

特別研究報告書

講師追跡撮影画像を基準とした
講義室の広視野画像合成手法

指導教員 美濃 導彦

京都大学工学部情報学科

八代 武大

平成18年2月10日

講師追跡撮影画像を基準とした 講義室の広視野画像合成手法

八代 武大

内容梗概

大学等の教育機関において講義をアーカイブし e-learning のコンテンツとして配信する試みが増加している．e-learning のコンテンツとして講義アーカイブが配信される際には，利用者は，講義の映像を視聴することにより遠隔地で行なわれている講義の情報を得る．このとき，ある時刻におけるアーカイブ映像の利用者の注目対象は，アーカイブ映像の利用者それぞれで異なる．一方で，従来配信されていたアーカイブ映像では，映像の被写体とその解像度は映像の配信者が決定していた．そのため，全てのアーカイブ映像の利用者の，注目対象とその詳細さに対する要求を満たすことはできなかった．

本研究では，アーカイブ映像の利用者の要求を満たすために，アーカイブ映像の利用者が要求する注目対象を全て含み，それぞれに対して十分な解像度をもつ画像を獲得することを考える．このような画像が得られれば，画像中から注目対象を任意に選択でき，また，画質を劣化させることなく注目領域を拡大できる．本研究では，上記のような画像として，講義室の後方中央の視点から講義室の講師方向を撮影した広視野画像の獲得を目標とする．

広視野画像を獲得する方法としては，複数枚の画像を合成する方法（イメージモザイク）があげられる．撮影しているシーンが三次元空間上で平面上にある場合には，平面射影変換という手法を用いてある画像を異なる画像の画像空間へと射影することにより，画像を合成することができる．本研究で対象とする講義室の大半の部分も平面で構成されていることから，本研究では平面射影変換を用いた広視野画像の獲得を考える．

しかし，平面射影変換には，対象とする平面上にない物体を含む画像を射影変換した時に平面上にない物体の領域で画像が歪んでしまうという問題がある．従来では，講師と背景を同時に撮影し，撮影された画像上の物体が全て黒板やスクリーンなどの背景の平面上にあるものとして，画像全体に射影変換を行っていた．そのため，画像の講師領域に歪みが生じ，その結果，複数枚の画像を合成した際に複数の講師の像が生成されてしまう，という問題が生じていた．

そこで本研究では，講師と講義室の背景部分をそれぞれ異なるカメラで撮影

し，背景を撮影した画像を講師を撮影した画像へと射影変換することで，画像を合成する際の講師領域での歪みの発生を防ぎ，自然な広視野高解像度画像を合成する手法を提案する．

まず，講師をパン，チルト，ズームの可能なカメラを用いて，講師の全身が含まれるような構図で追跡撮影することで，講師部分に歪みのない画像を獲得する．ここで，黒板やスクリーン等の講師以外の注目対象を，平面の集合で近似されると仮定する．このとき，講師以外の注目対象のみを含んだ画像には平面射影変換を適用することができるため，複数台の固定カメラを用いて講師以外の注目対象を撮影した画像から，講師以外の注目対象を含んだ広視野背景画像を合成することができる．広視野背景画像を講師追跡撮影画像の画像空間へと射影変換し，広視野背景画像に対して講師追跡画像を優先して合成することで，歪みのない広視野画像を合成する．

広視野背景画像を合成するために，講師追跡撮影カメラの基準姿勢を定め，固定カメラ画像と基準姿勢における講師追跡撮影画像との間の平面射影変換行列 H を，それぞれの画像間の特徴点の対応を手動で与えることにより計算する． H を用いて，広視野背景画像を講師追跡撮影カメラの画像空間上に獲得することで，広視野背景画像全体を一つの射影変換行列 M を用いて講師追跡撮影画像の画像空間へと射影することができる．次に，広視野背景画像と講師追跡撮影画像を合成するために，射影変換行列 M を推定する．射影変換行列 M の推定は，事前に講師追跡撮影カメラの姿勢を適当数サンプリングし，それぞれのサンプリング姿勢に対して事前に射影変換行列を与え，現在の姿勢の近傍のサンプリング姿勢における射影変換行列を用いて線形補間することにより行なう．

本稿ではまず，提案した手法により推定された M の精度を評価するため，講師追跡撮影画像の画像空間へと射影された広視野背景画像と講師追跡撮影画像との，対応する特徴点間のユークリッド距離を評価した．対応点間のユークリッド距離は10画素程度と，強校正を行なうことにより計算した M を用いて射影変換した場合のユークリッド距離と比較して同程度となり，本手法の有効性が確認できた．次に，実際の講義室における広視野画像の合成例を示した．合成された画像上では講師像やその他の領域において歪みの発生がなく，複数の講師像が生成されるという問題は解消されている一方で， M の推定誤差により，広視野背景画像と講師追跡撮影画像とのずれが生じている．また，合成された画像中から注目領域を抽出し，その領域における高解像度化を確認した．

Mosaicing Wide-Field Image in Lecture Room based on Lecturer Tracking Image

Takehiro YASHIRO

Abstract

In recent years, there have been a lot of trials to make archives of lectures and distribute them as e-learning content in educational facilities. When users watch the lecture video, the object in the video that users are interested in is different for each user. Also, the included objects and resolution of the video are determined by the producer of the video. Therefore, existing archiving systems cannot satisfy every user's demand regarding the desired objects and resolution of the video.

In order to be able to satisfy all users' demands, we think of acquiring an image which includes all objects that users could have interest in and that has high resolution of them. Acquiring such an image allows users to choose objects from the image and select the resolution of the objects. Therefore, we aim to obtain a wide-field image taken from the rear center of the lecture room.

Image mosaicing is often used for acquiring wide-field images. When the whole scene of a image is or is approximately a single plane in the 3D world, we can use planar projective transformation to synthesize images. Therefore, we use planar projective transformation for synthesizing the background image of a lecture room, because a lecture room is basically constructed from planes.

In the case when the image that we apply the planar projective transformation to includes non-planar object, the region including the object is distorted. In a lecture room, we can assume that screens and walls are planar, but cannot assume the lecturer to be a plane. Existing methods apply planar projective transformation to images which include lecturer, which results in distortion in the lecturer region.

In this paper, we propose a method for synthesizing a wide-field image of a lecture room which has no distortion in the lecturer region. Firstly, by shooting and tracking the lecturer using a camera the pose (pan, tilt, and zoom) of the which can be controlled, we acquire an image which has no distortion of the whole body of the lecturer. In addition, on the assumption that objects

except for the lecturer can be approximated a plane or a set of planes, we can apply planar projective transformation to images which are captured by several fixed cameras shooting all objects except for the lecturer and synthesize a wide-field background image of the lecture room. Finally, we project the wide-field background image onto the image plane of the lecturer tracking camera at a current pose. Then, we acquire the wide-field image of a lecture room.

In order to acquire a wide-field background image of the lecture room, we calculate the planar projective transformation matrix \mathbf{H} between the image of lecturer tracking camera at a chosen base pose and each image of the fixed cameras by assigning the coordinates of the corresponding points by hand. Using \mathbf{H} , we acquire a wide-field background image which is on the image plane of the lecturer tracking camera at the base pose. So, we can project the wide-field background image onto the lecturer tracking image plane at the current pose with single projective transformation matrix \mathbf{M} . Next, we estimate \mathbf{M} from the current pose of the lecturer tracking camera. We select some pose of the lecturer tracking camera and calculate \mathbf{M} corresponding to the pose in advance. We calculate \mathbf{M} corresponding to the current pose by bilinear interpolation from \mathbf{M} corresponding to neighbors' sampled pose.

Firstly, in order to evaluate the precision of the estimated \mathbf{M} , we calculate the Euclidean distance between corresponding points in the wide-field background image and lecturer tracking image. The Euclidean distance of the proposed method was about 10px, which was almost the same as the Euclidean distance corresponding to the \mathbf{M} calculated from the data of strong calibrated cameras. Next, we synthesized a wide-field image of a lecture scene using the proposed method. We acquired the image which has no distortion of the lecturer region and the other region. On the other hand, difference between the projected wide-field background image and lecturer tracking image occurred, which came from the error of the estimation of \mathbf{M} . Finally, in order to confirm that the image synthesized by our method has higher resolution than the image captured by a general camera, we compared the region of the lecturer's face and confirmed that the image acquired by our method has higher resolution.

講師追跡撮影画像を基準とした 講義室の広視野画像合成手法

目次

第1章	はじめに	1
第2章	平面射影変換による講義室の広視野画像合成	3
2.1	平面射影変換による広視野画像合成	3
2.1.1	平面射影変換行列とその求め方	3
2.1.2	平面射影変換を用いた広視野画像合成	5
2.2	講義室における平面射影変換の適用	6
2.2.1	講師を含む画像に平面射影変換を適用する際の問題点	6
2.2.2	講義室の広視野画像の合成	7
第3章	講義室における広視野画像合成手法	8
3.1	画像合成に用いるカメラとその画像	8
3.2	固定カメラを用いた広視野背景画像合成	9
3.2.1	広視野背景画像合成のための平面射影変換行列 H の計算	9
3.2.2	広視野背景画像の合成	10
3.3	広視野背景画像と講師追跡撮影画像との合成	15
3.3.1	広視野背景画像と講師追跡撮影画像の間の射影変換行列 M の推定	15
3.3.2	M を用いた広視野背景画像と講師追跡撮影画像の合成	17
第4章	実験と評価	17
4.1	実験環境	17
4.2	射影変換行列 M の推定精度の評価	18
4.3	合成された広視野画像の評価	19
4.3.1	講師追跡撮影カメラの姿勢およびサンプリング間隔	19
4.3.2	合成された広視野背景画像の評価	21
4.3.3	合成された広視野画像の位置ずれの評価	22
4.3.4	解像度に関する評価	22
第5章	まとめと今後の課題	24

謝辭	25
參考文獻	25

第1章 はじめに

大学等の教育機関において、講義をアーカイブし e-learning のコンテンツとして配信する試みが増えてきている。講義アーカイブを利用することで、時間的、空間的な制約により現場で講義を受講できない人々も教育を受けることができるという利点がある。

e-learning のコンテンツとしての講義アーカイブでは、講義を撮影した映像が提供される。講義アーカイブの利用者は、その映像を視聴することにより遠隔地で行われた講義に関してさまざまな情報を得ることができる。講義アーカイブ映像の利用者の注目対象は、実際に講義に出席している人々の注目対象と同様に、講師や黒板、ホワイトボード等の、講義において情報を発信している物体である。このとき、アーカイブ映像の利用者のある時刻における注目対象や、注目対象に要求する詳細さは、アーカイブ映像の利用者それぞれで異なる。しかし、従来の講義アーカイブシステムでは、配信される映像の被写体とその解像度は配信者が決定しているため、全ての利用者に対して、アーカイブ映像の注目対象とその詳細さに対する要求を同時に満たす映像を配信することはできなかった。

従来の講義アーカイブでは、注目対象を個々に撮影しそれを選択して配信していたため、上述の問題が生じていたが、注目対象を全て含んだ画像を得ることができれば、アーカイブ映像の利用者は画像中から注目対象を任意に選択できる。さらに、それぞれの注目対象に対して十分な解像度が得られていれば、画質を劣化させることなく注目領域を拡大でき、注目対象の詳細さに関する要求を満たすことができる [1]。そこで本研究では、アーカイブ映像の利用者の映像に対する要求を満たすために、注目対象を全て含み、かつ、それぞれの注目対象に対して十分な解像度をもつ画像 (以下、広視野画像と呼ぶ) を獲得することを考える。

講義アーカイブ映像において、利用者が最も注目すると考えられる対象は講師である。また、講師の移動範囲、講師以外の注目対象の存在範囲は、講義室の前方中央を中心に、講義室の左右に分布している。そのため、講義アーカイブのための広視野画像としては、講義室の後方中央視点から講師方向を撮影した画像が望ましい。このような広視野画像の獲得方法としては、単独で広視野画像を撮影することのできるカメラを利用する方法、またはイメージモザイク

ング手法 [2][3][4] を用いて単独もしくは複数のカメラで撮影した複数枚の画像を合成し、一枚の画像を生成する方法が考えられる。しかし、単独で広視野、高解像度画像を撮影できるカメラは一般に高価であり、要求される解像度と視野が大きくなるにつれてその価格はさらに高くなる。また、視野が広がるにつれ画像の端の領域で歪みが生じる。一方で、イメージモザイクング手法により画像を合成する際には、カメラの数を増やすことにより、画像の解像度を保ちつつ視野を広げることができる。そこで本研究では、イメージモザイクング手法を用いて広視野画像を合成する。

イメージモザイクングの際に用いられる射影変換の手法としては、円筒投影 [4] や平面射影変換 [5] が挙げられる。円筒投影は、任意の奥行きを含むシーンを撮影した画像を合成することができるが、画像を獲得する際にはカメラの光学中心を一致させる必要がある。これに対し、平面射影変換は、撮影しているシーンが三次元空間上で平面であるか、または十分遠景であり平面で近似できることが必要であるが、カメラの強校正を行う必要が無く、弱校正のみでモザイクングを行うことができる。本研究では、講義室のような広い空間に設置した複数のカメラで同時に撮影した画像を合成することを考えており、それらのカメラの光学中心を一致するように設置することは困難である。また、本研究での対象シーンである講義室は、壁面や黒板等、平面の集合で表すことができるため、平面射影変換を適用することが可能である。そのため、講義室の広視野画像を獲得するためには、平面射影変換を用いることが適切であり、本研究でも平面射影変換を用いる。

一方、一般に、三次元空間上で平面上にない物体を含むシーンを撮影した画像に対して平面射影変換を適用した場合、平面上にない物体の領域は実際に存在すべき位置には射影されない。そのため、射影変換により得られた画像を他の画像と合成する際に、歪みが生じたり、画像間での重複領域において複数の像が生成される等の問題が生じる。このように、講義が行なわれている状況で講義室の広視野画像を合成する際には、対象とするシーンは講師を含むため、平面の集合で表すことはできず、かつ、撮影するカメラを講義室内に配置する必要があるため対象シーンが十分遠景であるとも言えないことから、特に講師の部分の画像が歪んでしまうという問題を避けることができない。

実際、従来では、講義室において講師の後方にある黒板やスクリーンなどの背景領域を一枚の基準平面とし、撮影された画像中の全ての領域が基準平面上

にあるものとして，背景領域を撮影した画像へと射影変換することで広視野画像を合成していたために，講師領域など基準平面上に無い部分の画像が歪んでしまうという問題があった [6]．そこで本研究では，講師を追跡撮影するカメラを別途用意し，従来手法とは逆に，背景領域を撮影している複数のカメラの画像を用いて合成した広視野画像を，講師を追跡撮影しているカメラの画像空間へと射影変換することによって，講師領域を射影変換することによる歪みの発生を防ぎ，自然な広視野画像を合成する手法を提案する．

本稿の構成は次の通りである．2章では，従来の平面射影変換を用いた画像合成と，それらを講義室に適用する際の問題点，および講義室へ適用するための提案手法の概要について述べる．3章では，複数の固定カメラで得られる画像を用いた広視野な背景画像の合成手法と，その画像を講師を撮影しているカメラの画像空間へと射影することによる広視野画像の合成手法について述べる．4章では，実験によって得られた結果を元に提案手法についての評価を行う．最後に，5章で結論を述べる．

第2章 平面射影変換による講義室の広視野画像合成

2.1 平面射影変換による広視野画像合成

2.1.1 平面射影変換行列とその求め方

複数台のカメラで三次元空間上のある点 M を同時に撮影する際には，点 M の一方のカメラ画像での投影点の座標を得たときに，同一の点の他方のカメラ画像への投影点の座標に対する拘束式が与えられる．一般に，カメラ C_1, C_2 の画像 I_1, I_2 における M の投影点 m_1, m_2 に対して，それらの同次座標表現 \tilde{m}_1, \tilde{m}_2 の間には式 (1) の関係が成立する．

$$\tilde{m}_2^T F \tilde{m}_1 = 0 \quad (1)$$

以降では，点 m に対する同次座標表現を \tilde{m} で表す．ここで F は基礎行列と呼ばれ，画像 I_1 上の点 m_1 に対応する点が画像 I_2 上である直線上に存在する，という拘束を与える．

さらに，対象としている点が三次元空間上のある平面上に存在する場合には，画像 I_1 上の点 $m_1 = (x_1, y_1)^T$ に対応する画像 I_2 上での点 $m_2 = (x_2, y_2)^T$ を一意に定めることができる．この関係は，式 (2) または式 (3) で表される．

$$\tilde{m}_2 \approx H \tilde{m}_1 \quad (2)$$

$$x_2 = \frac{h_1x_1 + h_2y_1 + h_3}{h_7x_1 + h_8y_1 + 1}, y_2 = \frac{h_4x_1 + h_5y_1 + h_6}{h_7x_1 + h_8y_1 + 1} \quad (3)$$

ここで \approx は定数倍の不定性を残して等しいことを表す記号である．式 (2) における H は平面射影変換行列と呼ばれる， 3×3 の行列である．また， h_i は H における第 3 行 3 列要素が 1 となるように H を正規化したときの H の各要素であり， H は， h_i を用いて式 (4) のように表現される．

$$H \approx \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{pmatrix} \quad (4)$$

ここで， H は，カメラ C_1 の座標系におけるカメラ C_1 からカメラ C_2 への回転行列 R ，並進ベクトル t ，基準とする平面の法線ベクトル n と平面までの距離 d ，カメラ C_1, C_2 の投影行列 A_1, A_2 を用いて，式 (5) により表すこともできる．

$$H \equiv A_2 \left(R + \frac{tn^T}{d} \right) A_1^{-1} \quad (5)$$

すなわち，二台のカメラに対して強校正が行われており，空間上での平面の位置が既知である場合には，式 (5) により平面射影変換行列を求めることができる．このとき，利用される二台のカメラ間では回転，並進，焦点距離の変化の運動が許され，同一平面上にあるシーンを撮影しているかぎり任意のカメラ間で射影を行なうことができる．ここで，式 (5) において，カメラの並進ベクトル t を 0 とする場合を考える．これは，カメラの運動を回転と焦点距離の変化に制限することと同一であり，このとき，平面に関するパラメータ d と n は H に寄与しないため，シーン中の任意の点を同一の射影変換行列を用いて射影変換できる．つまり，二台のカメラ間の運動が回転と焦点距離の変化のみの場合には，撮影されたシーンが三次元空間において平面上にない場合でも，これらのカメラ間で射影変換を行うことができる．

一方，二台のカメラに対して強校正が行われていない場合でも，三次元空間上のある平面上に存在する複数の点について，画像間での対応を得ることができれば，式 (3) より，一組の対応点に対して $h_i (1 \leq i \leq 8)$ を未知数とした式を二つ得ることができる．よって，三次元空間上で同一平面上にある 4 つ以上の点の組に対して画像間での対応を得ることができれば，8 個の未知数に対して 8 個以上の方程式を得ることができ， $h_i (1 \leq i \leq 8)$ を計算することができる．

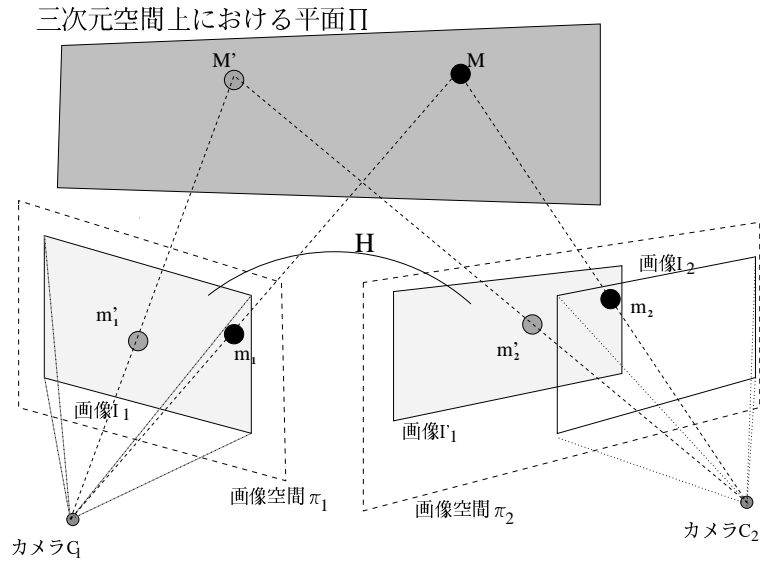


図 1: 平面射影変換による広視野画像合成

すなわち，二台のカメラ間の平面射影変換行列は，それぞれの画像上で4つ以上の点の対応を与えることにより，各カメラに対して強校正を行なうことなく，弱校正のみで求めることができる．

2.1.2 平面射影変換を用いた広視野画像合成

図 1 に示すように，シーンが一枚の平面 Π で近似でき，平面 Π 上のある点 M が画像 I_1 上の座標 m_1 に投影されているとする．画像 I_1, I_2 間の平面射影変換行列 H が与えられている場合には，式 (2) から，点 M の画像 I_2 上への投影点の座標 m_2 を得ることができる．このように，画像 I_1 上の全ての点は画像 I_2 の画像空間 π_2 上へ射影することができる．ここで図 1 における点 M' のような，画像 I_1 上には投影されるが，画像 I_2 上には投影されない三次元空間上の点を考える．このときも同様に，式 (2) に m_1' を代入することで，画像空間 π_2 上での投影点の座標 m_2' を計算することができる．ここで， m_1' と m_2' は三次元空間上での同一点をそれぞれ π_1, π_2 上に投影した点であり，点 M' を拡散反射面上の点であると仮定すると， m_1', m_2' の画素値を同一であるとみなすことができるため， m_1' の画素値を m_2' の画素値とすることができる．

このように，画像 I_1 上の全ての点に対して画像空間 π_2 での対応する点の座標を計算することで，画像 I_1 を，画像 I_2 と同じ画像空間 π_2 上の画像 I_1' に射影することができる．画像 I_1' と画像 I_2 を合成することにより，広視野画像を合成することができる．

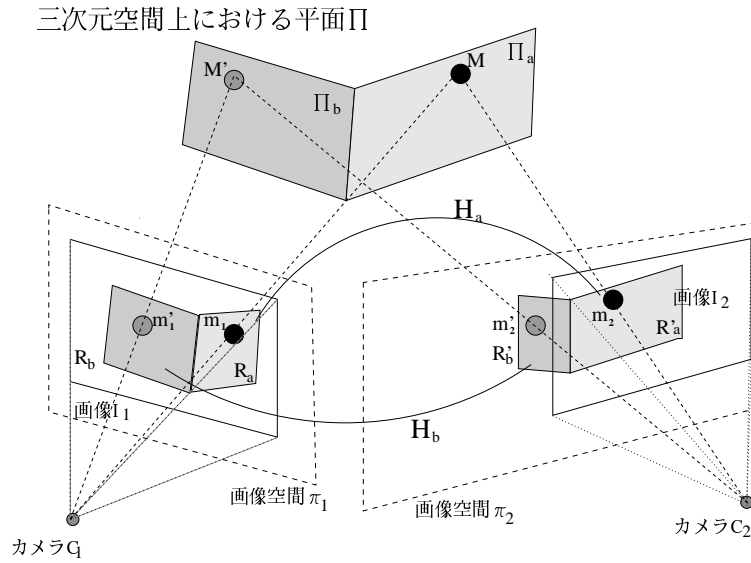


図 2: 平面領域ごとの平面射影変換を用いた画像合成

2.2 講義室における平面射影変換の適用

2.2.1 講師を含む画像に平面射影変換を適用する際の問題点

本研究で対象とする講義室は，スクリーンやホワイトボード，壁面などのような複数の平面から構成されており，図 2 のように，講義室を撮影した画像は実世界上での平面に対応する領域へと分割することができる．このとき，三次元空間上での平面 Π_a, Π_b に対して，画像 I_1, I_2 上でそれぞれの平面に対応する領域から特徴点を抽出することにより，それぞれの平面に対応する平面射影変換行列 H_a, H_b を計算することができる．このとき，領域 R_a に関しては， H_a を用いて画像空間 π_2 に射影することができる．また，同様に R_b に関しては， H_b を用いることで画像空間 π_2 に射影できる．このように，画像を三次元空間上のある平面に対応する領域に分割し，それぞれの領域ごとに射影変換行列を与えて射影することで，広視野画像を合成できる．

ここで，講師以外の注目対象は平面で構成されると仮定すると，講師以外の物体を撮影しているカメラ画像は三次元空間上で平面上にない領域を含まないことになるので，画像を壁面やスクリーン等の平面領域に分割し，それぞれの平面領域ごとに対応する平面射影変換を用いて射影することで，講義室の広視野画像を合成することができる．従って，講師以外の注目対象をそれぞれ撮影するカメラを用意することで，講師を除いた部分において歪みのない広視野画像を合成することが出来る．以降では，この方法で合成された画像を広視野背



図 3: 背景平面を基準とした平面射影変換における講師部分のずれ

景画像と呼ぶ．このように合成された広視野背景画像と講師を撮影した画像を合成することにより，講義での注目対象を含む広視野画像を得ることができる．

しかしながら，三次元空間上のある平面を基準とした平面射影変換行列を用いて，基準とする平面上にない部分を含む画像を射影すると，その平面上にない部分は射影先の画像上で存在するべき位置からずれてしまう．例えば，実際の講義では，講師が黑板やスクリーン等の前に立って講義を行うため，背景の平面と同一の平面上に存在するとは言えず，図 3 に示すように，講師の部分を背景の平面を基準とした平面射影変換行列で射影すると講師の部分に歪みやずれが生じるという問題がある．

2.2.2 講義室の広視野画像の合成

2.2.1 節で述べた問題を解決する方法の一つとして，講師を一つの平面で近似し，その平面と背景平面との間の距離を推定することにより，各時刻における講師平面の平面射影変換行列を計算し合成する手法 [6] も提案されている．しかしこの手法では，獲得された画像における講師像は奥行きが感じられない画像となり，講師領域の歪みが解消されているとはいえない．

講師部分の歪みは，背景平面上にない講師領域を，背景平面を基準とした平面射影変換を用いて射影変換していることが原因で生じる．一方，講師部分に歪みのない画像は，講師を撮影するカメラにより得ることができる．また，広視野背景画像では，被写体の大部分は黑板やスクリーン等，平面の集合で構成

されていると仮定できる物体であるため，広視野背景画像の講師を除いた部分では，平面射影変換を適用しても歪みは生じない．そこで本研究では，従来のように広視野背景画像の画像空間に講師画像を射影するのではなく，講師画像の画像空間に広視野背景画像を射影することにより，講師部分に歪みのない広視野画像を合成する手法を提案する．

第3章 講義室における広視野画像合成手法

3.1 画像合成に用いるカメラとその画像

本手法では，固定設置された，パン，チルト，ズームの制御可能なカメラ（以下講師追跡撮影カメラと呼ぶ）を用いて講師を追跡撮影し，講師部分に歪みのない画像を獲得する．また，固定カメラを用いて講師以外の注目対象を撮影し，固定カメラにより獲得された画像を合成することで講師以外の注目対象を含んだ広視野背景画像を獲得する．以下では，講師追跡撮影カメラにより撮影された画像を講師追跡撮影画像と呼ぶ．講師追跡撮影カメラを，全ての固定カメラ画像に対して重複領域をもつように光軸と画角を調整したときの講師追跡撮影カメラのパン角度，チルト角度，およびズーム値の組を，講師追跡撮影カメラの基準姿勢とする．講師追跡撮影カメラの，基準姿勢に対しての相対的なパン角度，チルト角度，ズーム値を講師追跡撮影カメラの姿勢とし，撮影中の各時刻における講師追跡撮影カメラの姿勢は常に獲得できるものとする．

固定カメラにより講師以外の注目対象を撮影する際には，全ての注目対象がいずれかの固定カメラ画像で撮影されるように固定カメラを配置する必要がある．また，広視野画像を合成する際には，固定カメラの画像中に講師が侵入することで，歪んだ講師像が生成されてしまうという問題を解決するために，本研究では，合成された広視野背景画像に対して，講師追跡撮影画像を優先して合成することで，画像中に歪んだ講師像が含まれないようにする．このとき，合成された広視野背景画像における講師部分が，講師追跡撮影カメラの撮影範囲に含まれる必要がある．そのためには，固定カメラ間の視差が少ないことが望ましく，また講師追跡撮影画像は，講師の全身を含むような構図で撮影される必要がある．

本研究では，上記の要求と，講義の観測，アーカイブの際に使用するための様々な要求を考慮し，図4及び表1のように，講義室に配置した固定カメラ，講

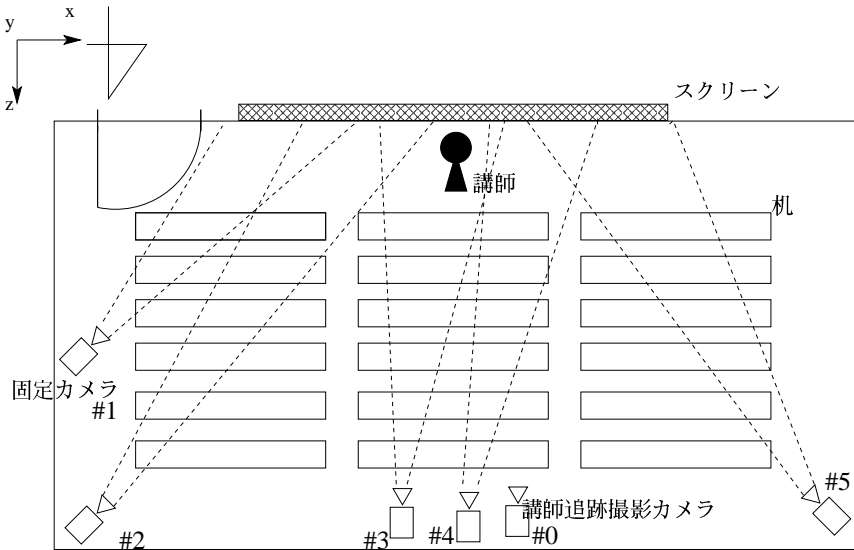


図 4: 講義室俯瞰図

表 1: 講義室における各カメラの位置

カメラ	#0	#1	#2
$(x(\text{m}), y(\text{m}), z(\text{m}))$	(7.1, 2.6, 8.6)	(-0.5, 2.4, 4.6)	(0.05, 2.5, 8.0)
カメラ	#3	#4	#5
$(x(\text{m}), y(\text{m}), z(\text{m}))$	(6.7, 2.6, 8.6)	(7.0, 2.6, 8.7)	(13.7, 2.5, 8.3)

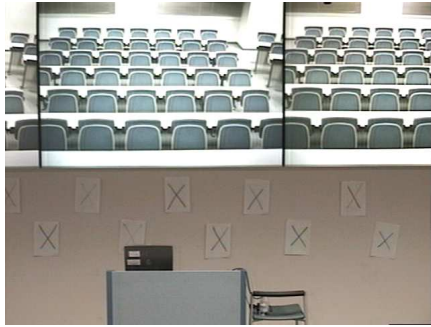
講師追跡撮影カメラを用いた。図 4 においてカメラ番号 #1 から #5 が固定カメラであり、#0 が講師追跡撮影カメラである。ここで、各カメラの座標は、原点を入口付近 (図 4 における左上) の床上にとり、東から西を指す方向に x 軸を、床から天井を指す方向に y 軸を、南から北を指す方向に z 軸をとった座標系のものである。

講師追跡撮影カメラ #0 の基準姿勢において撮影された画像と講師を追跡撮影した際の画像の例を図 5 に、固定カメラ #1 ~ #5 を利用してスクリーンとその周辺の壁面を撮影した画像の例を図 6 に示す。

3.2 固定カメラを用いた広視野背景画像合成

3.2.1 広視野背景画像合成のための平面射影変換行列 H の計算

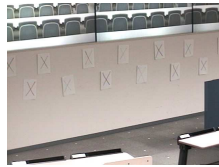
広視野背景画像 I_B を講師追跡撮影カメラ C_0 の現在の姿勢における画像空間上へと射影する場合、広視野背景画像 I_B が講師追跡撮影カメラの基準姿勢にお



カメラ#0の基準姿勢における画像

カメラ#0の講師を追跡撮影した画像の例

図 5: 講師追跡撮影カメラ画像の例



(i) カメラ#1

(ii) カメラ#2

(iii) カメラ#3

(iv) カメラ#4

(v) カメラ#5

図 6: 固定カメラ #0 ~ #5 の撮影例

ける画像空間 π_0 上であれば、講師追跡撮影画像 I_0 のカメラから広視野背景画像 I_B のカメラへの運動は、回転と焦点距離の変化のみであるため、2.1.1 節で述べたように、広視野背景画像 I_B 全体を単一の射影変換行列で射影することができる。そこで、広視野背景画像 I_B を合成するために、まず、固定カメラの画像それぞれに対して、講師追跡撮影カメラの基準姿勢における画像との平面射影変換行列を求める。以降では図 4 における講師追跡撮影カメラ #0 を C_0 とし、固定カメラ台数を n としてそれぞれを $C_i (1 \leq i \leq n)$ により表す。まず、固定カメラ $C_i (1 \leq i \leq n)$ により撮影された画像 I_i と I_i 上で観測される平面 j の組それぞれに対して、講師追跡カメラ C_0 の基準姿勢における画像 I_0 との平面射影変換行列 $H_{i,0}^j$ を計算する。平面射影変換行列の計算は、2.1.1 節で述べたように、二枚の画像間でそれぞれ対応する四組以上の特徴点から計算することが出来る。そこで本研究では、予め講師が存在しない状態で講義室を撮影し、撮影された画像中から手動で特徴点を抽出し対応づけることで $H_{i,0}^j$ を計算しておく。

3.2.2 広視野背景画像の合成

図 7 に示すように、各固定カメラ $C_i (1 \leq i \leq n)$ で撮影した複数の画像 I_i を、3.2.1 節で求められた平面射影変換行列を用いて講師追跡撮影カメラの基準姿勢

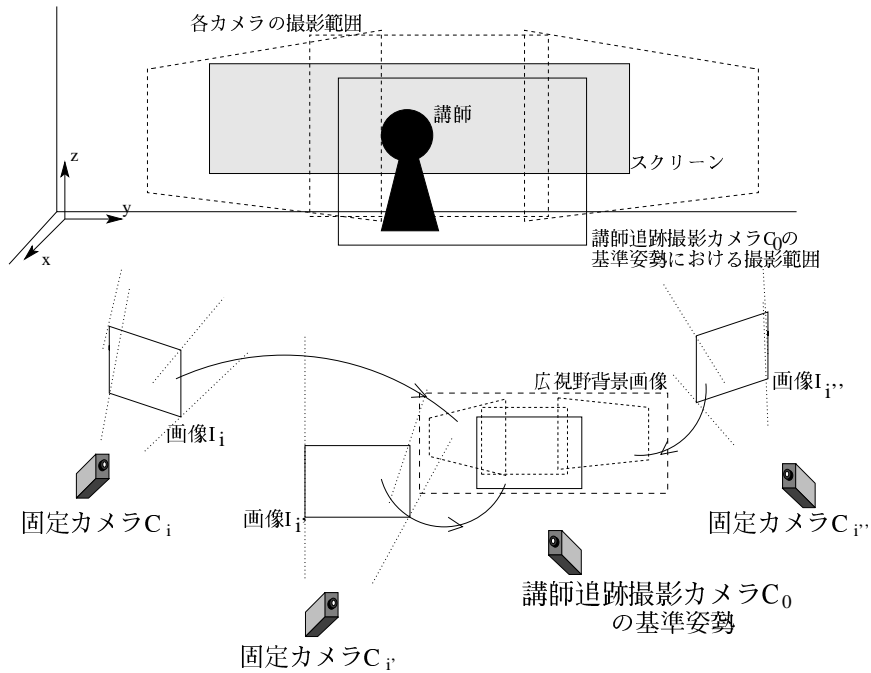


図 7: 広視野背景画像の合成

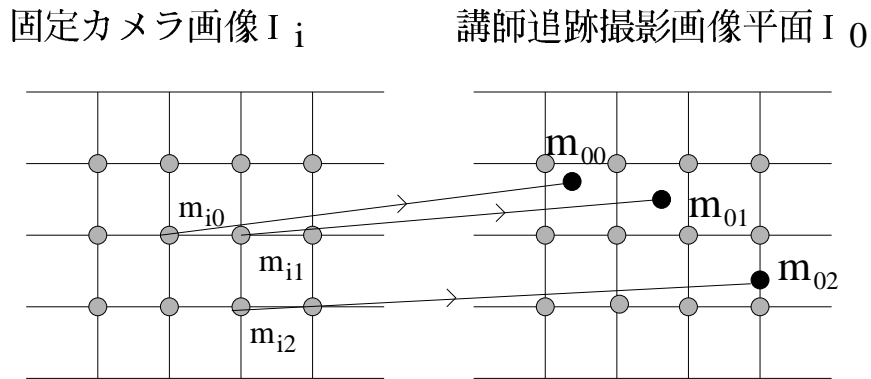


図 8: 射影変換後の画素の位置

における画像へと射影変換する．そして，射影変換された画像を合成することにより，広視野背景画像 I_B を合成する．

固定カメラ画像 I_i において，三次元空間上の平面 j に対応する領域を I_i^j とする． I_i^j 上の点 m_{ik}^j の同次座標 \tilde{m}_{ik}^j は，講師追跡撮影カメラの画像空間 π_0 の同次座標 $\tilde{m}_{0k} (= H_{i,0}^j \tilde{m}_{ik}^j)$ に射影される．しかしながら，図 8 に示すように，一般には \tilde{m}_{0k} に対応する画像座標 m_{0k} の各座標値は整数とならず， m_{0k} をもとに広視野背景画像 I_B における各画素の画素値を直接求めることはできない．

そこで本手法では，図 9 に示すように，画像を拡大，縮小する際に用い

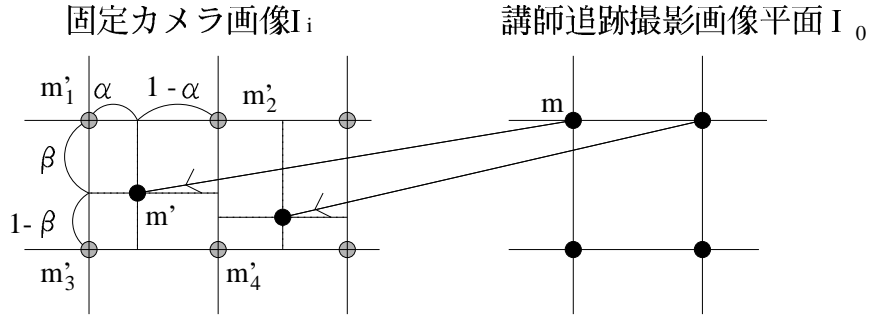


図 9: 線形補間による画素値の計算

られる線形補間の手法を用いて広視野背景画像 I_B の画素値を計算する．まず広視野背景画像 I_B の各画素 m に対して式 (6) を用いて逆変換を行い，固定カメラ画像 I_i 上で m に対応する座標 m' を計算する．

$$\tilde{m}' = H_{i,0}^j{}^{-1} \tilde{m} \quad (6)$$

座標 m' における画素値を，式 (8) に示すように m' の近傍 4 点の画素値を用いた線形補間により計算し，得られた画素値を m における画素値とする．

$$m'(c) = (1 - \alpha)(1 - \beta)m'_1(c) + \alpha(1 - \beta)m'_2(c) + (1 - \alpha)\beta m'_3(c) + \alpha\beta m'_4(c) \quad (7)$$

$$c \in \{R, G, B\}$$

ただし， $0 \leq \alpha < 1$ ， $0 \leq \beta < 1$ であり，それぞれ， m' の x ， y 座標の小数部である．また，画素 m の成分 c の値を $m(c)$ で表わすものとする．ここでは RGB 表色系を用いるため， $c \in \{R, G, B\}$ である．

固定カメラで撮影した画像 I_i と I_j が重複する領域や，固定カメラで撮影された異なる平面領域 I_i^k と I_j^l が重複する場合など， I_B 上の同一の画素 m に対して複数の画素値を計算することができる場合が存在する．このような画素に対して，計算された画素値の平均をその画素の画素値とする場合には，重複領域と非重複領域の境界上で画素値が不連続に変化し，シーン中に存在しない境界線が合成画像中に現れることがある (図 10)．これは，撮影時に使用したカメラ間の色温度の違いや照明条件により，対応点間で画素値が異なることに由来するものである．そこで本研究では，まず各固定カメラ画像の輝度値の平均が等しくなるように画素値を調整し，広視野背景画像全体で不自然に画素値が変化することを防ぐ．そして，複数の画像が重複する領域では，射影元の画素値に対する重み付き和を計算し，その値を画素値とすることで，重複領域において画

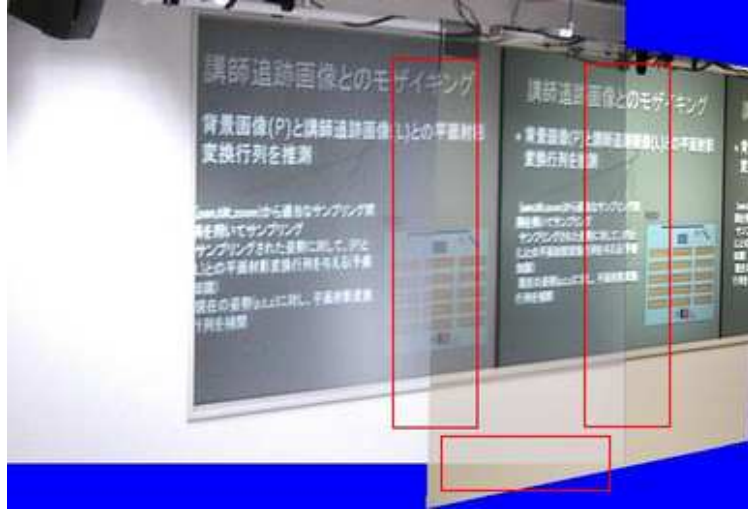


図 10: 境界線の生じた画像

素値をなめらかに変化させる。

上記のような画素値の調整を実現するために、本研究では、HSV 表色系を用いて調整処理を行う。HSV 表色系では、色を色相 (H)、彩度 (S)、輝度 (V) の三値で表現する。まず、入力画像 I_i の画素値を HSV 表色系へと変換する。そして、画像 I_i における輝度 (V) の平均 E_i を計算する (式 (8))。

$$E_i = \frac{\sum_{m \in I_i} m(V)}{w_i \times h_i} \quad (8)$$

ここで w_i, h_i は入力画像 I_i の幅と高さを表す。また、先ほどと同様に、画素 m の V 成分を $m(V)$ で表わす。全ての固定カメラ画像 I_i に対して E_i を計算し、さらにそれらの平均を計算する (式 (9))。

$$E = \frac{\sum_i E_i}{n} \quad (9)$$

固定カメラ画像 I_i の輝度値の平均が E となるように、式 (10) に示すように輝度値を調整する。

$$\forall i, \forall m \in I_i \\ m^{new}(V) = \frac{E}{E_i} m^{old}(V) \quad (10)$$

次に、画素値が複数計算できる領域では、それぞれの画素値に対して、対応する画像の境界に近づくにしながら 0 に近づくような重みをかけることで、画像の境界に対応する部分で滑らかに画素値が変化するようにする [3]。本研究

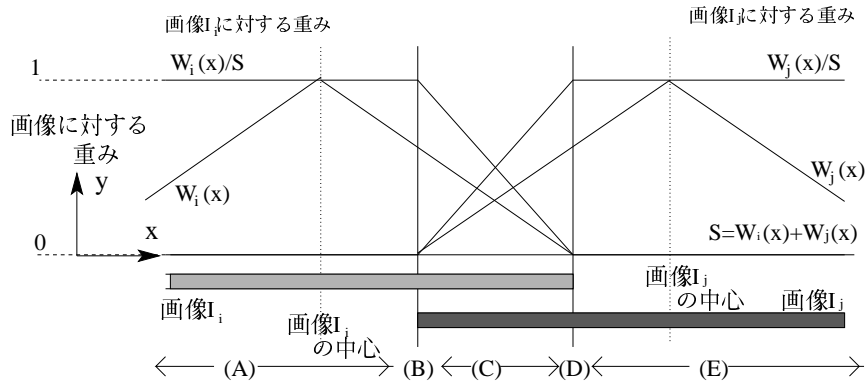


図 11: 重複領域における画素値の計算

では, 各固定カメラ C_j の画像 I_j に対して, 中心からの距離に応じた重み関数 $W_j(x, y)$ を式 (11) のように定義する.

$$W_j(x, y) = W_j(x)W_j(y) \quad (11)$$

$$W_j(x) = \begin{cases} \frac{x}{w_j} & \text{if } x \leq \frac{x_c^{(j)}}{2} \\ \frac{w_j - x}{w_j} & \text{if } x > \frac{x_c^{(j)}}{2} \end{cases}$$

$$W_j(y) = \begin{cases} \frac{y}{h_j} & \text{if } y \leq \frac{y_c^{(j)}}{2} \\ \frac{h_j - y}{h_j} & \text{if } y > \frac{y_c^{(j)}}{2} \end{cases}$$

ここで, $(x_c^{(j)}, y_c^{(j)})$ は画像 I_j の中心の座標である.

広視野背景画像 I_B 上の画素 m に対して n 個の画素値が計算されるとする. ここで, 画素 m に射影される画素 $m'_i (1 \leq i \leq n)$ が固定カメラ C_j 上の画素であるとき, この画素を $m'_i{}^{(j)}$ で表す. また, 画素 $m'_i{}^{(j)}$ の座標を $(x_i^{(j)}, y_i^{(j)})$ で表す. m に射影される全ての画素 $m'_i{}^{(j)}$ に対して $W_j(x_i^{(j)}, y_i^{(j)})$ を計算し, $W_j(x_i^{(j)}, y_i^{(j)})$ に応じて重み付き和をとることで画素値を計算する.

$$m(c) = \sum_{m'_i{}^{(j)}} \frac{W_j(x_i^{(j)}, y_i^{(j)})}{S} m'_i{}^{(j)}(c) \quad (12)$$

$$\text{ただし } S = \sum_i W_j(x_i^{(j)}, y_i^{(j)}) \quad (13)$$

図 11 に, 重複する画像が I_i と I_j の二枚である場合の例を示す. ここでは簡単のため, 画像の x 軸方向における重み付けを図示する. 図 11 における (A) の領域では, 画像 I_i と画像 I_j が重複しないため, 画像 I_i に対する重みは 1 となり, 画像 I_i の画素値がそのまま合成画像における画素値となる. 図 11 での (B) は,

画像 I_i と画像 I_j との重複が起こり始める点であり、画像 I_j の境界に対応する。この点では、画像 I_j に対する重み $W_j(x)/S$ は 0 となり、この点の画素値は I_i の画素値となる。(C) の領域では、 $W_i(x)$ と $W_j(x)$ の値を利用してそれぞれの重みを計算する。このとき、 $W_i(x)$ 、 $W_j(x)$ はそれぞれ連続的に変化し、またその和 S も連続的に変化する。従って、 $W_i(x)/S$ 、 $W_j(x)/S$ も連続的に変化することになり、領域 (C) では滑らかに画素値を混合することができる。(D)、(E) では (B)、(A) と同様のことが言える。以上より、(A) から (E) にかけて画素値を連続的に変化させることができる。

3.3 広視野背景画像と講師追跡撮影画像との合成

3.3.1 広視野背景画像と講師追跡撮影画像の間の射影変換行列 M の推定

3.2.2 節の方法により、広視野背景画像 I_B は講師追跡撮影カメラの基準姿勢における画像として獲得されている。従って、広視野背景画像から講師追跡撮影画像への射影変換行列はシーン中の平面の数によらず一つの射影変換行列 M で扱うことができる。本節では、広視野背景画像と講師追跡撮影画像を合成するための射影変換行列 M を求める手法について述べる。

射影変換行列 M を計算するには、講師以外の注目対象上にある特徴点を検出し、対応を求める必要がある。しかし、講師追跡カメラは常に講師を追跡撮影しているため、講師以外の領域において特徴点の存在が保証されず、またカメラや物体の移動に伴うブラーにより、特徴点を精度よく抽出することができないため、特徴点の座標を用いて射影変換行列 M を計算することも困難となる。そこで本研究では、撮影時の画像から抽出される特徴点に基づいて射影変換行列 M を計算するのではなく、撮影時の講師追跡カメラの姿勢に基づいて射影変換行列 M を推定する。

まず、講師が存在しない状況で、講師追跡撮影カメラの姿勢を適当な間隔でサンプリングする。パン、チルトに対してはその角度を、ズームにおいてはその倍率をカメラ姿勢の単位として用いる。サンプリングされた各姿勢に対して、広視野背景画像と講師追跡撮影画像との間の射影変換行列を、特徴点とその対応を与えることにより計算する。これにより、パン、チルト、ズームを軸とした三次元空間上で、格子状にサンプリングされた点に対して射影変換行列が与えられる。

講師追跡撮影カメラの各時刻における姿勢は、サンプリングされた姿勢と完

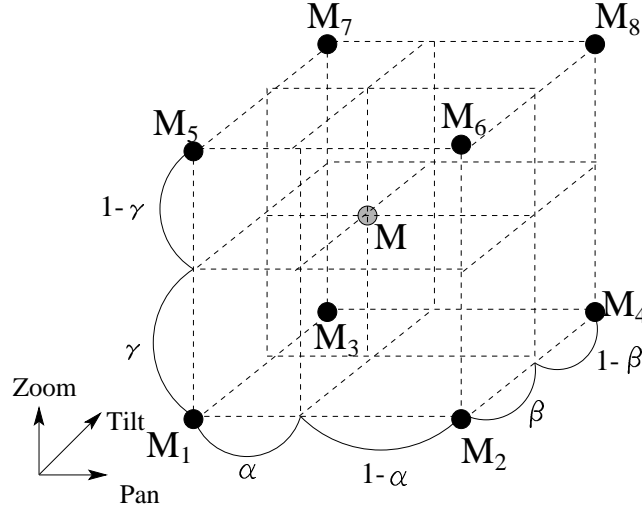


図 12: 線形補間による射影変換行列 M の計算

全には一致しないため，サンプリングされた姿勢に対して求めた射影変換行列をそのまま用いることはできない．そこで，現在の姿勢に対する射影変換行列 M を，近傍 8 点における射影変換行列の値から，パン，チルト角度，ズーム値に基づき式 (14) のように線形補間により計算する (図 12) ．

$$M = (1 - \alpha)(1 - \beta)(1 - \gamma)M_1 + \alpha(1 - \beta)(1 - \gamma)M_2 + \dots + \alpha\beta\gamma M_8 \quad (14)$$

式 (14) における α, β, γ は，現在の講師撮影カメラ C_0 の姿勢を (p, t, z) ，近傍のサンプリング点におけるパン，チルト角度，ズーム値を $p^-, p^+, t^-, t^+, z^-, z^+$ ($p^- \leq p < p^+, t^- \leq t < t^+, z^- \leq z < z^+$) としたときに，それぞれ式 (15) のように計算する．

$$\begin{aligned} \alpha &= \frac{p - p^-}{p^+ - p^-} \\ \beta &= \frac{t - t^-}{t^+ - t^-} \\ \gamma &= \frac{z - z^-}{z^+ - z^-} \end{aligned} \quad (15)$$

ただし，射影変換行列には定数倍の不定性があるため，線形補間を行う際に用いる近傍 8 点の射影変換行列 $M_i (1 \leq i \leq 8)$ は，各要素の二乗和が 1 となるようにあらかじめ正規化しておく．

3.3.2 M を用いた広視野背景画像と講師追跡撮影画像の合成

得られた射影変換行列 M をもとに広視野背景画像を講師追跡撮影画像へと射影変換し画像を合成する．広視野背景画像と講師追跡画像の重複領域では，両者の画素値から合成画像の画素値を決定する．3.2.2 節の方法で獲得される広視野背景画像では講師部分において歪みが生じているため，広視野背景画像の画素値を用いると合成画像上で歪みが残ってしまう．そこで，講師追跡撮影画像と広視野背景画像との重複領域での画素値には，講師追跡撮影画像の画素値を採用する (式 (16)) ．

$$m(x, y) = \begin{cases} m_0(x, y) & \text{if } m_0(x, y) \text{ exist} \\ m_{bg}(x, y) & \text{otherwise} \end{cases} \quad (16)$$

ここで， $m_0(x, y)$ ， $m_{bg}(x, y)$ は，それぞれ講師追跡撮影画像，広視野背景画像における画素 (x, y) での画素値である．これにより，重複領域において歪みのない画像を合成することが可能となる．

第4章 実験と評価

4.1 実験環境

3章で提案した広視野画像合成手法を評価するために，提案手法を用いて求めた射影変換行列 M の推定精度を評価した．また，実際の講義室で，提案手法を用いて広視野画像を合成し，そのずれに対する評価を行った．

本実験は，京都大学学術情報メディアセンター (南館)R202 講義室に複数台のカメラを設置して行った．カメラの配置は，図4，表1と同様である．講義室の寸法は，幅 $14m$ ，奥行き $9.8m$ ，高さ $3.2m$ である．講師追跡撮影カメラ #0 にはSONY 製の EVI-D30 を用いた．固定カメラには，カメラ #1，#2，#4，#5 にSONY 製 EVI-G20 を，#3 にはSONY 製 EVI-D30 を用いた．また，各カメラで獲得した画像の解像度は 640×480 画素である．

射影変換行列 M を計算するためのサンプリング点は，パンは -18.2° から 10.2° まで約 2.3° おきに 14 個，チルトは -2.5° と -3.3° の 2 個，ズームは画角 23.9° と 19.4° の 2 個をサンプル値とし，計 56 個の姿勢に対して平面射影変換行列を求めた．

また，広視野背景画像を合成する際に用いる，講師追跡撮影カメラの基準姿勢における画像と固定カメラ画像との平面射影変換行列 H と，広視野背景画像

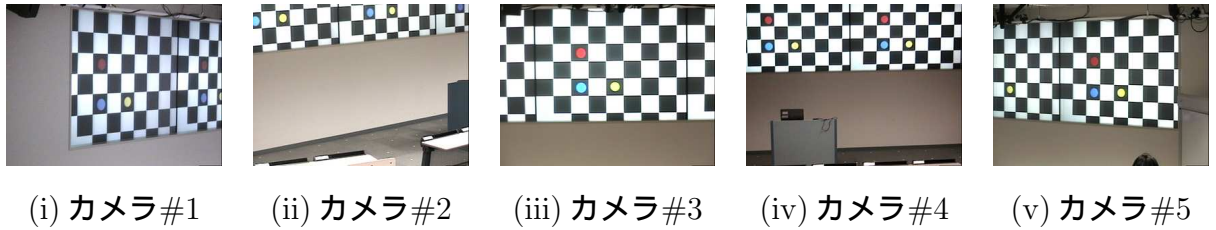


図 13: 固定カメラで撮影した格子パターン

表 2: 対応する特徴点間の平均ユークリッド距離

	講師追跡撮影カメラのパン角度			平均誤差 (画素)
	-17.0°	-3.4°	10.2°	
提案手法による推定誤差 (画素)	4.82	10.35	8.94	8.04
強校正下での誤差 (画素)	14.58	9.51	5.98	10.02

と講師追跡撮影画像間の射影変換行列 M は、それぞれの画像から手動で 10 組程度の特徴点の対応を選択し、金谷ら [7] により公開されているプログラムを用いて計算した。

4.2 射影変換行列 M の推定精度の評価

射影変換行列 M の推定精度を調べるために、推定された射影変換行列 M を用いて講師追跡撮影画像の画像空間へと射影された広視野背景画像上の特徴点と、その点に対応する講師追跡撮影画像上での特徴点との間の位置ずれの大きさを評価した。講義室前方に設置された、オンラインスライド提示用スクリーンに格子パターンを表示し、その様子を撮影した画像を評価に用いた (図 13)。

3.3 節で述べた手法により推定された射影変換行列 M を用いて、広視野背景画像を講師追跡撮影画像空間へと射影変換する。変換された広視野背景画像と講師追跡撮影画像において対応する特徴点間のユークリッド距離を求め、その平均を推定誤差とする。講師追跡撮影カメラのパン角度を 3 通りに変化させたときの推定誤差を表 2 の上段に示す。また、講師追跡撮影カメラに対して強校正を行い、カメラの姿勢から直接計算した M を用いて広視野背景画像に射影変換を行なった場合の誤差を表 2 の下段に示す。

提案手法では、講師追跡撮影カメラのサンプル姿勢において射影変換行列を事前に獲得するために、各サンプル姿勢に対して画像中から特徴点を抽出しそ

表 3: 実験に用いた講師追跡撮影カメラの姿勢とサンプリング間隔

実験 No.	講師追跡撮影カメラの姿勢 (パン角度, チルト角度, 画角)	サンプリング間隔 (パン角度, チルト角度, 画角)
(A)	(9.1°, 2.5°, 23.9°)	(2.3°, 0.83°, 4.5°)
(B)	(-13.6°, 2.5°, 23.9°)	(2.3°, 0.83°, 4.5°)
(C-i)	(10.2°, 2.92°, 21.5°)	(2.3°, 0.83°, 4.5°)
(C-ii)	(10.2°, 2.92°, 21.5°)	(4.6°, 0.83°, 4.5°)

の対応付けを行う必要があるが、三次元空間上におけるカメラの位置や、カメラの内部パラメータを知る必要がないという点で強校正と比較してその手間は軽減されている。一方で、表 2 から分かるように、提案手法によって得られた射影変換行列を用いた場合と強校正によって得られた射影変換行列を用いた場合とでは精度はほぼ等しいことが確認できる。

4.3 合成された広視野画像の評価

3 章で述べた手法に基づいて講義室における広視野画像を合成し、得られた画像について評価を行った。

4.3.1 講師追跡撮影カメラの姿勢およびサンプリング間隔

本実験では、表 3 に示した 4 通りの講師追跡撮影カメラの姿勢とサンプリング間隔の組み合わせに対して広視野画像を合成した。(A) は、講師追跡撮影カメラの姿勢が、講義室においてスクリーンに向かって右側を撮影しているサンプリング姿勢の一つと一致する場合である。(B) は、講師追跡撮影カメラの姿勢が、スクリーンに向かって左側を撮影しているサンプリング姿勢の一つと一致する場合である。(A)、(B) により、各サンプリング姿勢に対して手動で与えた射影変換行列 M による合成画像のずれと、講師追跡撮影カメラの姿勢の変化による合成画像全体の変化を調べた。(C) は、講師追跡撮影カメラの姿勢がサンプリング姿勢のどれとも一致しない場合であり、また、(C-i) と (C-ii) では射影変換行列 M を推定する際のサンプリング間隔を変化させて実験を行った。(C) により、射影変換行列 M の推定結果によるずれと、サンプリング間隔の変化の影響を調べた。図 14 にそれぞれの姿勢に対する講師追跡撮影カメラ、固定カメラそれぞれの入力画像を示す。

実験 No.	講師追跡撮影 カメラの画像	固定カメラの画像				
	#0	#1	#2	#3	#4	#5
(A)						
(B)						
(C-i), (C-ii)						

図 14: 実験に用いた画像列

	固定カメラ				
	#1	#2	#3	#4	#5
スクリーン 部					
壁面部					

図 15: 各固定カメラにおけるスクリーン部と壁面部 (明るい部分)

実験を行った講義室では、壁面に対してスクリーンが 10cm ほど奥に位置しているため、この二つの平面を同一平面として扱えば合成画像に歪みが生じる。この歪みは、特にカメラ #1, #2, #5 などの黒板、壁面の法線方向とカメラの光軸方向とが大きな角度をもつカメラの画像を射影変換したときに顕著に現れる。そこで、本実験ではシーンをスクリーン部分と壁面部分に分け、それぞれに対して平面射影変換行列を求めた。図 15 に、固定カメラ画像におけるスクリーン部と壁面部を示す。それぞれの画像における明るい部分がスクリーンまたは壁面の平面に対応する領域である。

実験 No.	広視野背景画像	講師部分の拡大図
(A)		
(B)		
(C-i), (C-ii)		

図 16: 合成された広視野背景画像

4.3.2 合成された広視野背景画像の評価

3.2 節で述べた手法により合成した広視野背景画像と、それぞれの画像の講師部分の拡大図を図 16 に示す．なお，合成される広視野背景画像は，講師追跡撮影カメラの姿勢に依存しないため，(C-i) と (C-ii) に対して合成される広視野背景画像は同一である．(A)，(C-i)，(C-ii) では固定カメラ #4 と #5 の画像中に，(B) では固定カメラ #1 と #2 の画像中に講師が含まれている．そのために，図 16 の右図で確認できるように，それぞれの広視野背景画像中で複数の講師像が生成されている．

また，それぞれの入力画像の明るさが異なっているが，画素値を連続的に変

化させることで境界線の出現を軽減できている。

4.3.3 合成された広視野画像の位置ずれの評価

3章で述べた手法によって合成した広視野画像を，図 17 に示す。

(A)，(B)，(C-i)，(C-ii) それぞれの広視野背景画像上に生成されていた講師画像の歪みは，講師追跡撮影画像を合成することにより広視野画像上では解消されており，講師領域で歪みのない画像を合成できていることが分かる。

(A)，(B) に対して得られた広視野画像では，講師追跡撮影カメラの画像空間上へと射影した広視野背景画像と講師追跡撮影画像とのずれは数画素程度である。これに対し，(C-i) に対して得られた広視野画像では，講師追跡撮影カメラの画像空間上へと射影した広視野背景画像と講師追跡撮影画像とのずれは 10 画素程度となり，(A)，(B) の場合と比較して大きくなっている。画像上では，スクリーン上に投影された文字は 2 画素から 3 画素の太さであり，この誤差は合成された画像を講義アーカイブ映像として利用するには無視できる大きさであるとはいえない。(A)，(B) の場合にも数画素の画素ずれが生じていることから，本手法を用いる際には，各サンプリング姿勢に対して，手動で対応点を与え平面射影変換行列を計算するにあたって十分な精度が必要となると考えられる。

また，(C-i) と比較して，サンプリング間隔を大きくとった (C-ii) の場合では，合成された画像上での位置ずれはより大きくなっている。本手法では，十分な数のサンプリング姿勢に対して，平面射影変換行列を与える必要があると考えられる。

4.3.4 解像度に関する評価

本手法で合成した広視野画像と，これと同等の画角をもつように一般のカメラで講義室を撮影した画像に対し，その解像度について比較，評価を行った。

本実験では (A) に対して 1346×808 画素，(B) に対して 1409×753 画素，(C-i) に対して 1550×838 画素の広視野画像を生成することができた。また，このうち画素値を計算することができた画素数は (A) では 761,712 画素，(B) では 715,627 画素，(C-i) では 962,657 画素であり，同等の画角で，一般的なカメラの画素数 640×480 画素で撮影した場合と比較すると，その画素数はおよそ 2.3 ～ 3.1 倍となっている。

ここで講師の顔部分がほぼ同等の大きさとなるように，(A) における広視野画像と，一般的なカメラで撮影した画像からそれぞれ注目領域を選択すると，本手法により合成された画像では 80×80 画素，一般的なカメラで撮影された画

(A)



(B)



(C-i)



(C-ii)



図 17: 広視野画像合成結果



図 18: 本手法で合成した画像からの顔領域の抽出

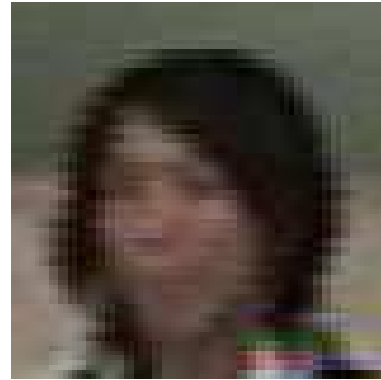


図 19: 一般的なカメラ画像からの顔領域の抽出

像では 40×40 画素となり、注目領域においてより大きな解像度を持つ画像を獲得できている。それぞれの画像での講師の顔領域を同等の大きさへと拡大した例を図 18, 図 19 に示す。図 19 では講師の目や鼻などがつぶれてしまっているが、図 18 では講師の目や鼻などを確認でき、注目領域において高い解像度が得られることが確認できる。

第5章 まとめと今後の課題

本研究では、講義における注目対象を講師とそれ以外へと分け、講師以外の注目対象を撮影した広視野背景画像を講師追跡撮影画像空間へと射影変換することで、強校正を行うことなく、歪みの無い広視野画像を合成する手法を提案した。

実験では、本手法を用いることにより講師部分に歪みの無い、自然な広視野画像を合成できることを確認した。一方で、本手法により推定された射影変換行列 M を用いて広視野背景画像と講師追跡撮影画像を合成する際には広視野背景画像と講師追跡撮影画像との間でずれが生じる場合も確認され、 M の推定に精度が必要であることが分かった。今後の課題としては、 M の推定方法の改善を行うことや、推定された射影変換行列 M に対して合成する画像間のずれをなくすように最適化手法を適用することが挙げられる。

また、本手法で合成された広視野画像は、講師追跡撮影カメラの画像空間上の画像として獲得される。そのため、図 17 における (A) と (B) の様に、講師追跡撮影カメラの姿勢の変化にともなって、広視野画像における背景領域の見

え方が変化する．このような画像をアーカイブ映像として提供する際のインタフェースの検討も今後の課題として挙げられる．

謝辞

本研究を行うにあたり多くの御教示，熱心な御指導を賜りました美濃導彦教授に深く感謝致します．日頃より熱心な御指導と御助言を賜りました角所考助教授に深く感謝いたします．また，日頃より親身な御指導と，本報告書の作製におきましても多くの御助言を頂きました大阪工業大学西口敏司講師に深く感謝いたします．最後に，日頃よりの熱心な御指導と多くの御助言を頂きました正司哲朗博士をはじめとするプレゼンスグループの方々，並びに美濃研究室の皆様にも深くお礼を申し上げます．

参考文献

- [1] 横井隆雄, 遠山聖司, 藤吉弘巨: イメージモザイクによる講義のデジタルアーカイブと再生, インタラクション 2004 (2004).
- [2] 千葉直樹, 蚊野浩, 美濃導彦, 安田昌司: 画像特徴に基づくイメージモザイク, 電子情報通信学会, Vol. J-82-D-II, pp. 1582–1589 (1999).
- [3] R.Szeliski: Video mosaics for virtual environment, *IEEE Comput. Graphics & Appl.*, Vol. 16, issue 2, pp. 22–30 (1996).
- [4] Chen, S. E.: QuickTime VR - An Image-Based Approach to Virtual Environment Navigation, *SIGGRAPH'95*, pp. 29–38 (1995).
- [5] 出口光一郎: ロボットビジョンの基礎, コロナ社.
- [6] 遠山聖司, 藤吉弘巨: 複数台のカメラによる画像合成のための前景物体の距離推定, 電気関係学会東海支部連合大会 (2005).
- [7] 清水慶行, 太田直哉, 金谷健一: 信頼性評価を備えた最適な射影変換の計算プログラム, Vol. 98-CVIM-111-5 (1998-05), pp. 33–40 (1998).