

人物追跡画像に基づく服装の実時間分類

郷原 邦男 馬場 功淳 江島 俊朗

九州工業大学

概要 多彩な行動をとる人物の情報を収集することは非常に興味深いため、実時間による人物追跡が盛んに行われている。特に、服装の情報を利用することは同一人物の判別や追跡に重要な手がかりになると考えられる。しかし、服の色や模様を分類する処理は光の当たり具合や形が一定でないことから処理が難しいものとされてきた。本論文では服の領域抽出法、少々の光の変化や服の動きに捕らわれない色の分類法、模様の分類法、形状の分類法、そして Probabilistic Voting について説明し、最後に動画像中に現れる服装の実時間抽出および分類結果を示す。

Real-time Clothes Classifier based on person tracking image

Kunio GOHARA Naruatsu BABA Toshiaki EJIMA

Kyushu Institute of Technology

Abstract Since, it is interesting to observe and collect the person activity. Many real-time person tracking systems have been studied. To make the system more intelligent, it is important to classify clothes that people wear. However, due to plasticity of clothes and variation of lighting condition, classification of clothes is not so easy. In this paper, we have proposed not only a real-time clothes segmentation from input images, which is based on statistical technique called discrimination analyses, but also a real-time mapping of segmented region to ten representative colors. This technique leads to a system in which to specify or to retrieve the people, simple color and simple texture information about clothes can be used.

1 はじめに

多彩な行動や表情をとる人物の情報を収集することは非常に興味深いため、実時間による人物追跡が盛んに行われている。特に顔の表情や向きの推定は多く行われ、人物の動きや個人の識別なども行われてきた。本研究は頭部と共に人物追跡に有用と思われる服装の実時間分類を行う手法を提案する。服の色や模様の分類は光の当たり具合による色の変化や人の動きによる形状の変化の為、処理が難しいものとされてきた。本論文では少々の光の変化や服の動きに因わず服装の実時間分類をロバストに行える手法を提案する。

本論文では色、模様、型に関して分類を行う。つまり図1のような入力に対して表のような出力を行うシステムを構築する。

実世界での画像処理は、常に変化する環境を考慮したロバストな処理方法が必要となる。そこでロバストに服装を分類する為の手順を図2に示す。動画像処理は大きく分けて1フレーム毎の画像空間と、その画像空間が集まった時系列で考える事ができる。

人間が見ると同じ色や模様をした服装でも、画像空間中では各フレームでその色合いや影が時々



模様	色	型
ストライプ	黒、黄色	長袖

図1: 服装の分類結果

刻々と変化し、図2の同一の服装の時間的变化に示されている様に、同一の服装でも画像空間中の一点に留まっている訳ではない。その為、服装の色、模様、袖をロバストに分類できる処理(1, 2, 3)が必要となる。ここで1は服装の色をロバストに分類する手法で本研究では3章で説明する一次元ヒストグラム法[1]を用いた。2は服装の模様をロバストに分類する手法で本研究では4章で説明するフーリエ変換を適用した。また、3はロバストに腕を追跡し半袖と長袖を分類する手法で、5章で説明する。

これらのロバストな手法を用い服装の分類を行っても、色や模様が誤って分類されてしまう場合がしばしばある。しかし、カメラ映像からの連続したフレームは同一の服装を捉えている可能性が高いことから過去 N フレームを利用することで真の値を推定する Probabilistic Voting を提案した。

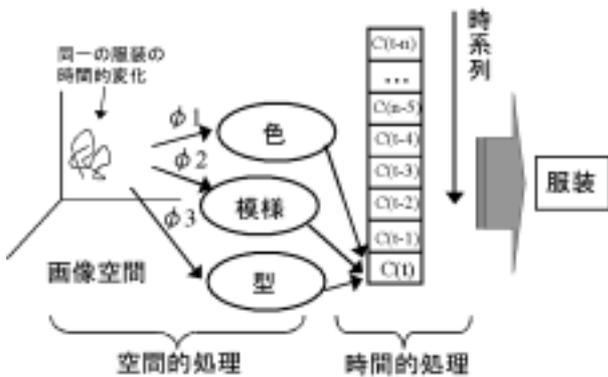


図 2: 服装の実時間分類を行う処理の流れ

2 服の抽出

服装の分類にあたって、動画像から服装領域の切り出しを行う手法が必要となる。そこで我々は頭部を実時間でロバストに追従できる HeadFinder [2] を利用し、頭部の下には必ず服装が存在するという仮定のもと、服装の抽出を行った。抽出矩形の大きさは縦が頭部直径の 2 倍、横が頭部直径と同じになるように設定した。服装の抽出した結果を図 4(a) に示す。

3 領域のセグメンテーション

3.1 表色系について

ビデオカメラからの出力である RGB 表色系では色の分布の偏りが激しく、物体抽出にはあまり向いていない。そのため色彩 T 、彩度 S 、明度 L の値で色空間を表現する TSL 表色系 [3] も同時に用いた。この表色系は色彩が独立した軸のため「赤い物体を取り出せ」といった、色による物体抽出により適した表色系であるといえる。RGB 表色系を TSL 表色系に変換する式を (1) に示す。

$$\left. \begin{aligned} T &= \tan^{-1}(r'/g')/2\pi + 1/4 & (g' > 0) \\ T &= \tan^{-1}(r'/g')/2\pi + 3/4 & (g' < 0) \\ T &= 0 & (g' = 0) \\ S &= \sqrt{9/5(r'^2 + g'^2)} \\ L &= 0.299R + 0.587G + 0.114B \end{aligned} \right\} (1)$$

ただし、 g' と r' は以下の式 (2) で正規化した値である。

$$\left. \begin{aligned} r' &= R/(R + G + B) - 1/3 \\ g' &= G/(R + G + B) - 1/3 \end{aligned} \right\} (2)$$

我々は服装を最終的に表 1 で示した代表的な 10 色で表すことにした。本研究では色を決定する手法として「最近隣法」と「一次元ヒストグラム法」について、比較検討した。

表 1: 代表的とした 10 色

無彩色	黒・明るい灰・暗い灰・白
有彩色	青・水色・赤・黄・紫・緑

3.2 最近隣法

色の分類を行う手法として各画素をあらかじめ決められた閾値により、各色クラスに割り当てていく手法が考えられる。この手法での色空間のイメージは図 3 の様になる。これはある表色系で赤青緑を分類する閾値のイメージである。このような単純な方法を最近隣法とする。

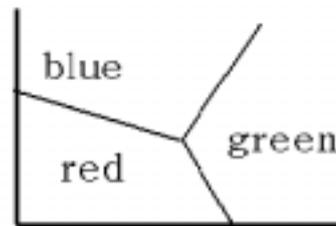


図 3: 固定された色の閾値

この手法により得た結果が図 4(b) である。

また、あらかじめ決めた閾値を表 2 に示す。この閾値を元に、最も近い色を決めるアルゴリズムを以下に示す。ここで P は画素、 C は有彩色、 G は無彩色とする。

1. Pixel の RGB から $T(P)$ 、 $S(P)$ 、 $L(P)$ を計算する。
2. $S(P) >$ 閾値なら 3 へ、そうでなければ 4 へ。
3. $T(P)$ と $T(C)$ で比較し近い色を割り当てる。
4. $L(P)$ と $L(G)$ で比較し近い色を割り当てる。

結果、図 4(b) から分かるように影の影響を受けていることが分かる。これは、全体の色分布を考慮せず、画素ごとに色を割り当てた為といえる。

表 2: 代表色の固定閾値

有彩色	緑	黄	赤
閾値 T	32-70	71-120	121-159
有彩色	紫	青	水
閾値 T	160-197	198-248	249-31

注)T は 0 ~ 255 で正規化

無彩色	黒	明るい灰	暗い灰	白
閾値 L	0-32	33-96	97-192	193-255



(a) 元画像 (b) 最近隣法

図 4: 固定閾値による分割

3.3 一次元ヒストグラム法

分割の段階で色の閾値を使わない手法として 1 次元ヒストグラムに基づく領域法が提案されている。この手法は、様々な色空間から領域を分割し、領域ができた後に色を割り当てる手法である。一次元ヒストグラム法のアルゴリズムを以下に示す。

1. 元画像から RGB、TSL のヒストグラムを生成。
2. 判別分析法により分割に適したヒストグラムを探す。
3. 判別分析法で求めた閾値で 2 により求めたヒストグラムを二分割する。
4. 分割されたそれぞれの領域に対して、1 からの処理を再帰的に行う。

本研究では RGB と TSL の二つの表色系を利用した。表色系を増やすことにより、多くの特徴空間から領域を分割することが可能である。図 5 では本手法を RGB 表色系のみで実行した結果と、RGB と TSL を使い実行した結果を比較している。RGB 表色系のみだと特徴量が少なすぎ服装の分類に適した分割ができていない。これに対し、RGB と TSL 表色系を用いると影に影響されず服装の分類に適したセグメンテーションがなされている。その反面、多くの表色系を利用することで処理コストは比例して増す。また、表色系を増やす際にお互いに独立な色空間を選択しなければ意味がない。RGB と TSL 表色系が非線形な関係であることにより、異なった特徴空間を増やすことで良い結果を得ることができた(式 1 参照)。またヒストグラムに

よって領域を分割するため、ヒストグラムの形にさほど影響しないような、影や一部分の模様を排除することに成功した。よって本研究では一次元ヒストグラム法を用いることにした。

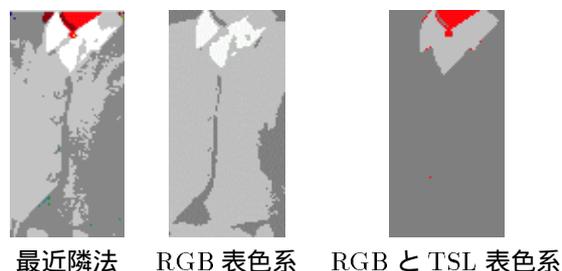


図 5: 一次元ヒストグラム法による分割

3.4 一次元ヒストグラム法の優位性

3.2 節と 3.3 節から一次元ヒストグラム法が最近隣法に比べて影などの影響を受けにくく優れていることを示した。このように一次元ヒストグラム法が優れている理由として、分割の段階で色の閾値を固定していないことが挙げられる。

最近隣法ではあらかじめ決められた閾値により色が決められる。しかし、一次元ヒストグラム法では全体の色の分布により領域を分ける閾値が動的に変化する。領域の分割例を図 6 に示す。この図から分かるように最近隣法ではたとえ一つの領域内で小さい色の変化(影によるグラデーションなど)があった場合でも、そこが閾値として定められていたら二つの領域として分類されてしまう。しかし、一次元ヒストグラム法では影などによるグラデーションの変化はヒストグラム上で一つの山となり、領域の分割は行われない。その為、影などの小さな変化に影響されず、安定した色の分類ができる。

色空間的には最近隣法が図 3 の様な決められた閾値に対して、一次元ヒストグラム法は図 7 の様に色の分割面を動的に変化させることのできる手法である。

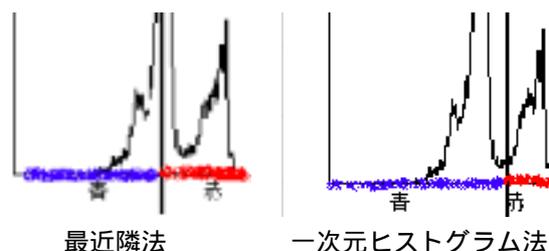


図 6: 各手法による色分類の閾値の違い

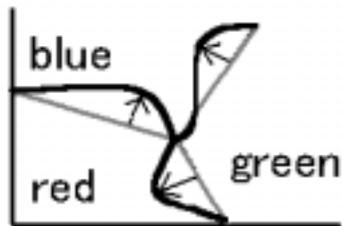


図 7: 色空間内での分割面の変化

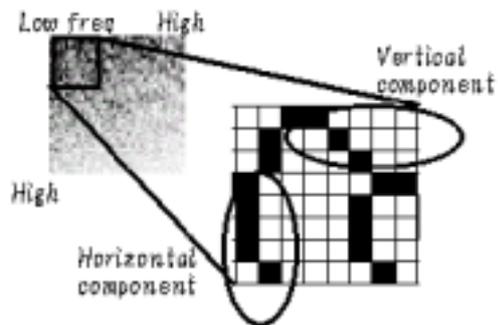


図 8: 周波数成分と分類の関係

4 模様分類

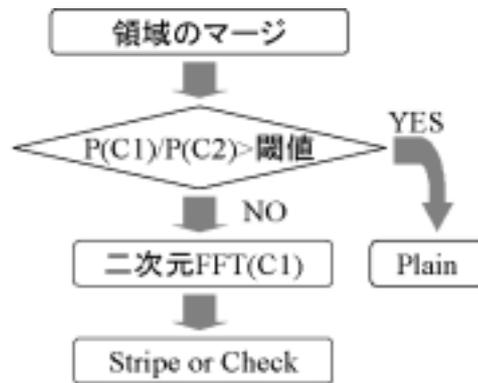
4.1 模様について

服装の模様を「プレーン」「縦ストライプ」「横ストライプ」「チェック」の4つのクラスに分類したい。しかし、服装の模様となる幅は一様でなく、また必ずしも直線ではなくゆがんでいる場合が多くある。その為、パターンマッチングのような、あらかじめ決められたパターンと比べる手法では実現が難しいと考えられる。

4.2 高速フーリエ変換

画像を周波数分析することにより、垂直および水平方向の成分をロバストに得ることができる。横と縦のストライプはそれぞれ垂直方向と水平方向の成分から判別する。また、各々の成分が現れた時チェックとして判別する。利用する周波数成分は図8に示すように、低周波成分の垂直成分と水平成分の一部とした。これは高周波成分がノイズと区別できない為である。なお、1枚の画像内で同じ色の領域が70%以上占めている時プレーンとした。

模様を分類するアルゴリズムを図9に示す。ここで領域のマージというのは、前章で分割された領域に対し、同じ代表色を割り当てられた領域を1つの領域にマージする処理である。図10の場合、白に割り当てられた領域を1とし、灰色に割り当てられた領域を2とすることで、少ない領域数により、横ストライプという特徴を表している。図10の場合も含め、服装の代表的な模様というのは視覚的に最も多い色で構成される場合が多い為、フーリエ変換は最も多い代表色、つまり最も大きい領域に対してのみ処理した。



C1:最も大きい領域
C2: 2番目に大きい領域
P(C):領域Cのピクセル数

図 9: フーリエ変換までの処理

4.3 Probabilistic Voting

ノイズ削減のために voting による推定を行う。少ない投票数で効率良くノイズの削減を行うため、システム特有の誤り確率等も考慮した Probabilistic Voting (以下 P-voting) を提案する。

ある模様 α_i が入力された時の出力 β とする。ある出力 β に対して、入力された模様 α_i を推測するには $P(\alpha_i|\beta)$ が最大であるものを選択すればよい。過去 n フレームを利用し現在入力されている模様 α_i を推測するには $P(\alpha_i|\beta(t), \beta(t-1), \beta(t-2), \dots, \beta(t-n))$ が最大のものを選べば良い。ここで $\beta(t)$ を時間 t で得られる出力とする。本手法を計算する式を以下に示す。

$$P(\alpha_i|\beta(t), \beta(t-2), \dots, \beta(t-n)) = \frac{P(\beta(t), \beta(t-2), \dots, \beta(t-n)|\alpha_i)P(\alpha_i)}{P(\beta(t), \beta(t-2), \dots, \beta(t-n))} \quad (3)$$

ここで分母は α_i によって変わらず、 $P(\alpha_i)$ は各模様が等確率で出現すると仮定でき、 $1/4$ の定数となるため省略できる。また、分母も入力に対し不変であるので省略できる。これらを考慮し、上式の対数をとると、

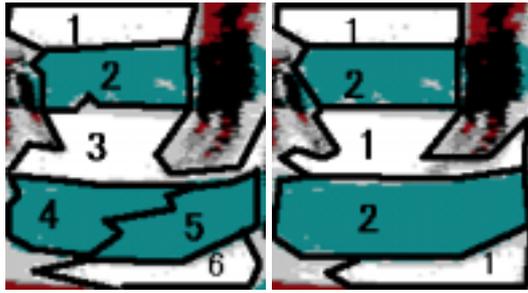


図 10: 領域のマージ

$$\begin{aligned}
 L_{\alpha_i} &= \log \prod_{j=0}^n P(\beta(t-j)|\alpha_i) \\
 &= \sum_{j=0}^n \log P(\beta(t-j)|\alpha_i) \quad (4)
 \end{aligned}$$

で求められる。ここで $P(\beta(t-j)|\alpha_i)$ はあらかじめシステム特有の事後確率として表 (3) のように実験的に求めておくことができる。式 (4) により求めた $L_{\alpha_{plain}}$ 、 $L_{\alpha_{h-stripe}}$ 、 $L_{\alpha_{v-stripe}}$ 、 $L_{\alpha_{check}}$ についてもっとも大きな確率を、入力された模様として選択する。

表 3: システム特有の事後確率

	β_{plain}	$\beta_{v-stripe}$	$\beta_{h-stripe}$	β_{check}
α_{plain}	0.878	0.019	0.007	0.001
$\alpha_{v-stripe}$	0.030	0.830	0.001	0.140
$\alpha_{h-stripe}$	0.011	0.001	0.833	0.156
α_{check}	0.136	0.010	0.239	0.615

5 服装の型分類

5.1 長袖と半袖の分類について

服装の型分類として長袖と半袖の分類を行った。袖は服装の胴体部分と違い必ず頭の下にあるような仮定は成り立たない。また、袖は胴体の後ろに隠れる場合も想定されるため、オクルージョンが発生することを考慮しなければならない。本論文では、ロバストに腕を検出し、その情報を基に長袖と半袖を分類する手法を提案する。

5.2 色空間での肌色の領域推定

腕を抽出するために肌色を検出することは有用である。しかし、光源の違いや人種の違い、影の発生により肌色も多様に変ってくる。そこで、ロ

バストな肌色検出手法として肌色の基準色を設け、その基準色に近い色を肌色として切り出す手法を用いた。

基準色の決定には図 11 に示すような、HeadFinder により抽出された頭部画像を用いる。これは、人物は頭部の皮膚の色と腕の皮膚の色が近いという仮定が成り立つためである。HeadFinder で抽出される頭部画像はフレーム間差分とその形状情報しか用いてないため、色情報に依存せず頭部を抽出することができる。人物の皮膚は HS 空間で表現した時に色相が 32 度の幅を持つことで知られている [4]。そこで我々は HS 空間より人物の皮膚の抽出に優れている TS 空間 [3] を利用した。TS 空間で人物の皮膚を表現すると、図 12 に示すように約 38 度の幅で表現できることが分かる。よって、腕の皮膚の色を抽出するには図 13 の様な、基準色から 19 度の幅の扇型内の色を抽出すれば良いことになる。次に、頭部から基準色を決める手法を以下に示す。

1. 頭部画像の各画素を TSL 色空間に変換する。
2. 輝度が低い画素は髪や影であるので消去する。
3. 残りの画素に対し、T と S の平均である T_a と S_a を計算する。

この時の T_a と S_a が基準色となる。なお、輝度に関しては光源の影響を受けやすいため利用しないことにした。



図 11: HeadFinder から得られた頭部画像

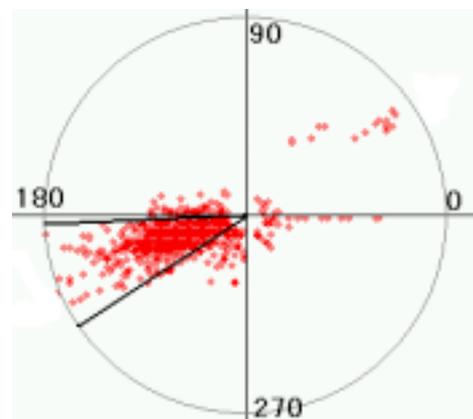


図 12: TS 空間でプロットされた皮膚の色

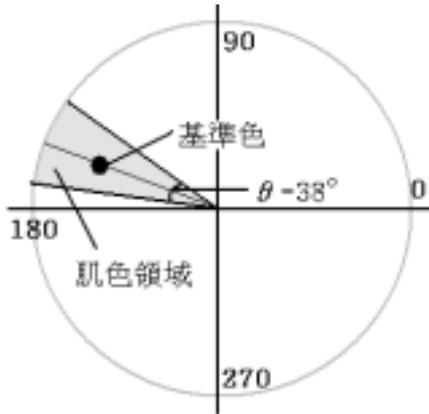


図 13: 基準色から推定される皮膚の色

5.3 腕の検出

腕の検出は、前節で求めた基準色をもとに、頭部以外の皮膚の対象画素を抽出する。抽出された画像をもとに、x 方向、y 方向のヒストグラムを生成し、その画素数から腕の有無を確認する。腕が存在する場合、人物の腕は1本もしくは2本であるので、判別分析法を用いヒストグラムの分割を試みる。分割可能であるなら、腕の本数は2本として処理し、分割不可能なら、腕の本数は1本として処理する。

上記の手順により、腕と思われるプロップを0~2個生成し、そのプロップを矩形として切り出し、表示したものを図14に示す。この例では画像中に腕が2本確認できているので、人物は半袖を着ていると推測できる。逆に、腕が確認できない場合、人物は長袖を着ていると推測できる。



図 14: 抽出された腕の画像

6 性能評価

6.1 システム構成

本システムの実験は表4に示す環境で構築されている。なお、本システムは服の抽出から分類までを秒間7フレームで処理する。

表 4: システム構成

計算機	AT 互換機
OS	FreeBSD4.4
CPU	Athlon700MHz
Camera	Canon VC-C4R
Capture Chip	Bt848

6.2 実験環境および内容

本システムを室内環境で評価した。ただし、照明など光の変化にロバストであることを確認できるように、いくつかの部屋で実験を行った。また、評価は服ごとの分類精度ではなく、記録画像枚数に対して行った。ただし、システムの得手、不得手に影響されないように服1枚に対しての記録画像枚数は、ほぼ一定にするようにした。

実験内容は人物が服を着てフレームインする。カメラの前で自由に歩いた後、360度回転してフレームアウトする。1枚の服に対しての実験時間はおよそ15秒、記録画像枚数として100枚を目安に行った。実験は服の色、袖の長さをとくに制限せず行った、ただし服の模様はプレーン、ストライプ、チェックに制限し、システムに対応していないスーツなどの場合は実験の対象外とした。また、動いている間に影や模様が頻繁に変わるのでその影響についても調べた。

6.3 実験結果

本節では前章までに挙げた服装の分類のための各手法を実際に行い性能を評価した結果を示す。ここで#sampleは記録画像枚数、vote(30)は過去30フレームで一番投票が多かったものを出力した結果である。

表5の色の分類はストライプとチェックの場合、服を構成する2色とも正しい時に分類できたとする。色の分類に関しては模様ごとの色の分類精度を求めた。

表6の模様の分類は各模様ごとにどの程度正しく分類できたかを示す。また、本論文で提案したvotingとp-votingとの分類精度の違いについても示す。

図15はvotingとp-votingで投票数の変化により、どれくらい分類精度が変化するか比較した結果である。

表7の袖の分類は半袖と長袖の分類精度を示す。腕はオクルージョンで隠れることがしばしばあり、1フレームだけの分類は意味をなさないため、30フレームでの分類精度を示した。

表 5: 色の分類に関する実験結果

texture	#sample	vote(30)
plain	837	95%
v-stripe	565	95%
h-stripe	692	97%
check	792	79%

表 6: 模様分類に関する実験結果

texture	#sample	vote(1)	vote(30)	p-vote(30)
plain	659	86%	95%	97%
v-stripe	446	78%	96%	98%
h-stripe	272	82%	94%	99%
check	1029	82%	88%	92%

7 考察

7.1 色の分類について

本研究では実環境で服装を分類するため、光の具合による色のグラデーションや影の影響に強い一次元ヒストグラム法を用いた。表 5 でも分かるようにプレーンやストライプに対して概ね 90%以上の認識率が得られることから、本手法が有用であることを示している。しかし、チェックに関しては認識率が他に比べて劣っている。これは 3色以上の色が服の中に含まれている場合、より大きい領域 2色のみを選択していることが挙げられる。この問題は複数色の候補がある場合すべて属性として抽出してしまうことで解決できそうだが、背景などの服でない部分によって得られた小さな領域までもが候補として挙がってしまう可能性がある。また、人間の心理に基づいた色の分類も考えている。例えば、緑・赤・灰の 3色で構成されたチェックの場合、人間が判断したら緑と赤のチェックに見えても、実際は灰の方が赤より少し大きかったりする。これは人間が無彩色より有彩色の方が印象に残るためである。このように人間の心理に基づいた色の分類をすることでチェックの認識率が向上すると思われる。

全体的な認識率の低下としては、影によるバースト誤りが認められた。影の影響に強いとされる一次元ヒストグラム法を利用しても、影の領域が大きくなりすぎると影が一つの領域となってしまう。この様な時、影で他の色の領域が生成され連続的に表れてしまうため、voting を使っても真の値を推定できなくなってしまった。

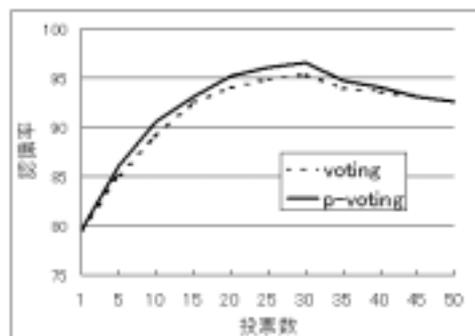


図 15: voting と p-voting の認識結果の違い

表 7: 袖の分類に関する実験結果

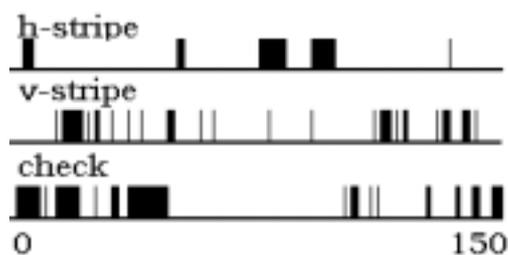
sleeve	#sample	vote(30)
long	208	93%
half	338	99%

7.2 模様分類について

模様分類の結果が表 6 に示されている。P-voting を用いることにより、voting に比べて良い結果となった。しかし、投票を用いない時でも 80%の認識率を示しているにもかかわらず、投票数を 30にしても場合によっては認識率が 90%程度までしか上がらないことや、投票数が 30 を越えた場合に認識率が低下するなどの現象が見られた。主な原因を調べてみると、認識率が極端に低下している時、図 16 に示すようなバースト誤りが起こっていることが確認できた。この様な場合、明らかに voting の様な手法では対処できず、逆に投票数が多いと誤りを引きずってしまう原因となっている。

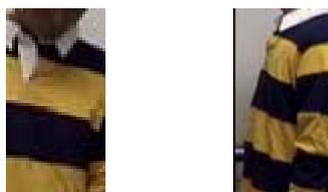
バースト誤りは大きく分けて図 17 から図 19 に示す 3つの原因により起こると確認できた。まず、図 17 のような服の不完全な抽出ある。人物が激しく動いている時、頭部と服の位置関係がずれることで連続して不完全な抽出が起こることを確認した。次に、図 18 のような、主に腕によるオクルージョンの発生である。人物は当然動いているので腕などのオクルージョンが発生し、バースト誤りが起こる。最後に、図 19 はチェックであるが、解像度の影響により、チェックとプレーンを誤認識する場合がしばしばある。特にチェックは模様が細かい場合が多く、この原因によるバースト誤りが顕著に表れた。

これらの 3つの原因を解決することで、voting の効果が表れ、認識率の向上につながると期待できる。



150 フレームまでのタイムチャート、黒い線が入ったところが誤ったフレームを示す。

図 16: パースト誤り



(h-stripe と判断) (v-stripe と判断)
(a) 正しい認識 (b) 誤った認識

図 17: 服の不完全な抽出による影響

7.3 型の分類について

腕の検出をベースとした長袖と半袖の結果は90%以上の認識率を示せたことから、有用な手法であることが確認できた。腕の色をあらかじめ定義せず、ロバストに抽出できる顔の色を基準値とし、腕を抽出する手法は光源の変化にも即座に対応できた。特に本手法のような動画像中からどれだけロバストに腕を検出できるかがポイントになるような場合、非常に有効であった。本手法は腕の大きさとなりえるプロップを生成しているため、多少のノイズには影響されないが、色のみを基本として腕を抽出しているため、壁全体が腕の色と近い場合は抽出できない。

8 まとめ

本論文では一次元ヒストグラム法とフーリエ変換を軸とした服装の実時間分類法を提案した。これらの手法を用いたシステムの特徴は以下の通りである。

- 実時間での服領域の抽出
- 服のしわや腕の影などによる色の変化に強い分割法の採用
- あらかじめ閾値を定めない各手法の採用によるロバストな分類
- 腕の検出をベースにした長袖、半袖の判定



(v-stripe と判断) (check と判断)
(a) 正しい認識 (b) 誤った認識

図 18: オクルージョンによる影響



(check と判断) (plain と判断)
(a) 正しい認識 (b) 誤った認識

図 19: 解像度による影響

今後の課題を以下に示す。

- オクルージョンなどによるパースト誤りの対応
- スーツなどの服装の種類の判別
- 動画像中から服装のデータベースの作成

参考文献

- [1] Jean-Christophe Terrillon and Shigeru Akamatsu, "Picture Segmentation Using a Recursive Region Splitting Method", *Computer Vision and Image Processing*, 8,3,1979.
- [2] 馬場功淳, 大橋健, 乃万司, 松尾英明, 江島俊朗, "HeadFinder:単眼視動画像を用いた複数人追跡", *画像センシングシンポジウム*, pp363-368,2001.
- [3] Jean-Christophe Terrillon and Shigeru Akamatsu, "Comparative Performance of Different Chrominance Spaces for Color Segmentation and Detection of Human Faces in Complex Scene Images", accepted at *Vision Interface '99(VI'99)*, to be held in Trois-Rivieres, P.Q., Canada, 18-21 may 1999.
- [4] Jamie Sherrah and Shaogang Gong, "Skin Colour Analysis", http://www.dai.ed.ac.uk/CVonline/LOCAL_COPIES/GONG1/cvOnline-skinColourAnalysis.html, 18,5,2001.