

顔検出・合成による 衣服コーディネート等身大表示システム

舩島一樹^{†1} 久保友香^{†2} 山崎俊彦^{†3} 相澤清晴^{†1†3}

本研究では、様々な衣服画像を前処理無しで擬似試着するシステムとして、大型ディスプレイを使った衣服コーディネート等身大表示システムを提案、実装した。このシステムは、過去の自分の服装や他人の服装にユーザの顔を合成しリアルタイムに等身大で表示することで、それらのコーディネートを追体験させることを目的としている。また、これらのシステムを実現するために顔合成の要素技術である Poisson Image Editing の画像シーケンス使用時の高速化、顔画像合成に特化した自然なマスクの高速作成法を提案した。これにより、リアルタイムでのシステムの稼働、よりリアルな追体験が可能となった。また、このシステムと合わせて服装を撮影するシステムを実装し、「変身」をテーマとした美術館の企画展に出展した。この企画展でシステムはユーザ約 13,000 人に使用された。展示中にユーザスタディを行い、提案した手法が衣服コーディネートの追体験に対して有用な手法であることを示した。

A Proposal and Implementation of Vicarious with Face Detection and Composition

KAZUKI HAJIMA^{†1} YUKA KUBO^{†1}
TOSHIHIKO YAMASAKI^{†2} KIYOHARU AIZAWA^{†3}

In this study, we construct system for displaying the past "clothes-log" on life-size, and synthesizing user's face image by shooting and real-time displaying, and reliving of experience of wearing past "clothes-log". And we displayed this system at an art exhibition to verify this system.

1. はじめに

1.1 背景

近年、自身の日常をデジタルデータとして記録・管理する「ライフログ」という活動が盛んである。デジタル機器の発展以前も日記や家計簿、体調記録などとして似たような活動は行われていたが、GPS 機能やカメラが搭載されたスマートフォンやクラウド技術の発展により、いつでもどこからでも手軽に日常の記録が可能になったことで、ライフログの普及は爆発的に加速した。

代表的なライフログとして「衣・食・住」の「食」を扱った Food-Log があるが、「衣」についても日々の衣服のコーディネートを記録することが多く行われている。この活動は特に若い女性の間で盛んである。

こういったファッションコーディネートのライフログは、現状スマートフォンや PC のディスプレイ上で閲覧を行うことが一般的である。しかしライフログの目的の 1 つである過去の追体験の効果を高めるにはこういった小さなディスプレイでは難しいことが懸念点としてあげられる。

また、衣服の着用体験をディスプレイ上で行うものとしてバーチャル試着システムが挙げられる。これはディスプレ

イ上にユーザの撮影動画と衣服の画像を合成してリアルタイムに映すことで擬似的な衣服の試着を行うというものである。これは大型のディスプレイを擬似的に用い、KINECT[1]等のセンサを用いてディスプレイ前のユーザの動きを検知することで重ね合わせる衣服データを同期させている。実用化された事例として実際の衣服販売店の店舗に試着システムとして導入されているものもある。

これらのシステムは試着を目的としているため、画像はデータベース等にあらかじめ用意されたものを用いており、それらの多くは複雑な処理を施されているものが多い。そのため、過去に自分が着用した衣服というように手軽に衣服のデータを作成できるものは少ないため、衣服データの汎用性に乏しい。

こうした背景から、様々な衣服データを対象に、強いリアリティ・追体験感をもって擬似衣服着用体験をもたらすシステムが求められている。

1.2 目的

本研究では、前述の様々な衣服データを対象とした追体験感の強い擬似試着システムとして、大型ディスプレイを使った衣服コーディネート等身大表示システムを構築する。このシステムは、過去の自分の服装や他人の服装にユーザの顔を合成してリアルタイムに等身大で表示することで、それらの衣服コーディネートを追体験することを目的としている。また、これらのシステムを実装するために要素技術の改善を行い評価した。

†1 東京大学大学院学祭学府・情報学環
The University of Tokyo, Inter faculty Initiative in Information Studies

†2 東京工科大学 メディア学部
Tokyo University of Technology, School of Media Science
Graduate School of Interdisciplinary Information Studies

†3 東京大学大学院情報理工学研究所
The University of Tokyo, Graduate School of Information Science and
Technology



図 1 提案システムの外観

2. 関連研究

衣服画像を表示しユーザに追体験感を与えるシステムの中でも代表的なものが、バーチャル試着システムである。これには KINECT を用いて実用化されているものがある [2], [3]。これらは服や靴の画像を撮影したユーザの画像の上に重ね合わせてリアルタイムに表示することで擬似的に試着体験をユーザに提供している。しかしこれらの例ではアイテムの自然な合成をするためにはアイテムの画像に高度な切り抜き処理などをはじめに行わなければならないため、ユーザが撮影した衣服の画像を気軽に用いることはできない。

また、UNIQLO で実用化された例 [4] は KINECT を用いて対象商品を着たユーザを検知し、ユーザの着用している対象商品の色相を変更することで様々な色の商品の試着を擬似的に行うことができるものである。これは表示結果も自然であるが、限られた商品のカラーバリエーションを表示し疑似体験する目的のみに限られるものである。

こうして見るように、現在行われている研究や開発システムは事前の衣服データに対する前処理が必要である。

3. システムの提案

関連研究で見たように、試着システムでは使う画像データに複雑な前処理が必要なものが多い。そこで、本研究では使用する衣服のデータに依存しない画像ベースでのアプローチによる試着システムを提案する。

このシステムでは、ユーザの顔部分を衣服データにはめ込み合成する事で上記の目的を実現する。具体的にはユーザを動画で撮影し、その画像シーケンス中の顔部分を衣服画像の顔部分と交換して合成する。この合成の際には Poisson Image Editing [5] を用いて自然な境界の合成画像を作成する。そして合成した画像を大型ディスプレイに等身大表示することで、ユーザへのその衣服の着せ替え体験を与えること

を目的としている。

4. システム ver.1 の実装と美術館展示

4.1 システム ver.1 の概要

本システムの ver.1 は大きく分けて「衣服撮影システム」と「表示システム」の2つに分かれ、その2つでユーザとの接点を持つ (図 2)。

衣服撮影システムでは、ユーザの衣服を自動で撮影し、撮影した画像を正規化する。ここで正規化された画像はデータベースに保存され、順次表示システムに衣服画像として送信される。

表示システムでは、データベース内の画像と表示システムの前に立ったユーザの顔部分の画像を合成して、ディスプレイ上に等身大で表示する。

図 3 に全体のユーザの動線を簡単に示す。

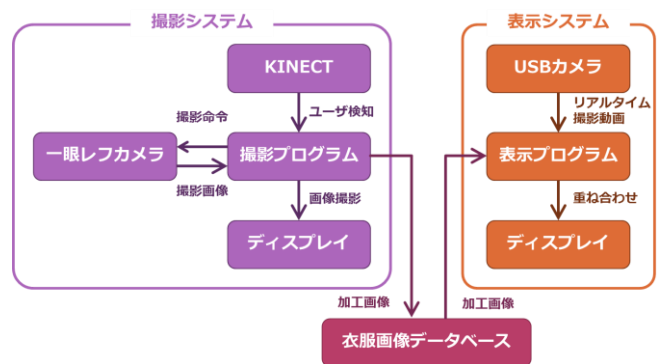


図 2 提案システム ver.1 の概要



図 3 プロトタイプを使うユーザの動線

また、このシステム ver.1 は 2013 年 6 月から 9 月にかけて東京都現代美術館の企画展「オバケとパンツとお星さま」に出展された [6]。この展示では「変身」がテーマとされているコーナーに設置されており、子供が自分でオリジナルの奇抜な衣装を作ることができるスペースが併設されていた。そのため、本システムで撮影された衣服データは来場者のオリジナルの衣装が多く、表示システムでは別のユーザのオリジナル衣装と自分の顔部分画像を合成できるため、他人の衣服の追体験・強い変身感を体験できるようになっていた。

4.2 衣服撮影システム

衣服撮影システムは撮影機能部分と撮影背景に分かれる(図4)。

撮影機能部分は, KINECTと一眼レフカメラ(Nikon D5100), 24インチのディスプレイ(DELL U2412M)によって構成されている(図5)。撮影背景上でユーザが静止すると, KINECTがそれを検知し自動的に一眼レフカメラによってユーザの全身を撮影する。ここでは自動で撮影を行うため, ユーザに撮影のタイミングを知らせる機能が必要となった。そこで, ディスプレイにメッセージや画像のキャプチャを表示することで, 撮影の補助としている。また, 背面と底面の境界を目立たなくさせるために撮影背景部分は緩やかな曲面となっている。

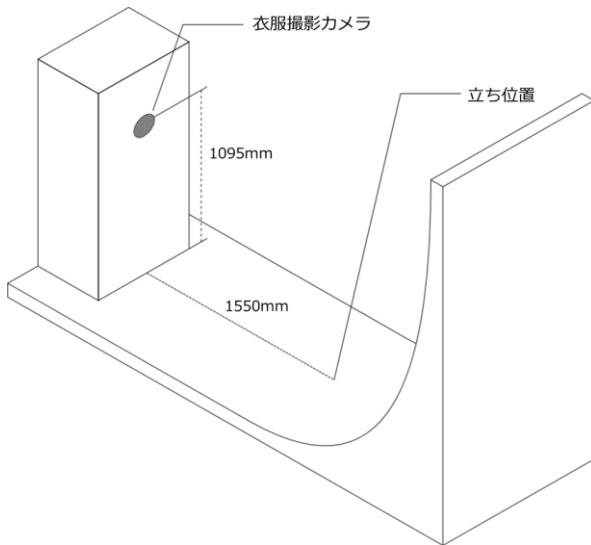


図4 衣服撮影システムの外観

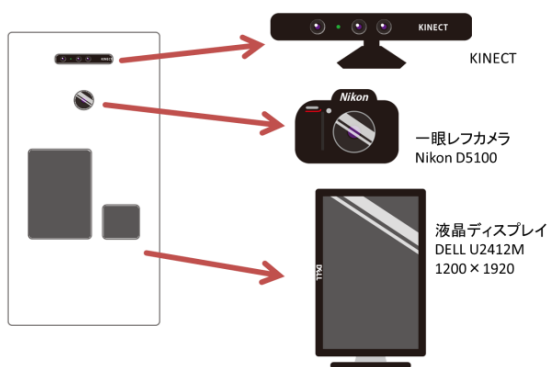


図5 撮影機能部の構成

4.3 合成画像表示システム

表示システムは, 高解像度 USB カメラ(PointGrey FL3-U3-32S2C-CS)と 80型の液晶ディスプレイ(SHARP PN-E802), 顔出し板を用いて構成されている。

ここでは, 撮影システムによって正規化された撮影画像の顔の部分に, USBカメラでキャプチャされたユーザの顔部分の動画を重ねることで, ユーザに衣装コーディネートの実験をもたらし、あたかも旅先に設置されている顔出し板のように顔を切り抜かれた全身画像に対して,

カメラで撮影するユーザの顔動画像がはめ込まれる。そうして合成された画像は大型ディスプレイに表示される。こうして合成された画像は, 1画像あたり5秒のスライドショーで順に表示される。

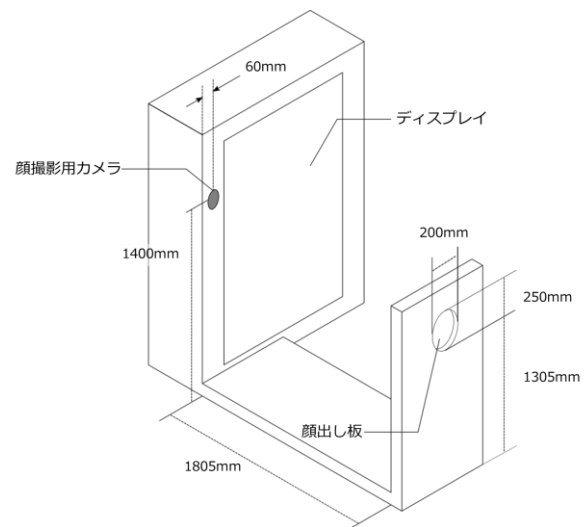


図6 表示システムの外観

4.4 東京都現代美術館の企画展への展示

この章で述べたシステム ver.1は, 実際に2013年6月から9月まで, 東京都現代美術館の企画展「オバケとパンツとお星さま」において展示された。

本展は, 幼児から小学生の子どもたちを主な対象としており, こどもが自由楽しんで鑑賞することをコンセプトとしていた。展示内容は5組のアーティストの作品で構成され, その中でファッションデザインを手がけるアーティスト「ゼロゼロエスエス」の担当する, 「変身」をテーマとしたコーナーに本システムを展示した。

上述のとおり, 来場者が実際に触って楽しむことをコンセプトとした展示であったため, 本システムは展示用にインタラクティブに設計され, 来場者合計約13,000人に実際に使用された[11]。

5. システムの改善

5.1 システム変更概要

美術館に展示していたシステム ver.1は, 企画展の1展示作品としていろいろと制約のある特殊な環境でのシステムであった。

例を挙げると, その場でのインタラクティブ性を強調するために, このシステム内でアクションを完結させる必要があったため衣服撮影システムを併設していた。しかし本来このシステムの目的は合成画像の等身大表示と, それに付随する体験であるため衣服撮影システムは必ずしも必要ではない。そのため, 改善システムでは衣服撮影システムを廃し, 表示システムのみとした。

そのため改善後のシステムの外観は図26のようになる。

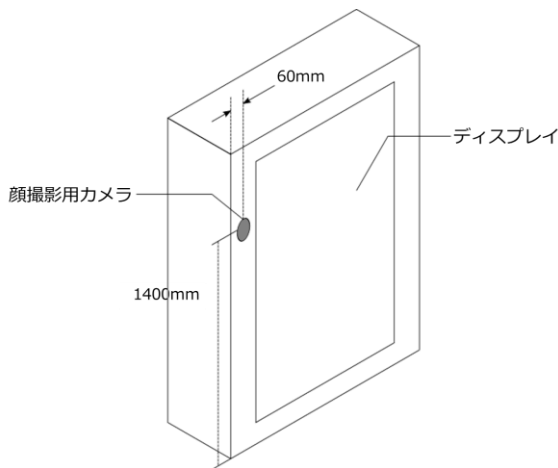


図 7 改善後のシステム外観
 Figure 1

また、ハードの変更に伴いシステム内部にもいくつかの変更・改善を施した。

5.2 顔の検出法の変更

システム ver.1 では、顔出し板によってユーザの顔の位置が固定されていたため、ユーザの撮影画像に対して顔検出を行う必要はなかった。しかし顔出し板を廃したため撮影している画像シーケンスからユーザの顔を検出する必要がある。本システムでは、画像シーケンスからの顔の検出には3章で紹介した Deformable Model Fitting を利用した。

教師データセットには Stephen Milborrow らによって作成されたデータセットを用いた。このデータセットは 3755 枚の顔画像で構成されており、76 個の特徴点が設定されている[12]。

5.3 顔の合成アルゴリズムの改善

顔の合成には Poisson Image Editing を用いている。

しかし、Poisson Image Editing をそのまま用いると、計算時間がかかるため、リアルタイムで動かすという提案システムの目的にはそぐわなかった。また合成後の画像の自然さについても、Poisson Image Editing を用いるとやや不自然な結果となったため、自然なマスクの自動生成法を提案し実装した。

5.3.1 Poisson Image Editing の問題点の整理

通常の Poisson Image Editing を次の 2 つの観点から改善した。

1. リアルタイムで合成するための処理速度
 2. 合成の自然さ
- の 2 つである。

それぞれの詳細を以下に述べる。

5.3.2 処理速度の向上

Poisson Image Editing は反復解法であるため、そのまま用いると遅延がおき、リアルタイムで動かすことができない。そこで、処理速度を向上するために、3 つの改善を施した。

SOR 法

まずは高速化の 1 つとして反復解法のアルゴリズムを、ガ

ウス・ザイデル法から、SOR 法に変更した。

並列化

また、マルチコア用にアルゴリズムを並列化した。

並列化の手法は Poisson Image Editing を解く際に、図 8 のようにいくつかの領域に縦に分割し、各領域をそれぞれのスレッドで SOR 法で反復処理する。

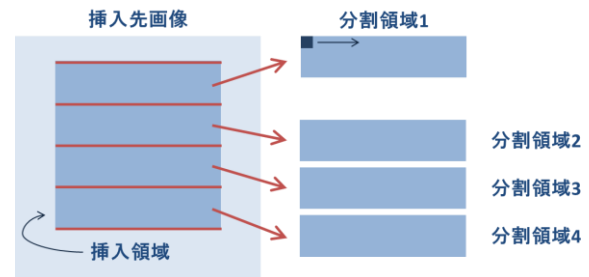


図 8 並列化のための領域分割

しかし、SOR 法では各ピクセルを逐次的に値を更新していくため、注目しているセルは上下左右のピクセルとの間に依存関係が生じる。そこで、領域を分割した際はそのまま並列で処理を進めると分割領域の境界でこの順序が崩れる可能性があるため、図 9 のように隣接する領域の処理が適切に行われるように工夫する必要がある。図 8 のように 4 つの領域に分割した場合は約半分の処理時間で処理が終了する。一般的に n 個の領域に分割すると、処理速度は最大 $n/2$ になる。

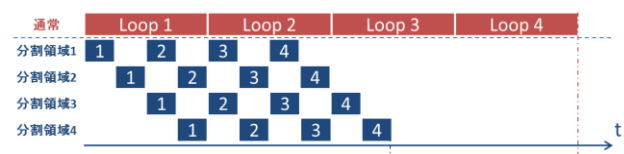


図 9 並列化の際のスレッドの実行順序

反復法初期値の前フレームの利用

さらに合成処理を高速化させるために、入力値である画像シーケンスの隣接するシーケンス間では照明環境などの色味や顔の角度などが大きく変わらないことに注目した。

提案システムでは、入力画像シーケンスに対して毎フレーム Poisson Image Editing を施すが、連続する画像間での色情報誤差が少ないと仮定すると、前フレームの合成結果を利用することで処理の反復回数の短縮が期待できる。

つまり、各フレームでは反復法によって合成結果を求めているため、合成結果の画像と初期値画像の差が小さいほど収束までの反復回数が少なくなり、合成処理の高速化が望める。よって、各フレームでの合成の初期値に前フレームの合成結果を利用することで高速化が期待できる。ここでは、その初期値の設定方法を提案する。

本研究では、具体的には前フレームと現在フレームの挿入元画像の差分を、前フレームの合成結果に足し合わせて初期値とする手法を提案する。

これは単に前フレームの合成結果を初期値として用いるだけでなく、前フレームの入力値と現在フレームの差分が、

フレーム間の合成結果の差につながるという予想に基づき提案した(図10)。初期値と合成結果の間には各フレームで同じ変換が行われている。そして、この変換は入力の差が小さければ出力の差も小さくなるので、出力の差分を入力差分で近似したものを初期値画像として用いた。

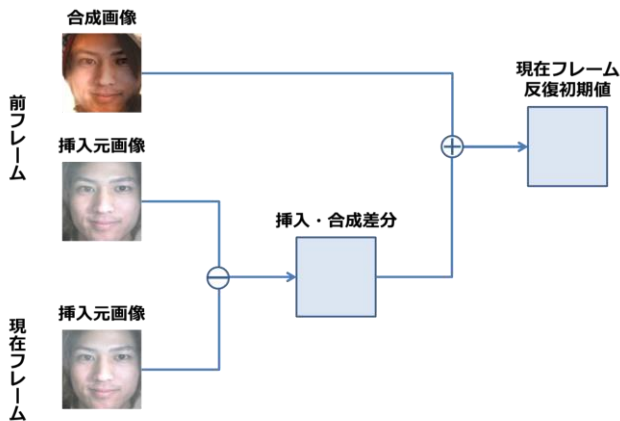


図10 フレーム間差分利用法

このように前回の合成結果を利用することで、合成完了までの反復回数の減少が期待される。

5.3.3 自然な合成法

エラー! 参照元が見つかりません。に見るように、挿入領域をマスクによって上手く指定しないと、顔以外の部分の影響を受けて合成画像の色味が不自然になってしまう事象が ver.1 では散見された。これは、合成には Poisson Image Editing を使っていて、領域境界部分のピクセルの画素値を Dirichlet 条件として用いているためこの画素値が内部まで影響してしまうためである。

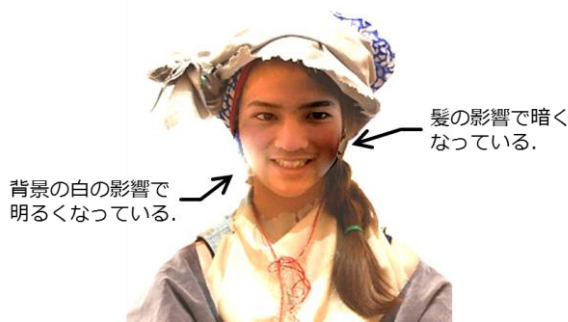


図11 画像の合成境界付近の色味が不自然な合成結果の例
 これを避けるためには、境界部分に背景や髪といった顔と極端に色味の違う画素値を持つピクセルが来ないように、マスクを上手く調整しなければならない。

システム ver.1 ではマスクの指定領域は単純な楕円としていたが、顔のパーツ部分が顔合成の主要な対象となるためその部分だけマスクされていれば良い。実際、図12のマスク2のようにT字型にして目・鼻・口のみをマスクしている事例も多く見られる。しかし、本手法では合成するユーザの顔のパーツの位置を前もって知ることはできず、顔の向きなどにより各フレームでパーツの位置は変わること

が予想されるため、小さい固定領域のマスクはフレームによってマスク領域内にパーツが入らないことがありうるため好ましくない。

また、自動で対象領域をマスクする研究[12]もあるが、これらは処理に時間がかかるためリアルタイムで動かすには適さない。そこで、ここでは各ピクセルの色情報から顔の肌の色の部分を検出して、毎フレーム自動でマスク領域を楕円から狭める手法を提案する(図12 マスク3)。

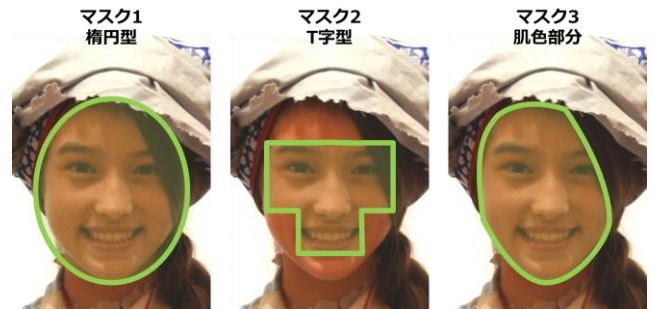


図12 マスクの形の種類

肌の色は、画像の色空間を RGB から HSV へ変更した後、ルールベースで肌色が存在する領域を検出した。

これによって、髪や背景部分といった図22で不自然な色味の原因となっていた部分がマスクされているのがわかる。

6. システムの改善に関する評価

この章では5章でみた、システムの改善による結果を考察する。

6.1 高速化の結果

6.1.1 前フレームを利用した初期値を用いる反復法の結果

まず高速化として前章で述べた、反復法初期値の前フレームの利用の結果に関して述べる。

比較対象として挿入領域の初期値を全て黒、挿入元画像、挿入先画像とする3つの手法の速度と比較した。速度の比較は収束に至るまでの反復回数を指標とした。サンプルデータを用いて各フレームを計算した際の各手法の反復回数を以下に示す(図13)。

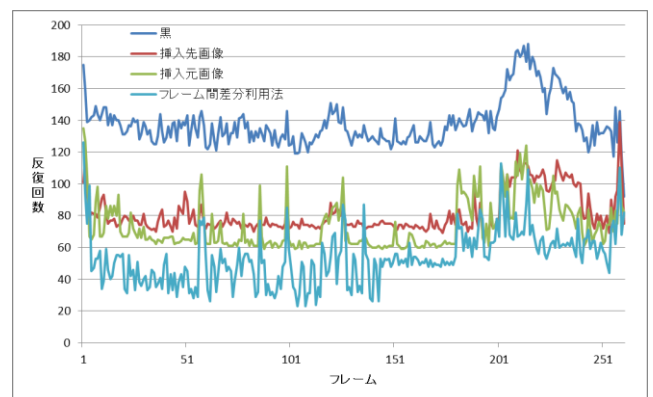


図13 各初期値での反復数の比較

これを見ると、提案手法が反復回数も少なく優れていることがわかる。

6.1.2 各手法のシステム全体の高速化の効果

検証に使用した PC のスペックは以下のとおりである。

CPU	Intel Corei7-264M
クロック数	2.80GHz
RAM	4.00GB

また、合成に用いた画像のサイズは 120px×120px である。SOR 法によって 5.79fps から 7.89fps へと 1.36 倍の高速化を達成した。また、並列化による高速化を用いた際には 10.02fps となった。フレーム間差分利用法を用いた場合は 8.04fps となった。それらの結果を図 14 に示す。

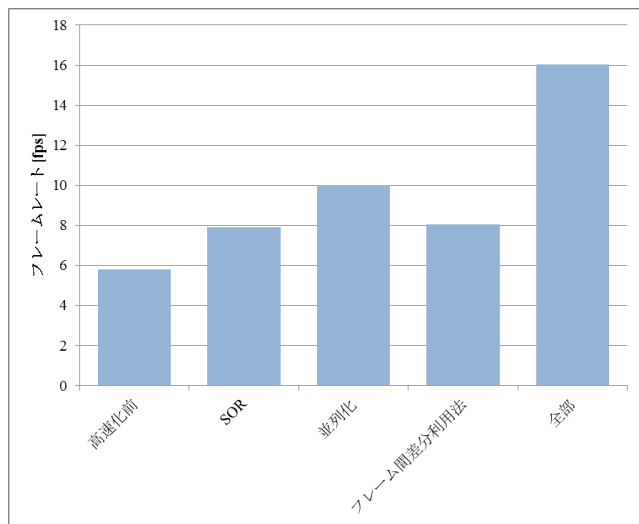


図 14 高速化手法の比較

上の図に見るようにこれらの手法は全て高速化に対して有効であった。前の章では並列化は理論的には 2 倍の高速化が期待されると書いたが、実際には 2 倍の速度になっていない。この理由は前処理部分や CPU 利用率などにより並列化効率が 100% になっていない結果である。また、今回は 4Core の環境で並列化を施したが、Core 数が増えれば速度も線形にスケールする。

これら 3 つの手法はそれぞれ独立であるため、全ての高速化を同時に施すことが可能である。それによって高速化前の約 2.5 倍の速度となり、15fps を超えることができた。また、これらの処理はすべての処理部分に対して高速化しているわけではないため、高速化率はすべての高速化率の積からは減少している。

6.2 合成の自然さの結果

6.2.1 定性的評価

マスクの改善によって、図 15 のような結果を得た。

具体的には髪や背景といった肌の色とは大きく外れた領域を合成領域から外すことで、顔内部の色味が影響を受けることを防いでいる。

また、挿入先画像内の髪を合成領域から排除することで髪と肌の境界をくっきり残すことができた。



図 15 マスクの改善による合成結果の変化

しかし、挿入元画像に髪などの肌以外の部分が多く存在している場合はこれにより、領域境界が不自然になる場合がある。

これを、顔のパーツの認識などにより改善することが今後の課題である。

7. 結論

本研究では、大型ディスプレイを使った衣服コーディネート等身大表示システムを提案・実装した。

実装に伴い、画像シーケンスにおける Poisson Image Editing の高速化と、顔部分の合成に特化した改善を行いこれが有用であったことを示した。

また、システム ver.1 を「変身」をテーマとした美術館の企画展に展示した。

また、システムの顔出し板の効果や、本研究では目的の 1 つとしてあげた「追体験感」、合成の自然さの評価などは、本研究では定性的なものにとどめたが、心理学やヒューマンインタラクションなど他の分野からのアプローチが今後の課題である。

本システムの利用法としては、背景でも触れたとおりファッションのライフログなどが考えられる。本研究では合成、表示システムのみに注力したためライフログとして使用するための機能等も今後の課題である。

参考文献

- 1) “Kinect for Windows | Voice, Movement & Gesture Recognition Technology.” [Online]. Available: <http://www.microsoft.com/en-us/kinectforwindows/>.
- 2) “Goertz Casefilm Virtual Shoe Fitting.” [Online]. Available: http://www.youtube.com/watch?feature=player_embedded&v=uSn7c1uw1_A.
- 3) “Swivel | Your Dressing Room Just Got a Whole Lot Bigger.” [Online]. Available: <http://www.facecake.com/swivel/>
- 4) “デジタルサイネージを活用した「バーチャル試着システム」を開発 | DNP 大日本印刷株式会社。” [Online]. Available: http://www.dnp.co.jp/news/10056353_2482.html.
- 5) P. Pérez, M. Gangnet, and A. Blake, “Poisson image editing,” *ACM Trans. Graph.*, vol. 22, no. 3, p. 313, Jul. 2003.
- 6) “東京都現代美術館 | MUSEUM OF CONTEMPORARY ART TOKYO.” [Online]. Available: <http://www.mot-art-museum.jp/exhibition/146/>.