論　文

状況理解と映像評価に基づく講義の知的自動撮影

大西　正輝[†]　　村上　昌史[†]　　福永　邦雄[†]

Computer-Controlled Camera Work at Lecture Scene Based on Situation Understanding and Evaluation of Video Images

Masaki ONISHI[†], Masashi MURAKAMI[†], and Kunio FUKUNAGA[†]

あらまし　本論文では，人手を介さずに板書主体の講義を自動的に撮影する手法として，講義の状況を理解することで撮影領域を決定し，得られた複数の映像を評価することで講義映像として最も適している映像にスイッチングを行う知的自動撮影手法を提案する．まず，講義の状況を理解するために，固定カメラによって撮影した講義映像から講義者と黒板の板書に関する情報を抽出し，それらの情報を用いて講義者の行動推定を行う．次に，講義者の行動に基づいて各カメラにおいて撮影領域を決定し，複数のカメラ位置から映像を取得する．最後に，得られた複数の映像をそれぞれ評価することにより，現在の講義状況を最も効果的に表している映像を選択する．実際に講義の自動撮影を行い，本手法の有効性を確認した．

キーワード　知的映像ハンドリング，講義映像，状況の理解，映像の評価，自動撮影

1. まえがき

ディジタル多チャネル時代を迎え，映像生成の省力化の要求が高まってきており，カメラマンが撮影するようなクオリティーの高い映像を自動的に生成する研究が行われている．これらの技術は，スタジオ番組やスポーツ番組などの撮影だけではなく，災害地のような危険な場所やカメラマンが撮影できないような位置からの映像生成が可能になるなど応用範囲は広い．ここでは，プロのカメラマンのような撮影知識を計算機に組み込み，自動的に最適な映像を取得することを，知的自動撮影と呼ぶことにする．

いかなる拘束条件ももたない一般的なシーンの撮影を自動化するためには，あらゆる条件下での撮影規則を用意されていることが前提となり，一般的には難しい．そこで，撮影対象を限定した知的撮影手法が数多く提案されている．能 [1] や料理番組 [2],[3] のようにあらかじめシナリオが決まっている（あるいはシナリオに相当するものを人が抽出する）シーンを想定した自動撮影手法では，シナリオや人の合図に同期したカメラワークを実現する研究が報告されており，スポーツ [4]〜[6] や講義 [7]〜[13]，プレゼンテーション [14] のようにシナリオが用意されていないシーンを想定した自動撮影手法では，シナリオに相当するイベントをシーン中から抽出して，この結果をもとにカメラワークを決定する研究が報告されている．中でも講義の自動撮影に関する研究は遠隔講義への利用などの需要が多いことから，盛んに研究が続けられている．

講義やスポーツなどを撮影する場合には，カメラがシーン中を自由に移動できる空間が限定されているため，複数台の撮影カメラをあらかじめ固定した位置に設置することが多い．このような条件下での，知的自動撮影に関する研究では，各カメラにおける撮影領域の決定方法と複数台のカメラを切り換えるスイッチング方法が問題となる．しかし，これまでに行われてきた知的自動撮影に関する研究では，どちらか一方に焦点を当てた研究が多く，両方を同時に扱った研究は少ない．通常，シナリオのないシーンを撮影対象とした番組制作においては，複数のカメラマンがシーンの状況を理解して各々の位置から最も適したアングルで撮影を行い，スイッチャが得られた複数の映像の中からそのシーンを表現するのに最も適した映像を選択することが多い．本論文では，撮影対象として黒板を利用する講義シーンを取り上げ，シーン中の状況理解によって決定した領域を撮影した後，得られた複数の映

[†] 大阪府立大学大学院工学研究科，堺市
Graduate School of Engineering, Osaka Prefecture University 1-1, Gakuen-cho, Sakai-shi, 599-8531 Japan

像の中から最も講義に適した映像を選択する知的自動撮影手法を提案する．また，実際に知的自動撮影システムを構成し，実験により本手法の有効性を確認する．

2. 講義映像におけるイベント抽出

スポーツや講義などを撮影する場合には，シーン中に起きている状況から，撮影領域を決定するトリガを抽出する必要がある．本論文では撮影領域を決定するためのトリガをイベントと呼ぶ．講義の撮影を想定したカメラワークを設定する上では，講義者の行動が重要なトリガとなる場合が多い．ここでは，固定カメラで撮影した講義映像から講義者や黒板の板書についての情報を抽出して，イベントとして設定した講義者の行動を認識する．

2.1 講義者と板書の分離

講義映像に現れる対象物として，講義者と板書文字の二つが考えられる．講義者や板書文字に関する情報を抽出するためには，講義映像からこれら二つの対象物を分離して抽出する必要がある．本研究では講義者は動いており板書文字は静止していることに注目し，時空間画像の断面図から抽出したエッジの方向を調べることで講義者と板書文字を分離して抽出する [15]．本論文でも文献 [15] と同様に，時空間画像の断面図から抽出した動物体によるエッジを動エッジ，静止する物体によるエッジを静止エッジと呼ぶ．

図 1 (a) の入力画像から，動エッジ（黒色）と静止エッジ（灰色）を抽出した例を図 1 (b) に示す．動エッジを生成する講義者と静止エッジを生成する板書文字が分離できているが確認できる．

2.2 講義者情報の抽出

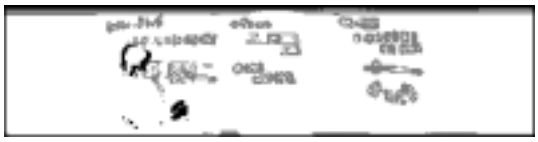
講義映像において動エッジを生成する対象は講義者であると考えられる．そこで，入力画像から得られる動エッジを用いて講義者に関する情報を抽出する．イベントとして設定した講義者の行動推定を行うための情報として，講義者の位置や顔の向き，更には手の位置を抽出する．

2.2.1 頭部の位置推定

動エッジを用いて講義者の頭部の位置を推定する．まず，得られたすべての動エッジ点の重心から講義者頭部の概略的な位置を調べる．次に，頭部は楕円形であると仮定して，動エッジの重心付近で楕円に近い形状で分布する動エッジを探索し，得られた楕円領域を講義者の頭部とする．ただし，ここで用いる楕円の短径と長径の比は 1 : 1.2 としており，頭部が傾いていないことを仮定して楕円の短軸方向は，水平方向と一致させて探索した．また，個人差やある程度のカメラの光軸方向への動きを吸収できるように，楕円の大きさは 3 パターン用意した．一方，講義者が動かなかった場合には動エッジがほとんど現れないため，動エッジ数がしきい値以下になった場合には講義者は動いていないものと考え，前フレームでの講義者の頭部位置を現在の推定位置とする．

以上の処理によって講義者の頭部位置を抽出した例を図 1 (c) に示す．図中の楕円が抽出した頭部を表しており，楕円の中心を講義者の位置とする．

(a) 入力画像
(a)Input image.



(b) 動エッジ（黒）と静止エッジ（灰）画像
(b) Dynamic (black) and static (gray) edge image.

(c) 講義者情報の抽出
(c) Extraction of the lecturer's feature.

(d) 肌色の抽出
(d) Extraction of the skin color.

(e) 板書情報の抽出
(e) Extraction of written region on the blackboard.

図 1 状況理解結果
Fig. 1 Results of situation understanding.

2.2.2 顔の向き推定

入力画像の各画素の色情報を調べることで人物の肌を抽出し，頭部抽出結果を用いて講義者の顔の向きを推定する．

人物の肌色を抽出するための色情報として比較的明るさの変化の影響を受けにくい CIE1976UCS 表色系を用いる．あらかじめ人手によって人物の肌を含む複数の画像から顔や手などの肌領域の画素を N 個抽出しておき，それらの画素の画素値を CIE1976UCS 色度図上に投影した $\boldsymbol{C}_i = (u_i, v_i)$ $(i = 1, 2, \ldots, N)$ に対し，人物の肌色の平均値 $\boldsymbol{\mu}_S$ 及び共分散行列 $\boldsymbol{\Sigma}_S$ を次式で求める．

$$\boldsymbol{\mu}_S = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{C}_i \tag{1}$$

$$\boldsymbol{\Sigma}_S = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{C}_i - \boldsymbol{\mu}_S)^T(\boldsymbol{C}_i - \boldsymbol{\mu}_S) \tag{2}$$

上式で求めた平均値 $\boldsymbol{\mu}_S$ と共分散行列 $\boldsymbol{\Sigma}_S$ から，入力画像中の各画素 $\boldsymbol{x}$ に対する画素値 $\boldsymbol{C_x}$ が肌色にどれくらい近いかを表す値 $P_S(\boldsymbol{C_x})$ を次式で求める．

$$P_S(\boldsymbol{C_x}) = \frac{1}{2\pi\,|\boldsymbol{\Sigma}_S|^{\frac{1}{2}}}\cdot \exp\left\{-\frac{1}{2}(\boldsymbol{C_x} - \boldsymbol{\mu}_S)\boldsymbol{\Sigma}_S{}^{-1}(\boldsymbol{C_x} - \boldsymbol{\mu}_S)^T\right\} \tag{3}$$

すべての画素において式 (3) を用いて求めた値をしきい値処理により 2 値化して肌色画素を抽出する．抽出した肌色画素を図 1 (d) に示す．黒い画素が肌色領域を示している．

次に，講義者の頭部（**2.2.1** で求めた楕円）領域に含まれている肌色画素の分布から顔の向きを推定する．ここでは概略的な向きの推定を目的としているため，頭部領域として抽出した楕円を左側と右側の二つの領域に分けてそれぞれの領域に含まれる肌色画素の数を数え，しきい値処理によってそれらの大小を調べることで，顔の向きを「正面」「右」「左」「後ろ」の四つに分類する．例えば，肌色画素が楕円の左半分に多く含まれているが，右半分にはあまり含まれていない場合には，講義者の顔の向きは「左」であると推定することにする．

2.2.3 手の位置推定

次に，**2.2.1** で得られた頭部領域と **2.2.2** で得られた肌色抽出結果を用いて，講義者の手の位置を推定する．

講義映像に現れる肌色領域は，講義者の顔領域と手領域であると考えられるため，抽出した肌色画素のうち，頭部領域に含まれない肌色画素が手領域を形成していると考えられる．そこで，頭部領域を除くすべての肌色画素を座標位置によってクラスタリングした後に，各クラスタの重心を求めれば手の位置を推定することができる．ここでは K–平均アルゴリズムを用いて肌色画素のクラスタリングを行う．求めるクラスタ数は両手を想定した 2 としてクラスタリングを行い，最終的に決定した二つのクラスタ中心の座標間距離がしきい値未満であった場合には手領域が一つしか映っていないものとして，二つのクラスタの平均座標を手の位置とする．また二つのクラスタ中心間の距離がしきい値以上であった場合には，手領域が二つ映っているものと考え，それぞれのクラスタ中心の座標を手の位置とする．

手の抽出結果を図 1 (c) に＋印で示す．入力画像中に映し出されている片手のみが抽出できている．

2.3 板書情報の抽出

2.3.1 板書文字量の推定

設定したイベントを抽出するためには，講義者に関する情報だけでは十分とはいえない．例えば，“板書している” という行動と “黒板を消している” という行動は，どちらも黒板の方を向いて手を動作させる行動であるため，講義者に関する情報のみからこれらの行動を区別するのは難しい．そこで，“板書している” ときには板書文字量が増加し，“黒板を消している” ときには板書文字量が減少することに注目して，板書文字量の増減から，これらの行動を区別する．講義シーンを考えた場合には，板書文字は黒板中の静止エッジで構成されていると考えられるため静止エッジのドット数を調べ，その数を板書文字量とする．

2.3.2 黒板消しの位置推定

“黒板を消している” という行動を認識するためには，板書文字量が減少したことを確認すればよいが，板書文字が講義者の陰に隠れてしまい，板書文字量が減少することも考えられるため，板書文字量の変化のみから “黒板を消している” ことを認識すると誤認識に結び付くことが多い．そこで，“黒板を消している” 場合には講義者が黒板消しを使用することから，黒板消しの色情報を用いて黒板下部のレール上における黒板消しの位置を推定し，黒板消しがレール上に存在するか否かを調べる．

まず，**2.2.2**で肌色画素を抽出した方法と同様に，

表 1　行動推定規則
Table 1　Rule of action estimation.

イベント（講義者の行動）	条件				
	POSI	FACE	HAND	CHAR	ERAS
1. “板書している”	停止	後ろ	動作	増加	＊
2. “黒板を消している”	停止	後ろ	動作	減少	なし
3. “板書について説明している”	停止	正面以外	動作	変化なし	＊
4. “学生に対して説明している”	停止	正面	動作	変化なし	＊
5. “左に移動している”	左へ	＊	＊	変化なし	＊
6. “右に移動している”	右へ	＊	＊	変化なし	＊

（＊：don’t care）

複数の画像から黒板消しを含む画素を抽出する．次に抽出した画素の画素値から黒板消しの色の平均値 $\boldsymbol{\mu}_E$ 及び共分散行列 $\boldsymbol{\Sigma}_E$ を求め，これらの値を用いて式 (3) と同様に黒板下部のレール上の画素 $\boldsymbol{x}$ の画素値 $\boldsymbol{C_x}$ が黒板消しの色にどの程度類似するかを表す値 $P_E(\boldsymbol{C_x})$ を求める．そして，求めた値をしきい値処理により 2 値化することで黒板消し領域を抽出し，その重心の座標値をレール上における黒板消しの位置とする．一方，黒板下部のレール上に黒板消し領域が抽出されなかった場合には，黒板消しがレール上に存在しないものと考える．黒板消し位置の抽出結果を図 1 (e) に＋印で示す．

2.4　講義者の行動推定

講義者，板書に関する情報をもとに講義者の行動推定を行い，その結果をイベントとする．まず，現在のフレームで得られた情報と，過去のフレームで得られた情報から，講義者の位置変化（POSI），顔の向き（FACE），手の位置変化（HAND），板書文字量の変化（CHAR），黒板消しの有無（ERAS）について調べる．そして，POSI，FACE，HAND，CHAR，ERAS の状態を条件とした if-then ルールを用いて講義者の行動を推定する．講義者の行動とその条件を表す行動推定規則は経験的に作成しており，表 1 に示す．ただし，表中の ＊ は don’t care を表す．

例えば，講義者の位置が変化せず，黒板の方を向いて手を動かし，板書文字量が増加していれば，“板書している” と推定され，講義者の位置が変化せず，黒板の方を向いて手を動かし，板書文字が減少していて黒板消しがレール上になければ，“黒板を消している” と推定されることになる．

3.　状況理解による撮影領域の決定

講義中に受講者が注目している領域は講義者の行動によって大きく左右される．例えば，講義者が “板書について説明している” 場合に受講者が注目する領域は，講義者と講義者が説明している板書になる場合が多い．このとき，受講者が注目する板書は講義者が直接説明している板書文字のみとは限らず，その板書文字を理解するために重要な情報を含む周囲の板書文字であることも考えられる．この場合に撮影領域に含むべき板書の範囲は講義者の手付近にある板書文字だけではなく，その板書文字と意味的につながりをもつ板書領域（以下，板書ブロックと呼ぶ）であると考えられる．そこで，文献 [15] で報告されている手法を用いて板書ブロックを抽出する．

板書ブロックの抽出例を図 1 (e) に示す．図中の黒線で囲まれた領域が板書ブロックを表している．講義者の手の位置に最も近い板書ブロックを講義者が使用している板書ブロックとすることで，講義者が “板書している” 場合や “板書について説明している” 場合には使用している板書ブロックを考慮した撮影領域が決定できる．具体的には，講義者と使用している板書ブロックが画面いっぱいに映るように撮影領域を決定する．ただし，板書を書き始めた場合などで板書ブロックが小さいときには，その板書ブロックを拡大しすぎて，全体的に窮屈な映像になってしまうため，最大ズーム率を定めて，板書ブロックが小さいときにはズーム率がその値よりも大きくならないようにした．また，本研究では板書ブロック内で板書の内容的なつながりを考慮した分割を行っていないため，板書ブロックが横に大きくなりすぎた場合には，それ以上に分割することができず，ズームアウトした映像になってしまう．このため，文字が小さくなり板書が読みづらくなってしまう可能性が生じる．これを解決する手法としては，板書の文字（静止エッジ）の密度などから適切な解像度を見つけ出す手法も考えられるが，ここでは考えないことにする．

本論文では，講義中に比較的多くの受講者が注目し

表 2 撮影規則
Table 2 Rule of camera work.

イベント	撮影要求
1.“板書している”	講義者と板書に使用している板書ブロック
2.“黒板を消している”	黒板全体
3.“板書について説明している”	講義者と説明に使用している板書ブロック
4.“学生に対して説明している”	講義者（中心となるように）
5.“左に移動している”	講義者（左側に空きを作るように）
6.“右に移動している”	講義者（右側に空きを作るように）

表 3 映像評価規則
Table 3 Rule of evaluation.

イベント	評価法
1.“板書している”	講義者が使用している板書がよく見える
2.“黒板を消している”	黒板と正対している
3.“板書について説明している”	講義者が使用している板書がよく見える
4.“学生に対して説明している”	講義者と目線が合っている
5.“左に移動している”	講義者とカメラの位置が近い
6.“右に移動している”	講義者とカメラの位置が近い

ていると予想される対象物を考え，知的自動撮影システムの撮影規則を作成した．撮影規則は，イベントとイベントをトリガとして発生する撮影要求を対応させたものである．講義を撮影する場合に，講義者の行動が撮影領域を決定するものと考え，講義者の行動をイベントとして，撮影規則を表 2 のように決める．

例えば，講義者が “板書している” 場合には，講義者と講義者が板書に使用している板書ブロック全体を撮影領域とする．一方，講義者が “黒板を消している” ときには，受講者が注目している領域は特にないと考えて，黒板全体を撮影領域としている．

また，撮影領域を決定する手法として，撮影要求に合致した撮影要求領域の加重平均を求めることによって撮影領域を決定するディジタルカメラワークを用いた手法 [6] を用いている．ディジタルカメラワークは，撮影された映像の一部分を拡大呈示することで，擬似的なカメラの動きを実現しているため，解像度が粗くなるという問題が残るが，機械的なカメラワークではなく自然なカメラワークを実現することが可能である．

4. 講義映像における映像評価

4.1 映像評価規則

次にスイッチング規則について考える．複数のカメラを用いて撮影を行う場合にでも，視聴者が同時に見ることのできる映像は一つであるため，複数のカメラから得られた映像の中から，最適な映像一つを選択するスイッチングが必要になる．スイッチャは，それぞれのカメラ位置から撮影された複数の映像を映し出すプレビューモニタを見比べて，今の状況から考えてどの映像を視聴者が要求しているかを判断し，映像を切り換えることが多い [16]．ここでは，このような実際の撮影現場でとられている手法を計算機で実現する．

例えば，講義シーンを撮影対象とする場合に，カメラを選択する基準としては，板書が見やすいことや，

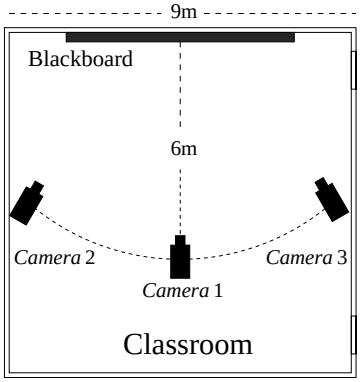


図 2 撮影カメラの配置
Fig. 2 Layout of cameras.

講義者と目線が合っていることなどが考えられる．板書が見やすいように撮影すべき講義状況は，講義者が “板書している” 場合や “板書について説明している” 場合であり，講義者と目線が合うように撮影するのは，講義者が “学生に対して説明している” 場合になることが多い．このように，設定したイベントがトリガとなって映像の評価規準も変化するものと考えられる．そこで，複数のカメラから得られた映像に対してシーン上で起きているイベントに合わせた映像評価法を設計する．講義を対象としたスイッチングのための映像評価規則を表 3 に示す．

ここでは，撮影に使用するカメラは 3 台として，図 2 のように中央・左・右にカメラが黒板中央から等距離になるように設置した．そして，それぞれのカメラから得られた映像を入力として講義者の行動推定を行い，その結果に基づいて撮影領域を決定する．次に，それぞれのカメラによって得られた映像に対する評価を表す映像評価値を求める．

映像評価値は 0〜1 の値をもつように設定することを考える．例えば講義者が “板書している”，または

“板書について説明している” 場合には，使用している板書ブロックと講義者の重なりが少なく，かつ離れすぎないほど使用している板書ブロックは受講者にとって見えやすく映る．そこで使用している板書ブロックの横端から外側に 20 画素離れた位置と講義者位置が重なるときに最大値 1 をとるようなガウス関数を考え，講義者の位置から映像評価値を求める．一方，講義者が “学生に対して説明している” 場合には，映像を見ている受講者と講義者の目線が合うように講義者の顔がより正面を向いている映像が得られた時に評価値が高くなるようにする．ここでは，頭部の肌色画素の数を頭部の画素数で割ったものを映像評価値としている．また，講義者が左または右に移動しているときには，図 2 において講義者が左付近にいるときは $Camera$ 2 ，中央付近にいるときは $Camera$ 1 ，右付近にいるときは $Camera$ 3 によって撮影された映像の評価値を最大値の 1 として，それ以外の評価値を 0 とする．講義者が “黒板を消している” 場合には，黒板全体が撮影されるように規則を作っているため， $Camera$ 1 で撮影した映像の評価値を最大値の 1 とする．

4.2 スイッチングのタイミング

映像評価値は，各撮影位置におけるその時点での映像の見やすさを評価しているが，各カメラから得られた映像を評価して評価値の高い映像にスイッチングを行った場合には，不必要に多くのスイッチングが行われる場合があり，視聴者にとって見づらい映像になることもある．

これは，スイッチングが生じてからしばらくの間は，スイッチングされた映像に視聴者の興味が向いていることに原因があると考えられる．言い換えれば，スイッチングされた映像以外の映像に対する視聴者の興味は失われていることになる．短い時間間隔で何度もスイッチングが行われると見づらい映像になってしまうが，ある程度の時間が経った後にスイッチングが起こる場合には，この難点が解消されることから，これらの失われた興味は時間とともに回復すると考えられる．そこで，時間とともに回復する映像への興味を正規分布関数，

$$\Psi(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{t} \exp\left\{-\frac{(y-m)^2}{2\sigma^2}\right\} dy \quad (4)$$

により興味回復関数として定式化する．興味回復関数は，スイッチング時に採用されなかった映像に対する失われた興味の回復度合を表したもので，スイッチングが生じた時刻から t 時刻後の興味の回復度は，$\Psi(t)$ であるとする．これは，現在見ている映像が飽きてきたことから生じる他の映像に対する興味の回復度合を定式化したもので，m に正の値を与えるとスイッチングが起こった直後の $\Psi(t)$ は 0 に近く，時間が経つにつれて 1 に近づく．また，多数の受講者に対して，固定したカメラ位置から得られる映像に飽きて，他のカメラ位置から撮影した映像を見たいと感じるまでの時間を測定することで，式 (4) の平均や分散を求めることができると考えられるが，ここでは経験的な値を用いる．

一方，一つのフィルムの断片（カット）には 1 ないし複数の演出上の区分（ショット）を含むことが一般的な撮影技法である．本論文におけるカットとは，同一のカメラで撮影している期間に相当し，ショットとは連続して同一のイベントが続いている期間に相当すると考えられることから，イベントの切り替わりに同期してスイッチングを行えば，違和感の少ない自然なスイッチングが期待できる．

そこで，映像評価値を比較評価する際に，前フレームと同一のイベントが継続している場合には，スイッチングされなかった映像に対して $\Psi(t)$ を映像評価値に乗算したものを比較評価する．ただし，イベントが変わった瞬間は，1 カット 1 ないし数ショットとなるように $\Psi(t)$ を乗算することなく映像評価値を比較し，最も映像評価値の高い映像を選択する．これらの処理によって，イベントの変わる瞬間にスイッチングが行われる可能性が高くなることから，カットとショットの同期をとることが可能となり，また，スイッチングが短い時間内に数多く繰り返されることを回避できる．

5. 実験及び考察

5.1 実　　験

本手法を用いて実際に講義の知的自動撮影システムを構築した．使用した計算機の CPU は Pentium III 800MHz，メモリは 256MB である．また，ビデオキャプチャカードは富士通のカラートラッキングビジョン TRV-CPW5，カメラはソニーの DCR-TRV900 を使用し，それぞれ 3 台ずつ用意した．また，ディジタルカメラワークによって撮影した領域を画像ファイルとして順に保存していき，ネットワークを介してスイッチングを行った．画像の処理領域は黒板全面が映る大きさとして 320 × 80 画素とした．また，カメラは

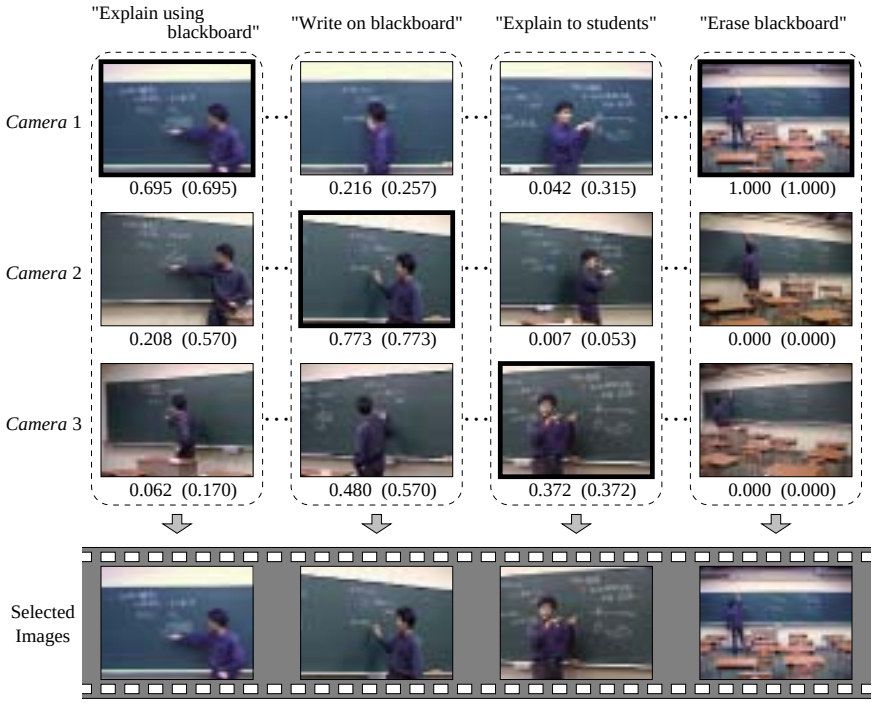


図 3　講義映像生成の例
Fig. 3　Example of lecture video images.

図 2 のように配置しており，左右に設置したカメラ $Camera$ 2，$Camera$ 3 から得られる映像は，黒板に対して正対しておらず，板書ブロックが方形で表せないため，黒板の淵が長方形となるようにアフィン変換を施した後に処理を行った．また，3 台の計算機を用いて同時にイベント抽出などの処理を行っているため，計算機によってイベント抽出や撮影領域の決定結果に違いが生じることもある．この場合には多数決によって最も多くの計算機に支持されている結果を採用した．また，すべての計算機において結果が異なる場合には中心に設置した $Camera$ 1 の映像から算出した結果を用いた．処理速度は毎秒 7 フレーム程度である．

実験によって得られた講義映像の一例を図 3 に示す．図 3 の映像は，左の列から順に講義者が “板書について説明している”，“板書している”，“学生に対して説明している”，“黒板を消している” 場合に各カメラで撮影された映像を表している．各映像の下には映像評価値に興味回復度を乗算した値を，括弧内にはその瞬間の映像評価値を示している．また，それぞれの時刻で最も高い評価が得られた映像を黒い太枠で囲み，受講者が視聴する映像を最下段に示している．

講義者が “板書について説明している” 場合や，“板書している” 場合には，使用している板書ブロックが講義者によって遮られることなく最も見やすいと思われる映像が選択できている．また，“学生に対して説明している” 場合には講義者の顔の向きがより正面となり目線が合っている $Camera$ 3 の映像が選択できている．

5.2　撮影映像の評価

本手法によって撮影した講義映像を受講者に見せて映像の評価を行った．本手法に関係する評価の項目として，撮影領域が適当であるかと撮影カメラの映像選択が適当であるかの 2 項目が必要であると考えた．そこで，大学生の受講者を被験者として，三つのカメラ位置から本論文で述べたカメラワークで撮影したビデオ映像を同時に呈示し，約 30 秒に 1 度すべての映像をストップさせ，それぞれのカメラ位置から得られた映像に順位を付けるように指示した．そして，本手法によって計算機が選択した映像が被験者によって 1 位に選ばれていた場合には 3 点，2 位に選ばれていた場合には 2 点，3 位に選ばれていた場合には 1 点として映像選択に関する点数を付けた．また，1 位のカメラ

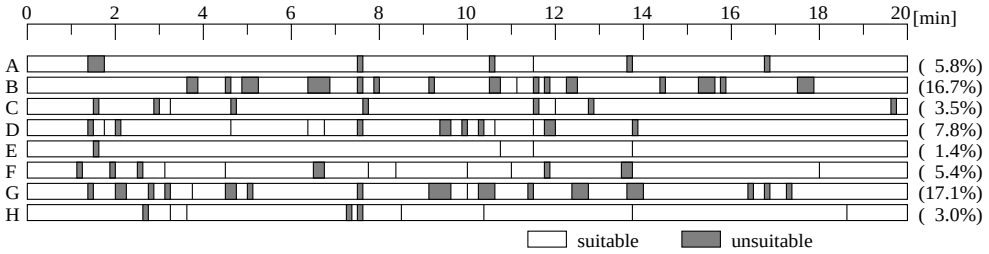


図 4　撮影映像の評価
Fig. 4　Subjective evaluation of lecture video.

位置から得られた映像の撮影領域について「良い」・「普通」・「悪い」の 3 段階評価をさせ，それぞれ 3 点・2 点・1 点として撮影領域に関する点数をつけた．

以上の採点方法で，10 人の被験者に各 30 箇所の映像（画像）300 枚分を評価させた．この結果，撮影領域に関する評価の平均点は 2.658 であり，映像選択に関する評価の平均点は 2.362 であった．両方の評価において，期待値の 2 に比べて十分に高い評価を得ることができた．一方，中央に設置したカメラから得られる映像の撮影位置に関する評価は 2.207 であった．期待値の 2 点に比べて高い評価が得られた理由は中央に設置したカメラからは一般的な映像が得られるため，最低点の 1 をとることがほとんどなかったことに起因している．本手法による映像選択基準では，中央に設置したカメラよりも高い点数を得ることができており，中央に設置した撮影カメラのみを用いて講義を撮影する場合に比べて，映像が単調になることを防ぐだけではなく，効果的な映像にスイッチングすることができていると結論できる．

以上の評価は，映像中の 1 枚の画像としての評価であるため，カメラの動き方などの動画像としての評価が行えていない．そこで，本手法によって撮影した映像を呈示して，受講者に講義映像として違和感を感じる時間帯にボタンを押し続けるように指示し，ボタンを押している時間を計測した．先の評価で撮影領域や撮影位置の評価が行えているものと考え，ここでは映像として違和感を感じる時間帯の測定のみを行い違和感を感じた原因については調べていない．20 分程度の講義映像を用いて評価実験を行い，被験者は A 〜 H の 8 人とした．図 4 には評価の結果をタイムチャートで示し，濃色が違和感を感じた時間帯を表している．右横の括弧内の数字は被験者が違和感に感じた時間の割合を表している．また，被験者の過半数（ 4 人以上）が映像として違和感を感じた時間帯は全映像中の 2.4% であった．多くの受講者が違和感を感じた映像例を図 5 に示す．違和感を感じた理由としては，板書が講義者に隠れて見にくいことや，カメラの動きにぶれがあることなどが考えられ，その原因としては，講義者の行動推定に誤りが生じることや，新しく書き始めた板書ブロックの特定に時間遅れが生じることなどが挙げられる．

図 5　評価の低かった映像例
Fig. 5　Example of unsuitable video images.

5.3 考　　察

提案した知的自動撮影システムについて考察する．

(1)　講義の知的自動撮影手法の提案

講義を撮影対象として，シーンの状況を理解することで撮影領域の決定を行い，異なる位置に配置した複数のカメラから得られた映像を評価することでスイッチングを行う知的自動撮影手法を提案した．また，実際に構築した知的自動撮影システムを用いて講義を撮影することで，板書などが見えやすいカメラ位置から講義に適した映像を撮影することができた．

(2)　撮影した講義映像の評価

本手法によって撮影した講義映像を受講者によって評価させたところ適切な撮影位置から適切な領域を撮影できていることがわかった．特に撮影位置については，教室中央に設置した 1 台のカメラを用いて撮影する場合に比べて，複数の位置に設置したカメラから最

適なカメラ位置を選択することで，より効果的な映像を撮影できていることが確認できた．

(3) 講義者の行動推定

本システムでは，講義状況をリアルタイムで認識する必要があることから，講義者や板書の情報を比較的単純な方法で抽出しており，講義者は肌色の服や半袖の服を着ていないという制約を設けている．このような制約のもとで，ランダムに選んだ画像 120 枚に対して頭部の位置・顔の向き・両手の位置が正しく抽出できているかを調べたところ，それぞれ 91.3% ・86.2% ・87.9% の割合で正しく認識できていた．また，目視によって 20 分間の映像について行動推定が正しく行えているかを判断したところ，映像中の 10.5% の時間で “次に書く（話す）内容を考えている” といった，設定したどの行動にも分類できない時間があり，想定した行動が行われている時間のみでの認識率は 66.6% であった．認識率が高くない原因は，“板書している” と “板書について説明している” の行動の類似性が高いことによる誤認識が多かったためである．しかし，本手法では “板書している” と “板書について説明している” をトリガとした撮影領域も評価規則も同じであるため，これらを同一とした場合の認識率は 92.7% であった．

(4) 撮影規則・映像評価規則の作成

今回用いた撮影規則と映像評価規則は，講義者の行動の中から主な 6 種類をイベントとして経験的に作成した．しかし実際にカメラマンが撮影したり，スイッチャが映像切換を行う際には，更に多くの講義者の行動を理解していると考えられる．認識対象とする講義者の行動（イベント）を増やすことができれば，更に効果的な映像生成が可能になると思われる．また，次に講義者が行う行動の予測ができれば，前もってカメラを動かすことができるため，更に自然な映像生成が可能になると考えられるが今後の課題とする．

(5) 受講者の満足度を満たす映像生成への応用

本研究では講義撮影すべての自動化を目的としている．しかし，双方向テレビなどを用いた遠隔講義への利用を考えた場合には，受講者の要求に合わせた映像生成の必要性も出てくると考えられる．受講者の要求を満たす映像を生成するためには，本手法で自動化している部分の一部である撮影要求の決定部分に受講者の要求を調停するなどの方法で容易に応用できると考えられる．

(6) 他シーンへの適用

今回は講義を撮影対象として知的自動撮影システムを構築した．しかし講義に限らず，撮影したいシーン中のイベントをそのシーン映像から抽出することができれば，スポーツなどの様々なシーンに本手法を適用することができる．

6. む す び

本論文では，シーンの状況理解に基づいて自動的に講義を撮影し，撮影された複数の映像を評価することで適切な映像を選択する知的自動撮影手法を提案した．本手法によって撮影した講義映像を用いて受講者による評価を行い，本手法の有効性を確認した．今後の課題として，行動や動きの予測による撮影領域の決定や講義以外のシーンへの応用などが挙げられる．

文　献

[1] 灰塚凡樹，井上誠喜，“カメラワーク生成に関する一考察，” 信学技報，PRMU96-201, March 1997．

[2] C.S.Pinhanez and A.F.Bobick, “Intelligent studios: Using computer vision to control TV cameras,” Proc. IJCAI’95 Workshop on Entertainment and AI/Alife, pp.69–76, Aug. 1995.

[3] C.S.Pinhanez and A.F.Bobick, “Approximate world models: Incorporating qualitative and linguistic information into vision systems, ” Proc. AAAI ’96, pp. 1116–1123, Aug. 1996.

[4] 松本圭介，須藤　智，斎藤英雄，小沢慎治，“サッカー放送における視点選択のための多視点画像の統合によるボール追跡，” 電学論，vol.121-C, no.10, pp.1530–1539, Oct. 2001．

[5] 井口泰典，土居元紀，眞鍋佳嗣，千原國宏，“スポーツ映像放送のための実時間映像解析によるマルチカメラの自動制御と自動スイッチング，” 映情学誌， vol.56, no.2, pp.271–279, Feb. 2002．

[6] 大西正輝，泉　正夫，福永邦雄，“デジタルカメラワークを用いた自動映像生成，” 画像の認識・理解シンポジウム (MIRU2000), pp.I-331–I-336, Jul. 2000．

[7] 大西正輝，泉　正夫，福永邦雄，“情報発生量の分布に基づく遠隔講義撮影の自動化，” 信学論 (D-II)，vol.J82-D-II, no.10, pp.1590–1597, Oct. 1999．

[8] 亀田能成，石塚健太郎，美濃導彦，“状況理解に基づく遠隔講義のための実時間映像化法，” 情処研報 CVIM-121-11, March 2000.

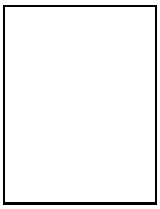
[9] 山口　達，吉川大弘，篠木 剛，鶴岡信治，“講師の動作認識に基づいた遠隔授業映像の自動撮影，” 信学技報 PRMU2000-181，Jan. 2001．

[10] 村上昌史，大西正輝，福永邦雄，“状況理解と映像評価を考慮した講義の知的自動撮影，” 情処研報，2001-CVIM-125-5， Jan. 2001．

[11] 先山卓朗，大野直樹，椋木雅之，池田克夫，“遠隔講義における講義状況に応じた送信映像選択，” 信学論 (D-II),

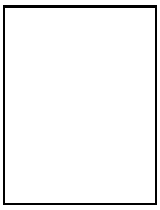
vol.J84-D-II, no.2, pp.248–257, Feb. 2001.
[12] 錦織修一郎，菅沼　明，谷口倫一郎，“黒板講義を対象とした講義自動撮影システムの構築,” 信学技報，PRMU2000-212，March 2001.
[13] M. Minoh and Y. Kameda, “Image a 3D lecture room by interpreting its dynamic situation,” Proc. 4th Int. Workshop on Cooperative Distributed Vision, pp.371–412, March 2001.
[14] 尾関基行，中村裕一，大田友一，“プレゼンテーションの知的撮影システム—手元作業を対象とした適応的カメラワーク,” 信学技報，PRMU2000-104, Nov. 2000.
[15] 大西正輝，泉　正夫，福永邦雄，“講義映像における板書領域のブロック分割とその応用,” 信学論 (D-I)，vol.J83-D-I, no.11, pp.1187–1195, Nov. 2000.
[16] 森田敏夫，伊藤清次（監修），テレビ番組の製作技術 —基礎からノウハウまで，兼六館出版，1991.

（平成 13 年 6 月 18 日受付，11 月 21 日再受付）



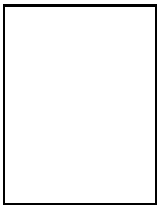
大西　正輝　（学生員）

平 9 阪府大・工・情報卒．平 11 同大大学院博士前期課程了．現在同大学院博士後期課程在学中．コンピュータビジョンに関する研究に従事．電気学会，日本ロボット学会，映像情報メディア学会各学生員．



村上　昌史　（学生員）

平 12 阪府大・工・情報卒．現在同大大学院博士前期課程在学中．画像の認識・理解に関する研究に従事．



福永　邦雄　（正員）

昭 42 阪府大・工・電気卒．昭 44 同大大学院修士課程了．同年同大・工・電気助手．現在同大・工・情報教授．コンピュータビジョン，グラフ理論とその応用などの研究に従事．情報処理学会，システム制御情報学会，IEEE 各会員．工博．