複数枚の顔写真を用いた発話による口の動きを持つ顔画像生成

高雷　向川康博　大田友一
筑波大学 電子・情報工学系

1　はじめに

発話による口の動きを持った顔画像を生成する技術は、相手の顔を見ながら会話できるマンマシンインターフェースや仮想俳優など様々なアプリケーションに用いることができる。しかし、発話による口の動きのモデル化は難しく、合成された顔画像は不自然になることが多かった。この問題に対して、本研究では、あらかじめ用意された複数枚の顔写真を組合せることで、発話による口の動きを持った顔画像を生成する手法を提案する。顔画像の生成が目的の場合には、発話における唇や口周辺部の動きのモデル化は必ずしも必要ではなく、複数枚の代表的な口の形の顔写真を適切に組み合せることで、発話中の任意の口の動きを画像として生成できれば十分である。

2　特徴点座標の算出

2.1　線形結合による特徴点の表現

口の形の異なる m 枚の顔写真 $B_j(1 \le \mathrm{j} \le \mathrm{m})$ が与えられ、顔表面上に配置された n 個の特徴点の画像間での対応付けがとれていると仮定し、これらをもとに様々な口の動きを持った顔画像を生成することを考える。顔画像 B_j 上での i 番目の特徴点の座標を (x_i^j, y_i^j) とする。 m 枚の入力画像上での n 個の特徴点の X,Y 座標を一定の順に並べたベクトルをそれぞれ $\boldsymbol{x}^j, \boldsymbol{y}^j$ とする。

$$
\begin{array}{rcl}
\boldsymbol{x}^j & = & [x_1^j, x_2^j, \ldots, x_n^j] \\
\boldsymbol{y}^j & = & [y_1^j, y_2^j, \ldots, y_n^j]
\end{array}
\tag{1}
$$

ここで、簡略化のため、平行移動成分は取り除いてあるものとする。任意の口の形は m 枚の入力画像の内挿・外挿で近似できるとすれば、任意の口の形を持った顔画像 $\hat{B}$ における特徴点座標を表すベクトル $\hat{x}, \hat{\boldsymbol{y}}$ は式 (2) のように線形結合で近似できる。これらの線形結合の係数 $c_1 \sim c_m$ を適切に決めることで、任意の口の動きを持った顔画像上のすべての特徴点の座標が求められる。

$$
\begin{array}{l}
\hat{\boldsymbol{x}} = \sum_{i=1}^m c_i \boldsymbol{x}^i \\
\hat{\boldsymbol{y}} = \sum_{i=1}^m c_i \boldsymbol{y}^i
\end{array}
\tag{2}
$$

しかし、入力画像の口の形によっては、特徴点座標ベクトルが１次従属なってしまう場合があり、線形結合の係数が一意に定まらず不安定になる。そこで、 m 枚の入力画像から得られる特徴点座標ベクトル $\boldsymbol{x}^1 \sim \boldsymbol{x}^m, \boldsymbol{y}^1 \sim \boldsymbol{y}^m$ の主成分分析を行う。得られた主成分 $\boldsymbol{P}^1 \sim \boldsymbol{P}^{2m}$ のうち、上位 k 個の主成分を用いて、式 (3) に示すように任意の口の形を持った顔画像 $\hat{B}$ における特徴点座標ベクトルを表現する。

$$
\begin{array}{l}
\hat{\boldsymbol{x}} = \sum_{l=1}^k C_l^x \boldsymbol{P}^l \\
\hat{\boldsymbol{y}} = \sum_{l=1}^k C_l^y \boldsymbol{P}^l
\end{array}
\tag{3}
$$

主成分分析によって、互いに１次独立な安定した基底ベクトルを用いることができ、寄与率の大きい上位の主成分のみを用いることで入力画像上での特徴点座標の誤差を低減することができる。

この方法では、生成できる口の形や口周辺部の動きの多様さは、入力として用いる画像にどれだけ多くの基底となりうる口の形が含まれているに依存する。つまり、より多くの入力画像を用いることで生成できる画像の種類や自然さが向上するということである。

2.2　線形結合の係数決定

生成したい口の形を指定し、その見え方に対応する線形結合の係数を決定するための一方法として、本研究ではサンプル画像を利用する。サンプル画像は、生成したい口の形を示す画像であり、この画像と同じ口の形を持った画像を新たに生成する。サンプル画像上では少数の代表的な特徴点が抽出されていれば十分であり、これらの点と入力画像上での同一の点が一致するような線形結合の係数を、最小自乗法によって求めることができる。

3　テクスチャの生成

3.1　重み付けテクスチャマッピング

前節の方法により、任意の口の形における特徴点座標が算出できる。次に、画像生成のために、特徴点間を埋めるテクスチャを生成する。テクスチャは特徴点を頂点とする３角形パッチごとに生成する。入力画像から対応する３角形領域のテクスチャを切り出し、生成する口の形にあわせてアフィン変換し、マッピングする。しかし、様々な口の動きを持った顔画像のテクスチャを生成するには、単一のテクスチャを無理に変形するより、もともと口の動きを持ったテクスチャを無理な変形なしに用いる方が、より自然な画像が得られる。そこで、本研究で

Synthesis of Face Images with Lip Motion from Several Photographs
Lei GAO, Yasuhiro MUKAIGAWA, Yuichi OHTA
University of Tsukuba