

2K-3

注視点を考慮した映像索引付け方式の提案と評価

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[†] 名古屋大学情報連携基盤センター[‡]

1. はじめに

今日、一般利用者が映像を扱う機会が増えている。撮影した大量の映像の中から自分の求めている部分映像を探し出すには、人手によりコンテンツ情報を書き加える手法が望ましいが、作業量が多く普及は期待できない。本稿では、GPS・電子方位計により取得した位置・方向情報を用いて、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ために注視点からの検索方式について提案する。本方式は、被写体とカメラの位置の幾何学的な位置関係を用いて、利用者の軌跡の部分に絞りを絞る方法である。本方式の検索精度を評価するため、撮影映像に対してプロトタイプを用いる。

2. 研究動機

デジタル映像機器の高性能化やネットワークの高速化より、映像を用いた情報交換が容易になりつつある。それに伴い長時間に渡る映像の中から、ある特定の被写体が映っている場面を検索することが必要となる。従来から、撮影した映像の中から見たい映像を探し出すために、さまざまな方法が提案されている[1][2]。本稿では、日常生活やスポーツイベントで撮影した映像が利用者間で交換、共有されることを想定し、映像に対して位置・方向情報を付加することで、位置・方向情報による被写体の位置を索引としての映像検索、または目標映像の抽出を目指す。本稿での目標映像とは、検索条件とした注視点映っている一連の部分映像とする。位置・方向情報は、映像収録と同時にGPSと電子方位計より取得する。GPS・電子方位計は表1に示すようなパラメータから成り、カメラの位置情報ならびにカメラを向けている3次元方向を記録する。また、日常生活やスポーツイベントなどで撮影される映像は、自由に動き回り撮影される。よって、撮影カメラの向きは一点に定まらず常に変化するため、位置・方向情報は1秒間隔で取得し、撮影者の動きに対応させる。

3. 注視点からの映像検索方式の提案

カメラの位置をキーとした視点からの映像検索は、GPSなどを用いてカメラの軌跡を記録しておき、検索者が指定した場所から撮影された部分映像を抽出する。しかし、自分の見たい部分映像を映像の中から探す場合、「どこから撮影したか」という視点からの検索ではなく、「どこを撮影しているか」という注視点からの検索が重要になることが多い。注視点からの検索は、注視点の緯度、経度と注視点からの検索範囲を用いて行う。検索手順を以下に示す。

1. 検索する被写体(図1中のA)の位置と検索範囲(図1中のr)を検索者が指定

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

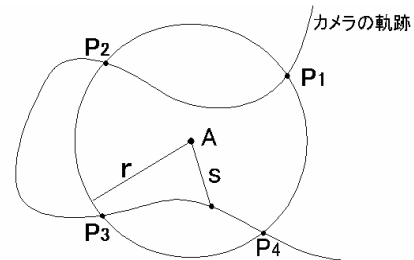


図1 2次元で表した検索範囲とカメラの軌跡

2. 指定された範囲内から撮影した映像を選び出す
3. 2で選び出された映像の中から、指定した被写体(図1中のA)が撮影カメラの画角内に入っている部分を検索結果とする。

1で検索したい被写体の位置Aと、検索範囲の半径rを指定する(図1参照)。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。2で、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。しかし、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ので、そのままでは距離sを計算することはできない。そこで、2点の緯度、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、指定領域内にカメラがあるかどうか分かる。図1では、P₁~P₂とP₃~P₄の部分映像を抽出することができる。3では、2で抽出したP₁~P₂間とP₃~P₄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を用いて、求めている被写体が映っている部分映像を検索結果とする。

4. 検索方式の評価

4.1 システム構成

検索者は、見たい被写体の緯度、経度と検索範囲を入力する。プロトタイプは、GPSと電子方位計から取得したデータを記述したXMLファイルを読み込み、3章で述べた手順に従い目標映像を抽出する。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ておく。プロトタイプの画角は水平方向、垂直方向とも60°とする。

4.2 評価(1)

プロトタイプに対し実験映像を用いて、撮影環境の違いが検索精度に影響するかどうかを検証する。検索精度の評価は、プロトタイプによって抽出した目標映像に対

A Video Retrieval Scheme with Gazing Points and its Evaluation

[†]Graduate School of Informatics, Kansai University

[‡]Information Technology Center, Nagoya University

表 2 実験映像の説明

	条件 1	条件 2	条件 3
映像 1	常に変化	2パターン	回転なし
映像 2	3パターン	一定	回転なし
映像 3	一定	動かない	3パターン

条件 1：撮影カメラと注視点の距離

条件 2：撮影カメラの移動速度

条件 3：撮影カメラの回転速度

表 3 映像 1 の検索結果

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%

し行う。検証する項目を以下に示す。

項目 1：撮影時のカメラの移動速度による影響

項目 2：撮影カメラと注視点との距離による影響

項目 3：撮影時のカメラの回転速度による影響

表 2 に、項目 1~3 を検証するために用いた映像 1~3 について示す。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。撮影場所の GPS の測位精度は 5~6m であった。なお、本稿のプロトタイプはカメラのズームアップに対応していないため、焦点距離一定で撮影を行った。検索精度の評価に用いた適合率、再現率について以下に定義する。

$$\text{適合率}(P) = \frac{\text{適合映像数}}{\text{抽出映像数}} \times 100 \quad (1)$$

$$\text{再現率}(R) = \frac{\text{適合映像数}}{\text{撮影映像数}} \times 100 \quad (2)$$

撮影映像数：映像内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：抽出した目標映像に注視点映っている数

表 4 よりカメラと注視点の距離が 15~40m では全ての場合において P は 75% 以上、 R は 67% 以上の精度となった。しかし、カメラと注視点の距離を自由にして撮影した場合では、普通に歩いて撮影したとき(表 3 参照)でも $P = 63%$ 、 $R = 60%$ となった。適合率、再現率が下がった原因としては、撮影カメラと注視点の距離が近づきすぎた場合にあると考えられる。緯度、経度は、地球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、映像撮影の場所で GPS により測った 0.1 秒における緯度、経度の距離は北緯方向に 5m30cm、東経方向に 5m80cm であった。同じ緯度、経度でこれだけの領域幅があるので、撮影カメラと注視点との距離が近くなれば、カメラの画面内に領域全体を映すことはできない。よって、カメラに映っている領域と注視点との位置関係により、指定した緯度、経度の領域が映っている部分映像を抽出しても注視点映っていない場合がある。表 5 より、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ると

表 4 映像 2 の検索結果

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%

表 5 映像 3 の検索結果

カメラの 1 秒間の回転角度	適合率	再現率
45°	86%	88%
90°	88%	60%
180°	72%	15%

表 6 実映像の検索結果

適合率	再現率
80%	63%

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。

4.3 評価(2)

実映像に対してプロトタイプを用いて注視点からの検索を行い、抽出した目標映像に対し評価を行う。用いた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。撮影場所の大学キャンパス内にある目印となりやすいものを注視点として選んだ。実映像に対する検索精度を表 6 に示す。 $P = 80%$ 、 $R = 63%$ と、前述した映像 1 を用いた場合に比べ良い精度となった。再現率が適合率と同じ程度の値になれば、日常生活やスポーツイベントで撮影された映像に対し適用できると考えられる。

5. おわりに

本稿では実験映像と実映像を用いて、提案方式を実装したプロトタイプにより抽出した目標映像に対し検索精度を評価した。実映像に対しての再現率が向上すれば、全てにコンテンツ情報を書き加える手法よりも簡便性の意味も踏まえ有効になると考えられる。今後の課題としては、検索インターフェースの改良や撮影カメラのズームアップへの対応が考えられる。

参考文献

- [1] 住野優ら, "3D 仮想空間を用いた動画蓄積システムの試作", 電子情報通信学会データ工学ワークショップ(DEWS2004)2-C-03
- [2] 兵清弘ら, "MPEG-7 を利用したウェアブルカメラ映像の索引付け手法", 電子情報通信学会第 13 回データ工学ワークショップ(DEWS2002)C2-17