講義映像要約のための撮影ルールの構築

Construction of the shooting rule for lecture video skimming

西口敏司 † 亀田能成 ‡ 美濃導彦 ‡ 池田克夫 †
NISHIGUCHI Satoshi KAMEDA Yoshinari MINOH Michihiko IKEDA Katsuo

† 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University

‡ 京都大学総合情報メディアセンター
Center for Information and Multimedia Studies, Kyoto University

1 はじめに

我々は講義映像を要約する手法の開発に取り組んでいる．我々が考える映像要約とは，要約対象である複数の映像から空間的，時間的，内容的に圧縮した映像を生成することである．1本の映像を対象とした映像要約の研究 [1] は，ニュース映像などの既に人間が見やすいように編集された映像を要約対象としており，編集済み映像の再編集のための手法の提案であった．これに対して本研究では，撮影の段階から映像要約を考慮して映像を撮影し，またそのときのカメラワークを記録することで，講義映像に適した映像要約を実現することを目指す．

本稿では，パン，チルト，ズームが可能な複数のカメラで講義を撮影して得られる複数の映像を要約対象とし，講義映像を要約するのに適した複数の映像を得るための撮影ルールについて考察する．

2 講義映像要約

本研究では，複数のカメラがパン，チルト，ズームを繰り返しながら講義の様子を撮影する．3台のカメラによって得られる映像の模式図を図 1 に示す．図の灰色の部分はカメラがパン，チルトなどの動作をしている状態での映像を表す．ニュースなどの映像はいくつかのシーンやショットから構成されている [2]．一方，本研究で使用するカメラで撮影される講義映像は，ニュース映像とは異なり，予めショットやシーンが定義されているわけではない．しかし，例えばパン，チルト動作をカットとみなせば，図 1 で示す各映像の白い部分はある意味でショットとみなすことができる．

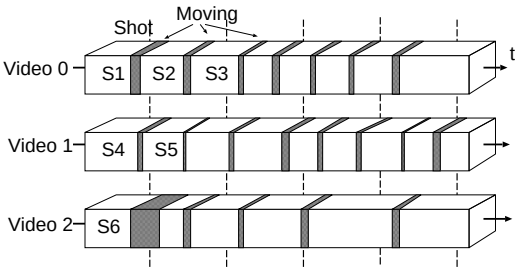

図 1: 複数カメラによる映像例

このように考えると，本研究で考えているような複数の映像を対象に要約する場合では，映像要約の手法として次のような基準を与えることができる．

空間的要約 例：S1, S4, S6 から一つを選択
時間的要約 例：Video 0 内で S1 と S3 を選択
内容的要約 例：S6 を関連する別の映像内の D6 で置換

これらの基準のうち，複数のカメラでどのような講義映像を撮影しておくかという問題と直接関係があるのは空間的要約である．すなわち，講義映像の要約に対しては，視聴者の目的に依存して様々な要求がありうるため，多くのショットが存在することが望ましい．以上のような考察から，本稿では講義映像に対してショットに相当する映像を撮影するカメラワークを定義する．そして，このような撮影を実現するための撮影ルールを提案する．

3 講義映像要約のための撮影ルール

本研究における撮影ルールとは，講義中の講義の状況に基づいて各カメラに空間的な多様性を持つ映像を撮影するようにカメラワークを割当てる方法のことである．また，あるカメラのカメラワークとは，そのカメラが撮影すべき被写体とカメラ動作と撮影サイズの組合せである．あるカメラワークのもとで撮影された映像は，2 章におけるショットに相当する．そして，空間的な多様性を持つ映像とは，複数台のカメラで互いに異なる被写体をいくつかのカメラ動作のもとで様々な撮影サイズで撮影した映像のことである．つまり，多くのショットを持つ映像に相当する．本研究では講義映像に空間的な多様性を持たせるため，パンニング，チルティング，ズーミングの遠隔操作が可能な複数台のカメラを講義室内に固定して設置する．講義室とカメラの設置例を図 2 に示す．

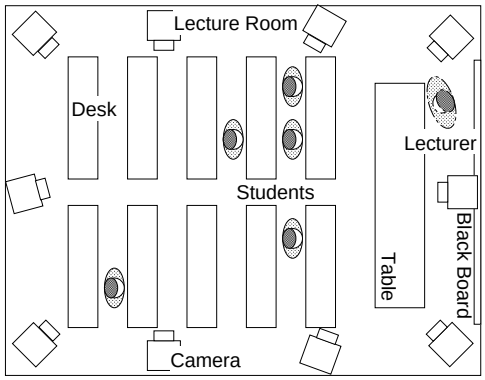

図 2: 講義室とカメラの設置例

本稿では，講義中に講義状況の把握，複数カメラへのカメラワークの割当という手順を繰り返すことによって，講義映像の要約に適した空間的な多様性を持つ映像を撮影することを考える．

3.1 講義状況の把握

各カメラにカメラワークを割当てるには，講義中の状況を把握する必要がある．本稿においてカメラワークを割当てる際に利用する講義状況は，ある時刻における講師の位置，講師の顔の向き，講師の指示行為，受講者の位置，受講者の動きの組で構成されるものとし，これらの情報はコンピュータビジョンの手法を利用して取得する．

3.2 複数カメラへのカメラワークの割当

実験環境が講義であることと，講義が講師から受講者への知識の伝達であることを考慮して，講義室内で撮影すべき被写体は人物 (講師，受講者) および教材 (スライド，OHP，黒板など) であるとする．また，ある被写体に対するカメラの動作としては，固定撮影と追跡撮影が考えられる．さらに，本研究で考えている講義の形態では，講師や受講者の腰から下は教壇や机によって隠れる場合が多いため，腰より上の身振り手振りを撮影できればよい．

そこで，講義室内に設置した各カメラのカメラワークは表 1 に示す構成要素からなるものとする．構成要素の内容を組み合わせることにより，様々なカメラワークをカメラに割当てることが可能となる．

表 1: カメラワークの構成要素と内容

構成要素	内　容
被写体	講師，受講者，スライドなど
カメラ動作	固定，追跡
撮影サイズ	ロングショット，ウエストショット バストショット，クローズアップ

講義状況に基づき，撮影ルールが各カメラにカメラワークを割当てが，各カメラが割当てられるカメラワークが似たものであれば，結果として似た映像を撮影することになり，空間的に多様な映像を撮影するという方針に反する．そこで，複数のカメラに割当てられたカメラワークによって多様な映像が得られるように，各カメラに割当てられるカメラワークの構成要素の内容の決定基準について考察する．

被写体の決定基準
: ある時刻にすべてのカメラが講師を撮影していたり，逆にある受講者のみを撮影していたりすれば，その時刻で撮影していない人物の動作や表情を写した映像を要約映像に使うことができない．そこで，同時刻にできるだけ多くの人物被写体を撮影するために，講師とある一人の受講者を少なくともそれぞれ 1 台以上のカメラの被写体とする．また，講義が進行していく際の状況の変化を映像から把握するには，講義室内の全体を写した映像を見ることが効果的である．そこで，少なくとも 1 台のカメラは講師や受講者の両者を同時に被写体としてみなすことにする．ただし空間的に多様な映像を撮影するという方針から，このカメラは特定のカメラに固定しないこととする．

カメラ動作の決定基準
: 受講者は席に着席しているものと考えるので，受講者を被写体とするときにはカメラ動作は固定とする．一方，講師を追跡撮影すべきかどうかを撮影の段階で客観的な基準によって判断することはできないため，本来であれば要約の段階における様々な要求に応えるために，両方のカメラ動作で撮影すべきである．しかし，カメラの台数などの物理的制約によって，本稿では講師を追跡撮影するかしないかは確率的に決定することにする．

撮影サイズの決定基準
: 撮影サイズについてもどのサイズで撮影すべきかを撮影の段階で客観的な基準によって判断することはできない．本来であれば，撮影可能なすべての撮影サイズで撮影すべきである．しかし，この場合も物理的制約から，本稿では次のような方針で撮影サイズを決定する．まず，時刻的に 1 つ前に割当てられたカメラワークで撮影していた被写体を連続して撮影する場合は，1 つ前のカメラワークにおける撮影サイズと，ひとまわり小さいかまたは大きい撮影サイズの中から確率的にどれかを決定する．次に，被写体が変わる場合も，新しく被写体となったものをどの撮影サイズで撮影するかは，確率的に決定することにする．

以上の決定基準によって空間的に多様な映像を撮影することができ，要約の段階で多くのショットを利用することが可能となる．

一方，これまで述べてきた撮影ルールは各カメラに対して互いに異なるカメラワークを割当てようとするものであった．しかし要約の段階で，互いに協調して動作したカメラからの映像を利用したいという要求などもある．例えば映画の分野では，経験則に基づいて人物を撮影するカメラワークの技法が提案されている．特に対話する二人の人物を結ぶ線や，一人の人物の顔の向きに関心を示す線 を仮定し，この線と平行な底辺を持つ三角形を構成するようなカメラの配置が最も心理的に安定した映像が得られるといわれている．これをカメラ配置の三角形原則 [3] と呼ぶ．そこで本研究では講師と受講者の間に関心を示す線が存在すると仮定し，上で述べた原則に基づいてカメラを動作させるという撮影ルールも追加する．図 3 に講師の顔の向きに基づく関心を示す線と，対応する安定した映像を得るためのカメラワークが割当てられたカメラの位置の例を示す．

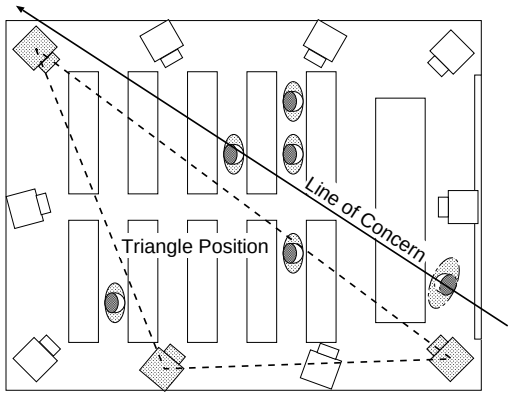

図 3: カメラ配置の三角形原則

4 おわりに

本稿では，講義映像を要約することを考慮して，空間的に多様な映像を取得する撮影ルールを提案した．今後は，本稿で示した撮影ルールを実際の講義室に適用して検証するために，講義状況の把握，関心を示す線の検出などの手法をシステム上に実装するとともに，得られる複数の映像や講師の音声から要約映像を生成する枠組について検討する予定である．

参考文献

[1] Smith, A. and Kanade, M.: Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques, *IEEE Proc. of Computer Vison and Pattern Recognition*, pp. 775–781 (1997).

[2] 西尾章治郎, 田中克己, 上原邦昭, 有木康雄, 加藤俊一, 河野浩之: 岩波講座マルチメディア情報学 8 情報の構造化と検索, 岩波書店 (2000).

[3] ダニエル・アリホン (著), 岩本憲児, 出口丈人 (訳): 映画の文法, 紀伊国屋書店 (1980).