J-002

ユーザの表情に基づく映像コンテンツへのタギング
Tagging for Video Contents Based on User's Facial Expression

宮原 正典 † 青木 政樹 † 滝口 哲也 ‡ 有木 康雄 ‡
Masanori Miyahara Masaki Aoki Tetsuya Takiguchi Yasuo Ariki

1. はじめに

近年，テレビでは多チャンネル化が進み，またインターネットでは，YouTube などに代表される動画共有サイトが発達してきたこともあり，ユーザが視聴できる映像コンテンツは莫大な量になっている．これに伴ってユーザが自分の見たい番組を簡単に探すのが困難になってきている．この問題に対して，ユーザの視聴行動をセンシングして，コンテンツへの関心度を推定し，それに基づき興味の発見・番組の推薦をするアプローチがある．視聴行動に関しては，リモコン操作履歴，顔方向，顔表情，視聴動作などを利用するものがある [1][2] が，本研究では，この中で顔表情に着目し，従来の手法よりも詳細な関心推定を行い，コンテンツにタギングする手法を提案する．

2. 提案システムの概要

本研究では，まず，図 1 に示すような，PC のディスプレイに映るコンテンツをユーザが 1 人で視聴している実験環境を構築した．ウェブカメラはユーザの顔を撮影していて，PC はコンテンツの再生とユーザの顔動画の解析処理を行う．

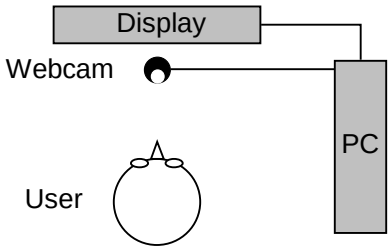

図 1: Top view of experimental environment

次に，ウェブカメラで撮影された顔動画の解析処理の流れを図 2 に示す．まず，顔動画から Haar-like 特徴を用いた AdaBoost 法 [3] によって，正確な顔領域を切り出す．そして切り出された顔領域に対して Gabor 特徴を用いた Elastic Bunch Graph Matching により，顔の特徴点座標を抽出する．あらかじめ用意されたユーザの無表情顔画像から抽出された特徴点座標と，毎フレーム抽出される特徴点座標の差分を特徴量として，Support Vector Machines で顔表情認識を行う．認識結果はコンテンツのフレームと同期して表情タグとして保存される．

†神戸大学大学院工学研究科
‡神戸大学自然科学系先端融合研究環

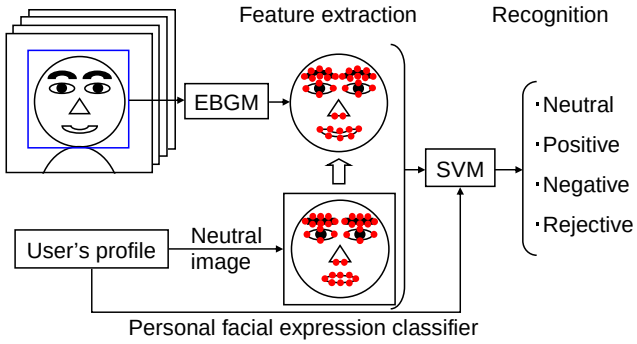

図 2: System flow

3. 提案システムの手法

3.1 EBGM による特徴点抽出

Elastic Bunch Graph Matching(EBGM)[4] は，まず，画像を様々な周波数と方向を持った Gabor フィルターで畳み込み，それらの応答の集合を Jet とし，Bunch Graph と呼ばれるモデルを作成する．この Bunch Graph は，複数人のデータから作成しておく．次に，Bunch Graph と，入力画像の各特徴点の Jet との間で類似度を計算し，特徴点の座標を推定する．本研究では，Bunch Graph の特徴点は，図 2，図 3 中に示すような 34 点を用いた．

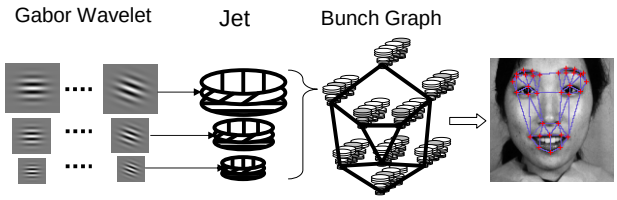

図 3: Elastic Bunch Graph Matching

3.2 SVM による顔表情認識

Support Vector Machines(SVM)[5] は，特徴空間における線形識別関数をマージン最大化基準で構成する手法であり，本論文では RBF カーネルを用いてカーネル化を行った．また，SVM は 2 クラスの識別器を構成する手法なので，One-Versus-Rest 法によって多クラスの識別器に拡張した．そして，EBGM によって抽出された 34 点の顔特徴点座標と，あらかじめ保持しておいた本人の無表情画像から抽出した顔特徴点座標の差分を取ることにより，68 次元の顔特徴点移動量ベクトルを求め，これを特徴ベクトルとした．

3.3 顔表情認識のクラス

顔表情認識のクラスとして，従来では関心の有無という 2 クラスで分類していたが，関心の中にも肯定的な関心と否定的な関心があると考え，より詳細な分類ができるように，Neutral，Positive，Negative という 3 クラスを用いた．また，表情変化ではないが，ユーザが正面を向いていない，顔の一部が隠れているなどの状態を表す，Rejective というクラスも用意した．各クラスの分類の詳細を表 1 に示す．

表 1: Facial expression classes

クラス	内容
Neutral(Neu)	無表情
Positive(Pos)	喜び，笑い，快，など
Negative(Neg)	怒り，嫌悪，不快，など
Rejective(Rej)	画面に顔を向けていない，顔の一部が隠れている，顔が傾いている，など

4. 実験

4.1 実験条件

図 1 に示すような実験環境を用いて，被験者 1 名に 1 回約 17 分間の映像コンテンツを 4 回分視聴させた．この際，被験者の顔動画を映像コンテンツと同期させながら，毎秒 15 フレームで記録した．その後，被験者に自分の顔動画と映像コンテンツを並べて見せ，表 1 に基づくタギングをさせた．タギングは，顔動画再生中に，各クラス指定のボタンを押すことで行う．ただし，何も押していないフレームは Neutral に分類される．そして，EBGM により抽出された特徴ベクトルのうち，コンテンツ 3 回分を SVM に学習させ，残りの 1 回分について認識を行い，その認識率を求めた．

4.2 実験結果

認識を行った全フレームに対して，フレームごとの認識結果を表す Confusion matrix を求めた．結果を表 2 に示す．

表 2: Confusion matrix

	Neu	Pos	Neg	Rej	sum	recall(%)
Neu	48275	443	525	622	49865	96.81
Pos	743	6907	1	14	7665	90.11
Neg	356	107	3250	6	3719	87.39
Rej	135	0	5	1326	1466	90.45
sum	49509	7457	3781	1968	62715	
precision(%)	97.51	92.62	85.96	67.38		

また，連続したフレームでの認識結果の例を図 4 に示す．

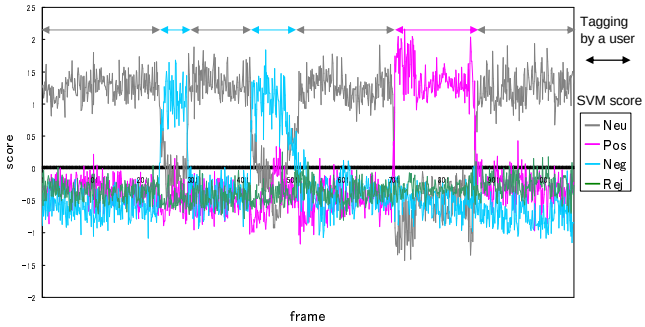

図 4: Example of recognition result

4.3 考察

各クラスについて，平均再現率は 91.19%，平均適合率は 85.87%となった．被験者が Positive や Negative と答えていても，表情表出の度合いが比較的小さい場合に，システムが Neutral と誤認識してしまうパターンが多かった．また，今回は 1 つのフレームにクラスは 1 つしかないと仮定しているので，中間的な表情をしている場合には誤認識が多く発生していた．

5. おわりに

本論文では，コンテンツを視聴するユーザの顔表情を認識することにより，コンテンツに自動的にタギングを行うシステムの提案と評価を行った．これにより，ユーザの興味区間を発見し，番組推薦に役立てることが可能となった．今後は事前学習なしに複数人の表情を認識したり，あるいは複数人同時にコンテンツを見ていても認識ができるように改良していく予定である．また，感情以外の様々なマルチモーダル情報も組み合わせて，ユーザに自動的に番組を推薦するシステムを構築することも検討中である．

参考文献

[1] 山本誠，新田直子，馬場口登，“個人的選好獲得のためのテレビ視聴時における興味区間の推定，” 画像の認識・理解シンポジウム (MIRU2006)，pp.37-42，2006

[2] 清水正浩，岩田満，田野俊一，“テレビ視聴時のマルチモーダル情報を利用した関心度推定システムの実現と評価，” ヒューマンインターフェース学会論文誌，Vol.5，No.1，2003

[3] P. Viola, M. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features,” In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, USA, pp.1-9, 2001

[4] Laurenz Wiskott, Jean-Marc Fellous, Norbert Kruger, Christoph von der Malsburg, “Face Recognition by Elastic Bunch Graph Matching,” IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL.19, NO.7, pp.775-779, JULY, 1997

[5] Vladimir N. Vapnik, “The Nature of Statistical Learning Theory,” Springer (1995)