

テクスチャブレンドによる皺の表現と口形アニメーション

Making Wrinkles Using Texture-Blending and Mouth Shape Animation

石山 慎一郎 高橋 光紀 大橋 俊介 森島 繁生
Shinichiro ISHIYAMA Mitunori TAKAHASHI Syunsuke OHASHI Shigeo MORISHIMA

成蹊大学工学部
Faculty of Engineering, SEIKI University

1. はじめに

現在、映画など多くの分野でCGが利用されている。その技術は驚くべき速さで進歩している。その中でも特に人間の表情をリアルに再現することは非常に困難である。本稿では人間の顔面をより精密に再現し、さらにコミュニケーションツールへの応用を目指し、人間の発話アニメーションを作成する。

従来はワイヤフレームで構築された3次元顔面モデルを基本顔と呼ばれる表情ごとに用意し、その各頂点の座標、貼り付けるテクスチャの画素値をブレンドすることによって、様々な感情状態、表情を再現することが行われてきた。^[1]

本稿ではこのシステムを応用し、発話時の顔面のアニメーションの作成を行う。このために各VISEMEごとに3次元顔面モデルを用意し、これを組み合わせることによって顔面、特に口形のアニメーションを作成し、人間の発話のリアルな再現を試みる。

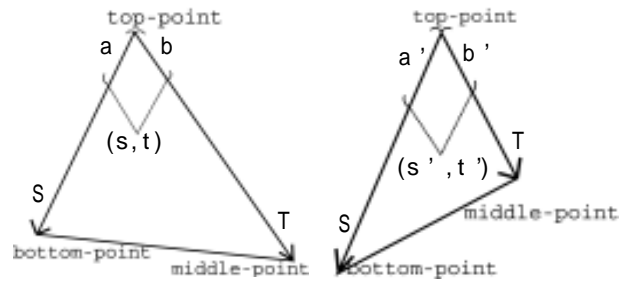
2. 3次元顔面モデル

本稿ではより実物に近いモデルを構築するためにレンジファインダ(製造: NEC、名称: Danae-R)を用いて、顔面の3次元データを取得している。従来までは作り込まれたモデルで、人物変更が不可能であったが、このレンジファインダの使用により人物変更が可能になっただけでなく、より正確な顔面の3次元情報を取得することが可能となった。しかしこの取得した3次元データは多数の3次元空間上の点群でしかないため、このデータに汎用ワイヤフレームを整合することにより、実物の構造を反映した個人ワイヤフレームモデルを構築する。

またワイヤフレームにマッピングするテクスチャはより高精細なものを取得するため、デジタルカメラを用いて3次元構造のデータとは別にテクスチャのデータを取得している。このテクスチャの撮影の際には取得するテクスチャのデータに撮影時の環境等の影響が反映されないよう、照明環境を配慮して撮影を行っている。



a) 整合されたワイヤフレーム b) 3次元顔面モデル
図1. 3次元顔面モデル



a) 無表情モデルのポリゴン b) 基本顔モデルのポリゴン

図2. 3次元顔面モデルのポリゴンの例



a) 無表情 b) 正規化前 c) 正規化後

図3. テクスチャの正規化

3. テクスチャブレンド

本稿では、ワイヤフレームの各頂点のブレンドを行う割合(以下ブレンド率と呼ぶ。)に伴い、各テクスチャの画素値もブレンドする。このブレンドを行ったテクスチャをマッピングすることにより実物に近い肌の質感の再現、特に皺の表現を可能にした。

このテクスチャのブレンドを行う際にはブレンドを行う各テクスチャの形、大きさを無表情のものへと統一する正規化を行っている。本稿で用いているワイヤフレームは全て三角形のポリゴンで構成されている。テクスチャの正規化とはこのポリゴン内のテクスチャ座標を対応させることによって行われている。図2に示すようなポリゴン内の任意のテクスチャの座標 (s, t)



a) ブレンドあり b) ブレンドなし

図4. テクスチャブレンド

はポリゴンを構成する2辺の比で表すことができる。従って形状が異なるポリゴンに対して正規化を行うことができる。

画素値のブレンディングにはRGB値を用いている。使用するテクスチャ(512×512pixels)のすべてのピクセルに関してRGB値をそれぞれブレンディングを行うことによってマッピングするためのテクスチャが完成する。

このテクスチャのブレンディングをワイヤフレームモデルの形状が変化するたびに繰り返しマッピングし、よりリアルな3次元顔面モデルの構築を実現している。

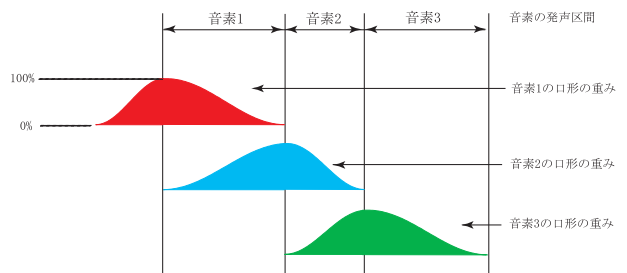


図 .5 正弦波補間の概念図

4. 顔面モデルと口内モデルの構築

本稿ではVISEMEを用いて発話アニメーションを作成している。VISEMEとは母音、子音といった音素のうち、音が異なっても、発音したときの口形が同じであれば同一のグループに分類する方法である。本稿ではこのVISEMEごとに3次元構造とテクスチャのデータを取得、ワイヤフレームのフィッティングを行い3次元顔面モデルを構築した。また3次元顔面モデルだけでなく、口内、歯、舌の3次元モデルも構築している。これらの3次元モデルもワイヤフレームにより構築されている。この3次元モデルは口形の変化、すなわち口の部分のワイヤフレームの頂点の移動に対応して、変形、移動する。

本稿ではこうして構築した3次元顔面モデルと歯、舌を含めた口内モデルを用いて発話のアニメーションを作成している。本稿ではリップシンクさせる音声波形のセグメンテーションを行うことにより、各VISEMEの継続長を取得した。口形のアニメーションは各VISEMEの先頭フレームにキーフレームを配置して、遷移する2つの口形のブレンディング率を徐々に変化させることによって作成する。このキーフレーム間の補間法は主観評価によって最も自然とされる正弦波補間を用いることで、より実際の発話に近い口形の変化を実現している。



図4. 口形アニメーション(無表情からu:へ)

5. まとめ

本稿ではリアルな発話アニメーションの実現方法を提案した。レンジファインダを用いることによって、対象の人物により近い印象を実現する顔面モデルの構築が可能となった。またレンジファインダの使用によって、奥行き精度の向上を実現しただけではなく、3次元構造とテクスチャのデータを入れ替える事により、カスタマイズも可能とし、システムとしての汎用性を向上させた。またテクスチャブレンディングを行うことにより、皺など肌の質感の表現を向上させた。

しかしながら本稿ではリップシンクする音声波形を手動によりセグメンテーションを行っているため、発話内容等の変更が不可能である。今後の課題としては、自動セグメンテーションの導入、またそれによる汎用性の向上が挙げられる。

参考文献

[1]柳澤、高橋、森島, "テクスチャブレンディングによる皺の表現と基本顔モデルによる感情空間の構築", 電子情報通信学会技術研究報告, Vol.101, No.693, HCS2001-49, pp.17-24, March 2002.

[2]鳥飼、伊藤、緒方、森島, "仮想人物の舌モデル構成と発話アニメーション作成", 電子情報通信学会2002年総合全国大会, A-16-34, March 2002.

[3]伊藤、森島, "複数アングル画像とレンジファインダを用いた3次元顔モデルの生成とその表情合成", 電子情報通信学会2002年総合全国大会, A-15-5, March 2002.

VISEME No.	音素表記
1	/æ/
2	/ʌ/ /ɑ/
3	/a/
4	/a:/
5	/ə:r/
6	/i/ /i:/
7	/u/
8	/u:/
9	/e/
10	/ɔ:/
11	/ə:r/
12	/l/
13	/r/
14	/p/ /b/
15	/t/
16	/d/ /n/
17	/k/ /g/ /ŋ/ /h/
18	/f/ /v/
19	/s/ /z/ /ʃ/ /ʒ/ /tʃ/ /dʒ/
20	/ / / / δ/
21	/j/
22	/w/

表1. 英語の発音記号とVISEMEの対応表