

## オーケストラ撮影における楽譜を利用したカメラワークの計画

柴 貞行\* 井上 亮文\* 平石 絢子\*  
高久 宗史\* 重野 寛† 岡田 謙一†

e-mail: shiba@mos.ics.keio.ac.jp

映像制作に関する専門知識や経験に乏しい一般ユーザでも撮影を行う機会が増えている。そこで本研究では、シナリオが存在するシーンの映像化において、複数カメラによる複数被写体の効果的な撮影を目的としたカメラワークの計画手法を提案する。まず、シナリオから撮影すべき被写体を選び出し、次に、被写体の役割と直前の撮影状態、カメラの配置状況などから算出される優先度に応じて各被写体へ割り振るカメラを決定する。これにより、撮影環境の変動と複数ユーザからの様々な編集要求に対応可能なカメラワークを自動的に計画・生成する。以上の手法をオーケストラ演奏を撮影対象とした場合の具体的実現方法について述べる。

## Score-Based Camerawork Planning for Orchestra Shooting

Sadayuki Shiba\*, Akifumi Inoue\*, Junko Hiraishi\*,  
Hirofumi Takaku\*, Hiroshi Shigeno†, and Ken-ichi Okada†

Shootings have been receiving increasing attention even from those who have poor experiences about video making. In this paper, we propose an effective camerawork planning method to shoot a scenario-based scene. First, subjects are picked out from the scenario. Second, a camera is allocated to each subject in accordance with the priority calculated from its role, previous shooting state and cameras location. Herewith camerawork which reacts to the shift of shooting environment and to the variety of users' edit requirements becomes possible. The proposal method applies to an orchestra shooting.

### 1 はじめに

高速ネットワークが手軽に利用できるようになった昨今、誰もが日常的に様々な Web コンテンツへアクセスするようになってきている。コンテンツには文字や静止画だけではなく、動画やアニメーション、音楽を盛り込むことが容易になり、リッチなコンテンツが普及しつつある。特に動画に関しては専門的な機材を用意せずともデジタルビデオカメラで気軽に撮影を行うことが可能になったため、個人が自ら撮影した映像を Web 上に発信する機会も増えていくであろう。これに伴い、

今後はクオリティの高い映像が求められるようになるが、撮影・編集に関する専門知識と経験に乏しい一般ユーザが撮影した映像は見る側に退屈感を与えてしまう。このため、一般ユーザを対象としたコンテンツ制作では、

- マルチメディア化：テキストから音声・動画へ
- 効率性：容易に制作・加工可能
- クオリティ：質を維持

という 3 つの要素を支援する自動撮影技術が必要となる。

この自動撮影において、ある程度ストーリー性のあるシーンの映像化を考えると、そのようなシーンにはほとんどの場合シナリオが存在する。通常シナリオに

\* 慶應義塾大学大学院 理工学研究科 開放環境科学専攻  
School for Open and Environmental Systems,  
Graduate School of Science and Technology,  
Keio University

† 慶應義塾大学 理工学部  
Faculty of Science and Technology, Keio University

はシーンごとに最も注目すべきメインパートや、脇役的な存在のマイナーパート、流れの中におけるパート間の関係や物理的な位置関係など、撮影に役立つ様々な情報が書き込まれているので、この情報に基づいてどの被写体を、どのビデオカメラで、どのように撮影するのかという撮影計画をあらかじめ立てておく撮影手法が考えられる。

しかし、一般ユーザによる映像制作においては、撮影環境や編集の仕方がその都度異なることが多い。このため、一般ユーザが適切な撮影計画を立てることは難しい。そこで本稿では、撮影環境の変動と複数ユーザからの様々な編集要求に対応可能なカメラワークを自動的に計画・生成する手法を提案する。撮影対象はシナリオとして楽譜が存在するオーケストラ演奏とする。

## 2 提案

### 2.1 本研究の位置付け

自動撮影による映像制作のプロセスは、図1に示すように、カメラワークの計画フェーズ、その計画に従った撮影の実行フェーズ、得られるストリームを編集する編集フェーズの3つのフェーズに分けることができる。

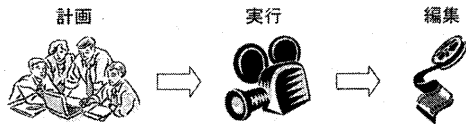


図1: 映像制作のプロセス

一方、撮影対象はその進行の仕方によって表1に示すイベント型とストーリー型の2つに分類することができる。このうち、ストーリー型にはほとんどの場合シナリオが存在し、プロのカメラマンによる撮影においても計画フェーズでシナリオが重要な役割を果たしている[1]。例えば番組制作においては、シナリオから絵コンテを書き出し、環境の整ったスタジオにおいてそのショットを撮影するのに適した位置にカメラを配置する、といった手順で撮影計画を立てることができる。

これに対し一般ユーザによる映像コンテンツ制作では、シーン内に複数の被写体が存在する場合には、注目する被写体や編集の仕方はユーザごとに様々であると考えられる。また、常に整った環境が用意できるとは限らないため、ビデオカメラの数や配置などの撮影

表1: 撮影対象の分類

イベント型	特徴	その場の状況に応じて進行
	例	会議・スポーツ中継など
ストーリー型	特徴	決まった流れに沿って進行
	例	演劇・コンサートなど

環境はその都度異なることが多い。つまり、一般ユーザによる映像コンテンツ制作においては、以下の2点を考慮した撮影計画が重要であると考えられる。

- シーンの状態から、編集時に発生する様々なショットの要求を予測する。
- 限られた撮影環境を考慮し、かつ撮影資源を有効に利用する。

そこで本稿ではシナリオが存在するストーリー型の撮影対象に着目し、シナリオ情報に基づいたカメラワークの計画手法を提案する。本手法ではシナリオ情報からシーンの状況を把握することで視線の集まる被写体を抽出し、これら複数の被写体（以下被写体候補）のうち撮影要求の高いものから順に撮影環境を考慮したカメラの割り当てを行う。これにより、編集時に発生する様々なショットの要求をなるべく満足し、また異なる撮影環境でもカメラの台数なりの動きにより撮影資源の有効利用を目指す。以降ではオーケストラを例として本手法の具体的な実現方法について述べる。

### 2.2 撮影対象

本稿において撮影対象とするオーケストラには基本となる編成（標準編成）があり、木管楽器、金管楽器、打楽器、高弦楽器、低弦楽器という5つのグループから成り立つのが一般的である[2]。通常のオーケストラの演奏会が行われるコンサートホールでは、図2のような配置が現在最もよく用いられている。舞台上には複数の楽器が存在するため、オーケストラ楽曲では複数の楽器が同時に演奏していることが多い。また、楽曲の流れとともに演奏する楽器の種類・組み合わせも様々に移り変わる。従って、同じ撮影対象であっても見る人によって視点が異なることが予想され、カメラワークを計画する撮影対象として適していると考えた。

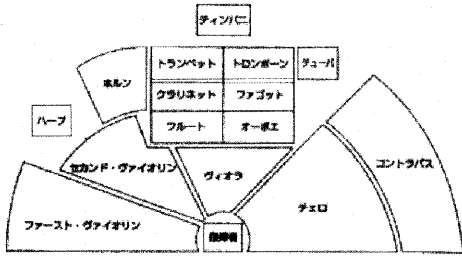


図 2: オーケストラの配置

## 2.3 シナリオ情報

本稿で扱うシナリオとは「“どこで、誰が（または何が）、何を”というイベントがストーリーの時間軸に沿って書かれているもの」と定義する。

オーケストラにはスコアと呼ばれる各楽器の楽譜をまとめた総譜が1つの楽曲につき1つ存在する。このスコアでは、各楽器がどのような音を演奏するかというイベントが音符によって表現されており、その曲の進行に従って時間軸は左から右へ流れている。以上の点からスコアはシナリオとしての性質を持つといえる。そこで、本稿ではスコアから次の2つの特徴を抽出し、シナリオとして扱うこととする。

- 楽器の編成および配置

オーケストラにおける楽器の編成およびその配置は楽曲によって決まるものであり、これらはスコアから抽出可能な情報である。

- 曲のまとまりと楽器の役割

楽曲にはフレーズと呼ばれる特徴のある型を持った一連の旋律群のひとまとまりが存在し、このフレーズがいくつか集まって楽式（ソナタ形式など）の流れに従って楽曲を構成している。このフレーズ中での各楽器は、ソロあるいはハーモニーでの中心的な位置を占める「メロディー」と、それを支え装飾する「伴奏」というように、役割によって大きく2つに分類できる。

## 2.4 映像の分析

以上のシナリオ情報が実際の映像にどのように利用されているかを調べるため、オーケストラ映像の分析を行った。手順は次の通りである。

1. スコアから楽曲をフレーズに分ける
2. フレーズごとに各楽器の役割を分類する
3. 映像からフレーズごとに撮影されている楽器を抽出、分析する

この結果、オーケストラ映像にはシナリオ情報に関連して次の2点の傾向があることが分かった。

- ショットの切り替えとショットサイズ

ショットの切り替えは固定の小節数単位ではなく、曲のまとまりであるフレーズ単位で起こる傾向があった。そこで、本手法ではフレーズ単位でカメラワークを計画をすることとする。

また、切り替え前後でのショットサイズは常に一定ではなく、段階的にズームインするショットへの切り替えや、ズームインショットとズームアウトショットの交互の切り替えが多くなることが分かった。一般に、イベント型のシーンにおけるショット切り替え時の急激なサイズの変化は見る側に違和感を与えるものだと考えられているが[5]、ストーリー型のシーンではある程度先を予測することができるため、ショットサイズの変化も演出の1つとして利用できるといえる。そこで、これをカメラワークの計画における優先度付けに利用する。

- 演奏楽器数とショット

音を出している楽器を演奏楽器とすると、舞台上に存在する全楽器中の演奏楽器の割合と、そのときのショットとの間には表2のような関係があることが分かった。

表 2: 演奏楽器数とショットの傾向

演奏楽器数/全楽器数	ショット
2割未満	メロディー楽器 指揮者
2割以上 6割未満	メロディー楽器 演奏楽器の組み合わせ
6割以上	メロディー楽器 演奏楽器の組み合わせ 指揮者または全体

同一のシーン内に複数の被写体が存在する場合には、カメラの台数に応じて撮影要求の高いものを選択する必要がある。そこで、これらの分析から得られた傾向をカメラワークの計画に利用する。

## 2.5 カメラワークの計画

本稿におけるカメラワークとは「どの被写体をどのカメラでどの大きさで撮るのか決定すること」と定義する。カメラワークの計画手順は、シナリオ情報から各シーンの状況を把握して被写体候補を抽出し、そのショットの大きさおよびカメラとの位置関係から決定する優先度に応じ、撮影要求の高い被写体を各カメラに割り当てる、といったものである(図3)。

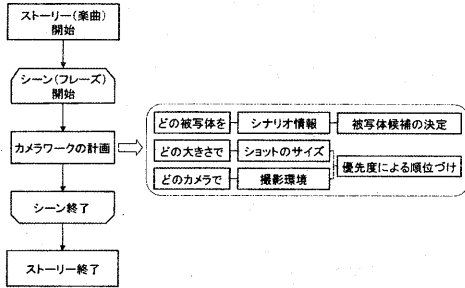


図3: カメラワークの計画

### 2.5.1 被写体の選出

まず、シーンの状況を把握するため、シナリオからフレーズ内の演奏楽器に関する情報を全て抽出する。次に先ほどの映像の分析結果を利用し、全楽器に対する演奏楽器の割合に応じて被写体候補を決定する。

#### ● 演奏楽器の割合: 2割未満の場合

メロディー楽器および指揮者を被写体候補とする。メロディー楽器同士が物理的に隣に位置する場合は同じ画面内にアップショットで捉えることができるが、この組み合わせも被写体候補に加える。

#### ● 演奏楽器の割合: 2割以上6割未満の場合

全てのメロディー楽器および演奏楽器の全ての組み合わせを被写体候補とする。演奏楽器の組み合わせを決定する際はオーケストラの楽器の階層関係から、

- 子ノードの半数以上を演奏楽器に含む場合
- 子ノードの1つでも演奏楽器に含む場合

という2つのポリシーのいずれかに従い、条件を満たすときには親ノードの楽器のグループを演奏楽器の組み合わせとして被写体候補に加える。

#### ● 演奏楽器の割合: 6割以上の場合

全てのメロディー楽器および演奏楽器の全ての組み合わせを被写体候補とする。さらに、オーケストラ全体または指揮者を被写体候補に加える。この選択には、オーケストラの標準編成である基本5グループにおける演奏楽器の分布を閾値として利用する。

##### - 演奏楽器が4グループ以上の場合

演奏楽器の組み合わせを作る段階で階層関係をたどるとオーケストラ全体までたどり着く。そこでこの場合は全体を選択する。この場合の被写体候補の例を図4に示す。

##### - 演奏楽器が3グループ以下の場合

階層関係は途中までしかたどられないため指揮者を選択する。

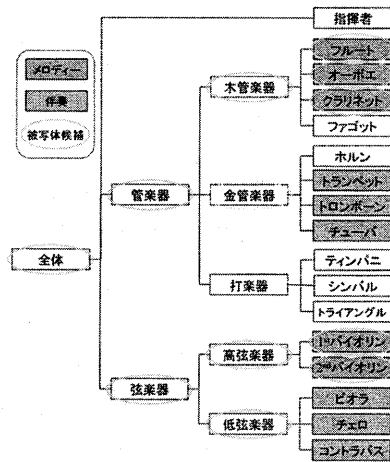


図4: 被写体候補: 6割以上の場合

以上のようにシナリオ情報からシーンの状況を把握し、被写体候補を決定する。この被写体候補の数が撮影環境において利用可能なカメラの台数より多い場合、撮影要求に応じて取捨選択する必要がある。また、カメラの台数に余裕がある場合でも、どのカメラから撮影するのがよりよいのかを決定する必要がある。そこで、これらの被写体候補に優先度付けを行う。

### 2.5.2 ショットサイズの設定

シナリオ情報から抽出した被写体候補に対し、画面内での被写体の大きさ（ショットサイズ）による優先度付けを行う。そこで図5に示すように、同一フレーズ内のカメラ間における被写体の大きさの差をカメラ間距離、時間軸で前後のフレーズにおける大きさの差をフレーズ間距離と定義する。

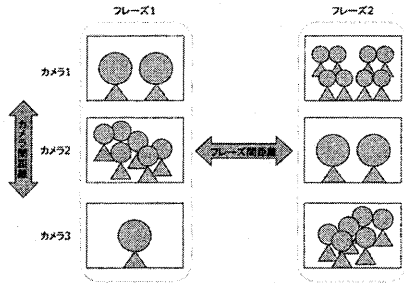


図5: カメラ間距離とフレーズ間距離

映像の分析結果から、ショットの切り替え前後では画面内における被写体の大きさは常に一定ではなく、これを変化させることで演出効果が得られると考えられる。そこで、カメラ間距離では被写体の距離の合計がなるべく大きくなるように優先度を付ける。また、フレーズ間距離では前回大きい被写体が撮影されていた場合は小さく、あるいはその逆になるように優先度を付け、全体としては平均的な距離を保つようにする。これらの距離の比較にはオーケストラの階層関係のレベルを利用し、大きさだけでなくその被写体のグループから楽器の種類も考慮した優先度を付ける。このように、様々な大きさおよび種類を考慮して被写体候補の優先度を変化させることにより、ユーザのショット選択の幅を広げることが可能になると考えられる。

## 2.6 ビデオカメラの割り振り

最後に撮影環境を考慮して被写体候補とカメラとの位置関係による優先度付けを行う。本稿では撮影環境はその都度異なると想定しているため、各撮影環境ごとにビデオカメラの識別子および3次元座標とパン・チルト角、ズーム値を記述したカメラマップを予め用意しておくものとする。

カメラマップの情報のうち、まずビデオカメラの3次元座標から撮影可能な被写体候補を選択する。撮影

環境として想定するオーケストラホールは大きく舞台と客席という2つに分けることができる。このうち、客席側に存在するビデオカメラは全ての被写体を撮影可能とする。また、舞台上に存在するビデオカメラについては被写体の向きを考慮する必要があるが、オーケストラでは各楽器が指揮者を中心として放射状に配置されていることを利用し、被写体に対して正面・斜め・横の順に「よりよく映せる」とする。図6のように舞台上を4つのエリアに分割し、各ビデオカメラは自身と異なるエリアに存在する被写体候補に対してよりよく映せる順に優先度を付ける。

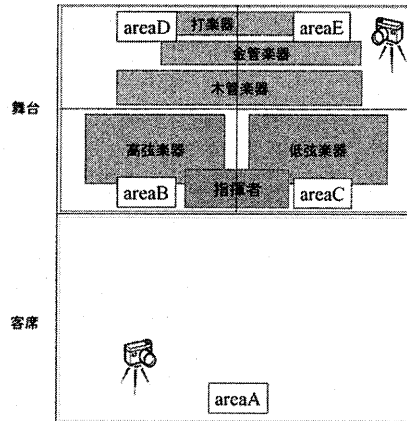


図6: 撮影領域の分割

また、カメラマップ中の各ビデオカメラのパン・チルト角、ズーム値から直前の撮影状態を把握し、単調な映像を避けるため、極力同じ被写体を特に同じビデオカメラで連続して撮影することがないように被写体の優先度を下げることとする。

## 2.7 被写体の決定

以上のようにして決定した優先度に従い、全体およびビデオカメラごとに被写体候補をソートする。そして全体で最も優先度の高い被写体を最もよく撮影可能なビデオカメラに割り当てる。優先度が同位の場合は異なるビデオカメラで撮影可能な場合はそれぞれに別のビデオカメラを割り当てる。また、同一のビデオカメラでしか撮影できない場合は過去の出現回数からどちらかを選択する。1つ被写体が決定するごとに優先度を付け直し、全てのカメラに被写体を割り当てる。

## 3 関連研究

### 3.1 TVML

シナリオを利用した映像コンテンツ制作に関する研究として、林らが提案した TVML がある [3]。TVML はセリフやタイトルのテキスト文といった情報を直接台本に記述するための番組記述言語である。

TVML は一般ユーザによるコンテンツ制作支援に注目した研究であるという点では本研究と目的が一致している。しかし、TVML はシナリオ記述言語の提案であり、またカメラのスイッチングの自動化は検討されているが、カメラワークの自動化についてはまだ実現されていない。さらに、これが実現したとしても仮想空間という視点に制約のない世界に閉じている。これに対し、本研究では実世界での撮影に対応できるカメラワークの自動的な計画を目指しており、その提案の範囲と実現方法が異なる。

### 3.2 イベント駆動型カメラワーク

田中らは動的に変動するシーンの状況を分かりやすく見飽きない効果的な映像として提示するため、シナリオなどの事前知識に基づいてカメラワークを計画、実行する仕組みを提案した [4]。計画では人物の位置・行動などが記述されたシナリオと、各ショットでの画面構成や撮影時のカメラの制御の種類、映像効果などが記述されたストーリーボード（絵コンテ）を利用し、具体的なカメラ配置やカメラ制御手順、画面構成を決定する。田中らは主に撮影の計画と実行段階におけるズレを調整する手法に主眼を置いている。カメラワークの自動的な計画についてはその手法の提案にとどまっており、実現されていない。これに対し本研究では、一般ユーザによる映像コンテンツ制作を前提としており、異なる撮影環境でも編集するユーザごとの様々な要求をなるべく満たすことのできるカメラワークの計画を自動的に行うことを目的としている点が異なる。

## 4 おわりに

本稿では提案の範囲を撮影の計画段階に限定したため、計画されたカメラワークの確認には TVML などを用いることが考えられる。将来的には実世界での利用を目的としているため、撮影の実行フェーズではシナリオの進行とシーンの進行との同期を取り、ズレが生じた際には調整を行う必要がある。この調整にはいく

つかの手法が考えられるが、カメラワークの制御を行うタイミングのみを把握できる程度の簡易な画像・音声認識を行えば十分である。オーケストラでは指揮者のタクトにセンサを付け、その動きを認識するといった既存の手法が利用でき、複雑な処理は必要としないと考えている。

また、本手法はシナリオが存在するシーンを撮影対象としている。しかし、スポーツ中継などにおいても、いわゆるフォーメーションなどの状況に応じて進行形式が決まる場合がある。このような「半ストーリー型」の撮影対象へも適応させることにより、本手法の利用例が広がると考えている。

## 謝辞

本研究の一部は 21 世紀 COE プログラム研究拠点形成費補助金のもとに行われた。

## 参考文献

- [1] 住吉 英樹, 有安 香子, 望月 祐一, 佐野 雅則, 井上 誠喜: データベースを中心とした番組制作支援システム. 情報処理学会研究報告 2000-DBS-120, pp. 83-90, 2000.
- [2] 鈴木 織衛: オーケストラを読む本. ヤマハミュージックメディア, 2000.
- [3] 林 正樹: テキスト台本からの自動番組制作～TVML の提案. 情報処理学会研究報告 DBS-120-13, pp. 91-98, 2000.
- [4] 田中 彰, 東海 彰吾, 松山 隆司: イベント駆動型カメラワークによる動的シーンの効果的映像化. 情報処理学会研究報告 CVIM-121-10, pp. 73-80, 2000.
- [5] 宮崎 英明, 亀田 能成, 美濃 導彦: 複数のカメラを用いた複数ユーザに対する講義の実時間映像化手法. 電子情報通信学会論文誌 J82-D-II, pp. 1598-1605, 1999.