

動画を用いたファッションカタログの複合マッチングへの拡張

石田正明^{†1} 市村哲^{†2}

電子ファッションカタログとしては現状では、写真からのシミュレーションや服を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ていないことや長袖等、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。この問題に対し著者らは過去において「Kinect ファッションカタログ」(K.F.C.)を提案した。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。しかし、上下のファッションコーディネートができないという問題があった。そこで今回、トップスとボトムスの登録および動きの認識を別々に行えるように拡張した。これにより、上下別々のモデルの服を組み合わせることが可能になった。評価実験として、本研究で作成したシステムを20代の被験者5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解答してもらった。その結果、本システムが他のシステムと比べてファッションコーディネートに役になったと回答した。

Extension to combinational matching for moving fashion catalog

MASAAKI ISHIDA^{†1} SATOSHI ICHIMURA^{†2}

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. In the past, the authors have proposed the "Kinect fashion catalog" (K.F.C.) for this problem. 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. However, there was a problem that you can't tops and bottoms fashion coordination. I have extended to allow separate registration and recognition of the movement of the tops and bottoms. This allows to combine the top and bottom clothes separate models are now available. In an experiment, the 5 men in their 20s were asked to use. As a result, it as useful to the coordinated fashion.

1. はじめに

服を購入する際に、カタログを参考にすることが多い。しかし、現状のカタログでは静止画像のため自分が見たい組み合わせや動きを見ることができない。この問題に対して2つの既存技術を参照する。1つ目の既存技術は、静止画像を用いたファッションコーディネートシステムの「suGATALOG」³⁾である。ユーザが着用した服を鏡で確認するタイミングで撮影し、その姿の静止画像をトップスとボトムスに分け、組み合わせることでシミュレーションするシステムである。2つ目の既存技術は、Kinect カメラ⁵⁾を利用したAR試着体験システム「Kinect Fitting Room」⁹⁾である。AR技術により、ディスプレイ上に3Dモデルで作成された服を自分の体に合成することができるシステムである。

しかし、「suGATALOG」は正面から撮影した静止画像のみで組み合わせを行うので、服の動きを見ることができない。「Kinect Fitting Room」では、3Dモデルの服を使用しているため、3Dモデルが存在する服しか利用できない。また、3Dモデルの作成をするには、専門的な知識、甚大なコスト、作成する時間を必要とする。長袖など服の形状は自分の体に合成することが難しい等の問題点もある。

この問題に対し著者らは過去において「Kinect ファッションカタログ」(K.F.C.)¹⁾を提案した。Kinect カメラ前で動くだけで画面上のカタログ内のモデルを人の動きに合わせて動かすことができるシステムである。また、カタログの画像はモデルが動くだけで作成できる。そのため、新しい服の登録が容易である他、服の形状に左右されず、また専門的な知識を用いることなく作成することができる。しかし、上下のファッションコーディネートができないという問題があった。また、カタログのモデルの動きがバタついて表示される頻度が多かった。

そこで今回、トップスとボトムスの登録および動きの認識を別々に行えるように拡張した。これにより、上下別々のモデルの服を組み合わせることが可能になった。また、バタつきの原因としてKinect カメラから取得できるデータ²⁾⁴⁾⁶⁾が不安定であることが実験でわかった。解決策として、取得データの標準偏差を求めて活用するなどした。

本論文では、本システムの試作について述べる。評価実験として、本システムを20代の男性5名に使用させた。

2. 既存技術と問題点

本論文で提案するシステムは、既存研究であるK.F.C.の問題点を解決したシステムである。そこで既存研究であるK.F.C.について述べる。

図1にK.F.C.のシステムの全体構成を示す。K.F.C.は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

^{†1} 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

^{†2} 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

アクションカタログシステムである。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つで構成されている。

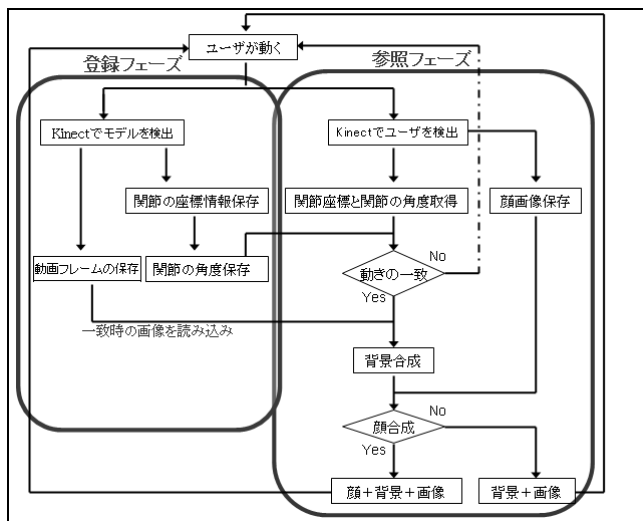


図1 K.F.C.のシステム構成 1)

Figure 1 K.F.C. system architecture

2.1 登録フェーズ

服を着て動くモデルの姿を1秒間に30コマの連続静止画像として保存する。Kinectカメラが取得した骨格情報(両手首・両肘・両肩・両肩の中央の座標)8と両肩・両肘の角度の値を保存した画像に対応付けて記録する。

2.2 参照フェーズ

Kinectカメラの前でユーザが動くと、その動きと一致する動きをしたモデルの画像を表示する。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。登録フェーズで保存した骨格情報から値を取得して、誤差15度以下の画像を表示する。図2は参照フェーズの実行画面である。左はユーザの画面で、右はユーザの動きと一致したモデルの画像である。図3は参照フェーズでのシステム動作である。①はユーザの動きの情報から、カタログ画像群と一致する動きの画像を検索する。②は一致した画像を表示する流れを示す。この検索と表示が高速に実行されるため、モデルの動きが動画像のように見える。

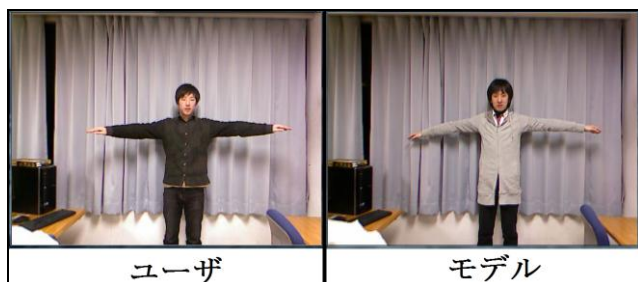


図2 K.F.C.の実行画面 1)

Figure 2 Execution screen of K.F.C.

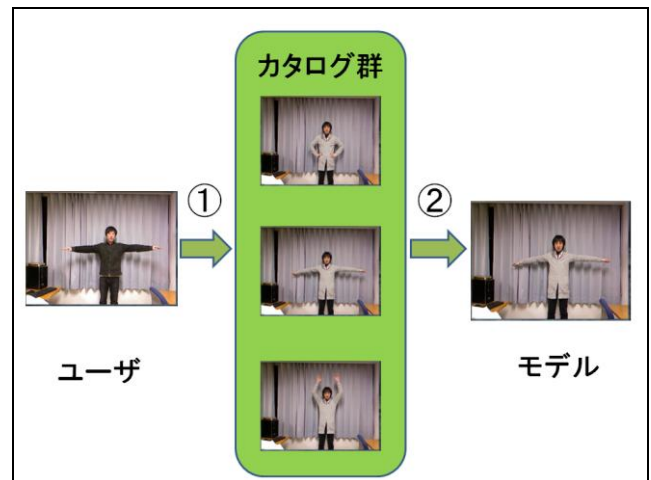


図3 参照フェーズの動作 1)

Figure 3 Operation of a reference phase

2.3 問題点

既存研究であるK.F.C.はトップスの動きしか認識しておらず、上下のファッションコーディネートが不可能なため、動くカタログとしては不十分であった。また、カタログのモデルの動きが、バタついて表示される頻度が多いのでユーザに違和感をあたえる問題点が挙げられた。

3. 提案

トップスとボトムスを別々にマッチング処理する拡張方法について提案する。マッチングとは、ユーザの動きから登録されているモデルの動きを検索し、該当する画像を特定することを示す。トップスとボトムスを、別々にマッチングを行うので色々な組み合わせをすることができる。また、全身の動きのマッチング方法と比較し、登録画像が少なく済むことが期待される。

バタついて表示される問題点は、後述する事前実験より、Kinectカメラが取得できる骨格情報にノイズが含まれることが原因であることがわかった。ノイズを軽減させるため標準偏差を活用する。

4. 事前実験

事前実験としてKinectカメラから取得できる骨格情報に対して2つの実験を行った。1つ目として、人を認識している範囲とRGBカメラに写っている被験者とのズレについて実験を行った。図4にKinectカメラが認識している人と被験者を透過合成して示す。図4より、Kinectカメラで取得した骨格情報と被験者の姿とにズレが生じていることが分かる。また、Kinectカメラが取得した腰の高さが被験者の腰の高さより高く認識されている。

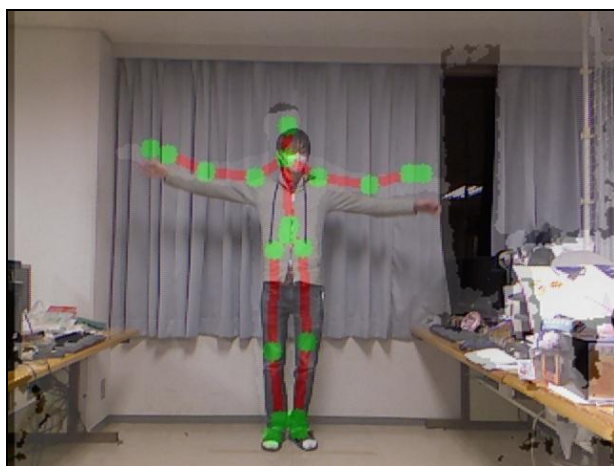


図 4 Kinect カメラのズレ

Figure 4 Gap of a Kinect

2 つ目として、Kinect カメラから骨格情報が安定して取得できるかを実験した。被験者が Kinect カメラ前に動かず立ち、本システムで使用する 14 点の座標の値を 10 秒間取得した。

表 1 に個々の骨格情報の標準偏差の値を示す。左膝は 0.44 とバラつきが小さく安定したデータを取得しているが、左足首では、37.73 とバラつきが多い。全体の平均が 9.22 と高いことから、取得する骨格情報にバラつきが多いことがわかった。

表 1 取得データの標準偏差

Table 1 Standard deviation of acquisition data

	X 座標	Y 座標	Z 座標	平均値
両肩の中央	1.93	13.79	6.20	7.31
臀部の中央	4.59	1.99	11.66	6.08
右肩	3.82	6.15	16.73	8.90
右肘	7.92	12.89	19.24	13.35
右手首	11.12	18.06	21.30	16.83
左肩	3.93	5.35	7.60	5.63
左肘	4.52	4.43	6.87	5.27
左手首	4.43	5.41	14.87	8.24
臀部の右側	4.87	2.12	16.51	7.83
右膝	0.46	3.37	11.47	5.10
右足首	4.11	15.22	7.60	8.98
臀部の左側	6.91	3.54	11.25	7.23
左膝	0.44	3.01	10.93	4.79
左足首	13.23	37.73	19.78	23.58
平均値	5.16	9.50	13.00	9.22

本システムではトップスとボトムスの合成を行うので、Kinect カメラから取得する腰の高さとユーザの腰の高さの誤差補正を考慮する必要がある。また、Kinect カメラから取得する骨格情報にバラつきが多いため、ノイズを除去する必要がある。

5. 実装

マッチング方法としては、図 5 で示す両肩・両肘・股関節・両膝の角度をマッチングするようにする。図 5 の①はトップス、②はボトムスの骨格情報をマッチングで使用する。本システムでは、Kinect カメラで取得できる関節情報は 20 点の内、両肩の中央・両肩・両肘・両手首・臀部の中央・両臀部・両膝・両足首の 14 点を使用する。まず、Kinect カメラから関節情報の XY 座標と Kinect カメラからの距離である Z 座標を取得する。角度を求めたい関節の座標と両隣接の関節の座標の 3 点の座標から、内積を求め関節の角度を算出した。なお、肩の角度は 180 度以上曲げることができるため、肩の Y 座標の値と肘の Y 座標を比較し、肩の Y 座標の値が大きい場合は、180 度以上の値となるようにした。また、Kinect が取得できる情報にノイズが含まれるため、標準偏差を活用し、ノイズを除去した。本システムは、登録モードと参照モードの 2 つから構成される。

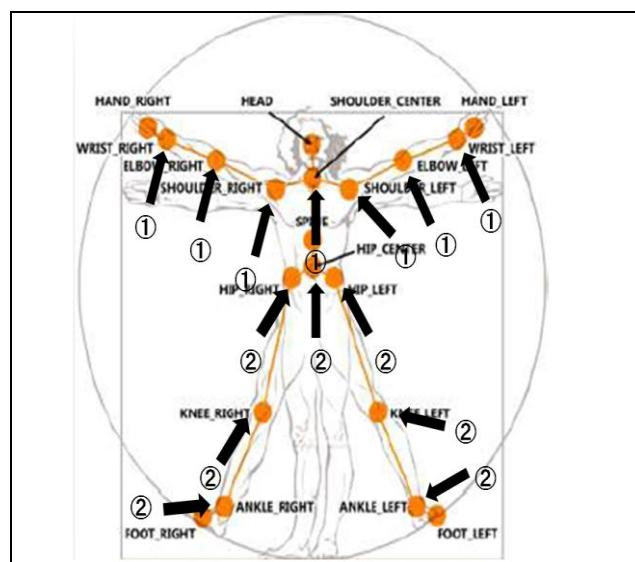


図 5 関節取得図 5)

Figure 5 Figure joint acquisition

5.1 登録モード

図 6 に登録モードのシステム構成を示す。服を着て動くモデルの静止画像を 1 秒間に 30 コマ保存する。Kinect カメラが取得する腰の高さがユーザの腰の高さより高いため、この問題を解決する必要がある。そこで、モデルとなった人にトップスとボトムスの境の高さを右手で示してもらったようにした。Kinect カメラが取得する腰の高さと示してもらった右手の高さの差分を算出する。その値を Kinect カメラが取得する腰の高さに反映させ、トップスとボトムスを区別するようにした。トップスのマッチング情報（両肩・両肘の角度、モデルの腰の X 座標・Y 座標）とボトムスのマッチング情報（股関節・両膝の角度、モデルの腰の X 座標・Y 座標）を保存した。各角度を 5 フレームごとに区切り平均値と標準偏差を求める。個々のフレームごとに平均

値と標準偏差の差分を算出する。標準偏差の値より小さければ、該当する画像とマッチング情報を保存する。

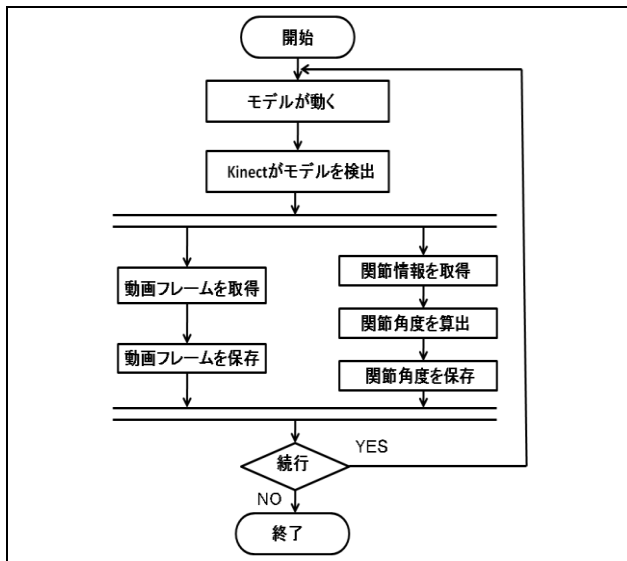


図 6 登録モードのシステム構成

Figure 6 System architecture of register mode

5.2 参照モード

図 7 に参照モードのシステム構成を示す。Kinect カメラの前にいるユーザの両肩・両肘・股関節・両膝の角度を 5 フレームごとに、中央値を算出する。ボトムスの場合、股関節と両膝の角度を誤差 10 度以内で一番近い値に対応する静止画像を表示する。ボトムス画像を表示する場合、マッチング情報からモデルの腰の X 座標と画像の X 座標中央の差分を算出する。その値から、出力する画像が中央に表示されるように横に移動して、腰の高さより低い範囲を出力する。トップス画像を表示する場合、マッチング情報からモデルの腰の X 座標と画像の X 座標中央の差分を算出する。その値から、出力する画像が中央に表示されるように横に移動して、腰の高さより高い範囲を出力する。そして、トップスとボトムスの画像を合成して表示する。実装画面は図 8 のようになる。左はユーザの画面で、右はトップスとボトムスを合成したモデルの画像である。ユーザが登録されていないモデルの動きをした場合は、その前に出力した画像のまま画面に表示する。

図 9 は参照モードの動作である。左が手と足を広げているユーザの画像、中央はカタログ群で、上が手を広げて足を閉じているモデルの画像、下が手を閉じて足を広げているモデルの画像、右がトップスとボトムスを合成した画像を示す。①はユーザの動きを元にカタログ画像群からトップスとボトムス個々に一致した動きの画像を検索し、②は個々に一致した画像を合成して表示する流れを示す。

トップスとボトムスを、別々にマッチングを行っているため、例えば、上下別々のモデルの服を組み合わせることが可能になった。この検索と表示が高速に実行されるため、モデルの動きが動画像のように見える。

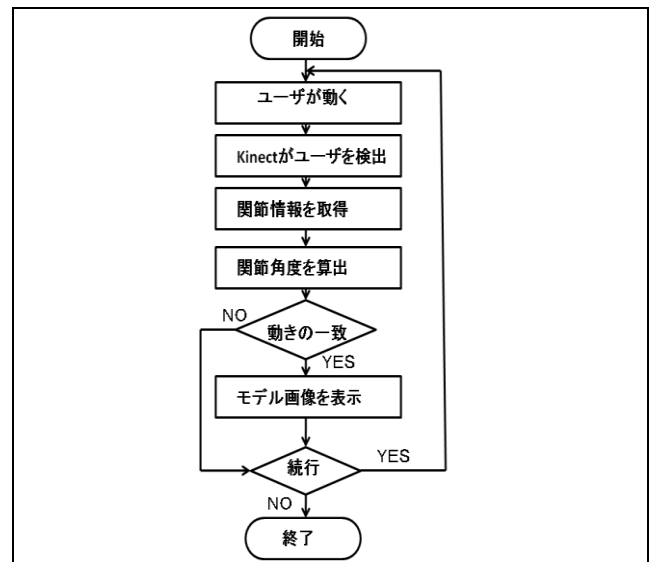


図 7 参照モードのシステム構成

Figure 7 System architecture of reference mode

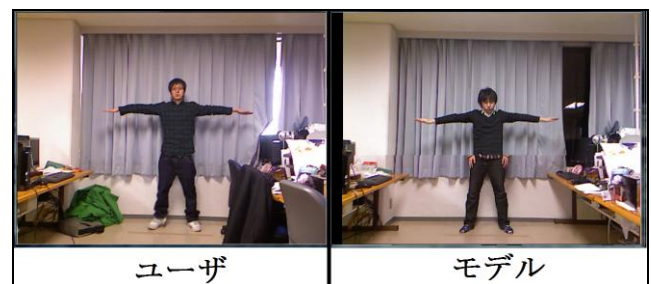


図 8 実装画面

Figure 8 Execution screen

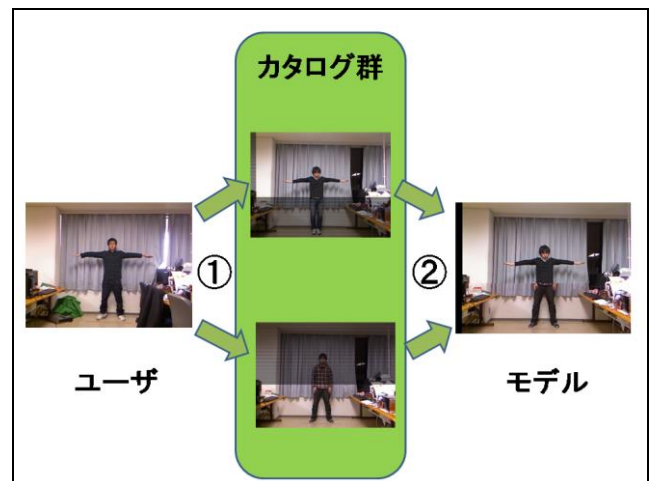


図 9 参照モードの動作

Figure 9 Operation in reference mode

6. 評価実験と結果

6.1 上下合成の評価

新システムで追加したトップスとボトムスの合成の必要性と合成した画像の正確性について20代の被験者5名を対象に5段階評価を行わせた。(5:とても良い~1:とても悪い)

表2に上下合成の評価の比較結果を示す。追加した上下合成機能は平均値5.0と高かった。しかし、合成した画像表示の正確性についての評価は平均値2.6と低かった。

表2 上下合成の評価

Table 2 Evaluation of top-and-bottom composition

被験者	A	B	C	D	E	平均
上下合成は必要性	5	5	5	5	5	5.0
合成した画像表示の正確性	4	3	1	3	2	2.6

6.2 バタつきについての評価

6.2.1 1秒間あたりのバタつき数

K.F.C.と本システムが一秒間のバタつき数について比較を行った。両手を横にする(ポーズA)、両手を挙げる(ポーズB)、両手を横に広げる(ポーズC)、両手を腰に当てる(ポーズD)、腕組みをする(ポーズE)の5種類のポーズを1分間維持して表示されるバタつきの数を計測した。K.F.C.はボトムスのマッチングは実装されていないので、トップスの結果のみを反映させる。

表3に1秒間あたりのバタつき数の比較結果を示す。平均値では、新システムのバタつき数が少なかった。しかし、ポーズAはK.F.C.のバタつき数が少なかった。

表3 1秒間あたりのバタつき数

Table 3 Number of errors per second

	K.F.C.	本システム
ポーズA	0.98	4.75
ポーズB	4.33	1.92
ポーズC	4.21	3.92
ポーズD	13.07	6.92
ポーズE	5.52	1.25
平均値	5.62	3.75

6.2.2 評価アンケート

マッチング精度について20代の被験者5名に5段階評価を行わせた。(5:とても良い~1:とても悪い)

表4にマッチング精度の比較結果を示す。K.F.C.の平均値2.0, 新システムの平均値3.8と新システムの評価が高かった。

表4 マッチング精度

Table 4 Matching accuracy

被験者	A	B	C	D	E	平均
K.F.C.のマッチング精度	2	2	2	2	2	2.0
新システムのマッチング精度	4	3	4	4	4	3.8

バタつきの度合いについて20代の被験者5名に5段階評価を行わせた。(5:バタつきが目立たない~1:バタつきが目立つ)

表5にバタつきの度合いの比較結果を示す。K.F.C.の平均値1.2, 新システムの平均値4.2と新システムの評価が高かった。

表5 バタつきの度合い

Table 5 Degree with error

被験者	A	B	C	D	E	平均
K.F.C.のバタつき度合い	1	2	1	1	1	1.2
新システムのバタつき度合い	4	4	4	5	4	4.2

6.3 マッチング成功率

20代の被験者5名に登録したモデルの動きと同じ動きをしてもらい、K.F.C.と本システムのマッチング成功率を比較した。被験者の動きに対応する画像が表示されたらマッチング成功と判定する。また、K.F.C.はボトムスのマッチングは実装されていないので、トップスのみの結果を反映させる。

表6にマッチング成功率の比較結果を示す。K.F.C.が平均94.75%, 本システムが平均95.08%となった。

表6 マッチング成功率

Table 6 Success rate matching

	K.F.C.(%)	本システム(%)
A	87.58	91.30
B	95.53	96.64
C	96.65	94.26
D	96.30	97.14
E	97.69	96.05
平均	94.75	95.08

6.4 他カタログとの比較

20代の被験者5名に服を着たモデルの静止画像と既存研究のK.F.C.と本システムの3つを、ファッションカタログとして参考になる順番に並べてもらった。

表7に他のシステムとの比較結果を示す。これにより、本システムが最もファッションカタログの有効性が高いことがわかった。

表7 他カタログとの比較結果(順位)

Table 7 Results of comparison with other catalog

被験者	A	B	C	D	E	平均(順位)
モデルの静止画像	2	3	1	2	1	1.8
旧システム(K.F.C.)	3	2	3	3	3	2.8
本システム	1	1	2	1	2	1.4

7. 考察

トップスとボトムスの合成の組合せ機能についてアンケート結果から K.F.C.の平均値が 5.0 であり被験者全員が上下合成を必要と答えた。しかし、合成した画像表示の正確性については平均値が 2.6 と低かった。その理由として、トップスとボトムスがズレて合成し、表示されるためと考えられる。上下合成に対して標準偏差を活用したノイズ除去処理は有効ではなく、他の方法を検討する必要がある。

既存技術の K.F.C.と本システムのバタつきについて比較を行った。本システムが全ポーズの 1 秒間あたりの平均バタつき数が 3.75 と K.F.C.と比較して少なかった。バタつき度合いについても K.F.C.は平均値が 1.2 で新システムは、4.2 と新システムの方が高かった。このことから、本システムは、K.F.C.と比べてバタつきについての評価が高かった。

マッチングの成功率の比較では、K.F.C.は 94.75%で新システムは 95.08%と差がないことがわかった。マッチング精度のアンケート結果から、K.F.C.は平均値が 2.0、新システムは平均値 3.8 と新システムの方が高い評価だった。このことから、今回加えたノイズ除去処理によって必要なデータまでもが削られてしまう問題はないことが確認できた。

本システムが他のシステムと比べてファッションコーディネートに役にたったかについて検討した結果、本システムが 1 番有用であるという結果を得た。このことから本システムはファッションコーディネートに有用であることが示された。

8. まとめ

本システムでは、トップスとボトムスを別々にマッチング処理する拡張方法について提案した。Kinect カメラが取得できる人間の関節の座標データを利用してカタログのモデルの動きを登録する登録モードと登録した動きを参照する参照モードの 2 つのモードで実装した。登録した服のマッチング方法は、トップスのマッチング情報（両肩・両肘の角度、モデルの腰の X 座標・Y 座標）とボトムスのマッチング情報（股関節・両膝の角度、モデルの腰の X 座標・Y 座標）を別々に検索して、一致したらモデルの画像を合成して表示した。

また、Kinect カメラから取得できる骨格情報に多くのノイズが含まれていることがわかった。ノイズを除去してデータの精度の向上の処理が必要であり、その点も考慮したシステム構築を行った。

Kinect カメラが取得できる骨格情報に含まれるノイズ除去のために、登録するデータを 5 フレームごとの標準偏差を活用し、ノイズを除去した。また、参照する場合は 5 フレームごとに中央値を求めて、その値を現在の情報としてマッチングを行った。

評価実験では、被験者に新システムで拡張した上下合成

の評価、バタつきについての評価、マッチング成功率、他のカタログとの比較を行った。結果として、上下合成への拡張性は有効であったが、ユーザが満足するまでにはいかなかった。トップスとボトムスの合成にズレが生じてしまったのが原因である。今後の課題として検討する必要がある。標準偏差を用いたノイズの除去については一定の評価を得られた。しかし、ポーズによっては標準偏差を活用しない方が良い結果もあったので、今後検討していく必要がある。マッチング成功率も高くノイズ以外の必要なデータは削られていなかった。また、本システムは K.F.C よりファッションコーディネートに役立つことがわかった。

参考文献

- 1) 中島剛史, 石田正明, 市村哲: 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, 情報処理学会研究報告. GN, グループウェアとネットワークサービス 2012-GN-83(19), 1-6.
- 2) 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」.
- 3) 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
- 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 (参照日: 2012 年 10 月) <http://www.xbox.com/ja-JP/kinect>
- 6) Kinect SDK XNA で骨格情報の取得 (参照日: 2012 年 10 月) <http://www.kosaka-lab.com/tips/2011/06/kinect-sdk-xna.php>
- 7) OpenCV.jp (参照日: 2012 年 9 月) <http://opencv.jp/>
- 8) 正常関節可動域と計測法 (参照日: 2012 年 10 月) http://www.koganet.ne.jp/~care/joint_angle.htm
- 9) Kinect Fitting Room for Topshop (参照日: 2011 年 9 月) http://www.youtube.com/watch?v=L_cYKfP1_0&feature=player_embedded