

固有空間法と位相限定相関法による 動画像からの唇領域のロバストな検出

Robust Detection of Labial Region for a Moving Image
by Using Eigen Space Method and Phase Only Correlation Method

古賀 崇了¹ 古川 翔大² 内野 英治^{2,1} 末竹 規哲²

Takanori Koga Shota Furukawa Eiji Uchino Noriaki Suetake

¹ 財団法人ファジィシステム研究所

Fuzzy Logic Systems Institute

² 山口大学大学院理工学研究科

Graduate School of Science and Engineering, Yamaguchi University

Abstract: We propose a labial area detection method, which is robust against undesirable shifting of detected area, as a pre-processing of computer-vision-based lip reading system. In the computer-vision-based lip reading system, accurate detection of labial area is very important because accuracy of pattern classification of subsequent stage is significantly influenced by its area detection result. In the proposed method, an eigen space method using a phase only correlation as distance measure is used. The proposed method has advantages over some conventional methods in the labial area detection accuracy and the robustness for undesirable shifting of detected area.

1 はじめに

近年、音声認識技術の進歩により、様々な場面において音声認識が使われるようになった。しかしながら、高騒音環境下などで音声認識を行う場合、雑音の影響による識別精度の低下が問題となる。そのような環境において、唇領域の動画像情報を用いて発話内容を認識する読唇処理を用いる研究が報告されている [1]。

動画像認識による読唇処理は、顔画像中から唇領域を検出する処理と、動作特徴量の抽出および発話内容の認識を行う処理から構成される。特に、唇領域の検出結果が後段のパターン認識処理の性能を大きく左右するため、高精度な検出手法が必要となる。唇領域の抽出には、Snakes等の幾何学形状モデルベース法と固有空間法を用いるイメージベース法の2つのアプローチがある。

文献 [1] では、色情報を基にした幾何学形状モデルベース法とイメージベース法の二つを組み合わせることにより、それぞれの欠点を補い、高精度に唇領域を抽出する手法を提案している。しかしながら、この手法においても複雑に変化する唇の形状の変化に対して、抽出位置のずれが生じる場合がある。これは、テンプレートと動画像のマッチングを行う際に、ユークリッド距離に基づいた類似度を用いていることが原因であると考えられる。

本研究では、[1]における、固有空間法の一つである Eigentemplate 法を用いて唇領域を検出する際に、位相限定相関法 [2] を用いることを提案する。位相限定相関法とは画像の位相情報を用いて画像間の相関を調べる手法である。提案手法により、従来手法と比較して、唇領域の検出精度が向上することを実験により確認したので、ここに報告する。

2 唇領域の検出法

本研究では、文献 [1] の幾何学形状モデルベース法とイメージベース法を組み合わせる方法を唇領域抽出に用いる。抽出の手順を図1に示す。この手法では、色情報により唇領域を大まかに抽出した後、Eigentemplate 法により正確な唇領域の抽出を行う。

2.1 輝度正規化 RGB 値

顔画像から唇領域を検出する際、全ての領域でテンプレートマッチングを行うと、計算コストが膨大になる。したがって、図1(a)に示すように、テンプレートマッチングにおける探索領域を限定するために、肌の色情報と唇の色情報を用いて、顔画像中から大まかな唇周辺領域をあらかじめ検出しておく。ここでは、中田らによって提案された輝度正規化 RGB 値 [1] を用いる。輝度正規化 RGB 値とは各画素の R,G,B の値をその画素の輝度値 Y で割った値である。輝度正規化 RGB

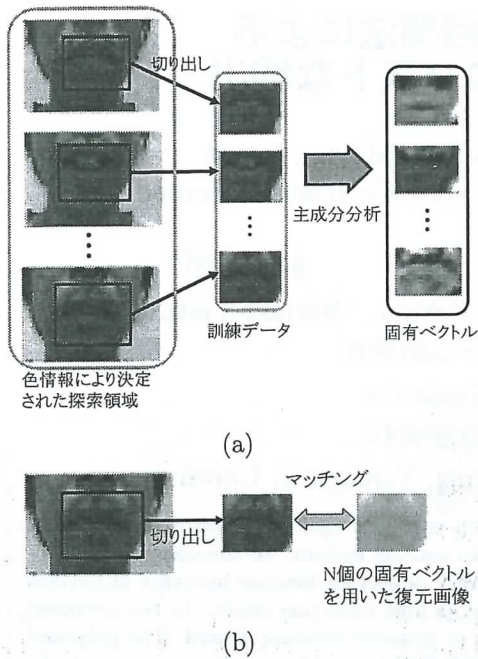


図 1: Eigentemplate 法による唇領域の抽出. (a) テンプレート作成のための固有画像作成の手順. (b) 実際の画像データに対するマッチングの手順.

値は、唇・肌などの同一の領域における画素値が輝度から受ける影響を低減でき、色情報のみによる領域分割がロバストになるという特長を有する。

2.2 Eigentemplate 法

唇領域の検出の際、単純なテンプレートマッチングを行おうとすると、様々な唇形状に対して大量のテンプレートが必要となる。Eigentemplate 法は基本的には画像と主成分分析により得られる固有ベクトルから復元される復元画像とのテンプレートマッチングである。

図 1(a) にテンプレートの作成手順を示す。まず、前述の色情報を用いて、口の端点座標と口の中心座標を算出し、求めた中心座標が画像の中心となるように一定サイズの画像を切り出す。ここで切り出した様々な形状をもつ唇画像に対して、主成分分析を適用し、以下の手順で固有ベクトルを計算する。

まず、切り出した k 番目の唇画像を 1 次元ベクトル x_k で表す。唇画像の大きさを $m \times n$ とすると、ベクトル x_k は i 行 j 列目の画素の輝度 $Y(i, j)$ を用いて次式で表される。

$$x_k = [Y(1,1) Y(1,2) \cdots Y(1,n) Y(2,1) \cdots Y(2,n) \cdots Y(m,1) \cdots Y(m,n)]^T \quad (1)$$

次に、画像の総枚数を k_{max} とし、すべての唇画像を

画像行列 A で表す。

$$A = [x_1 \ x_2 \ \cdots \ x_{k_{max}}] \quad (2)$$

次に、共分散行列 C を算出する。

$$C = \frac{1}{k_{max} - 1} \sum_{k=1}^{k_{max}} (x_k - \bar{a})(x_k - \bar{a})^T \quad (3)$$

ここで、 \bar{a} は k_{max} 個の画像の平均ベクトルである。

$$\bar{a} = \frac{1}{k_{max}} \sum_{k=1}^{k_{max}} x_k \quad (4)$$

このとき、固有ベクトル μ_k は固有値を λ_k とすると、以下の固有値問題を解くことで求まる。

$$C\mu_k = \lambda_k \mu_k \quad (5)$$

固有値 λ_k を値の大きい順に並べ、 λ_k に対応する固有ベクトル μ_k を N 個選択する ($N < k_{max}$)。固有ベクトルを画像に変換すると、図 1(a) に示す固有画像となる。

次に復元画像を用いて、口の位置を検出する方法について述べる。図 1(b) のように、探索対象の画像から、 $m \times n$ 画素の画像を切り出す。切り出した画像 y_k に対して、先に求めた固有画像を用いて復元画像を作成する。復元画像ベクトル t_k は、平均画像のベクトルを \bar{a} 、切り出した画像のベクトル y_k と固有ベクトル μ_i の内積を b_i とすると、次式で与えられる。

$$t_k = \bar{a} + \sum_{i=1}^N b_i \mu_i \quad (6)$$

ただし、 $b_i = y_k \cdot \mu_i$ 。このようにして求めた復元画像をテンプレートとして用い、任意の切り出し画像 y_k との類似度を算出する。任意の切り出し画像のうち、類似度が最大となるものを唇領域画像として抽出する。

3 提案手法

文献 [1] の手法では、任意の切り出し画像ベクトル y_k と復元画像ベクトル t_k とのユークリッド距離を基に、画像間の類似度を次式で算出している。

$$\xi = \frac{1}{|y_k - t_k|} \quad (7)$$

しかしながら、発話時と非発話時では唇の形状が大きく変化するため、正確な唇領域を取得することが難しい場合がある。

したがって、提案手法では、画像の位相情報に含まれる輪郭の情報を用いて相関係数を求める位相限定相

関法を類似度計算に用いることにより、唇領域の形状の変化に対して柔軟に対応し、唇領域を正確に抽出することを目指す。

3.1 位相限定相関法

位相限定相関法 [2] とは、画像の位相情報を使って画像間の相関を調べる方法である。位相限定相関は振幅スペクトルを定数に置き換えること以外は通常の相関の処理と同一である。

その手順は、まず2枚の同サイズの画像がある時に、それら2枚の画像に対してフーリエ変換を行う。得られた振幅スペクトルを定数に置き換えることにより位相情報のみを残す。2枚の位相情報のみの画像の内1枚は複素共役を取り、その2枚を掛け合わせることにより合成画像を得る。この合成画像を逆フーリエ変換し、2枚の画像同士の位相限定相関を求めることができる。2枚の $M \times L$ の画像を $f_1(m, l)$, $f_2(m, l)$ とする。ここで m, l は、 $m = 0, 1, 2, \dots, M-1$, $l = 0, 1, 2, \dots, L-1$ である。この $f_1(m, l)$, $f_2(m, l)$ の離散フーリエ変換を $F_1(u, v)$, $F_2(u, v)$ と表すと、

$$F_1(u, v) = \sum_{m=0}^{M-1} \sum_{l=0}^{L-1} f_1(m, l) e^{-j2\pi(mu/M+lv/L)} \quad (8)$$

$$= A(u, v) e^{j\theta(u, v)} \quad (9)$$

$$F_2(u, v) = \sum_{m=0}^{M-1} \sum_{l=0}^{L-1} f_2(m, l) e^{-j2\pi(mu/M+lv/L)} \quad (10)$$

$$= B(u, v) e^{j\theta(u, v)} \quad (11)$$

となる。ただし、 $u = 0, 1, 2, \dots, M-1$, $v = 0, 1, 2, \dots, L-1$ である。ここで、 $A(u, v)$, $B(u, v)$ は振幅スペクトルである。この振幅スペクトルを定数 (= 1) とすると、

$$F'_1(u, v) = e^{j\theta(u, v)} \quad (12)$$

$$F'_2(u, v) = e^{j\phi(u, v)} \quad (13)$$

となる。 $F'_2(u, v)$ の複素共役をとり、 $F'_1(u, v)$ と掛け合わせることで、合成画像 $H_{12}(u, v)$ を得る。

$$H_{12}(u, v) = F'_1(u, v) (F'_2(u, v))^* \quad (14)$$

$$= e^{j(\theta-\phi)} \quad (15)$$

この合成画像 $H_{12}(u, v)$ を逆フーリエ変換することにより、相関強度画像 $G_{12}(r, s)$ を得る。

$$G_{12}(r, s) = \sum_{u=0}^{M-1} \sum_{v=0}^{L-1} (H_{12}(u, v)) e^{j2\pi(ur/M+vs/L)} \quad (16)$$

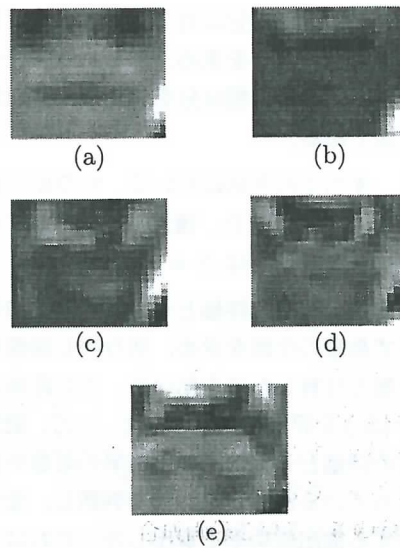


図 2: 窓幅 24×20 の実験で得られた固有画像。(a) 第1固有画像。(b) 第2固有画像。(c) 第3固有画像。(d) 第4固有画像。(e) 第5固有画像。

$$= \sum_{u=0}^{M-1} \sum_{v=0}^{L-1} (e^{j(\theta-\phi)}) e^{j2\pi(ur/M+vs/L)} \quad (17)$$

ただし、 $r = 0, 1, 2, \dots, M-1$, $s = 0, 1, 2, \dots, L-1$ である。ここで、逆フーリエ変換後に元の値に戻すために通常使われている定数 $1/ML$ は省略した。

4 実験

4.1 実験データ

実験データとして、一定の照明条件の下で、縦方向には目元からあごまでが写り、横方向には左右の頬が入るように撮影した動画画像を用いる。動画画像は、 320×240 [pixels] で撮影されたものを、 64×48 [pixels] に縮小したものを用いる。データ採取のための被験者は男性1名である。撮影時のフレームレートは 30 [fps] である。固有ベクトルを計算するために、「はい」、「いいえ」と繰り返し発話している状態を撮影した動画画像から 1900 フレーム分の画像を訓練データとして用いる。テストデータには、訓練データとは別の 500 フレーム分のデータを用いる。

4.2 実験内容

実験では、位相限定相関法を用いる提案手法の他に、文献 [1] の手法における色情報のみを用いた手法、文献 [1] の手法、類似度に周波数領域での相互相関を用いた手法、PCA 相関フィルタ [3] を用いた手法について唇領域の検出精度を比較した。PCA 相関フィルタは、

フーリエ変換後の位相成分に対して主成分分析を行うことにより固有ベクトルを求め、得られた固有ベクトルを用いて入力画像の位相成分を復元し、類似度の計算を行う手法である。

実験では、すべての手法において、切り出し画像の窓幅を 24×20 とした。また、復元画像を作成する際に用いる固有ベクトルの数は $N = 3$ とした。

水平方向の抽出精度の評価としては、抽出結果の画像から手で鼻頭の位置を求め、切り出し画像の中心と鼻との距離を計算した値を用いた。この距離の標準偏差の3倍 (3σ) を評価の基準とした。一方、垂直方向の抽出精度の評価としては、抽出結果の画像中に唇が十分に納まっているか否かを目視で判断し、全テストデータに対する抽出成功率を算出した。これは、口唇の動きにより口の中心位置がずれてしまい、水平方向と同様の評価基準を用いることができないためである。

4.3 実験結果

図2に示す固有画像に着目すると、唇の領域以外に頬の輪郭線の一部が特徴として表れていることがわかる。これにより、文献[1]の手法、相互相関および位相限定相関を用いた場合には、この頬の輪郭線情報が重要な特徴として作用し、図3(b)に示すように、 x 軸方向におけるずれが小さい値となっていると考えられる。

これに対してPCA相関フィルタでは画像の位相情報のみを用いて主成分分析を行うため、得られる固有ベクトルの性質が他の手法と異なる。すなわち、頬の輪郭線の情報が失われることで、 x 軸方向におけるずれが大きくなっていると考えられる。

また、図3(b)に示す y 軸方向における唇領域の検出成功率に着目すると、提案手法以外では x 軸のずれに関わらず、一様に低い値を示している。これは、提案手法以外では、頬の輪郭線を主要な特徴として領域を抽出していると考えられるため、唇領域を正確に検出する事が出来ていない。一方、PCA相関フィルタでは x 軸に対するずれは大きいがある程度唇領域の検出に成功している。これは、位相情報のみで主成分分析を行うため、頬の輪郭線の情報が失われ、唇の輪郭線の特徴をある程度捉えていることに起因すると考えられる。

提案手法では、位相限定相関法により画像の位相情報のみに着目することで、頬の輪郭線だけでなく唇の輪郭線が強調されたマッチングが行われており、高精度に唇領域を検出できていることがわかる。

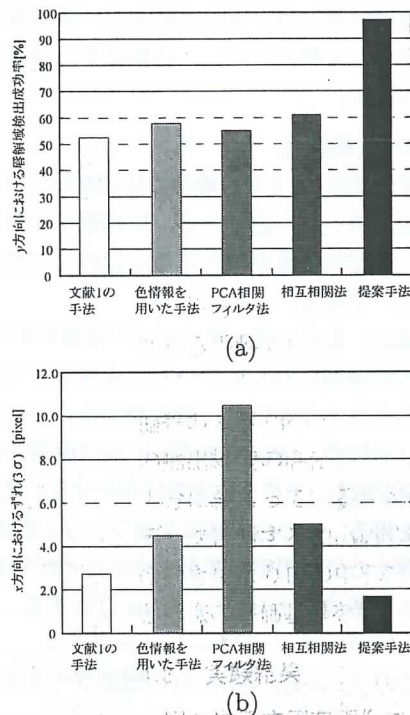


図3: 唇領域検出性能の評価。(a) x 方向に対する評価。(b) y 方向に対する評価。

5 おわりに

本研究では固有空間法を基にした Eigentemplate 法における類似度の計算に位相限定相関法を用いることにより、ロバストな唇領域検出法を提案した。実験結果より、既存の手法より高精度でロバストに唇領域を検出できることが確認された。

今後は、本提案手法を用いて得られる唇の動画像を用いて、動作特徴量の抽出および発話内容の認識を行う予定である。

参考文献

- [1] 中田他, “色抽出法と固有空間法を用いた読唇,” 信学論, Vol.J85-D-II, No.12, pp.1813-1822, 2002.
- [2] 小林他 “位相限定相関法の原理とその応用,” テレビジョン学会技報, Vol.20, No.41, pp.1-6, 1996.
- [3] 鈴木他, “PCA相関フィルタによる目領域の探索,” 第6回情報科学技術フォーラム予稿集, pp.37-38, 2007.

問合せ先

〒753-8512 山口県山口市吉田 1677-1
 山口大学大学院 理工学研究科
 内野英治 (うちの えいじ)
 Tel/Fax: 083-933-5699
 E-mail: uchino@yamaguchi-u.ac.jp