

## 監視カメラによる人物挙動の変化検知

影広 達彦<sup>†</sup> 緒方 健人<sup>†</sup> 酒匂 裕<sup>††</sup> Josef Kittler<sup>†††</sup>

<sup>†</sup> (株) 日立製作所 日立研究所 〒319-1292 茨城県日立市大みか町七丁目1番1号

<sup>††</sup> (株) 日立製作所 中央研究所 〒185-8601 東京都国分寺市東恋ヶ窪 1-280

<sup>†††</sup> Centre for Vision, Speech, and Signal Processing, University of Surrey, UK

あらまし： 監視カメラによって人物の異常な挙動を検知するために、挙動認識システムの研究を行った。本システムは、人物挙動の変化を検知することで、正常/異常の判定を行う。人物を追跡するために、最近傍法と Particle filter 法の2つの手法を動的に切り替え、複数の人物が重なり合う場合でも安定した追跡を可能にした。また、学習判定機能に自己組織化マップを適用し、事前学習なしで挙動の変化検知が可能になった。本報告では、歩行状態を正常、乱闘状態を異常と定義し、CAVIAR データベースを元に精度評価を行った。その結果、フレーム単位の評価で、歩行状態に対する虚報率は 1.8%、乱闘状態の検知率は 64.2%となった。また、オプティカルフローの方向成分の平均値が、最も有効な特徴量であることが分かった。

## Detection of Human Behavior Changing by Surveillance Camera

Tatsuhiko Kagehiro<sup>†</sup>, Takehito Ogata<sup>†</sup>, Hiroshi Sako<sup>††</sup>, Josef Kittler<sup>†††</sup>

<sup>†</sup> Hitachi Research Laboratory, Hitachi, Ltd., Ibaraki 319-1292, JAPAN

<sup>††</sup> Central Research Laboratory, Hitachi, Ltd., Tokyo 185-8601, JAPAN

<sup>†††</sup> Centre for Vision, Speech, and Signal Processing, University of Surrey, UK

**Abstract:** We developed the human behavior recognition system by using surveillance camera. This system detects human behavior changing. I adapted the hybrid method which can change the nearest tracking method and particle filter method dynamically, so this system can track some humans who are overlapped. Because of Self Organizing feature Maps, this system can output the classification result without the learning step in advance. In this report, I defined that human walking condition is normal, and the human fighting condition is abnormal. As a result, the rate of false alarm in the walking data set is 1.8%, and the sensitivity rate in the fighting data set is 64.2% in all frames. I found that the most effective feature is the average of the optical flow direction (sine and cosine).

## 1. はじめに

様々なテロや凶悪犯罪が発生する昨今、監視カメラによる犯罪予兆の検知や、迅速な犯人検挙の期待が高まっている。英国では、数百万台のカメラが公共の場に配置されており、2005年7月に発生したロンドンテロ事件では、地下鉄駅に配置された監視カメラの動画を人手で解析することで、素早く実行犯を特定することが出来た。このように、監視カメラから得られる動画は、犯人特定のために非常に有用な情報となる。ただし、監視カメラで撮像された動画の解析は、事件発生後に人手で行われることが多く、莫大なコストが掛かる。そのため、監視カメラ自体が、実行犯への抑止力のみとして期待されている場合が多い。もし、監視カメラの動画から人物の異常な挙動が自動検知出来れば、犯罪の予兆を警告することができ、効果的な監視システムの使い方を提示できると考えられる。

動画から人物の挙動を認識する研究は盛んになっており、特に監視カメラ画像を用いた研究は多数報告されている[1][2][3]。また、TV放映用に撮影されたスポーツ映像に対し、人物の挙動を基に自動的に分類する研究も行われている[4]。また、挙動認識のためには、動画中の人物の追跡が、大きな課題となっており、多くの報告がされている[5][6]。

本研究は、監視カメラから得られた動画を入力として、画像中の人物を追跡し、それぞれの人物の挙動を認識することを目標とする。ただし、人物挙動を、歩行、乱闘、走りなどに分類するのではなく、正常/異常の判定のために、挙動変化の検知を実現した。通常、認識処理には学習ステップが必要であるが、大きなコストが掛かる。そこで、予め教師付き信号による学習を行わず、入力された挙動特徴を徐々に自動学習し、異常な挙動を検知出来るシステムの構築を目指す。

## 2. 挙動の定義

本研究では、処理コストや認識結果の安定性を考慮し、カメラが単数かつ固定の監視カメラで撮像された動画を入力とする。また、認識対象となる移動物体は人物とし、人物の挙動の変化を検知して、正常/異常の

判別をすることを目的とする。下記に、本報告における人物挙動の正常/異常の定義を述べる。

正常：人物が動作を繰り返している。

異常：人物が当初の動作と異なる動作へ急に変更する。

上記の定義を元にとすると、画像中に歩いて進入してきた人物が、歩いて画像外に出て行った場合は、全て正常挙動であったとする。また、歩いて進入してきた人物が、途中で異なる挙動(乱闘など)を起こした場合、異常挙動であったとする。

人物の異常挙動は、ある瞬間に生じて終わるのではなく、乱闘などのように一定時間継続される。この一定時間中の一部でも異常挙動を検出できれば、犯罪予兆の検知や犯人検挙には有効である。

## 3. 本研究の課題

パターン認識においては、予め学習データを用意し、事前に学習を行い、パラメータ調整を行うことが多い。しかし、データ収集のコストや、カメラ設置状況が異なる場合のパラメータのロバスト性が問題になる。本研究では、システムは事前知識を持たず、処理経過で挙動特徴を学習し、正常/異常の判定を行うことを目標とする。これにより、監視カメラ設置のカスタマイズやメンテナンスコストが軽減され、容易な運用が可能になる。

挙動認識を実現するには、下記の課題が考えられる。

- A) 人物領域の抽出と追跡
- B) 挙動を分類するための判別器の選定
- C) 挙動分類に有効な特徴量の選定

上記のうち(A)に関しては、各時点で最適な背景画像を生成更新し入力画像と差分を取ることで、人物領域を抽出できる。また、複数の人物が同時に移動する場合、オクルージョン問題が発生する可能性が高いため、多少の隠れに対応できる追跡アルゴリズムを採用する必要がある。

(B) に関しては、事前知識無しで設置稼動するために、学習ステップを必要とせず入力信号の特徴をそのまま学習し判別できる手法が必要である。この機能を持つ判別器の1つに、自己組織化マップ (Self Organizing feature Map : SOM) がある。本研究では、SOMを採用し挙動認識システムを構築する。

(C) に関しては、有効な挙動特徴を反映させる入力ベクトルを選択する必要がある。そこで、CAVIAR データベース [8] を用いて定量的な精度評価を行い、虚報率が低く検知率の高い結果が得られる入力ベクトルを選択する。

## 4. 挙動認識システム

### 4.1. システム概要

本節では、挙動認識システムの概要について述べる。挙動認識システムは、監視カメラから得られるカラー動画像を入力とする。出力は、人物の挙動に対する正常/異常の判別結果である。

挙動認識システムは、図1で示すように、背景画像生成更新、移動物体抽出追跡、オプティカルフロー生成、挙動分類の4つのモジュールで構成されている。また、移動物体追跡結果を格納する追跡マップを持つ。

挙動認識システムは、入力された動画像から、最適な背景画像を生成更新し、背景画像を差し引くことで人物領域のみの画像を生成し、人物の追跡とオプティカルフローの算出を行う。人物の追跡結果の履歴は追

跡マップに書き込まれ、人物同士の相対的な位置関係のチェックに用いる。得られたオプティカルフローデータを挙動の特徴とし、挙動分類処理を行う。

### 4.2. 背景画像の動的更新

背景画像と入力画像の差異を観測することで、動画中の移動物体領域を抽出することができる。監視カメラでは、照明条件の変化に追従して感度特性が変化するため、静的な背景画像を適用できない。そのため、本システムでは、入力された動画から背景画像を自動生成し、フレームごとに動的更新を行う [9]。

まず、入力画像の複数フレームの平均画像を生成し、初期背景画像とする。背景画像の更新には、2つの情報を適用する。(A) フレーム間差分の情報、(B) 背景画像と入力画像の差分情報である。情報 (A) から移動物体の領域が得られ、情報 (B) より一時的に停止している人物領域が得られる。まず、情報 (A) による差分画像と、情報 (B) による差分画像を生成する。各差分画像において、各画素の近傍領域の輝度値合計を、背景更新の指標とする。この指標を、式(1)中の  $C_f$  と  $C_m$  に代入し、背景更新の度合いを決定する。 $C_f$  は、学習データのフレーム間差分によって得られる差分値の頻度分布を元に作成する。 $C_m$  は、学習データの背景差分によって得られる差分値の頻度分布を元に作成する。本手法は、近傍領域の情報を元に背景画像の更新度合いを決定するため、局所的なノイズによる悪影響を抑えることができる。

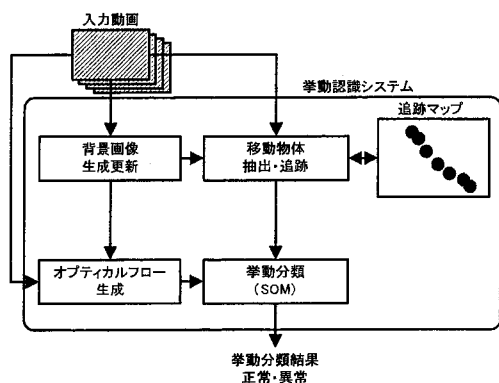


図 1: 挙動認識システム

$$B_{t+1} = B_t + \frac{C_f(S_f) + C_m(S_m)}{2} (I_t - B_t) \dots (1)$$

$B$ : 背景画像輝度値

$I$ : 入力画像輝度値

$C_f$ : フレーム間差分確信度

$C_m$ : 背景入力差分確信度

$S_f$ : フレーム間差分画像における近傍画素合計値

$S_m$ : 背景入力差分画像における近傍画素合計値

### 4.3. オプティカルフロー生成

画像中の挙動特徴を得るために、オプティカルフローを算出する。ただし、画面全体のオプティカルフローを算出すると、計算コストの増大や人物領域以外の外乱による誤差が発生するため、背景画像と入力画像の差分によって得られる人物領域のみにおいて、オプティカルフローを算出する。

オプティカルフローの算出には、大きく分けて勾配法とブロックマッチング法があるが、本報告では勾配法を適用した。勾配法は、“隣接するフレーム間では物体上の点の明るさは移動後も変化しない”という仮定から、時空間微分とオプティカルフローとの関係式を導出し、それを元に対象の動きを推定する。

### 4.4. 移動物体追跡

移動物体の追跡は、動画像処理の大きな課題であり、本システムにおいても重要な機能となっている。人物追跡では、複数の人物を同時に追跡した場合、隠れが生じる可能性がある。特に乱闘状態では、複数の人物が激しく動き、かつ、互いに絡み合う。そのため、隠れが生じる可能性が高い。隠れ対策に有効な方式として、Particle filter法が報告されている[6]。Particle filter法は、複数の仮説を立て、確率的に追跡対象物の位置やサイズを決定する。そのため、短時間、対象物全体が観測不能になった場合でも、再度出現する位置を仮説として生成し、追跡することが可能である。

しかし、Particle filter法にも複数の欠点がある。まず1つ目は、多数の仮説を立て探索を行うため、計算コストが大きい。また、論文[6]で報告された手法では、色情報を元にして追跡するため、追跡対象物の色の見え方が変化した場合、誤差が累積する場合がある。この問題に対し、追跡結果を元にテンプレートを随時アップデートする手段も存在するが、追跡誤差をテンプレートに織り込む危険性がある。

これらの問題を解決するため、本システムの移動物体追跡は、2つ手法を組み合わせ適用した。1つは、背景差分により得られた物体領域の最近傍をフレームごと一致させ追跡する最近傍法、もう1つは

表 1：最近傍法と Particle filter 法の利点と欠点

	利点	欠点
最近傍法	処理コスト小 誤差累積無し	隠れに弱い
Particle filter 法	隠れに強い	処理コスト大 誤差累積あり

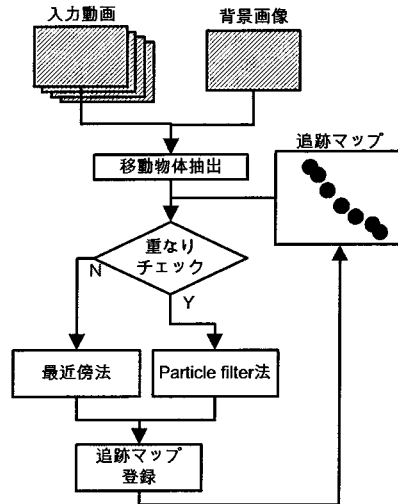


図 2：人物追跡手法

Particle filter法である。この2つの手法の利点と欠点を表1示す。表1を参照すると、最近傍法はオーソドックスなアプローチであるため、計算コストが小さい。また、人物領域の対応が取れれば、領域の誤差累積は発生しない。しかし、複数人物が接近し隠れが発生すると、人物領域の対応付けが破綻する。Particle filter法は、部分的な隠れに対しロバストであるが、計算コストが大きいことと、対象人物の見え方が変化することによる誤差の累積が問題となる。

表1の利点欠点を考慮し、本報告では、図2に示すように2つの手法を動的に切り替えて適用する。人物が個々に離れて移動している場合は、最近傍法を適用し、人物同士が接近し重なり隠れが発生する可能性がある場合は、Particle filter法に切り替える。Particle filter法で追跡中に、人物が個々に離れて移動し始めた場合、最近傍法に切り替えて追跡を続行する。手法の切り替えは、図1に示した追跡マップを観測し追跡

中の人物同士の相対的距離を測定することで実現できる。このように、特性の異なる手法を組み合わせることで、高速でロバスト性の高い複数人物の追跡が可能となる。

#### 4.5. 挙動分類モジュール

挙動を分類するために、オプティカルフローデータから特徴ベクトルを生成し、判別器へ入力する。オプティカルフローデータを特徴ベクトルとして用いる際に、下記のような課題が存在する。

- (1) 各画素にはノイズ成分が存在する。
- (2) 特徴ベクトルの次元数が大きすぎると、判別に悪影響が生じる。
- (3) 短時間の動作は、挙動の本質的な特徴を表現していない。

上記のうち(1)と(2)に関しては、人物領域内の平均値、標準偏差などを算出し、統計処理によりノイズを緩和し、かつ、次元数を減らすことで解決できる。(3)に関しては、フレームごとに人物の挙動を観測した場合、細かく特徴が変動して、大局的な挙動特徴を観測し難い。そこで、ある一定フレーム間隔で観測窓を設け、時間軸方向の平均を算出する。つまり、人物領域内の統計値を算出後、時間軸方向に平均値を取り、特徴ベクトルとして適用した。これにより、フレーム単位の細かい変動が吸収され、人物挙動の大局的な特徴が観測可能になる。

得られた特徴ベクトルを、自己組織化マップアルゴリズム (Self Organizing feature Map : SOM) に入力し、学習、判定を行う[7]。SOM は、入力ベクトルに応じて、特徴マップ上の各ニューロンと連結されたアーク上の重みを決定し、特徴マップ上の各ニューロンの誤差を決定する。誤差を持つ各ニューロンの分布が、入力ベクトルの特徴を表現する。各ニューロンのうち、誤差最小のものを勝ちニューロンと呼び、その位置が分類結果として重要な情報となる。

SOM は、初期状態で、特徴マップ上の各ニューロンと

表 2 : 人物ラベル表

ラベル	動画	挙動
W 1	歩行 動画 1	歩行者 1 名が右下部から左上部へ
W 2	歩行 動画 2	歩行者 1 名が右下部から右上部へ
W 3		歩行者 1 名が右上部から左上部へ
F 1	乱闘 動画 1	歩行者 2 名が遭遇→乱闘→1 名が左上部へ逃げる
F 2		1 名が左下部へ逃げる
F 3	乱闘 動画 2	歩行者 2 名が遭遇→乱闘→1 名が倒れる
F 4		1 名が逃げる



図 3 : 歩行者動画 1 (W1)



図 4 : 歩行者動画 2 (W2)

アーク上に乱数が配置されており、ベクトルを入力するにつれて学習してアーク上の重み変動し、特徴マップ上の各ニューロンの誤差が決定される。つまり、類似した特徴ベクトルが入力された後、異なる傾向の特徴ベクトルが入力されると、マップ上の傾向は変化し、勝ちニューロンの位置は移動する。この勝ちニューロン位置の移動を検知すれば、挙動の変化を捕らえることが出来る。また、徐々に変化する特徴ベクトル



図 5: 歩行者動画 2 (W3)



図 6: 乱闘動画 1 (F1,F2)



図 7: 乱闘動画 2 (F3,F4)

に対しては、各アーク上の重みが徐々に調整され、勝ちニューロンの位置は変化しない。これにより、人物の歩行状態の揺らぎは吸収され、乱闘による急激な変化は検知することが可能になる。

## 5. 評価実験

### 5.1. 入力データと実験条件

本システムを評価するために、CAVIAR データベース [8]に含まれる「歩行」と「乱闘」の動画を用いた。動

表 3: 評価フレーム数

	歩行	乱闘	その他
W1	230	0	0
W2	128	0	0
W3	98	0	0
F1	98	61	66
F2	50	56	58
F3	54	126	75
F4	64	138	43

表 4: 正誤表

		正解挙動		
		歩行	乱闘	その他
判別結果	歩行	True	False	Unknown
	乱闘	False	True	Unknown
		Negative	Negative	
		Positive	Positive	

表 5: 各ベクトルの統計値

略称	統計値
xav	オブティカルフローX 軸成分の平均値
xstd	オブティカルフローX 軸成分の標準偏差
yav	オブティカルフローY 軸成分の平均値
ystd	オブティカルフローY 軸成分の標準偏差
cosav	オブティカルフローcos 値の平均値
cosstd	オブティカルフローcos 値の標準偏差
sinav	オブティカルフローsin 値の平均値
sinstd	オブティカルフローsin 値の標準偏差
cx	人物領域重心の X 方向移動距離
cy	人物領域重心の Y 方向移動距離

画データは、RGB フルカラー、画像サイズは 320×240、フレームレートは 25fps である。CAVIAR データベースには、多数の挙動種の動画データがあり、各動画には人物に対するフレームごとの位置と挙動種の正解が付加されている。この正解を参照し、本システムの判定結果と比較することで、挙動認識結果の精度評価ができる。評価に用いた人物のラベルを表 2 に示し、各動画の 1 フレームを図 3 から図 7 に示す。

評価対象人物の W1, W2, W3, F1, F2, F3, F4 それぞれの挙動種ごとの評価フレーム数を、表 3 に示す。各人物のフレームごとの挙動判定結果を、CAVIAR データベースの正解と比較した。歩行、乱闘以外のフレームは、本実験では無視した。

本システムは、予め教師付きデータを用いた学習を



表 6 : 各ベクトルの評価結果

入力ベクトル	FP (%) (W1, W2, W3)	TP (%) (F1, F2, F3, F4)
xav xstd yav ystd cosav cosstd sinav sinstd cx cy	2.9	0
xav xstd yav ystd cosav cosstd sinav sinstd	0	0
xav xstd yav ystd cx cy	14.9	23.6
xav/cx yav/cy cosav sinav	8.9	22.1
xav/(cx+cy) yav/(cx+cy) cosav sinav	1.8	51.8
xav/(cx+cy) xstd yav/(cx+cy) ystd cosav cosstd sinav sinstd	0	0
xav/(cx+cy) yav/(cx+cy)	6.7	0
cosav sinav	1.8	64.2
cosav cosstd sinav sinstd	0	0

行わず、徐々に自動学習を行う。そのため、出力された認識結果と挙動種の対応を自動的に取ることが出来ない。そこで、人物が画像中に入ってきた際、必ず歩行していると仮定し、初期認識結果を歩行と設定し、その後、出力状態に変化があった場合を乱闘と定義した。つまり、W1, W2, W3 に対しては、初期認識結果の状態から変化無く画像から出て行くことが完全正解である。また、F1, F2, F3, F4 に対しては、初期認識結果の状態から、途中乱闘によって出力結果に変化が生じることが理想的である。そこで、下記の表 4 のように正誤表を定義し、判別結果の精度を測定する。表 4 中の False Positive (FP) と True Positive (TP) が、精度を評価する上で分かり易い尺度となり、FP が低く TP が高いほど、精度のよいシステムと考えられる。そこで、上記 2 つの尺度を用い、本システムの精度評価を行い、効果的な特徴量の選定を行った。

## 5.2. 特徴量の評価結果

SOM に入力される特徴量は、人物領域内の全オプティカルフローの統計値を用いる。オプティカルフローデータは 2 次元ベクトルの集合であり平均値と標準偏差を算出すると、表 5 のような特徴量が得られる。オプティカルフローの統計値に加えて、人物領域の重心位置の移動距離も特徴量として加えた。これらの特徴量を組み合わせて特徴ベクトルとし SOM に入力して、挙動認識結果の精度を測定した。

特徴量の統計値の組み合わせを変え、FP と TP を測定した。表 6 に評価実験の一覧と FP, TP の算出結果を示

す。FP は W1, W2, W3 を対象とした個々の実験結果の平均値、TP は F1, F2, F3, F4 を対象とした個々の実験結果の平均値を示している。

## 6. 考察

表 6 の評価結果から、最も FP が低く TP が高いのは、特徴量  $\cosav$ ,  $\sinav$  を組み合わせた特徴ベクトルであることが分かる。2 番目に好結果なのは、特徴量  $xav/(cx+cy)$ ,  $yav/(cx+cy)$ ,  $\cosav$ ,  $\sinav$  を組み合わせた特徴ベクトルである。標準偏差を特徴量に組み込んだ実験結果は比較的悪く、人物領域内のオプティカルフローのばらつき情報は、本システムの挙動認識の特徴として適していないことが分かる。最も結果の良かった特徴ベクトルの要素である  $\cosav$ ,  $\sinav$  は、2 番目に良い結果の特徴ベクトルにも含まれており、本システムにおいて有力な挙動特徴であることが分かる。

つまり、一定に歩いている際は、オプティカルフローの方向成分は、ほぼ一定であり、方向に変化があっても滑らかである。しかし、乱闘状態が生じると、オプティカルフローの方向成分は大きく変動し始め、これを SOM が異なる状態と判定する。また、SOM の特徴マップのサイズに対して、精度の変化はほとんど無かった。

最も結果の良かった特徴ベクトルである  $\cosav$ ,  $\sinav$  の結果を参照すると、歩行状態の動画に対し、FP が 1.8% であり虚報が少ないため、システム運用時に無駄な人的コストを抑えることが可能なことが分かる。乱闘状態の動画に対しては、フレーム単位の TP は 64.2%

となっている。これは、乱闘中の全フレームの半分以上を検知している。また、人物単位の評価では、全ての人物の乱闘を検知しており、実用上有効であると考えられる。

## 7. まとめ

監視カメラ画像を入力とした挙動認識システムの研究を行った。本システムは、予め学習データを用意することなく、人物の挙動を自動学習し、異常な挙動を検知することができる。また、特性の異なる追跡アルゴリズムを組み合わせることで、重なりが生じる複数の人物を安定して追跡できた。

本報告では、人物の歩行状態を正常、乱闘状態を異常と定義し、精度評価を行い、有効な特徴量を調査した。その結果、オプティカルフローの方向成分の平均値が、有力な特徴であることが分かった。上記特徴を用いると、フレーム単位の評価で、歩行状態の動画中の虚報率が 1.8%、乱闘状態の動画中の検出感度が 64.2%となった。本評価はフレーム単位であり、乱闘中の各人物は全て検知可能であった。

本挙動認識システムは、少人数の人物に対する挙動認識を試みており、群衆の挙動に関しては正確な判定は困難であると考えられる。また、精度評価として歩行と乱闘の2種類の挙動判別を行っており、他の挙動に対する判別能力は未知である。そこで、今後の課題として、他の挙動種に対する反応を調査し、更なる判定精度の向上を図る。また、他の動画データを本システムに入力し、実用上の課題について検討を行う。

## 8. 参考文献

- [1]. T. Nanri, N. Otsu, "Unsupervised Abnormality Detection in Video Surveillance", Proc. MVA2005, pp. 574-577.
- [2]. D. Tweed, W. Feng, R. Fisher, J. Bins, T. List, "Exploring techniques for behavior recognition via the CAVIAR module vision framework", International Workshop on Human Activity Recognition and Modeling, pp.97-104, September, 2005.
- [3]. J. W. Davis, M. A. Keck, "Modeling Behavior

Trend and Detecting Abnormal Events using Seasonal Kalman Filters", International Workshop on Human Activity Recognition and Modeling, pp. 9-16, September, 2005.

[4]. F. Cheng, "Human Motion Description in Multimedia Database", Ph.D Thesis, CVSSP, University of Surrey, 2004.

[5]. P. KaewTrakulPong, R. Bowden, "A real time adaptive visual surveillance system for tracking low-resolution color targets in dynamically changing scenes", Image and Vision Computing 21, pp. 913-929, 2003.

[6]. K Nummiaro, E. Koller-Meier, L. V. Gool, "An adaptive color-based particle filter", Image and Vision Computing 21, pp. 99-110, 2003.

[7]. 萩原将文, "ニューロ・ファジィ・遺伝的アルゴリズム", 産業図書, 1994年初版

[8]. <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/caviar.htm>

[9]. 影広達彦, 大田友一 "動画像からの背景画像の自動生成と適応的更新", 画像の認識・理解シンポジウム MIRU'94, II, pp.263-270.