

テレビ視聴者の興味状態推定に向けた顔表情変化度測定

Estimating Facial Expression Intensity for Inferring Status of TV Viewer's Interest

奥田 誠†
Makoto Okuda

苗村 昌秀†
Masahide Naemura

藤井 真人†
Mahito Fujii

佐藤 洋一†
Yoichi Sato

1. まえがき

近年、インターネットの普及に伴い、視聴したテレビ番組についての情報(出演者の紹介など)や、過去に放送されたテレビ番組を、通信を介して取得できるようになった。しかし、インターネット上の情報は膨大であり、必要とするコンテンツを、誰もが容易に取得できるわけではない。

このような背景の中、筆者らは、番組シーン毎に視聴者の興味状態を推定することで、各人の趣味・嗜好に応じたコンテンツ推薦を実現しようとしている。

視聴者が、笑ったり、驚いたりして、顔表情を変化させたときには、番組への興味度が高いと考えられる。本論文では、顔表情の変化度合を測定する手法について提案する。

2. 顔表情変化度合測定システム

2.1 開発方針

家庭環境で、視聴者の興味状態を推定することを想定すると、視聴者に接触センサを取り付けることは、現実的ではない。そこで、図1に示すように、テレビに単眼カメラを設置し、画像処理を行って視聴者の顔表情変化度を測定する。

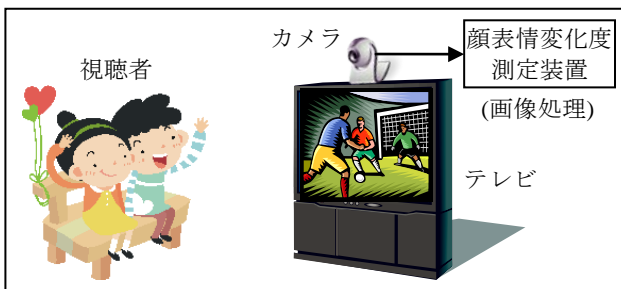


図1 視聴者の顔表情変化度測定イメージ

顔表情変化度の測定は、事前に、さまざまな人の顔表情を機械学習しておくことにより、視聴者が、顔表情の登録を行うなどの煩わしい作業を行う必要がないようにする。

これまで、Littlewort[1]らが、顔画像を Support Vector Machines(SVM)[2]により表情分類し、画像特徴量と SVM 境界面との距離を表情変化の大きさとすることを提案している。筆者らも、これに倣うが、システムを家庭環境で用いることを勘案し、照明変化にロバストなどの特徴をもつ画像特徴量を用いる。

2.2 システム

2.2.1 学習データ

†日本放送協会, Japan Broadcasting Corporation

‡東京大学, The University of Tokyo

顔表情の学習には、Extended Cohn-Kanade Dataset[3]を用いた。

Extended Cohn-Kanade Dataset には、123 人の人物が、ニュートラルな状態から、徐々にピーク方向へ向かって顔表情を変化させた 593 の画像シーケンスが存在する。これらのうち、Anger, Disgust, Fear, Happiness, Sadness, Surprise の 6 表情がラベリングされた 309 シーケンスについて、シーケンスの最初の画像(ニュートラルな表情)と最後の画像(ピーク表情)を学習データとした。

2.2.2 顔表情の学習

顔表情の学習手順を、以下に示す。

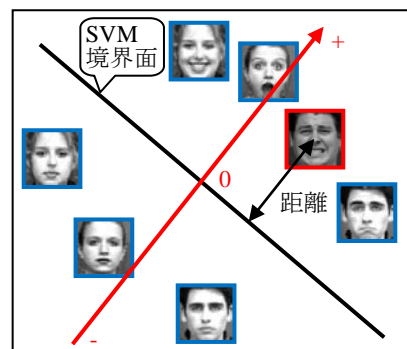
1. すべての学習画像について、顔領域の切り出し[4]とサイズの正規化を行った後、画像特徴量 Bag-of-Keypoints[5]を計算する。
2. 計算した画像特徴量 Bag-of-Keypoints に基づき、ニュートラル表情とピーク表情を分類できるように SVM による機械学習を行う。

Bag-of-Keypoints は、近年、一般物体認識で活発に利用されている画像特徴量であるが、Li ら[6]が、これを用いることにより、顔画像を Anger, Disgust, Fear, Happiness, Sadness, Surprise の 6 表情へ高精度に分類できることを発表している。このことから、Bag-of-Keypoints は、顔表情を識別するための情報を多く含んでいると考えられ、さらに、照明変化にロバストなどの利点をもつため、使用することにした。

2.2.3 顔表情変化度の測定

顔表情変化度の測定手順を、以下に示す。

1. 測定する画像について、顔領域の切り出しとサイズの正規化を行った後、画像特徴量 Bag-of-Keypoints を計算する。
2. 計算した Bag-of-Keypoints について、学習により得た SVM 境界面からの距離を計算し、ピーク表情側を正、ニュートラル表情側を負とした値を顔表情変化度とする(図2)。



※ 青枠画像: 学習データ, 赤枠画像: 測定データ

図2 顔表情変化度の計算

3. 評価実験

3.1 識別率

提案した顔表情変化度測定手法の妥当性を測る1つの目安として、まず、学習したSVMがニュートラル表情とピーク表情をどれだけ識別できるかをテストした。

顔領域は、半径160ピクセルの円に正規化した(図3)。Bag-of-Keypointsは、位置情報を導入するため、顔領域の上半分と下半分で別々に計算し(それぞれ350次元ベクトル, 250次元ベクトル), 各ベクトルを繋いで1つのBag-of-Keypoints(600次元ベクトル)とした。

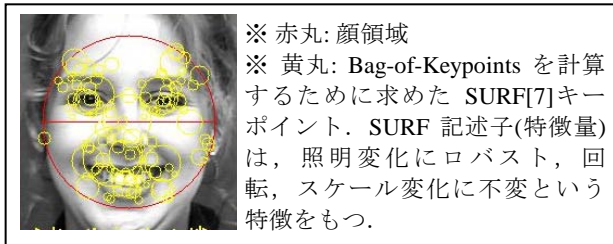


図3 顔領域の処理

Extended Cohn-Kanade Datasetのニュートラル表情309画像、ピーク表情309画像について、Leave-one-out法によりテストを行った結果、識別率は86.2%であった。

3.2 顔表情変化度測定テスト

Extended Cohn-Kanade datasetの各顔表情変化シーケンスについて、顔表情変化度の測定を行った。

図4は、被写体S052, S055, S074, S111, S113のSurpriseシーケンスについて、顔表情変化度を測定した結果である。図5は、被写体S055の各表情シーケンスについて、顔表情変化度を測定した結果である。いずれも、テスト画像は、学習には用いていない。

図4, 図5を見ると、顔表情が変化する(画像番号が増加する)に従い、顔表情変化度が増加する傾向にあり、良好な結果が得られた。

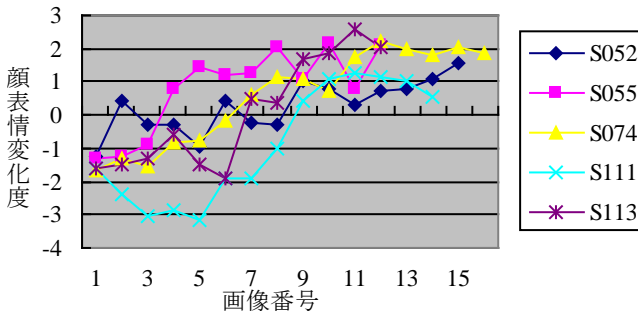


図4 Surpriseの顔表情変化度の推移

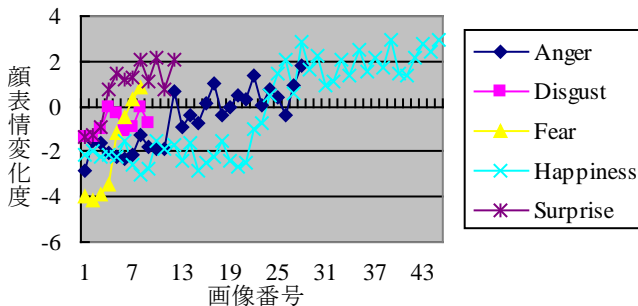


図5 被写体S055の顔表情変化度の推移

実際の家庭内でもテレビ視聴者を撮影し、顔表情変化度を測定した。図6は、視聴者がニュートラルな状態から、徐々に表情を変化させていった約10秒間を切り出した結果である。

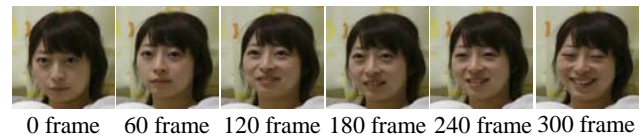
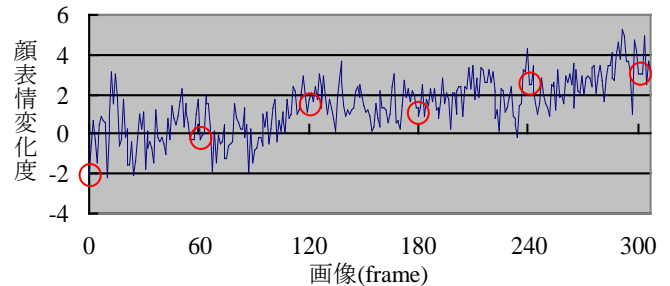


図6 視聴者の顔表情変化度の推移

フレーム間の変化度の変動が激しく、課題も残るが、学習データと異なる環境で撮影した画像でのテスト結果であるにも関わらず、視聴者の顔表情変化に合わせて、顔表情変化度が増加する傾向にあると言える。

解像度720×480の映像に対する処理速度は、約80ms/frame(CPU: Intel Xeon 3.20GHz, メモリ: 4GB, OS: Windows XP Professional)であった。

4. まとめ

テレビ視聴者の興味状態推定システム開発に向けて、顔表情変化度測定手法を提案した。

顔表情データセットを用いた評価実験を行い、顔表情変化が大きくなると、顔表情変化度の計算結果も大きくなる傾向にあることを確認した。また、家庭環境でもテストを行い、同様の結果を得た。

今後、提案手法について、さらに詳細な評価実験を行い、システムの改善を図っていく予定である。

参考文献

- [1] G. Littlewort *et al.*: "Dynamics of facial expression extracted automatically from video", *IVC*, Vol. 24, No. 6, pp. 615-625 (2006)
- [2] C. Cortes *et al.*: "Support-Vector Networks", *ML*, Vol. 20, No. 3, pp. 273-297 (1995)
- [3] P. Lucey *et al.*: "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression", *CVPR4HB*, pp. 94-101 (2010)
- [4] P. Viola *et al.*: "Robust Real-Time Face Detection", *IJCV*, Vol. 57, No. 2, pp. 137-154 (2004)
- [5] G. Ssurka *et al.*: "Visual Categorization with Bags of Keypoints", *ECCV Workshop on SLCV*, pp. 1-22 (2004)
- [6] Z. Li *et al.*: "Facial Expression Recognition Using Facial-component-based Bag of Words and PHOG Descriptors", *ITE*, Vol. 64, No. 2, pp. 230-236 (2010)
- [7] H. Bay *et al.*: "SURF: Speeded Up Robust Features", *CVIU*, Vol. 110, No. 3, pp. 346-359 (2008)