

仮想試着による動画 ファッションコーディネートシステム

中島剛史[†] 石田正明^{††} 市村哲^{††}

服を購入する際、カタログを参考にすることが多い。しかし、カタログでは自分が服を着たイメージがわかりづらい。現状では、写真を使ったシミュレーションや服をCG合成する手法がある。最近では、Kinectセンサーを用いたKinect Fitting Roomが話題になった。しかし、背面や服のねじれ等のリアリティが出せないことや長袖等、服によっては仮想試着を実装することが困難である等の問題点が挙げられる。そこで、服のねじれ等のリアリティがあり、服の種類に左右されないことに重点を置いた、仮想試着による動画ファッションコーディネートシステムを提案する。本研究では、Kinect SDK 及び OpenCV を用いてシステムを実装した。実装したシステムは、参照する動作を登録する「登録フェーズ」と登録された動作を参照する「参照フェーズ」の2つのフェーズから成る。また、色々な場所を想定した背景や登録した動作を参照している人物を合成する機能を付けた。評価実験として、本研究で作成したシステムを20~30代の男女10名に使用してもらった。被験者には、実験終了後アンケート項目に回答してもらった。その結果、被験者の90%が、本システムがファッションコーディネートに役に立ったと回答した。

An animated fashion coordination system with virtual fitting function

Takeshi Nakashima[†] Masaski Ishida^{††}
Satoshi Ichimura^{††}

There was a technique for synthesizing CGs and clothes. Recently, "Kinect Fitting Room" that used "Kinect sensor" got into the news. However, it was difficult to see the back, a twist of the clothes or long sleeves, etc. Thus, we propose an animated fashion coordination system with virtual fitting function. In this study, "Kinect SDK" and "OpenCV" were used for implementing the system. It consists of two phases, the "registration phase" which registers the action to be referred, and the "reference phase" which refers to the registered action. Also it has a background synthesis and face synthesis functions. In an experiment, the 10 men and women in their 20s to 30 were asked to use. As a result, 90% of subjects praised it as useful to the coordinated fashion.

1. はじめに

服を購入する際、カタログを参考にすることが多い。しかし、カタログでは自分が服を着たイメージがわかりづらい。一方、店で試着する場合は、多数の組み合わせを確認するには時間がかかるという問題がある。

近年、AR (拡張現実) 技術を用いた試着システムが提案されている。例えば、ゲームハード機 Xbox360 の Kinect センサー⁵⁾を用いて仮想的に試着できる Kinect Fitting Room⁴⁾がその例である。しかし、Kinect Fitting Room では、背面や服のねじれ等のリアリティが出せないことや、長袖等、服によっては仮想試着を実装することが困難である等の問題点が挙げられる。

そこで、Kinect センサー前で動くだけで画面上のカタログ内のモデルを自分の動きに合わせて動かすことができる仮想試着による動画ファッションコーディネートシステム「Kinect ファッションカタログ」(以降 K.F.C.) を提案する。

評価実験として、本研究で作成したシステムを20~30代の男女10名に、6つの実験項目に沿って使用してもらった。その結果、被験者の90%が、K.F.C.がファッションコーディネートに役に立ったと回答した。

2. 背景

図1は全国の20歳以上の男女1,574人(2008年4月)に対する世代別の服の購入場所に対するアンケートデータである³⁾。この図から、どの年代でも通信販売ではなく、現地での購買意欲が高いことが分かる。20代では「郊外ショッピングモール・アウトレット」(47.0%)、「駅ビル・ファッションビル」(29.8%)等、小売店舗が多数集まった商業施設での購入割合が高い。しかし、30代になると「インターネット通販」(20.6%)の利用傾向が増えており、仕事や育児等の影響であまり時間をかけたくないということが窺える。

上記の点から、今回、現地での服の購入の決め手である試着に着目した。この試着をゲーム的な要素を取り入れた楽しめる作業に変えることに重点を置き、あまり複雑でなく、より簡易的に試着を体感できるシステムを作成することを目標とした。

[†] 東京工科大学大学院バイオ・情報メディア研究科
Graduate School of Bionics, Computer and Media Sciences, Tokyo University of Technology

^{††} 東京工科大学コンピュータサイエンス学部
School of Computer Science, Tokyo University of Technology

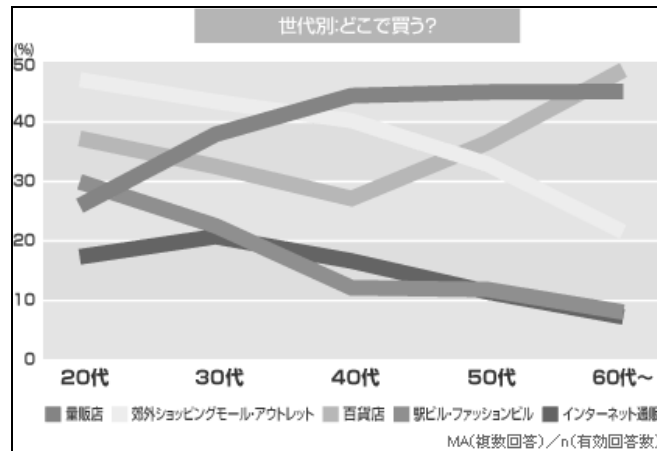


図 1 世代別服の購入動向³⁾

Figure 1 Buy of clothing trends by generation.

3. 既存技術と問題点

紙のカタログでは、ユーザ自身が着用した姿が分からない。そこで、実際に着用をするが、実際の試着となると多数の組み合わせを確認するには時間がかかってしまう。また、服装選びは、会う人・場所・予定等を考慮しなくてはならない。

既存技術に suGATALOG²⁾ というものがある。suGATALOG は、ユーザが服を着用した静止画をトップスとボトムスに分け、組み合わせのシミュレーションをして、ファッションをコーディネートするシステムである。他にも、複数のコーディネートと比較して、新しいコーディネートを発見する機会を提供する機能がある。

しかし、suGATALOG は静止画像のみであり、正面からの画像のみの組み合わせによるファッションコーディネートを行うので、服のねじれや側面の様子等、動いた時の服のイメージがわきづらい。

2つ目の既存技術は、Kinect Fitting Room¹⁴⁾ である。Kinect Fitting Room は、Kinect を使った試着体験システムである。AR (拡張現実) 技術によりユーザの体に CG の服を合成する仮想的な試着体験が可能になっている。

しかし、Kinect Fitting Room では CG を使用しているので、画像として提供された服しか試着することができない。また、服のパーツを 3D モデルとして作成するには時間がかかり、一般の店舗で服の 3D モデルを作成するとなると実現が困難である。他にも、長袖のような服の形状によっては対応していない等の問題点が挙げられる。

4. 提案

実際に試着を行わずにお気に入りのファッションを探し出すことが可能な対話型ファッションカタログ「Kinect ファッションカタログ」(K.F.C.) を提案する。Kinect センサーの前で動くだけで自分の動きに合わせてカタログ内のモデルが動く。これにより、購入のきっかけを作る。システムは、カタログで使用する服を着たモデルが服を登録する「登録フェーズ」と、モデルの服をユーザが参照する「参照フェーズ」の2つで構成される。登録フェーズでは、服を着て動くモデルの姿を登録し、参照フェーズでは、ユーザの動きと一致したモデルの姿を表示する。

また、モデルの背景を画像合成することで、場所や目的に合った服を探すことや、ユーザ自身の顔をモデルの顔に合成することにより、ユーザ自身が服を着て動いているというイメージをしやすくしている。

5. 実装

図 2 にシステムの全体構成を示す。Kinect センサーから関節の XYZ 座標を取得する。取得できる関節は 20 点あるが、本実装では現状トップスのみを対象としており、そのうち、図 3 の矢印で示された両肩の中央・両肩・両肘・両手首の関節座標を取得するようにした¹⁴⁾。その取得情報から、求めたい角の関節の座標と両隣接の関節の座標の 3 点の座標の内積を計算し、両肩と両肘の角度を算出した。座標から角度を求めた式を数式 1 に示す。右肩の関節座標 $A(X_a, Y_a, Z_a)$ 、右肘の関節座標 $B(X_b, Y_b, Z_b)$ 、右手首の関節座標 $C(X_c, Y_c, Z_c)$ 、両肩の中央の関節座標 $D(X_d, Y_d, Z_d)$ として角度を求めた式である。尚、肩は 180 度以上曲げることができるため、両肩の中央 D の Y 座標と肘 B の Y 座標を比較して、両肩の中央 D の Y 座標の値が大きいときに 180 度以上の値を取るようにした。K.F.C. は、登録フェーズと参照フェーズの 2 つのフェーズから成る。

5.1 登録フェーズ

服を着て動くモデルの姿を静止画像 (動画フレーム) として 1 秒間に 30 コマで保存する。加えて、Kinect センサーが取得した骨格情報 (両手首と両肘と両肩と両肩の中央の座標) と両肩と両肘の角度をファイルに保存した。

5.2 参照フェーズ

Kinect センサーの前でユーザが動いた時、ユーザの骨格情報 (両手首・両肘・両肩・両肩の中央の座標) から両肩と両肘の角度を算出する。登録されているモデルの両肩と両肘の角度をファイルから取得して誤差 15 度以内で一致したら対応する静止画像を表示する。図 4 は K.F.C. の実行画面である。左はマッチングしたモデルの画像で、右はユーザの画面を示す。

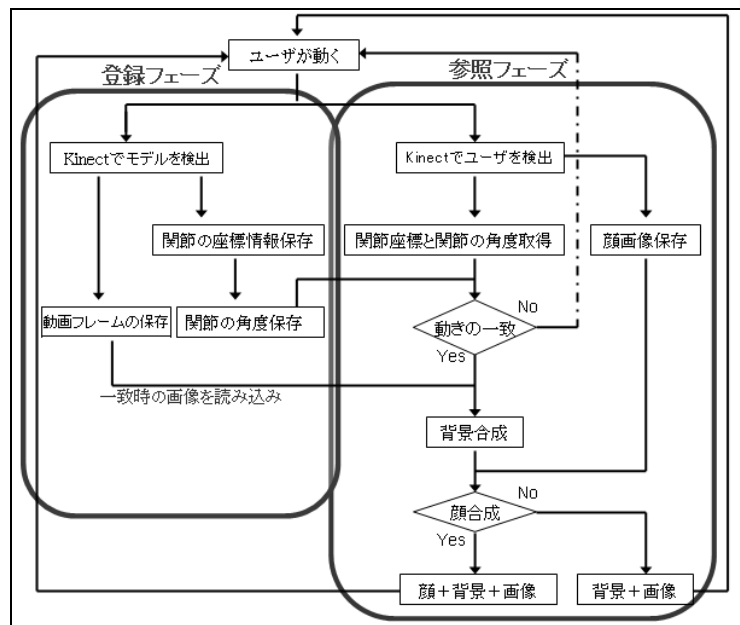


図 2 システム構成
Figure 2 System architecture

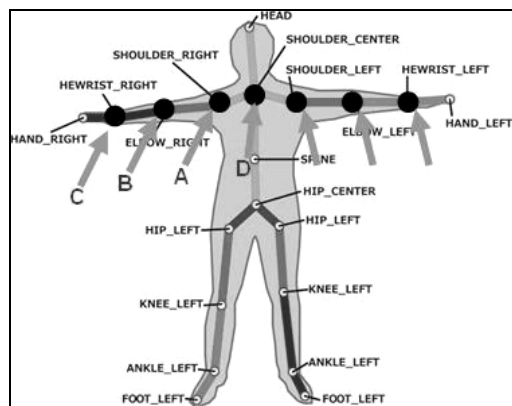


図 3 関節取得図⁶⁾
Figure 3 Figure Joint acquisition

$$\begin{aligned}
 &A(X_a, Y_a, Z_a) \quad B(X_b, Y_b, Z_b) \quad C(X_c, Y_c, Z_c) \quad D(X_d, Y_d, Z_d) \\
 &\vec{a} = (X_a - X_b, Y_a - Y_b, Z_a - Z_b) \\
 &\vec{b} = (X_c - X_b, Y_c - Y_b, Z_c - Z_b) \\
 &\vec{a} = (a_1, a_2, a_3) \\
 &\vec{b} = (b_1, b_2, b_3) \\
 &\cos \theta = \frac{a_1 \times b_1 + a_2 \times b_2 + a_3 \times b_3}{\|\vec{a}\| \|\vec{b}\|} \\
 &\text{rad} = \cos^{-1} \cos \theta \\
 &Y_b \geq Y_d \\
 &\theta = \text{rad} \times \frac{180}{\pi} \\
 &Y_b < Y_d \\
 &\theta = 360 - \text{rad} \times \frac{180}{\pi}
 \end{aligned}$$

数式 1 座標から角度を求める式
Equation 1 The equation which asks for an angle from coordinates

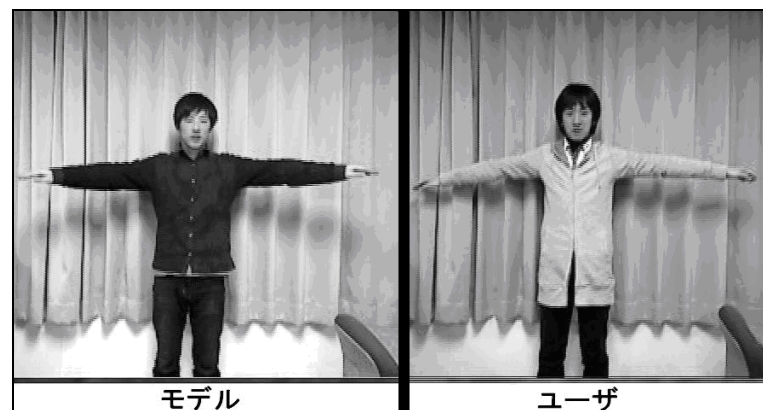


図 4 実装画面
Figure 4 Implementation screen.

背景合成は、OpenCV を用いて背景差分、背景を濃淡画像に変換、二値化、ノイズ除去、合成の順に処理することで実装した⁸⁾¹¹⁾¹³⁾。図 5 は背景合成前（左）と合成後（右）の画像である。

顔合成は、OpenCV の haar-like 特徴を利用したオブジェクト検出を用いて実装した⁷⁾⁹⁾¹²⁾。はじめに、ユーザの顔の画像を検出し保存しておく。次に、保存しておいた顔画像を、表示中のモデルの画像に合成することでユーザ自身が他の服を着て動いているように見せる。図 6 に顔合成の図を示す。



図 5 背景合成画像

Figure 5 Composite image background.

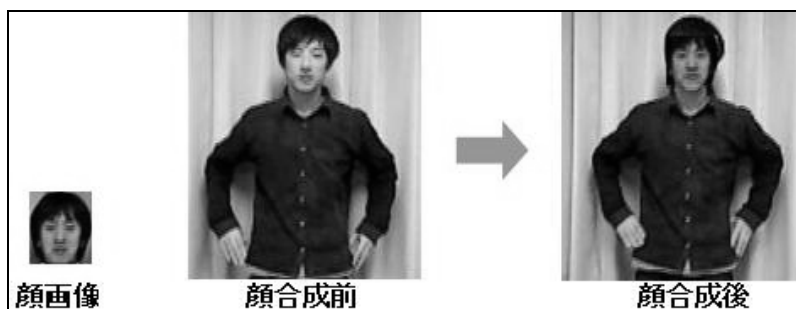


図 6 顔合成処理

Figure 6 Face synthesis process.

6. 評価

4 つの評価実験を行った。1 つ目は、他カタログとの比較である。20 代の男性 5 名に対して数種類のカatalogを見せ、自分がその服を着ているイメージを持ちやすい順番に並べてもらった。モデルが着ていた服だけの写真、モデルが服を着た写真、モデルが服を着て動くビデオ、K.F.C.の 4 カタログである。

表 1 に他カタログとの比較結果を示す。モデルが服を着て動くビデオも高い結果となったが、K.F.C.の評価が最も高いことが示された。

表 1 他カタログとの比較結果

Table 1 The results of the comparison with other catalogs.

被験者	A	B	C	D	E	平均 (順位)
服の写真	4	3	4	1	4	3.2
モデルが服を着た写真	3	4	3	4	2	3.2
モデルが服を着て動くビデオ	2	2	2	3	1	2.0
K.F.C.	1	1	1	2	3	2.0

2 つ目は、Kinect センサーの前で自由に動いてもらった時、どれぐらいの確率で一致した画像が表示されるかを実験した。実験回数は 5 回である (のべ 100 秒間)。表 2 にマッチング精度の結果を示す。平均で 70%を超えた。

表 2 マッチング精度

Table 2 Matching accuracy.

回数	1	2	3	4	5	平均
総フレーム数	521	276	392	473	494	
一致フレーム数	389	158	222	381	401	
割合 (%)	74.66	57.25	56.63	80.55	81.17	70.05

3 つ目は、処理速度の評価を行った。実験は全部で 10 回行い (のべ 200 秒間)、一致時の画像表示のみの場合と画像表示に背景と顔を合成した場合の 2 パターンで各 5 回、20 秒ずつ行った。Kinect の標準フレームレートが 30fps であることから、総フレーム数が 600 であれば全くコマ落ちが発生していないことになる。表 3 に結果を示す。画像のみの場合でも、約 59%コマ落ちしていた。また、画像に背景と顔を合成した場合では、約 93%のコマ落ちがあった。

表 3 処理速度
 Table 3 Processing speed.

回数	1	2	3	4	5	平均	フレームレート (平均/実行時間)
画像のみ	343	205	188	247	248	246.2	12.31
画像+背景+顔合成	39	37	39	44	47	41.2	2.06

4つ目の評価は、システムのアンケート評価である。20代から30代の男女10名に被験者としてK.F.C.を使用してもらい、下記のアンケート項目に対して回答してもらった。尚、①～③のアンケートは5段階評価とした。

- ① マッチングした場合の画面表示の速さ (速い 5・4・3・2・1 遅い)
- ② マッチングの精度 (良い 5・4・3・2・1 悪い)
- ③ システムの使い易さ (使いやすい 5・4・3・2・1 使いにくい) +自由記述
- ④ 顔合成は必要でしたか 必要・不必要
- ⑤ K.F.C.はファッションコーディネーターに役に立った なる・ならない+自由記述
- ⑥ 雑誌などのカタログと比較してK.F.C.の良い点・悪い点 (自由記述)

表 4 アンケート項目①から③の結果
 Table 4 The result of the questionnaire items 1-3.

被験者\質問項目	①	②	③
1	3	3	4
2	4	2	4
3	4	2	4
4	5	2	2
5	5	2	4
6	5	2	5
7	3	4	5
8	4	4	3
9	3	4	3
10	3	3	3
平均値	3.9	2.8	3.7

表 5 アンケート項目④から⑤の結果
 Table 5 The result of the questionnaire items 4-5.

被験者\質問項目	④	⑤
1	必要	なる
2	不必要	なる
3	不必要	なる
4	不必要	なる
5	必要	なる
6	必要	なる
7	必要	なる
8	必要	未回答
9	必要	なる
10	不必要	なる
最多	必要 (60%)	なる (90%)

- ・ファッションコーディネーターに役に立った
- ・動くだけで、色々な動きを見ることができ、参考になった (類似の意見他 3)
- ・動いている時の印象がわかる (類似の意見他 2)
- ・背景・顔合成によりイメージしやすい (類似の意見他 2)

7. 考察

アンケート項目③の平均値が 3.7 であり被験者の多数が使いやすいと回答した。しかし、K.F.C.はPCソフトの操作が必要なため、使いづらいという意見があった。そこで、Kinect センサーの前で特定の動きすることによってジェスチャで操作できるようにすることを検討する。

マッチングした場合の画面表示の速さにおいて、遅いと感じた人は少なく、アンケート項目①の平均値が 3.9 であり、人の直感として服の動きを参照する上では静止画の連続再生は有効であるといえる。

アンケート項目④について、顔合成が必要であると回答したのが 60%とあまり高い数字ではなかった。また、処理速度の実験において、背景+顔合成のフレームレートが93%低下してしまった。このため、マッチングした時の画面表示の速さが遅くなり、

ユーザに不快感を与えたと考えられる。

また、K.F.C.では、OpenCVによる顔検出と合成を行ったが、髪の高い女性や横向きの顔といった場合の時には認識が困難なため、Kinectセンサーが検出する骨格情報の一部である、頭の座標と深度データを利用して顔を抽出する方法を検討する。

アンケート項目⑤と⑥の自由記述から、服が動いているのでユーザ自身が試したいポーズを参照できることが良いという意見が多かった。このことから、マッチング精度が70%程度でもファッションコーディネートに役立つのではないかと考えられる。アンケート項目②の平均は2.8とかなり低いが、ユーザはファッションコーディネートにおいてマッチングにあまりこだわりがなく、イメージを大切にしていることが分かる。

今後の改善手法として、Kinectセンサーの音声認識を用いたインタフェースの追加を考える。例として、ユーザが見たい姿勢の際に、「ストップ」と発言をすることで、同じ姿勢の画像を表示し、固定するインタフェースを追加する。

他にも、現状ではトップスのみのマッチングとなっているため、ボトムスと組み合わせたコーディネートシステムを検討することや、参照したファッションの在庫を確認する機能、一つのファッションまたは複数のファッションについて参照できるように切り替えられるモード選択機能、背景選択機能、Kinectセンサーがあればどの場所でもK.F.C.を体験できるようにする等の機能を実装することを検討する。

8. まとめ

本研究では、Kinectセンサーを用いることにより、服のねじれ等のリアリティがあり、服の種類に捉われないことに重点を置いた、「仮想試着による動画ファッションコーディネートシステム」を提案し、システムの構築を行った。

K.F.C.では、Kinectセンサーで取得できる人間の関節の座標データを利用して、参照する動作を登録する登録フェーズと登録された動作を参照する参照フェーズの2つのフェーズを実装した。

また、参照フェーズで参照するファッションを自分自身として捉えるための顔合成や場所や状況に合わせたファッションをイメージしやすくするために、クロマキー合成を用いた背景合成を実装した。

評価実験では、被験者に6つの実験項目を行いアンケートに回答してもらった。結果としてK.F.C.はファッションコーディネートに役立つことがわかった。しかし、マッチングの精度に対して低い評価であった。今後の課題として、参照フェーズで表示される画像群の表示速度の向上やキネクトセンサーがあればどの場所でもK.F.C.を使用できるようにすることを検討する。

参考文献

- 1) 谷尻豊寿：「身体の動きがコントローラ C++で Kinect プログラミング KINECT センサー 画像処理プログラミング」.
- 2) 佐藤彩夏, 渡邊恵太, 安村通晃：「suGATALOG：ユーザの姿を利用したファッションコーディネート支援システム」情報処理学会 インタラクション 2011
- 3) ハピ研毎週アンケート：洋服購入時のポイントは？
<http://img.asahibeer.co.jp/enjoy/hapiken/maian/bn/200804/00231/>
- 4) プログラミング・ガイド
(Programming Guide: Getting Started with the Kinect for Windows SDK Beta)
- 5) Kinect - Xbox.com
<http://www.xbox.com/ja-JP/kinect>
- 6) Kinect SDK XNA で骨格情報の取得 その1
<http://www.kosaka-lab.com/tips/2011/06/kinect-sdk-xna.php>
- 7) OpenCV.jp
<http://opencv.jp/>
- 8) 縣弘樹, 山下敦, 金子透：「市松模様の背景を用いたクロマキーによる領域抽出」情報処理学会研究報告. [オーディオビジュアル複合情報処理] 2005(124), 57-62, 2005-12-12
- 9) 山下敦, 縣弘樹, 金子透：「動的輪郭モデルを用いたストライプ状クロマキー」映像情報メディア学会誌：映像情報メディア 61(2), 189-197, 2007-02-01
- 10) 「仮想試着による動画ファッションコーディネートシステム」
東京工科大学 石田正明 2011 年度卒業論文
- 11) コンピュータビジョン特論
<http://www.wakayama-u.ac.jp/~wuhy/CV03.pdf>
- 12) OpenCV で顔検出
http://www.aianet.ne.jp/~asada/prog_doc/opencv/opencv_obj_det_img.htm
- 13) ISP imaging developers：クロマキーとは
<http://image-d.isp.jp/commentary/chromakey/index.html>
- 14) Kinect Fitting Room for Topshop：
http://www.youtube.com/watch?v=L_cYKfP1_0