

チアリーディング映像における 多視点視聴支援のためのフォーメーション分析

View Stream Generation Using Objects' Formations Based on QOV for Multi-View Cheerleading

永井 有希*1 平野 靖 *2 梶田 将司 *3 間瀬 健二 *1
Yuki Nagai Yasushi Hirano Shoji Kajita Kenji Mase

*1名古屋大学大学院情報科学研究科 *2山口大学大学院医学系研究科
Graduate School of Information Science, Nagoya University Graduate School of Medicine, Yamaguchi University

*3名古屋大学情報連携統括本部
Information and Communication Technology Services, Nagoya University

Multi-view video makes users possible to watch sports events more freely. However, users should select a camera which captures the best view every time, and users who don't know well how to watch multi-viewer might be confused. In this paper, we propose an assistant method for multi-camera view for cheerleading. We define a confidence measure, QOV. This measure evaluate the camera's view performance for each time. By comparing each camera's QOVs, systems can select the most appropriate cameras to perform user's aim. Formations are specific features for cheerleading, so QOV takes into account a cheerleader's formation.

1. はじめに

多数のカメラで撮影され相互に時間同期された映像を多視点映像という。現在、多視点映像を用いたサービスが盛んに研究されており、様々な分野で関心を集めている [Lou 05]。放送における多視点映像の応用として、スポーツイベントなどの映像をユーザが好きな角度から視聴できるようなシステムが考えられている [Liu 09, Inamoto 03]。しかし多視点映像は視聴の自由度が高い反面、ユーザは多数のカメラから常に視聴する映像を選択しなければならない。また、慣れないユーザはどのように視聴したら目的に合った見方ができるか分からないだろう。このように、ユーザにとっては映像選択が負担となってしまうことが考えられる。

そこで本研究では、多視点映像からユーザの視聴目的に合った映像ストリームを自動的に生成し、提示することで多視点映像の視聴支援を行うことを目的とする。映像ストリームを生成する方法として、Quality-Of-View(QOV)[Shen 07]という評価尺度を用いる。QOVとは、視聴目的からフレームの価値を評価する尺度である。このQOVを用いて最適なカメラの映像を選択していくことで、多視点映像からユーザの視聴目的に合った映像ストリームの生成を行う。多視点映像として、本研究ではチアリーディング映像を対象とする。チアリーディング映像は、選手がフォーメーションを構成しそのフォーメーションを変化させていく。そこで本研究ではフォーメーションの情報を用いてQOVを評価することで映像選択を行い、視聴目的に合った映像ストリームを生成する方法について述べる。

2. QOVに基づくチアリーディング映像のフォーメーションを考慮した生成手法

本節では、チアリーディング映像を題材とした多視点映像を評価する手法について説明する。

文献 [Shen 07] では被写体を観測できる環境 [Svoboda 05] を用いて、QOVをカメラに対する被写体の角度と距離により

評価する。しかしチアリーディング映像は複数の被写体が存在し、それぞれの被写体間の関係を考慮する必要がある。またチアリーディング映像を視聴するユーザの目的も様々である。指導者の観点から視聴したい場合もあれば、エンターテイメントとして視聴したい場合もある。これらの場合で被写体への注目方法は異なってくる。文献 [Shen 07] におけるQOVは、このような被写体間の関係や多様なユーザの視聴目的を考慮していない。そこで本研究では文献 [Shen 07] により定められたQOVをより一般化し、再定義を行う。

2.1 一般的なQOVの定義

カメラ C_i ($i \in \{1, \dots, n\}$) が撮影する時刻 t での映像を $x_i = f_i(t)$ とし、 x_i に対するQOVを次式 (1) のように定義する。

$$QOV(x_i) = \left(\prod_{k=1}^{N_c} \text{Con}_k(x_i) \right) \times \left(\sum_{j=1}^{N_a} w_j \text{Asp}_j(x_i) \right) \quad (1)$$

ただし

$$\text{Con}_k(x_i) \in \{0, 1\}, \quad \text{Asp}_j(x_i) \in [0, 1], \quad \sum_{j=1}^{N_a} w_j = 1 \quad (2)$$

とする。Asp_j は計算機が映像を評価する一つの観点を示す。w_j は Asp_j の重みである。視聴目的に対して計算機が映像を評価する観点 Asp_j を複数組み合わせることで、多様な視聴目的を柔軟に評価可能とする。別の見方をすれば、Asp_j の組み合わせは、視聴目的という上位の概念を計算機で評価可能な尺度に落とし込んだものである。また Con_k は映像が満たすべき一つの制約を示す。例えば、制約 Con_k は A さんが映っている映像のみを視聴したい場合には、A さんが映っていれば 1、映っていなければ 0 と定義される。n_c, n_a はそれぞれ制約 Con_k と評価する観点 Asp_j の総数を示す。

2.2 フォーメーションパターンの表現

本稿で用いるチアリーディングにおけるフォーメーションパターンの表現について説明する。ここでフォーメーションパターンとは、理想的に想定されたフォーメーションの形であ

る。また、フォーメーションパターンに基づき実際に選手により構成されたものをフォーメーションと呼ぶ。

本研究では図1のようにフォーメーションパターンをグラフで表現する。選手一人をノード $n_k (k = 1, \dots, N_p)$ とし、選手 n_a と選手 n_p の間をエッジ e_{ab} として表現する。

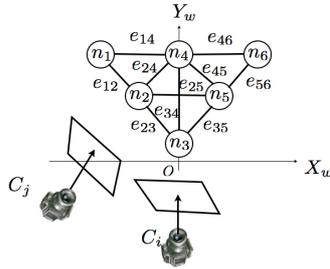


図1: フォーメーションパターンのグラフ表現

フォーメーションパターンは時間により変化していく。ひとつのフォーメーションパターンを Pat_j として表すと、フォーメーションは時間区間 $I_j = [t_{s_j}, t_{e_j}]$ においてフォーメーションパターン Pat_j を維持する。

2.3 QOV に基づく多視点映像からの映像ストリーム生成

多視点チアリーディング映像のフォーメーションパターン総数が n_f である場合には、全時間区間は $I = [t_s, t_e] = \{I_1, \dots, I_{n_f}\}$ と表現できる。すべての時刻 $t \in I$ において、各カメラの映像 $x_i = f_i(t) (i = \{1, \dots, n\})$ を QOV を用いて評価することで多視点映像から視聴目的に合った映像ストリームを生成する。つまり、 $I = [t_s, t_e]$ における多視点映像から生成した映像ストリームを $STR = (s_{t_s}, \dots, s_{t_e})$ とすると、時刻 $t \in I$ の映像は式 (3) で定義される。

$$s_t = \arg \max_{i \in \{1, \dots, n\}} QOV(x_i) \quad (3)$$

3. チアリーディング映像を評価する尺度

チアリーディング映像の特徴として、複数の選手がフォーメーションを形成し、それを変化させていくということがある。カメラ選択には、このフォーメーションを考慮すべきであると考えられる。本節ではフォーメーションを考慮した映像評価のための尺度について定義を行う。

3.1 対称性の尺度

チアリーディングのフォーメーションにおいては、対称性が重要な要素となる。そこで QOV の観点 Asp として対称性を評価する尺度をもうける。図1は上から見た場合のフォーメーションパターンを表すグラフである。カメラ画面においては、このグラフの形が撮影する方向や位置によって変化する。図1におけるカメラ C_i とカメラ C_j が映す映像のグラフを図2に示す。カメラ C_i には、ノード n_4 がノード n_3 に隠れて映っていない。そのような映像に映っていないノードとエッジは定義しない。

図1のフォーメーションパターンは、 Y_w 軸を中心に線対称であることが理想であるとする。この場合 $\{(e_{12}, e_{56}), (e_{23}, e_{35}), (e_{14}, e_{46}), (e_{24}, e_{45})\}$ が対称なエッジの組である。この対称なエッジの組の映像に映る長さが同じである程、よりフォーメーションの対称性が高い映像である。任意の対称なエッジの組 (e_{ab}, e_{cd}) の対称性の評価を式 (4) で定義する。

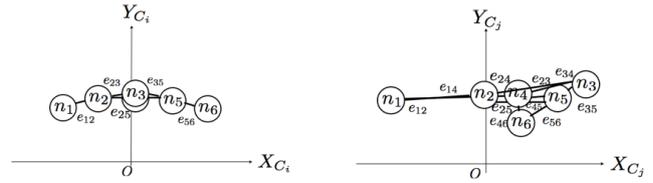


図2: [左] カメラ C_i の映像 [右] カメラ C_j の映像

$$\text{sym}((e_{ij}, e_{kl})) = \frac{\min\{|e_{ij}|, |e_{kl}|\}}{\max\{|e_{ij}|, |e_{kl}|\}} \quad (4)$$

フォーメーションがあるフォーメーションパターン Pat_j のとき、 Pat_j に含まれるすべての対称なエッジの組集合を Es_j とする。ただし $(e_{ab}, e_{cd}) \in Es_j$ は対称なエッジの組を表し、映像に映らなかったノードをつなぐエッジは長さ0とする。このときのカメラ C_i の映像 $x_i = f_i(t) (t \in I_j)$ における対称性の高さを式 (4) を用いて式 (5) と定義する。ただし m は対称なエッジの組数を示す。

$$\text{Asp}_1(x_i) = \frac{\sum_{(e_{ab}, e_{cd}) \in Es_j} \text{sym}(e_{ab}, e_{cd})}{m} \quad (5)$$

これは QOV を評価する観点の一つであり、映像に映る対称性の高さを評価する尺度である。

3.2 非対称性の尺度

フォーメーションの出来映えを確認したいときなど、フォーメーションが歪んでいる場合にそれがよく分かる映像を視聴したい場合がある。そこで3.1における対称性の尺度とは反対に、 Asp_1 を用いて非対称性を評価する尺度を定義する。

$$\text{Asp}_2(x_i) = 1 - \text{Asp}_1(x_i) \quad (6)$$

3.3 直列性の尺度

フォーメーションパターンにおいて、前後に選手が整列する場合がある。例えば図1において、ノード n_3 とノード n_4 は前後関係にあり、実際に構成されたフォーメーションが理想系に近ければ近いほど、 Y_w 軸方向から見た場合には n_4 は見えなくなる。左右の関係も同様で、 X_w 軸方向から見た場合には、 n_4 と n_6 は n_1 に、 n_5 は n_2 により見えなくなる。ここで、 $\{e_{34}\}$ の長さが最も短くなる映像が Y_w 軸方向における最も直列性の高い映像であり、 $\{e_{14}, e_{46}, e_{25}\}$ の合計の長さが最も短くなる映像が X_w 軸方向における最も直列性の高い映像であるとする。このような直列性の尺度を定式化する。

同軸方向に直列するエッジ集合を $El = \{e_1, \dots, e_k\}$ とすると、直列性の評価は式 (7) である。

$$\text{lin}(El) = \frac{1}{\exp(\sum_{e_l \in El} |e_l|)} \quad (7)$$

エッジの長さの合計 $\sum_{l=1}^k e_l$ は範囲 $[0, \infty)$ の間で定義される。 $\sum_{l=1}^k |e_l| = 0$ のとき最大値 $\text{lin}(El) = 1$ となり、合計の長さが増えるほど単調増加し $\sum_{l=1}^k |e_l| \rightarrow \infty$ で $\text{lin}(El) \rightarrow 0$ となる。

式 (7) を用いて、カメラ C_i の映像 $x_i = f_i(t) (t \in I_j)$ における同軸方向の直列性の高さを式 (8) と定義する。 El_j はフォー

メーションパターン Pat_j における同軸方向に直列するエッジ集合である.

$$Asp_3(x_i) = \text{lin}(El_j) \quad (8)$$

3.4 広域性の尺度

チアリーディングには見せ場となるシーンがいくつかあり、難易度の高い技や特殊な技により決めポーズをとる選手がいる. そのような選手を大きく映す映像が視聴したい場合など、被写体の大きさを評価する必要がある. そこでこれを広域性の尺度として評価する. フォーマーションパターン Pat_j において注目するノード間のエッジ集合を $E_w = \{e_1, \dots, e_k\}$ とすれば、広域性の尺度を式 (9) として定義する.

$$Asp_4(x_i) = 1 - \frac{1}{\exp(\sum_{|e_w \in E_w} |e_w|)} \quad (9)$$

3.5 ノード数の尺度

映像にどれだけ多くノードが映っているかを評価する尺度 $ps_5(x_i)$ を定義する. 映像 x_i に映るノード数を $\text{Num}(x_i)$ とし、選手の総数を N_p とする. このときのノード数の評価を式 (10) で定義する.

$$Asp_5(x_i) = \frac{\text{Num}(x_i)}{n_p} \quad (10)$$

4. 提案手法の評価

実際に撮影した多視点映像データを基に、エンターテイメントと指導者のための映像ストリーム生成について述べ、本手法の考察を行う.

4.1 収集データ

3 節で本研究で提案した QOV に基づく映像ストリーム生成手法を評価するために、7 名の選手で構成される多視点チアリーディング映像を撮影した. 図 3 に示すように半円状に並べたハイスピードカメラ (CASIO EX-F1, 300fps, 512×384 pixels) を 16 台使って撮影を行った. 各カメラは原点 O を中心した半径 537cm の半円状に並び、視線は原点 O を向く. 各カメラ間の角度は約 12 度である. 図 4 に撮影した多視点映像の一部を示す.

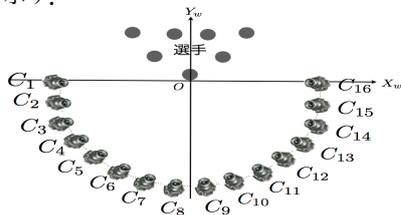


図 3: データ収集におけるカメラ配置

フォーマーションは全部で 9 パターン観測された. 観測されたフォーマーションを図 5 に示す. なお、図 5 には簡略化のためにエッジの表記を省略した. Pattern8 と Pattern9 の二



図 4: 撮影した多視点チアリーディング映像の例

重丸は、上下に重なるノードを表す. これらのノードは、実際には内側のノードを示す選手を外側のノードを示す選手が下となり支える組体操になっている.

4.2 QOV の評価式

3 節で定義した尺度から、式 (5), (6), (8), (6), (9),(10) を用いて QOV の評価式を式 (11) のように定める.

$$QOV(x_i) = w_1 Asp_1(x_i) + w_2 Asp_2(x_i) + w_3 Asp_3(x_i) + w_4 Asp_4(x_i) + w_5 Asp_5(x_i) \quad (11)$$

この評価式の重みベクトル $\mathbf{w} = (w_1, w_2, w_3, w_4, w_5)$ の値を変えることと評価するエッジ集合を定めることにより、様々な視聴目的に応じた映像ストリーム生成を実現する. 制約 Con は今回は考慮しない. 式 (11) を用いてエンターテイメントでの視聴と指導者のための映像ストリーム生成について評価を行う.

エンターテイメントとして視聴する場合には、フォーマーションがきれい選手全体が大きく画面に映っているような映像を視聴できるようにしたい. そこで映像のきれいさと対象の大きさをはかるために、対称性、直列性、広域性の尺度を用いる. エンターテイメントの映像を評価する重みベクトルは、 $\mathbf{w} = (\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3}, 0)$ として定める.

一方で指導者として視聴する場合には、選手が多数確認でき、実際に構成したフォーマーションが非対称的であるような場合にはそれを分かりやすく映すことが必要であるとする. また、初心者の選手や他の選手に比べ難易度が高い技をする選手など、注目したい対象があると考える. そこで指導者が視聴する場合には非対称性、ノード数、広域性の尺度を用いて映像を評価する. 指導者のための映像を評価する重みベクトルは $\mathbf{w} = (0, \frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3})$ と定める.

直列性を評価するエッジ集合は、各フォーマーションパターンにおいて Y_w 軸方向に直列しているノード間のエッジ集合として定める. 広域性を評価するエッジ集合は、エンターテイメントでは選手全体の広がり評価するために隣接するノード間の全エッジ集合として定める. 指導者の場合は各フォーマーションパターンにおいてそれぞれ定める. 例えば Pattern1 では全エッジ集合の、Pattern7 ではエッジ集合 $\{e_{35}, e_{56}, e_{24}, e_{47}\}$ の広域性を評価する. また直列性は Y_w 軸に直列に並ぶエッジ集合に対して評価する.

4.3 映像選択の考察

4.2 で定めた評価式により映像選択を行った場合の結果について考察を行う. 図 6 は Pattern1 においてエンターテイメントと指導者の尺度で選択された映像である. 図中には検出されたノードと対称性の評価に関わるエッジが示してある. 左の図はカメラ 16 台のうち、エンターテイメントの尺度で最も評価が高かったカメラ C_9 の映像である. 右の図は指導者の尺度で最も評価が高かったカメラ C_{16} の映像である. カメラ C_6 の映像からは、対称性に関わるエッジの組が (e_{34}, e_{56}) の一組しか検出されなかったが、このエッジの対称性が良く、高評価となった. さらに Pattern1 には Y_w 軸に直列なノード集合がないため、直列性の尺度はすべてのカメラで 0 となる. それゆえ直列性は評価に関わらず、エッジの対称性の尺度に依存する結果となった. また指導者で選択された C_{16} の映像は、対称性に関わる全てのエッジが検出された. しかしそれらの対称性が悪かったため非対称性の尺度で高評価となり、ノードも多く検出されたので、高く評価された.

エンターテイメントの尺度により評価された映像は構図が整っており、よい映像選択が出来た. 一方、指導者の尺度では

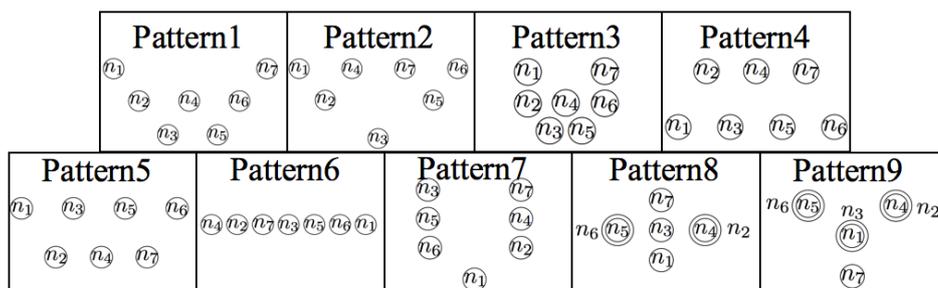


図 5: 観測されたフォーメーションパターン

非対称性によりフォーメーションが歪んでいないかを確認し、歪んでいる場合にはそれがよく分かる映像を選択したい。それに対し、選択された映像は確かに全員が閲覧できる画面にはなっているが、フォーメーションの歪みは分かりにくい角度となってしまった。広域性の尺度はエンターテイメントと指導者のどちらの評価にも、ほとんど影響を与えなかった。広域性の尺度のみによる評価結果はどのカメラの映像にも値にばらつきがでなかったためである。今後は評価結果が分散するような広域性の尺度を検討する必要がある。

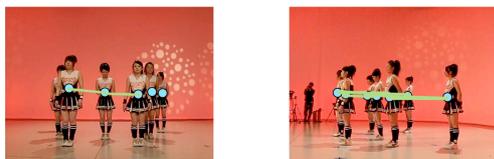


図 6: Pattern1:エンターテイメント (左) と視聴者 (右) の映像

図 7 は Pattern7 においてエンターテイメントと指導者の尺度で選択された映像である。右の図に記したエッジは対称性に関わるエッジであり、左の図に記したエッジは指導者は指導者の尺度において広域性を評価するエッジである。Pattern7 では直列に並ぶノード $n_2, n_3, n_4, n_5, n_6, n_7$ は組体操を完成させる準備をしている。エンターテイメントの尺度で最も高評価だったカメラ C_9 の映像は、直列性を評価するエッジ集合は一つも検出されず直列性が高く評価された。また図 6 の場合と同様に対称性も高く評価された。指導者の尺度では、カメラ C_{14} の映像が最も評価が高かった。広域性をはかるエッジ集合を $\{e_{35}, e_{56}, e_{24}, e_{47}\}$ として定めたことで、エンターテイメントの場合とは対称的に組体操の準備の様子がよく分かる映像を選択することができた。

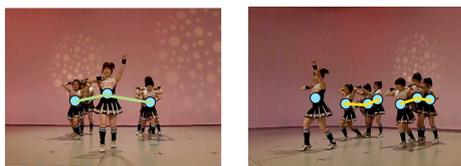


図 7: Pattern7:エンターテイメント (左) と視聴者 (右) の映像

今回は各尺度に対して重みを同じとして定めたが、尺度によって評価結果の値にばらつきがある。たとえば広域性は評価するエッジの数を増やすとほぼ 0 に近く評価されしまう。それゆえ他の尺度と同じ重みにしたためほとんど結果に影響しなくなってしまった。また対称性の尺度も、ノードが検出されない場合には対称的であっても正しく評価することができない。このように、今後は各尺度の評価値のばらつきを同程度にすること、重みベクトルの値や、各尺度における評価方法の再検

討を行う必要がある。

5. おわりに

本稿ではチャリーディングにおける多視点視聴を支援することを目的に、ユーザの視聴目的に合った映像ストリームを生成するための映像評価手法を提案した。本手法では、ユーザの視聴目的から映像を評価する尺度として QOV を再定義し、この QOV を用いて映像選択を行う。チャリーディングにおけるフォーメーションをグラフで表現し、QOV は各映像に映るグラフの形状を評価することで定められる。

今後は各評価尺度をより精度よくすることが課題である。今回、広域性はエッジの長さの和としたが、本来は被写体全体の面積で評価する必要がある。また更にに多様な視聴目的に対応できるように、評価する新しい尺度とその組み合わせ方法を検討する必要がある。

謝辞

本研究の一部は総務省 SCOPE の委託研究 (082306005) および独立行政法人情報通信研究機構 (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」によって行われた。

参考文献

- [Inamoto 03] Inamoto, N. and Saito, H.: Fly-through viewpoint video system for multi-view soccer movie using viewpoint interpolation, *Visual Communications and Image Processing*, pp.1143 - 1151(2003).
- [Liu 09] Liu, J., Tong, X., Li, W., Wang, T., Zhang, Y., Wang, H., Yang, B., Sun, L. and Yang, S.: Automatic Player Detection, Labeling and Tracking in Broadcast Soccer Video, *Pattern Recognition Letters*, Vol.30, Issue 2, pp. 103-113(2009).
- [Lou 05] Lou, J.G., Cai, H. and Li, J.: A real-time interactive multi-view video system, *Proceedings of the 13th annual ACM international conference on Multimedia*, pp.161 -170, New York, NY, USA (2009).
- [Shen 07] Shen, C., Zhang, C. and Fels., S.S.: A Multi-Camera Surveillance System that Estimates Quality-of-View Measurement, *The IEEE International Conference on Image Processing (ICIP)*, pp. 193-196 (2007).
- [Svoboda 05] Svoboda, T., Martinec, D. and Pajdla, T.: A convenient multi-camera self-calibration for virtual environments, *Teleoperators and Virtual Environments*, pp.405-422 (2005).