>
        <MediaIncrDuration timeUnit="PT1S">
          210
        </MediaIncrDuration>
      </MediaTime>
    </AudioVisualSegment>
    :
    <AudioVisualSegment id="annotation2">
      <GeographicPointType longitude="135.442125" latitude="34.438639" />
      <PointOfView viewpoint="工事中">
        <Importance><Value>重要度</Value></Importance>
      </PointOfView>
      <MediaTime>
        <MediaTimePoint>
          2002-02-08T14:40:00:0F30
        </MediaTimePoint>
        <MediaIncrDuration timeUnit="PT1S">
          180
        </MediaIncrDuration>
      </MediaTime>
    </AudioVisualSegment>
    :
  </MultimediaContent>
</ContentDescription>
</Mpeg7>
```

図 7: 検索に使用する MPEG-7 ファイル例

を生成する。

4. 生成された MPEG-7 ファイル中の日付、日付毎の地理オブジェクト名を利用者に提示。
5. 映像区間の提示・ダイジェスト作成または日付、キーワードによる検索を選ぶ。

図 7 のファイルにおいて、高山サイエンスプラザで検索した場合、id 属性が place2 の AudioVisualSegment が新たな MPEG-7 ファイルに記述されることになる。また、複数の地理オブジェクトを検索したい場合には、1 において複数入力する。

4.3 キーワードによる検索

キーワードによる検索の場合にも、基本的な処理の流れとしては地理オブジェクト名による検索と同じである。以下に処理の流れについて述べる。

1. 検索するキーワードを入力
2. AudioVisualSegment 記述子の id 属性が annotation* の中から, PointOfView 記述子の viewpoint 属性に検索語が含まれる AudioVisualSegment を抽出.
3. 2 で抽出された AudioVisualSegment と, 映像に関する情報が記述されている部分 (AudioVisual タグ以下), AudioVisualSegment 記述子の id 属性が Place* のもの全てを用いて, 新たな MPEG-7 ファイルを生成する.
4. 生成された MPEG-7 ファイル中の日付, 日付毎の地理オブジェクト名を利用者に提示.
5. 映像区間の提示・ダイジェスト作成または日付, 地理オブジェクト名による検索を選ぶ.

図 7 のファイルにおいて, 桜で検索した場合, id 属性が annotation1 の AudioVisualSegment が新たな MPEG-7 ファイルに記述されることになる. また複数のキーワードによる検索を行いたい場合には, 1 において複数のキーワードを入力する.

4.4 動的に生成された MPEG-7 ファイルの利用

利用者との対話的な検索システムにするため, 前節までの地理オブジェクト名やキーワードによる検索により生成された MPEG-7 ファイルを利用し, さらに次の検索を行う.

- 日付による検索
この場合は, 該当する日付の部分の AudioVisualSegment のみを抽出し, MPEG-7 ファイルを作成.
- 映像区間の提示・ダイジェスト作成
これを選択した場合は, 次節の処理に移る.
- このファイルを基に検索
最初に地理オブジェクト名による検索を行った場合にはキーワードによる検索, キーワードによる検索を行った場合には地理オブジェクト名による検索が可能とする. この場合返される映像区間は最初と 2 回目の映像区間が重なる部分とする. 例えば図 7 のファイルで, 地理オブジェクト名を奈良先端科学技術大学院大学として検索した後, キーワードとして桜を入力した場合, 図 8 のような結果となる.

4.5 映像区間の提示・ダイジェストの作成

日付による検索や地理オブジェクト名・キーワードによる検索のいずれの場合においても, 検索を行った後には, MPEG-7 ファイルが特定または生成され, 利用者にはそのファイルの中から日付・地理オブジェクト名が提示される. 利用者がある映像区間を見たい場合には, 地理オブジェクト名を一つ選択, ダイジェストを見たい場

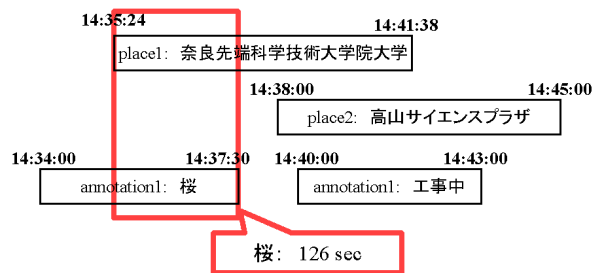


図 8: 検索結果例

合には地理オブジェクト名から見たい場面を複数選択し, 再生時間を指定する.

4.5.1 映像区間の提示

本研究では地理オブジェクトを基にして映像の分割を行っており, ある地理オブジェクト名を選択すると該当する映像区間は一意に決まる. そのため利用者がある地理オブジェクト名を一つ選択した場合には, システムは該当する映像区間の再生を行う.

4.5.2 ダイジェストの作成

利用者は提示された地理オブジェクト名から, 希望する場面を複数選択する. そしてダイジェストの再生時間を指定することで, システムは該当する場面のダイジェストを自動で作成する.

以下に, 処理の流れを述べる.

1. 地理オブジェクト名を選択.
2. ダイジェストの再生時間を指定.
3. 該当する映像区間の重要度 (PointOfView 記述子内の Value の値) を比較し, 再生時間を各映像区間に振り分ける.
4. 各映像区間に割り当てられた再生時間で, 重要度が最も大きくなる区間を基礎データ (図 4) から計算. その区間を再生する映像区間とする (図 9).
5. 4 で特定された各映像区間中の再生する部分を全てつなげてダイジェストとして提示.

ここで一番重要な点は, 4 の処理である. この方法であれば, ダイジェストの再生時間に合わせて各映像区間の重要な部分を動的に特定できる. ただし重要度が最も大きくなる区間を特定する処理に時間がかかるようであれば, 次のような方法も検討する.

- MPEG-7 ファイルに各映像区間中の最も重要度が大きい時刻を予め記述しておく.
- 再生時間が割り当てられると, 上記の時刻を中心としてその前後の映像を返す.

提案する手法では, ある 1 日の映像のダイジェストだけではなく, 複数の日にまたがったダイジェストも作成

「桜」の映像区間に割り当てられた時間を4秒と仮定

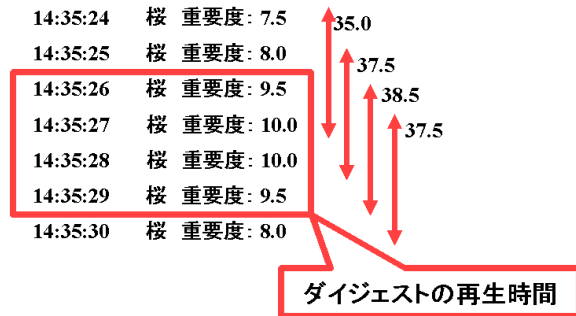


図 9: 再生する映像区間の特定

できる。すなわち、去年1年間に見た桜の映像といったダイジェストの作成も可能となる。

5 関連研究

ウェアラブルカメラに関する研究としては、Healey 等 [6] の発汗情報をトリガーにした映像取得の提案がある。また石島等 [7] はウェアラブルカメラで取得した映像を対象とし、脳波を用いて自動編集する試みを行なっている。これらは位置情報を用いたものではないが、映像の構造化に関して有効となり得るものであり、本研究と組み合わせることで利用者に合った映像への索引付けが可能となろう。

位置情報と地理情報を用いてウェアラブルカメラ映像のダイジェスト作成を行う研究としては上田等 [4] のものが挙げられる。本研究においてもこのなかで提案されている手法を利用している。

MPEG-7 に関するものとしては、寺口等 [5] の移動体通信に利用者の嗜好に合わせたダイジェスト映像を配信するものや、Harmony [8] や DICEMAN [9] などのプロジェクトがある。これらの研究での MPEG-7 の使用法は本研究とは異なるものであり、ダイジェストの作成手法などは本研究においても検討すべきものである。しかし、位置情報に基づいて MPEG-7 ファイルを自動で生成する手法はまだない。

6 まとめと今後の課題

本研究では、ウェアラブルカメラでの映像撮影と同時に時間、位置、アノテーションなどの情報を取得し、これらを用いて MPEG-7 ファイルを自動生成する手法を提案した。さらにウェアラブルカメラ映像から所望の場面を素早く検索するため、利用者の検索要求に合致した MPEG-7 ファイルを動的に生成することにより、利用者との対話的な検索システムについても提案を行った。また同時に、対話的な検索システムの一部として、ダイジェストを作成する手法についても議論を行った。

今後の課題としては、次のような事項が挙げられる。

● システムの実装

システムの実装を行い、提案する手法が利用者が希望する場面をどの程度の精度で提示することができるのか確認を行う。

● 領域をもった地理オブジェクトへの対応

現状では地理オブジェクトが点で表されているデータを用いており、奈良公園など大きな領域をもった地理オブジェクトの場合には対応できない。そのため領域をもった地理オブジェクトの重要度に関して新たな定義を行うか、現在の地理情報データベースの代わりに GIS(地理情報システム)の利用などを考慮する必要がある。

● アノテーションの取扱い

提案した手法では、アノテーション付けを行った地点について地理オブジェクトと同様に重要度を計算し、映像の分割を行っている。これは対象となる地理オブジェクトがすぐ近くの場合には有効だが、少し距離が離れた地理オブジェクトの場合には必ずしも有効でない。この点についてより検討する必要がある。

謝辞 本研究の一部は、文部科学省科学技術研究費基盤研究(課題番号: 11480088, 12680417, 12780309)、ならびに科学技術振興事業団(JST)の戦略的基礎研究推進事業(CREST)「高度メディア社会の生活情報技術」プログラムの支援によるものである。また同じ研究室の上田氏には重要度の計算手法の解説などで多大な助言を頂いた。ここに記して謝意を表す。

参考文献

- [1] M. Eldridge, M. Lamming, and M. Flynn. "does a video diary help recall". Proceedings of the HCI'92 Conference on People and Computers VII, Cambridge University Press, Cambridge.
- [2] 南憲一, 阿久津明人, 浜田洋, 外村佳伸. "音情報を用いた映像インデクシングとその応用". 信学論 (D-II), Vol. 81, No. 3, pp. 529-537, 1998.
- [3] 谷口行信, 南憲一, 佐藤隆, 桑野秀豪, 児島治彦, 外村佳伸. "scenecabinet: 映像解析技術を統合した映像インデクシングシステム". 信学論 (D-II), Vol. 84, No. 6, pp. 1112-1121, 2001.
- [4] 上田隆正, 天笠俊之, 植村俊亮, 吉川正俊. "位置情報と地理情報を用いたウェアラブルカメラ映像のダイジェスト作成". 信学技報, DE, Vol. 2001, No. 78-104, pp. 175-182, 2001.
- [5] 寺口正義, 益満健, 越後富夫, 関口俊一. "時区間インデックス生成によるパーソナルビデオダイジェスト". 信学技報, PRMU, Vol. 2001, No. 85-93, pp. 43-50, 2001.
- [6] Jennifer Healey and Rosalind Picard. "startlecam: A cybernetic wearable camera". Proc.ISWC, 1998.
- [7] 石島健一郎, 相澤清晴. "ウェアラブルによる長時間個人体験記録の編集 脳波を利用した映像の自動編集の試み". 信学技報, PRMU, Vol. 2000, No. 147-159, pp. 85-92, 2001.
- [8] Harmony. <http://www.ilrt.bris.ac.uk/discovery/harmony/>.
- [9] Diceman. <http://www.teltec.dcu.ie/diceman/>.