Kuroko:話者シルエットを活用するプレゼンツール

梅村　恭司[†]　　梅村　真由[‡]

話者のシルエットを背景に薄い影絵として表示するプレゼンテーションシステムのプロトタイプを開発した．これにより，講演のポインティングデバイスに利用できるのはもちろんのこと，講演者のジェスチャを表示できる．これによって，話者がスクリーンの前にたってプレゼンしているかのような効果を提供しながら，講演者によってスクリーンが覆われるようなことを避けることができる．

Kuroko: A Presentation Tool with Presenter's Silhouette

KYOJI UMEMURA[†]　　MAYU UMEMURA[‡]

We have developed a prototype of presentation system that displays the gray silhouette of presenter in the background of the slides. This system provides not only the functionality of pointing devices, but also a method to convey the gesture of the presenter. As the result, the audience feels as the presenter stands before the screen. Unlike the real situation, the presentation does not suffer from the undesired effect, that is, some part of presentation may be covered by the presenter.

1.　はじめに

近年，学会のプレセンテーションが録画されたり，ネットワークに配信されることが多くなっている．このようなときにレーザポインタなどのポインティングデバイスを使用すると，コンピュータの出力信号を処理している場合，指し示している場所の記録が正しく扱われないという問題がある．投影されているスクリーンをカメラで処理する場合には，ポインティングの場所が処理される代わりに，プレゼンテーション情報が明瞭性が失われるという問題がある．

また，録画や配信の状況では，話者のジェスチャを処理することにも問題がある．プレゼンテーションスクリーンと話者のジェスチャを同時に録画する場合には，プレゼンテーションの画面が小さくなり，プレゼンテーションの明瞭性が低下する．２つを切り替えて処理する場合には，切り替えを自動で行うことは難しいため，専用のオペレータを配置しなければいけなくなり，コストが高くなるという問題がある．

録画や配信ということを行わない場合でも，話者とスクリーンが離れて配置されるような大人数教室でのプレゼンテーションにおいては，話者のジェスチャを見ることとスクリーンを見ることに視線の移動があり，タイミングによっては，プレゼンテーションの情報を読み損ねる危険性がある．スクリーンと講演者の距離を近くにするという方法は，スクリーンの情報が講演者の陰になって判読できないタイミングが生じるという危険性がある．

このような問題に対処するために，話者の画像を処理して，動きのある場所を特定し，それに対応した影の画像を作成した後，その画像をプレゼンテーションの背景に表示するというシステムを開発した．背景に影が表示されても文字の判別に問題がないように影は薄い色とした（図１）．さらに，動きがないときには，色を薄くする．この図では，中央に話者の顔があり，説明箇所を指で示している．本稿では，システムの要求分析を行った上，その要求にあった効果的な話者のシルエットの計算方法と提示方法について報告する．

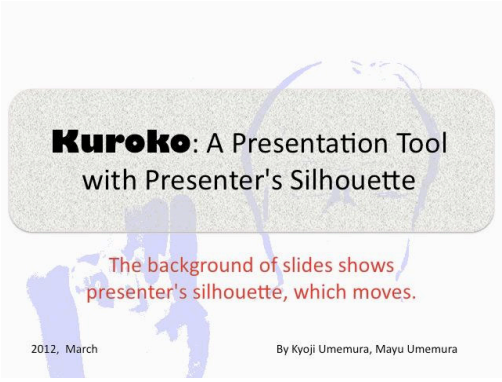

図１：話者シルエットを背景としたスライド表示

† 豊橋技術科学大学　情報・知能工学系
Department of Computer Science and Engineering, Toyohashi University of Technology
‡ 名古屋工業大学 情報工学科
Department of Computer Science, Nagoya Institute of Technology

2. プレゼンテーションツールの要求分析

想定するプレゼンテーションの環境として，以下のような状況を考える．特に S1, S2 の状況は，近年において，需要性が高まっているケースと考えられる．

- S1:プレゼンテーションの配信をインターネットで送るような会議.
- S2:プレゼンテーションを録画することで，オンラインの教材の資料を作成するような状況.
- S3:聴衆の数が多く，大きなスクリーンに投影して行うプレゼンテーション.

2.1 状況分析

想定している状況に共通していることは，話者のプレゼンテーションの動画と，本来のプレゼンテーションの二つを同時を扱うことが難しい状況と整理できる．つまり．S1 のストリーム配信に転送するときには，受信側では動画のフレームを一つで受信しなければならないケースが多い．S2 の録画については，録画の一つのストリームで録画をするケースが多い．S3 については，十分拡大されたスクリーンを視聴することはできるが，直接見る話者の姿は小さいために効果的でない.

通常のプレゼンテーションを準備するときには，スクリーンと同時に話者の姿が聴衆に見えることを仮定することが多い．このため話者の説明の間，スクリーンに何も変化がない時間が長い．話者の姿が見えない状況では，スクリーンに変化がない状態が長く続くことがプレゼンテーションの効果を下げることになる.

2.2 既存の方法による解決法と問題点

話者が直接は見えないことを想定しているケースにおいては，プレゼンテーションにアニメーションを組み入れて，聴衆の注意を持続するような工夫がなされる．しかしながら，話者の存在感そのものがプレゼンを構成する情報 (文献 3)であり，それを利用しないプレゼンテーションの方法ということになり，数十人のプレゼンテーションに比べ，プレゼンテーションの準備のコストが高くやアドリブ性を組み込めないプレゼンテーション方法といえる.

既存のプレゼンテーション資料を変更しないアプローチでは，話者の映像とスクリーンの情報を含めて一つの画像をつくり，それを録画するようなツールが市販されている．しかしながら，このためには，オリジナルのプレゼンテーションを縮小することが必要となり，また，二つの画像の同期と，プレゼンテーションで使用するレーザポインタの情報の扱いがうまくできない.

プレゼンテーションの補助者を用意できるという状況では，画面を適切に切り替えて，効果のある画面を選んで放送または録画するというアプローチがある．この場合，補助者の画像選択の能力と補助者の内容についての理解のレベルによって，プレゼンテーションの効果が左右される．そして，補助者を用意するコストは高いと考えられる.

直接的な解決方法として，著者のプレゼンテーション画像，スクリーンをカメラで撮影した画像，コンピュータから出力する画像，プレゼンテータの手書きの情報（書画カメラ，または，白板のカメラ撮影）の４つを同時に録画し，再生するときは，高画像の状態のまま４つのストリームを同時に扱い，それを表示するほうも高解像度のディスプレイを使うなどを想定し４つのストリームを同時に再生するというアプローチがある（文献２）．技術としては，画像の同期と管理を行うだけの単純な方法であり，プレゼンテーションの準備方法も少人数のものと同様に用意すればよいのでコストが低い．しかしながら，実際にそのようなプレゼンテーションを実施できる場所と視聴できる場所が限られるという問題がある.

2.3 話者映像の機能

授業でのプレゼンテーションや，学会等のプレゼンテーションにおいても，話者映像が必要であること事実であるが，話者映像を利用できなくてもプレゼンテーションが成立しないとまではいえない．逆に，このようなプレゼンテーションにおいては，話者の映像だけの映像ではプレゼンテーションは成立しない．これが示唆するところは，話者映像はプレゼンテーションの補助情報であり，スライドの情報と話者映像は対等ではないということである.

補助情報の最大のものは，話者の説明している場所を明示する機能である．口頭での説明とスライドを関連つける指し棒やレーザポインタなどが，必ず用意されるのは，この機能がプレゼンテーションで重要であることの証拠といえる.

追加の補助情報としては，会話における相手の表情や仕草と同様な情報である．講演者の様子の画像から，強調したい程度や，自信の程度等の微妙なニュアンスを受け取れることがあり，これが有効なのは，相手が見えない電話と，ビデオチャットなどによるコミュニケーションの差があることからも明白である.

3. アプローチ

われわれは，プレゼンテーションを主役として，話者映像を補助と考えているため，話者の映像をそのまま映すのではなく，シルエットの映像に加工し，ポインティングの情報を伝え，同時に，話者の存在感と動作で補助情報をできるだけまかなうことにした．シルエットで表示された楽器を使い，演奏者がスクリーンの前で演奏するというシステムの報告（文献１）がある．この会議のデモにおいてシルエットで表示された楽器が十分に印象を伝えていたことがシルエットを使うきっかけの一つになっている．また，近年，シルエットによる表示でも，人物の存在感がつたわるという報告（文献４）に影響を受けた．また，音楽の演奏会において，演奏者は白または黒の服装で演奏し，演奏している動作をつたえるものの，演奏の音に注意をむけることを邪魔しないように，演奏者の個性についてはむしろ抑制していることにも影響を受けた．

われわれの想定しているプレゼンテーションでは，話者の服装や，話者の性別，話者の年齢などはプレゼンテーションの理解を助ける情報とはいえないので，それが判明しないシルエットにしても補助情報を伝えることができると考えた．

逆に，シルエットにすることによって，プレゼンテーションに重ね合わせて表示してもスライドへの影響を抑制できる．そして，スライドと話者情報の重ね合わせ表示できれば，話者が自分のポーズで説明してる箇所を明確にすることができ，スライド情報だけを配信・記録しても有用なプレゼンテーションになると考えた．

4. 提案システム：Kuroko

提案システムは，ノート型コンピュータに標準的に装備されるようになったカメラを利用するプレゼンテーションツールとして実装した．プレゼンテーションツールとしては，キーボードの操作によって格納されたスライドを順番または逆順に表示していく最小限の実装であるが，カメラの画像をリアルタイムで加工して，人物のシルエットを作り出し，それをプレゼンテーションの背景に重ねて表示する（図１）．人物シルエットは白と黒の二値画像であることと，裏方でプレゼンを支えていることから，システムを舞台で黒装束で登場する人に例えて Kuroko と命名した．

シルエットの抽出はリアルタイムで処理する必要から，単純な方法となる．まず，システムは最近のある時間幅の画像から計算される背景画像を保持している．現在のカメラ画像を RBG それぞれ 8bit で表現し，それを 3 次元の点と考えたときに，画素毎の L1 ノルム（差の絶対値の合計）が閾値を越えたときにに，シルエットの影の部分にするという方法である．具体的には，図２のようなコードでピクセルごとのシルエットの判定をしている．図２において，prevR, prevB, prevG が保持されている背景画素の変数であり，currR, currB, currG が現在のカメラがその変数である．しきい値の 80 は実験により，調整した．

```
int diffR = abs(currR  - prevR);
int diffG = abs(currG  - prevG);
int diffB = abs(currB  - prevB);

if ((diffR+diffG+diffB) > 80) {
  // light purple
  video_image.pixels[j] = shadow_color; }
else {
  // white

  video_image.pixels[j] = white_color;
}
```

図２ 背景とカメラ画像から影を生成するコード

動きのない部分は，背景画像と現在画像とは近い色になるので，白色になり，動きのある部分だけ影色となる．閾値の値が大きいので大きな色の変化が移動しているところだけが影になるため，結果として，動いている場所の輪郭がとれる．体を水平報告に移動したときに得られる画像を図３に示す．ここで，大部分の背景が白色となっていることは重要で，スライドに対して影響を与える部分が小さくなるようにシルエットを作成している．また，顔と手が体よりも速く動いているので，影となる領域が広くなり，結果として強調される．図３では，白黒で表示しているが，生成するときはプレゼンテーションの邪魔にならないように薄いグレーで描画する．また，画面を指差してポインティングデバイスするときに方向を直感的な向きにするために，画像は左右を反転させる必要があった．

背景画像の更新時間も実験によって定めたが，体を止めているときには，シルエットがなくなるようにすることと，比較的にゆっくり動いているときに輪郭がわかるようにすることとを念頭に調整した．これは，体を動かさないことで，スライド本来の情報に注意をむけるように制御できることと，ゆっくりとしたジェスチャでも，それが伝わることにトレードオフがあり，両方が必要と考えたからである

カメラ画像とプレゼンテーションを重ね合わせるときに，背景以外のプレゼンテーション画素が変更され

ることは望ましくない．本来ならば，プレゼンテーションの画像の背景を特定して重ね合わせるべきであるが，現状ではカメラ画像が薄い色で構成されているので，濃い色を優先するという重ね合わせて，プレゼンテーション画面を優先させる処理を実現した．

図３　計算されたシルエット画像の例

実際に表示される場合には，影は動いているので，図１の例のように，薄い色であっても，また，輪郭だけの表示であっても，話者の存在は感じられる．最後に，話者の動きがとまって，影の領域の面積が小さくなったときに、影の色を薄くするという処理をして，動作をしていないときへのプレゼンテーションへの悪影響を抑制する工夫をした．

5.　シルエット画像の計算における問題

シルエット画像を計算するときに，問題となるのは背景画像の更新の方法である．初期のプロトタイプにおいては，プログラムを起動したときに背景画像を取得し，その差分を表示するという方法をとった．この場合，最初の数秒くらいは，カメラの明るさが安定しない（自動調整の途中と考えられる）ために，背景を消すことができないという問題が生じた．これを回避するために，１秒程度，カメラを初期化してから背景画像を取得する方法にしたが，さらに，以下のような問題点が生じた．

- P1:最初に話者がカメラに映っていると，そのイメージがずっと残ってしまう．
- P2:周囲の明るさが変化するか，またはカメラの自動の明度調整が作用すると，大きな領域が影と成ってしまう．
- P3:何らかのはずみで，コンピュータの位置がずれると，背景が表示され，それがずっと残ってしまう．
- P4:人物全体が影になる傾向に成るが，実際は内部で背景の一部が映り込むケースが多かった

これらを解決するために，「キーの操作で背景情報を再度取り込む」という機能を実装したが，実際にはプレゼンテーションの操作に別の要因による操作を含めるために，操作性が悪くなってしまった．

次の改良として，背景画像はつねに更新することとし，一つ前のタイミングの画像を背景画像とするように変更した．このように毎回更新していると，動いたときだけ瞬間的に細い影がでるようになった．これでは，十分に動作を見ることができなかった．

最終的には，ある時間範囲の重み付き平均をとった画像を背景画像とすることで，その時間範囲の間に静止しているものは，影とならないようになった．

画像の重み付け平均の重み付けについては，時間によって影響が指数関数的に減少するように平均することにした．これを実現するためには，更新しようとする画像と，背景の画像のちょうど中間の色素を新しい背景画素とする操作で実装でき，メモリ負荷も少ない．

背景画像がダイナミックに更新されることに伴い，生成されるシルエット画像は，いわゆるシルエット画像というより，シルエットの輪郭画像となった（図３）．人物の内部が白抜きになることについては，むしろ，プレゼンテーションの妨害にならないためには良いことであろうと考えた．

6.　評価

システムの効果を確かめるために，シルエット表示の on/off を設定できようにし，かつ，話者の前にカーテンを用意し，話者の映像が聴衆に見える／見えないを設定し，下記の条件で５枚づつのスライド（文字だけのスライド）を使い条件ごとにある話題の説明をした．内容は統計の初歩で，すべての話題の構成は類似しているが，内容は異なっているようにした．プレゼンテーションに図は用いないものとした．プレゼンテーションにおいてポインティングデバイスは重要なので，マウスポインタを常時表示し，ポインティングデバイスは全部の条件で使う条件とした．したがって，実験の目的は，ジェスチャ情報の効果を測定することになる．

- A:シルエット off，話者映像 on
- B:シルエット off，話者映像 off
- C:シルエット on，話者映像 off
- A(再度）:シルエット off，話者映像 on

そして，A と B，B と C，C と A の条件について，それぞれ，B の終わり，C の終わり，A(再度）の終わり

に直ちに比較してもらい記入してもらった．比較は，直前の２つを比較してもらい，印象が正確に残っているものとするようにした．評価の設問としては，「伝わりやすい条件はどちらか？」とし，コンテンツの内容を調査する物ではないことを強調した．実際のアンケートでは，「条件 X が良い」，「迷うが条件 X が良い」のように程度を記入してもらった．迷った場合でも，どちらかの条件を選ぶこと強制した．集計においては，「迷いがある」ということを無視し，選んだ条件がどれかを判定する．結果の分析の検定は符号検定を使用する．

被験者は，成人の 50 人である．片側検定の符号検定をする．最初に，A と B の比較は，36 対 14 で A が優れているという結果であり，p 値は，0.0013 で危険率 1%と考えたときに有意であり，話者が見えない状況では情報が伝わりにくいという結果は有意である．

つぎの比較は B と C であり，話者が見えない条件でシステムの効果があるかがわかる重要な比較である．B と C の比較は，18 対 32 で C が優れているという結果であり，p 値は，0.032 で，危険率 5%と考えたときに差が有意であるということがわかった．これによって， 提案するシステムが話者が見えない状況で効果があることが示唆された．

最後に，C と A の比較は，7 対 43 で A が優れており，p 値は，10^{-7} 以下であり，シルエットの提示は，話者が直接見える状態よりは不自然な状態であることが分かった．

同時に実施した自由記述においては，シルエットの動きに注意を取られてしまうケースがあることが指摘された．動きのないときに見えないように工夫ただけでは十分でなく，控えめ／滑らかに動くことが期待されるようである．処理速度のとの兼ね合いがあるので，簡単には解決できないが，プレゼンテーションのコンテンツへの影響を小さくできる方向に，今後さらに改良したい．

7. 議論

7.1 シルエット化の妥当性

話者画像をシルエットにして，背景に表示することが妥当かどうかについては，状況や考え方によって意見が分かれるところである．まず，このシステムを利用した場合，話者の動作がメインスクリーンに投影されるので，視線の移動の量が少なくなる．これは，スクリーンの前でプレゼンテーションを行うことを支援するシステム（文献５）と同様な効果が得られることが期待できる．

文献５のシステムは，話者が直接にスクリーンの前に立てることによる利点を追求しているために，ジェスチャ以外に，表情や動き等を通常のプレゼンテーションより有効に利用することを意図しており，シルエット化することと逆の方向であるようにも解釈できる．しかしながら，本稿の前提には，ネットワーク配信や録画などのを想定しており，その環境での話者動画を利用する方法のためのシルエット化であり，シルエット化しなくても効果的に画面に組み込むことができるならば，シルエット化は必然ではない．文献５のシステムでは，話者がスクリーンに立つことによるスクリーンの隠蔽に対して，プレゼンのコンテンツを変化させるというアイディアを提示しているが，このような機能が効果的に実装できれば，シルエット化しないで，そのままの情報をスライドと重ねることも可能かもしれない．

7.2 表情の扱いについて

本来の目的からすれば，話者の表情は重要な情報であるはずだが，実装したシステムでの話者のシルエット画像において話者の表情の情報は失われている．

表情をうまく取り出すようにすると，シルエットで表示されるところが増える問題がある．また，表情が仮にうまくとりだされたとしても，顔の部分がスライドのコンテンツに隠れると，顔の表情の情報が表示されないことになる．表情の問題については，現状は利用したいが，満足できる実装が難しいのが現状である

7.3 フレームレートと解像度と処理能力

プロトタイプシステムは，MacBook Air (2.13GHz Core2 duo)で実行している．フレームレートは毎秒 30 フレームとし，カメラ画像からシルエット画像を生成して表示している．そして，フレームの大きさは，幅 640 ドット，高さ 480 ドットとした．

通常のプレゼンテーションシステムでは，一つのスライドを表示している間は，コンピュータは計算をしていないのであるが，本システムでは定常的に処理をすることが必要である．

実際にシステムを使用すると，演算装置の使用率は 80%を越える．多少の余裕はあるとはいえ，しばらくすると放熱ファンが回転を始める．研究という意味では，実現さえできれば実装上の効率は問題ではないが，シルエットを表示して効果を得るのに必要なフレーム間隔は毎秒 30 フレーム以下である可能性は大きい．評価実験を併用して，効果を維持したまま，シルエット生成のフレームレートを下げ，コンピュータの負荷を下げることの検討は行うべきことといえる．

7.4　コンテンツへの影響

本来ならば，どんなプレゼンテーション資料についても効果的に使えるシステムが望ましいが，現状では背景を利用するので，背景に濃い色を配置してあるコンテンツでは本システムは利用できない．同様に，全面に写真を利用したプレゼンテーション資料については，シルエット画像が表示されない．コンテンツの図や文字の配置や，要求されるアニメーションのレベルについては，通常のプレゼンテーションと差はないので，変更は比較的に容易と考えられる．写真を利用する場合には，適切に白い場所を用意することで，効果がでくように工夫することもできる．そのような例を図４に示す．これは，２枚の写真の間に空白の場所を意図的に要しておき，そこから写真の片方を指差しているような状況である．

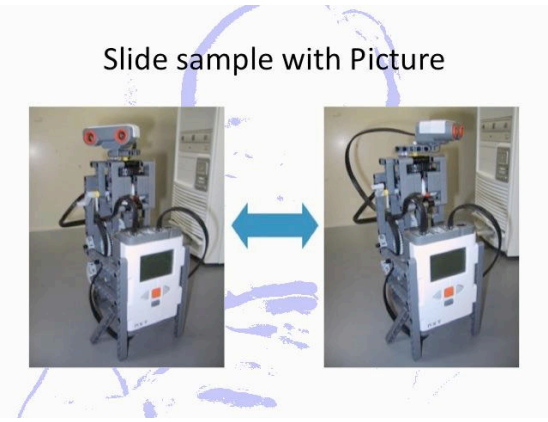

図４　写真を利用したときのスライド表示

7.5　プレゼンテーションの会場の影響

このシステムでシルエットイメージをつくる原理は，停止していて変化のないものを映さないことにある．逆に，変化するものは，人物でなくても映り込んでしまう．発表者の後ろにファンがあったり，定期的に変化するチャット画面がながれていたりすると，それ表示される可能性がある．また，会場が暗すぎて，カメラが動作しないことも考えられる．ただし，通常のプレゼンテーション会場で問題になることは少ないと考えられる．

もうひとつの懸案は，背景のシルエットに使う色である．色が濃すぎると，プレゼンテーションから注意を奪い，薄いと認識性がわるいということから，注意深く決定した．このような中間色については，プロジェクターやディスプレイの性質によって影響を受けやすいということも考えられる．

最後に，プレゼンテーションの会場が暗すぎて，内蔵カメラが十分に映像を取り込めない場合に，システム上の問題が生じることも考えられる．

7.6　システムの設計改善のための評価

このシステムは，シルエットの生成において，多くのパラメータがあり，その多くは作成者の直感によって決定されている．本来ならば，そのようなものを改善して，より効果の高いシステムに改良する必要があり，多くの比較実験が必要となる．そして，話者の姿が見える状態でも，シルエットを利用してかまわないようにすることは，システムが普及するために必要なことであると考えられる．

8.　まとめ

プレゼンテーションの配信・録画を念頭に，話者の映像をシルエットに変換してプレゼンテーションの背景に表示するプレゼンテーションツールを作成した．システムの要求分析を説明し，システムの構成と調整について説明した．最後に，話者が見えない条件を設定し，提案するシステムを使うことで効果がえられることを確認する実験を行った．

謝辞　本システムの設計の妥当性については，豊橋技術科学大学で学会のネットワーク配信を行うときに中心となって作業していただいている情報メディア基盤センターの入部百合絵助教に検討に加わっていただいた．ここに感謝したい．

参　考　文　献

1) Schiphorst, Thecla, Calvert, Tom W., Lee, C., Welman, Chris and Gaudet, S.: Tools for Interaction with the Creative Process of Composition. ACM CHI 90, Seattle, Washington,USA. pp. 167-174, (1990).
2) 梅村恭司，岡部正幸，手間の少ない講義録画を目的とした 4 チャンネル高解像度同時録画設備，情報処理学会研究会報告，OS 2005(48), pp157-158, (2005).
3) 渡辺富夫，身体的メディアによるメディア芸術創造支援，情報処理，Vol.48, No.12, pp.1327-1334, (2007).
4) 中森玲奈，青木貴司，椎尾一朗，ご近所知るえっと-身近い他人との緩やかなコミュニケーション支援-、日本ソフトウェア科学会研究会 WISS2009,pp193-194, (2009).
5) 前田晴己，黒沢祐也，栗原一貴，宮下芳明，MAEDE: スクリーン前でのプレゼンテーションスタイル，日本ソフトウェア科学会研究会 WISS2011,pp164-166, (2011).