

位置・方向情報付き映像の検索方式の提案

西岡 欣亮[†] 住野 優^{††} 大西 真晶[†]

上田 真由美^{†††} 上島 紳一[†]

[†]関西大学大学院総合情報学研究科 〒569-1095 大阪府高槻市雲仙寺町 2-1-1

^{†††}名古屋大学情報連携基盤センター 〒464-8601 愛知県名古屋市千種区不老町 1

E-mail: [†]{ fb4m133, fa4d001, ueshima }@edu.kutc.kansai-u.ac.jp ,

^{††}yutaka.s@mb.newweb.ne.jp , ^{†††}ueda@itc.nagoya-u.ac.jp

あらまし 近年, デジタル映像機器やウェアラブルカメラの普及により日常生活, イベント, スポーツ観戦などにおいて, 一般利用者が映像を撮影する機会が増えている. 手元の映像ファイルが増加し大容量化するに従って, 映像を対象とした簡単な索引付け機構ならびに効率的な部分映像検索を備えた方式の実現要求が高まっている. 本稿では, 利用者の視点ならびに注視点の幾何学的な位置関係を利用した映像検索手法について議論する. ここでは, 利用者は, 映像収録の間, 自由に移動し方向を変更しながら様々な対象物を撮影し続けることを想定している. また, 検索に必要な位置・方向情報は, 映像収録と同時に GPS・電子方位計を用いて取得する. これらの情報を用いて撮影者の軌跡の部分を絞り込み, 適切な部分映像を抽出することができる. 更に, 本方式の有効性を評価するため 現在開発中のプロトタイプを用いて実映像に対して本方式を適用し 検索適合率と再現率の観点から検討する.

キーワード マルチメディア処理, メタデータ管理, 情報検索, 映像検索

Proposing a Video Retrieval Scheme utilizing Location and Direction Information

Yoshiaki NISHIOKA[†], Yutaka SUMINO^{††}, Masaaki OHNISHI[†]

Mayumi UEDA^{†††}, Shinichi UESHIMA[†]

[†]Graduate School of Informatics, Kansai University, 2-1-1 Reizenji, Takatsuki, Osaka, 569-1095 Japan

^{†††}Information Technology Center, Nagoya University, 1 Hurou, Chikusa, Aichi, 464-8601 Japan

E-mail: [†]{fb4m133, fa4d001, ueshima }@edu.kutc.kansai-u.ac.jp ,

^{††}yutaka.s@mb.newweb.ne.jp , ^{†††}ueda@itc.nagoya-u.ac.jp

Abstract Due to the recent popularization of digital videos, we have more chances to record audio-visually our daily life, social events, sports, using video cameras, wearable cameras. As the volume of video files becomes enormous, a simple and efficient video indexing/retrieval scheme is required to handle and manage them. This paper proposes and discusses a retrieval scheme utilizing geometry of user's gazing points and viewpoints. We assume a user move freely and turn around when he/she records various targets during the recording time, and try to extract appropriate video portions by a combination of location information and headed direction of a video camera and to put geometrical information to video data necessary for retrievals. To obtain these geometrical data, we use a handy GPS and a 3D compass together with a camera. Using these data, the proposed scheme cuts out user's trace-segments, and squeezes them to retrieve appropriate portions. Further, we evaluate our scheme using actual video data in terms of precision and recall factors with our prototype systems.

Keyword Multimedia processing , Metadata management , Information retrieval , Video data retrieval

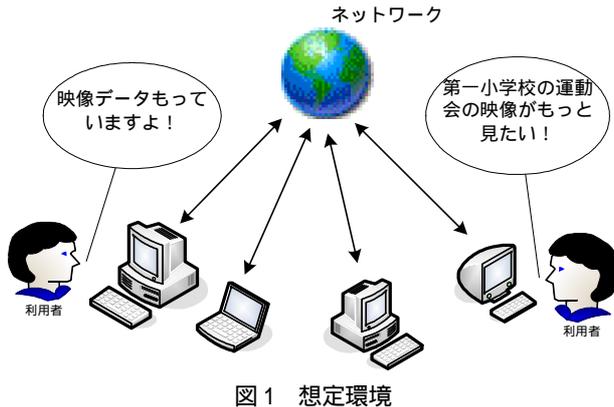


表1 位置・方向情報

	使用パラメータ	計測機器
カメラの位置	緯度, 経度, 高度	GPS
カメラの方向	方位角, 仰角, 回転角	電子方位計



1.はじめに

デジタルカメラやデジタルビデオの普及により、一般利用者が日常生活やスポーツイベントなどの種々のイベントにおいて気軽に映像収録する機会が増加している。これらの映像は日常の生活を記録したり、更に、得られた複数の映像をテーマに沿って集約したり様々な用途に用いられている。このように利用者が大量の映像を保有するにつれ、このような長時間で大量の映像への索引付け手法や、また、インターネットを介した利用者間で互いの映像を用いた集約手法の技術開発がますます重要度を増している。従来、映像を対象とした索引付け手法は、映像の内容の特徴に着目して索引を構成し必要な部分映像を切り出す手法[1]や、カットの変わり目を抽出する手法[2]、多視点映像から最適なカットを切り出す手法、テキスト情報と映像情報を同時に用いて映像を構造化する技術など多くの研究が進められている。更に、それらの手法により得られたデータを映像のメタ情報として記述する MPEG-7 や、更に、ネットワーク上での集約機構、利用者環境、社会での高度な利用までも考慮した MPEG-21 形式も層毎に標準化されつつある現状である。しかし、個人が映像作成者・発信者となるユビキタス社会では、従来の映像索引付け手法と併用して、撮影時に同時に検索に利用できるメタ情報をネットワーク利用できる形式で記録し、検索に用いることができることが望ましい。現状のビデオカメラを用いた映像収録では、収録時間、フレーム番号、映像データ形式などが用いられているが、今後、種々の物理データなどを感知するセンサー機器の高性能化・小型化に伴ってセンサー付きのビデオが一般利用者の手元に来る状況が考えられ[3]、映像索引として適切なデータの選択やその記述方法の研究が重要であると考えられる。これらの視点からの研究として、ウェアラブルカメラなどを利用して各人の見た日常生活のすべてを撮影した映像に対して索引情報を付加する方法[4]、[5]や、多数の人がスポーツイベントなどで撮影した多視点動画から代表映像を検索する方法[6]、[7]などが提案されている。

しかし、著者らの知る限り、各利用者が大量に蓄積した映像を検索し、相互に利用する方法は、まだ十分に確立されたとは言えない。本稿では映像データを利用するシナリオとして、スポーツイベントや日常生活で撮影された映像を利用者間で共有、

交換することを想定する。例えば、日常生活、スポーツイベントなどで撮影された映像を検索する場合、次のような問い合わせ、つまり、

1. “東京タワーが映っている映像が見たい”
2. “10月11日に撮影した映像が見たい”
3. “×小学校の運動会で撮影された映像が見たい”
4. “君が映っている場面が見たい”

といった日常的に収録した映像群に対して簡単に利用者が検索できる技術を想定している。このような問い合わせに対応するため本稿では、ビデオカメラの進展を想定して、ビデオカメラの位置・方向情報の利用に着目する。位置・方向情報は、映像収録と同時に GPS・電子方位計を用いて取得する。本方式では、被写体と撮影カメラの位置の幾何学的な位置関係によって撮影者の軌跡の部分を絞り込む。それにより見たい被写体の位置を索引として映像検索、または目標映像を抽出することを目指している。本研究での目標映像とは、見たい被写体が映っている一連の部分映像とする。すなわち、見たい被写体が撮影カメラのレンズにフレームインしてからフレームアウトするまでの一連の部分映像である。本稿では以下のような構成で述べる。2章で本研究の動機を説明する。3章で提案する検索方式について述べ、4章では本方式を実装したプロトタイプを用いて検索精度の評価を行う。5章では、関連研究について述べ、6章でまとめと今後の課題について述べる。

2. 研究動機

一般利用者がスポーツイベントや日常生活で撮影する場合、映像のメタ情報としては、映像全体のタイトル、おおまかな撮影内容と撮影時間、映像データの形式などに限られる場合が多く、収録した映像に対して撮影後に映像データ処理を行って内容検索のためのメタ情報を付与することは現実的とは言えない。既存の映像の内容検索技術は、必ずしもこれらの情報を対象に索引を構成しているのではなく、固定カメラや大規模なビデオ撮影環境を用いるものが多い。しかも長時間の映像に対して人手によるインデキシングを行うことは、多くの時間を費やすため現実的ではない。このような状況での撮影では、様々な撮影

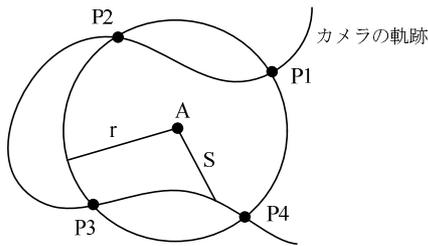


図3 2次元で表した検索範囲とカメラの軌跡

環境に応じて簡単なセンサーをビデオカメラに取り付けることで、簡単にデータが入手できる情報を扱うことが望ましい。これらの撮影の形態としては、利用者が一定の場所でビデオカメラの方向を回転させて撮影する場合や、利用者自身がビデオカメラを構えて移動しながら撮影する場合などが特徴的であるため、ビデオカメラの位置と方向が重要な情報である[4]。また、映像のオリジナルデータからでは、直接的に内容を識別することは困難であり、一般利用者が何を検索条件として問い合わせるかは明確でない。映像データを利用するシナリオとしては、スポーツイベントや日常生活で撮影された映像を利用者間で共有、交換することを想定する。本研究ではこのような問い合わせに対応するために、映像の索引として以下の、位置・方向データ、内容記述データ、映像データ型式を用いる。[カメラの位置・方向データ]

問い合わせ例1, 4のような映像の中に含まれるある特定の被写体を探し出せるように位置・方向情報を用いる。映像の収録と同時にGPSで取得するカメラ位置と電子方位計で取得するカメラの向きを位置・方向情報として付加する。カメラの位置・方向情報は、表1の6つのパラメータから成る。

[内容記述データ]

問い合わせ例2, 3のように、大量の映像の中から個々の映像ファイルを探し出せるようにコンテンツ情報と動画メタ情報を用いる。内容記述データは、映像の全ての部分に記述すると多くの時間を費やすため、映像のタイトルやあらすじ、日時など、1つの映像ファイル全体に共通する情報と映像撮影者が重要と感じた部分に対してのみコンテンツ情報として記述する。

[映像データ型式]

映像データ型式としては、コーデックや解像度、アスペクト比、フレームレート、オーディオ情報などを用いる。個々の映像ファイルを検索するとき映像データ型式によって利用者は、自分の環境に合った映像ファイルに検索結果を絞ることができる。カメラの位置・方向データは、撮影者の動きに追随するため1秒間隔でGPS、電子方位計のデータを取得する。取得したデータを映像に付加するためXMLで記述する。後節で示す注視点からの検索では、プロトタイプによりこれらのXMLを読み込んで映像検索を行う。

3. 映像検索手法

本節では、提案する幾何計算による検索手法について述べる。また、抽出した映像の中から見たい被写体が画面の中心部に映

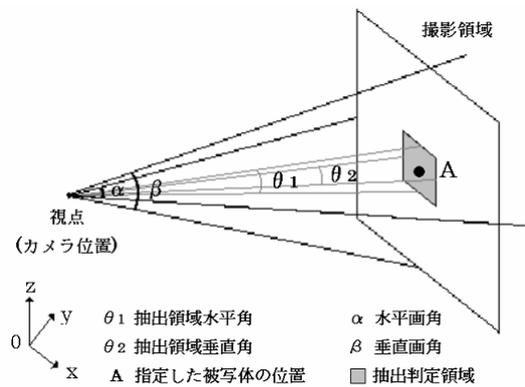


図4 抽出領域の指定

っている部分映像だけを抽出する方法についても述べる。後述するプロトタイプでは、見たい被写体の位置を入力することにより、映像の中からその被写体が映っている部分映像を抽出することを目指す。

3.1 視点からの検索

視点からの映像検索は、撮影カメラの撮影位置を検索条件として行う。GPSなどを用いて撮影カメラの軌跡を保存し、検索者はカメラの撮影位置(緯度、経度)を指定する。検索結果として表示される映像は、指定された場所から撮影されているが、撮影カメラがどちらの方向を撮影しているかわからないため何が映っているかはわからない。ウェアラブルカメラを用いて撮影された映像においては、撮影者がどこを見ているかなどを知ることができる。しかし、スポーツイベントなどで撮影された映像に対してはカメラの位置による検索では、見たい映像を探すには不十分である。

3.2 注視点からの検索

視点からの検索が撮影しているカメラの位置を検索条件とするのに対し、注視点からの検索はカメラに映っている被写体の位置を検索条件とする。一般的に映像検索をする場合、前述した問い合わせ例のように、“どこから撮影された映像なのか”ということよりも、“どこを撮影しているのか”ということが重要になる。たとえば、東京タワーが映っている場面を見たいという場合、視点からの検索では、東京タワー近くの場所を指定することにより映っている部分映像を見つけることができるかもしれない。しかし、必ずしも指定した地点からの映像の中に東京タワーが映っているという保障はなく、東京タワーが映っている部分映像を見つけるまで何度も地点を変えて検索しなければならない。また、映像の中から東京タワーが映っている部分映像の全てを探し出すためにはさらに時間がかかる。カメラに映っている被写体をキーとする注視点からの検索では、このような煩わしさが無い。注視点からの検索方式は、撮影カメラの位置、被写体の位置、撮影カメラの方向を用いて行う。検索手順を以下に示す。

1. 検索する被写体(図3中のA)の位置と被写体からカメラまでの距離(図3中のr)を検索者が指定

2. 指定された領域内から撮影した映像を絞り込む
3. 2 で絞り込まれた映像の中から、指定した被写体(図 3 中の A)が撮影カメラのレンズ内に入っている部分映像を抽出する

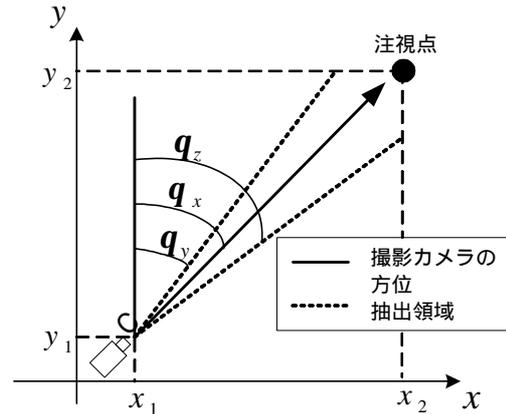
手順 1 で検索したい被写体の位置 A と、検索領域の半径 r を指定する(図 3 参照)．本研究では、焦点距離を考慮しておらず検索項目に距離を加えることで検索時に撮影カメラと注視点までの距離を指定できるようにしている．手順 2 で、GPS より取得した撮影カメラの位置情報を用いて、検索者が指定した被写体の位置から撮影位置までの距離 S を求め、撮影位置が指定された半径 r の円内に含まれるか調べる．被写体の位置とカメラの位置は緯度、経度で表現されているため、そのまま距離 S を計算することはできない．そこで、2 点の緯経度を平面直角座標に直してから S を求める．この S と利用者が指定した r を比較することで、指定領域内に撮影カメラがあるかどうか分かる．図 3 では、 $P_1 \sim P_2$ と $P_3 \sim P_4$ の部分映像に絞り込むこと

なる．手順 3 では、手順 2 で絞り込まれた $P_1 \sim P_2$ 間と $P_3 \sim P_4$ 間

の部分映像の中から、電子方位計より取得した撮影カメラの方向情報と撮影カメラの画角を用いて、カメラのレンズ内に求めている被写体が写っている部分映像を検索結果として抽出する．後節でプロトタイプを用いて注視点からの検索により検索精度を評価する．

3.3 注視点の抽出領域指定

撮影された映像には、主に撮影される撮影対象物とその対象物の周りや風景として映る対象物がある．例えば、運動会では自分の息子や娘を主に撮影し、その周りには同じ競技に参加している友人やクラスメートと一緒に映っている．主に撮影される被写体は、撮影時にある程度画面の中心で捉えるように撮影されている．そのような対象物を検索する場合は、検索した映像にも画面の中心に映っていると考えられる．利用者間で映像の交換、共有を考えた場合、主に撮影された被写体の周りの風景にも注目が集まる．多くの映像を交換、共有できれば注視点からの検索によって、自分の持つ映像以外から見たい被写体が映っている部分映像を検索することができる．しかし、探し出した映像の中で自分の見たい被写体がどのような状態(画面に映っている位置や見えている角度)で映っているかは分からない．その被写体が主な対象物として撮影されていなければ画面の端に映るような映像が検索結果として返される．抽出された映像に対して“良い”、“悪い”という評価は人それぞれではあるが、なるべく画面の中心に映っているものが見たいというのが一般的である．本研究では次の方法で、画面の中心に被写体が映っている部分映像だけを抽出する．検索者は、指定した被写体が画面の中心を基準としてどれくらいの範囲に映っているかを図 4 のように抽出領域水平角 q_1 と抽出領域垂直角 q_2 を指定することによって行う．但し $q_1 \leq a$, $q_2 \leq b$ とする．



$$q_x = \text{Arc tan} \frac{y_2 - y_1}{x_2 - x_1} \quad q_y = \text{Arc tan} \frac{y_2 - y_1 - \frac{q_1}{2}}{x_2 - x_1}$$

$$q_z = \text{Arc tan} \frac{y_2 - y_1 + \frac{q_1}{2}}{x_2 - x_1}$$

図5 水平方向二次元でのカメラ位置と注視点の関係

$$\text{Arc tan} \frac{y_2 - y_1 - \frac{q_1}{2}}{x_2 - x_1} \leq q_A \quad (1)$$

$$q_A \leq \text{Arc tan} \frac{y_2 - y_1 + \frac{q_1}{2}}{x_2 - x_1} \quad (2)$$

$$\text{Arc tan} \frac{z_2 - z_1 - \frac{q_2}{2}}{s} \leq q_P \quad (3)$$

$$q_P \leq \text{Arc tan} \frac{z_2 - z_1 + \frac{q_2}{2}}{s} \quad (4)$$

撮影カメラの方位角を q_A , 仰角を q_P とするとき、式(1)～(4)によって求める範囲に被写体が映っている映像を抽出する．式(1)～(4)は撮影カメラの位置を平面直角座標 (x_1, y_1, z_1) , 注視点の位置を平面直角座標 (x_2, y_2, z_2) とする．

式(1), (2)は、指定された抽出領域水平角 q_1 内に被写体が映っている部分映像を探し出すものであり図 3 のようになる．撮影カメラ (x_1, y_1) と被写体の位置 (x_2, y_2) がなす角が図 3 中の q_x であり、指定された抽出領域水平角 q_1 の半角を足し引きしたものが q_y, q_z になる．この角度内に撮影カメラの方位角が入った

ときの映像を抽出する．式(3), (4)も同様の方法で抽出領域垂直角 q_2 内にカメラの仰角が入っているときの映像を抽出する．本方式では、撮影カメラと注視点とする被写体の位置と距離、撮影カメラの方向により検索を行うので画面に映る被写体の大きさ、または被写体の実際の大きさは考慮していない．後述のプロトタイプでは注視点が存在する緯度、経度を 0.1 秒まで入力することにより検索を行うので、その緯度、経度の領域が撮影カメラの画角内に入ったときの部分映像が抽出される．よって、緯度、経度 0.1 秒での北緯・東経方向の距離以上の大きさである被写体では、その被写体全てが映る映像を抽出することはできない．また、撮影者は撮影中に被写体の大きさや撮影距離にあわせてズームアップを行い、被写体を見やすいように撮影することが考えられるが、本研究の現段階では、ズームアップに

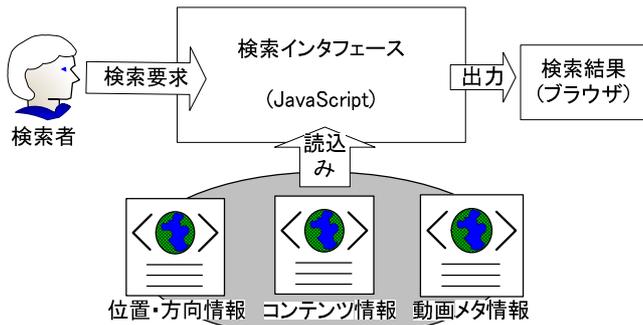


図6 システム構成

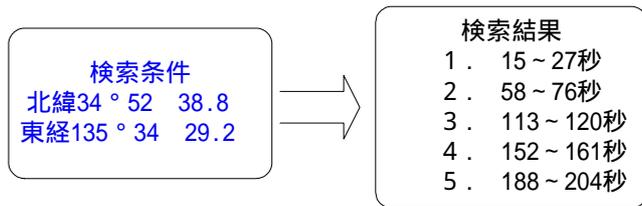


図7 検索例

対応していないことを述べておく．ズームアップへの対応については後節で述べる．

4. プロトタイプによる評価

本節では，作成したプロトタイプのシステム構成について述べ，それを用いてそれぞれの映像に対する適合率，再現率について示し，注視点からの検索手法による検索精度の評価を行う．

4.1 システム構成

撮影者または利用者により記述される動画メタ情報，コンテンツ情報は，利用者間で共有できるように MPEG-7 形式で XML に記述する．撮影時に 1 秒間隔で GPS，電子方位計から取得した位置・方向情報は，XML に独自の形式で記述する．位置・方向情報の XML はデータ量が多いため，MPEG-7 の AudioVisual タグの属性に ID 番号を持たせ，対応する ID 番号の位置・方向情報の XML を外部ファイルとして実装する．作成した検索インタフェースより検索者は，検索キーとなるキーワードや見たい被写体の緯度 経度を入力する JavaScript が XML を読み込み，入力条件にあった映像ファイルまたは指定された注視点映っている目標映像を検索する[8]．プロトタイプでは，位置情報と方向情報を読み取る際に，撮影映像と時間同期を持たせる．

4.2 評価(1)

本節では，日常生活やスポーツイベントで映像が撮影されることを想定し，プロトタイプを用いて提案方式がどれほどの検索精度になるかを検証する．実映像に対しての検索精度を見る前に，撮影に用いるカメラ，GPS，電子方位計の撮影時の環境による検索精度の変化を見る実験をした．異なる環境の下で撮影された実験映像に対して，本方式を実装したプロトタイプにより注視点からの検索を行い，抽出した目標映像により評価を行う．データ取得に使用する GPS は GAMIN 社の eTrex Venture を使用した．測位精度は約 15m 未満であるが，建物近くや

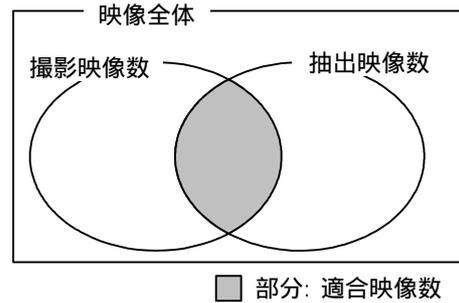


図8 評価値の関係性

表2 移動速度を変えた場合

撮影カメラの速度	適合率	再現率
普通に歩いた場合	63%	60%
早く歩いた場合	44%	44%

表3 距離を変えた場合

撮影カメラと注視点の距離	適合率	再現率
15m	75%	67%
30m	88%	78%
45m	88%	88%

障害物がある場所では，衛星からの電波が妨げられるために位置精度が大幅に落ちてしまう．本方式は，GPS から取得する撮影カメラと被写体の位置による幾何計算により検索を行うので，GPS の精度が直接検索精度に影響する．実験映像では，他のパラメータの変化による特徴を詳しく見るために，なるべく GPS の位置精度が良い場所で撮影を行った．それぞれの環境で用いる映像は 1 つにつき 5～10 分であり，用いるプロトタイプの画角は水平方向に 60°，垂直方向に 60° とした．それぞれの環境で撮影した映像に対して注視点からの検索を行い適合率，再現率を求める．適合率，再現率についての定義を以下に示す．

$$\text{適合率}(P) = \frac{\text{適合映像数}}{\text{抽出映像数}} \times 100 \quad (5)$$

$$\text{再現率}(R) = \frac{\text{適合映像数}}{\text{撮影映像数}} \times 100 \quad (6)$$

撮影映像数：映像の中で注視点映っている部分映像数
抽出映像数：プロトタイプにより抽出した目標映像数
適合映像数：抽出した目標映像内に注視点映っている数
実験映像を用いてプロトタイプへの影響を見る項目は，撮影時のカメラの移動速度による検索精度の変化，撮影カメラと注視点の距離の違いによる検索精度の変化，撮影カメラの回転速度の違いによる検索精度の変化である．以下に実験映像を用いてこれらの項目によるプロトタイプへの影響を調べた結果を示す．

4.2.1 移動速度による影響

以下に示すような環境で撮影された映像に対してプロトタイプを用いて検索を行い，撮影時におけるカメラの移動速度の変化によるプロトタイプへの影響をみる．

[撮影環境]

カメラと注視点の距離：常に変化

表4 回転速度を変えた場合

回転速度	適合率	再現率
45°	86%	88%
90°	88%	60%
180°	72%	15%

の2種類

撮影カメラの移動速度：普通で歩く場合と早く歩く場合
 撮影カメラの回転速度：回転なし．進行方向にまっすぐ

測位精度：5～6m

表2より早く歩いたほうが適合率，再現率とも約15%～20%ほど低くなった．人間の標準的な歩行速度が1.1m/sであることを考えると，GPS，電子方位計からのデータ取得は1秒間隔であるので今回のように進行方向にまっすぐ進んだ場合，距離で考えると1.1m間隔でデータ取得を行っている．早歩きしたときの速度を普通に歩いた場合の倍以上になると考えるとデータ取得間隔は2～3mになる．つまり，移動速度が速くなればなるほどデータ取得の距離間隔は大きくなるのでそれに伴い適合率，再現率は低下すると考えられる．よって，走りながら撮影した映像などに対しては適用できないと考えられる．

4.2.2 撮影距離による影響

以下に示すような環境で撮影した映像に対してプロトタイプを用いて検索を行い，撮影カメラの位置と検索条件とする注視点の位置との距離の変化によるプロトタイプへの影響をみる．

[撮影環境]

カメラと注視点の距離：15m，30m，45mの3種類

撮影カメラの移動速度：普通で歩く

撮影カメラの回転速度：回転なし．進行方向にまっすぐ

測位精度：5～6m

表3より距離が遠くなるにつれて適合率，再現率の結果が良くなるのがわかる．15m～40mの距離では，全て60%以上の結果となった．検索キーに用いる緯度，経度は，地球上の位置を表すものであり，日本では1秒=約30mで扱われることが多い．プロトタイプでは指定する緯度，経度を0.1秒まで入力できるようにしており，GPSにより測定した0.1秒における映像撮影の場所の緯度，経度の距離は北緯方向に5m83cm，東経方向に5m80cmであった．本方式では注視点の位置を考慮おらずその領域内であればどこにいても同じ位置にいると扱われる．よって，撮影カメラのレンズ内に緯度，経度0.1秒における領域全体が映らない距離では検索精度は低下する．距離が離れば撮影カメラのレンズに映る範囲は大きくなり適合率，再現率は上がると考えられる．しかし，あまり離れすぎると焦点距離を一定で撮影した場合，被写体が見えにくくなるので，30m～50mぐらいの場合に有効であると考えられる．

4.2.3 回転速度による影響

以下に示すような環境で撮影した映像に対してプロトタイプを用いて検索を行い，撮影カメラの回転速度によるプロトタイプへの影響をみる．

[撮影環境]

カメラと注視点の距離：一定(30m)

表5 検索結果

注視点	撮影映像数	抽出映像数	適合映像数
ポイント1	21	12	10
ポイント2	10	9	9
ポイント3	14	11	9
ポイント4	14	13	9
ポイント5	6	6	5
ポイント6	5	4	2

表6 適合率

抽出映像数	適合映像数	適合率
55	44	80%

表7 再現率

撮影映像数	適合映像数	再現率
70	44	63%

撮影カメラの移動速度：動かない

撮影カメラの回転速度：1秒間の回転角度が45度，90度，180度の3種類

測位精度：5～6m

表4に示す結果を考察すると，適合率はどの場合でもあまり変わらないが，回転速度が速くなるにつれ再現率が低くなることわかる．この原因は，GPS，電子方位計のデータを1秒間隔で取得していることにありと考えられる．撮影カメラを1秒間に180°回転させると方位角が1秒前の反対方向を向くことになる．データ取得のタイミングにもよるが，カメラが回転している間に注視点映っていたとしても抽出できない．よって，1秒間に90度以上の回転をしながら撮影した時には適用できないと考えられる．

4.3 評価(2)

実映像に対してプロトタイプを用い注視点からの検索を行い，抽出した目標映像より評価を行う．実映像について以下に示す．

- 撮影場所：大学キャンパス内
- 撮影内容：風景
- 撮影時間：6440秒

映像の撮影は，GPSと電子方位計から位置情報を1秒間隔で自動的に取得しながら，大学キャンパス内を自由に動き回って行った．GPSの測位精度は，場所により様々ではあるが障害物などの影響から6m以上の精度誤差がある．今回の実験では撮影範囲を50mとした．プロトタイプの画角は，評価(1)と同じ水平方向に60°，垂直方向に60°とした．プロトタイプにより抽出できた目標映像を表5に示す．注視点として選んだものは，大学キャンパス内にある消火栓や道沿いのミラーなど映像の中で目印となりやすいものがある6つ地点を選んだ．表6，7に適合率と再現率を示す．P=80%，R=63%となり，適合率は比較的高い値を示したが，再現率は低かった．今回用いた実映像では風景を撮影したのであまり撮影カメラを激しく動かすことは

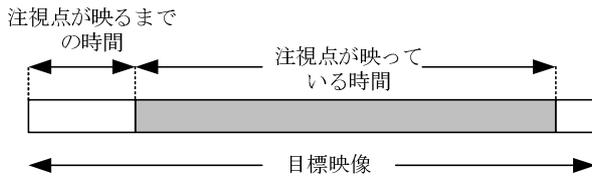


図 10 抽出した目標映像の説明

表 8 注視点が映る割合

注視点	T1	T2	T3
ポイント 1	220 秒	179 秒	81%
ポイント 2	99 秒	60 秒	61%
ポイント 3	86 秒	65 秒	76%
ポイント 4	73 秒	31 秒	42%
ポイント 5	14 秒	8 秒	57%
ポイント 6	18 秒	17 秒	94%
合計	510 秒	360 秒	71%

T1：プロトタイプにより抽出された目標映像の総時間

T2：抽出された目標映像の中で注視点が映っている総時間

T3：抽出された目標映像の中で注視点が映っている割合

なかったが、動きを激しくなることが予想されるスポーツイベントなどでは前述した実験映像の結果より検索精度は低下する。適合した目標映像の中で注視点が、映っている時間の割合を表 8 に示し、図 11 に抽出した目標映像における注視点が現れるまでの時間を示す。適合した目標映像の総時間に対する注視点が映る割合は 71%であった。注視点が現れるまでの時間は、8 秒以内に全ての場合において注視点が現れ、3 秒以内に 84% が現れる結果となった。目標映像の総時間の約 30%において注視点が映らない理由は、GPS の精度誤差と前述した緯度、経度の 0.1 秒における領域の問題が考えられる。今回、注視点として選んだものは 0.1 秒の領域内に十分に収まる大きさであったが、逆に検索条件とする被写体が 0.1 秒での領域に比べて大きいものであるときは、抽出された目標映像に注視点が映る割合は大きくなると考えられ、注視点が映るまでの時間も短くなると考えられる。再現率が上がり、適合率と同じ程度の値となれば、映像検索の索引として用いることができると考えられる。

5. 関連研究

本研究では、位置・方向情報、コンテンツ情報、動画メタ情報を映像に対して付加することで映像検索を行った。同様に位置情報を用いている[4]では、MPEG-7 で記述した位置(緯度・経度)、時間、アノテーションといった情報を、ウェアラブルカメラを用いて撮影した日常生活の映像に付加し映像検索を行っている。映像の索引として付与される地理オブジェクトに対し時刻ごとの重要度を計算し、基礎データとして記述する。この基礎データを使用して、地理オブジェクトを基にした映像分割を行うために MPEG-7 ファイルを作成する。重要度は、地理オブジ

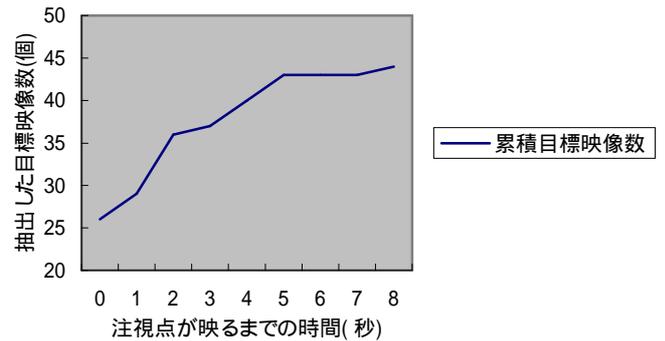


図 11 注視点が現れるまでの時間

ェクトと撮影カメラとの距離に基づく重みとカメラの向きに基づく重みを足すことにより求められている。どちらの重みを重視するかは、利用者がパラメータを指定することにより決定される。本稿が、常に注視点の位置と撮影カメラの向きと方向を考慮しながら目標映像を抽出するのに対し、[4]は、1 秒間隔ごとに映像に対して重要度を計算し、重要度を用いて映像を分割している。また、[4]では二次元でのカメラの向きを考えているのに対し、本研究では水平方向と垂直方向の 3 次元での撮影カメラの方向を考慮している。[3]では、時系列の撮影範囲と注釈情報を伴った多視点映像データから、より良い映像区間を得るために捕捉状態という速度を持ち、他のカメラの時系列な撮影範囲データと比較することで検索する手法を提案している。また、キーワードの出現密度やカメラの切り替えを考慮した映像の検索手法も提案している。[5]は、ウェアラブル機器で撮影した長期間の体験映像を視点移動のパラメータを利用してショットごとに分割している。さらに分割したショットに対し、同期記録した撮影者の脳波で評価付けを行い、撮影者が興味を持ったシーンの抽出を実現している。[9]は、GPS と電子コンパスを用いて注視点を考慮した写真検索を行っている。映像の特徴量を用いる方法として、[10]では、MPEG 符号化情報から直接的に求められる映像の特徴情報を用いて映像の類似シーンを検出し、それらに一括してタグ情報を付与することでインデキシング作業の効率化を図っている。[11]は、映像やシーンの内容を表す画像から抽出した色や形状などの特徴量の似ているものが近くに集まるように配置することで、視覚的に映像検索を行っている。[6]は、映像が映されたパレット上に直接描画することで問い合わせを行いサッカー映像の中から描画された軌跡と類似しているシーンを検索することを考案している。索引情報の付加、特徴量を用いる方法以外では、[7]は、運動会やコンサート、スポーツ中継の映像から問い合わせ範囲と撮影範囲の時空間での類似度を求め映像検索を行う。被写体がどのように写っているかを示す捕捉状態も考慮している。[12]は、GPS やジャイロセンサを用いてウェアラブルカメラで撮影した映像に、滞在時間を考慮した重み付けを行い映像の要約を行っている。

6. おわりに

6.1 まとめ

本稿では、見たい被写体の位置を索引として映像検索、または見たい被写体が映っている一連の部分映像を抽出することを目指し、撮影カメラと被写体との位置を用いた幾何計算による映像検索方式について述べた。また、本方式を実装したプロトタイプにより注視点からの検索を行い、映像の撮影環境による影響と実映像に対する検索精度を評価した。再現率の精度が上がれば、GPS、電子方位計を組み込んだウェアラブルカメラやビデオカメラにより日常生活を撮影した映像に対して有効に検索を行えると考えられる。また将来的には、GPSの測地精度が向上することや撮影カメラの画角が記録可能になることが期待される。

6.2 今後の課題

本研究での今後の課題を以下に示す。

1. ズームアップへの対応

映像を撮影する際、撮影者は被写体に合わせてズームアップを行うと考えられる。本稿で用いた映像では、ズームアップなどを行わず一定の焦点距離で撮影を行った。プロトタイプでは、焦点距離を考慮しておらず検索項目に距離を加える事で、撮影カメラと注視点までの距離を指定できるようにしている。従って、現在のプロトタイプでは焦点距離を考慮していないので、撮影カメラの位置と注視点の位置が遠く離れていて注視点が画面にほとんど映らない状態でも、指定された距離内でカメラの画角内に入っていれば検索条件に当てはまるものとして扱ってしまう。撮影カメラからカメラデータを取得できれば、それにより画角を変えることでズームアップに対応できると思われる。

2. 人間などの動きのあるものへの対応

本稿のプロトタイプでは、検索する注視点として動かない被写体を選んだが、前述した問い合わせ例のように映像の中から人物を検索することが考えられる。特にスポーツイベントなどでは重要になる。特定の人物への対応は、眼鏡などにGPS、電子方位計を組み込むことができれば可能であると思われる。

3. インタフェースの改良

本研究では被写体の位置を検索条件として、緯度・経度の入力が必要としている。しかし、緯度・経度による被写体の位置の指定は慣れ親しんだものではない。よって、緯度・経度の代わりに、インタフェースとして地図を用いることが考えられる。

4. 障害物への対応

本研究の方式では、撮影カメラと注視点の位置と距離、撮影カメラの方向を用いて検索を行う。よって、検索条件を満たしていても、間に障害物があり映像に注視点が映らなくなるという問題がある。よってこれに対応する手段が必要となる。

謝辞

本研究の一部は文部科学省私立大学学術研究高度化推進事業オープン・リサーチ・センター整備事業(平成15年度-20年度)並びに同サイバーキャンパス整備事業(平成14年度-16年度)の一環として、また、同省の「知的資産の電子的な保存・活用を支援するソフトウェア基盤技術の構築」プロジェクトの支援

により行われた。

参考文献

- [1] 有木 康雄, "DCT特徴のクラスタリングに基づくニュース映像のカット検出と記事切出し", 電子情報通信学会論文誌 Vol. J80-D2 No. 9 pp. 2421-2427 (1997)
- [2] 長坂晃朗, 田中謙, "カラービデオ映像における自動索引付け法と物体探索法", 情報処理学会論文誌 Vol. 33 No. 4 pp. 543-550 (1992)
- [3] 中西吉洋, 廣瀬龍男, 田中克己, "多視点映像データの概念モデリングと代表映像の検索", 情報処理学会論文誌: データベース Vol. 43 No. SIG 5 (TOD 14), pp. 54-65 (2002)
- [4] 兵清弘, 天笠俊之, 吉川正俊, 植村俊亮, "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", 電子情報通信学会データ工学ワークショップ(DEWS2002)C2-17
- [5] 相澤清晴, 石島健一郎, 椎名誠, "ウェアラブル映像の構造化と要約: 個人の主観を考慮した要約生成の試み", 電子情報通信学会論文誌 D-II Vol. J86-D-II No. 6 pp. 807-815 (2003)
- [6] 矢島 史, 角谷 和俊, 田中 克己, "映像上での動きの直接描画によるサッカー映像検索", 研究報告: データベースシステム, 情報学基礎 No. 041, pp. 33-40 (2002)
- [7] 秦 淑彦 廣瀬 竜男 中西 吉洋 田中 克己, "カメラメタファによる多視点映像の検索", 情報処理学会論文誌: データベース Vol. 42 No. SIG4 (TOD 9), pp. 14-26 (2001)
- [8] 住野優, 明石康久, 上島紳一, "3D仮想空間を用いた動画蓄積システムの試作", 電子情報処理学会データ工学ワークショップ(DEWS2004)2-c-03
- [9] 藤田秀之, 有川正俊, 岡村耕二, "注視点を考慮したデジタル写真の検索インタフェース", 電子情報通信学会データ工学ワークショップ (DEWS2003) 6-P-03 (2003)
- [10] 片岡良治, 遠藤育, "MPEG符号化に基づく類似シーン検出方式", 情報処理学会論文誌: データベース Vol. 41 No. SIG3 (TOD 6), pp. 37-45 (2000)
- [11] 椎谷秀一, 馬場孝之, 遠藤進, 上原祐介, 増本大器, 長田茂美, "画像情報とテキスト情報を統合的に利用したインタラクティブな映像検索システム", 研究報告: データベースシステム, 情報学基礎 No. 127, pp. 129-136 (2002)
- [12] 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", 研究報告: データベースシステム No. 125, pp. 177-184 (2001)
- [13] 国土地理院 <http://www.gsi.go.jp>