

自然動画像中の対象物位置検出方式

正会員 富高忠房[†],関谷庸男^{††},景山浩二^{†††}

The Detection Method of Object Position in Moving Images

Tadafusa Tomitaka[†], Tsuneo Sekiya[†] and Kouji Kageyama^{††}

Abstract An algorithm for tracking the position of a moving object in a naturally moving image is proposed. In this algorithm, the first step is to simply identify the object. Then, the object is segmented from the background image by the local histogram growing method. The histograms in the perceptual color system (hue, saturation, brightness) are insensitive to video-camera movement and rotation, as well as deformation, shrinkage, and enlargement of the object. Last, the position of the object is detected by the optimal histogram matching method. Real-time experiments have verified the effectiveness of this technique.

1. ま え が き

近年、民生用ビデオカメラの進歩は著しく、デジタル処理技術の発展とあいまって、AF(Auto Focus), AE(Auto Exposure), AWB(Auto White Balance), 手ぶれ補正が開発され、ユーザがビデオカメラの操作に惑わされることなく手軽に質の高い映像を撮影することが可能になってきた。しかし現状のシステムにおいて、さまざまな不具合点が指摘されている。例えば、AFは画面の中央に合焦するために、被写体が画面の端にいた場合、背景に合焦する欠点がある。また、AEにおいては、画面内の被写体の位置とは無関係に、画面内の輝度を場所ごとに決まる固定の係数で重み付け加算した値が一定になるようにアイリスを制御している。そのため、被写体が画面中央に

いない場合、過順光、逆光状態で、被写体の明るさを最適にすることが困難である¹⁾。これらの問題を解決するには、前処理として被写体を抽出することが必要である。また、被写体を抽出し、パン・チルト機構や電子ズーム機能と併用することにより、被写体を自動追尾し撮影することが可能になる。これは、TV会議、TV電話など、自動的に被写体を画面内に収めることが必要な用途に対して重要な技術となる。

これまで動画像の中から対象物を抽出する手法について、さまざまなアプローチがされている。この技術を応用した商品の使われ方に注目すると

(1) カメラが固定されているかどうか

(2) 追尾する対象物を特定するかどうか

という要因で、開発すべきアルゴリズムの内容も大きく変わる。

キーワード: 画像認識, 画像処理, ヒストグラム, メンバシップ関数

1993年7月15日, 情報処理学会グラフィックスとCAD研究会(テレビジョン学会 画像処理・コンピュータビジョン研究会共催)で発表
1994年7月18日受付, 1995年2月17日再受付

[†] ソニー株式会社 CAVカンパニー商品開発研究所 (〒141 東京都品川区北品川5-9-12, TEL 03-5488-6353)

^{††} ソニー株式会社 CAVカンパニーPAV部門(同上)

^{†††} ソニー株式会社 中央研究所(〒141 東京都品川区北品川6-7-35, TEL 03-3448-5901)

[†] Consumer A & V Products Company, Sony Corporation (5-9-12, Kita-shinagawa, Shinagawa-ku, Tokyo 141, Japan)

^{††} PAV Div., CAV Company, Sony Corporation (Same as above)

^{†††} Research Center, Sony Corporation (6-7-35, Kita-shinagawa, Shinagawa-ku, Tokyo 141, Japan)

カメラが固定の場合、背景画像との画素単位の輝度差分を用いることで被写体を抽出することが可能になる。この問題点として、背景画像との残差が0にならないことや、時間経過に伴う照明や天候変化に弱いことが指摘されており、安居院ら²⁾によりこの改善手法が提案されている。

本論文では、手持ちのビデオカメラを対象にしているため“カメラを固定しない”用途での手法を開発する必要がある。このような場合、背景、対象物とも手ぶれやパン、チルトで動くため、背景画像との差分では対象物を抽出できない。FA分野では、テンプレートとの正規化相関を用いたパターンマッチングで対象物を抽出する方法が実用化されている³⁾が、適用条件として、テンプレートと抽出される対象物との形の変化がほとんどなく平行移動だけの動きで、また、照明も専用の固定照明であるために、今回の応用には使えない。他の手法として宮脇らは、色彩情報を用いて領域分割する方法⁴⁾を提案している。この手法は光源の明暗に強いという利点をもつが、背景と対象物が近い色彩の場合、分離が困難である。この問題に対して筆者らは、色相と輝度情報を使うと共に、マッチングに用いる類似度計算法を撮影シーンごとに最適化する(最適ヒストグラムマッチング法)ことにより上記の問題点を解決する。

もうひとつの問題点として“追尾する対象物の特定方法”がある。後処理を楽にするために、ユーザにできるだけ詳しく対象物を示してもらうことが望ましい。例えば、対象物の輪郭を太い閉曲線で荒くなぞる方法が提案されている⁵⁾。しかし、ビデオカメラを片手で持って操作するユーザの便宜をはかるためには、この方法は複雑すぎる。そこで今回、特定したい対象物を画面の中央におき、トリガボタンを押すことで対象物の特定が完了する操作を採用する。そのために必要な初期画面での対象物領域抽出アルゴリズムとして、局所領域でのヒストグラムを成長させ、対象物の領域を抽出する局所ヒストグラム成長法を提案する。

また、できるだけハードウェア規模を少なくするために、デシメーション画像を利用することによりメモリーを減らし、演算処理は専用ハードをもたずマイコンで処理できるアルゴリズム開発を行った。

本論文では、まず全体のアルゴリズムを示し、初期画面からの対象物領域の抽出を行う局所ヒストグラム成長法と、任意の画像で最適な類似度計算法を求め対象物の位置を検出する最適ヒストグラムマッチング法について説明する。最後に、リアルタイム実験の結果を報告し、その有効性を示す。

2. 前処理とアルゴリズムの流れ

本手法は以下の手順からなる。前処理として4:1:1のビデオ信号を輝度信号Yを14.3 MHz ($4f_{sc}$)、色差信号R-Y, B-Yを3.5 MHz (f_{sc})でサンプリングし、1フィールドで768×240 pixelのYと192×240 pixelの色差信号R-Y, B-Yを得る。階調は各8ビットで表現する。この信号に対して、エリアジリング防止のため、水平方向・垂直方向にローパスフィルタLPFを通した後デシメーションを行い、48×30 pixelのY, R-Y, B-Yを得る。デシメーションを行う理由は以下の4項目である。

- ① 上記のカラー画像を人間が見て充分、対象物を抽出できるという筆者らの実験事実
- ② ①を裏づける Harman(1973)の人間の認知能力に対する見解⁶⁾
- ③ 本アルゴリズムが均一な領域を抽出しやすくなるためのノイズ成分除去
- ④ ハードウェアとしてコストに最も影響するメモリー量の削減

次に、照明の明暗、照明と対象物の位置で変化する照明むら、陰影の影響を受けにくい色相成分をもつ知覚的表色系の輝度 $Y(x, y, f)$ 、色相 $Hue(x, y, f)$ 、飽和度 $Sat(x, y, f)$ に変換する。ここで x, y は、図1に示すデシメーション画像の水平・垂直位置、 f は撮影シーンのフィールド数($f=0$ が初期画面)を示す。

- (1) 特定したい対象物を画面の中央(指示枠の上)に置き(図1)、トリガボタンを押す(ユーザによる対象物の指定)。
- (2) “局所ヒストグラム成長法(Local Histogram Growing Method)”により初期画面の中央から対象物のヒストグラムテンプレート(大きさ

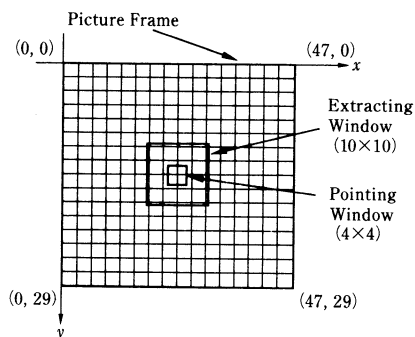


図1 画面内の抽出枠と指示枠
Extracting window and pointing window in a image.

10×10 pixel)をつくる．

- (3) 画面を水平方向に5分割，垂直方向に3分割し，各領域のY, Hueの各ヒストグラムを求める．この時点では，(1)の条件を仮定しているため，対象物は必ず画面中央にある．画面を5×3に分割すると，1つの領域がデシメーション画像で10×10 pixel (一部，9×10 pixel) になり，それがテンプレートの大きさとなる．このテンプレートの大きさは，これまで開発されたビデオカメラのAF, AEがもつ中央の枠の大きさを参考にし，また，画面内を同じ大きさで分割できることも考慮して決定した．対象物の大きさは，これより大きくても小さくてもさしつかえない．
- (4) (3)の各領域を用いて，特定した対象物と背景を最もよく分離できる最適なヒストグラム類似度演算を選択する．ここまでの処理は初期画面($f=0$)で行う．
- (5) 2フィールド目($f=1$)以降の画面において，ヒストグラムテンプレートの位置を動かしながら，各点での画像との類似度を(4)の演算法で求める．類似度が最小(最も似ている)の点が次画面での特定対象物の位置となる．これを最適ヒストグラムマッチング法(Optimal Histogram Matchin Method)と呼ぶことにする．
- (6) 対象物の位置の決定は，3×3のメディアンフィルタ処理を施した後行う．これは決定位置のふらつきをなくすためである．
- (7) (5)にもどる．
- 以下では，これらの各手順について説明する．

3. 局所ヒストグラム成長法による対象物のヒストグラムテンプレートの抽出

画像中の対象物を特定するために，初期画面から対象物の領域を対象背景と分離して抽出する必要がある(セグメンテーション)．

今回，新たに開発した局所ヒストグラム成長法(Local Histogram Growing Method)を用いてセグメンテーションを行う．つぎに，この方法について説明する．

初期画面の中央に抽出枠，その中に指示枠を設ける(図1)．抽出枠の大きさは，デシメーションした48×30 pixelの画面で10×10 pixel，指示枠は4×4 pixelとする．抽出枠の大きさは画面を5×3に分割した1領域の大きさである．指示枠は対象物を指示するために小さい方がよいが，その枠内の情報が対象物の特徴を代表している必要があるため，対象物内の特異な画素だけを抽出しないために，1×1 pixelでなく，4×4 pixelとした．この大きさは，対象物の大きさによらず一定とする．

指示枠は，対象物を特定するとき，その対象物上に1/60 s(1フィールド)以上置き(画面の中央に対象物をもってくる動作)トリガボタンを押す．この操作で対象物がポインティングされる．その時得られた指示枠内の4×4 pixelの画素からY, HueのヒストグラムYPoint(i), HuePoint(i)をつくる(図2(a))． i はヒストグラムの要素になる．このとき，Sat.がレベル T 以下のHueに関しては頻度をカウントしない．これは，無彩色の対象物ではHueデータがノイズに埋もれて使用できないからである．このSat.レ

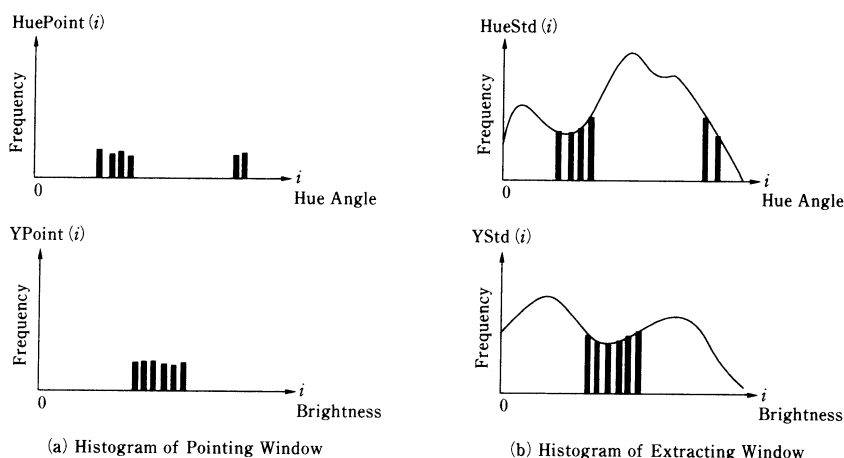


図2 局所ヒストグラム成長法
Local Histogram Growing method.



図 3 セグメンテーション前
Before segmentation.

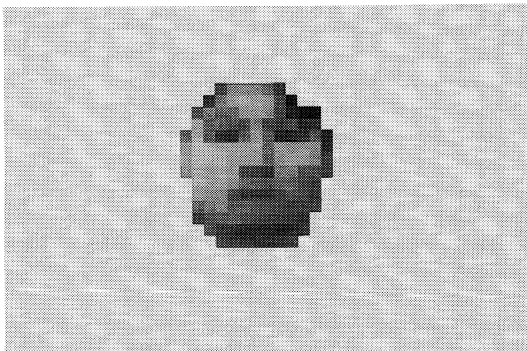


図 4 セグメンテーション後
After segmentation.

ベルの設定はカメラシステムの違いにより変化するものであるため、各システムごとに固有の値が決められる。

つぎに、指示枠のヒストグラムに成分をもつ要素 i が、特定したい対象物の特徴を反映しているわけであるから、抽出枠で指示枠と同じ要素 i だけについてヒストグラムをつくる(図 2 (b))。カウントされた要素は縦の棒線で示している。この過程で、抽出枠内にある特定したい対象物以外の背景(曲線で示した部分)が自動的に除外される。ここで得られた抽出枠の Y , Hue のヒストグラム $YStd(i)$, $HueStd(i)$ が特定したい対象物の特徴量となる。これが対象物のヒストグラムテンプレートになる。

ヒストグラムテンプレートは、2次元の画像情報を1次元に変換し情報を減らすと共に、形の情報を保持しない特徴がある。手持ちのビデオカメラで顔を対象物として人を撮影するような場合、上下左右の向きの変化により顔の大きさ・形が変化するので、このヒストグラムによる特徴量が有効となる。特に、照明むらや陰影の影響を受けにくい Hue によるヒストグラム

が対象物の連続的な位置検出には有効である。
本手法の効果を図 3, 図 4 に示す。

4. 対象物テンプレートの最適化と ヒストグラムマッチング

—最適ヒストグラムマッチング法—

(OHM: Optimal Histogram Matching)

対象物の位置を画面内で正確に検出するには、対象物の特徴量とそれ以外の場所での特徴量との類似度が最大(似ていない)になり、初期画面で得た対象物のヒストグラムテンプレートと、それ以後の画面での対象物の特徴量との類似度が最小(似ている)になる特徴量を使う必要がある。この特徴量は、背景と特定する対象物に対して最適化することで位置検出精度を上げることができる。

単純な例として Hue , Y どちらの情報が有効かという判断は、対象物が $Hue=100$, $Y=50$, 背景が $Hue=50$, $Y=50$ の場合、 Hue だけの特徴量にする方が対象物の位置検出精度は上がる。このように、背景と対象物の関係は自然画では無限に存在し、普遍的な特徴量をただ1つ決めることが不可能である。そこで今回、3章で得られた対象物のテンプレートと画面の他の領域とを差別化できるように、ヒストグラム類似度演算法として最適なものを選択する。対象物はユーザが指定するために追尾中に他のものと入れ替わることはないが、背景はパンニング、チルティングをすることで変わることが予想される。しかし、一般的な自然の背景では、それほど急激に背景の特徴は変化せず、初期画面と似たような特徴が継続するため、本手法で最適化する効果は充分あると考えられる。

最適なヒストグラム類似度演算法の選択は以下のように行う。図 5 に示すように、局所ヒストグラム成長法で得られた対象物のヒストグラムテンプレート $HueStd(i)$ と、画面を 5×3 に分割した1つの領域から得られるヒストグラム $HueHist(i, x, y, f)$ の成分をベクトルと考え、2つのベクトルの距離を計算した値を類似度とする。

類似度演算として以下の3種類の方法を使う。

- (A) Histogram Euclid Distance 2つのベクトル間の距離の二乗を計算したもの。
- (B) Histogram Hamming Distance それぞれのベクトルの長さを2値化して求めた距離の二乗を計算したもの。
- (C) Integral Histogram Distance 2つの n 次元ベクトル(n は標準パターンで0でないベクトルの次元)の各要素を加算することで1次

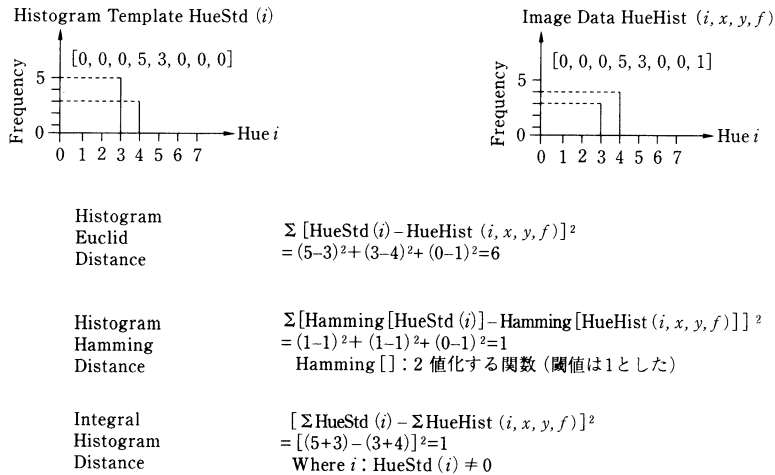


図 5 最適ヒストグラムマッチングの類似度計算法
Similarity calculations of optimal histogram matching.

元ベクトルにし、距離の二乗を計算したもの。
これらの類似度計算は、次のような対象物の特徴を反映している。

(A)は、ヒストグラムの要素を次元とする2つのベクトル HueStd と HueHist のユークリッド距離を示している。ベクトル HueStd と HueHist の大きさは、その要素の画素数つまり面積を、向きは Hue の分布状況つまり対象物の構成色を示している。つまり、この類似度演算は、対象物とそれ以外の画面(背景)との、面積、構成色を同時に表現したものになる。もし対象物の大きさが変わった場合、その影響が面積の変化としてあらわれるために、距離も大きくなる。

(B)は、ベクトル HueStd と HueHist の大きさを無視して、対象物の構成色だけに注目した距離演算である。

(C)は、逆にベクトル HueStd と HueHist の大きさだけに注目している。つまり、標準パターンに選ばれた画素(セグメンテーションされた領域)についての面積だけの特徴量にしている。

つまり、対象物と背景の分離に最適な情報が、面積である場合、(C)が選ばれ、色合いの場合は(B)が選ばれ、その中間の場合は(A)が選ばれる。

図5は Hue について示したが、色のないグレイスケールの対象物に対して Y, 両方の特徴をもつことが有効な対象物に対して Y+Hue を画素情報として持つことにする。その結果、次の9通りを類似度演算候補としてあげる。

- (1) Hue Histogram Euclid Distance (H-HED)

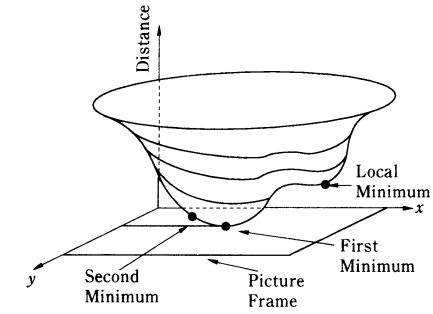


図 6 対象物テンプレートと画像内各領域間の類似度表示

Similarity map between object template and any area in an image.

- (2) Hue Histogram Hamming Distance (H-HHD)
- (3) Hue Integral Histogram Distance (H-IHD)
- (4) Y Histogram Euclid Distance (Y-HED)
- (5) Y Histogram Hamming Distance (Y-HHD)
- (6) Y Integral Histogram Distance (Y-IHD)
- (7) Y + Hue Histogram Euclid Distance (YH-HED)
- (8) Y + Hue Histogram Hamming Distance (YH-HHD)
- (9) Y + Hue Integral Histogram Distance (YH-IHD)

この9種類の類似度演算法に対して、画面全体の5×3の各領域とテンプレートの類似度を計算する。

ある1つの類似度演算法を用いて、ある画像の1フィールドとそのシーンのテンプレートとの類似度計算した結果を図6に示す。この図で、 x, y は、図1と同様に画像の水平・垂直位置を示し、Distance軸は、ある類似度演算を使って求めた距離を示す。画面を 5×3 に分割するため、距離として15カ所の値が得られるが、作図の都合上、図6ではなめらかな曲線で結んでいる。First Minimumは、画面内で距離が最小(似ている)の位置で、Second Minimumは2番目に距離が小さい位置である。Local Minimumは距離が最小で

ない極小値をもつ位置である。
 最適な類似度演算法は、以下の条件を満たすものを1つ選択する。
 (条件) 9つの類似度演算法の中で、初期画面の距離演算結果の第二最小値/第一最小値が最大の類似度演算法
 ただし、第一最小値が0の時は、第一最小値を1とする。
 この方法により、初期画面の対象物と背景を最もよく分離する類似度演算法が選択される。以後の画面で

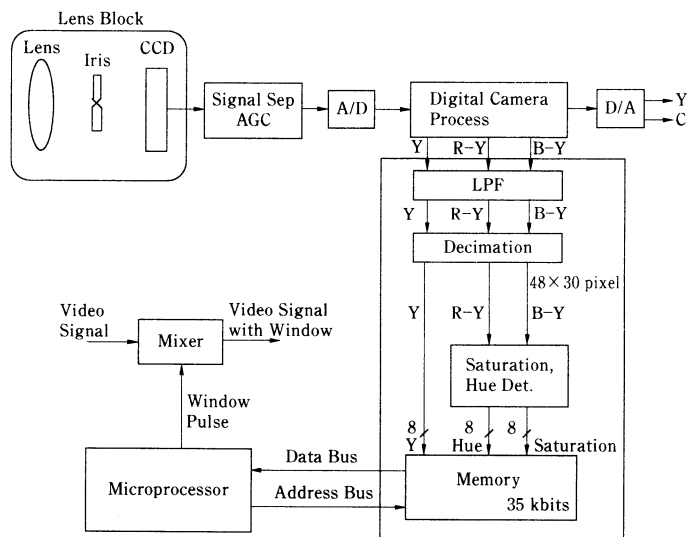


図 7 実験機のブロック図
 Block diagram of experimental system.

表 1 5のシーンによる実験結果
 Experimental results at 5 scenes.

	対象物	背 景	選択された類似度演算	2nd Min/1st Min	認識率
シーン 1	玩具の汽車 (青, 赤, 黄)	レール(赤, 黄, 緑) 机(グレイ) モニタ(白, 黒) 板(青)	H-HHD	21.3	100%
シーン 2	ウーロン茶のパッケージ (茶色)	机(薄茶色) 本, モニタ(白)	YH-HHD	39.3	100%
シーン 3	顔の石膏像 (白)	モニタ(白, 黒) カーテン(肌色)	Y-IHD	53.1	100%
シーン 4	座っている人 (肌色)	窓(逆光) 機材(黒) カーテン(肌色) 壁, ラック(薄茶色)	YH-HED	18.5	100%
シーン 5	人間の顔 (肌色)	カーテン(肌色) ガラス越しの壁(肌色)	YH-HED	3.3	95%

は、この類似度演算法を用いて対象物の位置を検出していく。位置精度は、デシメーション画面48×30 pixel内で大きき10×10 pixelのテンプレートをスキャニングしているので、画面で横48、縦30分割の精度をもつ。

5. 実験結果

図7に示す実験機を試作し、本アルゴリズムを用いた対象物位置検出の実験を行った。検出結果は2フィールドごとに白枠として出力される。今回の実験では、各画素の $Y(x, y, f)$, $Hue(x, y, f)$ は256階調で、各領域のヒストグラム $HueHist(i)$, $YHist(i)$ の横軸 i は32階調で表現した。実験に用いたシーンは、1シーン30秒(1800フィールド)の連続シーンで、背景は対象物の移動に伴い変わっていく。また、カメラを片手で持っているために完全に固定されてなく手ぶれをおこしている。実験に用いたカメラはソニー製CCD-TR 205で、AF, AE, AWBは通常の状態で

動作している。実験は、蛍光灯下のオフィスで、窓からは晴天の外光が差し込んでいる日中に行った。

今回の実験で用いたシーンと結果を表1に示す。

写真1～写真5に対象物の位置を検出している様子

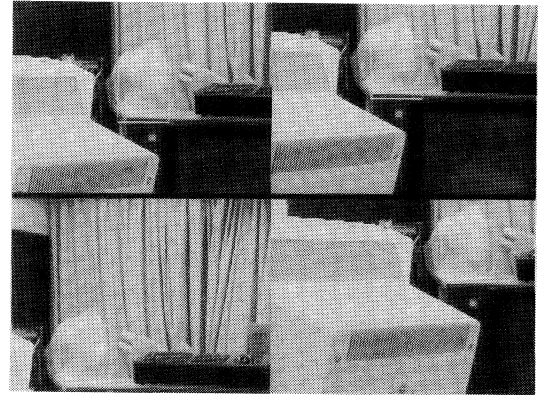


写真3 シーン3
Scene No. 3.

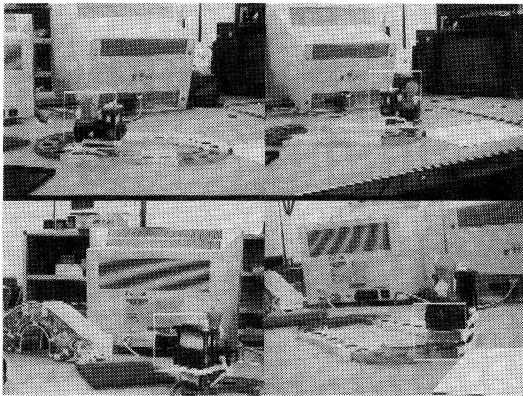


写真1 シーン1
Scene No. 1.



写真4 シーン4
Scene No. 4.

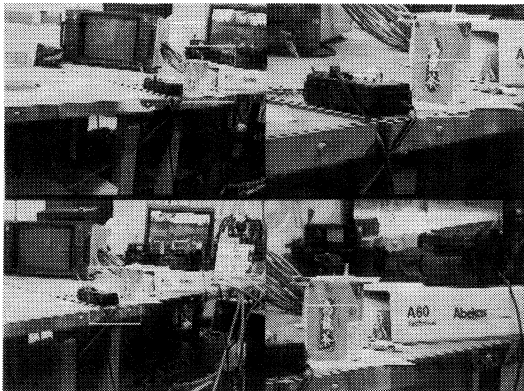


写真2 シーン2
Scene No. 2.

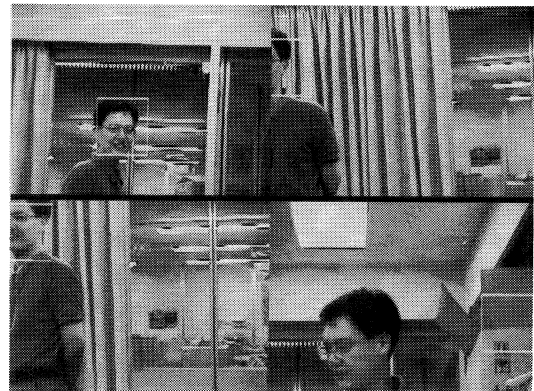


写真5 シーン5
Scene No. 5.

表 2 シーン 1, 5 の各類似度計算による評価値
Evaluation value of each method at scene no. 1 and
no. 5.

類似度演算法	シーン 1 の評価値	シーン 5 の評価値
H-HED	18.2	1.1
H-HHD	21.3	2.3
H-IHD	14.6	0.5
Y-HED	4.5	2.5
Y-HHD	5.8	1.3
Y-IHD	9.3	2.6
YH-HED	14.5	3.3
YH-HHD	13.2	1.7
YH-IHD	11.9	1.9

を示す。これは、リアルタイムで位置を検出しているところを任意の時間でフリーズし、ビデオプリンタに記録したものである。白枠が対象物を検出した位置を示す。

認識率は、30 秒のシーンで追尾が成功している時間の割合を示す。追尾が成功している定義は、白枠内に対象物があることとする。第二最小値/第一最小値は、各シーンの初期画面において図 6 の類似度表示から対象物テンプレートとそれ以外の領域を最もよく分離する類似度計算法を選ぶための評価値である。ここでは、この評価値が最大の計算方法を選択する。また、この評価法は対象物と背景との領域分割の容易さを示している。つまり、対象物と似ている色の物体が同じ画面内にあれば評価値は小さくなる。シーン 1 から 4 に関しては、うまく追尾が成功している。シーン 5 では、対象物である顔と背景のカーテンが似た色であり、評価値からも分離の難しさを示している。シーンが変化して蛍光灯が画面内に入り、やや逆光ぎみになり、なおかつ顔と同じ色と輝度のダンボールが画面内に現れたことにより、ダンボールを対象物と判断し誤認識しているのがわかる。シーン 1, 5 に関して、すべての類似度計算法に対する評価値(第二最小値/第一最小値)を表 2 に示す。これからもシーン 5 が対象物を抽出しにくいことがわかるが、この状況下でも最適な類似度計算法を選択することにより、少しでも誤認識をなくす配慮がなされている。

6. む す び

自然動画像中の対象物の連続的な位置検出の方式を提案し、リアルタイム実験を行った。本方式の特徴を以下に示す。

- (1) 簡単な対象物のポインティング。
- (2) 局所ヒストグラム成長法による対象物のセグメンテーション(背景との分離)。
- (3) 照明の変化、ビデオカメラのぶれ、対象物の回転、変形、縮小、拡大に強い象物の特徴量としての知覚的表色系のヒストグラム。
- (4) 任意の対象物、背景に対応するため、常に最適な画素情報、距離計算法を選び、マッチングを行う最適ヒストグラムマッチング。

本アルゴリズムにより、自然動画像中の対象物の位置を連続的に検出することが、高い確率で可能になった。今後は、(1)照明の色温度が急に大きく変化した場合、(2)照明の明るさが急に大きく変化した場合への対応が必要である。また、認識の性能を定量的に評価する手法を確立していく。

最後に、本研究の機会を与えて下さった CAV カンパニー商品開発研究所第 4 開発部 平井部長、高橋課長、中央研究所情報通信研究部門 羽根部門長、また、数々の助言を与えて下さった皆様に感謝致します。

〔参 考 文 献〕

- 1) 春木ほか：“画面分割を用いたディジタル AE, AF システム”，テレビ技報, TEBS 89-8(1989)
- 2) 安居院, 長尾, 斉藤, 中島：“ブロック特徴を用いた歩行人物の抽出と追跡”，テレビ誌, 45, 10, pp. 1213-1220(1991)
- 3) 尾上：“画像処理ハンドブック”，pp. 304-305, 昭晃堂
- 4) 宮脇, 石橋, 岸野：“色彩情報を用いたカラー画像の領域分割”，信学技報, IE9 89-50(1989)
- 5) 井上：“自然画像からの対象物抽出”，NHK 技研 R & D, 19(1992)
- 6) Vicki Bruce(著), 吉川左紀子(訳)：“顔の認知と情報処理”，pp. 59-64, サイエンス社(1990)



とみたか たかふみ
富高 忠房 昭和 58 年, 信州大学大学院工学研究科修士課程修了。同年, ソニー(株)に入社。VTR サーボ系の開発に従事。平成 1~2 年, カリフォルニア大学バークレー校 EECS 学科客員研究員。現在, ビデオカメラの画像処理, 認識, メカニズム制御の開発に従事。正会員。



せきや つお
関谷 庸男 平成 3 年, 法政大学工学部電気工学科(計測制御専攻)卒業。同年, ソニー(株)に入社。8 mm ビデオカメラの信号処理システムの開発に従事。



かげやま こうじ
景山 浩二 昭和 56 年, 京都大学理学部卒業。同年, ソニー(株)に入社。以来, 数値解析(主としてフィルタの設計解析)の研究, ならびにディジタル信号処理装置, 画像処理システムの研究開発に従事。現在, 中央研究所情報通信研究部門に所属。