

視聴者の興味モデリングを用いた 多視点映像視聴インタフェース

間瀬 健二^{†1} 永井 有希^{†1} 丹羽 宏介^{†1}
藤井 俊彰^{†2} 丸谷 宜史^{†1} 川本 哲也^{†3}

多数のカメラで撮影した多視点映像や自由視点映像を、視聴者がそれぞれの興味や目的に基づいてゆったりと映像を鑑賞できるインタフェースを提供するために、対象とするコンテンツの内容と関連づけながら、視聴者の興味をモデリングする。視聴品質 (Quality of View, QoV) 尺度を提案し、その定量的計算方法を提示する。QoV を統合した定義を与え、重回帰モデルを援用して定式化する方法と、視聴履歴を収集して獲得する方法を事例を示しながら統合モデルで説明する。

Multiview Video Viewing Interface Based on Viewer Interest Modeling

KENJI MASE,^{†1} YUKI NAGAI,^{†1} KOSUKE NIWA,^{†1}
TOSHIAKI FUJII,^{†2} TAKAFUMI MARUTANI,^{†1}
and TETSUYA KAWAMOTO^{†3}

Modeling viewers' interest is indispensable for multi-video viewing interface to provide comfortable free access to the best viewing positions. We propose a Quality-of-View modeling method and its measures based on top-down and bottom-up approaches. We show the computed models for the beach volleyball and the cheer leading cases.

^{†1} 名古屋大学大学院 情報科学研究科

Graduate School of Information Science, Nagoya University

^{†2} 名古屋大学大学院 工学研究科

Graduate School of Engineering, Nagoya University

^{†3} 中京テレビ放送

Chukyo Television

1. はじめに

多数のカメラで撮影した多視点映像や自由視点映像によるスポーツなどのコンテンツが大量に流通し、視聴者がそれぞれの興味や目的に基づいてインタラクティブに映像を鑑賞することが、今後予想される。しかしながら、視聴者は自由度が増える一方で、その自由度に応じて視点の操作が煩雑になり、豊かな鑑賞という本来の目的が阻害されることが懸念される。本稿では、視聴者の興味や目的を、対象とするコンテンツの内容と関連づけることにより、視聴者の興味をモデリングして視聴品質 (Quality of View, QoV) として尺度化することを提案し、さらにその利用技術について考察する。QoV の計算により、各コンテンツに対する視聴者の興味等にそった画面構成を、システムが推薦し自動誘導することが可能となり、新しい臨場感のある視聴が可能となることが期待できる。

複数のカメラを用いて多視点映像を記録し、そこから自分の好きな位置のカメラ映像を自由に再生し、さらに映像を補間して自由な視点からの映像を合成するための研究などが熱心に行われている。CMU の Kanade らが 2001 年の米国スーパーボウルにてアメリカンフットボールの自由視点映像放送として実用化した Eye-Vision システム¹⁾ は有名である。コンピュータ制御による複数のロボットカメラが多方向から選手を撮影する中から、映像ディレクターが最適の映像を居間に送出した。これは力技であり膨大なコストがかかる。そこで、より低いコストで実現するため、あらかじめ多数の固定カメラで撮影した素材を処理して、任意視点の映像生成をする自由視点 TV の研究²⁾ が行われている。

しかしながら、Eye-Vision で映像ディレクターが行っていたような「最適な映像の選択」技術の開発は未着手といえる。視聴者が将来インタラクティブに映像を閲覧できるようになるからといって、四六時中視点を制御したいわけではない。映像ディレクターのような知能をもったエージェントがそばにいて視聴支援・視聴推薦をしてもらいつつ、ここぞというときには、視聴者自身が視点を操作するようなシステムが目指すべき姿と考える。このようなシステムで、視聴者は視点の選択に明け暮れず、映像の内容を存分に楽しむことができるようになる。本稿では、この意図に任せる際に自分の興味を反映した意図を選ぶ自由度を与え、さらに、ちょっとした操作が随時可能なインタフェースを目指している。

そこで、これを実現するため、我々は、視聴者の興味を計算機の内部表現として記述する必要性に着目し、視聴者の興味や目的に依存した視聴品質 (QoV) を定義して、その定量的計算方法を提示することを試みる。具体的には、まず、第 1 の方法として QoV の一般化モデルを定義する。そのうえで、QoV を重回帰モデルを援用してトップダウンに定式化す

る．具体的には，被験者調査実験により，視聴被験者の興味に基づく評価を獲得し，予め定めてあるコンテンツ内対象要素の要素評価値との線形重回帰分析を行いパラメータを同定した^{3),4)}．ここでは実験に携わった被験者の興味モデルが内部表現として記述された．次に，第2の方法としてQoVをボトムアップに構築する．ここでは，視聴者の視聴履歴を詳細に獲得でき，かつ視聴者の興味を誘引する手掛かりを提供できる視聴インタフェースを準備し，被験者調査実験により興味ごとの視聴履歴を集めた⁵⁾．本稿では，これらQoVの計算対象は個々の画像フレーム内の対象物として，視聴履歴の頻度によってQoVを記述できることをしめす．以上のように，これまで別々のアプローチにより，コンテンツと視聴者の興味・目的を想定した内部表現と具体的な計算手順の設計と評価の枠組みを提案してきたが，それらを統合した観点から整理し説明しなおすことを試みる．本稿では，永井らのアプローチ^{3),4)}と丹羽らのアプローチ⁵⁾をまとめて引用しつつ統合したモデルの提案に注力し，詳細はそれぞれの文献に委ねる．

以下，2節では関連研究について述べ，3節ではQoVの一般化モデルと，トップダウン，ボトムアップの2つのアプローチを提案する．4節はそれぞれの実験結果を示し，5節で結論を述べる．

2. 関連研究

多視点で撮影した映像から自由視点映像を生成し提供する研究は多数ある．南谷ら⁶⁾は，サッカーを多視点で撮影しておいて，自由な方向から視聴する技術を開発している．おもに人物ビルボード形式と呼ぶ，多視点で撮影した人影を新しい視点の方向に少し傾けることによる任意視点の映像を擬似的に生成する手法を用いている．また，多視点で撮影した映像から交差体積法⁷⁾と呼ばれるアプローチで対象物体を3次元的に切り出して対象モデルを構築することで，任意視点からの映像を生成する手法も開発されている．福島ら²⁾は，光線空間法による多視点映像からの自由視聴点映像の生成技術を提案し，MPEG標準化に貢献している．

多視点や自由視点の映像コンテンツは，自由度が非常に高い．3次元空間を自由に動ける仮想視点の場合，少なくとも視点位置 (x, y, z) ，視方向 $(yaw : \theta, pitch : \phi, roll : \psi)$ ，視野角(焦点距離 f)の7自由度がある．実カメラを考えると，画像フレームのアスペクト比や被写界深度なども考慮する必要がある．これらのパラメータを同時に制御することは一般的な視聴者には無理難題である．稲本らが提案した⁸⁾箱庭スタジアム方式は，Mixed Reality技術を使って，手軽なインタフェースで，自由視点コンテンツを視聴する方式を提案してい

る．箱庭スタジアムは視点操作インタフェースとしての小型カメラを常に操作する必要があり，安定した映像を表示する工夫が必要である．そこで，北原ら⁹⁾は，人物ビルボード手法と箱庭スタジアムを組み合わせて，サッカーの試合の自由な視点からの視聴インタフェースを構築しつつ，各ゲームイベントごとの最適視点を複数の被験者による試行実験から代表点を求めて提示する方法を提案している．間瀬ら¹⁰⁾は多視点画像の視聴インタフェースで，釘付け視聴方式として制御の自由度を下げたインタフェースを提案している．一方，Shenら¹¹⁾は，多視点の監視カメラ映像のうち犯人の顔がもっともよく映るカメラ映像を選ぶための尺度としてQoVを定義している．

コンピュータグラフィックスの分野では，世界モデルが完全に記述されているために，仮想視点からの映像評価が容易であるが，視聴者の興味モデリングには至っていない．例えばコンピュータゲームのサッカー，格闘技，自動車レースの視点制御も同じ問題を抱える．多くは，ドライバなどアクターの一人称視点や鳥瞰視点などの擬似三人称視点のモード切り替え選択ができるようになっている．いずれもアクターの視点 (x, y, z) はすでに定まっていたので視方向の制御のみで十分である．多くはアクターの着目すべき対象が画面中心にくるように自動的に制御されている．そのうえで，プレイヤーが自由に視方向を切り替えられる機能がついていたり，ゲームクリエイターのデザインにおいて，適切な仮想視点を用意しておいてレンダリングをするようになっている．仮想3D空間をレンダリングする際の仮想カメラ制御については，カメラマン技法の知識を用いたVirtual Cinematographerの研究がある¹²⁾．

ひとつの動画内の重要部分の抽出や要約には，多くの関連研究が存在する．なかでも，ソーシャルアノテーション(同一映像を視聴することの可能な複数ユーザによるタグ付け)が注目を集めている．Miyamori¹³⁾らはインターネット掲示板の書込をTVへのアノテーションとみなすことで，アメリカンフットボールの試合の要約を試みている．また，青木ら¹⁴⁾は動画に対してリアルタイムにコメントがつけられるサービスであるニコニコ動画を取りあげ，そのコメントをアノテーションと捉えて，量的に解釈し，映像要約と動画の見所の検出を試みた．これらは単一映像に対するアノテーションであるが，我々は，一つのコンテンツを多視点で記録した映像群に対するソーシャルアノテーションからQoVを計算し視聴支援に用いることにする．

また，多視点画像群からその場を要約して記述する研究としては，Mase¹⁵⁾らのウェアラブル・ユビキタスコンピュータを使った体験記録がある．ポスター会場などの場を，来訪者が装着したカメラや会場に設置したカメラを使って多視点で記録し，来訪者の興味に沿って最適なカメラ映像セグメントを選び，要約映像を構築している．映像セグメントの評価に，

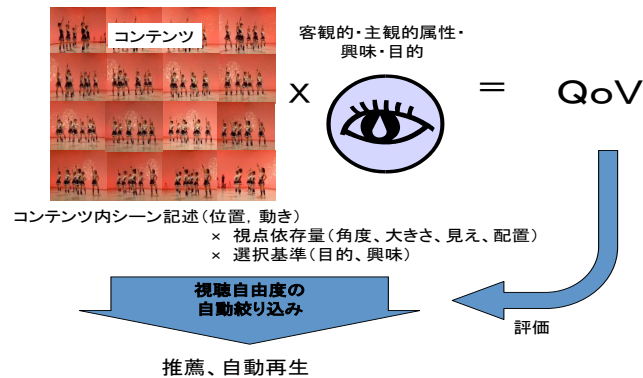


図 1 Quality-of-View:コンテンツの視点を定める評価尺度
Fig.1 Quality-of-View: a criteria to define a quality of content view

IrLED センサなどを使ったコンテキスト認識を用いて、トップダウンの尺度を用いており、利用者の目的によって要約の仕方を変更させる可能性を言及するにとどまっている。

3. Quality of View

放送局の映像プロデューサや映画監督（以下、ここでは演出者と呼ぶ）は、それぞれのコンテンツ（視聴プログラムや視聴対象）に対して最適なカメラ配置を考えて、そのカットやシーンを切り取ってつなぎ、番組や映画を編集している。演出者は、何らかの映像の評価尺度を体得しており、あるシーンをどの位置から映像フレームに納めると最適かを判断しつつ、決めていると考えられる。この評価尺度が定義できれば、スポーツやエンタテインメントなどの自由視点で視聴できるコンテンツがあたえられたときに、演出者をまねた仮想視点での視聴フレームを自動生成することも可能である。

ここで、従来型のマスメディアでは演出家の考える最適な観点からの最善の映像を制作することが必要とされたが、今後は、自由視点映像技術により、1つのコンテンツをさまざまな興味や観点で鑑賞できるようになる。すなわち、多様性のある評価尺度を与えることによって、1つのコンテンツを違う角度から楽しむことが望まれる。例えば、サッカーの試合であれば、戦略を確認するためにプレーを分析的に見ることもあれば、ファンである選手

がもっとも格好良く映る角度から見ることであろう。すなわち、興味が異なれば3次元コンテンツを鑑賞する際の視点の最適性は異なる。コンテンツを記述しておいて、視聴者の興味に基づいて、その見え方である画像フレームの視点と切り方を評価する尺度が必要である。

3.1 QoV の定義

そこで、視聴者の興味によって変動する評価尺度を、Quality-of-View(QoV) という視聴品質尺度でモデル化し定めることを提案する。具体的には図 1 に示すように、コンテンツ内の視対象の要素を記述し、その要素を物理量で評価しておき、一方、視聴者の目的や興味がそれらの要素をどのように価値づけて選択するかを計算機の内部表現として記述し定式化することを目指す。すなわち、次式のように時刻 t 、視点 $viewport(vp)$ 、視聴者特性 $viewer_attrib(va)$ を入力変数とし、これらを変数とするシーン記述 ($scene_desc$)、視点依存量 ($view_dependent$)、選択基準量 ($eval_criteria$) の3つを中間パラメータとして QoV を定義する。

$$QoV(t, vp, va) = h(scene_desc, view_dependent, eval_criteria) \quad (1)$$

ここで、時刻 t は、ある時間区間を代表する時刻である。QoV は静止画フレーム単位で定義することもあれば、画像系列で定義することもある。視点 ($viewport$) は、視点位置、視線方向ベクトル、画像フレームおよび焦点距離などの視聴フレームを定義する変数である。また、視聴者特性は、視聴者の目的や興味を記述する。さらに、各中間パラメータは各入力変

表 1 Quality-of-View 中間パラメータの例
Table 1 Examples of Quality-of-View Parameters

中間パラメータ	シーン記述	視点依存量	選択基準量
定義	シーンの 3 次元空間固有記述	任意視点からの見かけ量	目的・対象への興味
スポーツコンテンツの例	選手・ボール・ゴールの ID・位置・姿勢、コートの配置、プレーの内容、等	視点に対する画面上の選手の大きさ、顔の向き、選手の配置（見え隠れ、プレーの見やすさ、プレーの長さ、等	特定の選手のファン、コーチの視点、ゲームを楽しみたい、等
エンタテインメントの例（歌謡ショーや演劇）	出演者の ID・位置・姿勢・ジェスチャ・インタラクション、演技内容、ステージ形状、等	画面上の出演者の大きさ、顔の向き、出演者の配置、インタラクションの様子、など	特定の出演者のファン、全体を楽しみたい、演技の評価
遠隔医療・防犯・事件ニュースなど事象説明	人やモノの ID・位置・配置、作業の内容、事象の内容	モノの大きさや鮮明さ、画面構成、連続性	事象の詳細理解、経過の理解、関係の理解

数に依存する関数であり、「シーン記述 (scene_desc)」はシーン内の視聴対象の3次元世界における記述であり、サッカーなどのスポーツの場合、選手、ボール、フィールドであり、エンタテインメントであれば出演者、楽器、背景などで、場合によっては音楽など音響要素も含まれるであろう。これらの対象の3次元空間内での、位置、大きさ、方向、配置も記述される。つぎに「視点依存量 (view_dependent)」は、これらの対象をある視点から鑑賞したときの、画像フレーム内での、見かけの位置、見かけの大きさ、見え、配置などとなる。さらに「選択基準量 (eval_criteria)」は、コンテンツをどう見るといふ目的や興味であり、ある選手や演者への興味、分析・技術獲得のためのプレーの詳細、座席視点のシミュレーションなどがある。まとめを、表1に示す。

このQoVは、Shenら¹¹⁾が提案したQoVを拡張したものと考えることができる。Shenらは、多視点の監視カメラ映像のうち、監視の目的に沿った映像、すなわち被疑者の顔がもっともよく映る画像を選ぶための尺度としてQoVを定義している。上記の表現によれば、

$$QoV_{Shen}(t, viewport, surveillance) = h(scene_desc, view_dependent, SUSPECT)$$

であり、シーン記述と視点依存量を変数として評価するのみで、視聴者の“選択基準”は固定であったと考えることができる。

3.2 QoVの線形重回帰モデル

まず、QoV関数を線形重回帰分析で求める手法を検討した³⁾。多視点カメラでとった動画像の同時刻の画像(または同時区間の画像系列)を対象に、カメラごとの画像のQoV値を、時刻 t (またはある時刻 tt を含む時区間)、カメラ番号 i 、フレーム中の興味対象 O をもちいて、 $QoV(t, i, O)$ として求める。予め用意した因子関数で与えられる因子ベクトル $\mathbf{G} = (g_1, g_2, \dots, g_K)$ 、ただし $g_k = g_k(t, n, O)$ と、因子関数重みベクトル $\mathbf{w} = (w_1, w_2, \dots, w_K)$ を用いてその内積(線形加重和)で以下のように表現する。

$$QoV \approx \widehat{QoV}(t, i, O) = \mathbf{w} \cdot \mathbf{G} + c = \sum_{k=1, \dots, K} w_k g_k(t, i, O) + c \quad (2)$$

ここで、 K は因子関数の個数、 c は線形重回帰における定数である。こうしておいて、映像素材に対する視点に依存した様々な物理的観測値を使って因子関数を表現しておき、被験者実験によりQoV値を与え、重回帰分析で因子関数重みベクトル \mathbf{w} を同定して、モデルを完成する。実験に携わった被験者の観点で興味モデルが計算機の内部表現として得られる。因子関数としては、動画像を対象に出現度、広域度、中心度⁴⁾を、静止映像を対象に広域度、中心度、正面度³⁾を用意し、モデル化を試みた。

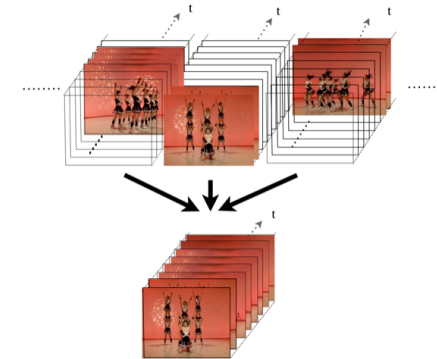


図2 多視点映像からの映像ストリーム構成: QoV値をもとに、最適な映像カットをつなぐ
Fig.2 Video stream composition from multiple video source based on QoV values

3.3 QoVの頻度値モデル

つぎに、QoVをボトムアップに構築する手法を示す。ここでは、視聴者の興味を誘引する手掛かりを提供しつつ、視聴者の視聴履歴を詳細に獲得できる視聴インタフェースを準備し、被験者実験により興味視聴対象ごとの視聴履歴を集める。前述の線形重回帰モデルが、コンテンツ内の対象に関する視点依存の物理量 $view_dependent$ をもとに因子関数をモデル化して興味の高さを明示的に尋ねたのに対し、このモデルは、視聴行動の履歴を視聴者の興味の発現ととらえて、ボトムアップに次式で $QoV(t, i, O)$ を直接求める方法である。

$$QoV \approx \widehat{QoV}(t, i, O) = histogram(t, i, O) \quad (3)$$

ここで、 $histogram(t, i, O)$ は、視聴対象 O を選んだ状況での、時刻 t 、カメラ番号 i の映像フレームの視聴頻度である。得られた視聴頻度分布を可視化したものを我々はとくにビューグラム($viewgram$)と呼ぶ。

3.4 QoVを利用した映像カット編集と自動生成

QoVが定義されると、図1に示したように、その値を評価して、ある時刻の、ある視聴者の目的にとって最適な視点を計算することができる。それに基づいて自動的に映像を編集して再生したり、視聴者の指定された好みにしたがって、最適な映像シーケンスを推薦するように使うことが可能となる。図2は、多視点で予め撮影されたコンテンツが与えられたときのQoV値利用のイメージである。多視点映像の場合には、撮影映像素材に限られるので、QoV値によって各時刻ごとにどのカメラ映像を選ぶかという比較的単純な問題になる。

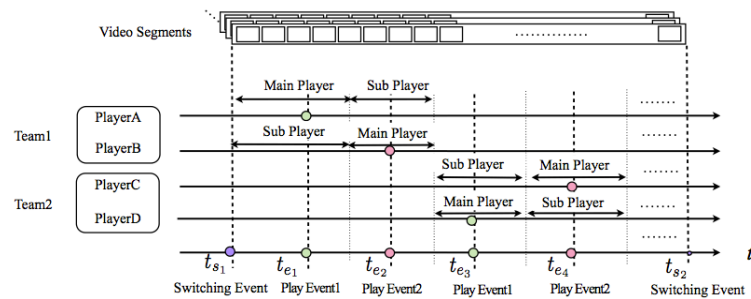


図 3 ビーチバレーの時系列モデル
Fig. 3 Flow Models of a Beach Volleyball.

が、カメラ映像のフレーム構成まで考慮すると、問題は複雑化していく。また、一般の映像演出では、大きな制作目的のなかで小ゴールが次々と設定され切り替えられる。この演出や制作の戦術やテクニックは脚本家や演出者の芸術的感覚や技量によるところが大きく、さまざまな手法が存在していると考えられるが、将来 QoV 値の評価が進むとこれらの演出に関する知識をつかって、飽きの来ない映像番組を自動的に生成することも可能になると思われる。これらは将来課題であり、本稿では、QoV 値の導出に焦点をあてる。

4. QoV の計算事例

前記の 2 つの方法で実際に QoV を求めた例を示す。

4.1 線形重回帰による QoV 計算

プレーヤ数が少なくプレイイベントが明確なビーチバレーを対象に多視点映像を記録し、プレーに着目した場合の評価モデルを求めた⁴⁾。

4.1.1 ビーチバレーの時系列モデル

ビーチバレーは、選手 2 人がチームとなって試合が行われる。選手同士は連携しながら刻々と動作内容を変化させる。このようなビーチバレーの試合の時系列モデルを図 3 のように表現する。ビーチバレーのプレイイベントとしては主に、サーブ、レシーブ、トス、スパイクの 4 つがある。プレイイベントを起こす選手を映像におけるメイン選手とし、メイン選手と同じチームの選手を映像におけるサブ選手として捉える。ここで、それぞれのプレイイベントをひとつの映像セグメントとして、ボールがネットを越える時刻をセグメントの切れ目とし、各セグメントに対して QoV を評価する。

4.1.2 映像の『良さ』に影響する因子の候補

前述の時系列モデルから、ビーチバレーの映像を評価する場合の QoV 評価関数を定義する。本稿では映像の物理的な観測値から評価できる映像の『良さ』に影響する因子の候補として、i) 出現性, ii) 広域性, iii) 中心性 および iv) 正面性 に注目し、それぞれ、1) 出現度, 2) 広域度, 3) 中心度 および 4) 正面度 を評価するための因子関数を定義する。これらを実験するための因子関数は、おもに興味対象の面積と位置および向きを用いて計算することができる。興味対象の位置は、フレームにおける興味対象の領域の重心位置とする。また、興味対象として、メイン選手、サブ選手、ボール、メイン選手とサブ選手を足合わせた領域を対象に注目し、因子関数をモデル化し、重み係数を主観評価によって得た QoV 値から重回帰分析で同定した。

4.1.2.1 出現度

出現度は対象とする時刻 t で代表する映像セグメントにおいて、興味対象 O が、カメラ i の映像において時間的割合としてどの程度映っていたかを示す尺度である。出現度を評価する因子関数を $g_{app}(t, i, O)$ として定義する。

4.1.2.2 広域度

広域度は、同様に、対象とする映像セグメント中に興味対象 O が映像に映っていた場合に、どの程度大きく映っていたかを示す尺度である。映像セグメント中の各フレームにおける興味対象 O の面積を正規化して当該セグメント全体について積算する。広域度を評価する因子関数を $g_{area}(t, i, O)$ として定義する。

4.1.2.3 中心度

中心度は、同様に、対象とする映像セグメントにおいて、興味対象 O がどの程度フレームの中心に映っていたかを示す尺度であり、映像の x 軸方向と y 軸方向のそれぞれに関して定められる。 x 軸方向と y 軸方向の中心度を評価する因子関数はそれぞれ $g_x(t, i, O)$, $g_y(t, i, O)$ として定義する。

4.1.2.4 正面度

正面度は、対象とする映像フレームの中で、興味対象 O がどの程度、視聴者と正対しているかを示す尺度である。水平方向の角度 (yaw) θ 、垂直方向の角度 ($pitch$) ϕ のそれぞれで定義する。 $g_\theta(t, i, O)$, $g_\phi(t, i, O)$ として定義する。なお、正面度は、文献³⁾において、ビーチバレー選手個人に興味があるときについてのみモデルを評価し、静止画フレーム単位でパラメータ同定をした。本稿では紙面の都合で省略する。



図 4 撮影した多視点ビーチバレー映像の例
Fig.4 Sample: Multi-View Video of Beach Volleyball.



図 5 PegScopeViewer の画面スナップショット
Fig.5 PegScopeViewer and its screen snapshot

4.1.3 被験者実験による QoV 値の収集とパラメータ同定

ビーチバレーを対象に実験を行った。コート半周囲にデジタルカメラ (CASIO EX-F1, 30fps, 640 × 480 pixels) を 9 台並べて撮影した多視点映像を用いた。図 4 に撮影した多視点映像の一部を示す。

撮影した多視点映像を時刻同期させ、セグメント抽出した映像 63 本 (7 セグメント × カメラ 9 台) に対して評価を行った。主観評価実験の被験者は 20 代の男性 10 名である。映像セグメントに対して、「各動画はプレーをどの程度分かりやすく表していますか？」と質

問し、11 件法で点数付けしてもらった。

得られた被験者の評価値はそれぞれ、各被験者ごとに正規化し、被験者間のばらつきを最小化した。主観評価による点数は順序尺度であるが、便宜的に比例尺度として扱った。さらに、各映像セグメントに対する被験者間のばらつきを調べ、分散の小さい上位 21 本を評価がそろったとして用いた。

上記の正規化で得られた評価値を QoV 値の目的変数、そして映像の各因子関数の値を説明変数としてまず単回帰分析を行い、自由度補正決定係数が 0.25 を越える因子を候補として抽出した。その結果、メイン選手の出現度と x 軸方向の中心度、サブ選手の出現度と x 軸、y 軸方向の中心度、ボールの出現度と x 軸方向の中心度が有用な因子の候補として抽出された。抽出した因子関数の中で相関が 0.4 以上ある因子関数同士について、決定係数の低い方の因子関数を QoV 評価関数を構成する因子関数から外し、結果的にメイン選手の出現度、サブ選手の x 軸方向と y 軸方向の中心度、ボールの出現度を有用な因子として抽出した。

さらに、評価の定まった映像に対する QoV 値を目的変数、抽出された因子関数を説明変数とし重回帰分析をおこなった。得られた各因子関数の重みを式 (2) にあてはめることで、カメラ i が映す時区間 t の映像に対するプレー全般 $Play_{all}$ に着目した場合の QoV 評価関数は式 (4) として定義される。

$$\widehat{QoV}(t, i, Play) = 0.16g_{app}(t, i, MainPlayer) + 0.44g_x(t, i, SubPlayer) + 0.21g_y(t, i, SubPlayer) + 0.33g_{app}(t, i, Ball) + 0.15 \quad (4)$$

各因子関数の重みは、平均 0、分散 1 に正規化した値で重回帰分析を行った標準偏回帰係数であり、QoV 値への影響度を表す値である。式 (4) のあてはまりの良さは、相関が 0.91、自由度補正決定係数が 0.77 と、十分に良い精度で適合できたと言える。この方法により、様々な因子関数でシーンを記述し興味対象に依存した QoV 尺度を定義することが可能となる。なお、因子関数を与えるための画像特徴は手作業で入力している。

4.2 Viewgram による QoV 計算

ここでは、エンタテインメントの題材として、チャリーディングを多視点で撮影したものを対象とした。複数のチャリーダが演技をするため、興味対象を個々の演技者とする事で、QoV の効果を示しやすいと考えられる。

4.2.1 多視点映像視聴インタフェース

単純にカメラを切り替えながらフレーム内の対象にフォーカスする視聴インタフェースでは、

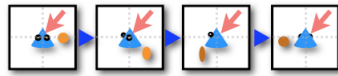


図 6 釘付け視聴方式:釘付け対象を常に画面の一定位置に配置する
Fig. 6 PesScope viewing method

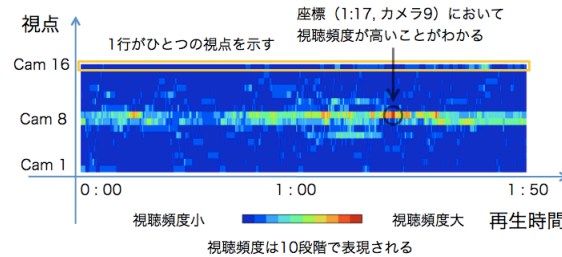


図 7 Viewgram の一例
Fig. 7 An example of Viewgram

目的の映像を得るために非常に煩雑な操作が必要となる．そこで、まず我々は、PegScope-Viewer と呼ぶ、被視対象者（被視対象物）を先に指定したのちにカメラ選択やズーム操作する多視点映像視聴インタフェースを開発した¹⁶⁾．このインタフェースは利用が簡便であり、とくに興味対象が明確な場合はペグ点リストと呼ぶあらかじめ定義された視対象の直接操作による選択が快適なインタフェースを提供できる．図 5 にそのインタフェース画面を示す．多視点映像として取得したコンテンツにおいて、多方向からの簡便な視聴操作を支援する際に、視聴に先んじて注目対象物を選び、それを画面の一定した場所に固定したまま、視点位置を移動させて映像を提示することができる（図 6）．

この視聴方式を用いることで、視聴者の興味の対象が特定でき、視聴者の興味の対象に沿った視聴頻度データ $histogram(t, i, O)$ を収集することが可能となる．こうして作られた視聴者の興味の対象ごとの視聴頻度データは、将来、多くの視聴者の間で互いに構築・共有する事によって、多視点映像視聴を広く支援するという展開もある．

4.2.2 被験者実験による Viewgram の構築

図 7 に実験で得られた viewgram の例を示す．グラフの横軸は映像の動画内時刻 t を、縦軸はカメラ視点 i を示す．グラフ上で視聴頻度の大きさを 10 段階の色相によって表現している．視聴頻度データを View-gram という形で提示する事で視聴者は再生時間と視点に対



図 8 釘付け点 girlE について生成された Viewgram: $\widetilde{QoV}(t, i, girlE)$
Fig. 8 Viewgram for girlE: $\widetilde{QoV}(t, i, girlE)$

する視聴頻度を直感的に理解することが可能になる．

多視点映像コンテンツとして、チアリーディングの一連の演技（1分50秒、300fps）を用いた．20代～30代の16名（男13名、女3名）に、釘付け視聴方式を用いてそれぞれ15分ずつ自由に何度でも繰り返してコンテンツを視聴してもらい、その全視聴履歴を記録した．釘付け点視聴方式において、あらかじめ、image center（画像の中心）、stage center（チアが演技している空間の中心）、そして、7人のチアリーダ各々に注目することができる girlA から girlG、の合計9つの釘付け点を用意しそれらを自由に選んでもらいつつ視聴してもらった．

図 8 に、釘付け点 girlE を選択して視聴した履歴についてのみ集約した viewgram ($\widetilde{QoV}(t, i, girlE) = histogram(t, i, girlE)$) を示す．前半部分（図上、左側）では正面から見て右側の視点（図中の上方）の視聴頻度が高くなっている事が分かる．写真左下（ステージは画像左方向）からも分かるように、これは girlE が見やすい視点を的確に示している．中盤では視聴頻度の高い部分が正面に移行するが、これは手のひらで形作る人文字のシーンで、girlE も見やすい位置の視点であり、位置と演技の意味を考慮した結果一番良い視点選ばれていると言える．また、viewgram 上で、各フレームで最大の QoV 値を与えるカメラを選びつつピタビアルゴリズムを援用してなめらかに接続した軌跡を白線で描画している．これは、今回の実験に参加した被験者の視聴傾向に基づく、girlE に興味をもつ人向けの推薦映像シーケンスとして利用することができる．

5. おわりに

多数のカメラで撮影した多視点映像の視聴インタフェースでの利用を前提として、コンテンツに対する視聴者の興味と視点の関係をモデリングし、定量的な尺度として視聴品質 (QoV) を導いた。時刻、視点、視聴者特性を一般化した入力変数とし、それらを変数とするシーン記述、視点依存量、選択基準量の3つを具体的な中間パラメータとした定式化を提案した。そして、重回帰モデルと視聴頻度によるアプローチをそれぞれ、トップダウンとボトムアップの両面からのアプローチと位置づけ、QoV モデルへのあてはめを行い、それぞれのアプローチについて、本提案の QoV の定義に沿った一貫した説明ができることを示した。

重回帰モデルによるアプローチが内容記述と因子関数に用いる物理特徴の定義を必要とするのに対し、視聴頻度アプローチでは、視対象を記述するアノテーション (ベグ点) の付与さえあれば、視聴傾向から QoV が直接定義できるという利点がある。一方視聴頻度アプローチは、中間パラメータが直接制御できないため、視聴者の視聴振る舞いに大きく依存する。今回は、重回帰モデルでは、映像セグメントの評価による内部パラメータの推定を行ったが、さらに上位からのアプローチとして演出家の知識を直接、式に書き起こすことも考えられる。

いずれの方法も、興味対象として、人物やボールだったり、プレーイベントなど人物の配置にかかわるものなどを前提にして考えるため、これらの3次元空間アノテーション情報の自動的な抽出法と画像特徴の選別方法を今後確立することが重要である。また、同時に、これらの情報を利用者が自ら付与するインタフェースを用意して、集合知として収集することも考えられる。

謝辞 本研究を進めるにあたって、多視点映像の撮影に協力いただいたプロジェクトメンバーに感謝する。本研究の一部は独立行政法人情報通信 (NICT) の委託研究「三次元映像通信・放送のための中核的要素技術」および独立行政法人科学技術振興機構 CREST の支援による。

参 考 文 献

- 1) Kanade, T. and Narayanan, P.: Virtualized Reality: Perspectives on 4D Digitization of Dynamic Events, *IEEE Computer Graphics and Applications*, Vol.27, pp.32-40 (2007).
- 2) 福嶋慶繁, 他: 多視点・多聴点データ取得システムを用いた自由視聴映像生成, 電子情報通信学会論文誌 D, Vol.J91-D, No.8, pp.2039-2041 (2008).
- 3) 永井有希, 丸谷宜史, 梶田将司, 間瀬健二: 視聴者の関心を考慮した Quality-of-View に基づく多視点画像の評価, 信学技報, PRMU210-231, 埼玉大学, pp.147-152 (2011-02).
- 4) 永井有希, 丸谷宜史, 梶田将司, 間瀬健二: プレーに着目したスポーツ多視点映像の評価尺度, 情報処理学会技術研究報告, Vol.2011-EC-19, 慶応大学, pp.1-6 (2011-03).
- 5) 丹羽宏介, 東海彰吾, 川本哲也, 藤井俊彰, 丸谷宜史, 梶田将司, 間瀬健二: 多視点映像視聴支援の為にソーシャルな視聴履歴の利用, 情報処理学会技術研究報告, Vol.2011-EC-19, 慶応大学, pp.1-6 (2011-03).
- 6) 南谷真哉, 北原 格, 亀田能成, 大田友一: 変型ビルボードを用いた人物像の提示による複合現実感卓上作業の遠隔共有, 日本バーチャルリアリティ学会論文誌, Vol.13, No.3, pp.363-373 (2008).
- 7) Moezzi, S., Katkere, A., Kuramura, D.Y. and Jain, R.: Immersive Video, *Virtual Reality Annual International Symposium*, Vol.0, Los Alamitos, CA, USA, IEEE Computer Society, p.17 (1996).
- 8) Inamoto, N. and Saito, H.: Intermediate View Generation of Soccer Scene from Multiple Videos., *ICPR (2)'02*, pp.713-716 (2002).
- 9) 北原 格, 橋本浩一郎, 亀田能成, 大田友一: サッカーの自由視点映像提示における気の利いた視点選択手法, 日本バーチャルリアリティ学会論文誌, Vol.12, No.2, pp.171-180 (2007).
- 10) 間瀬健二, 東海彰吾, 川本哲也, 藤井俊彰: 多視点画像の釘付け視聴方式と操作インタフェースのデザインに関する考察, ヒューマンインタフェース学会研究会報告, 11-1, Vol.11, No.1, pp.7-12 (2009-03).
- 11) Shen, C., Zhang, C. and Fels, S.: A Multi-Camera Surveillance System that Estimates Quality-of-View Measurement, *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, pp.III-193 - III-196 (online) (2007).
- 12) He, L.-w., Cohen, M. F. and Salesin, D. H.: The virtual cinematographer: a paradigm for automatic real-time camera control and directing, *SIGGRAPH '96*, New York, NY, USA, ACM, pp.217-224 (1996).
- 13) Miyamori, H., Nakamura, S. and Tanaka, K.: Generation of views of TV content using TV viewers' perspectives expressed in live chats on the web, *Proceedings of the 13th annual ACM international conference on Multimedia*, pp.853-861 (2005).
- 14) 青木秀憲, 宮下芳明: ニコニコ動画における映像要約とサビ検出の試み, 情報処理学会研究報告, Vol.MUS-50, pp.37-42 (2008-05-21).
- 15) Mase, K., Sumi, Y., Toriyama, T., Tsuchikawa, M., Ito, S., Iwasawa, S., Kogure, K. and Hagita, N.: Ubiquitous Experience Media, *IEEE Multimedia*, Vol.13, pp.20-29, (2006).
- 16) 丸谷宜史, 間瀬健二, 東海彰吾: タグ情報に基づく多視点カメラ映像の視聴支援, 第13回画像の認識・理解シンポジウム (MIRU2010), 北海道釧路市 (2010).