情報処理学会
グループウェアとネットワークサービス研究会
2005年11月10日, 11日

講義コンテンツ自動作成システムのための板書講義撮影方法

○市村哲; 福井登志也; 井上亮文; 松下温
東京工科大学

講師が講義前にビデオカメラを 1 台設置しておくだけで，その収録映像から自動的に講師の動画映像と板書の静止画とを作成し，即座にインターネット配信可能とする講義自動収録システム ChalkTalk-OSS を構築した．板書静止画については，画像処理を施して講師の姿を消去する工夫を施している．1 台の固定ビデオカメラのみで黒板全体を記録できるように，解像度の高い民生用ハイビジョンカメラを用いて実装した．講義室環境に適した講師位置検出処理，人間が行うカメラワークに近い講師動画作成処理等について提案する．

A Method Suitable for Recording ChalkTalk lectures.

Satoshi Ichimura; Toshiya Fukui; Akifumi Inoue; Yutaka Matsushita
Tokyo University of Technology

We developed ChalkTalk-OSS that automatically produces E-learning materials from videotaped chalk talks. The system separately extracts a lecturer's image and writing on blackboard from video images recorded with a single high-definition digital camcorder, and stores the former as a streaming video and the latter as a series of snapshot images. Methods for locating the lecturer in the class room environment and creating lecture-videos with smooth camera-work are described.

1. はじめに

電子プレゼンテーションが普及した現在でも，大学や予備校等の教育機関において黒板を用いた講義は根強い支持を得ている．たとえば，電子プレゼンテーション中心の講義であっても，重要箇所の説明の際や受講者の理解ペースにあわせて説明する必要がある際に板書が頻繁に用いられている．また，大学や予備校において，数学や英語では，ほとんどが黒板を用いて授業が行われているのが現状である．一方，近年，インターネットを利用した新しい情報サービスとして WBT(Web-Based Training) などの E-ラーニングが注目されている．いつでもどこでも自分の進捗状況に合わせて学習を進められることや，通学費や施設費などの経費を削減できるなどの利点がある．

以上のような背景から，著者らは，黒板の板書を用いた講義を E-ラーニング教材として提供できるようにすることが必要であると考え，これを目的とした講義自動撮影システムの試作および提案を行ってきた[1]．本試作システムを用いれば，講師が講義前にビデオカメラを 1 台設置しておくだけで講義を自動撮影できるため，専属カメラマンが不要となり，講義を担当する講師自らが極めて簡単に講義撮影できる利点がある．実装方法としては，従来のビデオカメラ（DV カメラ等）で黒板全体（著者らの大学において最も一般的な黒板のサイズは 750cm ×120cm）を撮影した場合には解像度が絶対的に足りないために，解像度の高い民生用ハイビジョンカメラ（HDV カメラ）を利用した．たとえば，DV カメラ画像では判別困難な文字や講師の顔が，ハイビジョンカメラ画像では明確に判別できる．

しかしながら，講義映像をインターネット経由で配信することを考慮した場合，ハイビジョン映像をそのままインターネットに送出するためには極めて多くのネットワーク帯域が必要になるという問題がある．90 分の講義をハイビジョンカメラで記録すると容量は 17GB に達する．そこで著者らは，板書の映像は変化が少ないことに着目し，講師近辺の映像は動画として保存し，板書全体は静止画として保存することで従来の問題点を緩和した．板書に変化があったのみ静止画を作成し，板書を静止画アニメーションとしてインターネット配信できるよ

うになっている．さらに，この試作システムでは，画像処理を施して板書静止画から講師の姿を消去し，講師の陰のない静止画を作成して保存するようになっている．動画であれば，板書の一部が講師の陰となって見えない場合にでも，少し待てば講師がその部分の前から立ち去るために問題とならないが，板書を静止画アニメーションとして記録してしまうと，その見えない部分が長時間にわたって読めない状態になってしまうためである．

本稿では，以前提案した上記試作システムの実用性と信頼性を向上させ，Linux 上で動作するオープンソースソフトウェアとして再構築した「ChalkTalk-OSS」システムの提案を行う．特に，講師の姿や動きを検出する画像処理アルゴリズムを大幅に改良し，講義室環境により適したロバストネスの高い講師検出処理を実現した．また，講師近辺の映像をなめらかな動画として保存するための機構を新たに実装し，人間が行うカメラワークに近い講師動画作成を可能とした．本ソフトウェアのソースコードはオープンソースリポジトリである OSCJ.NET(Open Source Collaboration Joint Network)に公開されている．

2. 従来研究と研究の背景

ビデオカメラによって板書映像を記録しようとすると，講師撮影映像の一部として板書を記録する必要があるが，黒板全体をカバーする領域を動画として保存またはインターネット配信しようとするとネットワーク帯域を多く必要とするという問題点がある．このような理由から，講義における板書の自動撮影に関する研究の一環として，複数のカメラで講義を撮影して最も講義状況に適した映像を選択して記録する研究や，固定カメラとズーム制御可能なカメラとを併用し重要箇所に限定して板書を自動ズーム撮影する研究がなされてきた[2,3,4,5]．しかしながら，装置が大掛かりとなるため，どの教室にでも手軽に持ち運んで使えるシステムとすることが難しいという問題や，セットアップに手間がかかるという問題があった．くわえて，講義の場合は，少し前に説明された内容を参照しながら現在の内容を理解するということがしばしばあり，黒板全体を同時に見渡したいという要求が少なからず存在する．このため，複数カメラを切り替えて撮影したり，黒板をパンして撮影したりした映像には不満がでることが予測される．

また，講義映像収録以外の分野において，たとえば，テレビ会議参加者を自動撮影するシステムが提案されている．市販されているテレビ会議製品の多くは 1 台の光学ズーム式カメラを備え，手動で方向や拡大率を制御できるようになっているが，先進的な研究としては，これらを全て自動化しようとする動きがある．しかしながら，光学ズーム式カメラについては，首振り速度やズーム速度が間に合わなくなる恐れが常に存在する．例えば，一旦ズームする箇所を間違ってしまうと，それから復帰するためには多大な時間がかかるという問題がある．

そこで著者らは，過去において，講師近辺の映像を動画として保存すると同時に板書全体を静止画として保存する講義自動収録システムを提案した．1 台の固定ビデオカメラのみで黒板全体を明瞭に記録するためにハイビジョンカメラ(HDV カメラ)を用い，また，板書静止画については，画像処理を施して講師の姿を消去する工夫を施したことが特徴である．ハイビジョンカメラとしては，記録画素を 1440×1080（約 155 万画素）確保できる民生用の SONY「HDR-HC1」[6]を使用した．HDR-HC1 は，HDV 規格（ミニ DV カセットテープにハイビジョン映像の記録再生を可能にするための規格）に準拠したデジタルビデオカメラであり，映像記録フォーマットは MPEG2-TS である．

試作システムは Windows 上に構築された．ユーザ評価において，黒板の死角をなくしていることや，常に黒板全体を見渡せることから，試作システムが作成した板書静止画は見やすいとの結果を得ることができた[1]．また，ハイビジョンカメラで黒板領域全体を動画として保存したファイル容量に比較して，試作システムが作成した全静止画ファイル(1440×1080 ピクセル)と講師近傍動画ファイル（320×240 ピクセル 700Kbps WindowsMedia 形式）の合計容量は平均 1/44 であり，有効に帯域を圧縮できることが確認できた[1]．

しかしながら，この試作システムは，実現可能性を検証するためのプロトタイプとして構築されたものであり，次のような重要な課題を未解決のままとしていた．

1. 黒板領域を撮影の都度手作業で指定する必要があった
2. 講師領域を誤って認識する状況が少なからず発生した
3. 自動作成された講師動画が不自然であり見づらかった

試作システムでは，1に関し，黒板の四隅座標（認識対象領域）をユーザが手作業で入力する必要があった．また2に関し，比較的単純な背景差分法とフレーム間差分法の組み合わせによって講師領域を検出していたために，認識システムが講師を見失ってしまう問題や，講師が殆ど動かなかった場合に講師位置の特定が曖昧になってしまう問題があった．さらに3に関し，自動生成された講師動画は，一定間隔（4秒毎）で，講師位置を中心とした映像に強制的に切り替える方式で作成されていたために，映像に不連続感があり見づらいという問題があった．

3. 実装と実験評価

ChalkTalk-OSSの典型的な使用手順について述べる．講師は，講義開始前に教室の後方にハイビジョンカメラ1台を設置し，黒板全体を撮影できるよう画角を調整して撮影を開始する．講義終了後，撮影に使ったビデオテープを巻き戻し，ハイビジョンカメラとLinux PCとをIEEE1394ケーブルで接続してMPEG2-TS形式でデジタルビデオキャプチャを行う（このビデオキャプチャツールについてもOSCJ.NETで公開している）．そして，Linux PC上に作成された動画ファイルを入力ファイルとしてChalkTalk-OSSを実行する．これだけの操作で，講師の姿を消去して常に黒板の板書が見える状態を保った静止画として板書を保存し，同時に，講師を自動追尾した映像を動画として保存できる．また，これらのコンテンツをストリーミング配信するためのWebページも自動作成されるため，Webサーバにコピーして直ちにコンテンツ配信ができるようになっている．

3.1. 黒板領域自動検出

以前の試作システムでは，黒板の四隅座標（認識対象領域）をユーザが手作業で指定する必要があり手間がかかるという問題があった．そこで今回，黒板領域を自動検出する機能をChalkTalk-OSSに新たに設けることで，撮影準備作業を簡略化することを目指した．

黒板を検出する方法として，黒板の形状と色に着目して検出するように方針を定めた．まず色に関し，RGB表色系から黒板領域の特徴を抽出することを試みた．しかしながら，RGBの各要素の画素値を調べた結果，同じ黒板を撮影したシーンでも，部屋の明るさの変化等によってRGBの値が大きく異なり，その際の法則性がわかりにくいという結果となった．

そこで次に，HSV表色系から黒板の色を特定することを試みた．HSV表色系では，色合いを表す色相:H，色の鮮やかさを表す彩度:S，色の明るさを表す明度:Vで色が表現される．Hの値（色相）は，輝度の変化に依存せずに色を識別できることが特徴であり，今回の実験でも，部屋の明るさがかなり大きく違う環境でもHの値に大きな差は見られなかった．実験の結果，黒板としては，$90 \leqq H < 150$を黒板色と定めることが適当であることがわかった．このHの範囲は12色相環の黄緑から青緑に相当する領域である．

一方，黒板の形状に関しては長方形であることを仮定した．撮影画像の上下の端，または，左右の端からそれぞれ水平，または，垂直方向に画素を走査して，上記HSV系の色特性が黒板である領域を絞り込むという方法によっておこなった．例えば，撮影画像上端（Y座標=0）から走査する場合，Y座標を1つずつ増やしながら，その都度，水平方向に黒板色と認識される色が何ピクセル存在するかを調べる．そして，そのピクセル数が閾値を超えたところを黒板領域の上端と定めた．

3.2. 講師領域自動検出

黒板領域が特定されると，次は，その領域に対し講師領域検出処理が行われる．講師領域は，その後の，講師の姿を消去した板書静止画を作成する処理，および，ハイビジョン映像から講師近傍動画を切り出す処理に使用されるため，本システムにおいては，講師領域を正確に検出することが極めて重要である．しかしながら，以前の試作システムでは，比較的単純な背景差分法とフレーム間差分法の組み合わせによって講師領域を検出していたためノイズ等弱く，講師領域を誤って認識してしまう状況が少なからず発生していた．

ここでは，今回新たに実装した講師領域自動検出方式について述べる．本方式は，1．移動物体領域検出処理，2．板書撮影時に発生するノイズの除去，および，3．講師の胴体と手の関係を考慮した講師領域検出，から構成されている．以下，それぞれについて述べる．

[移動物体領域検出処理]

まず，移動物体領域検出処理の基本となるフレーム間差分法について説明する．フレーム間差分法は，撮影された時間の近い2つの画像で差分を計算する方法であり，画素変化のあった領域を移動物体が

映っている領域と定める方法である．輝度によって2値化した画像が処理対象とされることが多い．ただし，一般的なフレーム間差分法では，動いている物体の領域としては，時刻tにおける位置と，時刻t-1における位置の両方が抽出されてしまうという問題がある．実際に，講師が早く移動したような場合に，時刻tにおける位置と時刻t-1における位置の両方の領域が移動領域として抽出されてしまう不具合が観察された．また，講師の動きが遅い場合にでも，移動物体が部分的に重なっているため，移動物体の輪郭が抽出されにくいという問題点があることが観察された．

そこで今回，時間的に連続した3フレームの画像（輝度によって2値化した画像）に対し，2段階のフレーム間差分法を施すことによって移動物体領域を検出するようにした（本実装において，フレーム間差分抽出処理は8x8ピクセルのマクロブロック毎に行われる）．連続したフレーム画像をf(1), f(2), f(3)とするとき，最初に，f(1)とf(2)の差分画像h(1,2)と，f(2)とf(3)の差分画像h(2,3)とを作成する．次に，差分画像h(1,2)と差分画像h(2,3)の差分画像h(h(1,2), h(2,3))を作成するようにした．そして，この差分画像h(h(1,2), h(2,3))に消えずに存在している画像領域を，画像f(1), f(2), f(3)から得られた移動物体領域として定めた．これによって，時刻tにおける位置と時刻t-1における位置の両方が移動物体として抽出されてしまうという問題や，移動物体の動きが遅い場合に移動物体の輪郭が抽出されにくいという問題を大幅に緩和することができた．

[板書撮影時に発生するノイズの除去]

しかしながら，移動物体領域検出処理を終えた段階で，講師が存在しない領域を移動物体領域と誤認する例が少なからず見受けられた．そこで原因を確認するために，移動物体領域検出処理を1ピクセル毎に施し，移動物体と認識されたピクセルを着色して画像表示する実験を行った．複数の講義収録ビデオを対象に実験を行った結果，黒板にチョークで書かれた文字が移動物体として誤認識されることが多いこと，さらに，同じ文字でも白色チョークで書かれた文字の誤認識率が高いことがわかった．次に，白色文字の誤認識率が高い理由を調べるために，屋外で撮影したビデオや研究室内で撮影したビデオ等，様々なビデオをChalkTalk-OSSに入力して移動物体検出実験を行った．しかしながらこの実験においては，白色の部分の誤認識率が高いと判定されたビデオは少なく，むしろ，そのような傾向は見られない場合がほとんどであった．

そこで，ハイスピードカメラを用いてビデオ映像の分析を行うと共に，講義収録ビデオに立ち返って再実験をすることにした．200フレーム/秒で撮影できるハイスピードカメラを用いて映像を分析した結果，肉眼ではわからなかった蛍光灯によるフリッカが，板書の明度を大きく変化させていることがわかった．実際に，講義収録ビデオに対する再実験の結果，黒板に書かれた白色文字の中でも，黒板灯（黒板を照らす目的で設置された蛍光灯）に近い黒板上部にかかれた白色文字の誤認識率が高いことが確認できた．白色は他の色よりも蛍光灯の光を反射する率が高いため，その影響が顕著に現れ，白色文字が移動物体として誤認識されたと推測できる．

このことを検証するために，白色の文字部分と濃緑色の黒板部分とで，黒板灯のフリッカによる影響度合の違いを比較する実験を行った．講義ビデオ映像を輝度256段階のグレースケール映像に変換して実験したところ，黒板上の同じ場所（黒板灯に比較的近い場所）において，濃緑色の黒板部分ではフリッカの影響による輝度変化は平均9.0であったが，一方，白色の文字部分の輝度変化は平均18.0であり，輝度変化量が平均2倍であることがわかった．板書の視認性をあげるために設置されている黒板灯が，移動物体領域検出処理に悪影響を与えていたと言える．

この実験結果に基づき，システムの改良を行った．移動物体を漏れなく検出するためには，輝度変化の閾値は極力小さい方が好ましいが，一方で，白色文字部分に関して閾値を小さくすると，板書文字が移動物体として誤認識される可能性が高くなる．これらの条件を鑑み，フリッカによる影響を受け易い白色部分（RGB値がほぼ均等に大きい値(210以上)を示す部分）に関しては，他の色の部分に比べて約2倍の輝度変化を許容するようにシステムを変更した．評価実験の結果，白色文字を移動物体と誤認識する率を有意に低下させることができた．

[講師の胴体と手の関係を考慮した講師領域検出]

板書撮影時に発生するノイズの除去が実行された後，講師の胴体と手の位置関係を考慮した講師領域検出処理が行われる．具体的には，移動物体検出処理によって移動体物体領域と特定されたマクロブロックの数（単位時間あたりの度数）を黒板の横軸座標で整理してヒストグラム化し，以下のルールに

従って保存領域を特定するようになっている.

1. 移動物体領域が X 軸方向に最も広い範囲で存在する領域を講師の胴体部分とみなす．胴体部分の中心付近を撮影の中心点とする
2. 講師の胴体部分から左右に 120 ピクセル（15 マクロブロック）以内に存在する狭い移動物体領域は講師の手部分とみなして，胴体部分と領域を繋げて講師領域とし，講師領域の中心付近を撮影の中心点とする
3. X 軸方向に極めて狭い移動物体領域が単独で存在，または，講師の胴体部分から左右に 120 ピクセル以上離れて存在する時，ノイズとみなして講師領域とはしない

図 1に，黒板の横軸座標方向に移動物体領域のマクロブロック数をプロットした実際のデータを示す．これは，黒板の色と比較的似た色の服を着ている講師を撮影した際のデータであるが，図に示されるように，胴体と手の部分の間に移動物体領域が全く存在しない部分があることがわかる．講師が何を指し示したか，または，どのようなジェスチャをしたかは講義の中で非常に重要な情報であり，今回のように，講師の手部分と予測される領域と胴体部分と予測される領域とを繋げて保存することができれば，講義内容がよりわかりやすくなると考えられる.

くわえて，重要箇所を指し示している時や板書をしている時は，必然的に手の方向に顔が向いているため，手と顔（胴体）の中心付近を撮影の中心点とすることで，講師の顔が撮影画像の内側を向いている場面が多くなる．映像撮影や写真撮影のテクニックとして，顔の向いている方向に空間を開けた方が画面に奥行き感がでて見易くなる効果があると言われている．また，撮影したい被写体をあえて中心からずらして撮影する「三分割法」は，良い写真を撮るための基本技法として知られている．このような撮影技法の観点からも，講義中の講師の手と顔の中心付近を撮影の中心点とすることは好ましいと考えられる.

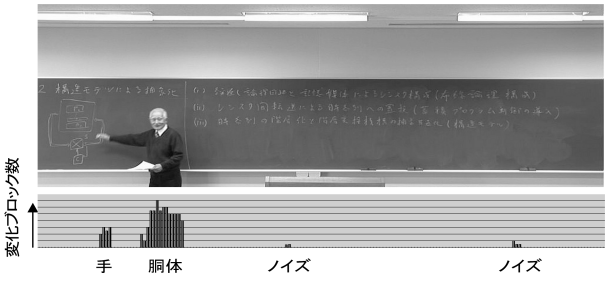


図 1　移動物体検出

3.3.　板書静止画と講師動画の作成

講師領域が特定されると，次に，板書静止画と講師動画の作成が行われる．またこれと同時に，講師領域として特定された領域は，ハイビジョン映像から切り出されて講師動画としてストリーミングビデオ形式で出力される.

[板書静止画からの講師の姿の除去]

特定された講師領域の画像が，講師が存在しなかった時点の画像（講師がその領域に入る直前の画像）によって置き換えられ，あたかもその時点で講師が存在しなかったような画像が作成される．具体的には，講師領域の画像を，講師が存在しなかった時点の画像（本実装の場合は，8 秒前の講師映像除去画像）によって置き換えるようになっている．したがって，作成された板書静止画は，見かけ上は同じ時点に撮影された一枚の画像のように見えるが，実際は，それぞれの部分が異なる時間に生成された画像の貼り合わせであると言える．さらに，板書静止画が保存されるのは板書内容に変更があったと認識された場合のみであり，板書が書き加えられたり消されたりしない間は新たな静止画は保存されない．これによってファイル保存容量を小さくしている.

なお，8 秒というタイミングは，「講師が通常の板書をし，学生に説明して消すまでにかかった時間」の最短時間であり，講義撮影ビデオを観察しこの値を定めた．当然，記憶容量に余裕がある場合には，8 秒より短い間隔で上記動作を繰り返すようにしても構わない.

[自然なカメラワークを有する講師動画の作成]

特定された講師領域は，その後，講師近傍動画を作成するためにも利用される．具体的には，ハイビジョン映像から講師領域部分が切り出され，講師動画としてストリーミングビデオ形式で出力される．しかしながら以前の試作システムでは，一定間隔（4 秒間隔）で講師位置を中心とした映像に強制的に切り替えていたために，映像に不連続感があり見づらいという問題があった.

この問題を解決するため，今回，2 パス方

式で講師動画を作成する方式を導入し，カメラがパンしているような映像効果を加えるようにした．1パス目で，講義ビデオ全体を通しての 1 秒毎の講師の位置をインデックスファイルに書出し，2 パス目で，そのインデックスファイルを参照して，撮影の中心点を線形補完によって滑らかに移動させるようにした．例えば，1 パス目で，t1 時点の講師座標が x1 であり，その 1 秒後の t2 時点の講師座標が x2 であるとする．この時，2 パス目では，t1 から t2 の間の任意時刻 t における撮影の中心点 x は，x1 と x2 の座標を線形補完した座標となり，次式で求まる．

$$x = x1 + (x2-x1)/(t2-t1)\times(t-t1)$$
$$(\text{ただし，} t1 \leqq t < t2)$$

以上のように，任意時点の撮影の中心点を決定するために，最大 1 秒先の講師座標を利用する構成となっている．なお 1 秒という値は，講師が比較的素早く動いた場合に，約 1 秒間で撮影範囲から出てしまうという実験結果から求められた値である．この変更によりカメラワークはスムーズになり，映像に不連続感が有るという問題は解決された．

しかしながら，生成映像を長時間学生に視聴させたところ，「撮影の中心点が常に小刻みに変化する（パンが止まらない）ため，映像に酔ってしまうような感覚がある」という指摘を受けた．そこでさらにシステムを改良し，講師領域の移動が少ない場合には，あえてカメラワークを行わないように変更した．具体的には，講師が左右に 160 ピクセル分以上動いた場合にのみ，1 秒先の講師位置に（パンしながら）撮影の中心点を移動させるようにした．160 ピクセルという値は，講師動画の横幅（320 ピクセル）の半分であり，「講師が画面からはみ出しそうになったらパンをする」という意味合いを有している．

以上の方法によって自動生成した講師動画を被験者 11 名に視聴させ，見やすさに関して 5 段階評価（1:とても見づらい～5:とても見やすい）する実験を行った．具体的には，人が撮影した講師動画 Mh と，ChalkTalk-OSS が生成した講師動画 Mc とを，被験者にはどの動画がどの方法で作成された動画かを知らせずに視聴させた．なお，Mh は，三脚に固定した DV カメラを撮影者（ビデオ撮影に慣れている 30 歳代男性）が手動でパンさせて講師を撮影したものである．

評価の結果，手動撮影 Mh の平均点が 2.7 に対し，自動生成 Mc の平均点が 3.3 であり，自動生成した講義動画の方が優位との結果になった．自由筆記によるアンケートからは，「Mh は細かいカメラ振れや急速なパンが多く，見ていてつかれる」，「常にカメラが動いていて，映像に酔ってしまう」という意見が得られた．実際，Mh と Mc のパンの回数はそれぞれ 38 回と 16 回であった．ChalkTalk-OSS の場合は，講師が次どこに移動するかを把握しているためにカメラワークが最低限の回数で済む．このことが影響している．

4. おわりに

板書を使用する講義の映像から，マルチメディア WBT 教材を自動的に作成する講義収録システム ChalkTalk-OSS を提案した．現状では，講師の胴体部分から左右に 120 ピクセル以内に存在する移動物体領域は講師の手部分とみなして，胴体部分と領域を繋げて講師領域としているが，映像から講師の胴体幅を認識して，繋げるべき幅を自動設定する機能が望まれる．また，なめらかな講師動画を作成するために線形補完によって撮影中心点を定めているが，スプライン補完などの他の方法についても検討したい．くわえて，ChalkTalk-OSS を電子プレゼンテーションの収録にも使いたいというユーザの声もあり，現在検討を進めている段階である．

参考文献

1. 市村,冨野,井上,松下：講師映像と板書静止画の記録が可能な講義自動収録システム，情報処理学会,グループウェアとネットワークサービス研究会報告, GN-56-2, 2005.
2. 板宮,林,千代倉：ワンマン録画可能な講義ビデオ作成システム，情報処理学会コンピュータと教育 研究報告, No.70, pp.17-20, (2003).
3. 大西，村上，福永：状況理解と映像評価に基づく講義の知的自動撮影，電子情報通信学会論文誌 (D-II), vol.J85-D-II, no.4, pp.594-603, 2002.
4. 大西，泉，福永：講義映像における板書領域のブロック分割とその応用,電子情報通信学会論文誌 (D-I)，vol.J83-D-I, no.11, pp.1187-1195, 2000.
5. 宮崎，亀田，美濃：複数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法，電子情報通信学会論文誌 Vol.J82-D-II,No.10,pp.1598-1605, (1999).
6. http://www.sony.jp/products/Consumer/handycam/PRODUCTS/HDR-HC1/