

# モバイル端末における板書講義映像の表示手法の提案

佐藤 雄一<sup>1,a)</sup> 森澤 竜<sup>1</sup> 市村 哲<sup>1</sup>

受付日 2013年3月27日, 採録日 2013年10月9日

**概要:** e-ラーニングの普及や、携帯電話、スマートフォンなどのモバイル端末の流行とともに、モバイルラーニングが注目を集めている。筆者らは、PC 向けの学習用講義映像と板書の静止画を作成する自動講義収録システム ChalkTalk の開発を行ってきた。本稿では、モバイル端末向けの板書講義映像の表示手法として、板書中心のカメラワークと講師のピクチャインピクチャ表示をする手法を提案する。実際に、この表示手法に基づく講義動画を生成するシステムを実装し、講義動画の評価を行った。評価の結果、受講者が見たい板書を重点的に表示しながら講師の様子も分かりやすく表示できることを確認した。

**キーワード:** 講義映像, 板書認識, カメラワーク, ピクチャインピクチャ, モバイルラーニング

## A Proposal of Display Style in Lecture Video for Mobile Devices

YUICHI SATO<sup>1,a)</sup> RYO MORISAWA<sup>1</sup> SATOSHI ICHIMURA<sup>1</sup>

Received: March 27, 2013, Accepted: October 9, 2013

**Abstract:** Mobile-learning has become popular as mobile devices, such as tablet or Smartphone, become widely used. In the past, we had developed ChalkTalk, which automatically produces E-learning materials including a lecturer-tracking video and entire blackboard image from videotaped chalk talks. In this paper we propose a new method for mobile devices to improve the ChalkTalk system. The methods include blackboard centric camerawork and picture-in-picture camerawork. As a result of the evaluation, it was found that the new camerawork can better display the area that users want to see.

**Keywords:** lecture video, blackboard recognition, camerawork, picture-in-picture, mobile-learning

### 1. はじめに

コンピュータやインターネットを利用した学習形態に e-ラーニングがある。受講者は映像や音声、文章などから構成される学習用コンテンツを視聴、閲覧することができる。e-ラーニングには、講師と受講者が 1 カ所に集まる必要がない、自分のペースで学習できるといった利点があり、大学や企業などで多く採用されている。また近年、Android や iPhone などのスマートフォンが流行し、これらのモバイル端末を利用した学習形態であるモバイルラーニングが注目されている。モバイルラーニングでは、e-ラーニングの利点に加え、受講者が自宅や通学中などいつでもどこでも

学習ができるという利点がある。ただしこれらには欠点もあり、学習用コンテンツの作成や配信にコストがかかる点などがあげられる。たとえば、講義映像コンテンツの作成を考えた場合、通常、専属カメラマンによる講義の撮影が必要となる。また、コンテンツを配信するには配信用サーバが必要であり、大容量のコンテンツを配信するには十分な設備も必要である。

学習用コンテンツの自動作成を行うシステムとして、筆者らが過去に開発した ChalkTalk がある [1]。ChalkTalk は、板書を用いた講義を教室後方に設置したハイビジョンビデオカメラで撮影し、その映像から講師の追尾動画（講師付近のみを拡大した動画）と黒板全体の静止画を自動作成する講義収録システムである。作成したコンテンツはデスクトップ PC やノート PC で視聴でき、受講者は動画と静止画で、講師と板書の両方を見ながら学習することがで

<sup>1</sup> 東京工科大学  
School of Computer Science, Tokyo University of Technol-  
ogy, Hachioji, Tokyo 192-0982, Japan

a) yuichi.sato.public@gmail.com

きる。しかし、このような講義映像コンテンツをモバイル端末で視聴しようとした場合、モバイル端末ゆえの問題が発生する。スマートフォンや携帯電話などは画面が小さく、たとえば、動画で黒板全体を映した場合、文字が小さくて読めないことがある。動画で講師付近をズームして映した場合は、板書が十分に映らないこともある。ChalkTalk が生成したコンテンツのような動画と静止画の複合コンテンツの表示を考えた場合もやはり画面が小さく両方を十分に表示することができない。

筆者らは過去に、ChalkTalk が出力したコンテンツのモバイル端末向け様式に関する実験を行った [2]。実験は、横方向にスクロール可能な動画プレイヤーを用いて、受講者が講義映像の中で何をしたいのか調査するものである。実験の結果、以下のような知見を得た。

- 受講者は講師の姿より、板書を重点的に見たい。特に、書いている途中および書かれた直後の板書をよく見たい。
- 受講者は講師が画面上に映っていないとき、講師がどこにいて何をしているのか気になる。

したがって、モバイル端末向けの板書講義映像コンテンツは、講師と板書の両方を表示することが必要である。

この知見に基づき、本稿では、ChalkTalk が作成する講義動画のモバイル端末向け表示手法について検討および提案を行う。提案する表示手法は、講義動画において板書を中心にズームして表示しながら、講師が映らないときはピクチャインピクチャで講師を表示するものである（以後、ピクチャインピクチャのことを PinP と表す）。板書を中心に表示しながら画面端に PinP で講師を表示することで、受講者は講師と板書を確実に見ることができる。本表示手法に基づく講義動画を作成するシステムを実装し、評価実験を行った。評価の結果、本手法は、受講者の見たい箇所を見せることができるということや、板書と講師の両方を表示することができ、受講者の満足度が高いことが示唆された。

## 2. 既存研究

### 2.1 e-ラーニングシステム

モバイル端末での講義受講システムとしては iPod, iPhone とポッドキャストを利用した大学の講義映像視聴システムが存在する [3], [4]。撮影方法としては、専属カメラマンが講師をビデオ撮影するものや、固定カメラや首振りカメラを用いて板書全体を撮影するもの [5], [6], [7] などがある。撮影した映像を iPod 向けにエンコードし、ポッドキャストコンテンツとして Web サーバに配置している。学生は iTunes などのポッドキャスト受信ツールにコンテンツを登録することで、講義映像を自動的に iPod にダウンロードし、通学中などの空いた時間に視聴することができる。しかし、使用するモバイル端末のディスプレイが PC のディ

スプレイに比べてきわめて小さく、そのため、黒板全体を撮影すると板書の文字が小さくなり読みとれないという問題や、講師を中心にズームして撮影すると講師付近の板書しか見えないという問題がある。モバイル端末のディスプレイ面積が小さいため、講師が板書しながら黒板の前を歩き回るような映像を収録するためには、通常、専属カメラマンがビデオカメラを操作し、講師や板書を適宜ズームしながら撮影しなければならなかった。一方、専属カメラマンを導入しない場合には、黒板全体が映るようにビデオカメラを固定設置して撮影し、それをモバイル端末で表示することが考えられるが、板書の文字が小さくなり読み取れないという問題や、講師の表情が分からないという問題が生じる。

Heck ら [8] は e-ラーニング用の板書講義映像コンテンツを生成するシステムを提案している。このシステムが生成する講義映像は、講師を追尾するようにカメラワークを行うもので、講師と板書が重なり一部の板書が見えにくくなったときは、講師の透明化や PinP などの手法を用いて板書が見えるようにしている。また、講師と板書が離れたときは、ズームアウトを行い講師と板書の両方を表示している。モバイル端末で視聴しているときに、ズームアウトを行うとやはり板書の文字が小さくなり読めなくなるという問題が発生する。

### 2.2 ChalkTalk

ChalkTalk はハイビジョン撮影した講義映像から講師の自動追尾動画と黒板全体の静止画アニメーションを出力するツールである。まず、教室の後方にハイビジョンビデオカメラを黒板全体が映るように固定設置し、講義を撮影する。次に撮影した映像を ChalkTalk で処理する。ハイビジョン撮影した講義動画から黒板領域内の移動物体を検出し講師領域とする。その際、移動物体の検出にはフレーム間差分法を用い、移動物体領域として特定されたマクロブロックの数を黒板の横軸座標ごとに集計してヒストグラム化し、以下のルールに従って講師領域を特定しカメラワークを行っている。

- (1) 移動物体領域が横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見なす。
- (2) 講師の胴体部分から左右に特定のピクセル以内に存在する動きの大きい移動物体領域は講師の手部分と見なして、胴体部分と領域をつなげて講師領域とし、講師領域の中心付近を撮影の中心点とする。
- (3) 映像酔いを防ぐため、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ず、講師の胴体や手が画面からはみ出しそうになったらパンをする。
- (4) 人間が行うカメラワークに似せるため、講師が移動したタイミングから意図的に 0.5 秒遅らせてパンをする。これらの処理を経て、講師近傍領域の小さな動画を出力

すると同時に H.264/MPEG4 AVC 技術を用いた動画圧縮を行う。また黒板全体の静止画アニメーションは、講師領域が変化したときに黒板領域の講師がいない部分を更新することにより作成される。元ハイビジョン動画に比べ、これらのコンテンツのサイズは約 1/20 に圧縮されている。また、受講者はデスクトップ PC やノート PC などのブラウザ上でこれらのコンテンツを視聴することができる。

PC などの環境ではディスプレイが大きく、講師の動画と黒板全体の静止画アニメーションの両方を容易に表示することができたが、スマートフォンや携帯電話などのモバイル端末ではディスプレイが小さく、黒板全体を表示するのは適切ではない。したがって、モバイル端末においては、ChalkTalk が出力した動画のみを表示することが考えられるが、この場合、講師付近の板書しか見えないという問題が発生する。

### 2.3 ChalkTalk コンテンツのモバイル向け様式の検討

筆者らは過去に、ChalkTalk コンテンツのモバイル端末向け様式について実験を行った [2]。実験は、図 1 のような横方向にスクロール可能な動画プレイヤーを用いてハイビジョンの講義動画を被験者に見てもらい、受講者が講義動画中の何を見たいのか評価するものである。講義動画の縦方向はスマートフォンの画面いっぱいに表示されるが横方向は画面からはみ出ており、はみ出ている部分を見るには画面を横方向に指でなぞりスクロールさせる必要がある。実験の結果、受講者の傾向について以下の知見を得た。

- 講師の姿より、書かれた板書を重点的に見たい。また、書いている途中および書かれた直後の板書をよく見たい。
- 講師が画面上に映っていないとき、講師がどこにいて何をしているのか気になる。

以上より、本研究において解決すべき課題は以下のとおりである。

**課題 1** 板書講義映像コンテンツは、受講者の見たい箇所

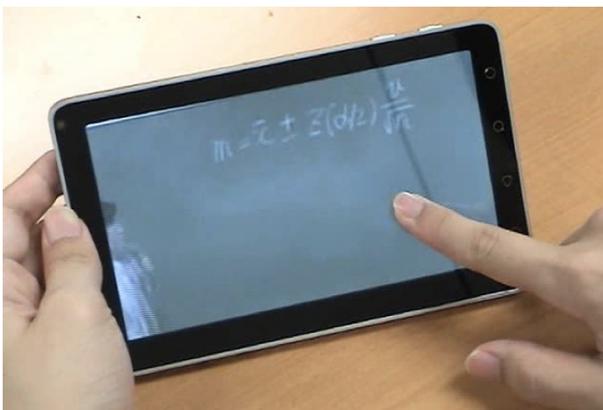


図 1 スクロール操作の様子  
Fig. 1 Scroll action by a user.

を重点的に表示する必要がある。

**課題 2** 受講者の見たい板書に加え、講師の様子も分かるようにし、見やすく表示する必要がある。

### 3. 提案

前章最後にあげた課題を解決するため、講師と板書の両方を十分に表示する手法を提案する。提案手法は、板書を中央に映るようにカメラワークを行いながら、端に PinP で講師を表示するものである。以下にその詳細を示す。

ChalkTalk の講師追尾動画は、講師を中心にカメラワークを行っており、板書を考慮していなかった。本稿では、モバイル端末において動画だけで板書も十分に表示するための板書中心のカメラワークと講師を表示する PinP を提案する [9]。提案する表示手法の概要を以下に示す。

- (1) 講師が板書を書いているときは、講師を中心にズームする。これは ChalkTalk のカメラワークと同じである。
- (2) 講師が板書を書き始めるとき、これから書かれるひとままとりの板書が中央に映るようにカメラをパンし、板書終了後、一定時間経過するまで固定する。
- (3) 一定時間経過し、新たな板書が検出されれば、(2)と同様にその板書が映るようにカメラをパンし一定時間映す。その後、また新たな板書があれば、これを繰り返す。新たな板書が無ければ、(1)のカメラワークに戻る。
- (4) カメラを板書に固定しているときに講師が板書から離れ、画面内に映らなくなった場合、講師付近の様子を PinP で動画内に表示し、一定時間後に (1) のカメラワークに戻る。

(1) は課題 1 に対応し、板書内容に変化がないときは主に先生が映っていることが望ましいと考えられるためである。(2)、(3) は課題 2 に対応し、書いている途中および書かれた直後の板書をよく見たいという要望があったためである。(4) は課題 1 に対応し、講師が画面から消えると、どこで何をしているのか気になるという意見のためである。これらを合わせることで、利用者は講師と板書の両方を見ることができる。

(2) における、ひとままとりの板書とは、講師が黒板上に距離的、および時間的に連続して書いた文字や記号の集合のことである。たとえば、黒板上に書かれた文字列や数式などがひとままとりの板書である。(2) では、このひとままとりの板書領域が画面中央に映るようにカメラをパンする。しかし、板書を書き始めたときには、これから書かれる板書がどこにどれだけ書かれるかは分からない。したがって、あらかじめ、その板書を書き終わるところまで動画を読み込み、板書領域の認識を行っておくようにした。(4) では、板書を中央に映しながら、端に PinP で講師を表示する。PinP の表示位置は画面四隅のうち板書がない

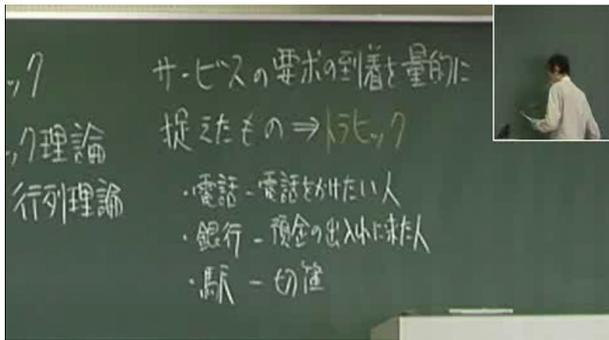


図 2 PinP を用いた講義動画 (コンピュータネットワーク) の例  
 Fig. 2 Example of PinP (computer network).

ところに表示するのが望ましい。図 2 は PinP の例で、画面中央に書き終わった板書を一定時間経過するまで表示しながら、画面右端に PinP で講師の様子を表示するようにした。講義動画において PinP を用いる点は Heck ら [8] の手法と同じであるが、彼らのシステムでは板書が講師と重なり見えにくくなったときに PinP で板書を表示しているのに対し、本システムでは板書と講師が離れたときに PinP で講師を表示する点で異なる。これにより講師が何をしているか確認しつつ板書を中心に見たいという要求に応えられると考えられる。

本研究では、iPhone [13] や Xperia [14] などのディスプレイが 3.5 インチから 5 インチ程度のスマートフォンデバイスを主な対象としているが、これらの小画面上に板書を見やすく表示するには、文字の表示条件を考慮する必要がある。そこで 3 つの講義動画の板書静止画像を用いて必要条件の調査を行った。12 名の被験者に、iPhone および Xperia で板書静止画像を見てもらい、板書文字の読みやすさについて 5 段階の評価してもらうものである。評価 1 は「読みにくい」、評価 5 は「読みやすい」とした。なお、本システムでは講義動画をスマートフォンのディスプレイ全体に拡大して表示するようになっている。そこで、本実験においても講義動画の板書静止画像を iPhone (3.5 インチディスプレイ) および Xperia (4.3 インチディスプレイ) のディスプレイ全体に拡大して表示するようにした。3 つの講義動画から文字の大きさが異なる板書画像を 6 枚ずつ作成し、1 枚の画像につき 6 人の被験者に見てもらった。結果を表 1 に示す。表中の文字の大きさ (mm) は、板書静止画中に現れる文字の横幅の平均である。結果より板書文字の大きさは、平均評価が 4 以上となる 2.0 mm 以上であれば問題がないと判断し、提案システムでは、3.5 インチのディスプレイに文字の大きさがおおむね 2.0 mm 以上となって表示されるように、1920 × 1080 の加工前の動画から 480 × 272 ピクセル切り出して動画を作成できるようにした。

表 1 文字の大きさによる読みやすさの印象評価

Table 1 Impression Evaluation of size of letters in lecture video

文字の大きさ (mm)	評価平均
1.4 以上 1.6 未満	3.00
1.6 以上 1.8 未満	2.83
1.8 以上 2.0 未満	3.67
2.0 以上 2.2 未満	4.83
2.2 以上 2.4 未満	4.17
2.4 以上 2.6 未満	4.67

## 4. 実装

筆者らは、提案表示手法に基づく講義動画収録システムを実装した。これには、まず講師の位置や板書の位置、講師の板書の書き始め、書き終わりタイミングの認識が必要となった。

板書の位置に関しては、背景モデルを用いる手法 [10] や、時空間画像からエッジ抽出を行い分析を行う手法 [11] がある。前者は背景差分法、後者はフレーム間差分法にそれぞれ基づく手法である。講師の板書の書き始め、書き終わりタイミングの認識については、フレーム間差分法やチョーク音を用いる手法 [12] がある。

今回の実装では、講師の位置、板書の位置、板書の書き始め、書き終わりの認識について、時空間画像からエッジ抽出を行い分析を行う手法を用いた。時空間画像やエッジ抽出を用いる既存研究 [11] と類似しているが、エッジ抽出の結果を利用して、板書の書き始めと書き終わりの認識を行う点が新しい。また講義動画作成時にカメラワークを決定するには、3 章で述べたように事前に板書位置の認識が必要である。そこで、動画作成前に 1 度、元のハイビジョン動画を読み込んで、板書位置、板書の書き始め、書き終わりなど必要な情報の認識を行うようにした。

### 4.1 時空間画像

認識アルゴリズムの詳細を述べる。本稿における時空間画像とは、図 3 のように動画の各フレーム画像を時系列順に並べた断面画像のことである。図中の  $x-t$  で示されたある  $y$  座標における断面画像を  $x-t$  時空間画像、 $y-t$  で示されたある  $x$  座標における断面画像を  $y-t$  時空間画像とする。図 4 と図 5 はそれぞれ図 3 をもとにした  $x-t$  時空間画像と  $y-t$  時空間画像の例である。図 4 の横方向は図 3 における  $x$  方向、縦方向は図 3 における時間方向  $t$  を表している。図 5 の横方向は図 3 における時間方向  $t$ 、縦方向は図 3 における  $y$  方向を表している。

また、時空間画像の作成においては、長い動画になると多大な記憶領域や処理時間が必要となる。そのため今回の実装では、30 fps の動画を想定し、10 フレームにつき 1 フレーム (0.33 秒につき 1 フレーム) ずつ抽出し時空間画像を作成することとした。

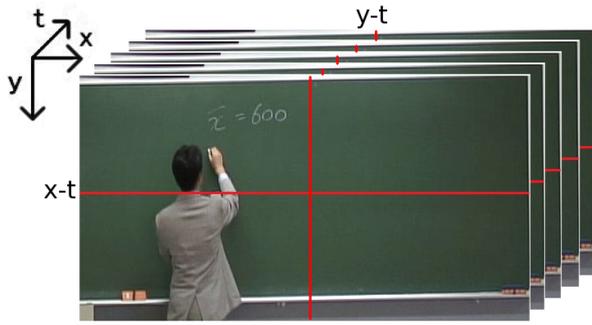


図 3 時空間画像の例 (統計学)

Fig. 3 Example of spatiotemporal image (statistics).

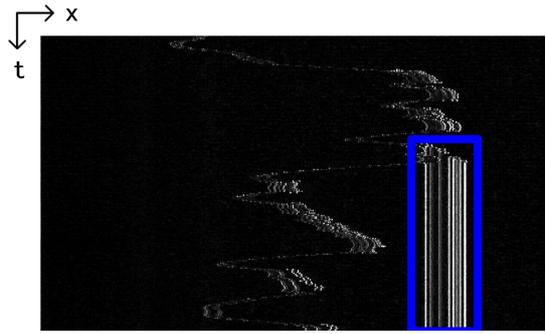


図 6 x-t 静エッジ画像 (統計学)

Fig. 6 Example of x-t static edge image (statistics).

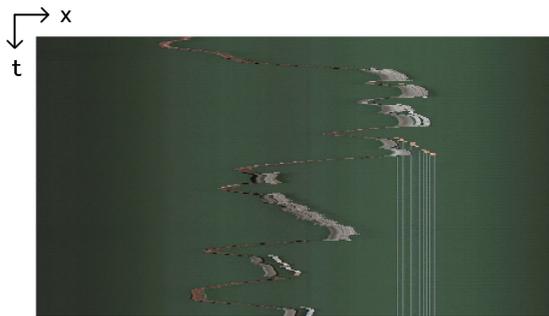


図 4 x-t 時空間画像の例

Fig. 4 Example of x-t spatiotemporal image.

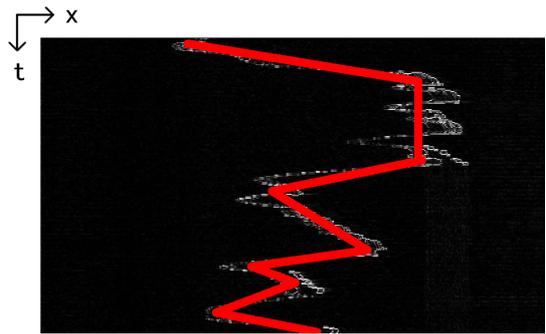


図 7 x-t 動エッジ画像 (統計学)

Fig. 7 Example of x-t dynamic image (statistics).



図 5 y-t 時空間画像の例

Fig. 5 Example of y-t spatiotemporal image.

#### 4.2 エッジ抽出

時空間画像の作成後、256 階調グレイスケールに変換し、エッジ抽出を行う。エッジ抽出は縦と横の 2 種類の Sobel オペレータ [11] を用いて行う。以下に示す  $s_x$  が縦方向のエッジを、 $s_y$  が横方向のエッジをそれぞれ抽出する Sobel オペレータである。

$$s_x : \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad s_y : \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}$$

時空間画像において、縦横 2 種類のエッジ抽出オペレータを用いるのは、静止物体 (板書) と動体 (講師) を認識するためである。x-t 時空間画像において縦方向のエッジを抽出すると、x 方向における静止物体が濃く現れ、横方向のエッジを抽出すると、x 方向における動体が濃く現れる。以後これらを x-t 静エッジ画像、x-t 動エッジ画像とそれぞれ表す。y-t 時空間画像においては縦方向のエッジを抽出すると、y 方向における動体が濃く現れ、横方向のエッジを抽出すると、y 方向における静止物体が濃く現れ

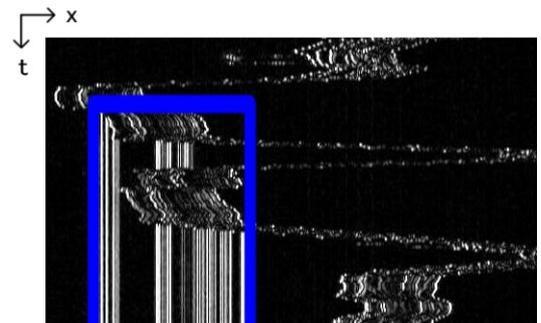


図 8 x-t 静エッジ画像 (コンピュータネットワーク)

Fig. 8 Example of x-t static edge image (computer network).

る。以後これらを y-t 動エッジ画像、y-t 静エッジ画像とそれぞれ表す。図 6、図 7 は、それぞれ図 4 から作成した x-t 静エッジ画像、x-t 動エッジ画像である。図 6 の右部分には、縦方向にまっすぐ伸びるエッジが濃く現れている。図中では、青色の矩形で示された領域である。これが板書の文字の一部を表している。図 7 では、図 6 で現れていた縦方向にまっすぐ伸びるエッジが消え、講師位置と思われるエッジが残っており、濃く現れている。図中では、赤色の軌跡で示された部分である。y-t 時空間画像からも同様のエッジ画像が作成でき、これらのエッジ画像をもとに講師の位置、板書の位置、板書の書き始め、書き終わりタイミングを認識する。同様に、図 8、図 9 は図 2 の講義動画から作成した x-t 静エッジ画像、x-t 動エッジ画像である。



図 9 x-t 動エッジ画像 (コンピュータネットワーク)

Fig. 9 Example of x-t dynamic image (computer network).

### 4.3 講師位置の認識

講師位置は、図 7 のような、x-t 動エッジ画像を用いて認識する。x-t 時空間画像は、動画の縦の長さの数だけ作成されるが、作成時に y 座標を指定し、その y 座標の x-t 時空間画像 1 つだけを用いる。これは、経験的に 1 つだけで十分に講師位置を認識できたためである。また y-t 動エッジ画像を用いないのは、講師は縦方向には移動せず、縦方向の動きを見る必要がないためである。x-t 動エッジ画像においては、講師はある時間 t において濃いエッジが多く出ている x 座標付近にいると考えられる。これより、以下のように各時間における講師位置を決定する。

- (1) すべての x 座標について、その x 座標とその左右各 5 ピクセルのエッジ量の合計値を計算する。
- (2) 最大の合計値を探し、その値が閾値 (256 階調グレイスケールで 1800) 以上ならば、その x 座標を講師位置とする。
- (3) 閾値以下ならば、前後の時間から線形補間する。

(2), (3) で閾値を用いるのは、講師が動いていない場合エッジ量の合計値が小さくなる傾向があり、他に偶発的に動画内に入る受講者の頭などが講師位置として認識されることがあったためである。閾値を用いることで、講師の移動のようなより大きな動きのみを認識することができる。また閾値以下のときは、講師はほとんど動いておらず、線形補間で問題ないと考えられる。

### 4.4 板書位置の認識

板書位置は、静エッジ画像を用いて認識する。まず x-t 静エッジ画像と y-t 静エッジ画像を組み合わせ、静エッジ直方体を作成する。x-t 静エッジ画像の各ピクセル値  $f_x(x, t)$ 、y-t 静エッジ画像の各ピクセル値  $f_y(y, t)$  から、以下の値を計算する。

$$f(x, y, t) = \sqrt{f_x(x, t)^2 + f_y(y, t)^2}$$

この  $f(x, y, t)$  を 2 値化したものが静エッジ直方体である。本実装では閾値を 120 に定めて 2 値化を行った。図 10, 図 11 は、静エッジ直方体の例である。これは静止物体を

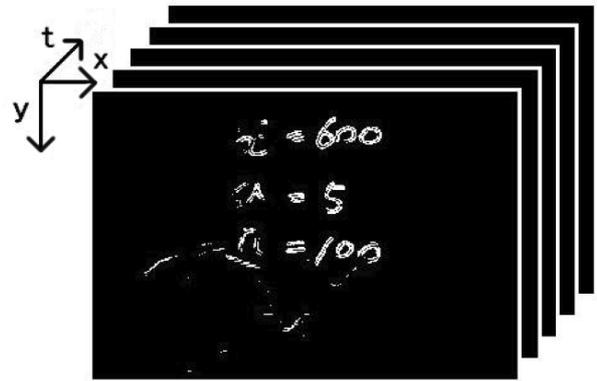


図 10 静エッジ直方体の例 (統計学)

Fig. 10 Example of static edge cuboid (statistics).

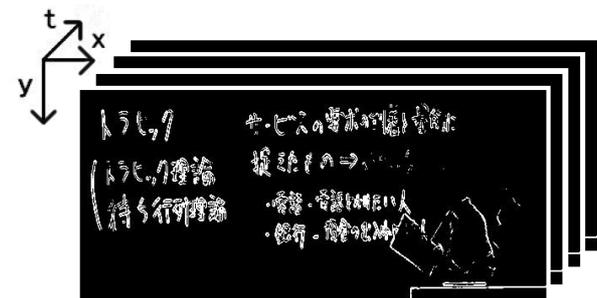


図 11 静エッジ直方体の例 (コンピュータネットワーク)

Fig. 11 Example of static edge cuboid (computer network).

エッジとして抽出した画像を時系列順に並べたもの、すなわち静止物体のみを表示する動画のようなものである。図 10 では、板書文字がエッジとして現れているのが分かる。板書の下には講師の肩と腕の輪郭と思われるエッジも現れている。時系列的に次の画像を見ると、多くの場合、板書文字エッジの位置はそのまま、講師の肩と腕のエッジの位置は移動すると考えられる。したがって、板書文字を認識するには、時系列的に見て一定時間同じ位置に存在するエッジを探せばよい。以下に、板書位置の認識手順を示す。

- (1) 静エッジ直方体のすべて xy 座標について、t 方向にエッジを走査し、エッジが 70% 以上存在する長さ 60 以上 (時空間画像は 0.33 秒につき 1 枚作成されているため、すなわち 20 秒分以上に相当) の区間を探す。区間は x 座標, y 座標, エッジ開始時間, エッジ終了時間で表せ、静エッジ直方体内の t 方向の線分と見なせる。以後、これをエッジ線分と表す。
- (2) (1) で求めたエッジ線分のうち近くにあるものを連結し仮板書とする。連結するエッジ線分はエッジ開始時間が 21 以内 (すなわち 7 秒分に相当)、かつ、xy 座標におけるユークリッド距離が 40 ピクセル以下 (黒板の横幅 7.5 m、動画中黒板サイズが 1,719 ピクセルの環境で、おおよそ 17.4 cm 以下に相当) のものである。また、3 つ以上連結する際には、それぞれのエッジ線

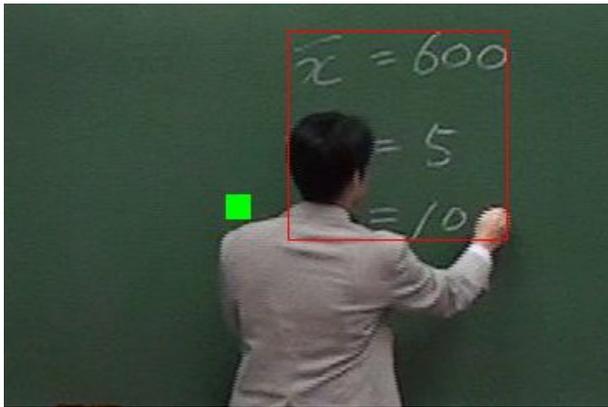


図 12 板書領域，講師位置の認識例（統計学）  
Fig. 12 Example of recognition (statistics).

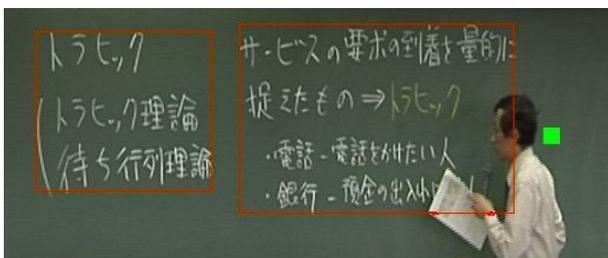


図 13 板書領域，講師位置の認識例（コンピュータネットワーク）  
Fig. 13 Example of recognition (computer network).

分がお互いに近い必要はなく，どれか 2 つのエッジ線分が条件を満たせば連結する

(3) (2) で求めた各仮板書の領域を求める．連結したエッジ線分のうち， $xy$  座標において左端と右端にあるエッジ線分の  $x$  座標をそれぞれ板書領域の左端と右端，上端と下端にあるエッジ線分の  $y$  座標をそれぞれ板書領域の上端と下端とする．

(4) (3) で求めた仮板書の領域の面積が 1,000 ピクセル以上（黒板の横幅 7.5m，動画中黒板サイズが 1,719 ピクセルの環境で，おおよそ 4363.0  $\text{cm}^2$  以上に相当）のものを真の板書とする．

(1) でエッジが一定割合以上存在する区間を探しているのは，板書が講師の体によって隠れることがあり，途中隠れることがあっても板書を構成する文字を認識できるようにするためである．(2) によって距離的に時間的に近いエッジ線分群は連結され 1 つの仮板書となり，遠いエッジ線分群は，別の仮板書として連結される．(4) で板書領域の大きさを調べ，小さいものはノイズとして板書と認めない．またこれらの手順での具体的な数値は実際の講義動画を使って実験し決定した．図 12，図 13，図 14 は，提案システムが認識した板書領域と講師位置の例である．図中の四角の枠と四角の点が，それぞれ認識された板書領域と講師位置を表している．図 14 のように，指数や小数点のような小さい文字でも，周りの文字と一緒に認識できる．



図 14 板書領域，講師位置の認識例（情報理論）  
Fig. 14 Example of recognition (information theory).

#### 4.5 板書タイミングの認識

板書の書き始め，書き終わりタイミングは，板書位置の認識で用いたエッジ線分を用いて認識する．板書位置の認識においては，板書は複数のエッジ線分から構成され，各エッジ線分を表す情報の 1 つにエッジ開始時間があった．これはエッジが出始めた時間，すなわち，講師がそのエッジ線分の  $xy$  座標に板書しそれが初めてカメラに映った時間である．そこで，板書を構成するエッジ線分のエッジ開始時間で，最も早いものを板書の書き始め，最も遅いものを板書の書き終わりタイミングとした．

#### 5. 評価

提案した表示手法に基づく講義動画を作成するシステムを実装した．このシステムが作成する講義動画が，本研究の課題を解決できているか確かめるため 2 種類の評価実験を行った．実験の目的は，以下のとおりである．

**実験目的 1** 提案カメラワークを用いた講義動画が受講者の見たい箇所を表示できているか確認する．

**実験目的 2** 板書と講師の同時表示による講義動画の見やすさを評価する．

実験目的 1 は課題 1 に，実験目的 2 は課題 2 にそれぞれ対応している．以後，ChalkTalk のことを旧 ChalkTalk，今回実装したシステムのことを新 ChalkTalk とそれぞれ表すことにする．また ChalkTalk で変換する前の動画である元のハイビジョンの動画のことを元動画と表す．各実験に用いた，元動画の詳細を表 2 に示す．統計学および物理学は数式主体，プログラミングは文章主体の講義である．図 3，4，5，6，7，8，10，12 は実験対象となった統計学講義の図である．一方，図 2，9，11，13 はコンピュータネットワークの図である．コンピュータネットワークは本実験の実験対象ではないが，実験対象となったプログラミングと同様に文章主体の講義である．これらの図は，数式主体の講義と文章主体の講義の両方において，正しく板書を認識可能であることを示している．実黒板サイズは実際の黒板の横幅，動画中黒板サイズは動画で見たときの黒板の横幅を表している．

表 2 実験に用いた元の講義動画

Table 2 Original HD lecture video used in experiment.

	サイズ [ピクセル]	長さ [sec]	講義内容	実黑板サイズ	動画中黑板サイズ
動画 1	1920 × 1080	180	プログラミング	7.5 m	1719 ピクセル
動画 2	1920 × 1080	180	統計学	7.5 m	1685 ピクセル
動画 3	1920 × 1080	180	物理学	4.5 m	1596 ピクセル

表 3 被験者の手動カメラワークの分析

Table 3 Anlyzation of users' manual camerawork.

被験者	元動画	パンニング回数	特徴
A	動画 1	7	パンニング回数が少なく、板書が映るように停止している時間が長い
B	動画 1	11	被験者 A と同様
C	動画 1	11	被験者 A と同様 加えて、過去の板書を見直すようなパンニングが見られた
D	動画 1	16	板書が映るように停止している時間が長い. 2 つの板書を交互に見比べるようなパンニングが複数回見られた
E	動画 2	46	板書を主にしながら講師に合わせて少しずらすようなパンニング, および, 過去の板書と現在書いている途中の板書を見比べるようなパンニングが複数回見られた
F	動画 2	22	板書を主にしながら講師に合わせて少しずらすようなパンニング, および, 過去の板書を見直すようなパンニングが見られた
G	動画 2	20	講師の動きに合わせたパンニングが多く見られた
H	動画 3	29	板書中および板書した直後は板書を, それ以外の身振りでの説明中のときは講師を見るようなパンニングが見られた
I	動画 3	18	被験者 H と同様
J	動画 3	16	板書見ており講師が映っていない時間が長い

### 5.1 実験 1

実験 1 は実験目的 1 に対応しており, 新 ChalkTalk の講義動画が講義動画中における被験者が見たい箇所を表示できているか評価する実験である. まず被験者が見たい箇所を調査するため, 被験者に過去の実験で用いた横方向にスクロール可能な動画プレイヤーと同様のソフトウェアを用いて, 被験者に元動画を視聴してもらった. このときのスクロールの履歴は自動的に保存されており, 各フレームにおいて被験者が見ていた箇所を分析できる. 次に動画の各フレームにおいて, 新および旧 ChalkTalk が作成した講義動画の自動カメラワーク位置と, スクロール履歴から分析した被験者の手動カメラワーク位置との距離を求め, 全フレームにおける平均値を求めた. この距離の平均値が小さいほど被験者が見たい箇所を表示できていると考えられる.

#### 5.1.1 実験条件

実験の被験者は大学生 10 人である. 実験に用いた元の講義動画は, 表 2 で示した 3 種類の動画である. 動画プレイヤーおよび新旧 ChalkTalk が作成する講義動画のサイズは, とともに 480 × 272 ピクセルとした. 実験機器の制約上, 本実験は一般的な PC モニタを用いて実施したが, スマートフォンで表示した場合と差異が生じないように, 講義動画を iPhone4S のディスプレイサイズと同じになるように拡大して PC モニタに表示し被験者に示した.

#### 5.1.2 実験結果

被験者の手動カメラワークを分析した結果を表 3 に示

表 4 各被験者の手動カメラワーク位置と新旧 ChalkTalk 動画の自動カメラワーク位置の平均距離 (単位: ピクセル)

Table 4 Difference between users' camerawork and system-generated camerawork.

	新 ChalkTalk	旧 ChalkTalk
平均	93.69	158.81
標準偏差	41.43	82.94

す. いずれの被験者も板書と講師両方を同時に見ようとするパンニングが見られた. 加えて, 過去の板書を見直すようなパンニングも被験者 C, D, E, F の 4 人で見られた. また, 各被験者の手動カメラワーク位置と講師動画の自動カメラワーク位置の平均距離を表 4 に示す. 旧 ChalkTalk の講義動画に比べ, 新 ChalkTalk 講義動画の自動カメラワーク位置は, いずれも被験者の手動カメラワーク位置により近くなった. また平均距離の平均について, 片側有意水準 5% の t 検定を行ったところ有意差が認められた. これは新 ChalkTalk の方が人間のカメラワークに近いことを意味し, よって, 新 ChalkTalk は, より人間にとって見やすい講義動画が作成できると考えられる.

### 5.2 実験 2

実験 2 は実験目的 2 に対応しており, 板書と講師の同時表示による講義動画の見やすさを評価する実験である. まず被験者に新旧 ChalkTalk が作成した講義動画の 5 段階印象評価を行ってもらった. 評価項目は「勉強を行うのにこ



図 15 実験 2 の様子

Fig. 15 A scene of Experiment2.

表 5 印象評価の結果

Table 5 Result of impression evaluation.

	平均	分散
旧 ChalkTalk	2.11	1.11
新 ChalkTalk (PinP あり)	3.89	0.61
新 ChalkTalk (PinP なし)	3.33	0.75

の講義動画は「見やすいか」で、5が見やすい、1が見にくいとした。印象評価のあと、動画の見やすさについてどう感じたかインタビューを行った。また評価してもらう講義動画は新旧 ChalkTalk の 2 種類に、新 ChalkTalk の講義動画の PinP なしのものを作成しこれを加えて 3 種類とした。

### 5.2.1 実験条件

実験の被験者は大学生 9 人である。9 人それぞれに旧 ChalkTalk と新 ChalkTalk (PinP あり) と新 ChalkTalk (PinP なし) の 3 種類の講義動画を視聴して評価を行った。視聴する講義動画の順番や元動画の組合せは、極力異なるようにした。実験に用いた元の講義動画は、表 2 で示した 3 種類の動画である。講義動画のサイズは 480 ピクセル × 272 ピクセルで、新 ChalkTalk の講義動画における PinP のサイズは 50 ピクセル × 50 ピクセルとした。また、新 ChalkTalk のカメラワークにおいて各板書を表示する時間は 10 秒とした。講義動画は 3.5 インチディスプレイの iPhone4S でディスプレイ全体に拡大して再生した (図 15)。

### 5.2.2 実験結果

実験を行った結果、各講義動画の評価平均および分散は表 5 のようになった。この結果について、両側有意水準 5% でマンホイットニーの U 検定を行ったところ、旧 ChalkTalk と新 ChalkTalk (PinP あり)、旧 ChalkTalk と新 ChalkTalk (PinP なし) で、有意差が認められた。このため、板書重視カメラワークの講義動画は、講師追尾動画と比べ見やすいといえる。PinP ありとなしの差について有意差は認められなかった。

表 6 旧 ChalkTalk 講義動画に関するインタビュー結果

Table 6 Answer to interview about old ChalkTalk.

件数	内容
6	講師を追尾する講義動画で板書が見にくい
1	講師をちゃんと追いかけていて良かった

表 7 新 ChalkTalk 講義動画 (PinP あり) に関するインタビュー結果

Table 7 Answer to interview about new ChalkTalk (with PinP)

件数	内容
6	PinP で講師が映っており見やすかった
2	PinP で板書が見えにくくなるがあった
1	講師が見えていなくても良い

表 8 新 ChalkTalk 講義動画 (PinP なし) に関するインタビュー結果

Table 8 Answer to interview about new ChalkTalk (without PinP).

件数	内容
4	講師が見えないのが気になる
2	板書の出現が時間的に連続した場合のカメラ遷移が遅れている・速すぎる

インタビューの結果を表 6、表 7、表 8 に示す。それぞれ、旧 ChalkTalk、新 ChalkTalk (PinP あり)、新 ChalkTalk (PinP なし) の講義動画に対するインタビュー結果をまとめたものである。旧 ChalkTalk 講義動画に対しては、講師のみを表示するだけで板書が見にくいという否定的な意見が多い一方で、講師が見られてよかったという肯定的な意見もあった。新 ChalkTalk (PinP あり) 講義動画については、PinP で講師が映っていることに対して肯定的な意見が多く、講師を見たいという被験者の欲求に応えられていると考えられる。また PinP によって板書が見にくかったという問題の指摘や、板書さえ見えていればよいという否定的な意見もあった。PinP なしの講義動画に対しては、PinP ありのものとは比べ、やはり講師が見えないことを指摘する意見が多く見られた。また板書重視のカメラワークについてカメラ遷移の問題を指摘する意見もあった。階印象評価においては PinP ありとなしで有意差は認められなかったにもかかわらず、このような意見が新 ChalkTalk (PinP あり) のときに出なかったのは、板書を中心に表示し講師が映っていないときでも PinP によって講師の様子が分かり被験者を安心させられたためではないかと考えられる。新 ChalkTalk (PinP あり) の印象評価の平均評価が 3.89 と高水準であること、また、インタビュー結果も満足な意見が多いことから、板書の見やすさ、読みやすさが十分に向上したといえる。

## 6. 考察

### 6.1 実験結果について

課題 1 について評価した実験 1 の結果より、提案したカ

メラワークは被験者の見たい板書により近づけられたと考えられる。多くの被験者が、板書と講師が両方映せるときは両方が映るように手動カメラワーク位置を調整しており板書が画面端の方になることがたびたびあった。実装した板書中心のカメラワークでは、板書をカメラの中心に表示し PinP で講師を同時に表示するため、板書が端になることは少ない。したがって、板書は十分表示できていると考えられる。

課題 2 について評価した実験 2 の結果より、板書を中心にカメラワークを行いながら、PinP で講師の様子を表示する手法は被験者から高い評価を得られた。特に PinP で講師が映っていて見やすかったという意見が多く、一定の効果があつたと考えられる。また板書文字の大きさについては、事前の調査により、またインタビューの結果不満の意見はなかったため、問題ないと考えられる。

## 6.2 今後の課題

インタビューの結果から板書の表示に関する問題点が見つかった。インタビューで得られた意見の「PinP で板書が見えにくくなることもある」というものである。提案したカメラワークでは PinP の表示位置を考慮していなかったためである。そのため、PinP の表示位置と板書位置を考慮したカメラ位置の決定を行えば解決すると考えられる。

本研究では、「ひとまとまりの板書」ごとに表示する手法を提案した。本実装では、1つの「ひとまとまりの板書」を表示する際、講師が書き終わってから一定時間（10 秒間）表示するようにした。今回の実験では板書は左から右へあるいは右から左へ順番に行われており問題は生じなかったが、黒板上の離れた箇所に短時間に交互に板書が行われたときに、同じ箇所の書き込みが別々の板書として認識されると、画面遷移が遅れて実行され、次の板書が表示されるまでに時間がかかる場合があると考えられる。多くの板書スタイルに対応できると考えるが、今後はひとまとまりの板書の量に応じて板書を表示する時間を変更することなどを検討したい。

## 7. おわりに

本稿では、板書講義映像コンテンツにおけるモバイル端末向けの表示手法を検討および提案し、その実装方法について論じた。実際に実装を行い提案した表示手法の評価を行った結果、板書と講師の両方を十分に表示できおり、被験者からは高い評価を得られた。また一方で、板書の表示に関してカメラワーク上の問題点も見つかった。今後は板書を受講者にとって、より見やすく表示することが課題である。

## 参考文献

- [1] 市村 哲, 井上亮文, 宇田隆哉, 伊藤雅仁, 田胡和哉, 松下温: ChalkTalk: 講師動画と板書静止画の同時記録が可能講義自動収録システム, 情報処理学会論文誌, Vol.47, No.3, pp.924–931 (2006).
- [2] 市村 哲, 中村亮太, 梶並知記: 講師動画と板書画像を含むモバイル学習用コンテンツ様式の検討, 情報処理学会 GN ワークショップ 2011, pp.41–46 (2011).
- [3] Chandra, S.: Lecture Video Capture for the Masses, *Proc. ACM IT&CSE'07*, pp.276–280 (2007).
- [4] Hurst, W., Welte, M. and Jung, S.: An Evaluation of the Mobile Usage of e-lecture Podcasts, *Proc. ACM Mobility'07*, pp.16–23 (2007).
- [5] 大西正輝, 村上昌史, 福永邦雄: 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).
- [6] Liu, Q., Rui, Y., Gupta, A. and Cadiz, J.: Automating Camera Management for Lecture Room Environments, *Proc. ACM Conference on Human Factors in Computing System (CHI 2001)*, Vol.3, pp.442–449 (2001).
- [7] Zhang, C., Crawford, J., Rui, Y. and He, L.: An Automated End-to-End Lecture Capturing and Broadcasting System, *Proc. ACM Multimedia 05*, pp.808–809 (2005).
- [8] Heck, R., Wallick, M. and Gleicher, M.: Virtual Videography, *ACM Trans. Multimedia Comput. Commun.*, Vol.3, No.1, Article 4 (Feb. 2007).
- [9] 佐藤雄一, 森澤 竜, 市村 哲: 板書講義映像コンテンツにおけるモバイル環境向けカメラワークの提案, 情報処理学会 GN 研究会報告, GN-87, pp.1–8 (2013).
- [10] 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, 電子情報通信学会技術報告, PRMU, Vol.100, No.701, pp.79–86 (2001).
- [11] 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, 電子情報通信学会論文誌 (D-I), Vol.J83-D-1, No.11, pp.1187–1195 (2000).
- [12] 芦川 平, 菅沼 明, 谷口倫一郎: チョーク音による板書動作の開始と終了の判定, 情報技術レターズ 1, pp.245–246 (2002).
- [13] アップル iPhone: 入手先 (<http://www.apple.com/jp/iphone/>) (2013).
- [14] ソニーモバイルコミュニケーションズ Xperia: 入手先 (<http://www.sonymobile.co.jp/xperia/>) (2013).



佐藤 雄一

1989 年生。2013 年東京工科大学コンピュータサイエンス学部卒業。同年東京工業大学大学院進学。教育工学、自然言語処理の研究に従事。



森澤 竜

1990 年生。2013 年東京工科大学コンピュータサイエンス学部卒業。同年シグマトロン (株) 入社。在学中、教育工学の研究に従事。



市村 哲 (正会員)

1966年生. 1989年慶応義塾大学工学部計測工学科卒業. 1994年同大学大学院理工学研究科博士後期課程修了. 博士(工学). 同年富士ゼロックス(株)入社. 1997~1999年富士ゼロックスパロアルト研究所駐在. 2002

年4月より東京工科大学助教授, 2011年4月より同大学教授, 現在に至る. グループウェア, ネットワークサービス, 生体情報活用等の研究に従事. 『IT TEXT 基礎 Web 技術』, 『IT TEXT 応用 Web 技術』(オーム社). DICOMO 2003 優秀論文賞受賞. ACM 会員.