コンピュータビジョンと
イメージメディア 106－2
（1997. 7. 24）

プレゼンテーションのための実写映像ハンドリング

山口敬人　戸田真志　川嶋稔夫　青木由直

violin@huie.hokudai.ac.jp

北海道大学 工学部
〒060　札幌市北区北13条西8丁目

あらまし　文書や図の実写映像に対するハンドリングが可能な、コンピュータ支援のプレゼンテーションシステムを提案する。このシステムにおいて、映像の入力はコンピュータ制御可能なアクティブカメラを用いて行われる。画像中における濃度分布および濃度変化の解析は、ズームが必要な領域抽出および適切なズーム値決定の指標となり得る。また、パン・チルトによって得られた画像群を、イメージモザイキングによって合成することにより、画像を統合的に扱うことが可能となる。本稿では文書などの実画像に基づく最適ズームの決定および、スキャニング画像の統合に関する手法について述べる。

キーワード　インテリジェントプレゼンテーションシステム、アクティブセンシング、イメージモザイキング

Live Image Handling for Computer-assisted Presentation System

Takahito Yamaguchi,Masashi Toda,Toshio Kawashima and Yoshinao Aoki

violin@huie.hokudai.ac.jp

Hokkaido University,Nishi 8,Kita 13,Kita-ku,Sapporo

Abstract　We propose a computer-assisted presentation system which can handle live images of documents and drawings. In the system, documents are input using a computer-controlled camera; the camera can actively determine zoom parameters and viewing angles from the gray-level histogram of viewing area in order to capture the whole document image in sufficient resolution. Scanned images are mosaiced into a large image. This report shows the algorthm to select optimal zoom parameter from document images, and the method to combine adaptively scanned images.

key words　Inteligent Presentation System,Active Sensing,Image Mosaicing

1　はじめに

近年、画像や音声など様々なメディアを統合することにより、マルチメディアプレゼンテーションを可能にするシステムが開発されてきた。また、個人レベルで、各メディアを統合したマルチメディアスライドを作成するパッケージソフトも販売されている。これらのツールはあらかじめ取り込んだ各データを加工・編集することにより、高度なプレゼンテーションを実現する。

その一方で、コンピュータへのデータ取り込み作業やソフトウェアスライド作成の手間を省いて、手持ちの資料などをそのまま提示したいという要求がある。また、こうした実画像を用いるプレゼンテーションにおいては、講演の際に、講演者と提示資料とのインタラクションが重要な場面も多いが、あらかじめ取り込まれたデータではその対応が困難である。このような場合、カメラを用いた観測ベースのプレゼンテーションシステムが必要となる。

観測ベースのプレゼンテーションにおいて提示される、カメラからの映像は、適切なハンドリングを施すことにより、情報の高度化が図られ、視聴者への効果的な情報呈示を可能にする。我々は現在、各種書類など、特に紙媒体を基本とした文書や図の実写映像ハンドリングに対する計算機の支援を考えることにより、効率的且つインテリジェントなプレゼンテーションを可能とする提示システムの開発を目指している（図1)。

このようなシステムを実現する上で、

1. 効果的な映像の取得
2. システムーユーザー間のインタラクション
3. 文書解析による映像のコンテンツ化
4. コンテンツに基づく他情報とのリンク

といった課題が考えられる。この中でも、第1の課題に関しては、得られた映像がその他全てのハンドリング部分の入力となるため、非常に重要な技術であり、我々の目指すシステムの基盤となる。本稿では、この問題を解決するためのハンドリング手法について検討する。

プレゼンテーションのための映像入出力システムとしては、ＣＣＤカメラを用いた提示装置が広く普及している。教材提示装置をはじめとするこのようなシステムは、一般にＮＴＳＣ対応の単眼固定ＣＣＤカメラを用いていることから、カメラの画素数に制限がある。このため講演者は、解像度を上げるためのズームの調整および、それに伴う提示部分の位置合わせを強いられることになり、このことはプレゼンーションの品質や効率を著しく低下させる。

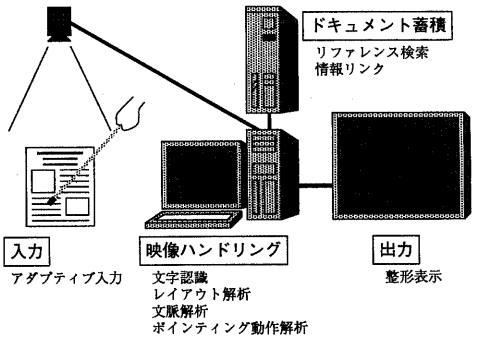


図 1: インテリジェントプレゼンテーションシステム

この問題を解決するために、解像度の高いハイビジョンカメラを用いることが考えられる。最近、メーカーからハイビジョンカメラを搭載したプロジェクターが発売されているが、現段階ではコストが高い。

こうしたことから、比較的安価で且つ高度なプレゼンテーションシステムを実現するためには、ＮＴＳＣカメラ画像のハンドリングに基づく、アクティブな映像の取得が重要になると考えられる。そこで我々は、ズームと視線方向の制御が可能な首振ＣＣＤカメラを利用し、画像群の統合によるシステムの構築を図る。

2　システムの概要

今回実現するシステムの概略を図2に示す。原稿台の上に置かれた資料原稿を、首振ＣＣＤカメラで観測し、その映像をシステムに送る。システムは送られてきた映像のハンドリング結果に基づいて必要な制御をカメラに送り、それによって得

られる画像を整形してディスプレイに表示する。

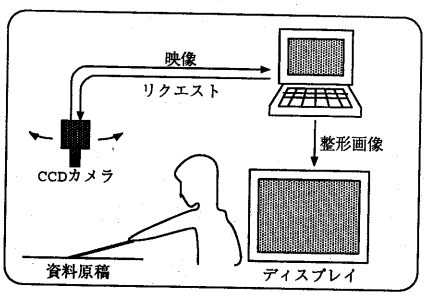


図 2: 目指すシステムの概略図

図 3はシステム全体における処理の流れを表している。

まず、提示資料全体を撮影した画像中からズームアップが必要な領域を抽出する。これらをズーム解析に基づいて空間的にクラスタリングし、クラスタリングされた各領域ごとに必要なズーム値を決定する。決定されたズームの下では、領域が1枚の画像に入り切らない場合が多いため、カメラのパン・チルトによる領域のスキャニング及び得られた画像群の統合により、合成画像を生成する。

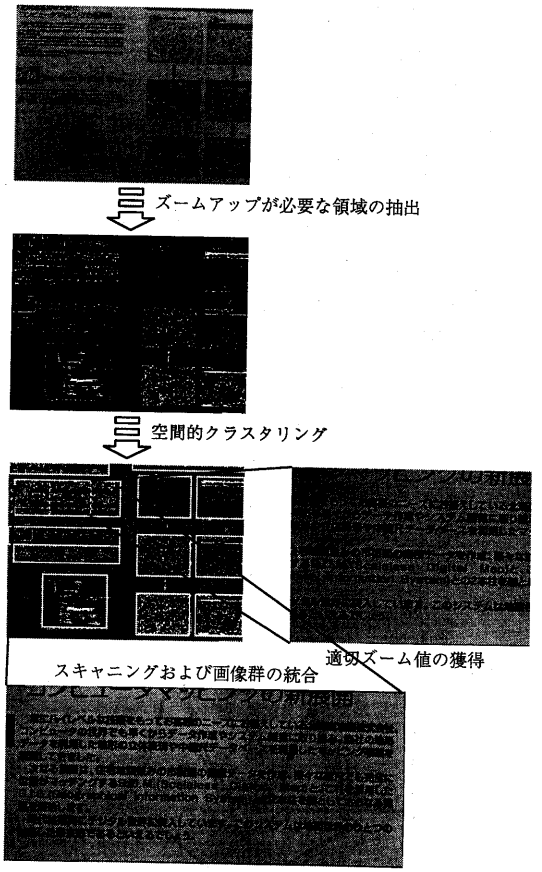


図 3: システムにおける処理の流れ

3 システム実現の具体的手法

3.1 ズームアップが必要な領域の抽出

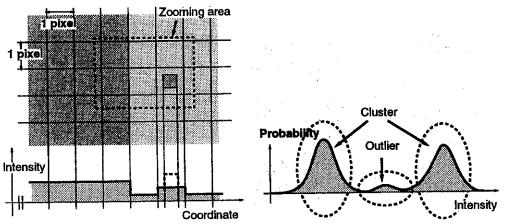
3.1.1 濃度分布に基づく孤立はずれ値点解析

注目する領域に対するズームが不足している場合、その領域内にある文字や記号といった構成要素は、背景や他の構成要素と融合もしくは埋没してしまうため、その領域に対応する画素の濃度値は、本来とは異なる値を示す（図4：状態1）。さらなるズームアップに伴い、注目する領域内の要素は、画像空間内で十分な面積を有し、本来の濃度値を示すようになる（図4：状態2）。

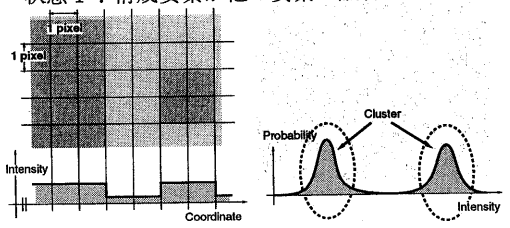
ここで、画像全体における濃度分布を調べ、その分布に基づいてクラスタリングを行なうと、注目する画素の濃度値は、状態2の時には主要クラスタとして知覚されるが、状態1の時には、はずれ値として取り扱われることが予想される。

我々は、画像中濃度の出現確率分布を、最尤法に基づいたロバストクラスタリングに適応フィッテイングの処理を加えた手法を用いることにより、主要特徴とはずれ値とに分離する（図5）。

また、文字の輪郭など、各構成要素の端点の濃度は、要素本来のものとは異なる値を示すため、はずれ値として抽出されてしまう。そこで、はずれ値となる濃度を有する画素（以下はずれ値点）から、このような端点を除外する処理を施す（図6）。こうして最終的に得られた画素（以下孤立はずれ値点）に基づいて、ズームアップが必要な領域を抽出することができる [1]。



状態１：構成要素が他の要素と融合している



状態２：構成要素が知覚される

図 4: ズームの変化に伴う画像状態の推移

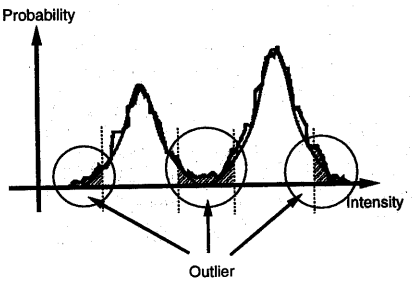
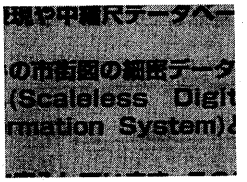
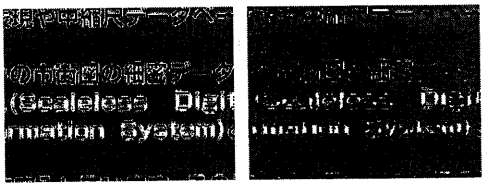


図 5: 最尤法に基づくロバストクラスタリング



原画像



はずれ値点　　孤立はずれ値点

図 6: 孤立はずれ値点の抽出

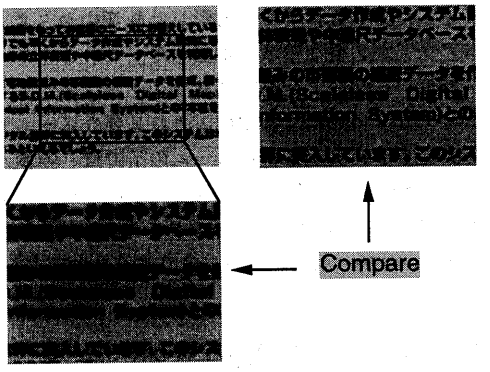


図 7: ズームに伴う濃度変化量の比較

3.1.2　局所的な濃度変化量の解析

前述の手法は、画像全体における濃度分布を考えるため、局所的には主要クラスタとして知覚される濃度値が、画像全体に占める面積の割合が小さいことによって、はずれ値としてとり扱われてしまう場合がある。

ところで、ズーミングに対する対応点の濃度変化に着目すると、ズームが十分になった段階で、対応点間の濃度が変化しなくなると期待できる。そこで、ズームに対して、各画素における濃度の変化量が大きい場合は、ズームが不足していると考えることにする。この処理は濃度を局所的に取り扱うため、上述したような問題が起こらない。

ズームアップに伴って得られる各画像に関して、１つ前の画像を現画像に合わせて拡大したものとの比較を行ない（図7）、その差分画像を生成する。差分画像における各画素を、前述のアルゴリズム同様の手法を用いて、主要特徴とはずれ値とに分離する。

3.1.3　実験

実画像に対して２つのアルゴリズムを適用し、はずれ値を検出することによって、それぞれズームが不足している画素の候補が抽出される。抽出さ

れた候補に関して、その論理積を取ることでズームアップが必要な画素を決定した（図8）。

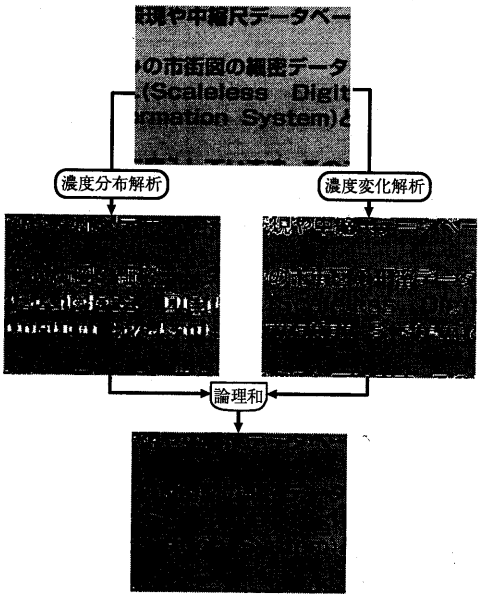


図 8: 2つのアルゴリズムの融合によるズーム不足領域の抽出

3.2 画像情報の空間的クラスタリング

ズームアップが必要な画素が抽出された画像に対して、垂直、水平それぞれの方向に走査することで、抽出画素数に関するヒストグラムを生成する。これにメディアンフィルタをかけ、行間など細かい境界部分を除去した上で、ヒストグラム分布に基づく空間的クラスタリングを行なう（図9）。

3.3 適切ズーム値の獲得

前述のアルゴリズムによって得られる、ズームアップが必要な画素数に関して、ズームの変化に伴う推移をモニタリングし、停滞開始時をもって、適切なズーム値と考える。

各ズームの下でズームアップが必要な画素を抽出した結果を図 10 に示す。白い部分がズーム不足の画素であり、その数をモニタリングした結果が図 11である。

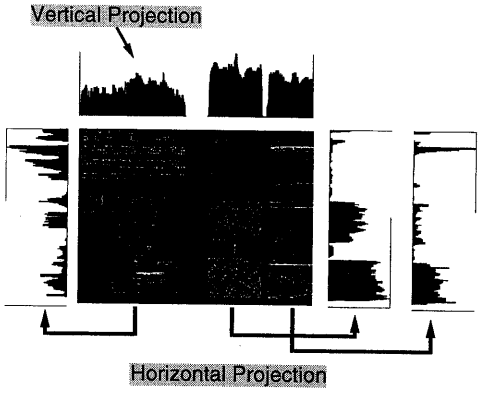


図 9: 空間的クラスタリング

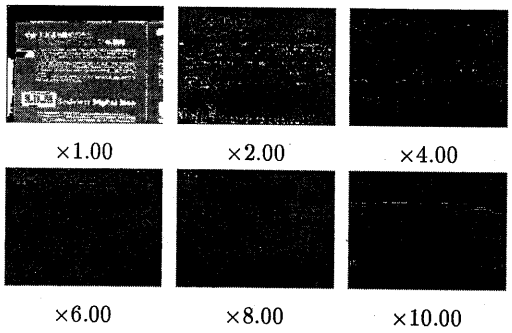


図 10: ズームアップが必要な画素の抽出

3.4 画像の統合

一般的なＣＣＤカメラを用いる場合、情報を認識するのに必要なズーム値の下では、十分な観測領域を確保できない場合が多い。我々は視覚デバイスとして首振ＣＣＤカメラを用いているため、カメラをパン・チルトさせて必要な領域を観測することが可能である。

このようにして得られた画像群に関して、各画像間の対応付けを図り、それに基づいて画像統合を行なう。出力デバイスとして高解像度のディスプレイを用いた場合、合成画像を表示することによって、視聴者に対してより多くの情報を一度に提示することが可能となる。また、統合された画像を格納しておくことは、その領域におけるデータの扱いを容易にする。さらに、インテリジェン

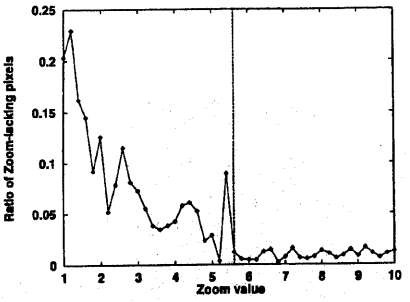


図 11: ズームの不足している画素数の推移

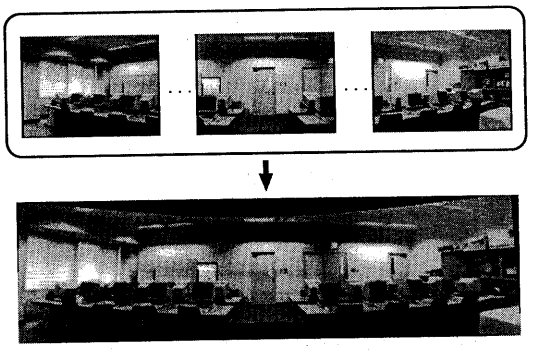


図 12: イメージモザイキング

トプレゼンテーションシステムにおいては、文字処理や図形処理などを行う上で重要である。

ここで用いられる画像統合の手段としては、必要領域に関して大きさや形状の制限を受けないような手法が望ましい。このような条件を満たす手法としてイメージモザイキングがある。

3.4.1 イメージモザイキング

イメージモザイキングとは、視点の異なる画像群から、元のシーンをモザイク集合として再構成する技術である（図 12)。画像間の動きをパラメータとして推定し、そのパラメータによって画像間の対応付けを図ることが出来る（[2][3])。そのアルゴリズムを以下に述べる。

視点の異なる 2 画像間の動き $\mathbf{u}$ がパラメータ $\mathbf{m}$ によって記述されるとすると、ピクセル i における濃度差分 $e_i = I'(\mathbf{x}_i + \mathbf{u}(\mathbf{m})) - I(\mathbf{x}_i)$ の 2 乗和 $E = \sum_i e_i$ に関する最小化問題を解くことにより、$\mathbf{m}$ を推定する。

$\mathbf{m} = (m_0 \ldots m_k)$ に関する近似ヘッセ行列及び勾配を、それぞれ $\mathbf{A}$,$\mathbf{b}$ とすると、勾配方向のパラメータの変化$\Delta\mathbf{m}$ は (1) 式のように記述できる（$\mathbf{A}$,$\mathbf{b}$ の成分は (3) 式に示す)。求められた$\Delta\mathbf{m}$ を現段階の試行解に加えたものを新たな試行解とし（式 (2))、再び同じ処理を実行。これを収束するまで繰り返すことによってパラメータの当てはめ値が求まる。

$$\Delta\mathbf{m} = \mathbf{A}^{-1}\mathbf{b} \qquad (1)$$

$$\mathbf{m}^{t+1} = \mathbf{m}^t + \Delta\mathbf{m} \qquad (2)$$

$$a_{kl} = \sum_i \frac{\partial e_i}{\partial m_k}\frac{\partial e_i}{\partial m_l} \quad b_k = -\sum_i e_i \frac{\partial e_i}{\partial m_k} \qquad (3)$$

3.4.2 動きモデルの選択

イメージモザイキングで用いられるような大域的動きモデルとして、代表的なものには次のようなものがある。

1. 平行移動モデル
2. アフィンモデル
3. 透視投影モデル

現在のシステムにおいては、カメラと撮影対象との距離に比べ、カメラの回転角が十分小さいことから、自由度の低い平行移動モデルでも対応することができると考えられる。しかし我々は、もっと自由な環境で動くシステムを考えているため、対象がより近くに置かれる可能性を考慮する必要がある。この場合、スキャニングにおけるカメラの回転角がより大きくなることが予想され、それによって生じる視覚的な歪みに対応できなくてはならない。このようなことから我々は画像間の動きを記述するモデルとして、平面の 3 次元的な変換が可能な透視投影モデル（式 (4)、ただし $\mathbf{u}(\mathbf{m}) = (u, v)$）を用いることにする。

$$\begin{aligned} u &= \frac{m_0 x + m_1 y + m_2}{m_6 x + m_7 y + 1} - x \\ v &= \frac{m_3 x + m_4 y + m_5}{m_6 x + m_7 y + 1} - y \end{aligned} \qquad (4)$$

3.4.3 パラメータ初期値の設定

パラメータ $\mathbf{m}$ 推定のアルゴリズムにおいて、収束時間はパラメータの初期値に大きく依存する。また、画像間の動きが大きい場合、収束途中で局所解に陥ってしまう可能性が非常に高い。我々は、カメラの動きに基づいてパラメータに初期値を与えることにより、これらの問題を回避する。

現在我々が用いているアクティブカメラは、計算機からメッセージを送ることにより、その動き（パン・チルト・ズーム）を１０ビット値で取得することが可能である。そこで、実際にプレゼンテーションを行なう前に、あらかじめ簡易的なキャリブレーションを行なうことにより、画像上での動きとカメラ制御系との対応関係をモデル化しておく。ここでは、画像上での動きとカメラの回転とが、ズームに依存した線形関係にあると仮定し、モデルを以下のように設定、モデルパラメータを実験的に求めた。

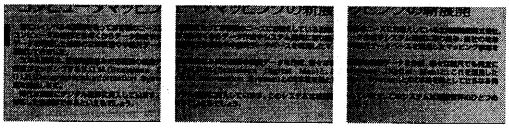
$$
\begin{aligned}
truezoom &= P_0 exp(P_1 zoom + P_2) \\
t_x &= truezoom(P_3 pan + P_4 tilt) \\
t_y &= truezoom(P_5 pan + P_6 tilt)
\end{aligned}
$$

$(P_0 \ldots P_6)$ はモデルパラメータ、$pan, tilt, zoom$ はカメラの動きを表す１０ビット値、t_x, t_y は画像平面上の動きをそれぞれ表している。実際のデータに対してイメージモザイキングを行なう際、上記のモデルを用いてカメラの動きを画像上の動きに変換し、その値を初期値にすることによって、収束の安定化、高速化を図ることができる。

3.4.4 実験

ズーム解析に基づき空間的にクラスタリングされた領域を、必要なズームの下でスキャンすることによって得られた画像群のサンプルを図13に示す。これら１０枚の画像を前述のアルゴリズムを用いて合成した（図14）。

視覚デバイスとして、ソニー製の首振ＣＣＤカメラ EVI-D30 を用いた。このカメラはホストコンピュータから、RS232-C インタフェースを介して、雲台のパン・チルト、カメラ部のズーム、フォーカス、アイリスを制御することが可能である。また、カメラからの映像はキャプチャーボードを用いてホストコンピュータに取り込まれる。



フレーム０　　フレーム５　　フレーム９

図 13: サンプル画像

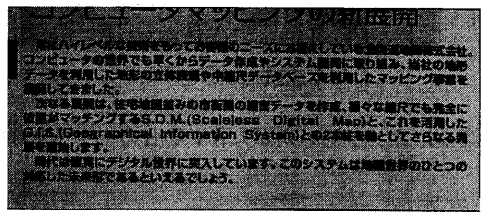


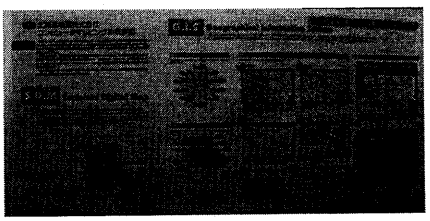
図 14: モザイク合成画像

この例においては、カメラの回転角が非常に小さいため、平行移動モデルあるいはアフィンモデルを用いた場合も、視覚的にほぼ等しい画像が生成された。現システムでは、カメラのメカニカルな制限のため、対象をこれ以上近くに置くことは不可能である。そこで、対象との距離がより近い場合を想定して、比較的大きな文書を撮影し、実験を行なった。以下にその結果を示す。図 15.a は平行移動モデルを用いて合成を行なった結果である。カメラの回転角が大きいことによる、パースペクティブな歪みが生じている。これに対し、透視投影モデルを用いて生成された合成画像においては、このような歪みがある程度補正されているのが分る（図 15.b）。

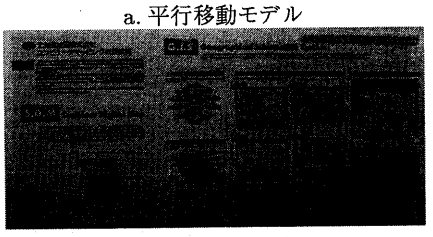
透視投影モデルを用いた場合においても、完全に歪みを取りきれていないのは、各画像においてレンズ周差による歪みが生じているためである。

4 終わりに

カメラからの映像をハンドリングすることによって実現される、インテリジェントなプレゼンテー



a. 平行移動モデル



b. 透視投影モデル

図 15: カメラ回転角が大きい場合のモザイク合成

ションシステムを提案した。システムの基盤となる映像入力部における要素技術として、

1. ズームアップが必要な領域の抽出
2. 紙面の空間的クラスタリング
3. 適切ズームの決定
4. 必要領域の網羅と画像統合

に関する手法について述べ、それぞれ実画像を用いて実験を行なった。

今後の課題として、

1. 原稿のトラッキング
2. ポインティング動作の認識
3. 映像のハイパーテキスト化

などが挙げられる。

参考文献

[1] 戸田 真志, 川嶋 稔夫, 青木 由直, "視覚ズームにおける能動的観測", 第 3 回画像センシングシンポジウム (SII'97) 講演論文集, pp.265-270, 1997.

[2] R.Szeliski, Image Mozaicing for Tele-Reality Applications. Technical Report CRL 94/2, DEC Cambridge Research Lab., 1994.

[3] R.Szeliski, Spline-Based Image Registration. Technical Report CRL 94/1, DEC Cambridge Research Lab., 1994.