



## 2008 年度下期未踏 IT 人材発掘・育成事業 採択案件評価書

### 1. 担当PM

笈 捷彦 PM(早稲田大学 基幹理工学部 情報理工学科 教授)

### 2. 採択者氏名

チーフクリエイター: 本田 浩一郎(東京大学 大学院情報理工学系研究科  
コンピュータ科学専攻)

コクリエイター: なし

### 3. プロジェクト管理組織

株式会社 創夢

### 4. 委託金支払額

3,000,000 円

### 5. テーマ名

大規模な 3 次元写真データの編集システムの開発

### 6. 関連Webサイト

なし

## 7. テーマ概要

### ■ 背景

近年、インターネットの高速化やストレージの大容量化に伴い、写真共有サイトが発達してきている。そしてアップロードされた写真の一部は、再利用可能な著作権の形態によって公開されており、誰にでも編集が可能になっている。こうしたウェブ上の資源を有効活用することは、従来のコンテンツ制作とは違う、新たな編集スタイルともいえるだろう。だが、取得した大量の写真进行处理しようとするとき、一貫性をもった処理をする必要が出てくると考えられる。既存のシステムでは、主に一枚の写真に対して編集を行うように機能が設計されており、大量の写真进行处理するような場合には向かない。

提案するシステムでは、主に複数枚の写真进行处理するための手法を提案、実装を行う。

具体的には、写真群を3次元位置を推定して空間上にマップし、その上で編集を行うことで、一貫性を保った処理を行う。

### ■ 作成するもの

複数の写真に対して、3次元的な位置を計算し、その位置関係を考慮して一貫性のある編集を行う。

主な機能としては次の2つを考えている。

#### 1. 3次元オブジェクトの挿入

3次元モデルを各写真に合成する。

写真の向きと画角によって、うつり方が変わる。

#### 2. 編集の伝播

写真合成などの処理を一括して行う。

一枚の写真のある部分に処理を行おうとすると、他の写真の対応する部分に整合性のとれるような処理が行われる。

## 8. 採択理由

本田君は、ボランティア活動など多彩に行なっているし、学部はシステム創成学科で管理工学を学んでいたというように、守備範囲がなかなか広い。その提案は、Flickrサイト(写真共有10億枚以上)などの中にcreative commonsで公開されている多量の画像を活用する「新編集」技法の道具を作りたい、というものである。その主要なものは、同一被写体の一群の画像から3次元モデルを自動生成すること、そのモデル

に対して様々な操作(他物体の追加、一部物体の除去など)が行なえること、特定対象モデルに対して施した操作を他の同様な対象に適用できるようにすること、である。第1の機能についてはデモ実演ができるところまで試作が進んでいる。デモでできたことを実用に堪えるレベルにまで仕上げるには相応の工夫と努力を要するが、新しい画像編集ソフトウェアとして広く使ってもらえるものを目指して仕上げてくれることを期待する。

## 9. 開発目標

デジタルカメラが普及するにつれて、インターネット上でのユーザ発信型のコンテンツ共有サイトが急激に増えてきている。このような環境の中では、コンテンツを制作・編集するツールが重要な役割を持つてくる。

ところが、写真の編集ツール自体は、あまり変化してきていない。1枚の写真に対する編集でも、一般的な写真編集ツールでは、難しいものが存在する。まして、いくつかの関連した写真を同時に編集する場合、一貫性を保って編集を行うのは非常に難しい。同じ物体を撮った写真を編集するとき、1枚の画像のある点が、他の画像のどの点に対応するのかを、常に考えながら編集をしなければならない。このように、複数の写真の中の「同じ位置」を考慮して変形を行う場合、従来のツールでは飛躍的に難しい編集作業となってしまう。既存のツールが、主に1枚の写真に対して編集を行うように機能が設計されているからである。複数の写真に対しての処理として実装されていても、画像の拡大縮小や、色調補正など、単純な処理を繰り返すバッチ処理くらいにとどまっている。このプロジェクトでは、このような複数の写真に対する、効果的な写真編集ツールを制作することを目的とする。単純なバッチ処理を超えるより複雑な処理を、複数画像に渡って「一貫性を保ちながら」効率的でかつ正確に行えるシステムを実現する。そのシステムの機能概要を図1に示す。

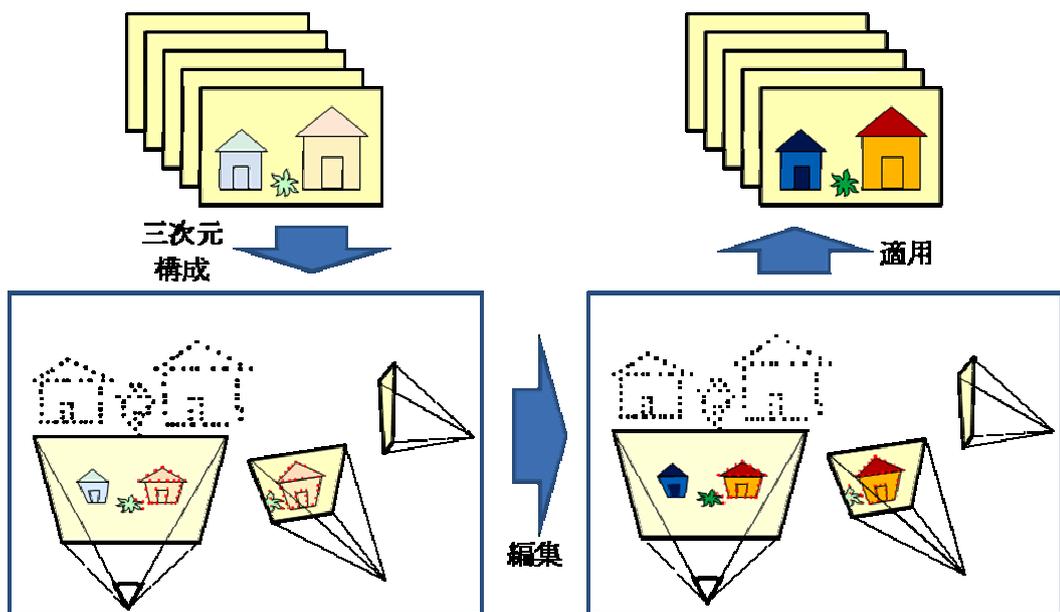


図1 システムの概要

このプロジェクトでは、はじめに特徴量の抽出と3次元構成を実装し、その後に編集の伝搬機能を実装することを目標とした。伝搬させる具体的な編集項目として、つぎの二つのものを実装することを目標とした。

- ・ モザイク合成(写真の合成) (図2)
- ・ 3次元ポリゴンモデルの写真への写り込み (図3)

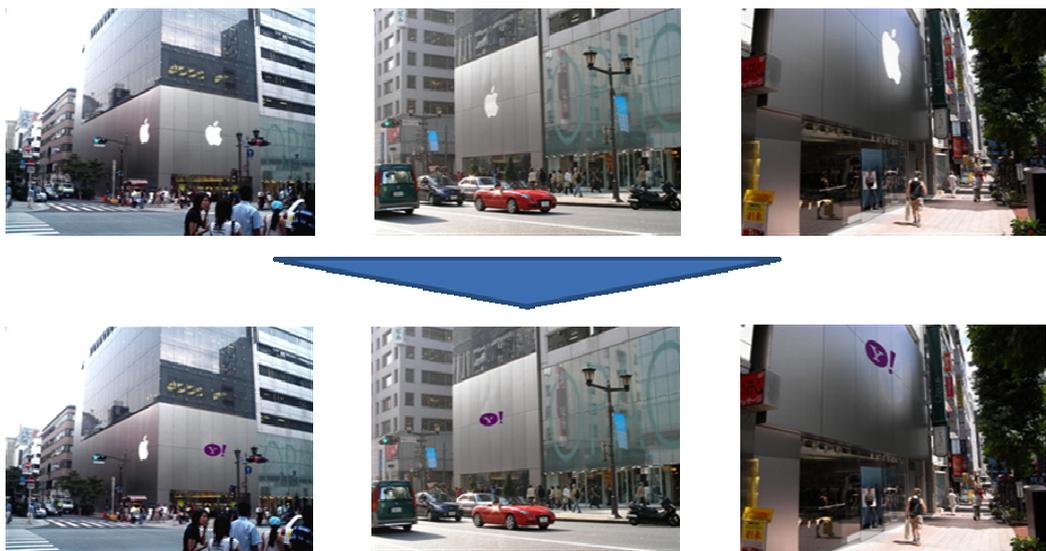


図2 モザイク合成のイメージ図

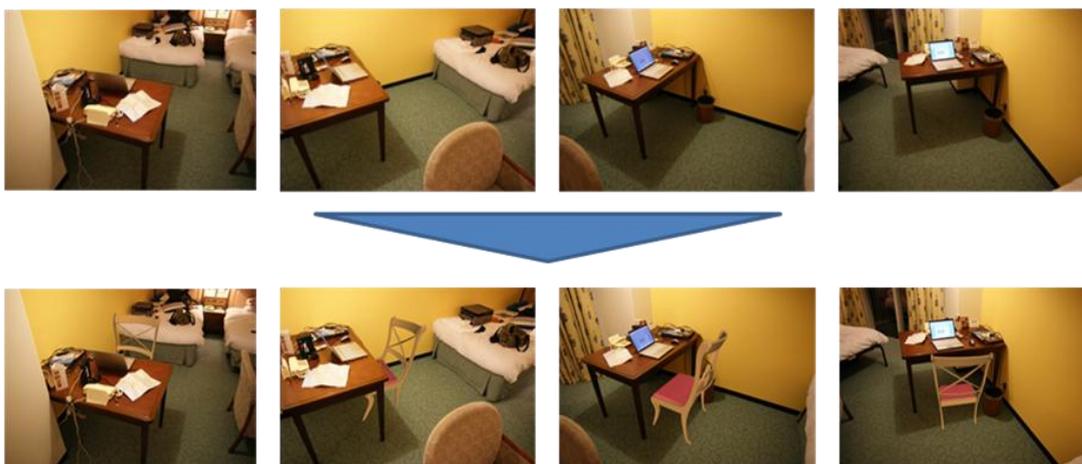


図 3 3 次元ポリゴンモデルの写真への写り込みのイメージ図

## 10. 進捗概要

2 月いっぱいにかけて、クリエイターは、目標を達成するべく、3 次元情報の復元に資すると思われる諸技術を幅広く調査した。その結果、オクルージョン(視点から見て、ある物体が他の物体を隠蔽することを指す。)をすべての場合にアルゴリズム的に解決することは困難であると判断するに至った。その結果を踏まえてブースト会議で様々な示唆を受け、実装としてはモザイク処理の伝搬に集中して取り組むことにし、2 次元画像のテクスチャマップやマッチングの実装に取りかかった。4 月は、前半を国際会議への出席に時間をとられたこともあり、実作業としてはすでに着手した部分のデバッグやリファクタリングにあて、各種技法の学習に時間を割いている。5 月にもまた国内学会出席があり、3 次元位置の対応付けやテクスチャマッピングに合わせての編集伝搬の実装を進めるもなかなか進まない状況にあった。6 月下旬の中間発表会に向けてなんとかモザイク合成の編集をまとめるところまでこぎ着けたが、システム細部の仕上げに時間がかかり、この編集システムを使えば何ができるのか、何が高度に行えるようになるのかの適切な例を探すこともままならなかった。7 月に入って、ようやくモザイク合成の編集ができるシステムが出来上がったが、ついに 3 次元ポリゴンモデルの写真への写り込みについてシステムの実現には取りかかることができなかつた。

## 11. 成果

目標設定した課題に対して様々な調査と試行の結果から、つぎの通りの方式を採用した。

- ・ 特徴点の記述 特徴点記述には、最近考案された SIFT (Scale-Invariant Feature Transform) を採用している。画像の拡大縮小や、回転についての耐性が非常に高く、異なる画像間のマッチングに適している一方で、計算量が多いという特徴をもつ。開発するシステムでは、3次元構成を復元するのに一度計算してしまえば画像編集には再計算する必要がないことから、計算量の多さがさほど欠点とはならないので、採用したものである。
- ・ 特徴点のマッチング これには、最近傍探索を採用した。計算を高速に行う必要から、近似最近傍探索で代替している。
- ・ 3次元情報の構成 マッチングのとれた2枚の画像から3次元情報を構成する(これを弱校正という)手法としては、8点アルゴリズムが従来から使われているが、原理的には5点あれば計算可能であり、ライブラリも入手できたので、5点アルゴリズムを使っている。

さらに、画像群すべてに対して3次元的に対応関係を取って組み立てるには、Structure from Motion と総称されている方法が知られている。各画像対について弱校正を行い、その結果を非線形の最適化計算によって洗練する(束調整という)というものである。

ところで、このプロジェクトで扱う画像上の特徴点が、構成し復元しようとする3次元物体の手前にあるのか、後ろにあるのか、はたまた内部にあるのかは未知である。しかも、そうした情報なしでは2次元の写真に射影することができないので、何とかして、前後関係を考えなければならない。特に3次元モデルの挿入では確実にこの問題の解決が必要になるので、早期に解決しなければならなかった。クリエイターは、初期段階でにあつては、ある程度まとまった点群に対してクラスタリング(分類)を行い、点群からメッシュモデルを生成する方法で解決できるのではと期待していたようである。メッシュは、凸包として生成してもいいし、MLS法を使っていいだろうと想定していたらしいが、実際に実装を行ってみたところ、空間中に存在する特徴点数が想定よりも遙かに少なく、メッシュ生成は困難であることが判明した。それでもなお解決策を求めてクリエイターは模索を続け、Multi-view Stereo とよばれる技術を使って、非常に計算時間はかかるが写真群からメッシュモデルを直接生成してしまう方法が有力であることを見つけるところまでは進んだものの、3次元ポリゴンモデルの写真への写り込みへの編集システムの実現に取りかかる時間はもはや残されていなかった。

出来上がったシステムは、モザイク合成の編集を行うものである。そこでの画像からの特徴量の抽出と3次元情報の構成は、図4に示すものとなっている。

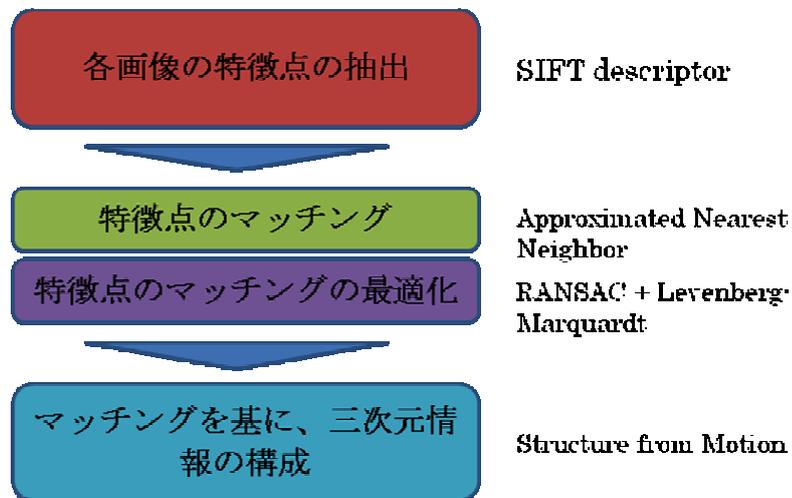


図 4 特徴量抽出と 3 次元情報の構成

ここでは、特徴点の記述に SIFT を用い、2 画像に対する弱校正において、特徴点のマッチングをとるのには近似最近傍 (approximated nearest neighbour) 探索を用いた上で、5 点アルゴリズムの解について RANSAC を用いてより良い解を求めた上で、Levenberg-Marquardt 法を用いてさらに洗練する方法を取っている。画像群に対しては、次の画像に対してすでに構成した 3 次元情報との間で弱校正を行い、その都度 Levenberg-Marquardt 法を用いてさらに洗練するという、逐次的な束調整で済ませている。

図 5 に SIFT による特徴点とそのマッチングの例を示す。



図 5 SIFT 特徴量記述子によるマッチング

3次元情報の構成を基に、モザイク合成の編集機能を実装した。ある写真上で、写真の合成を行うと、その情報が他の画像に伝搬し、一貫性を保ってテクスチャが貼付されるというものである。そこでの位置伝搬の概要を図6に示す。

ユーザがいずれかの画像をクリックすると、クリックした点の近傍の特徴点(ユークリッド距離で最も近い特徴点)3点を選び出し、クリックされた位置と特徴点との位置関係を保存する。そして、空間上で、保存した位置関係から場所を推定し、他画像においても、同様の推定を施す。モザイク合成には最低4点の制御点が必要であるので、ユーザに4点以上を指定してもらい、それぞれの点について同様の処理を行う。

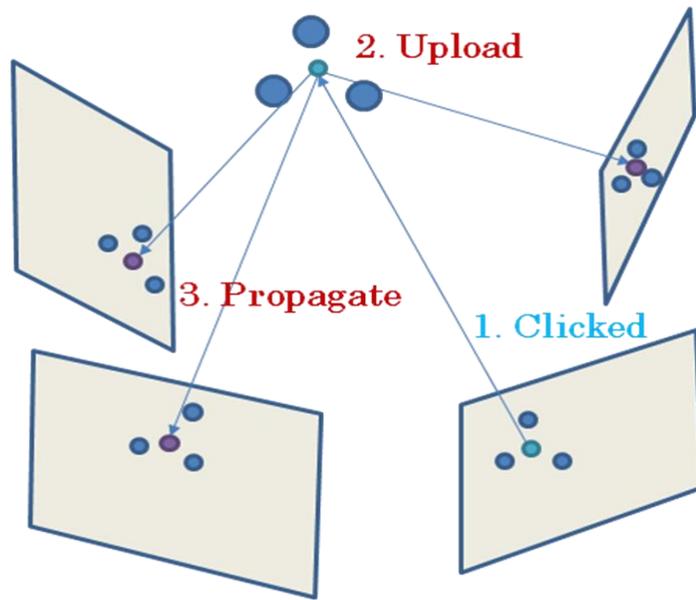


図 6 位置伝搬の概要

このシステムを使って、ドアに貼られたポスターを別のポスターに差し替える編集を行った例を示す(図 7～図 11)。

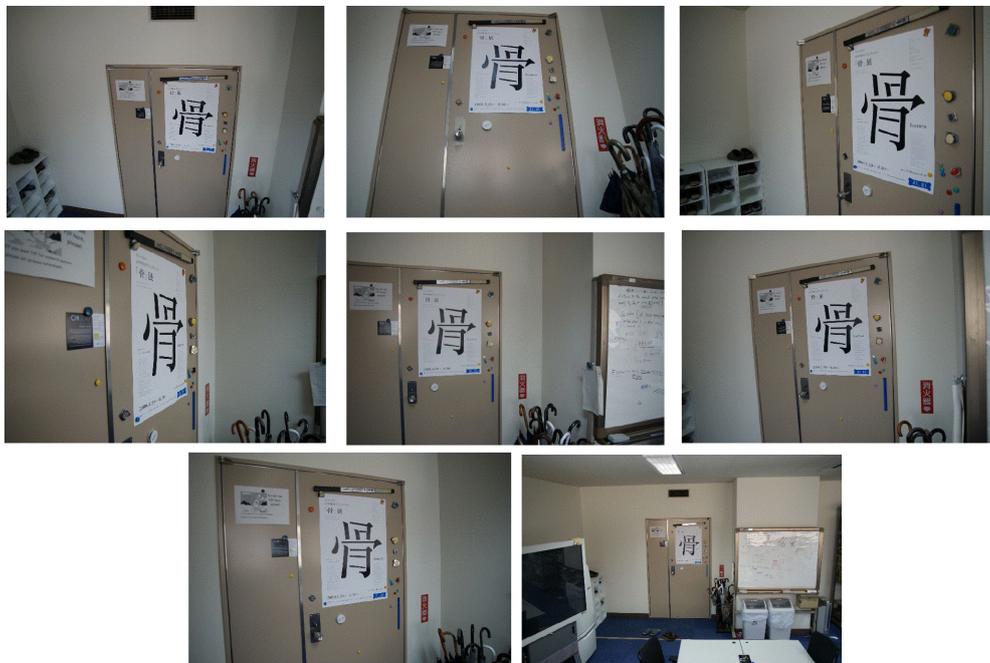


図 7 ドアに貼られたポスター

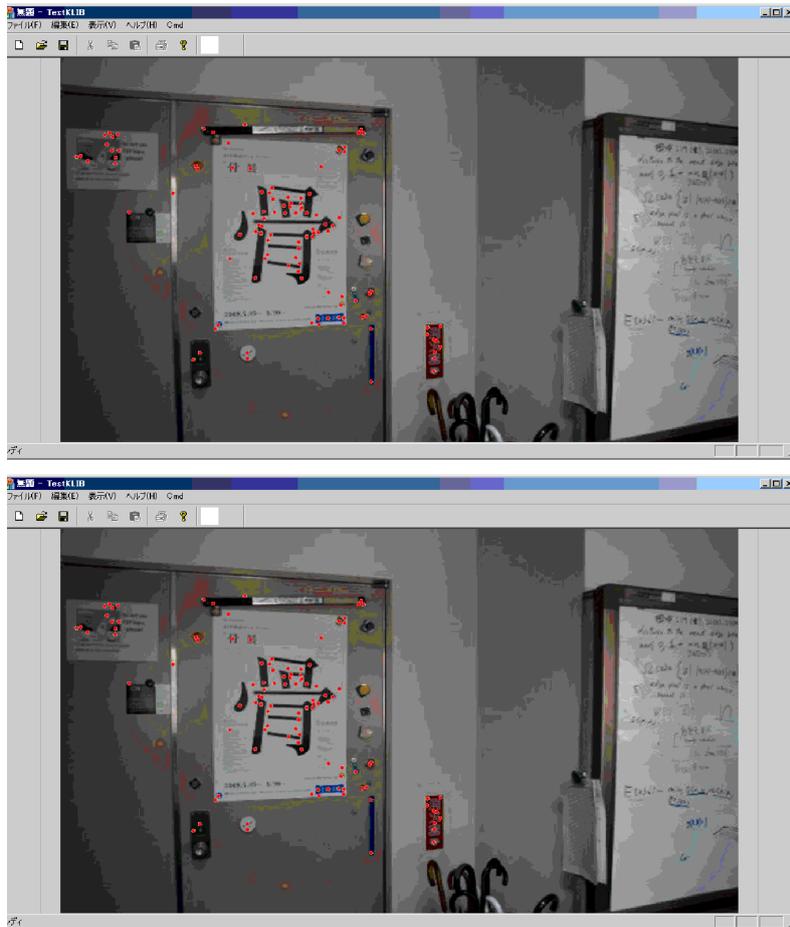


図 8 マッチングされた SIFT 特徴点



図 9 3次元構成されたポスター



図 10 テキスチャ(ポスター)の編集

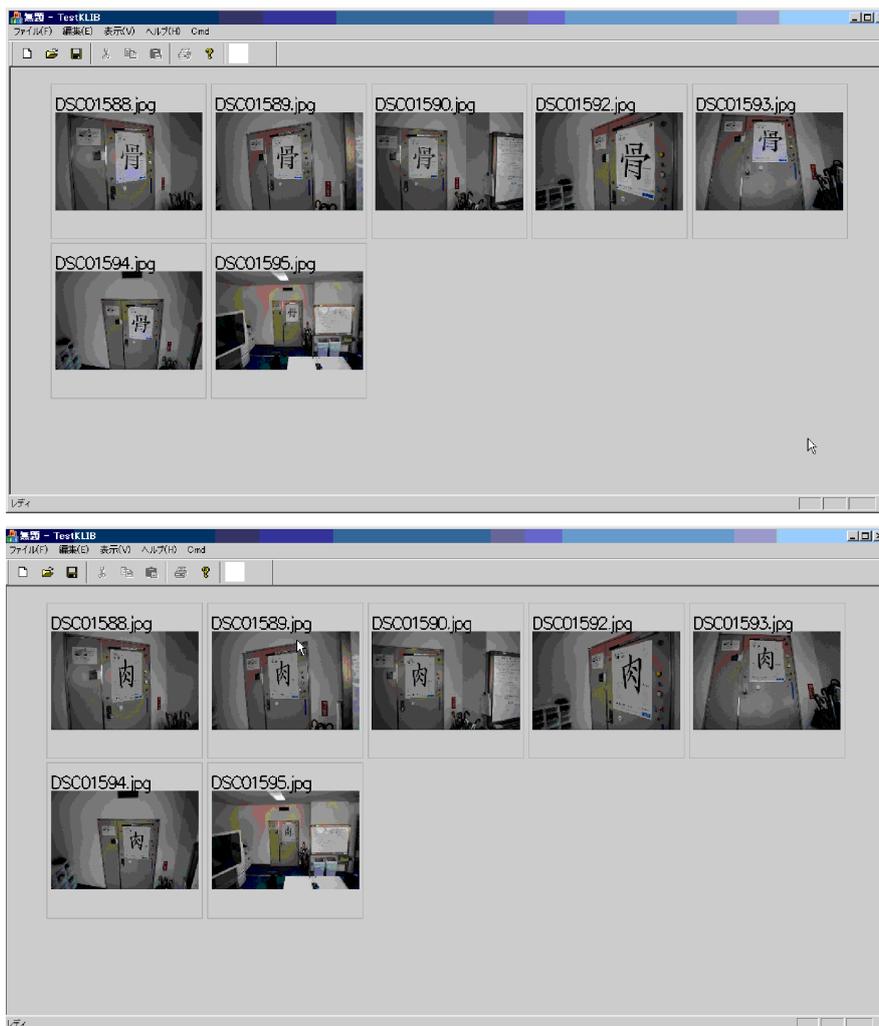


図 11 モザイク合成の編集結果(ポスターの差し替え)

## 12. プロジェクト評価

クリエイターは、なかなかの勉強家であり、各種実験を重ねながらソフトウェアをくみ上げていく力も相当のものを持っている。ただ、今回のプロジェクトについていえば、見通しが外れた、ということに尽きる。提案の時点から、モザイク合成編集と、3次元ポリゴンモデルの写真への写り込み編集という、相当に異なる二つのことがらを結びつけてのプロジェクトには PM も含め、多くの人が首をひねっていた。ただ、クリエイターは、それなりにプロジェクト提案に先立っての知識と経験から、両者を通じて3次元情報の復元ということに共通性を見、その難しさも知ったうえで、適当なクラスタリングによって十分に対応できそうだと踏んでいたようである。

しかし、未踏ユースの開発期間は、少々クリエイターにとって短すぎたようである。途中で PM からの示唆も含め、3次元ポリゴンモデルの写真への写り込み編集の方は後回しにしたものの、なお様々に調査を重ね、実現可能性の高い Multi-view Stereo 方式を見いだしたが、この開発期間中に実現するには及ばなかった。

未踏ユースプロジェクトの中では仕上げることはできなかったものの、クリエイターの力をもってすれば、Multi-view Stereo 方式による試行を行い、実用に耐える形での3次元ポリゴンモデルの写真への写り込み編集システムに仕上げることも不可能でないはずである。

一方で、モザイク合成編集については、一通り使える形のシステムを仕上げる事ができた。システムを仕上げるという点ではかろうじて合格点ということになるだろう。

## 13. 今後の課題

モザイク合成編集システムについては、さらに洗練し、使い勝手のよいものに仕上げ世の中に押し出してほしい。それには、写真中のポスター差し替えに類する編集例に加えて、なるほど、すばらしいことができるものだと多くの人が感心してくれる例を幾つも見つけて、そのどれに対しても要領よく編集ができるようにシステムを改良することが必要となる。とくに、ユーザインタフェースの洗練が求められる。

一方で、着手できなかった3次元ポリゴンモデルの写真への写り込み編集についても、見つけ出した方式を実際に試し、工夫を重ねて実用化レベルにもって行って欲しい。そして、今回の提案の時に目指した、“写真群に対する高度な編集処理”を可能とするシステムをそう遠くない次期に作ってみせて欲しい。大いにクリエイターに期待している。