

インテリジェントルームにおける ジェスチャ認識のための肌色登録手法

Skin Color Registration for Gesture Recognition in an Intelligent Room

入江 耕太 (中央大/JST CREST) 荻島 秀敏 (中央大)
○学 高橋 真人 (中央大) 正 寺林 賢司 (中央大/JST CREST)
正 梅田 和昇 (中央大/JST CREST)

Kota IRIE, Chuo University/CREST, JST, kotairie@sensor.mech.chuo-u.ac.jp
Hidetoshi OGISHIMA, Chuo University
Masahito TAKAHASHI, Chuo University
Kenji TERABAYASHI, Chuo University/CREST, JST
Kazunori UMEDA, Chuo University/CREST, JST

This paper proposes skin color registration in an intelligent room. We apply the gesture recognition to the operation of the household electrical appliances. In order to recognize the gestures, the skin color information is used in this system. Existing method we had constructed took time to register the skin color. That means, it was difficult to operate household electrical appliances smoothly. Therefore we developed the new method that improved the skin color registration, it recognize the gestures more smoothly by any skin colors. The new method uses a result of waving hands recognition that identifies an operator. It was confirmed that the method could resister the skin color smoothly and it was valid for an intelligent room system.

Key Words: Image Processing, Gesture Recognition, Color Extraction, Intelligent Room, Hand Waving

1. 緒言

近年、コンピュータの処理能力や画像処理技術の発達に伴い、ジェスチャをはじめ人間の動作認識に関する研究が数多く行われている。我々は、ジェスチャをヒューマンマシンインタフェースとして用いて家電機器の操作を行うインテリジェントルームを構築してきた[1] (図1参照)。人物の動作や状態を認識するためには、肌色情報を用いた手法が数多く提案されており、その抽出方法の研究等も盛んに行われている[2][3]。

インテリジェントルームでは動作認識のために肌色情報を用いており、照明環境や個人差に対してロバストな肌色抽出を行うために、操作を開始する際に毎回操作者の肌色登録処理を行っている。現在インテリジェントルームに実装されている肌色登録では、手領域を抽出する際に手のひらをカメラに向けて一定時間静止させる必要がある[4]。システムとしてスムーズな操作を実現するためには、ユーザーにとって煩わしい動作を極力減らすことが要求される。そこで本研究では、手振り認識を利用した肌色登録手法を提案する。インテリジェントルームにおける手振り認識[5]は、システムに操作者の位置を特定させるために用いられている。この手振り認識と肌色登録を統合することで、ユーザーへの負担を軽減させ、よりスムーズな操作を実現することを目指す。

2. 肌色登録手法

肌色登録の流れを図2に示す。まず、リアルタイムで背景画像の更新と手振り認識処理を行う。手振り認識後、手振り領域内における過去数フレーム分の画像と、登録した背景画像との差分を取得する。差分画素値の集合内で外れ値を除外した後、色相(H)と彩度(S)の平均ベクトルと共分散行列を求め、最終的な肌色データとして登録する。

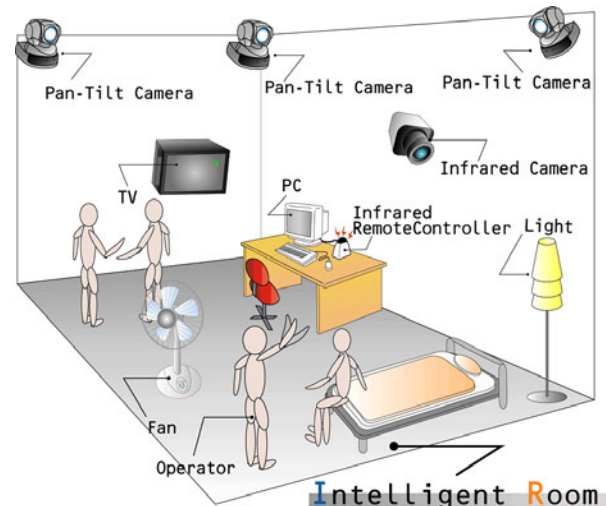


Fig.1 Basic concept of an intelligent room

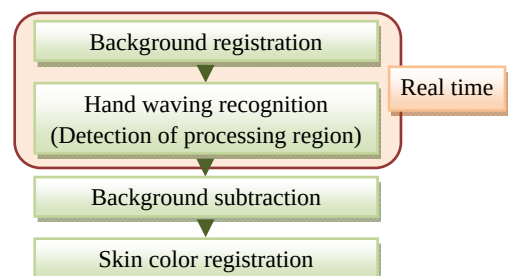
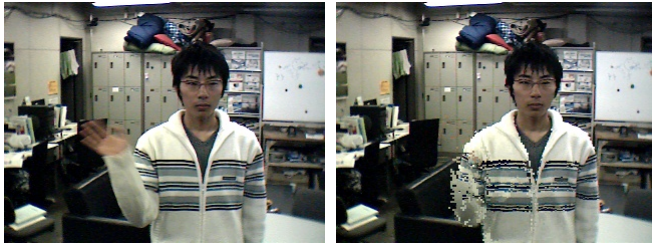


Fig.2 Flowchart of skin color registration

2.1 背景登録

手振り動作中の手の画素値を抽出するために背景登録を行う。数フレーム間変化のない画素値を背景画像として登録し、リアルタイムで更新する。短時間で変化の有無を判断するこ

とで、家具や壁などの一般的な静止物体のほか、手振りを行う際に一時的に静止する体を背景として登録する。一方、手振り動作は相対的に高速であるため、図 3 で示すように手振り領域内では背景更新が行われず、手は背景として登録されない。手振りの最低周波数は 1.9[Hz]と定義している。



(a) Hand waving image (b) Background image
Fig.3 Background image registration

2.2 手振り認識

インテリジェントルームにおける操作開始の合図である手振り動作を認識する。まず、処理負荷軽減ならびに振動のスムーズ化のため、キャプチャ画像を 40×30 [pixel]に低解像度化する。次に、各画素における輝度値の時系列変化を取得し、図 4 で示すように最大値と最小値の差を基準に正規化する。正規化した輝度値変化に対して FFT (Fast Fourier Transform) を行い、得られた周波数のパワースペクトルから最大、平均、歪度、尖度などの要約統計量を算出する。これらを特徴量とした多変量のマハラノビス距離による判別分析[5]を行うことで、手振りを認識する。さらに、手振り動作領域も同時に抽出し、以後の処理領域として定義する (図 5 参照)。

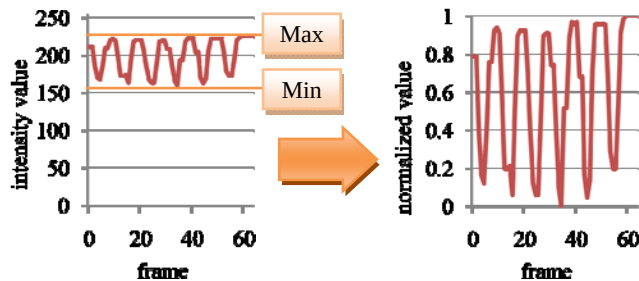


Fig.4 Normalize for intensity change

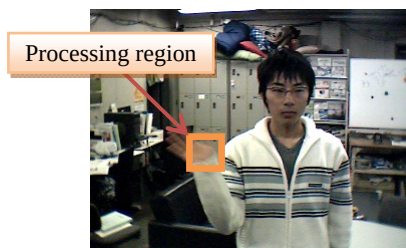


Fig.5 Hand waving recognition

2.3 背景差分

手振り領域内で、背景画像と過去数フレームのキャプチャ画像との差分を取得する。差分の評価値には、正規化した色相 (H)、彩度 (S)、輝度 (V) それぞれの差の絶対値を合計した S_d を用い、時系列方向で比較する。過去 n フレームに対して、背景の色相成分 $I_H(0)$ 、彩度成分 $I_S(0)$ 、輝度成分 $I_V(0)$ 、キャプチャ画像の色相成分 $I_H(t)$ 、彩度成分 $I_S(t)$ 、輝度成分 $I_V(t)$ として、

$$S_d(t) = |I_H(t) - I_H(0)| + |I_S(t) - I_S(0)| + |I_V(t) - I_V(0)|$$

$$S_d(t) \geq \max(S_d)/2 \quad (t = 1, 2, \dots, n) \quad (1)$$

を満たす場合に肌色とする。背景画像には手が含まれていないため、手が領域内を通過する際に差分値が大きくなる。そこで、フレームごとに手振り領域内の画素値を平均化し、肌色フレームを選択する (図 6 参照)。次に、元の解像度 (320×240 [pixel]) において画素単位で肌色を選択し、肌色候補として登録する (図 7 参照)。このとき、体の揺れなどによる背景画像とキャプチャ画像との小さなずれを軽減するために、背景画像をガウシアンフィルタによって平滑化する。差分処理によって除外された画素値は背景データとして登録する。

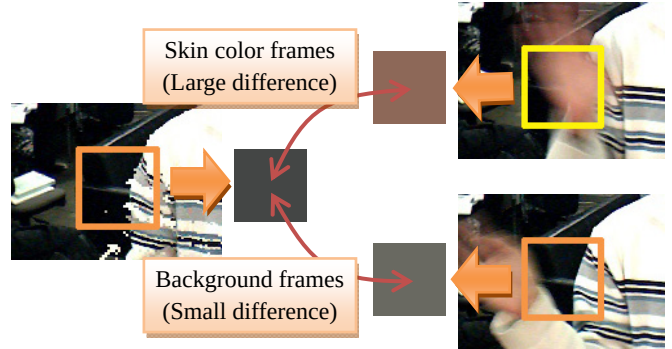


Fig.6 Background difference in frame

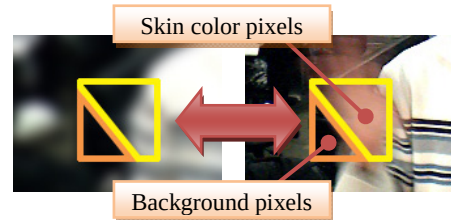


Fig.7 Background difference in pixel

2.4 肌色登録

登録する肌色情報の精度を向上させるために、誤って登録した服の画素値やノイズによる外れ値を除外する。H, S, V それぞれの中央値を基準に一定範囲で制限し、さらに各値のヒストグラムから両端から一定割合を外れ値として除外し、肌色画素値として登録する。

登録した肌色画素値のうち、V は同じ環境でも照明の当たり方によって値が大きく変動する。ロバストな肌色登録を行うために、照明に影響されにくい H, S を特徴量ベクトル $\mathbf{X}$ とする。式(2)、式(3)より、マハラノビス距離を計算するための平均ベクトル $\mathbf{A}$ および共分散行列 $\mathbf{V}$ を求め、最終的な肌色データとして登録する。

$$\mathbf{A} = \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i \quad (2)$$

$$\mathbf{V} = \begin{bmatrix} \sigma_H^2 & \sigma_{HS} \\ \sigma_{HS} & \sigma_S^2 \end{bmatrix} \quad (3)$$

N は登録する画素値集合の要素数である。キャプチャ画像から肌色を抽出する際は、

$$dm^2 = (\mathbf{X} - \mathbf{A})^T \mathbf{V}^{-1} (\mathbf{X} - \mathbf{A}) \quad (4)$$

から肌色データと背景データそれぞれに対するマハラノビス距離 dm_{true} , dm_{false} を求め,

$$dm_{true}^2 \leq dm_{false}^2 \times c \quad (5)$$

$$dm_{true}^2 \leq k \quad (6)$$

を満たす画素値を肌色として抽出する．ここで， c は背景に対する補正係数で， k は肌色に十分近いことを保証するための閾値である．抽出処理は，操作者が手振りと同位置でジェスチャを行うことを前提として，手振り認識領域の縦横 8 倍の範囲に限定する．また，抽出結果に対してモルフォロジ処理とラベリング処理を行い，画像中のノイズを除去し肌色領域を補正する．

3. 肌色登録検証実験

3.1 実験システム

実験システムは PC (DELL XPS 420, Intel Core2 Quad Q6600), キャプチャソフト (Microsoft DirectShow), 画像処理ソフト (Intel OpenCV), USB カメラ (ELECOM UCAM-E130, 30[fps]) により構成される．

3.2 実験条件

各処理のパラメータは以下のように設定した．背景登録は 5 フレーム (約 0.17[s]) で変化の有無を判断した．FFT のサンプリングフレーム数は 64 フレームとし，背景差分は $n=64$ とした．中央値による制限では H を ± 9 , S および V を ± 50 とし，ヒストグラムの両端 5[%]を除外した．マハラノビス距離による肌色の判別閾値は $c=4$, $k=10$ とし，登録したデータで適切に肌色領域を抽出できた場合を成功とした．

3.2.1 距離

距離に対する性能を評価するために，カメラから 1.5[m], 3[m], 5[m]の距離で各 20 回ずつ手振りを行い，肌色登録の成功率を検証した．

3.2.3 環境

環境の色に対する性能を評価するために，衣服と背景の色を変えて実験を行った．距離 1.5[m]で灰色，青色，クリーム色の環境で各 20 回ずつ手振りを行い，肌色登録の成功率を検証した．手動で取得した各環境の色を図 8 に示す．赤が肌色の分布を表し，その他は各環境の色に対応している．

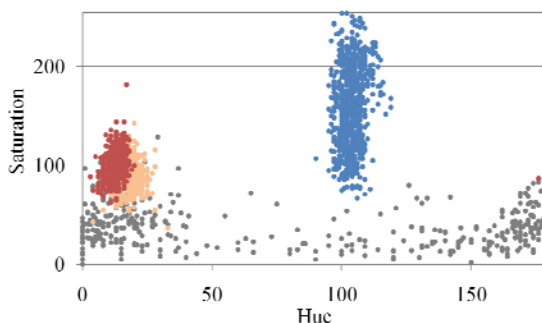


Fig.8 Distribution of ambient color and skin color
(Red: skin color, others: ambient color)

3.3 実験結果

3.3.1 距離に対する性能評価

距離に対する性能評価結果を表 1 に示す．失敗したケースは，認識した手振り領域が実際の手振り位置とずれている場

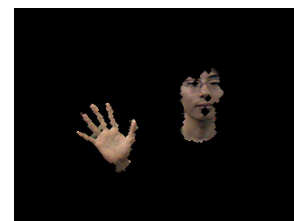
合と適切に肌色を抽出できない場合とに分けた．距離によらず手振り認識領域のずれの影響が最も大きく，手首や腕，手の影といった手振りと同時に振動する領域を認識してしまう場合や，手振り領域中のごく一部の画素のみ認識してしまう場合があった．手のひら付近が認識されないと，その後の背景差分で肌色を適切に抽出することができなかった．

また，遠距離になると領域が適切に認識されても肌色登録が失敗することがあった．肌色らしい色を登録できていても，背景との分離が不十分な場合や，逆に肌の一部分しか識別できない場合があった．解像度が相対的に低下して情報量が減少し，肌色画素値を必要十分に取得できずに偏った値が登録されたことが原因だと考えられる．これにより，遠距離になるほど精度が低下した．距離 3[m]以上ではキャプチャ画像 (320×240[pixel]) 中の手の縦幅が 12~23[pixel]程度，指の幅が 3[pixel]以下となり，適切に肌色を登録できた場合でも図 9 で示すように指本数などの細かい形状を識別できる画像は得られなかった．ただし，手領域の運動方向などの大まかな動きの識別には十分だと考えられる．

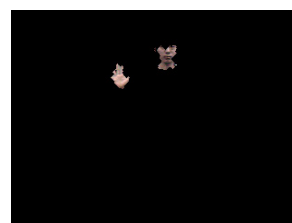
手振り開始から手振り認識までの時間は 1.5[m]で 1.8[s], 3[m]で 2.0[s], 5[m]で 2.1[s]となり，距離が遠くなると若干時間がかかる傾向があった．

Table 1 Success and failure rates of distances

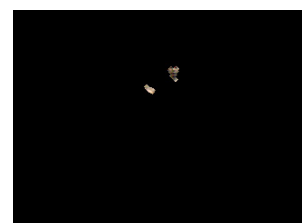
Distance	Success	Failure	
		Position	Skin color extraction
1.5[m]	85%	15%	0%
3[m]	75%	15%	10%
5[m]	65%	20%	15%



(a) 1.5[m]



(b) 3[m]



(c) 5[m]

Fig.9 Extraction result of skin color regions

3.3.2 環境の色に対する性能評価

環境の色に対する性能評価結果を表 2 に示す．背景色によらず手振り認識領域のずれの影響が最も大きかった．灰色および青色では，領域が適切に認識されれば全て成功したが，クリーム色では失敗する場合があった．色の分布が類似しているために，登録したデータが偏っていた場合の影響が大きいことが考えられる．失敗した 2 回ではいずれも手領域に背景が混ざった領域が抽出されたが，実験条件で述べた c および k を調節すれば適切に肌色を抽出することができた．登録値によって動的に適切な閾値を決定することができれば，改善する可能性がある．

Table 2 Success and failure rates for difference of ambient colors

Color	Success	Failure	
		Position	Skin color extraction
gray	85%	15%	0%
blue	90%	10%	0%
cream	75%	15%	10%

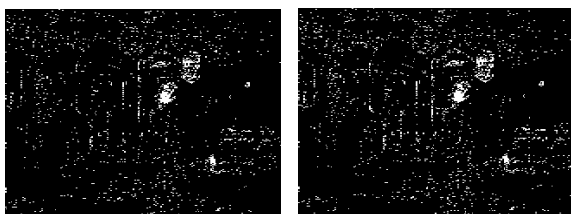
3.3.3 肌色登録値の評価

本手法で登録した肌色と、手動で手領域を選択して取得した肌色とを比較する。肌色登録が成功したデータの H および S の最大値 max , 最小値 min , 平均値 Ave , 標準偏差 SD を求め、自動登録値と手動登録値の差の絶対値を求めて平均した結果が表 3 である。また図 10, 図 11 は自動登録値と手動登録値とによる実際の肌色抽出結果を比較した図で、白黒画像はノイズ除去を行う前の画像である。図 10 はずれ幅が表 2 の平均程度の場合で、図 11 は Ave_H で 2.2, Ave_S で 7.3 と大きくずれていた場合である。ノイズ除去後の肌色抽出結果では顔部分が欠けているが、手領域は問題なく認識できた。

図 12(a)は、図 10 における具体的な画素値データを図示したものである。背景画像が青、背景差分で除外された値が緑、閾値処理で除外された値が黄色、自動登録した肌色がオレンジ、手動登録した肌色が赤でプロットされている。背景差分時には大きく分散していた値が、閾値処理によって手動登録値とほぼ同じ領域まで収束しているのがわかる。図 12(b)は、手振りの先端部分が選択されて肌色登録が失敗した際の値である。先端付近は動きが早く、画像中の手がぼけて背景と混ざった色になるため、肌色領域から大きくずれた値になった。

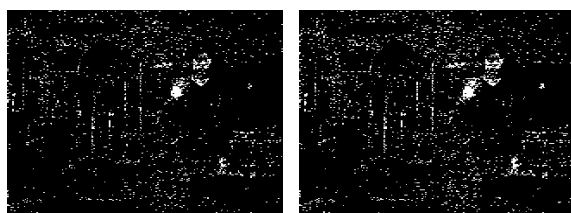
Table 3 Absolute value of the difference between true values and extracted values

	max	min	Ave	SD
H	3.0	2.8	1.2	0.6
S	11.8	9.8	4.1	2.2



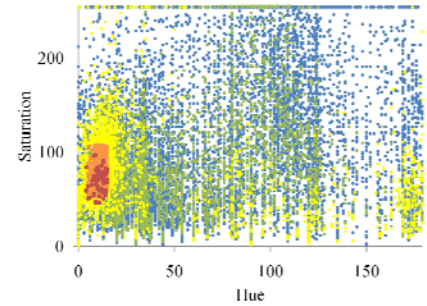
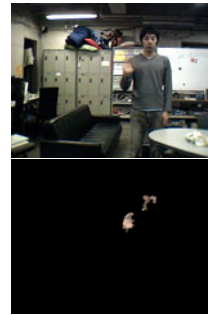
(a) Calculated color value (b) Manual color value

Fig.10 Extraction results with standard difference

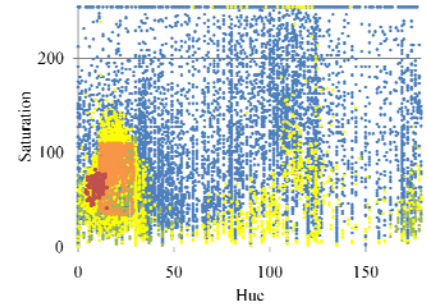


(a) Calculated color value (b) Manual color value

Fig.11 Extraction results with large difference



(a) Success case



(b) Failure case

Fig.12 Color distribution

(Blue: background region, green: result of background subtraction, yellow: result of thresholding using eq.(1), orange: registered value as skin color, red: true values of skin color)

3.3.4 処理時間

手振り認識から肌色登録完了までの処理時間は 0.05 ~ 0.10[s]で、肌色登録を意識せずに行えるようになった。手振り開始から認識までの時間を合わせ、平均して約 2[s]でジェスチャ認識状態へ移行した。

4. 結言

手振り認識を用いた肌色登録手法を提案し、実験によって提案手法の有効性を確認した。序論で述べた肌色登録の煩わしさは解消できたが、認識精度は改善の必要がある。今後の課題として、手振り認識領域の補正、肌色抽出閾値の動的決定、高解像度かつ低ノイズなカメラへの切り替えが挙げられる。

文 献

- [1] 入江耕太, 若村直弘, 梅田和昇, "ジェスチャ認識に基づくインテリジェントルームの構築", 機論 C 編, vol.73, no.725, pp.258-265, 2007.
- [2] P. Kakumanua, S. Makrogiannisa and N. Bourbakis, "A survey of skin-color modeling and detection methods," Pattern Recognition vol.40, no.3, pp.1106-1122, 2007.
- [3] S. L. Phung, A. Bouzerdoum, D. Chai, "Skin segmentation using color pixel classification: analysis and comparison," IEEE trans. Pattern Analysis and Machine Intelligence, vol.27, no.1, pp.148-154, 2005.
- [4] 若村直弘, 鈴木健一郎, 梅田和昇, "直観的なジェスチャの認識を利用したインテリジェントルームの構築", 日本機械学会ロボティクス・メカトロニクス講演会, 2A1-N-051, 2005.
- [5] 入江耕太, 梅田和昇, "濃淡値の時系列変化を利用した画像からの手振りの検出", 日本ロボット学会誌, vol.21, no.8, pp.923-931, 2003.
- [6] 末広啓史郎, 高橋武則, "マハラノビスの距離による判別の研究", (社)日本品質管理学会第33回年次大会, vol.34, no.1, pp.62, 2004.