

### 1. はじめに

藤吉研究室では、現在イメージモザイクによる講義のデジタルアーカイブとその映像を用いた e-learning システムの実現を目指している。本研究では、非同期カメラ映像間の同期フレーム検出と音声情報を基にした講義のインデキシングについて検討する。

### 2. 非同期カメラ映像によるイメージモザイク

複数のカメラで撮影された映像からイメージモザイクにより高解像動画像を生成し講義のデジタルアーカイブを行う。本研究では講義室に 3 台の非同期カメラ (DV カメラ) を配置して撮影した非同期映像を対象とする (図 1 参照)。

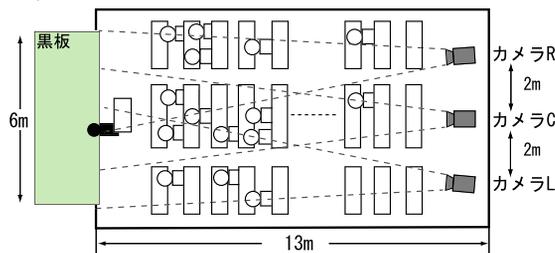


図 1：講義室におけるカメラ配置

#### 2.1. 音声信号による同期フレームの検出

3 台の DV カメラ間は同期されていないため、モザイク処理を施す前にカメラ映像間の同時刻のフレームを検出する必要がある。一般に動画像は 1 秒間に 30 フレーム、音声信号は 48kHz でサンプリングされるため、同期フレームの検出には時間分解能が高い音声信号を用いる。以下に同期フレーム検出法を示す (図 2)。

step1 基準カメラ映像 (カメラ C) の音声信号から 1 秒間の平均パワーを求め、そのパワーが最も高い場所から 5 秒間を基準信号区間とする。基準信号をフーリエ変換により短時間パワースペクトルを求める。

step2 カメラ R,L の音声信号のパワースペクトルを求める。基準区間のパワースペクトルと他のカメラのパワースペクトルとの類似度を SAD (Sum of absolute differences) により 9[msec] 毎計算する。最も類似度の高いフレームを同期フレームとする。

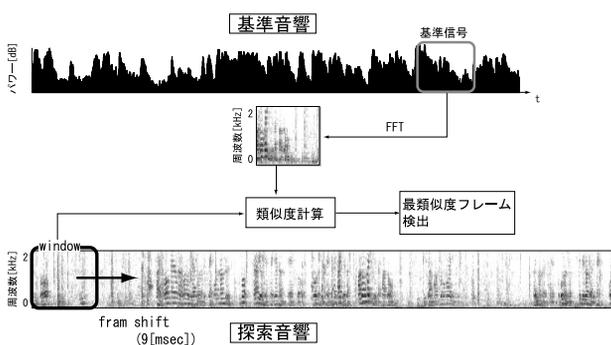


図 2：同期フレーム検出の概要

#### 2.2. 同期フレーム検出実験

本手法の有効性について検討するため、講師がマイクを使う講義とマイクを使わない講義の映像をそれぞれ 20 パターン撮影し、同期フレーム検出の評価実験を行った。基準信号の決定には、パワーの分散を用いる方法と比較した。その結果を表 1 に示す。ここでは、検出したフレームの時刻が正しい時刻の ± 0.1 秒以内なら正解とした。

表 1 より、類似度を計算する周波数範囲を狭くすることで、正解率を向上させることができた。また、基準信号を決定する際、分散値ではなく最大パワーを用いる方が良いことが分かる。

表 1：同期フレーム検出正解率

	5kHz 以下		2kHz 以下	
	分散	最音量	分散	最音量
マイクなし	25	85	45	95
マイクあり	30	95	55	100

### 3. 音声情報を基にした講義のインデキシング

デジタルアーカイブされた講義映像データを閲覧する際に、全てを繰り返し再生すると非常に時間がかかるため、必要な部分を選択的に試聴する機能が必要となる。そこで、講義モザイク映像より発話区間の検出を行い、これをインデックスに利用する。

#### 3.1. 発話区間の検出

発話区間の検出には、ダウンサンプリングした音響信号の振幅対数パワーを求め閾値処理により検出する。3 人の講師による講義映像 (各 10 分間) の発話区間の検出結果を表 2 に示す。

表 2：発話区間検出結果

	再現率 (%)	適合率 (%)
Movie1	84.7	97.9
Movie2	78.4	99.2
Movie3	83.2	92.2
平均	82.1	96.4

実験の評価には、次式に示す再現率と適合率を用いる。

$$\text{再現率} = \frac{\text{正検出}}{\text{正検出} + \text{検出浅れ}}, \quad \text{適合率} = \frac{\text{正検出}}{\text{正検出} + \text{誤検出}}$$

表 2 より、約 96.4% の適合率を得ることができた。Movie3 の適合率がやや低い理由は、講義中にカメラの近くの生徒の咳や話し声を講師の発話と誤検出したことが原因である。

#### 3.2. 講義のインデキシング

講義における講師の状態は、“説明のみ”、“板書のみ”、“説明 + 板書”、“その他 (見回り)” の 4 状態を遷移していると考えられる。これらの状態遷移を抽出することは、講義のインデキシングに有効である。図 3 に、講義の一部 (50[sec]) に対して振幅対数パワー、閾値処理により検出した発話区間とフレーム間差分により得られた講師の動きの有無を示す。図 3 より、発話の有無と講師の動きから“板書のみ”と“その他 (見回り)”の状態を推定することが可能であることが分かる。現在、黑板画像、講師の動き分析と音声信号を複合的に用いた“板書”状態の検出について検討している。

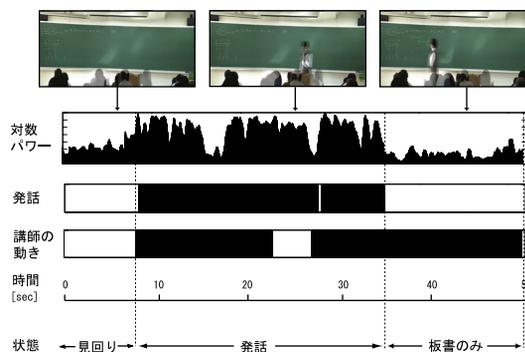


図 3：講義における発話区間と講師の動き検出例

### 4. まとめ

本研究では、音声を用いた同期フレーム検出と発話区間の検出について検討した。今後は同期フレームの検出を向上させるとともに正確なインデキシングを検討する予定である。本研究を基にしたビジネスプランは、愛知県学生 IT ベンチャーアイデアグランプリに入賞した。