

複数人物の顔方向・表情認識に基づく 映像中の観客の満足度の自動定量評価

松川 徹^{†1} 日高章理^{†2} 栗田多喜夫^{†3,†1}

スポーツやお笑いなどのエンタテインメント産業では、観客の観覧内容に対する満足度を知ることがサービスの質を評価するうえで重要となる。通常は観客に対するアンケート調査などで満足度評価を行うが、そのような方法では多くの人手や費用が必要となる。そのため観客の満足度調査をビデオ映像などから自動的に行う手法の開発が求められている。本論文では顔認識技術によって観客の表情や顔の向きなどの情報を自動的に取得し、それらに基づいて観客の満足度を機械的に推定するシステムを提案する。提案システムでは観客が映っているシーンからまず観客の顔の検出と向きの推定を行い、それから表情識別器によって検出された顔を笑顔と非笑顔に分類する。次に、あるシーンから検出された顔の向きと表情の組合せの出現頻度を数えあげたヒストグラムを作成する。このヒストグラムを特徴ベクトルとした識別器を用いて各シーンの状況（観客が喜んでいるシーンかどうかなど）を判別し、“喜んでいる”または“真剣に観戦している”度合いとしての満足度の推定を行う。実際のスポーツ観戦動画をを用いた実験により、提案するヒストグラム特徴でサポートベクタマシンを学習し、観客が“喜んでいる”シーンかどうか、“真剣に観戦している”シーンかどうかなどの判別およびその満足度推定を有効に行えることを確認した。

Automatically Evaluation of Degree of Spectators' Satisfaction in Video Sequences Based on Their Facial Expressions and Face Directions

TETSU MATSUKAWA,^{†1} AKINORI HIDAKA^{†2}
and TAKIO KURITA^{†3,†1}

In the entertainment industry which treats sports or comedy show, understanding whether their spectators have been satisfied or not is important to evaluate the quality of their services. Currently, questionnaire survey has been used to evaluate the degree of their satisfaction. But great cares of cost and time are required for questionnaire survey. So it is desired to automatically evaluate spectator's satisfaction degree from video sequences. This paper presents

a system for automatically evaluation of degree of spectators' satisfaction in video sequences. The proposed system is designed by a “bag-of-visual-words” approach based on face recognitions. First, the multiview (left-profile, front, right-profile) faces are detected from each image in the given video sequence. Then the detected faces are classified into the two expressions, smile or not. The classification results of the face directions and the facial expressions are voted to each classes' histogram over the video sequence. Finally, the state of the spectators is classified by using the kernel SVM on the voted histograms. The degree of spectators' satisfaction is estimated by the classification score of “Positive Scene” or “Watching Seriously”. Our approach demonstrated promising results for classifications of “Positive Scene” and “Negative Scene” or “Watching Seriously” and “Not Watching Seriously”.

1. はじめに

スポーツやお笑いなどのエンタテインメント産業では、観客の観覧内容に対する満足度を知ることがサービスの質を評価するうえで重要となる。通常は観客に対するアンケート調査などで満足度評価を行うが、そのような方法では多くの人手や費用が必要となる。そのため観客の満足度調査をビデオ映像などから自動的に行う手法の開発が求められている。そこで本論文では、映像中から観客の顔の向きと表情をパターン認識手法によって自動的に認識し、その認識結果に基づいて撮影されている観客の満足度を推定するシステムを提案する。

現在、画像から顔を検出する手法は数多く提案されている。Viola らは、ブースティングにより特徴選択した弱識別器をカスケード型に接続し、インテグラルイメージを用いて Harr-like 特徴を高速に計算することにより高速かつ高精度に顔を検出する手法を提案した¹⁾。また、Haung らは、Viola らの手法を Width-First-Search tree structure と sparse features in granular space を用いて異なる向きの顔を検出する手法を提案した²⁾。また、Hidaka らは粒子群最適化法を用いることで、Viola らの手法の学習時間を高速化する手法を提案⁵⁾しているなど、Viola らの顔検出手法には数多くの拡張が提案されおり、現在では

^{†1} 筑波大学大学院システム情報工学研究科

Graduate School of Systems and Information Engineering, University of Tsukuba

^{†2} 東京電機大学理工学部理学系

Division of Science, School of Science and Engineering, Tokyo Denki University

^{†3} 産業技術総合研究所脳神経情報研究部門

Neuroscience Research Institute, National Institute of Advanced Industrial Science and Technology

顔を検出することは容易に行える状況である。

画像から顔表情を認識する研究も数多く行われている。Shinohara らは、フィッシャー判別基準に基づく局所特徴の最適な重みマップを提案した⁶⁾。Hu らは、異なる向きの顔表情を認識するための数種類の局所特徴と次元圧縮法を比較した⁷⁾。Chen らは、選択された Harr-like 特徴と Gabor 特徴を用いて顔の検出と顔表情認識を一体化した⁸⁾。

このように、顔検出や表情認識の手法はこれまで数多く提案されているが、スポーツ観戦で多数の観客がどの程度試合を楽しんでいるかの度合いといった群衆全体の満足度を顔の情報に基づいて推定する研究はあまり行われていない。

そこで本論文では観客の顔の向きと表情に着目した満足度推定法を提案する。提案システムではまず顔検出器によって動画の中から観客の顔を複数見つけだし、次に表情識別器によってそれらの顔の表情を認識する。この作業を顔の向き（左向き、正面向き、右向き）ごとに個別に行った後、認識された顔の向きと表情の頻度を数えあげてヒストグラム特徴を生成する。これは顔識別器（顔検出器 + 表情識別器）の識別結果に基づいて画像パッチを量子化することに相当し、すなわち提案特徴は顔識別器ベースの bag-of-visual-words 法に基づいているといえる。本論文では、この提案特徴をもとにサポートベクタマシンなどの識別器を訓練することで、シーンに含まれる観客の顔の向きと表情の出現頻度から提供サービスに対する観客の満足度を自動的に推定するシステムを構築することを目指す。

ここで観客の顔情報と観覧内容に対する満足度をどのように結びつけるかが問題となる。このような満足度は提供されるサービスの分野や内容に応じて適切に定めるべきである。本論文ではスポーツ観戦時の観客の満足度推定という問題を取り扱うため、次の 2 つの指標を満足度の定義として用いる。1 つ目は“観客の喜び度合い”であり、これは多くの観客が笑顔などの喜ばしげな表情をしているときを満足状態と見なすものである。2 つ目は“観客の真剣さの度合い”であり、これは多くの観客が真剣な面持ちで一心不乱に試合を観戦しているときを満足状態と見なすものである。これらの満足度は、提案する顔の向き・表情のヒストグラム特徴で構築した SVM の識別面と入力特徴の間の距離として定量的に計算可能である。

近年、映像コンテンツを視聴するユーザの表情を解析することで、ユーザの映像コンテンツに対する関心度を推定する研究が行われている^{17),18)} が、彼らのシステムは 1 人の人間が正面を見ている状況を仮定し、1 人の人間の関心度を顔表情認識に基づいて推定するものである。これらに対する提案システムの相違点は、撮影されている多数の人間の表情の満足度を推定する点であり、正面向きに加えて横向きの顔認識器も用いることで撮影状況の制約が

低く、また、顔の向きも満足度推定に利用することで顔表情認識のみよりも豊富な情報を用いることができるという利点が存在する。

以下、2 章で提案システムの概要について説明し、3, 4, 5 章で提案手法の詳細について述べ、6 章でシステムの画面例を示す。7 章では実験結果を示し、8 章ではまとめを述べる。

2. 提案システムの概要

2.1 提案システムの扱う問題

提案システムは、観客映像に対して機械学習により満足度を推定するシステムである。提案システムを訓練するにはあらかじめ正解となる満足度が付与された学習用の映像データセット $\{x_i, y_i\}_{1:N}$ が必要となる (N はサンプル数)。ここで x_i は観客動画中の連続する M フレーム (1 フレームは $W \times H$ ピクセル) からなる画像シーケンスである (本研究では具体的には 30 フレーム、 720×480 ピクセルとしている)。また y_i は画像シーケンス x_i 内の観客が観覧対象に感じている“満足度”であり、訓練データの作成者が手動で与えるものである。

ここで画像シーケンスの満足度をどのように与えるかが問題となる。仮にその道の専門家などが訓練用シーケンス x_i を観て観客の満足度 y_i を 0.8, 0, 2 などのように実数値で与えることが可能な場合、この問題を回帰問題の枠組みで取り扱うことが可能になる。またそのような複雑な定量化が難しい場合には、訓練用シーケンス x_i に対して単純に“満足している ($y_i = +1$)”か“満足していない ($y_i = -1$)”かというラベルを割り当てる方法もある。この場合は満足度の推定問題を 2 クラスの識別問題として取り扱い、観客の満足度合いとしては識別時のスコア (識別境界からの距離など) を割り当てればよいと考えられる。本論文では後者の方法で満足度推定を行う。

何をもちて観客が満足しているか見なすかは、ユーザの目的や応用事例によって異なるものと考えられる。本論文では簡単な例として、“観客が満足している” = “観客が喜んでいる”と見なす場合 (満足度 A)、および“観客が満足している” = “観客が真剣に観戦している”と見なす場合 (満足度 B) の 2 例を取り扱う。すなわち前者のケースでは観客が喜んでいる (と訓練データ作成者が判断したシーケンス) に満足度 $y_i = +1$ が割り当てられ、逆に観客が喜んでいないシーケンスには満足度 $y_i = -1$ が割り当てられる。同様に後者のケースでは観客が真剣に観戦しているシーケンスとそうでないシーケンスにそれぞれ満足度 $y_i = +1$ と $y_i = -1$ を割り当てる。図 1 に対象シーンの例を示す。満足度推定方法の詳細は 5.3 節に述べる。



図 1 観客の画像例。(a): “喜んでいる”, (b): “喜んでいない” - “真剣に観戦していない”, (c): “喜んでいない” - “真剣に観戦している”. これらは説明用であり, 実験には 2008 年北京オリンピックの観客映像を用いた
Fig. 1 Example of spectators' scene. (a): “Positive Scene”, (b): “Negative Scene” - “Not Watching Seriously”, (c): “Negative Scene” - “Watching Seriously”. These are for explanation and not the actual Olympic scenes used in this paper.

本論文で実際に扱う観客動画は, スポーツを観戦している観客の動画である. 観客動画は, 2008 年北京オリンピックの TV 放映映像から, 観客が映る場面を手動でサーチして切り出した映像を用いた. 映像は 30 fps のビデオレートで撮影されている. ここでは “観客が喜んでいる” に対応するデータとして, 主に映っている観客の選手が勝ったときや優勢な状況の観客映像を 50 サンプル用意した. “喜んでいない” に対応するデータとして, 競技が始まっていないときや主に映っている観客の国の選手が優勢でない状況の観客映像を 50 サンプル用意した. なおスポーツの競技の種類及びカメラの撮影状況によっては, 同一シーンにおいても勝利したチームを応援していた観客と敗北したチームを応援していた観客が混在して撮影される可能性があるが, 本論文で用いる “喜んでいる” に対応するデータは, 特定のチームを応援していた観客の団を撮影されたものが多くそのような満足度の大きく異なる観客が混在しているデータは用いていない. また “喜んでいない” データをさらに “真剣に観戦している”, “真剣に観戦していない” に分類する. “真剣に観戦している” に対応するデータは, 主に映っている観客の国の選手が優勢でない状況の観客映像が 25 サンプルである. “真剣に観戦していない” に対応するデータは, 競技が始まっていないときの観客映像が 25 サンプルである. 顔検出・表情認識するときには, 画像をグレイスケールに変換する. それぞれのフレームには 3~30 人の観客が映っている.

2.2 提案システムの構成概要

提案システムでは, 観客が “喜んでいる” か “喜んでいない” かを識別することに観客が笑っているかいないかが重要と考え, 笑顔か非笑顔かの情報を利用する. また, “真剣に観戦している” か “真剣に観戦していない” かを識別するために観客の視線の集中度が重要と考え, 映像中の顔の大きな向きも利用する. 提案システムの構成概要を図 2 に示す. 初めに Viola らによって提案された顔検出法によって, 顔の向きごと (左向き, 正

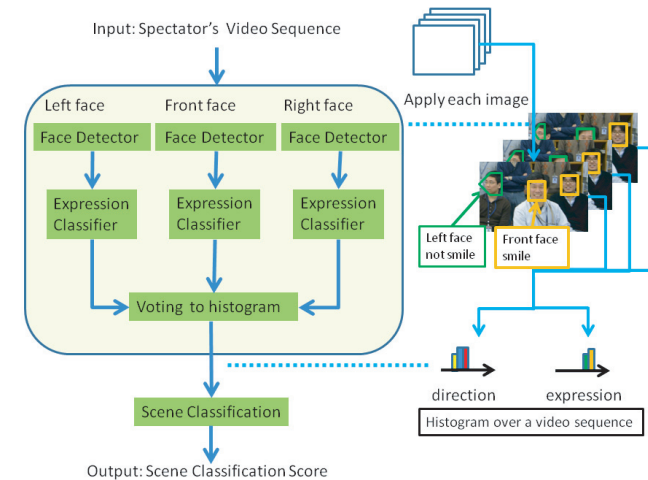


図 2 提案システムの構成概要
Fig. 2 Overview of proposed system.

面, 右向き) の顔を観客動画の各フレームで検出する. 次に検出された顔を笑顔, 非笑顔の 2 種類の表情に識別する. 1 つの画像シーケンスにおいて顔の向きと表情をそれぞれのヒストグラムに投票する. 投票されたヒストグラムを特徴ベクトルに用いたサポートベクタマシン (SVM) により, 各状態に識別したときの識別スコアを出力する. 近年, 一般物体認識の分野ではベクトル量子化された局所特徴のヒストグラムを用いて画像を認識する “bag-of-visual-words” という手法が注目されている¹³⁾⁻¹⁵⁾. 提案手法は, 顔識別器ベースの bag-of-visual-words と呼べる手法であり, 本論文で用いるサポートベクタマシンのほかにも probabilistic Latent Semantic Analysis (pLSA) や Latent Dirichlet Allocation (LDA) などの bag-of-visual-words の識別手法が適用可能と考えられる.

3. 顔検出手法

顔の向きと表情を抽出するために, 向きごと (左向き, 正面, 右向き) の顔検出器と表情認識器を用いる. これらは, 入力画像シーケンスの各フレームに適用する. 顔検出手法を図 3 に示す. 顔検出には, Viola と Jones により提案された Harr-like 特徴のカスケード接続による顔検出器 (VJ-detector)¹⁾ と, SVM に基づく非顔フィルタを組み合わせ用いている. VJ-detector は, Harr-like 特徴という輝度値の差分特徴を AdaBoost により

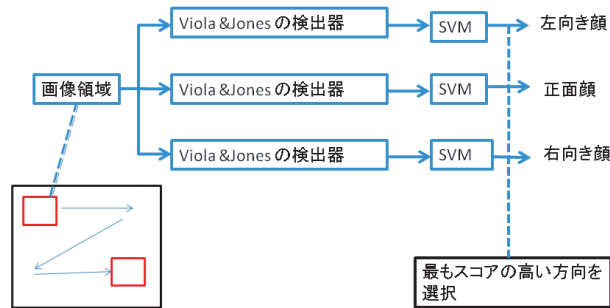


図3 顔検出手法
Fig. 3 Face detection method.



図4 VJ-detector による誤検出領域の例
Fig. 4 Example of false positive by VJ-detector.

統合した検出器であり、高速に顔検出が行える手法であることから採用している。ただし、満足度を推定する目的においてはなるべく誤検出を減らすことが望ましい。そこで、誤検出を減らすための識別器として輝度ヒストグラムとSVMを後段に用意している。同様にVJ-detectorの性能を向上させる手法として、VJ-detectorの最終段階でAdaBoostで選別した特徴をカーネルSVMで識別する手法も提案されている⁹⁾が、ここでは、非顔フィルタでは、VJ-detectorと異なる特徴量を用いることによって、誤検出を減らしている。これは、服や旗などに現れる勾配に着目すると顔らしいパターンは、輝度に着目すると顔ではないことが容易に分かる場合があるという観察に基づく。VJ-detectorにより誤検出した領域の例を図4に示す。非顔フィルタの画像特徴量とSVMのカーネルは表情認識手法と同じであり、それらは4章で説明する。異なる向きの顔が近い場所で検出された場合、非顔フィルタの識別スコアが最も高い向きの顔を採用する。この選択の実装は次のとおりである。まず全方向の顔を検出し、識別スコアと顔の位置と大きさをスタックする。次に、識別スコアが大きい順に顔を選び、前に選んだ顔の矩形領域内に新しく選ぶ顔の中心がないときに顔とする。また、横向きの顔は、正面の顔を半分含む場所で検出されることが多いため、そのような位置ずれした顔も非顔サンプルとして学習して除去する。

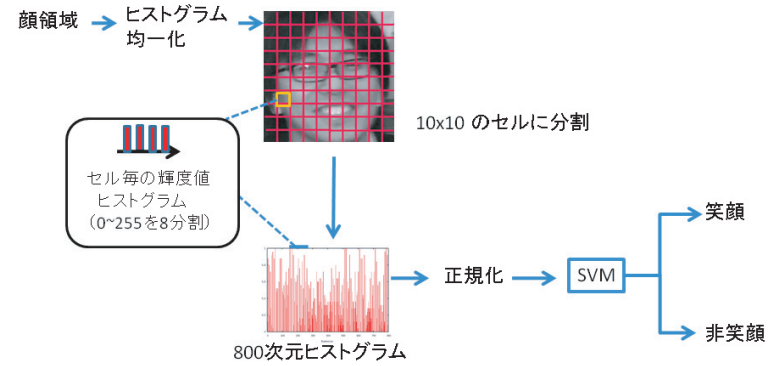


図5 表情認識手法
Fig. 5 Facial expression recognition method.

4. 表情認識手法

4.1 局所特徴量

顔画像を2種類の表情(笑顔, 非笑顔)に識別する識別器を顔の向きごとに用意する。現在研究されている表情認識手法は、目の中心や鼻の頂点などから数十の特徴的な部位を抽出して用いる手法¹⁸⁾とパターンベースで認識する手法⁶⁾とに大別されるが、笑顔か非笑顔かの表情を認識する場合、笑顔においては、歯に相当する部分が白くなっており、非笑顔ではそうでないなど、輝度値のみの情報でも判別が行えると思われる。そこで本研究では、単純に実装することのできるパターンベースの手法を用いる。シンプルなパターンベースの手法であれば、複数人物の表情認識への適用も容易に行える。図5に表情認識手法を示す。画像の特徴量として、5×5pixel内の輝度ヒストグラム特徴(Histogram of Intensity: HI)を用いる。これは、VJ-detectorで検出された顔は位置が完全に合っておらず、小さな位置変動に頑健性を持たせるためである。特徴抽出の手順は次のとおりである。まず、VJ-detectorで検出された顔領域を50×50pixelにリサイズし、ヒストグラム均一化を施す。次に、10×10の重複のないセルに分割し、セル内でヒストグラムを計算する。ここで輝度値は、0~255を8階調に分割する。HI特徴の次元数は10×10(cells)×8(intensitylevel)=800次元である。特徴量の正規化は、セルごとにセル内のピクセル値の個数で割ることで行った。HI特徴はきわめて単純な特徴であるが、表情を認識するために十分な次元数を持っており、多数のサンプルを用いることで高い表情認識性能を持たせることができる。

近年, HOG (Histogram of Oriented Gradient) 特徴が注目されている¹⁰⁾. HOG は, ブロック内の勾配ヒストグラムであるため, HI 特徴のような明度変動の影響を受にくい. また Hu らは HOG を局所性保存射影 (Locality Preserving Projection: LPP) により次元圧縮し識別性能を高めている⁷⁾ が, HOG や LPP は, 勾配のピンの切り方, 正規化ブロックの取り方, 正規化方法, 圧縮する次元数などの設定パラメータが多い反面, HI は, 設定パラメータが少なく, 上記設定で十分な識別性能を持っていた.

4.2 サポートベクタマシン

識別器には, 現在最も識別能力の高い識別器の 1 つとして知られているサポートベクタマシン (SVM)¹¹⁾ を用いる. SVM は, 識別面と学習サンプルとの最小距離 (マージン) を最大化するように識別面を決定する. 線形 SVM の識別関数は以下のように表される.

$$f(\mathbf{H}) = \text{sgn} \left(\sum_{i \in SV} \alpha_i y_i \mathbf{H}_i \mathbf{H} - b \right),$$

ここで, SV はサポートベクタの集合, \mathbf{H}_i, \mathbf{H} は学習サンプルとテストサンプル, y_i はサンプル \mathbf{H}_i のクラスラベル (+1 or -1), α_i は学習サンプル \mathbf{H}_i に対する重み, b は学習された閾値を表す.

ここまでは, 線形分離可能な場合を仮定していた. 線形分離不可能な場合には, 非線形写像 $\psi(\mathbf{x})$ を用いることにより高次元空間にデータを射影し, 高次元空間上でマージンを最大化することで対処できる. 高次元空間での内積 $\psi(\mathbf{x})^T \psi(\mathbf{y})$ が, カーネル $K(\mathbf{x}, \mathbf{y})$ で計算可能な場合, 高次元空間に射影することなく学習や識別が行える. これは, カーネルトリックと呼ばれる. カーネル SVM の識別関数は以下で表される.

$$f(\mathbf{H}) = \text{sgn} \left(\sum_{i \in SV} \alpha_i y_i K(\mathbf{H}_i, \mathbf{H}) - b \right),$$

ここで, $K(\mathbf{H}_i, \mathbf{H})$ は学習サンプル \mathbf{H}_i とテストサンプル \mathbf{H} のカーネル関数値を表す. カーネル SVM は非線形の識別面が獲得できるため高精度に認識が行える. ヒストグラム特徴を用いた識別問題では, χ^2 カーネル¹²⁾ が有効であることが一般に知られている. そこで本論文では, ヒストグラム特徴を用いて, 表情を認識するために χ^2 カーネルを用いる. ヒストグラム $\mathbf{H}_X = \{H_X(1), \dots, H_X(N)\}$ とヒストグラム $\mathbf{H}_Y = \{H_Y(1), \dots, H_Y(N)\}$ に対する χ^2 カーネル $K_{\chi^2}(\mathbf{H}_X, \mathbf{H}_Y)$ は以下のように定義される.

$$K_{\chi^2}(\mathbf{H}_X, \mathbf{H}_Y) = \exp \left(-\frac{1}{2\sigma} \sum_{i=1}^N \frac{(H_X(i) - H_Y(i))^2}{H_X(i) + H_Y(i)} \right).$$

カーネルパラメータ σ と SVM の正則化パラメータ C ¹¹⁾ は 5 分割交差確認法に基づくグリッドサーチによって決定する.

5. 満足度推定手法

5.1 投票方法

本研究では全体の満足度を推定することが目的であるので, 個人個人の顔向きと表情認識結果を全体の頻度分布であるヒストグラムに投票する. 3 章と 4 章で述べた顔検出と表情認識は, 入力動画の各フレームに対して適用されるものであったが, 顔向きと表情の認識結果は, 1 つのビデオシーケンスに対してヒストグラムに投票する. これは投票をビデオシーケンスごとに行えば, 1 フレームごとに行うより多くの顔が拾え, 誤認識も平滑化されると考えるからである. 本論文では, 顔表情 \mathbf{H}^E (笑顔, 非笑顔の 2 次元), 顔向き \mathbf{H}^D (左向き, 正面, 右向きの 3 次元) と顔表情と顔向きの共起 $\mathbf{H}^{D \wedge E}$ (左向き \wedge 笑顔, 左向き \wedge 非笑顔, 正面 \wedge 笑顔, 正面 \wedge 非笑顔, 右向き \wedge 笑顔, 右向き \wedge 非笑顔の 6 次元) の 3 つのヒストグラムを用いる. 各ヒストグラムはノルムが 1 になるように顔の個数で正規化する. ここで, 顔の向きは観客の視線の向きの集中度を表現するために用いている. カメラの設定がシーンによって異なるため, 最も生起頻度の高い顔の向きの次元を揃える目的で, 顔の向きのヒストグラム \mathbf{H}^D の次元の並びを以下のようにデータによって置き換えを行うという改良を行う. 左向き顔の数が最も多いシーンでは, 顔向きの次元の順序は (右向き, 左向き, 正面) とする. 正面顔の数が最も多いシーンでは, 顔向きの次元の順序は (左向き, 正面, 右向き) とする. 右向き顔の数が最も多いシーンでは, 顔向きの次元の順序は (正面, 右向き, 左向き) とする. 共起ヒストグラム $\mathbf{H}^{D \wedge E}$ の順序も同様に並べ替える. この並べ替えは, 3 つ向きの頻度ベクトル (次元 1, 次元 2, 次元 3) が, 次元 1-次元 2-次元 3-次元 1-... のように循環して隣接しているとして, 基準となる (左向き, 正面, 右向き) の隣接関係を保存したまま最も生起頻度の高い顔向きを中央になるように次元をシフトさせることに相当する. この並べ替えの基準は, 右向き 30 度, 右向き 60 度といったように向きの方向を細かく拡張していくことを考えた場合でも同様に向きの隣接関係を保存したまま向きの正規化を行うことができる. これらは, フレームごとに行う. このような並べ替えによって向き不変性が生じ, 識別精度が向上すると期待される. その効果は 7.3 節で示す.

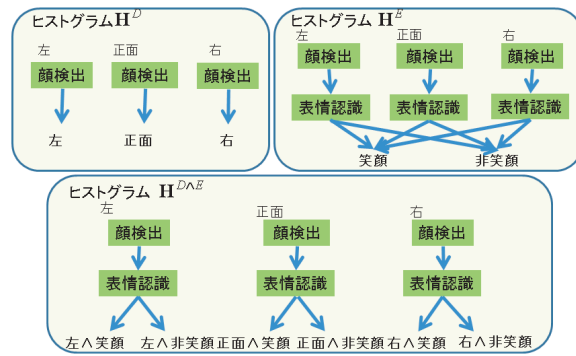


図 6 投票方法
Fig. 6 Flow of voting method.

共起ヒストグラムを用いたのは、試合を見ている人が笑っているのか、見ていない人が笑っているのかといった詳細な特徴をとらえるためである。以上の投票方法を識別器に対応させると図 6 のような流れになっている。

5.2 サポートベクタマシンによる判別

映像中の観客の状態を識別するために、ヒストグラムインターセクションカーネル¹⁵⁾ に対して重み付きで各ヒストグラムの重要度を調整するという改良を施したカーネル SVM で識別を行う。識別対象に応じて、各ヒストグラムの重要性は異なると期待される。たとえば、顔の表情は“喜んでいる”と“喜んでいない”を識別する際に重要であると考えられ、反対に顔の方向は観客が“真剣に観戦している”か“真剣に観戦していない”かを識別する際に重要であると考えられる。そのため、各ヒストグラムを単純に連結させたヒストグラムを特徴ベクトルとして 1 つのカーネルを用いる場合、特徴ベクトル間の類似度が適切に計算できない可能性があるが、各識別タスクに応じて各ヒストグラムの重要性を調整できるカーネルを用いれば適切な類似度を計算できる。ここでは、各ヒストグラムの重要性を調整するために、各ヒストグラムの重みを以下のように導入する。

$$K_B(\mathbf{H}_X, \mathbf{H}_Y) = \alpha_D K_A(\mathbf{H}_X^D, \mathbf{H}_Y^D) + (1 - \alpha_D) K_A(\mathbf{H}_X^E, \mathbf{H}_Y^E)$$

where $0 \leq \alpha_D \leq 1$.

ここで $K_A(\mathbf{H}_X, \mathbf{H}_Y)$ はヒストグラム間の重なり度合いを表すヒストグラムインターセクションカーネルであり、次式で表される。

$$K_A(\mathbf{H}_X, \mathbf{H}_Y) = \sum_{i=1}^N \min(H_X(i), H_Y(i)),$$

ここで N は各ヒストグラムの次元数を示す。本論文ではヒストグラムインターセクションカーネルをハイパーパラメータが存在せず、各ヒストグラムの重みのみの調整を行えばよいという理由で採用している。3 つのヒストグラムを用いたマルチチャンネルカーネルを以下に示す。

$$K_C(\mathbf{H}_X, \mathbf{H}_Y) = \alpha_D K_A(\mathbf{H}_X^D, \mathbf{H}_Y^D) + \alpha_E K_A(\mathbf{H}_X^E, \mathbf{H}_Y^E) + (1 - (\alpha_D + \alpha_E)) K_A(\mathbf{H}_X^{D \wedge E}, \mathbf{H}_Y^{D \wedge E})$$

where $0 \leq \alpha_D + \alpha_E \leq 1$.

ヒストグラムインターセクションカーネルは Mercer カーネルである¹⁵⁾。Mercer カーネルの非負重み付き和カーネルも Mercer カーネルである¹⁶⁾ ことが知られているので、 K_B と K_C も Mercer カーネルである。

各ヒストグラムの重み α_D と α_E は交差確認法により自動的に決定することにより、各ヒストグラムの重要度に対するヒューリスティックは不要となる。

5.3 満足度推定

2.1 節で述べたように本論文では、“観客が満足している” = “観客が喜んでいる” に対応する満足度 A と “観客が満足している” = “観客が真剣に観戦している” に対応する満足度 B の 2 つの満足度を推定する。具体的に満足度 A は “喜んでいる” か “喜んでいない” を SVM により識別したときに “喜んでいる” 状態に近い度合いで表す。すなわち、“喜んでいる” データをクラスラベル 1, “喜んでいない” をクラスラベル -1 として学習した SVM を用いて、入力データ \mathbf{H} の満足度 A は、満足度 $A(\mathbf{H}) = \sum_{i \in SV} \alpha_i y_i K(\mathbf{H}_i, \mathbf{H}) - b$ とする。同様に満足度 B を “真剣に観戦している” か “真剣に観戦していない” を SVM により識別したときのスコアで表す。

6. システム

3~5 章で述べた識別手法を linux-2.6 に実装した。システムは入力動画象に対して、顔検出結果、表情認識結果、および満足度推定結果を表示するものである。VJ-detector には OpenCV^{*1} に実装済みのものを利用した。VJ-detector の局所特徴量として、拡張 Harr-like

*1 Software available at <http://www.sourceforge.net/projects/opencvlibrary>



図 7 システム画面
Fig. 7 Screen of system.

feature⁴⁾を用いた．SVMの実装には LIBSVM^{*1}を用いている．左顔と右顔は対象であるため，左顔の識別器を全識別器（VJ-detectors，非顔フィルタ，表情認識器）で用意し，右顔の認識は入力画像を水平に反転させることで実現した．システム画面を図 7 に示す．四角形が正面顔を表し，五角形が左向きと右向きを表している．また，橙色が笑顔を表し，黄色が非笑顔を表している．

7. 実験

7.1 識別器の学習

SVMの学習に用いたサンプルは2章で説明したデータセットとTVからキャプチャした他の映像にVJ-detectorを適用することで用意した．それらは，10frame間隔の映像を用いている．顔の表情は手動により“笑顔”と“非笑顔”にラベル付けがされている．

表情認識において，正面顔は6,460サンプル（笑顔2,730，非笑顔3,730）と側面顔は6,440サンプル（笑顔2,270，非笑顔4,170）を用いた．正面顔の非顔フィルタには，3,000サンプルの顔画像と3,000サンプルの非顔画像を用いた．側面顔の非顔フィルタには，2,000サンプルの顔画像と2,000サンプルの非顔画像を用いた．ここで，VJ-detectorで誤って検出された顔領域が非顔画像として用いられている．各識別器の5分割交差確認法による識別率を表1に示す．

7.2 顔検出・表情認識実験

構築した識別器を用いた実際のシーンでの識別率を表2に示す．表2において，Accuracyは検出された顔において，表情が正しく正解した割合を表す．横顔の表情の認識率は正面顔に比べて少し悪い．これは，横顔のパリエーションの多さによると考えられる．誤認識の主

*1 Software available at <http://www.csie.edu.tw/~cjlin/libsvm>

表 1 SVMの識別率（学習データの5分割交差確認法による平均）

Table 1 Classification performance of SVM (average of 5-fold cross validation on training data).

Direction	NonFace Filter	Expression (HI)
Front	95.62%	92.3%
Side	94.05%	93.5%

表 2 識別データに対する識別性能．FP (a)：間違った表情が混入する率，FP (b)：非顔が混入する率

Table 2 Accuracy on classification scene. FP (a): The incorporation rate of false expression, FP (b): The incorporation rate of non face.

	Accuracy	FP (a)	FP (b)
Front (Smile)	96.3%	2.0%	1.7%
Front (Not Smile)	99.4%	0.3%	0.3%
Left (Smile)	93.8%	3.3%	2.9%
Left (Not Smile)	98.8%	0.2%	1.0%
Right (Smile)	89.7%	4.7%	5.6%
Right (Plane)	98.5%	0.2%	1.3%



図 8 隠れの存在する顔の例
Fig. 8 Example of occluded face.

な原因は他の人間の頭や腕などによる隠れであった．隠れ領域が小さい場合には，本論文で用いた大域のカイ2乗カーネルに代わり局所カイ2乗カーネルの和¹⁶⁾を用いることで改善が可能と考えられる．しかし，人間の眼でも識別が困難な大きな隠れを含む顔も存在していた．隠れの存在する顔の例を図8に示す．隠れに対応するためには，隠れているかを識別して，隠れの存在しない顔のみを用いることが必要と考えられる．これは今後の課題とする．隠れの場合を除いて，表情の識別率は高いといえる．

また未検出顔数の評価を行った．ここで，検出したい顔を大きな隠れや画面からはみ出しなどが存在せず，人間でも表情判別が行える顔と定義し，未検出率を $100 \times (\text{検出したい顔のうち未検出の顔の総数}) / (\text{検出したい顔の総数})$ とする．未検出率は1シーケンス中から最も時間のあく4フレームを用いて1シーケンスごとの未検出率を評価している．1シーケンスごとで未検出率を評価したのは，シーケンスによって未検出の多いシーケンスと少ないシーケンスがあるためである．未検出の分布を表3に示す．未検出の原因には以下の3通

表 3 未検出率の分布。(A): 0%以上 5%未満, (B): 5%以上 10%未満, (C): 10%以上 20%未満, (D): 20%以上 30%未満, (E): 30%以上 40%未満, (F): 40%以上 50%未満, (G): 50%以上, Mean: 平均未検出率

Table 3 Distribution of False negative rates. (A): more than 0% and less than 5%, (B): more than 5% and less than 10%, (C): more than 10% and less than 20%, (D): more than 20% and less than 30%, (E): more than 30% and less than 40%, (F): more than 40% and less than 50%, (G): more than 50%, Mean: mean false negative rate.

False negative rates	A	B	C	D	E	F	G	Mean
Number of Sequence	12	12	28	27	12	7	2	20.64%

りが存在した。1 つ目は、画面に映る顔が小さい場合である。この場合、観客とカメラとの位置が離れておりシーンに映る人数が多く、画面内に多くの顔があるため、多くの顔ができており全体の満足度推定に問題はないと考えられる。2 つ目の原因は、応援のため、顔に特殊なメイクをしている顔であるが、これは特殊な場合と考える。3 つ目の原因は、正面、側面以外の顔の角度（斜め方向への回転、下向きの顔など）をしている顔である。これには、本論文の手法を拡張して、様々な向きの顔の検出器を用いて顔を検出することで対処できると思われる。

各シーケンスの未検出率を全シーケンスで平均化すると 20.64%であった。本研究目的のアンケート調査などの自動化という観点から考えると実際のアンケートにおいても観客全員から回答が得られるわけではなく、全体の満足度の部分集合から全体の満足度を推定することになるため、全体の 8 割程度の人間の顔を検出すれば十分であると考えられる。未検出以上に全体の満足度推定に影響すると思われるものは顔以外のものを顔とする誤検出であり、誤検出率は表 2 に示したように 5%未満と低い。

7.3 SVM による観客動画識別実験

観客の状態を線形 SVM とマルチチャンネル和カーネルを用いた SVM で識別した。誤識別率を leave-one-out 法で測定した。ベストパラメータにおける各手法の誤識別率を表 4 に示す。線形 SVM が、カーネル SVM に匹敵する性能が得られている。顔向き histograms の順序を並べ替えることによって、多くの場合識別率は 2~4%改善された。“喜んでいる”が“喜んでいない”、“真剣に観戦している”が“真剣に観戦していない”かのどちらかの識別タスクにおいても、重み付き histogram インターセクションカーネル K_B , K_C は重みなし histogram インターセクション K_A より 2~3%識別率が改善された。3 チャンネル histogram (H^D , H^E , $H^{D \wedge E}$) は 2 チャンネル histogram (H^D , H^E) と同等以上の識別率であった。

線形 SVM で 3 チャンネル histogram (H^D , H^E , $H^{D \wedge E}$) を用いた場合の各次元の重み

表 4 シーン識別のエラー率 (%). (a): “喜んでいる” vs. “喜んでいない”, (b): “真剣に観戦している” vs. “真剣に観戦していない”, []: 向きの並べ替えを行わない場合のエラー率

Table 4 Error rates of scene classification (%), (a): “Positive Scene” vs. “Negative Scene”, (b): “Watching Seriously” vs. “Not Watching Seriously”, []: error rate of non aligned direction.

kernel	histogram	(a)	(b)
Linear	H^D	53 [44]	28 [40]
Linear	H^E	13 [-]	68 [-]
Linear	$H^{D \wedge E}$	10 [13]	32 [32]
Linear	H^D , H^E	13 [14]	30 [40]
Linear	H^D , H^E , $H^{D \wedge E}$	10 [13]	30 [32]
K_A	H^D	40 [39]	32 [36]
K_A	H^E	10 [-]	40 [-]
K_A	$H^{D \wedge E}$	13 [16]	34 [30]
K_A	H^D , H^E	12 [12]	32 [36]
K_A	H^D , H^E , $H^{D \wedge E}$	12 [13]	32 [34]
K_B	H^D , H^E	9 [12]	30 [34]
K_C	H^D , H^E , $H^{D \wedge E}$	9 [11]	28 [30]

を図 9 に示す。“喜んでいる”が“喜んでいない”かの識別においては、向きの並べ替えを行うことによって、最も多く向いている方向の人が、笑顔であることを示す次元が“喜んでいる”を最もよく支持している。一方で並べ替えを行っていないときには、左向きの顔が笑顔であることを示す次元が高い重みとなっている。また、“真剣に観戦している”と“真剣に観戦していない”においても向きの並べ替えを行うことによって、最も多く向いている方向を示す次元と最も多く向いている方向の顔が非笑顔であることを示す次元が“真剣に観戦している”を最もよく支持している。このように最も多くの人に向いている方向の向きを揃えるように並べ替えを行うことによって適切な重みを獲得することができ、識別率が向上したと考えられる。

SVM の同じ C パラメータにおいて α_D に対応したエラー曲線を図 10 に示す。図 10 においてパラメータ α_D は 0.05 間隔で変化させている。“喜んでいる”が“喜んでいない”かを識別する場合、最適なパラメータは $\alpha_D = 0.25$ であり、誤識別率は α_D を大きくするに従い増加している。これは、この識別問題に対して顔の表情の重要性が高いことを意味している。反対に“真剣に観戦している”が“真剣に観戦していない”かの識別の場合、最適な重みは $\alpha_D = 0.55$ であり、 α_D を小さくするに従って誤識別率は増大している。これは、このタスクにおいて顔の向きがより重要であることを意味している。これらの重要性は、我々の予測と一致した。それは、表 4 において、histogram H^D を用いた場合では、“真剣

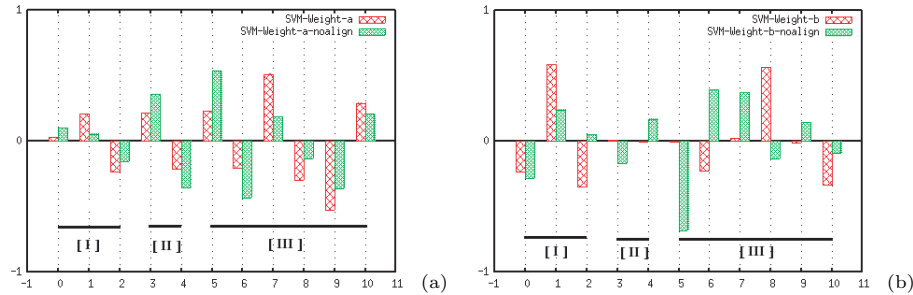


図 9 線形 SVM の H^D , H^E , $H^{D \wedge E}$ に対する重み. [I]: H^D - 左から順に, 左, 正面, 右, [II]: H^E - 左から順に, 笑顔, 非笑顔 [III]: $H^{D \wedge E}$ - 左から順に, 左 \wedge 笑顔, 左 \wedge 非笑顔, 正面 \wedge 笑顔, 正面 \wedge 非笑顔, 右 \wedge 笑顔, 右 \wedge 非笑顔, 赤棒: 向きの並べ替えあり, 緑棒: 向きの並べ替えなし, (a): “喜んでい” vs. “喜んでいない”, (b): “真剣に観戦している” vs. “真剣に観戦していない”

Fig.9 Weights of linear SVM for each dimension of histograms, H^D , H^E , $H^{D \wedge E}$. [I]: H^D - the order is left, front, right, [II]: H^E - the order is smile, not smile [III]: $H^{D \wedge E}$ - the order is left \wedge smile, left \wedge not smile, front \wedge smile, front \wedge not smile, right \wedge smile, right \wedge not smile, red bar: aligned direction, green bar: non aligned direction, (a): “Positive Scene” vs. “Negative Scene”, (b): “Watching Seriously” vs. “Not Watching Seriously”.

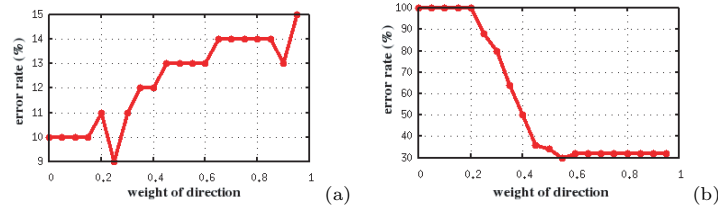


図 10 向きの重み α_D に対応するカーネル K_B のエラー率 (leave-one-out). (a): “喜んでい” vs. “喜んでいない” ($C=32$), (b): “真剣に観戦している” vs. “真剣に観戦していない” ($C=0.5$)

Fig.10 Error rates for K_B corresponding to direction weight α_D (leave-one-out). (a): “Positive Scene” vs. “Negative Scene” ($C=32$), (b): “Watching Seriously” vs. “Not Watching Seriously” ($C=0.5$).

に観戦している”か“真剣に観戦していない”かの識別率が高く, H^E を用いた場合には, “喜んでい”か“喜んでいない”かの識別率が高いことから分かる. また, カーネル K_C の最適重みは, “喜んでい”か“喜んでいない”かを識別する場合, $\alpha_D = 0$, $\alpha_E = 0.55$ であり, “真剣に観戦している”か“真剣に観戦していない”かを識別する場合, $\alpha_D = 0.4$, $\alpha_E = 0.45$ であった.

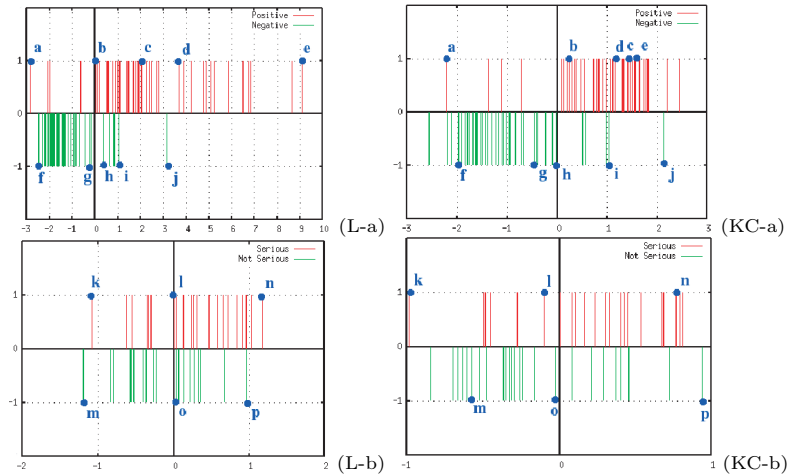


図 11 満足度の分布. 横軸が識別スコアを表し, 縦軸が正解クラスラベルを表す. (L-a): 線形 SVM, “喜んでい” vs. “喜んでいない”, (KC-a): カーネル K_C , “喜んでい” vs. “喜んでいない”, (L-b): 線形 SVM, “真剣に観戦している” vs. “真剣に観戦していない”, (KC-a): カーネル K_C , “真剣に観戦” vs. “真剣に観戦していない”

Fig. 11 Distribution of the degree of spectators' satisfaction: horizontal axis indicates classification score, vertical axis indicates correct class label. (L-a): Linear SVM, “Positive Scene” vs. “Negative Scene”, (KC-a): Kernel K_C , “Positive Scene” vs. “Negative Scene” (L-b): Linear SVM, “Watching Seriously” vs. “Wacting Not Seriously”, (KC-a): Kernel K_C , “Watching Seriously” vs. “Not Watching Seriously”.

誤識別したデータは, “喜んでい” 状況で観客は騒いで喜んでいるが笑っていない, “真剣に観戦している” シーンで異なる方向を見ているなど, 試合の状況に応じてラベル付けを行ったことや visual-words の種類が不足しているために起こっていた. 提案手法は, 他の顔表情の種類や顔以外の visual-words を用いることもできる. 騒いでいる状態においても “喜んでい” と識別することは, 叫びの表情を追加したり, 拍手するなど動きの情報を追加したりすることで容易に行えると考えられる.

7.4 満足度推定結果に関する考察

次に満足度の推定結果がどのようなデータに対して満足度が高くなっているか, または低くなっているかを調査することを目的として考察を行う. 図 11 は線形 SVM と 7.3 節における最適パラメータを用いた場合のカーネル K_C における満足度の分布である. これらは, 7.3 節の識別タスク (leave-one-out) におけるテストデータの分布である. “喜んでい

る”か“喜んでいない”の識別スコア(満足度 A)は,“真剣に観戦している”と“真剣に観戦していない”の識別スコア(満足度 B)に比べ,正解ラベルの方によく分離して分布している。しかし,図 11 a のように飛び跳ねたり,拍手したり,叫んだりして喜んでいたり笑ってはいない人が多いデータに対しては,低い満足度が与えられるなど,正解が“喜んでいない”データが“喜んでいない”と判定されるようなデータに対し,不自然な満足度となっていた。“喜んでいない”か“喜んでいない”を識別する場合(L-a),(KC-a)において図 11 b のように満足度が低いと思われるシーンでは満足度が低く図 11 c, d のように,笑っている人の数が中程度であれば,中程度の満足度となっており図 11 e のように全員笑っているようなシーンでは高い満足度となった。正解が“喜んでいない”データに対しては,図 11 f のように楽しそうでない場合に満足度が低く,図 11 g のように楽しそうになるに従って,満足度は高くなっていき図 11 h, i, j のように楽しそうにしている人が多いデータに対し,満足度が高くなった。これは,試合前に楽しそうにしている観客のデータが“喜んでいない”サンプルに入っているからであり,これらが満足していると判定されるのは,自然であると考えられる。喜んでいないデータに対して,中立的な立場と思われる観客が映っているデータが存在したが,その混入率は低く,満足度推定にあまり影響していないと思われる。仮にスポーツ競技の種類やカメラの設定によって中立的な人物や敗北したチームを応援していた観客の数が増えることによって,全体の満足度が低下するとしても「画面全体の観客の満足度」を評価するうえでは正しい推定結果であると考えられる。また,そのカメラ内で中心的に撮影している人物の満足度に対する影響を大きくしたい場合,顔の大きさなどによって重み付け投票を行うなどの対策も考えられる。

次に“真剣に観戦している”と“真剣に観戦していない”の識別タスク(L-b),(KC-b)において考察する。図 11 k は,一部の人は真剣に観戦しているデータであるが,いろいろな方向を向いている人がいるため真剣でないとして判定されてしまっている。図 11 l, n においては見た目の真剣度合いが上がるに従い満足度 B が上がっている。“真剣に観戦していない”データに関しても図 11 m は,まったく真剣に観戦していないと思われるデータであり,真剣でないとして正しく判定されている。図 11 o においては,手を振ったり,横を向いたり真剣に観戦していない人も真剣に観戦している人もいて,中間的な真剣な度合いと考えられるデータであったが,正しく満足度 B が推定されている。図 11 p は正解が“真剣に観戦していない”データであるが,真剣に観戦していない人もいるが,真剣に観戦している人も多かったため,真剣な度合いを表す満足度 B が高まったと考えられる。

8. ま と め

本論文では,映像中の顔に着目して,観客の満足度を推定するシステムを提案した。提案システムは,顔識別器ベースの bag-of-visual-words に基づいており,顔表情の識別器を画像上に検出されたすべての顔に適用することにより複数の人物の表情を認識し,表情と顔の向きを投票したヒストグラムを作る。このヒストグラムを特徴ベクトルとした識別器を用いて観客の満足している行動の度合いとしての満足度を自動的に推定する。提案システムを用いて TV から収集した観客の映像を用いた識別実験を行い,“喜んでいない”または“真剣に観戦している”と“真剣に観戦していない”場面を有効に識別できることを示し,“喜んでいない度合い”または“真剣に観戦している度合い”としての満足度を各識別問題の識別スコアを用いて推定した。“喜んでいない”と“喜んでいない”を識別する場合には顔表情が重要であり,“真剣に観戦している”と“真剣に観戦していない”を識別する場合には顔向きが重要であるということも確認できた。

提案手法を実際のアンケート調査の代わりに適用するためには,検出可能な顔方向の種類や笑顔の度合いなどをさらに増やす必要があると考えられる。これらを実装した場合の性能評価を行い,本手法の限界や制限を明らかにすることは今後の課題とする。また,今後はより多くの観客の状況を識別することや教師なし学習で観客の状況を分類するなど,より適切な観客の状況のラベル付け法を検討することも課題としてあげられる。

謝辞 本研究の一部は,魅力ある大学院教育イニシアティブ「実践 IT 力を備えた高度情報学人材育成プログラム」による。

参 考 文 献

- 1) Viola, P. and Jones, M.: Robust Real Time Face Detection, *International Journal of Computer Vision*, Vol.57, Issue 2, pp.137-154 (2004).
- 2) Huang, C., Ai, H., Li, Y. and Lao, S.: High-Performance Rotation Invariant Multiview Face Detection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.29, No.4, pp.671-686 (2007).
- 3) Wu, B., Ai, H., Huang, C. and Lao, S.: Fast Rotation Invariant Multi-View Face Detection Based on Real Adaboost, *Proc. IEEE Conf. on Automatic Face and Gesture Recognition* (2004).
- 4) Lienhart, R. and Maydt, J.: An Extended Set of Haar-like Features for Rapid Object Detection, *Proc. International Conf. on Image Processing*, Vol.1, pp.900-903 (Sep. 2002).

- 5) Hidaka, A. and Kurita, T.: Non-Neighboring Rectangular Feature Selection Using Particle Swarm Optimization, *Proc. International Conf. on Pattern Recognition* (2008).
- 6) Shinohara, Y. and Otsu, N.: Facial Expression Recognition Using Fisher Weight Maps, *Proc. IEEE Conf. on Automatic Face and Gesture Recognition*, pp.499–504 (2004).
- 7) Hu, Y., Zeng, Z., Yin, L., Wei, X., Zhou, X. and Haung, T.S.: Multi-View Facial Expression Recognition, *Proc. IEEE Conf. on Automatic Face and Gesture Recognition* (2008).
- 8) Chen, H.-Y., Haung, C.-L. and Fu, C.-M.: Hybrid-boost learning for multi-pose face detection and facial expression recognition, *Pattern Recognition*, Vol.41, No.3, pp.1173–1185 (2008).
- 9) Le, D.D. and Satou, S.: Feature Selection By AdaBoost For SVM-Based Face Detection, *Proc. Forum on Information Technology*, pp.183–186 (2004).
- 10) Dalal, N. and Triggs, B.: Histograms of Oriented Gradients for Human Detection, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp.886–893 (2005).
- 11) Vapnik, V.: *Statistical Learning Theory*, John Wiley & Sons (1998).
- 12) Chapelle, O., Haffner, P. and Vapnik, V.: SVMs for Histogram Based Image Classification, *IEEE Trans. Neural Networks*, Vol.10, Issue 5, pp.1055–1064 (1999).
- 13) Csurca, G., Dance, C., Fan, L., Willamowski, J. and Bray, C.: Visual categorization with bags of keypoints, *Proc. ECCV International Workshop on Statistical Learning in Computer Vision* (2004).
- 14) Lazebnik, S., Schmid, C. and Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp.2169–2178 (2006).
- 15) Grauman, K. and Darrell, T.: The Pyramid Match Kernel: Discriminative Classification with Set of Image Features, *Proc. IEEE Conf. on Computer Vision*, pp.1458–1465 (2005).
- 16) Hotta, K.: Robust face recognition under partial occlusion based on support vector machine with local Gaussian summation kernel, *Image and Vision Computing*, Vol.26, Issue 11, 1, pp.1490–1498 (2008).
- 17) 山本 誠, 谷本弘昭, 新田直子, 馬場口登: 個人的選好獲得のための特定人物のテレビ

視聴時における興味区間推定, 電子情報通信学会論文誌, Vol.J90-D-II, No.8, pp.2202–2211 (2007).

- 18) 宮原正典, 青木政樹, 滝口哲也, 有木康雄: 顔表情からの関心度推定に基づく映像コンテンツへのタギング, 情報処理学会論文誌, Vol.49, No.10, pp.3694–3702 (2008).

(平成 21 年 4 月 17 日受付)

(平成 21 年 9 月 11 日採録)



松川 徹

2006 年筑波大学第三学群工学システム学類卒業。2008 年同大学大学院システム情報工学研究科博士前期課程修了。現在、同研究科博士後期課程在学中。修士（工学）。パターン認識の理論とその画像認識への応用の研究に従事。電子情報通信学会学生会員。



日高 章理

2004 年茨城大学理学部卒業。2006 年筑波大学大学院システム情報工学研究科修士（工学）取得。2009 年同研究科博士（工学）取得。現在は東京電機大学理工学部理学系助教。パターン認識と機械学習に基づいたコンピュータビジョン研究に従事。主な研究関心事は、物体検出、追跡、一般物体認識、ロボットビジョン等。IEEE および電子情報通信学会会員。



栗田多喜夫（正会員）

1958 年生。1981 年名古屋工業大学卒業。同年電子技術総合研究所入所。1990～1991 年カナダ NRC 客員研究員。博士（工学）。現在（独）産業技術総合研究所脳神経情報研究部門副研究部門長。統計的パターン認識とその画像認識への応用に関する研究に従事。IEEE, 電子情報通信学会, 神経回路学会, 人工知能学会各会員。