

口唇の動きによるヒューマンインターフェースのための動画像認識

井上 哲理¹・斎藤 譲²・関 靖夫¹

¹情報工学科

²日産セキュリティ・サービス(株) (情報工学科平成 10 年 3 月卒業)

Dynamic Image Recognition Experiment for Human Interface by Lips' Movement

Tetsuri INOUE¹⁾, Yuzuru SAITO²⁾, Yasuo SEKI¹⁾

Abstract

This paper describes a method of recognizing lips' movement from image sequence for an input method to the user-friendly computer system. In the proposed method, the computer compares two images of the face to detect the change of lip's positions. The procedure of the method is as follows: (1) lips' region is extracted from each input image by estimating HSV (hue, saturation, value) values of the pixels. (2) each input image is transformed to a binary image representing lips' region. (3) the average and variance of the binary image are calculated in the horizontal and vertical directions. (4) the lips' movement is recognized by comparing the averages and the variances of two images. The experimental results show the proposed method recognizes various kinds of lips' movements correctly. The method is expected to be useful for an input method to computer systems.

Key Words : Human interface, Dynamic images recognition, Lips' movement

1. はじめに

今日、コンピュータ技術を用いたさまざまな機器が我々の生活の中に広く浸透してきている。これらの機器が対象とするユーザは多様であり、これまでコンピュータに触ることのなかった人々も含まれる。その結果、コンピュータ技術の発展には、これまでのハードウェア面での発展に加え、ユーザにとって使いやすいシステムを実現するヒューマンインターフェース面での進歩がますます重要になっている。

コンピュータシステムに馴染みのないユーザを対象としたインターフェースを実現する上では、できるだけコンピュータ操作を意識させないことが必要である。そのためには特別の機器を用いないことや特別な操作方法を必要としないことなどが要求される。つまり、普段の生活で経験する行為・行動をできるだけ変えない操作が求

められる。その方法のひとつとして、人物像を介した操作システムは非常に有効と考えられ、この実現を目指す多くの研究が行われている¹⁾。著者らのグループも以前、手指画像を用いたシステムに関する研究を発表した²⁾。

ところで、我々は日常生活の中で顔からさまざまな情報を得ている。人間どうしが言葉によらない情報伝達を行う場合、顔が情報伝達媒体として大きな役割を果たしていると考えられ、人間から計算機への自然な情報伝達の実現を目指す上で顔画像の認識技術は重要である。顔画像認識を用いた例として、表情の認識システムや個人識別システムなどがあげられ、古くから研究されている。その中で唇の動きをコンピュータに取り込もうとする研究は、コンピュータによる読唇や音声認識のための補助手段としての利用、あるいは高齢者や身体障害者の意志伝達手段などの目的で研究が進められている。著者らもこの点に注目し、顔画像中の口唇の動きを用いたシステ

ムを検討し、その動きを認識する方法を考えた。

本研究で検討したシステムは、顔の動画像より口の開閉を認識し、口の開閉動作およびその回数でコンピュータへの指示を行うものである。例えば、メニュー画面での選択を口の動きにより行うシステムなどが考えられる。システムにおいては、マーカなどの認識を容易するものを顔に付加することなく、顔画像のみで口の開閉を認識することを目指した。同システムの処理のうち、今回は口の開閉判別の方法を検討し、実際の画像を用いて実験的に有効性を検証した。

判別処理では時間的に前後する2枚の顔画像を用いる。まず、それぞれの画像から口唇部分を抽出し、2値化画像を作成する。その2値化された画像を用いて開閉の判定を行う。判定方法として、今回は差分画像により開閉を判断する方法と、平均と分散により開閉を判断する方法を検討した。差分による判定法は、2値化された2枚の画像上で同じ位置の画素値を比べて、差があった画素の割合により動きを判別するものである。一方、平均と分散による動きの判定法では、2値化画像の横方向と縦方向の画素値の平均と分散を求める。そして、2つの画像の平均・分散の大小により開閉を判断する。

カメラ入力画像を用いた実験の結果より、差分による判別法では動きの有無は判定できたが、開閉の順番は判

断できなかった。一方、平均と分散による判定法では2画像間における開閉の順序、顔のうなずきの判断が可能になったことがわかった。

2. 口唇の動き認識処理

2.1 処理全体の流れ

本処理は、時間的に前後する2枚の顔画像を比較し、口の開閉などを認識するものである。処理の流れを図1に示す。実際のシステムでは2枚の画像を入力するタイミングが問題となるが、今回の実験では、あらかじめ入力した2枚の画像を用いた。主な処理は、顔画像から口唇部分を抽出した2値化画像の作成と、2値化画像による開閉の判定処理である。開閉の判定方法は2つの方法を検討した。次にそれぞれの処理について説明する。

2.2 口唇の抽出・2値化処理

(1)濃淡値およびRGB値による口唇抽出の検討

画像からの特徴抽出として、モノクロ濃淡画像を対象として濃淡値の変化に着目したフィルタリングや2値化などの処理を用いるさまざまな方法がこれまでに提案されている。しかし、顔画像から口唇部分を抽出するには口唇と周辺の皮膚との明度差が小さいこと、照明による明度変化やむらが避けられないことが今回用いた画像より確認された。従って、自然な状態において濃淡値の変化のみにより口唇形状を抽出をすることは困難であることがわかった。そこで本処理ではカラー画像を扱い、色情報をもとに顔画像からの口唇の抽出を行なった。

コンピュータでは通常、カラー画像はRGB(赤(red), 緑(green), 青(blue))の3成分に分けて扱われる。画素のRGB成分の割合、大きさが色に対応する。従って、画像中の口唇部分のRGB値に近い範囲のRGB値を持つ画素を抽出することで、口唇部分の抽出が可能と考えられる。そこで、実際の顔画像をもとに、顔の各部分がどのようなRGB値を持つかを調べた。今回調べた結果を表1に示す。これは口唇周辺の典型的な部位の相対的強度の平均値を表している。これより口唇周辺のRGB成分は、次のような性質を持つことが観測される。

- ・R成分は照明がとどかない口腔や鼻孔などを除いて各部ほぼ均等である。

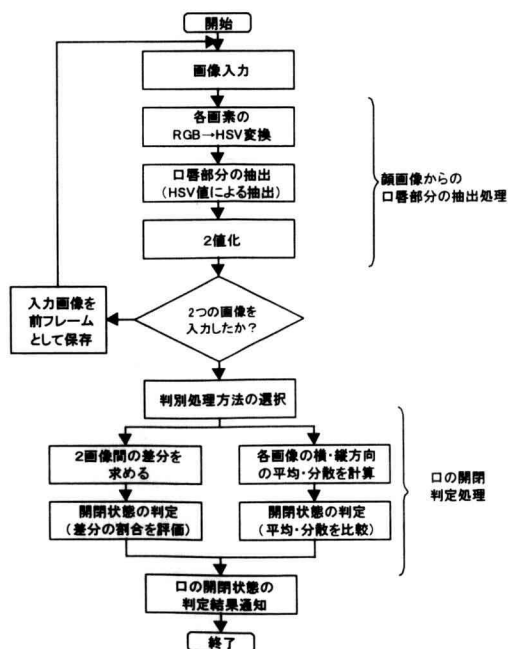


図1 処理全体の流れ

表 1 口唇周辺部の RGB 成分の強度比

部位	R	G	B
口唇	0.17	0.12	0.08
皮膚	0.17	0.14	0.10
口腔	0.10	0.09	0.08
歯	0.16	0.15	0.11
舌	0.17	0.12	0.10
鼻孔	0.10	0.09	0.08

RGB 各成分は 0~1 の値をとる

- ・ G 成分は周辺の皮膚や歯と比較して、口唇で明らかに小さな値となる。また口唇と口腔の差も大きい。
- ・ B 成分は、口唇と周辺の皮膚や舌で僅かな差が認められるが分散値が大きく、また唇と口腔の差が認められない。

以上より、唇の形状抽出には G 成分が重要な要素であると考察される。しかし次のような問題も発生した。

- ・ 周辺の皮膚で分離できない部分が残る。
- ・ 小鼻や顔の輪郭部分が明瞭に残る。
- ・ 照明むらの影響を直接受ける。

従って、RGB 成分による抽出処理では、口唇部分以外が抽出される可能性があり、照明条件の影響を受けやすいなど結果が不安定になる可能性が大きい。これらの考察の結果、今回の処理では RGB 成分による抽出方法は用いなかった。

(2) HSV による方法

顔画像からの口唇部分の抽出処理では、色の識別が容易な表色系で各画素の値を表現する方が適していると考えられる。そこで、本研究では HSV(色相(Hue)、彩度(Saturation)、明度(Value)) 表色系を用いた抽出処理を行った。HSV 表色系では、皮膚の肌色に関して H 値に明確なピークを持つ特徴がある。また、HSV 表色系を用いると明度の変化による H 値の変動が小さいことが実験的に確かめられている。なお各画素の HSV 値は RGB 値より以下の式で導かれる。

$$V = \max(R, G, B) \quad (1)$$

$$S = 1 - \frac{3 \min(R, G, B)}{R + G + B} \quad (2)$$

$$H = \cos^{-1} \left\{ \frac{1/2[(R-G) + (R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right\}$$

ただし、 $B > G$ なら

$$H = 2\pi - H$$

(3)

抽出・2 値化処理は以下になる。まず、あらかじめ数種類の顔画像から口唇部分の HSV 値の平均値を求めておく。入力画像の各画素の RGB 値を HSV 値に変換し、その値と先に求めた口唇部分の平均的な HSV 値との差が一定範囲内であれば、その画素の値を 1 とし、そうでない場合は値を 0 とする。これにより、2 値化画像ができあがり、口唇部分の抽出が行われる。

2.3 口の開閉の判定処理

(1) 差分画像による判定方法

本方法は、2 枚の 2 値化画像の各画素での値 (0 か 1) を比較することで口の開閉を判定するものである。

まず、先に読み込んだ画像 (前フレーム) 内で値が 1 の画素をカウントする。2 値化処理の結果、値が 1 の画素は唇の部分を示している。したがって、この数が前フレーム画像の口唇部分の面積を表わす。次に 2 枚の画像の差分をとるため、もう 1 つの画像 (現フレーム) を左上から走査していく。前フレーム画像における同一位置の画素と比較して、画素値が異なる画素をカウントする。すべての画素を走査し、カウントされた画素数を差分画素数とする。そして差分画素数が、はじめに求めた口唇の画素数の何割あるかを計算する。

口の開閉を行えば差分画素数は多くなると考えられるので、差分画素数の割合がある値を越えれば口に動きがあると判定することが可能である。

ただし、本処理のみでは口の開閉順序 (開→閉か閉→開) を判定できない。また、口の開閉状態は一定で首を左右、上下に振った場合も差分画素数が多くなり、開閉動作と誤認識されるなどの欠点がある。

(2) 平均・分散による判定方法

本方法は、2 値化した画像をもとに、各画像の X 方向 (水平方向)、Y 方向 (垂直方向) の画素値を重みとした平均と分散を計算し、それらの値を画像間で比較するこ

とで口唇の動きを判定する。画素値を重みづけとした平均値と分散は口唇部分の画素の分布状態を表していると考えられ、口の開閉や口部分の移動（首振り）の判定が可能と思われる。

k 番目フレームの画像の X 方向、Y 方向の平均値 $X_a(k)$, $Y_a(k)$ は 2 値化画像から、次の式(4), (5)で求める。ここで、画素 (i, j) の画素値を $f(i, j, k)$ とする。

$$X_a(k) = \sum_i \sum_j i \cdot f(i, j, k) / D \quad (4)$$

$$Y_a(k) = \sum_i \sum_j j \cdot f(i, j, k) / D \quad (5)$$

$$\text{ただし, } D = \sum_i \sum_j f(i, j, k) \quad (6)$$

X 方向の分散は $\sigma_x^2(k)$, Y 方向の分散は $\sigma_y^2(k)$ は次式で求める。

$$\sigma_x^2(k) = \sum_i \sum_j (i - X_a(k))^2 f(i, j, k) / D \quad (7)$$

$$\sigma_y^2(k) = \sum_i \sum_j (j - Y_a(k))^2 f(i, j, k) / D \quad (8)$$

実際の処理では、計算を行いやすくするために式(7), (8)の代わりに次の式(9), (10)を用いた。

$$\sigma_x^2(k) = \sum_i \sum_j i^2 \cdot f(i, j, k) / D - X_a(k)^2 \sum_i \sum_j f(i, j, k) / D \quad (9)$$

$$\sigma_y^2(k) = \sum_i \sum_j j^2 \cdot f(i, j, k) / D - Y_a(k)^2 \sum_i \sum_j f(i, j, k) / D \quad (10)$$

次に前フレーム画像と現フレーム画像の分散の差を求める。X 方向、Y 方向の分散の差をそれぞれ求め、差の大小によって開閉の有無を判断する。この時、開閉の判断をする条件には次のものが考えられる。ここで、 σ_{th} は差があると判断する閾値である。

(条件 1) X 方向に有意な移動あり

$$|\sigma_x(k) - \sigma_x(k+1)| \geq \sigma_{th} \text{ かつ } |\sigma_y(k) - \sigma_y(k+1)| < \sigma_{th} \quad (11)$$

(条件 2) Y 方向に有意な移動あり

$$|\sigma_x(k) - \sigma_x(k+1)| < \sigma_{th} \text{ かつ } |\sigma_y(k) - \sigma_y(k+1)| \geq \sigma_{th} \quad (12)$$

(条件 3) 開閉なし

$$|\sigma_x(k) - \sigma_x(k+1)| < \sigma_{th} \text{ かつ } |\sigma_y(k) - \sigma_y(k+1)| < \sigma_{th} \quad (13)$$

(条件 4) この状態は基本的に発生しない

$$|\sigma_x(k) - \sigma_x(k+1)| \geq \sigma_{th} \text{ かつ } |\sigma_y(k) - \sigma_y(k+1)| \geq \sigma_{th} \quad (14)$$

口の開閉を行なった場合、縦の動きが横の動きよりも大きいと考えられる。従って、Y 方向の分散の差が X 方向の分散の差より大きい値を示した場合、画像間で口の開閉が行なわれたと判断する。(条件 2) がこの場合にあたる。

(条件 2) の場合、前フレーム画像と現フレーム画像での Y 方向の分散の大小より開閉の順番を判定できる。この時の判定条件を示す。

(条件 2) を満たす場合、

(条件 2-1) 閉→開

$$\sigma_y(k+1) > \sigma_y(k) \quad (15)$$

(条件 2-2) 開→閉

$$\sigma_y(k+1) < \sigma_y(k) \quad (16)$$

(条件 2-3) 動きなし

$$\sigma_y(k+1) \equiv \sigma_y(k) \quad (17)$$

なお、(条件 2-3) の動きなしは、口を開ききった場合と閉じきった場合を意味する。

また、(条件 3) の開閉無しの場合、顔のうなずき、顔の横振りのどちらを行っているかを判定する条件を以下に示す。

(条件 3-1) 顔のうなずき

$$|Y_a(k+1) - Y_a(k)| > \Delta Y_{ath} \quad (18)$$

(条件 3-2) 顔の横振り

$$|X_a(k+1) - X_a(k)| > \Delta X_{ath} \quad (19)$$

以上の処理により、前フレームと現在のフレームとの間で、口にどのような動きがあったかを判別できる。

3. 実験

今回提案した顔画像から口の開閉を判定する処理の有効性を実際の画像を用いて検証した。

口を開いた状態、閉じた状態の画像をいくつか撮影し、入力画像とした。撮影に際しては、特別な照明は用いず、部屋の照明のみを用い、また口唇にマーカなどの付加物は用いていない。画像は約 32,000 色のカラー画像で、画像サイズは 320×240 画素である。

入力画像の例（画像 A～D）を図 2～5 に示す。各図の (a) が原画像（実際はカラー画像）であり、(b) が HSV 分解を用いた 2 値化画像で、黒い部分が抽出された口唇部分を示す。

結果より HSV 分解を用いた 2 値化により、口唇部分がほぼ抽出されていることがわかる。また、抽出画素上に口唇以外の部位が少なく、今回用いた HSV 分解によ

る口唇部分の抽出は適切な処理であったといえる。

次に、得られた 2 値画像から口の開閉判定処理を行った。まず、差分画像を用いた判定処理を行った結果を述べる。図 6 に前フレームを画像 A、現フレームを画像 B とした場合の差分画像を示す。黒い部分が画像間に差があった画素である。この結果では、前フレームの口唇部分の画素数に対して差分画素数が 56.6% あり、口の開閉があったと判定できる。しかし、差分情報のみでは開閉の順序が判定できない。また、他の画像を用いた場合は、正しい判定ができなかったなど問題があった。

次に 2 値化画像の平均・分散を用いた判定処理を行った。表 2 に画像 A～D のいくつかの組み合わせに対して得られた結果を示す。結果より本方法により口の開閉が正しく判定されていることがわかる。また、開閉の順序や首の横振りなど差分画像を用いた処理では得られない情報も正しく得られることも示された。

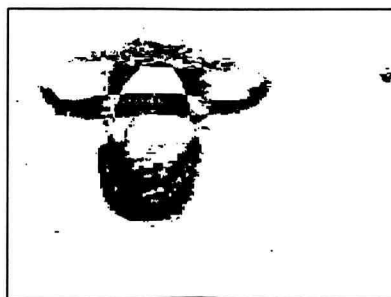
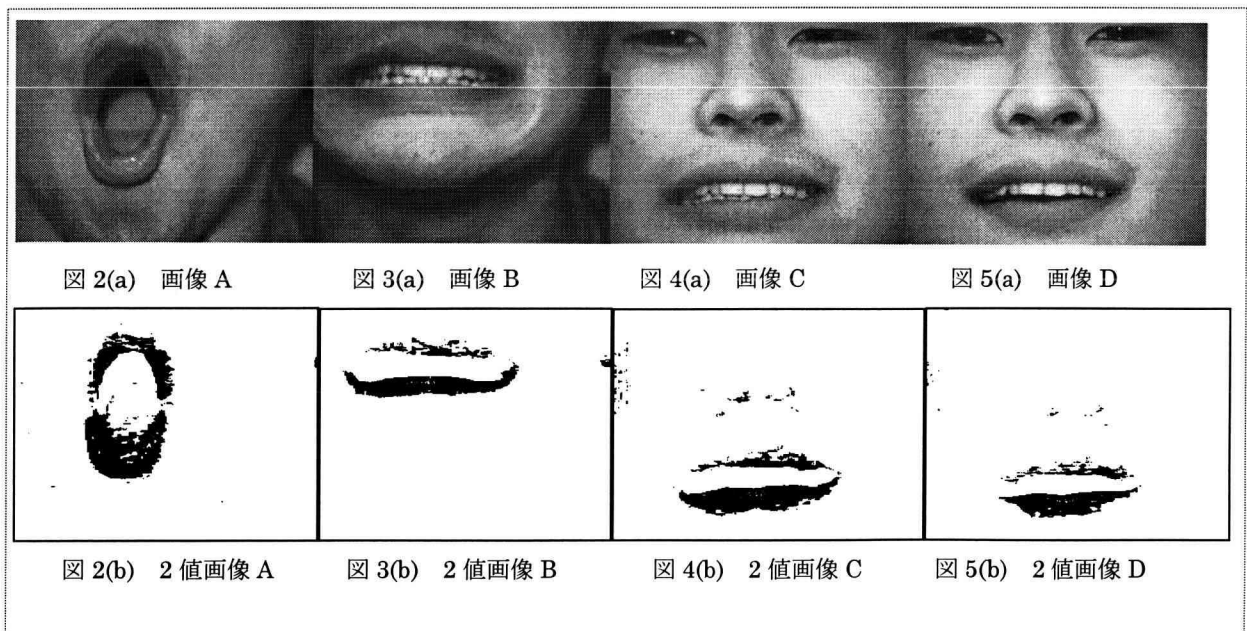


図 6 画像 A, B に対する差分画像

表 2 平均値・分散による開閉判定結果

前フレーム 画像	現フレーム 画像	分散の差		平均の差		適合条件	判定結果
		X	Y	X	Y		
A	B	25.2	29.7	4.8	40.4	(条件2-2)	開→閉
B	A	25.2	29.7	4.8	40.4	(条件2-1)	閉→開
C	D	9.7	10.8	2.2	3.1	(条件3-1)	顔のうなずき
B	C	8.7	5.1	24.0	117.4	(条件3)	動きなし

以上の結果より、口の開閉動作を判定する方法として
2 値化画像の平均・分散を用いる処理が有効であることが確認されたと考える。

4. まとめ

本研究では、人にやさしいインタフェースとして口唇の動きによりコンピュータに指示を与えるシステムに関して、顔画像より口唇部分を抽出する処理および口の開閉を判定する処理について提案した。また、実際の画像を用いた実験を行い、その有効性を評価した。

動きの判定処理として、差分画像による方法は動きが行われたことは判別できるが、細かな動きや開閉の順番には対応できなかった。従って、本研究で検討するシステムの判定処理として不十分であることがわかった。一方、2 値化画像を利用して画素値を重みとした平均と分

散を求め、画像間の差により判定する処理では、口の動きの有無に加え、開閉の順番が判別できること、首のうなずきや横振りも判定可能なことがわかった。この処理方法は、唇の完全な形状を抽出せず、おおまかな形の 2 値画像を重みとして行うため照明条件にある程度の影響を受けても、動きを判別することができる可能性を持っているなど、システムに適した処理と考えられる。

今後は読み込むフレーム画像を増やし、口の開閉の回数に基づく指示法を確立することが課題となる。

参考文献

- 1) 原島博, 黒川隆雄, “ノンバーバルインターフェース”, オーム社 (1994)
- 2) 井上哲理, 大嶺信孝, 関靖夫, “ヒューマンインターフェースのための手指屈伸認識”, 神奈川工科大学研究報告, B-21, 105-108 (1997)