移動オブジェクトの追跡による動画像インデクシングシステム

Video Indexing System using Moving Objects Tracking

正会員 宮内 新[†],笹代 孝次[†],渡邊 露文[†],石川 知雄[†]Arata Miyauchi[†], Kouji Sasadai[†], Tsuyufumi Watanabe[†] and Tomoo Isikawa[†]

あらまし 一般に、監視カメラの映像は、移動物体の動きの内容が最も重要な情報であり、移動物体が存在する時間のほうが短いという性質がある。そこで移動物体の動きの内容が一目で理解できるようになることが望ましい。本論文では、MPEG 動画像中の移動オブジェクトを追跡し、その追跡によって静止画を生成するという方式でインデクシングを行う手法を提案する。

キーワード：MPEG, インデクシング, 監視カメラ

1. ま え が き

監視カメラの映像では、移動物体の動きが最も重要な情報であり、図1のように移動物体が存在する時間のほうが短いという性質がある。そのため、動画像中の移動物体とその動きの内容およびその動きが発生している時間が一目で理解できるようになることが望ましい。しかし、従来のインデクシング手法では、移動物体の動きの内容を知ることができない¹⁾²⁾³⁾。そこで本研究では、固定カメラを想定し、MPEG 動画像中の移動オブジェクトを追跡し⁴⁾⁵⁾、移動オブジェクトの動きの変化を検出して、その場面から静止画を生成する方式でインデクスを自動で作成し⁶⁾、動画像中の動きを一目で理解する手法を提案し、実画像を用いて実験を行ったので報告する。

2. 本手法の概要

本手法では、図2に示す全体的な処理を全フレームに対して行う。まず、MPEG データから動き補償のための動きベクトル（以下 MV）を抽出し、MV をグルーピングして移動領域を抽出する。その後、移動領域のうち閾値より大きい移動領域を移動オブジェクトとみなして追跡し、各移動オブジェクトの動きに変化が生じた際に静止画を残すことによってインデクシングを行う。また、

位置情報を用いてオブジェクトの識別を行い、冗長なインデクス項目を削減する。

2.1 動きベクトル (MV) の抽出

MPEG 動画像のデータから直接 P ピクチャまたは B ピクチャを読み出し、MV が定義されているマクロブロック（以下 MB）を検出し、その部分だけを抽出する。

予備実験より、実際には動きの存在しない MB にも非常に小さな MV（そのほとんどは MV の大きさが 1）が含まれていることが分かっており、そのため本手法では、MV の大きさの閾値 ($Th1$) を設定し、これより MV の値が小さい MB は動きのない MB と推定して破棄する。

2.2 動きベクトルのグルーピング

画像上のオブジェクトは複数の MB にまたがる大きさを持っている可能性が大きい。そこで、同一の移動領域であると考えられる MV を持つ複数の MB をグルーピングし、1つの移動領域に対する MV を算出する。その MV のグルーピングの概念を図3に示す。

背景に近い色の移動領域や光源の関係などで、実際に動いている領域であっても MV の検出されない領域も存

図1 監視カメラの映像における移動物体
Moving object in the image sequence of the watch camera.

図2 1フレームの全体的な処理の流れ
Flow chart of the one frame process.

1999年12月3日受付, 2000年1月18日採録

[†] 武蔵工業大学

(〒158-8557 東京都世田谷区玉堤 1-28-1, TEL 03-3703-3111)

[†] Musashi Institute of Technology

(1-28-1, Tamazutsumi, Setagaya-ku, Tokyo 158-8557, Japan)

図3 動きベクトルのグルーピング

図4 オブジェクトの追跡及びインデクシング
Object tracking and indexing.

在する。そのため、実際は同一の移動領域であっても、領域が分かれていることがある。このような状況も考慮し、注目するMBの周囲2ブロックまで同一移動領域になりうるとみなし、接していない場合もグルーピングの対象とする。グルーピングは以下の手法で行う。

各MVの水平成分(MV_x)、垂直成分(MV_y)から、そのMVの角度を(1)式により算出する。

$$\theta = \tan^{-1} \frac{MV_y}{MV_x} \quad (1)$$

1ピクチャ内でMVが存在するすべてのMBについて周囲2ブロックを走査し、(2)式を満たすものを同一領域として統合する。

$$\theta - \phi \leq \theta' \leq \theta + \phi \quad (2)$$

θ : 走査対象の角度, θ' : 注目MVの角度, ϕ : 閾値

1フレームすべてのMBを探索しグルーピングが終わった後、それぞれの移動領域内の各MBのMVの平均 $\overline{MV}$ および重心の座標 $C = (x_c, y_c)$ を計算し、移動オブジェクトの動きは C に $\overline{MV}$ が存在するものとする。

2.3 オブジェクトの追跡およびインデクシング

移動領域の大きさの閾値($Th2$)を、カメラ位置から設定し、それより大きい移動領域を移動オブジェクトとみなす。ここでは、監視用にカメラを固定することにしたため、監視対象物(人など)の大きさから画像上での大きさを定めることができる。この値を用いて大きさの閾値を決定し、移動オブジェクトか否かを決定し、各移動オブジェクトの動きによって追跡を行う。この追跡で、オブジェクトの動きの出現、消失、複数オブジェクトの交差を検出する。複数のオブジェクトの交差については、交差によって生じる一番手前のオブジェクト以外のオブジェクトの動きが消失し、新たに出現すると捉えることができるため、一番手前のオブジェクト以外のオブジェクトの動きの消失、別のオブジェクトの動きの出現として検出する。

本手法におけるオブジェクトの追跡の概念図を図4に示す。この図は、右から左に動く物体1と上から下に動く物体2が交差する内容の動画で、物体2が出現したと

ころから交差が終了したところまでを表した図である。(a)では物体2の出現を検出し、(b)では2つの物体に出現、交差、消失がないため、何も検出しない。(c)で物体1と物体2が交差しているので、これを陰に隠れた物体2の消失として検出する。(d)で陰から出てきた交差後の物体2を新たな物体3の出現として検出する。このオブジェクトの追跡を、現フレームのオブジェクトの $\overline{MV}$ と前フレームのオブジェクトの $\overline{MV}$ との間でマッチングを行うことによって実現する。前フレームでオブジェクトの $\overline{MV}$ が検出され、現フレームで検出されない場合がオブジェクトの動きの消失となる。また、前フレームではオブジェクトの $\overline{MV}$ が検出されず、現フレームで検出される場合がオブジェクトの動きの出現となる。この場合に生成されるインデクスには、3枚の静止画(a), (c), (d)が残る。

本手法では、これらの移動オブジェクトの動きの変化が生じた場合に、復号の計算量を減少させるため、その直後のIピクチャを静止画像として残し、それを繰り返すことによってインデクシングを行う。静止画の形式は、多数の静止画を生成するためファイルサイズを考慮し、JPEG形式とする。

2.4 位置情報によるオブジェクトの識別

これまでに述べた方法では、オブジェクトが奥行方向への動きや、小さな動きをした際に、オブジェクトのMVを含むMB数が閾値($Th2$)に満たないフレームが発生する。これによって、実際の動きには変化が生じていないものでもオブジェクトの動きの消失・出現が検出され、重要でない冗長なインデクス項目がインデクスに残される。また、2.3節のオブジェクトの追跡は、移動オブジェクトそのものではなく移動オブジェクトの動きを追跡したものであるため、1つのオブジェクトの一連の複数の動きがそれぞれ別のオブジェクトの動きとみなされる。そのため、オブジェクト数も冗長になる。

そこで、消失したオブジェクトの動きと出現したオブジェクトの動きとのそれぞれの消失・出現位置によって、オブジェクトの識別を行い、2つのオブジェクトの動きが同一オブジェクトのものである場合に冗長なインデクス項目の削除を行う。新たなオブジェクトの動きが出現した際に、過去に消失したオブジェクトの動きとマッチングを行い、マッチングが成功した2つのオブジェクトの動きを同一オブジェクトとみなすことによってオブジェクトの識別を行う。マッチングに次の5つの条件を用いる。

- (1) 消失時の重心と出現時の重心との距離がオブジェクトとみなす大きさの閾値 $Th2$ の半分($Th2/2$)未満
- (2) 消失時と出現時のオブジェクトの大きさの差が出現時の大きさの閾値 $Th3$ 以内
- (3) 消失から出現までの時間が閾値 T 以内

表 1 エンコードパラメータ
Encoding parameter.

パラメータ	実験画像 1~4	実験画像 5~8
画面サイズ	704×480	
フレーム間隔	1 / 30 [秒]	
GOP 構造	I P P P P	I B P B P B P B
GOP 内のピクチャ数 (N)	5	8
I または P ピクチャの出現周期 (M)	1	2
伝送速度	6 Mbps	

表 2 実験パラメータ
Encoding parameter.

パラメータ	実験画像 1~4	実験画像 5~8
Th1	2	
Th2	50(60cm×60cm)	40(60cm×60cm)
φ	90°	
Th3 (%)	50	
T	60 秒	
対象フレーム	P	

(4) 消失したオブジェクト, 出現したオブジェクト
がともに画面の枠に触れていない

(5) 消失位置を他の物体が横切っていない

以上で述べた各閾値は本システムの使用目的に応じて定める. インデクスを作成する際には JPEG 画像の他に, オブジェクトの動きの変化の内容とそれが生じたフレーム番号を, インデクスの項目として, HTML ファイルに記述する. 1つの動画画像で, インデクス項目が記述された HTML ファイルを 1つ生成する. この HTML ファイルで JPEG 画像を表示することによって, WWW ブラウザを用いて移動物体の動きの内容とその場面を一目で理解することが可能となる.

3. 実験および結果

3.1 実験内容

デジタルビデオカメラ (SONY DCR-VX1000) で撮影され, MPEG-2 エンコーダ (富士通 FMMPE-201) でエンコードされた MPEG-2 動画画像を用いて実験を行った. 動画画像の内容は, 室内で撮影したもの 4 本 (実験画像 1~4) と, 室外から撮影したもの 4 本 (実験画像 5~8, 天候:曇り) の計 8 本である. エンコードパラメータを表 1 に示す. また, 実験では, ドアを監視する画像を用いてドアを出入りする人物を検出してインデクスを作成することとし, ドア位置での人物の画像上での大きさと, ドア開閉時の人物の動きの大きさから各パラメータを表 2 に示すように定めた.

3.2 実験結果と評価

本手法の定量的評価の基準値として, 理想値という値を設ける. これは, 筆者の判断によって定める検出されるオブジェクト数と生成される静止画の枚数の必要十分な値である. この理想値は, 画面内の 1つの物体を 1つのオブジェクトとみなし, オブジェクトの動きの変化 (出現または消失) が発生した際に 1枚の静止画が生成されるものとして決定した. ただし, オブジェクトの動きの変化が同時に複数発生した場合には静止画は 1枚生成されるものとした.

表 3 実験画像 6 のインデクスの一部
A part of index pictures of image sequence No.6.

フレーム	インデクス項目	
	オブジェクトの識別なし	オブジェクトの識別あり
68	オブジェクトの動き 1 出現	オブジェクト 1 出現
72	静止画 1 (図 5(1)) の生成	静止画 1' (図 5(1)) の生成
74	オブジェクトの動き 1 消失	オブジェクトの識別
76	オブジェクトの動き 2 出現	により削除
80	静止画 2 (図 5(2)) の生成	
98	オブジェクトの動き 2 消失	オブジェクト 1 消失
100	オブジェクトの動き 3 出現	オブジェクト 2 出現
104	静止画 3 (図 5(3)) の生成	静止画 2' (図 5(3)) の生成
138	オブジェクトの動き 3 消失	オブジェクト 2 消失
144	静止画 4 (図 5(4)) の生成	静止画 3' (図 5(4)) の生成

図 5 実験画像 6 のインデクシングで生成された静止画の一部
A part of JPEG pictures generated by indexing of image sequence No.6.

表 4 インデクスに残る静止画の理想値との比較
Comparison of experimental results and ideal results.

実験画像	静止画の枚数		枚数の差 (枚)	生成率
	理想値	実験値		
1	5	2	-3	0.40
2	9	11	2	1.22
3	14	19	5	1.36
4	5	9	4	1.80
5	12	22	10	1.83
6	4	6	2	1.50
7	6	16	10	2.67
8	20	49	29	2.45

本手法によるインデクシングで生成される静止画, インデクス項目の例を, 図 5, 表 3 に示す. オブジェクトの識別なしでは, オブジェクトの動きの出現・消失が検出され, その次の I ピクチャ (フレーム 72, 80, 104, 144) から静止画が生成された. オブジェクトの識別ありでは, 識別なしでの処理に加えて, オブジェクトの動きの出現時にオブジェクトの識別が行われ, オブジェクトの動き 2 がオブジェクト 1 と識別されたオブジェクトの動き 1 と同一オブジェクトとみなされた. そのため, 識別なしにおける “オブジェクト 1 の動きの消失” と “オブジェクト 2 の動きの出現” に相当する項目が削除され, “オブジェクトの動き 2 の消失” に相当する項目が “オブジェクト 1 の消失” とされた. これらから, 冗長なインデクス 2 が削除され, またオブジェクトの動きの変化を取りこぼすことなく, インデクシングが行われていることがわかる.

図 6 に結果の出力例, 表 4 にオブジェクトの識別を行った後, インデクスに残った静止画の枚数と理想値を比較したものを示す. 表中の枚数の差, 生成率は, それ

図 6 結果出力例
An example of output.

ぞれ次式によって計算した。

$$\text{枚数の差 (枚)} = \text{実験値} - \text{理想値} \quad (3)$$

$$\text{生成率} = \text{実験値} / \text{理想値} \quad (4)$$

ほとんどの実験画像で、オブジェクトの動きを取りこぼすことなく、インデクシングが行われていた。室内で撮影した画像である実験画像 1~4 の結果は、生成率が最高で 1.80 であり、良好な結果といえる。

実験画像 1 は、(1) 人物が出現し、(2) ドアを開け（ドアの動きの出現）、(3) ドアが開いて止まり（ドアの動きの消失）、(4) 人物がドアの死角に消え、(5) ドアが閉まるものであり、上記 5 枚のインデクス画像ができるものを理想値とした。実験結果では、ドアを開ける動作をした人物の体が大きく、ドアに被さるようにしてドアを開けたため、ドアを開ける動作とそれ以前の動作を識別することができなかった。さらに、ドアを開けた直後、人物がドアの右の死角に消えたため、消失位置がドアの位置と重なり、ドアが閉まる動作がこの人物と同一オブジェクトであると誤認識された。これらの 2 つの不具合が重なったため、理想値より 3 枚少なくなった。しかし、インデクスには人物の出現とドアが閉まっている状態が静止画として残されていた。この 2 枚の静止画があれば、この 2 枚の静止画が生成される間に 1 枚目の静止画に写っている人物が消失したことが推測できるため、移動物体の動きが発生した時間と移動物体を理解することができる。ゆえに、移動物体の動きが発生する場面を一目で理解するという目的は果たしているといえる。

室外で撮影された実験画像 5~8 の実験結果は、生成率が 2.00 以上のものが 2 つあった。この要因としては、自然光の強弱の影響がある。実験画像 5~8 が撮影された環境では、一方向だけが明るく、オブジェクトの影が発生する。光が弱く周りが暗いときには影は薄くなり、影がオブジェクトと認識されることはない。影をオブジェクトと認識すると、影の動きによって静止画が生成され、冗長な静止画の生成につながる。さらに、光源が自然光

であるために明るさの変動が大きく、影がオブジェクトと認識される場合と認識されない場合が同一動画中に混在した。実験の結果から、本手法は戸外のような自然光の環境下でもインデクスを理想の 3 倍未満で、照明条件が一定な室内ではインデクスを理想の 2 倍未満で作成することができ、動画像の内容を容易に把握することができることが分かった。

4. む す び

本研究では、動画像中の移動物体の動きを一目で理解することを可能とする手法の提案および検討を行った。本手法では、MPEG データから直接 MV を抽出し、グルーピングされた MV によってオブジェクトの追跡を行い、オブジェクトの動きに変化が生じた際にその場面を静止画として残すことによってインデクシングを行った。そして、実画像を用いた実験の結果、移動オブジェクトの動きの変化をほぼ取りこぼすことなくインデクシングを行うことができ、動画像中の移動物体の動きを一目で理解することができた。

今後の課題は、DCT 係数を用いたオブジェクト識別の精度向上やパン・ズームへの対応、アルゴリズムの改良による高速化などがある。

【文 献】

- 1) 長坂晃朗, 田中譲: “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”, 情報学論, 33, 4 (Apr. 1992)
- 2) 谷口行信, 外村佳伸, 浜田洋: “映像ショット切換え検出法とその映像アクセスインターフェースへの応用”, 信学技報, PRMU-96-99 (1996) (Apr. 1996)
- 3) 佐藤隆, 新倉康巨, 谷口行信, 阿久津明人, 外村佳伸: “MPEG2 映像からのカット点とテロップの高速検出法”, 信学技報, PRMU-96-99 (1996)
- 4) 馬淵睦実, 櫻村雅章, 宮内新, 全炳東, 小沢慎治: “MPEG 符号の直接操作に基づく動画解析”, 信学技報, PRMU96-102 (1996)
- 5) Yasuo Ishii: “Estimation of 3 Dimensional Motion Vector via MPEG Picture”, Proc. IEEE, ITC-CSCC '97, pp.1215-1218 (1997)
- 6) 渡邊露文, 宮内新, 石川知雄, 全炳東: “移動オブジェクトの追跡による MPEG 動画像のインデクシング”, 情報処理学会 全国大会 5k2 (Mar. 1999)

みやうち あらた
宮内 新 1983 年、慶應義塾大学大学院工学研究科博士課程修了。同年、相模工業大学情報工学科専任教師。1990 年より、武蔵工業大学電子通信工学科助教授。1997 年、武蔵工業大学電子情報工学科助教授。画像、音声の符号化・認識の研究に従事。正会員

さいだい こうじ
笹代 孝次 1999 年、北海道工業大学工学部電気工学科卒業。1999 年、武蔵工業大学大学院工学研究科入学。画像処理の研究に従事。

わたなべ とうふみ
渡邊 露文 1997 年、武蔵工業大学工学部電子通信工学科卒業。1999 年、同大学院電気工学専攻修士課程修了。在学中、画像処理の研究に従事。現在、富士ソフト ABC に勤務。

いしかわ ともお
石川 知雄 1954 年、東北大学工学部通信工学科卒業。1956 年、米国スタンフォード大学大学院修士課程修了。1960 年、同大学大学院博士課程修了。PhD。同年、日立製作所入社。1988 年、武蔵工業大学電子情報工学科教授。計算機アーキテクチャ・オペレーティングシステムの研究に従事。