テレビ会議自動撮影のための全方位カメラを用いた撮影機器設定手法

市村 哲[†] 福井 登志也[†]
井上 亮文[†] 星 徹[†]

従来の 1 拠点多人数型テレビ会議システムは，表示サイズや表示解像度が限られているため「誰が発言者か分かりにくい」または「発言者の表情をとらえにくい」などの問題点があった．著者らは，過去の研究において，会議室内の参加者の中から発言者を自動的に検出し，検出した発言者を拡大表示して遠隔地に伝送することが可能なテレビ会議システムを構築し，通常の会議室環境において 10 名程度の会議参加者から発言者の方向を検出して拡大表示することを可能とした．マイクロフォンアレイを用いて発言者方向の推定を行い，2 つの DV カメラから得られた映像から発言者を検出して拡大表示する機能を提供した．しかしながら，過去に開発したシステムでは，DV カメラ映像上の発言者の位置とマイクロフォンアレイの検出する音声到来方向の対応付け作業（キャリブレーション）をすべて手作業で行わなければならず，正確さにかける問題や，時間と手間がかかるという問題があった．そこで本論文では，全方位カメラの利用により短時間で正確にキャリブレーションを行える方法についての提案を行う．評価実験の結果，従来 2 名以上の作業者によって行わなければならなかったキャリブレーション作業が，1 名のみで迅速かつ正確に行えることが確認できた．

A Setup Method for Automatic TV-conference Recording Using Omnidirectional Camera

Satoshi Ichimura,[†] Toshiya Fukui,[†] Akifumi Inoue[†] and Toru Hoshi[†]

Due to the limitation of display size or resolution of traditional TV conference systems, we can not see detailed expression on speaker's face shown on a TV screen, or can not even see who is speaking from the remote site. Last year, we developed a TV conference system where a speaker was automatically identified and his/her face was zoomed in on the TV screen, so that remote participants could better read speaker's facial expression. Speakers were identified with a microphone array, and speakers' images were extracted from two fixed DV cameras. When using the system, however, system administrators had to manually perform a burdensome camera calibration task prior to the conferences. In the task, image positions on the two DV cameras and angles from the microphone array must be precisely corresponded. In this paper, we propose a camera calibration mechanism using omnidirectional cameras. After introducing the mechanism, one person can easily and accurately perform the calibration task that formerly required more than two people.

1. はじめに

遠隔地への会議出席は，多大な交通費や宿泊費が必要となる，移動時間が無駄である，不在時に業務が停滞するなど，仕事の効率面において多くの問題がある[1)]．さらに，企業の多国籍化が進むなか，安全確保やリスク回避の面から役員やエグゼクティブの国外出張の回数を減らしたいという社会的な要請も存在している[2)]．これらのことから，近年，テレビ会議システムを導入する例が増えており，テレビ会議室どうしを接続して，役員会議，意志決定会議，報告会，レビュー会などを行うようなことが多数企業において日常的に行われている[3)]．従来は ISDN 網や専用通信網を用いて構築されることの多かったテレビ会議システムであるが，現在ではその多くがインターネットを介して通信できるようになっており，テレビ会議サーバ機能をネットワークサービスとして契約者に提供する ASP（Application Service Provider）企業も増えてきている[4)]．

テレビ会議システムの形態は，Web カメラとヘッドセットを備えた個人用パソコンを用いる「1 拠点単独参加型」と，会議室にテレビ会議専用のカメラ，テレビモニタ，通信装置などを設置して用いる「1 拠点

† 東京工科大学
Tokyo University of Technology

多人数参加型」とに大別できる[1]が，本論文では，主にエグゼクティブや役員による会議を支援することを狙いとした1拠点多人数参加型テレビ会議のための提案を行う．1拠点に参加できる会議参加者数とテレビ会議システム形態の呼称との関係について明確なコンセンサスはいまだ存在していないが，現在グローバルマーケットにおいてテレビ会議システム販売台数シェア2位（売上高シェア1位）のTANDBERG社が，エグゼクティブ会議および役員会議向け商品として1拠点に8名から12名が参加できる商品をラインナップしていること[3]や，他メーカも1拠点10名まで参加できるテレビ会議システムを会議室用商品として開発していること[5]などから，本研究においても，1拠点10名程度が参加できるテレビ会議を支援対象としている．また，便宜上，本論文中においては1拠点10名程度が参加できるテレビ会議を「多人数参加型テレビ会議」または「1拠点多人数参加型テレビ会議」と呼ぶこととする．

1拠点多人数参加型のテレビ会議システムは従来からテレビ会議の主流となっているが，1拠点に多人数が参加した場合，画面の表示サイズが限られているなどの理由から各参加者の映像が小さくなってしまい，「誰が発言者か分かりにくい」または「発言者の表情をとらえにくい」などの問題点があった．首振り型の光学ズーム式カメラを導入し，専用オペレータが手動で操作する場合もあるが，コストが多大にかかるため多くのテレビ会議では実施できないのが実情である[6]．

著者らは，発言者を容易に特定でき，その表情が分かりやすい多人数参加型テレビ会議システムを構築することを目的として研究を実施しており，過去の研究[7),8)]において，複数の会議参加者の中から発言者を自動的に検出し，検出した発言者を遠隔地のテレビ会議画面に拡大表示することができるテレビ会議システムを構築した．同時に3名までの会議参加者を拡大表示できるようになっており，拡大映像がウィンドウ上部に，会議室全体映像がウィンドウ下部に表示され，1地点の会議映像として**図1**のような映像が別の地点に送信される．

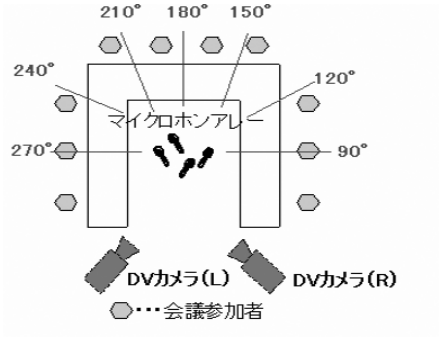
本システムは，マイクロフォンアレイを用いて音声到来方向を認識することにより発言者方向の推定を行い，さらに，2台の固定DVカメラから得られた映像から発言者の姿を検出してデジタルズーム表示する機能を備えている（マクロフォンアレイとDVカメラの配置については図5参照）．システムが新しい発言を検出すると，それまで拡大表示されていた3名のうち発言が最も少ない人を取り除き，代わりに新しく発言を行った人を拡大表示するようになっている．また，マイクロフォンアレイは著者らが開発したものであり，PC 1台と安価な市販音響機器（サウンドカード2枚，マイクロフォン4個）のみから構成されていることが特徴である[7]．そして，マイクロフォンアレイによる音声到来方向認識と，フレーム間差分認識および肌色認識による人物位置認識とを組み合わせて発言者方向検出精度を向上させることにより，1拠点10名程度の参加者が参加するテレビ会議環境において，特殊なハードウェアを使用せずに発言者の特定を行うことができた[8]．

図1 テレビ会議実行画面
Fig. 1 Screen image of TV conference.

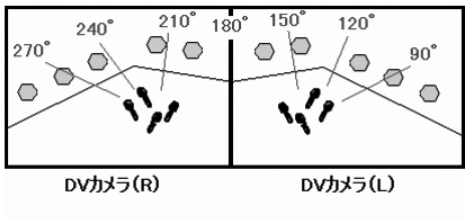
しかしながら，過去に開発したシステムでは，DVカメラ映像上の発言者の位置とマイクロフォンアレイの検出する音声到来方向の対応付け作業（キャリブレーション）をすべて目視によって手作業で行わなければならず，正確さにかける問題や，時間と手間が多大にかかるという問題があった．また，キャリブレーション作業に習熟した複数作業員が必要となり，1名のみで作業を行うことが困難であるといった問題があった．そこで本論文では，全方位カメラ[9]の利用により，短時間で誰もが正確にキャリブレーション作業を行える方法について提案を行う．またその実装として，全方位カメラを利用した手動キャリブレーション方法および自動キャリブレーション方法について述べる．評価実験の結果，従来2名以上の作業者によって行わなければならなかった撮影機器設定作業が，1名のみで迅速かつ正確に行えることが確認できた．

2. 過去に開発したシステムの問題点

ビデオカメラ映像上の発言者の位置とマイクロフォンアレイの検出する音声到来方向の対応付けをここでは「キャリブレーション」と呼ぶ．話者位置推定のた



(a) マイクロフォンアレイを中心とした角度



(b) DVカメラ2台の入力映像上での角度

図 2 キャリブレーション
Fig. 2 Calibration.

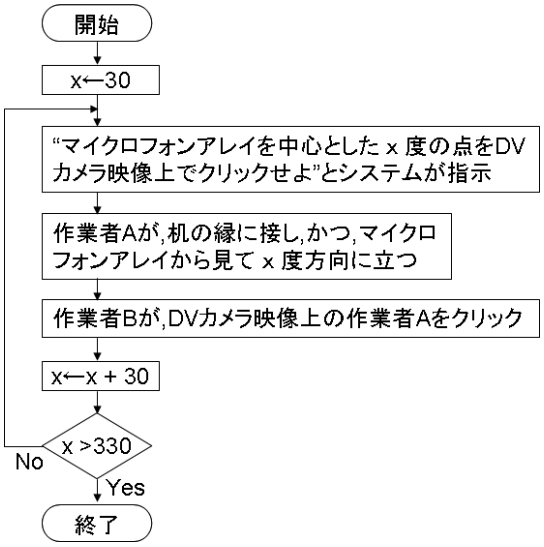


図 3 従来のキャリブレーション方法
Fig. 3 Existing calibration method.

めのマイクロフォンアレイを中心とした角度と会議参加者映像を取り込んでいる DV カメラからの角度とは非線形に対応しているため，会議開始前にキャリブレーションが不可欠となる．本テレビ会議システムを他の部屋に持ち運んで用いる際はもちろん，同じ部屋で用いる場合でも，マイクロフォンアレイの位置・角度，DV カメラの位置・画角，会議机の位置・レイアウトのいずれかが変更された場合にキャリブレーションが必要である．

キャリブレーションは，マイクロフォンアレイを中心とした角度（**図 2** (a)）を DV カメラの映像上の角度（図 2 (b)）と一致させる必要不可欠な作業であるが，過去に開発したシステムではこの作業を正確に行うことがきわめて難しいという問題があった．

具体的には，「マイクロフォンアレイを中心とした X 度の点を DV カメラ映像上でクリックしてください」（X は 30 から 330 まで 30 ずつ増える値）という指示がシステムから 11 回与えられ，この作業を作業者が実行するようになっていた（図 2 (b) で示されるように，0° 方向は DV カメラで撮影不可能な方向）．しかしながら，正常なキャリブレーション結果を得るためには，会議机の縁に近接している箇所をクリックする必要があり，クリックすべき 30° ごとの方向を DV カメラ 2 台の映像を並べた画像（図 1 下部の全体映像）から探すことはきわめて難しいという問題があった（会議机の縁から離れて存在する物体を選定した場合には不正なキャリブレーション点となってしまう）．また，クリックすべき 30° ごとの方向には目印となる物体が存在しないことが普通であり，作業者の勘だけでクリックが行われることが多かった．

このような問題から，過去に開発したシステムにおける撮影機材設定作業では，1 名（作業者 A）が 30° ごとの点と思われる箇所に立ち，他の 1 名（作業者 B）がその立っている人を DV カメラ映像上（図 1 下部の全体映像上）でクリックするという方法によって行うことが実質的に必要となっていた（作業者 2 名によるキャリブレーション作業のフローを**図 3** に示す）．しかしながら，「会議机の縁に近接し，かつ，マイクロフォンアレイから見て 30° ごとの箇所」に正確に人が立つこと自体が難しく，作業者が少なくとも 2 名必要になるという問題だけでなく，2 名で行ったとしても正確さにかけるという問題があった．キャリブレーションの正確さは本テレビ会議システムの発言者拡大性能に大きく影響を与えるため，キャリブレーション作業を正確かつ迅速に行える手段を提供することが過去に開発したシステムにおいて大きな課題となっていた．

3. 提　　案

3.1 提案手法

全方位カメラを利用することにより，本来，音声方向と映像位置を対応づける必要があるキャリブレーション作業を，映像方向と映像位置を対応づける作業

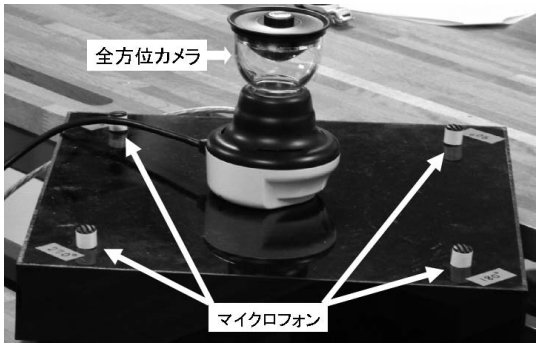


図 4 マイクロフォンアレイに固定された全方位カメラ
Fig. 4 Omnidirectional camera attached to microphone-array.

で済ませられるようにする方法を提案する．

全方位カメラ（**図 4**）は双曲面ミラーと CCD カメラから構成され，周囲 360° から双曲面に届く光を双曲面ミラーによって反射させることで円画像を生成する装置である．全方位カメラの下部には CCD カメラが取り付けられており，円映像を PC に出力できるようになっている．今回用いた全方位カメラ[9] の場合，PC には IEEE1394 カメラ映像として入力されるようになっている．

著者らは，図 4 のように，全方位カメラの中心が，マイクロフォンアレイ（四隅に無指向性マイクが 1 つずつ取り付けられている図 4 中の黒い箱）の中心になるように固定して取り付けた．このように取り付けることによって，マイクロフォンアレイを中心とした角度と，全方位カメラが取り込む画像の角度が等価となり，全方位カメラ映像を用いて，マイクロフォンアレイのキャリブレーション作業を実行できるようになっている．なお，キャリブレーション完了後は全方位カメラを用いる必要はないため，会議中などはマイクロフォンアレイの筐体から取り外し可能となっている．

また本システムは，**図 5** のような一般的な会議室環境において使用されることを想定している．ローカルな会議参加者はコの字型の会議机を囲んで着席し，遠隔地参加者の映像は会議室前方に設置されたテレビモニタに表示される．そして，テレビモニタに表示された遠隔地参加者と会話する際，会議参加者はテレビモニタの方向に顔を向けて話をする．

また機器設置の際は，マイクロフォンアレイは必要な音声到来方向認識精度を確保するために会議机の中央付近に設置され，2 台の DV カメラはテレビモニタの両端に近接するようにして設置される．このように設置された機器を用い，マイクロフォンアレイから得られた話者位置推定データに基づいて，DV カメラからの入力映像中の発言者を特定し拡大撮影するようになっている．

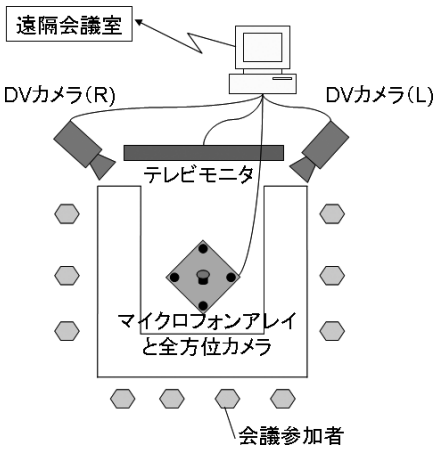


図 5 システム構成
Fig. 5 System architecture.

全方位カメラ 1 台と DV カメラ 2 台は 1 台の 3.4 GHz Pentium4 PC にすべて IEEE1394 接続されており，すべてのカメラから 1 秒間に 30 フレームの画像が当該 PC に取り込まれて処理される．また，マイクロフォンアレイに取り付けられた 4 つのマイクロフォンからの音声入力信号は，当該 PC に取り付けられた 2 枚のサウンドカードにそれぞれ 2 系統ずつ入力され，音声到来方向認識用途に利用される[7]．

3.2 関連研究

ここで提案手法の新規性について議論する．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，これらを自動化しようとする動きがある．

マイクロフォンアレイによって発言者を特定し，首振り型の光学ズーム式カメラを制御して発言者を撮影する会議システム[6] や，マイクロフォンアレイによって講師や質問者の位置を特定し，首振り型の光学ズーム式カメラを制御して自動撮影する講義自動収録システム[10),11)] が存在する．しかしながらこれらの既存研究では，マイクロフォンアレイを中心とした角度と光学ズーム式カメラの首振り角度との対応関係をあらかじめパラメータとして与える必要があり，キャリブレーション作業に関する支援機能を提供していない．このため，マイクロフォンアレイの位置，撮影カメラの位置，会議机の位置・レイアウトなどが変化した場合に，利用者によるキャリブレーションが非常に困難であるという問題がある．また，これら既存研究には，マイクロフォンアレイとビデオカメラのキャリブレーショ

ンのために全方位カメラを用いた例は見当たらない．

上記以外の関連研究としては，固定カメラとズーム制御可能なカメラとを併用し，特定重要箇所に限定して自動ズーム撮影する研究がある[12)~16)]．しかしながら，これら従来研究においてはマイクロフォンアレイを用いていないため，音声到来方向と撮影方向のキャリブレーションについて言及していない．

全方位カメラを利用したキャリブレーション機構に関し，次章以降，手動キャリブレーション構成と自動キャリブレーション構成について詳細に述べる．

4. 手動キャリブレーション構成

全方位カメラを利用したキャリブレーション機構として，まず，キャリブレーションウィンドウ（**図 6**）をユーザがマウスクリックして行う手動キャリブレーション構成を実装した．キャリブレーションウィンドウには，全方位カメラの 360° 映像（全方位カメラが生成した円映像を，映像処理をほどこして矩形に引き伸ばした映像）が上部に，DV カメラ 2 台の映像が下部に表示される．そして，作業者は，全方位カメラ画像と DV カメラ画像の対応点をクリックすることでキャリブレーションを行えるようになっている．

図 6 は，キャリブレーションを実行している最中のキャリブレーションウィンドウの表示例である．上部の全方位カメラ映像，下部の DV カメラ映像ともに，1 秒間に 5 回程度更新される準リアルタイム映像である．キャリブレーションウィンドウのタイトルバーには，現在のキャリブレーション点数が表示され，また，ウィンドウ下部には現在のキャリブレーション状態を表すグラフ（以下，キャリブレーション状態グラフ）が DV カメラ映像の上にオーバレイ表示されている．キャリブレーション点が追加されるごとに，図 6 のように，キャリブレーションウィンドウ下部にキャリブレーション点を表す白丸が追加され，キャリブレーション状態グラフの形が更新される．キャリブレーション状態グラフの形状を確認しながら作業できるため，作業者が異常なキャリブレーション点を発見しやすくなっている．

図 6 キャリブレーションウィンドウ
Fig. 6 Calibration window.

作業者は，上下映像の中に共通に表示されている物体を見つけ出し，この物体を上部映像中で左クリックし，続いて下部映像中で左クリックすることで，キャリブレーション点を追加できるようになっている．選定されるべき物体としては，会議机の周りに座っている人物や会議机の周りに配置されている椅子の背もたれなど，会議机の縁に接して存在する物体である．

なお，会議参加者が会議中にまったく同じ位置にとどまっていることはまれであり，座っている椅子を横方向や前後方向に動かすことがしばしば存在する．このため，あらかじめキャリブレーションした場所のみ拡大表示するようにすると，会議中に椅子を動かした人の映像が拡大領域からはみ出てしまうという問題や，予定外の参加者が途中で会議に加わったような場合に対応できないという問題が生じる．

そこで提案方式では，任意の座標点に対応する全方位カメラ映像上の角度を近似値によって求めるようになっている．具体的には，クリックされなかった座標点については，線形補間の変換式によって近似値が求められる．DV カメラ映像上の座標点 $X1$ と全方位カメラ映像上の角度 $Y1$ とが対応点としてクリックされ，また，DV カメラ映像上の点 $X2$ と全方位カメラ映像上の角度 $Y2$ とが対応点としてクリックされたとすると，DV カメラ映像上の任意の座標点 x（ただし，$X1 \leq x < X2$）に対応する全方位カメラ映像上の角度 y は以下の式で算出される（$X1$ と $X2$ の間に他のクリック点が存在しないと仮定）．

$$y = Y1 + \frac{Y2 - Y1}{X2 - X1}(x - X1)$$
$$\text{（ただし，} X1 \leq x < X2\text{）}$$

さらに，提案方式では，拡大領域に人物を表示する際，画像処理による人物位置認識（フレーム間差分認識，および，肌色認識を組み合わせて認識）の結果に基づき，人物が拡大領域の中心に表示されるようにズーム撮影位置の微調整が行われるようになっている．

キャリブレーション状態グラフの横軸は DV カメラ映像の X 座標（左端が 0，右端が 719）と一致している．一方，グラフの縦軸は，全方位カメラ映像上の角度（DV カメラ映像の最下部が最小値 0° であり，DV カメラ映像の最上部が最大値 359°）である．グラフ上の白丸を SHIFT キーを押しながら左クリックすると，

当該キャリブレーション点を削除することができる．また，キャリブレーションウィンドウ上で C キーを押すとキャリブレーション点が 0 個の初期状態に戻る．

5. 自動キャリブレーション構成

次に著者らは，全方位カメラを利用したキャリブレーション機構として，マーカ（色紙）を画像認識することにより行う自動キャリブレーション構成を実装した．自動キャリブレーションは，全方位カメラと DV カメラとで同じマーカ（色紙）を自動追尾して，キャリブレーションを自動的に行う仕組みであり，マーカを会議机の縁に沿って一周させるだけでキャリブレーションが完了するようになっている．

作業者は，会議参加者が座ると予想される座席の上をマーカが通るようにして，マーカを持って会議机の周りを一周する（**図 7**）．このとき，システムは全方位カメラ映像からマーカの位置を検出し，さらにこれと同期して DV カメラ映像から同一のマーカを検出することを試みる．そして，両カメラ映像から同一のマーカが検出できた瞬間において，検出時刻を記録するとともに，全方位カメラ映像から得られたマーカの角度と DV カメラから得られたマーカの座標を記録する．このようにして，キャリブレーションウィンドウ上をクリックすることなしに，全方位カメラと DV カメラとで位置あわせを自動的に行うことができるようになっている．そして，手動キャリブレーション構成と同様，対応付けされなかった座標点については，線形補間によって近似値が求められる．

自動キャリブレーション処理の実装について述べる．マーカの認識は，フレーム間差分[17),18)] による動体認識と，HSV 表色系[19)] を用いた色認識を基本としている．

図 7 自動キャリブレーション作業
Fig. 7 Auto calibration task.

まず，マーカの色としては会議室で用いられることが少ない色を選定することが望ましいと考えた．調査の結果，一般的なオフィスや会議室では，白やグレーを基調とした淡い明るい色や，沈静色である青色が好まれ，一方，彩度の高い赤は，自律神経系統を刺激し血圧・脈拍・呼吸数の増加を誘発することから使われることがきわめて少ないことが分かった[20)]．著者らが所属する大学の会議室 5 つを調査した結果からも，人を興奮させる色とされる赤色は床・壁・什器の色としてはいっさい使われていなかったことから，マーカとして利用するのに適している色であると考えた．

次に，本システムのマーカ認識方法（**図 8**）について述べる．まず，全方位カメラから入力された映像と，DV カメラから入力された映像の両方に対し，フレーム間差分による動体認識を行っている．これにより，会議室内に赤色の物体が存在したとしても，それが静止している場合には認識対象から除外できる．

次に，動体と認識された部分について HSV 表色系に基づく色認識を行い，「鮮やかな赤色」とそれ以外の色に 2 値化する．色認識には HSV 表色系を用いて H（色相）の値が 20 以下または 330 以上で，S（彩度）の値が 0.7 以上であるものを鮮やかな赤色として特定した．HSV 表色系の H および S の値は輝度（V）から独立しており，時間帯などでさまざまに変化する会議室の明るさに影響されにくいという特徴がある．

しかしながら，実際にマーカを認識させてみたところ，全方位カメラの双曲面ミラーを囲むガラス面に赤色マーカの映りこみが存在し，これがノイズとなって出力映像が乱れてしまうことが分かった．

そこでこの問題を解消するために，メディアンフィルタ，モザイク処理，などの各種ノイズ除去手法を試みた．結果としては，収縮・膨張によるノイズ除去処理をほどこすこととした．収縮とは，ある 2 値化画素の近傍に 1 つでも 0 があればその画素を 0 に，そのほかは 1 にする処理であり，一方，膨張とは，ある画

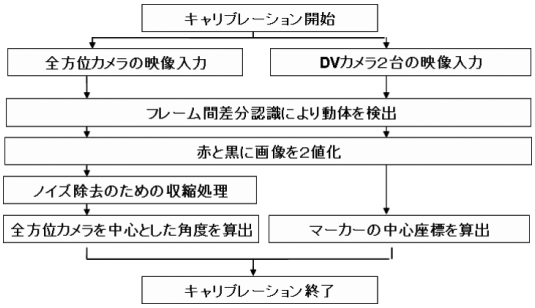


図 8 マーカ認識方法
Fig. 8 Marker recognition.

素の近傍に 1 つでも 1 があればその画素を 1 に，そのほかは 0 にする処理である．この処理を複数回繰り返し行うことでノイズを除去することができた．処理が単純であり，PC にかかる負荷を小さくできるという利点がある．

ノイズ除去処理が終わると，最後に，全方位カメラ画像と，DV カメラ画像の両方からマーカ位置の中心座標が算出され，キャリブレーション結果として出力される．そして，この自動キャリブレーション結果は，キャリブレーションウィンドウ上（図 6）で F キーを押すと，テレビ会議システムに取り込まれる．

6. 評価・考察

6.1 手動キャリブレーション

手動キャリブレーションの評価について述べる．過去にキャリブレーション作業を行った経験のない被験者（大学生）8 名を対象として，過去に開発したシステムの方法と今回提案した手動キャリブレーション方法の両方でキャリブレーション作業を行わせ，改善効果を検証する実験を行った．4 名の被験者は，提案した手動キャリブレーション手法，従来手法の順で作業を行い，残り 4 名の被験者は，従来手法，提案した手動キャリブレーション手法の順で作業を行った．

まず，従来のキャリブレーション方法の正確さと時間を計測する実験（実験 1a および 1b）について述べる．2.4 m × 2.0 m の長方形会議机の上にマイクロフォンアレイを配置し，被験者に対し，このマイクロフォンアレイを中心とした 30° から 330° までの 30° ごとの方向を特定する作業を課した．実験 1a は，1 名によるキャリブレーションを想定したものであり，各被験者には「マイクロフォンアレイを中心とした 30° ごとの方向を DV カメラ映像から見つけ出す」というタスクが課せられた．実験 1b は，2 名によるキャリブレーションを想定したものであり，各被験者には「マイクロフォンアレイを中心とした 30° ごとの方向であり，かつ，会議机の縁に接する位置に立つ」というタスクが課せられた．そして実験 1a および 1b において，被験者が目測で選定した角度と正しい角度とを比較した．

実験 1a の結果，被験者が DV カメラ映像上で選定した角度と正しい角度との誤差は，平均 21.3°（標準偏差 = 22.1°）であった（n = 88）．また実験 1b の結果，被験者が立った角度と正しい角度との誤差は，平均 9.3°（標準偏差 = 12.4°）であった（n = 88）．実験 1a の結果である 21.3° という値は，仮にマイクロフォンアレイが発言者からの音声方向を 100%正しく出力できたとしても，発言者の隣の人を拡大表示してしまう確率がきわめて高い値である（360° 方向に 10 名が均等に着席した状況を想定）．また，実験 1b の結果である 9.3° という値は，実験 1 と比較してかなり誤差が減少しているものの標準偏差が 12.4° と依然大きく，キャリブレーションの精度としては十分ではない値である．現状，マイクロフォンアレイの音声到来方向推定能力が限定的[8] であることを考慮すれば，マイクロフォンアレイの出力誤差とキャリブレーション時の誤差が足し合わされてしまうため，わずかであってもキャリブレーション結果に誤差を含むことは望ましくない．

1 回のキャリブレーション作業（30° から 330° までの 11 点のキャリブレーション作業）にかかった時間は，実験 1a においては被験者 1 名あたり平均 149.0 秒であり，実験 1b においては被験者 1 名あたり平均 133.5 秒であった．ただし，過去に開発したシステムが実際のテレビ会議で使われている様子を観察してみると，撮影機器セッティングを終えた後に机やカメラ位置を動かしてしまい，キャリブレーション作業をやり直すことが多々あり，キャリブレーションに 15 分以上かかるケースがたびたび見受けられた．

次に，今回提案したキャリブレーション方法の正確さと時間を計測する実験（実験 2）について述べる．実験 1a および 1b と極力実験条件をそろえるために，同じ長方形会議机の周りに 11 個の椅子を適当に配置し，この椅子の背もたれをクリック点として，キャリブレーションウィンドウの上部と下部をそれぞれ 11 回ずつクリックさせた．

実験 2 の結果，被験者が DV カメラ映像上で選定した角度と正しい角度との誤差は，平均 1.9°（標準偏差 = 1.5°）であった（n = 88）．また，1 回のキャリブレーション作業にかかった時間は，平均 28.5 秒であった．実験 1a および 1b の結果と実験 2 の結果とを比較して，今回提案した方法においては，キャリブレーション作業時間は大幅に短縮され，かつ，作業者 1 名のみで十分精度良く実施できるようになったことが確認できた．

くわえて，従来方式においてはつねに 11 個の点をクリックする必要があったが，今回提案した方式においては必ずしもキャリブレーション点を 11 点とる必要はなく，会議参加予定者数を考慮してキャリブレーション点数を決定すればよいという特徴がある．また，従来方式においては会議参加者が座る位置をキャリブレーション点とすることができず誤差が生じていたが，今回提案した方式においては椅子の置かれた位置に

表 1 従来方式と提案方式（手動キャリブレーション）の特徴の比較
Table 1 Comparison between existing and proposed manual methods.

	従来方式	提案方式
マイクロフォンアレイから見た角度の指定方法	DVカメラ映像上でクリック	全方位カメラ映像上でクリック
キャリブレーション点数	常に11点	参加予定者数
キャリブレーション位置	30° 毎の位置	参加者着席予定位置

おいて正確にキャリブレーションすることができるという特徴がある．従来方式と提案方式（手動キャリブレーション）の特徴の比較について**表 1** に示す．

たとえば，会議参加予定者数（+アルファ）分の椅子を，各参加者が座りそうな場所に配置しておき，その椅子の背もたれをクリック点とすれば精度の高いキャリブレーションが可能となる．会議参加人数はほぼ分かっていることが多いため，このような方法は現実的であると考えられる．さらに本実装においては，提案した手動キャリブレーションによって会議中にでもキャリブレーション点を追加することが可能であり，会議に途中参加した人をキャリブレーション点とするようなことができる．

6.2 自動キャリブレーション

自動キャリブレーションの評価について述べる．24 cm × 24 cm の大きさの赤色紙をマーカとして用い，自動キャリブレーションを実施する実験を行った．過去に開発したシステムにおいて 360° 範囲の 11 点をクリックすることでキャリブレーションを行っていたことから，本実験においても，11 点以上のキャリブレーション点が自動取得できた場合に成功と判断することとした．

4.8 m × 3.2 m の長方形会議机がある会議室で実験を行った．この会議机を一周する時間長が 8 秒，12 秒，20 秒となるようにマーカの移動速度を変えて実験したところ，一周する時間長が 8 秒，12 秒，20 秒に対し，取得できたキャリブレーション点数はそれぞれ平均 7 点，22 点，44 点となった．これにより，時間長が 12 秒以上（移動速度 1.3 m/s 以下に相当）の場合に必要な数のキャリブレーション点を取得できることが分かった．1.3 m/s は「会議机の周りを少し速く歩く」速さに相当する移動速度である．

なお，より大きい 10 人掛け会議机に対し同様の自動キャリブレーション実験を行ったが認識率の低下はみられなかった．また，24 cm × 24 cm より大きいサイズのマーカを用いて同様の実験を実施したが，認識率が向上する効果はみられなかった．マーカを大きくすることより，マーカの移動速度を遅くすることが認識率を上げるためには重要であることが分かった．

6.3 考　　察

まず，本論文で提案した手動キャリブレーションと自動キャリブレーションの精度について考察する．提案方式では，任意の座標点に対応する全方位カメラ映像上の角度を線形補間によって求めているが，線形補間を行っている性質上，既知点（手動キャリブレーションにおけるクリック点）が密であれば分解能が高まり近似の精度が良くなるといえる．

手動キャリブレーションの場合，クリックすべき「会議机の縁に接しており，かつ，DV カメラ映像と全方位カメラ映像の両方に表示される物体」は，通常，机の周りに置かれた椅子の背もたれしか存在しないことが多く，そのような場合取得できる既知点の数は置かれた椅子の数と等しくなる．また，全方位カメラ画像を矩形に引き伸ばした映像は解像度が低いうえに極度に歪んでおり，クリックすべき物体が画面の端に表示されている場合，人間の目で見つけにくかったり，会議机の縁に接しているかどうかを判断できなかったりするという問題が生じることがある．

一方自動キャリブレーションの場合，マーカを会議机の縁に接して移動させることによって既知点を無数に取得することができる．このことから，手動キャリブレーションと自動キャリブレーションとを比較すると，自動キャリブレーションの方が既知点を多く取得できることから近似精度がより高くなることが期待できる．実際，自動キャリブレーションの評価実験の結果，通常の歩行速度でマーカを移動させることで，手動キャリブレーションより多くの既知点を取得することができることが分かった．このようなことから，自動キャリブレーションが実行可能な状況であれば，極力自動キャリブレーションの手法を用いることが望ましいと考察できる．

ただし，会議前に自動キャリブレーションができなかった場合や，会議中に机・マイクロフォン・ビデオカメラのいずれかの配置が変更されたような場合には，手動キャリブレーションによって対応する必要がある．また，画像認識による自動キャリブレーションの結果が明らかに不正な値であった場合，手動キャリブレーションによって不正値を修正する手段を提供することが望ましい．以上のことから，現状の実装では，手動キャリブレーションと，自動キャリブレーションの両

方がつねに利用可能な状態となっており，必要に応じて選択または併用できるようになっている．

次に，全方位カメラを会議参加者撮影のために用いる案について考察する．全方位カメラをキャリブレーション用途に導入するに際し，キャリブレーション用途だけでなく会議参加者および発言者の撮影のためにも用いることができないか検討した．しかしながら，結果としては，机の中心付近に置いた全方位カメラで発言者を撮影した場合，会話している当人同士が顔を合わせられないという問題があることが分かり，そのような利用は行わないようにした．

図 5 に示されるような一般的なテレビ会議を想定した場合，テレビモニタに表示された遠隔地参加者と会話する際，会議参加者はテレビモニタの方向に顔を向けて話をする．しかしながら，テレビモニタの表示面に対向して着席していない会議参加者が，遠隔地参加者に話しかけるためにテレビモニタ方向に顔を向けた場合，机の中心付近に置いた全方位カメラはその発言者の顔を横方向からしか撮影できないという問題が生じる．このことから，遠隔地参加者と会話する際に極力視線が合う（顔が向き合う）ようにするために，テレビモニタと撮影用カメラを近接させて配置する必要があり，今回のように，テレビモニタと近接した DV カメラによって会議参加者を撮影する構成とした．この際，マイクロフォンアレイは必要な音声到来方向認識精度を確保するために必ず会議机の中央付近に設置される必要があることから，必然的にマイクロフォンアレイと撮影用カメラの位置は離れて設置されることとなり，今回提案したようなキャリブレーションが必須となる．

また，高解像度の全方位カメラをテレビモニタに近づけて配置し会議参加者撮影用に利用することも可能ではあるが，マイクロフォンアレイと全方位カメラの設置場所が離れてしまうために，やはり本論文で述べている撮影方向あわせのためのキャリブレーション操作が別途必要となり，キャリブレーション用の全方位カメラとあわせて 2 台の全方位カメラが同時に必要となる．また，DV カメラと同等の解像度で撮像を生成できる全方位カメラは現状非常に高価であり，安価にテレビ会議システムを構成することを目指した本研究の目的[7] に合わないという問題もある．

なお，DV カメラ 1 台のみをテレビモニタの上部または下部に固定して設置したとすると，ローカルな会議参加者同士が会話している場面を撮影する際に，テレビモニタの表示面に向かい合って着席していない会議参加者を横方向からしか撮影できなくなってしまうという問題がある．このことから，本提案方式では，ローカルな会議参加者と遠隔地参加者とが会話している場面と，ローカルな会議参加者同士が会話している場面の両方をカバーするために，テレビモニタの両端に 2 台の DV カメラを設置する構成としている．

7. おわりに

過去に開発した 1 拠点多人数参加型テレビ会議システムの課題であった撮影機器設定作業時の問題点を解消する方法について述べた．本論文では特に，全方位カメラの利用により短時間で正確にキャリブレーションを行う方法について提案した．実験評価の結果，過去に開発したシステムと比較して改善がみられたことを確認した．また現状の実装においては，提案した手動キャリブレーションと自動キャリブレーションの両方が利用可能であり，必要に応じて選択または併用できるようになっている．

テレビ会議機能がネットワークサービスとして提供されるようになり，テレビ会議の利用者層は中小企業に拡大している．このような利用者層においては，テレビ会議のために専用オペレータを雇用するなどのコストをかけることは難しいため，本論文で提案した自動撮影機能・自動機器設定機能を有したテレビ会議システムが特に適していると考えられる．

参考文献

1) CNA レポート—電話会議システム業界専門マーケットリサーチ&コンサルティング (2006). http://cnar.jp/
2) 橋本啓介：テレビ会議市場についての一考察，CNA レポートジャパン (2002). http://cnar.jp/TV-Kaigi.pdf
3) TANDBERG，導入事例 (2006). http://www.tandbergpartner.jp/
4) V-cube Inc.: Nice to Meet You (2006). http://www.nice2meet.us/ja/
5) ギンガネット (2006). http:/www.ginganet.co.jp/business.html
6) 大西，影林，福永：視聴情報統合による会議映像の自動撮影，電子情報通信学会論文誌，Vol.J85-D-II, No.3, pp.537–542 (2002).
7) 冨野，井上，市村，松下：マイクロホンアレーと映像処理を用いた多人数参加型テレビ会議システム，情報処理学会 DICOMO 2005 論文集 (2005).
8) 冨野，井上，市村，松下：多人数参加型テレビ会議システムにおける発言者拡大映像の作成，情報処理学会論文誌，Vol.47, No.7, pp.2091–2098 (2006).
9) 全方位カメラ，映蔵 (2006).

http://www.eizoh.co.jp/quality/index.html
10) Liu, Q., Rui, Y., Gupta, A. and Cadiz, J.J.: Automating Camera Management for Lecture Room Environments, *Proc. ACM CHI 2001*, Vol.3, Issue 1, pp.442–449 (2001).
11) 西口，東，亀田，角所，美濃：講義自動撮影における話者位置推定のための視聴覚情報の統合，電気学会論文誌 C，Vol.124, No.3, pp.729–739 (2004).
12) Cruz, G. and Hill, R.: Capturing and Playing Multimedia Events with STREAMS, *Proc. ACM Multimedia 94*, pp.193–200 (1994).
13) Chiu, P., Kapuskar, A., Reitmeier, S. and Wilcox, L.: NoteLook: taking notes in meetings with digital video and ink, *Proc. ACM Multimedia 99*, pp.149–158 (1999).
14) Uchihashi, S.: Improvising camera control for capturing meeting activities using a floor plan, *Proc. ACM Multimedia 01*, pp.12–18 (2001).
15) 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).
16) 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).
17) 市村，井上，宇田，伊藤，田胡，松下：ChalkTalk：講師動画と板書静止画の同時記録が可能な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).
18) 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).
19) 橋本 聡，藤本研司，中村 納，南 敏：顔領域抽出に有効な修正 HSV 表色系の提案，テレビ誌，Vol.49, No.6, pp.787–797 (1995).
20) 中野，冨野，福井，市村，松下：多人数テレビ会議システムにおけるキャリブレーションの自動化と映像表示の工夫，第 68 回情報処理学会全国大会，5T-9 (2006).

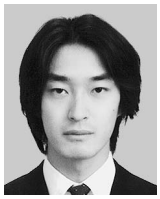
(平成 18 年 5 月 29 日受付)
(平成 18 年 11 月 2 日採録)

市村 哲（正会員）

1989 年慶應義塾大学理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（FXPAL）駐在．2002 年より東京工科大学助教授．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 & DICOMO 2005 優秀論文賞受賞．ACM，電子情報通信学会各会員．

福井登志也

1989 年日立製作所半導体事業部入社．1992 年日立茨城工業専門学院管理工学科卒業．1997 年東京ソフト販売入社．1997〜2003 年富士ゼロックスに派遣，ソフトウェア開発業務に従事．2003〜2005 年蝶理情報システムに派遣，官公庁システムの開発業務に従事．2005 年より東京工科大学 Linux オープンソースソフトウェアセンターにてソフトウェアの研究開発業務に従事，現在に至る．



井上 亮文（正会員）

1999 年慶應義塾大学理工学部計測工学科卒業．2001 年同大学大学院理工学研究科前期博士課程修了．2005 年同大学院理工学研究科後期博士課程修了．博士（工学）．現在，東京工科大学コンピュータサイエンス学部助手．グループウェア，マルチメディアコンテンツ処理の研究に従事．DICOMO 2006 ヤングリサーチャー賞受賞．

星　　徹（フェロー）

1969 年東京工業大学工学部電気工学科卒業．同年（株）日立製作所入社．1975 年カリフォルニア大学ロサンゼルス校（UCLA）大学院コンピュータサイエンス学科修了．交換システム，マルチメディア LAN，リアルタイムグループウエア，CTI，IP テレフォニー，RFID 応用，ユビキタスネットワーク等の研究に従事．現在，東京工科大学コンピュータサイエンス学部教授．IEEE，ACM，電子情報通信学会各会員．情報処理学会フェロー．工学博士．