

膝関節ラテラルスラスト映像の合成への Motion Transfer の適用可能性の検証

岡村 淳介[†] 三上 弾[†] 井原 拓哉[‡] 藤田 浩二[‡]

† 工学院大学情報学部コンピュータ科学科 〒163-8677 東京都新宿区西新宿1丁目24-2

‡ 東京医科歯科大学 〒113-8510 東京都文京区湯島1丁目5-45

E-mail: †j119069@g.kogakuin.jp, mikami.dan@cc.kogakuin.ac.jp, ‡ibara.fana@tdm.ac.jp, fujiorth@tdm.ac.jp

あらまし 本研究では、近年注目を集める Motion Transfer の技術を利用して、ラテラルスラストの所見を有する人物の歩行を他者に付与する。ラテラルスラストは、変形性膝関節症の所見として知られ、悪化の要因ともされる。そのため、早期発見が望まれるが、ラテラルスラストの動きは微小であり、診断基準が確立されておらず、この発見は難しい。本研究は、この問題の解決に向けて、若手医師・理学療法士(PT)がより多くの症例に触れる機会を創出することを目指すものである。そのために本研究では、Motion Transferにより、ラテラルスラスト症例の Data Augmentationを行う。具体的には患者の映像からラテラルスラストの動きを抽出し、これを他者に Motion Transferにより付与する。実験では、ラテラルスラスト患者の歩行映像とその動きを合成した映像の双方に対し、ラテラルスラストの度合いを、医師・PTに評価させ比較した。実験の結果、患者の実映像への評価と合成動画への評価に高い相関がみられ、ラテラルスラストが適切に合成できていることが確認できた。

キーワード Motion Transfer, 変形性膝関節症, ラテラルスラスト, Data Augmentation

Applicability of Motion Transfer to Generation of Knee Joint Lateral Thrust Video

Yusuke OKAMURA[†] Dan MIKAMI[†] Takuya IBARA[‡] Koji FUJITA[‡]

† Faculty of Engineering, Kogakuin University 1-24-2 Nishi-Shinjuku, Shinjuku-ku, Tokyo, 163-8677 Japan

‡ Tokyo Medical And Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo, 113-8510 Japan

E-mail: †j119069@g.kogakuin.jp, mikami.dan@cc.kogakuin.ac.jp, ‡ibara.fana@tdm.ac.jp, fujiorth@tdm.ac.jp

Abstract This study uses Motion Transfer technology, which has been gaining attention in recent years, to simulate the gait of a person with "lateral thrust". Lateral thrust is a typical symptom of knee osteoarthritis and is a factor in its deterioration. Early detection is desirable, but it is difficult because the movement of lateral thrust is minute and diagnostic criteria have not been established. This study aims to create opportunities for young physicians and physical therapists (PTs) to be exposed to more cases in order to solve this problem. Specifically, lateral thrust movements are extracted from the patient's video, and these motions are transferred to other people using "Motion Transfer". In the experiment, doctors and PTs evaluated the degree of lateral thrust in both a original and generated, i.e., motion transferred, videos, and compared them. The results showed a high correlation between the patient's evaluation of the original video and the evaluation of the generated video, confirming that the lateral thrust was synthesized appropriately.

Keywords Motion Transfer, osteoarthritis of the knee, lateral thrust, Data Augmentation

1. はじめに

高齢化が加速的に進行し、社会問題となっている。年齢を重ねることで発症しやすい病気のひとつとして変形性膝関節症が知られている。膝の痛みにより活動量が低下し、筋力が低下することで転倒のリスクが上がったり、出歩く機会が減ることで認知能力の低下につながったりと、多くの課題のきっかけになっているとも言われている。そのため、変形性膝関節症を早期

に発見し、介入により状態を維持する、あるいは進行を遅らせることには大きな意義がある。しかし変形性膝関節症には明らかな原因はなく、加齢に慢性的な機械的刺激が加わることで発症する[1]。そのため、早期発見、早期介入が難しいという側面がある。

変形性膝関節症の所見としてラテラルスラストがある。ラテラルスラストとは歩行時に見られる膝関節が横にぶれる現象であり、変形性膝関節症の悪化要因

及び重症度の指標として用いられる。ただしラテラルスラストは臨床の専門家（主に整形外科医師および理学療法士。以降、簡単のため単に専門家とする）の視認により同定されているのが現状である[2]。そのため、ラテラルスラストの発見は専門家のスキルに左右されてしまう。

このような背景から、ラテラルスラストを判断可能な専門家の育成、さらには、非侵襲的な手法によりラテラルスラストを判別可能なシステムの開発、などが求められる。専門家を育成する場合においても、機械学習システムをトレーニングする場合においても共通して重要なのは学習データの量であろう。多くの症例に触ることは良いトレーニングになるのは明らかであり、また、機械学習においても学習データの量は非常に重要である。しかしながら、実際の医療の現場において大量のデータを取得することは非常に困難であり、プライバシーの問題もあいまって現実的ではない。

そこで本稿では、近年研究が急速に進んでいる Motion Transfer の技術を用いてラテラルスラストの動きを健常者に付与することでラテラルスラストを有する歩行動画を合成することを提案する。以降、2 で関連研究を述べた後に、3 で本稿における Motion transfer の実装を示す。4 で被験者実験について述べた後、まとめる。

2. 関連研究

本稿では、ラテラルスラストを有する患者の動きを健常者に付与した映像を合成することを目指す。そこで本節では関連が深い Motion transfer に関する取り組みを述べる。

この取り組みの代表的な例のひとつとして Chan らによる”Everybody Dance Now”が挙げられる[3]。この研究では、骨格情報を経由して、異なる人物にモーションを合成することが可能である。モーション合成先の人物について、骨格画像と人物画像との対応関係を学習したモデルを作成する。その学習モデルにモーション元の骨格画像を入力することによりモーションの付与を実現している。なお、この研究の問題点として大きな計算機コストが挙げられる。実装には 12GB の GPU が必要であり簡易に手に入るとは言えない。

Chan ら[3]は、動画を対象としているが、フレーム毎に考えると、骨格画像と人物画像の対応関係をモデル化し、骨格画像から人物画像を合成可能な手法として pix2pix が挙げられる[4]。pix2pix は GAN ベースの Style transfer 手法である。線画から実画像、セマンティックセグメンテーションされた状態から実画像を推定するなど幅広いスタイル変換を実現している。



図 1 接近時と遠ざかる時の様子

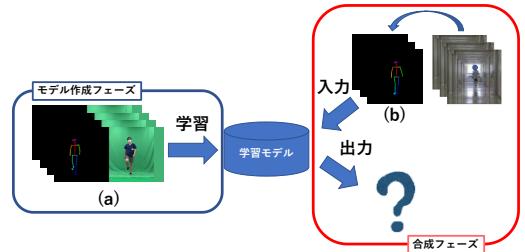


図 2 pix2pix による Motion Transfer の流れ

3. Motion transfer の実装

本研究では、ラテラルスラストを有する患者の動きの健常者への付与により、ラテラルスラスト動画を合成することを目指している。本稿の範囲で対象とする歩行は、画面内遠方から、カメラに向かって真っすぐ近づいてくる、および、画面内近くからまっすぐ遠ざかるの 2 種類であり、いずれも歩行者は前進している（図 1 参照）。歩行中の照明環境も安定しており背景に大きな変化もないことから、本稿の範囲では、実装のしやすさを重視し、フレーム毎に pix2pix を用いて画像の合成を行い、それらをつなぎ合わせることで合成動画とする。以下に詳細を述べる。

3.1. 全体構成

今回採用する実装の基本的な流れを図 2 に示す。実装は大きく 2 つのフェーズに分けることができる。モデル作成フェーズと、合成フェーズである。モデル作成フェーズは、モーション合成先人物に対して行う。人物画像と、当該画像から推定した骨格画像とをペアとし学習を行うことで、骨格画像から人物画像を合成可能とする。合成フェーズでは、まず、ラテラルスラストを有する患者の映像から骨格情報を抽出する。次に、それを作成済みのモデルへ入力することで、当該骨格情報を有する人物画像を合成する。これを各フレームに対して行うことで動画を合成することができる。

3.2. モデル作成フェーズ

モデル作成フェーズでは、人物画像に対して骨格抽出を行い、その対応関係を学習する。本稿の範囲では、骨格抽出には OpenPose を利用した[5]。またモデルの学習には pix2pix を利用した。学習用の画像のサンプ

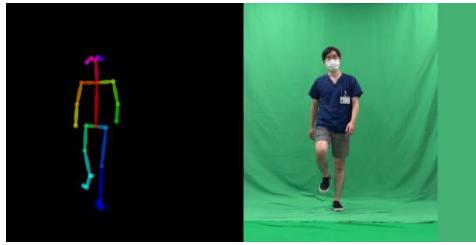


図 3 学習用画像サンプル



図 6 出力結果



図 4 足踏みの様子



図 7 クロマキー処理後



図 5 膝の横移動動作の様子

ルを図 3 に示す。実際の学習には、人物 1 名当たりは 3000 程度枚用意した。

モデル作成にあたり、モーション合成先人物には、グリーンバックの前で所定の動作を実施させた。ここで所定動作は、足踏み（図 4）および簡単な膝の横移動（図 5）である。これはモデルにラテラルスラストと類似した姿勢を含めるためである。これらを、前進に対応する前向きと、後退に対応する後ろ向きとで実施させた。1 名分の撮影に必要な時間は 1 分未満であった。

3.3. 合成フェーズ

合成フェーズでは、まず、ラテラルスラストを有する患者の映像から、骨格情報を抽出する。ここでの骨格抽出も当然ながら OpenPose を用いる。得られた骨格情報に基づいてフレーム毎に、3.2 に記載のモデルに適用し合成画像を生成する。なおこの段階では背景は図 6 に示す通り緑である。これを処理し、元の患者動画と同一の背景に貼りこむことで合成動画としている（図 7 参照）。

3.4. 奥行き変化への対応

本研究で対象とする動画は、奥行き方向に大きな移

動を伴う。それに対して、モデル作成時の撮影の手間を考えると、奥行き方向に大きな移動を伴う撮影は好ましくない（必要となるグリーンバックの領域やスペースの確保など）。モデル作成時に奥行の変化を制約した状態で、ここまで述べた手続きを実行した場合、図 8 に示す通りの結果となる。図 8 左はモーション元、右はモーション合成先である。上段と下段の違いは奥行きである。図 8 に示した通り、モデル作成時と同等の奥行においては適切な合成ができているものの、奥行きが大きく変化した下段では不適切な合成となっていることが見て取れる。

そこで、本実装では、モデル作成時に敢えて奥行変化を伴わない撮影を行い、モデル作成することとした。加えて合成時には、モーション元画像に対して以下の手続きを踏む。

1. OpenPose により骨格推定を行う
 2. 関節座標に基づき身長（ピクセル単位）を取得する
 3. モデル作成画像の身長と同等になるようスケーリング
 4. 改めて OpenPose により骨格情報を抽出
 5. 学習済みモデルを適用し合成
 6. 上記ステップ 3 の逆のスケールで変換
- 以上の各ステップにより、奥行き変化があっても良好な合成が実現する。その例を図 9 に示す。図 9 では幅広い奥行に対して適切な合成がなされているのが確認できる。



図 8 出力結果



図 9 出力結果(歩行変化対応)

4. 実験

4.1. 実験概要

提案手法の有用性を検証するため実験を行った。検証の主項目は、専門家の観点で適切なラテラルスラストが合成できているのか、である。

まず、モーション元として、左右の膝少なくともどちらかにラテラルスラストの所見がある患者 5 名の歩行映像を撮影した。そして、モーション合成先として、健常者 5 名についてモデル作成を行った。

評価実験では、被験者に、歩行映像を視聴させ、左ひざ、右ひざのそれぞれについて、ラテラルスラストの状態を回答させた。なお回答は、ラテラルスラストなしを意味する 0、および、症状のレベル 1~3 の 4 段階にて回答させた。

評価実験では、評価者を 2 グループに分けて実施した。第 1 グループはベテラン群であり、経験を 6 年以上有する整形外科専門医 5 名である。第 2 グループはルーキー群であり、経験の浅い整形外科医および PT から構成される 5 名である。

4.2. 実験の流れ

実験では、患者 5 名の元動画および合成動画患者 1 名につき 3 種類、つまり 5 人 × 4 (オリジナル + 合成 × 3) = 20 本の動画をランダムな順番で提示し、回答させ

図 10 アンケートサンプル

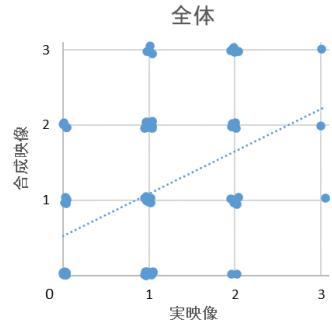


図 10 ベテラン群による評価

た。なお、回答にあたって映像視聴回数に制限は加えていない。アンケートは Google Forms を行って実施しておりそのサンプルを図 10 に示す。

4.3. ベテラン群による評価結果

ラテラルスラストが適切に合成されていれば、患者元動画に対する評価と、合成映像に対する評価には高い相関がみられるはずである。そこで、まず、患者オリジナル動画に対する評価と、当該患者の動きが付与された合成動画に対する評価の相関を調査した。その結果を図 10 に示す。図 10 で横軸は、患者元映像に対する評価値であり、縦軸は合成映像に対する評価値である。なお、プロットは格子状にしか存在しないため、頻繁に同じ位置にプロットが重なる。そこで分かりやすさのために小さなノイズを加えてプロットしている。図 10 について相関係数を求めるとき 0.456 となり、相関ありといえる水準であることが分かった。

ここで以降の詳細な分析にあたり、前提を再確認する。ラテラルスラストは動きが微小であり、専門家においても診断が困難である。このことから、まず、本実験で対象とした各膝について、元映像においてどの

表 1 実映像に対するベテラン群の評価

	回答者 1	回答者 2	回答者 3	回答者 4	回答者 5
膝 ID1	1	2	0	1	1
膝 ID2	1	1	0	0	0
膝 ID3	0	1	0	1	0
膝 ID4	2	1	1	2	2
膝 ID5	0	1	0	1	1
膝 ID6	1	3	1	2	2
膝 ID7	0	0	0	0	0
膝 ID8	0	0	0	0	0
膝 ID9	0	1	0	1	1
膝 ID10	2	1	1	1	2

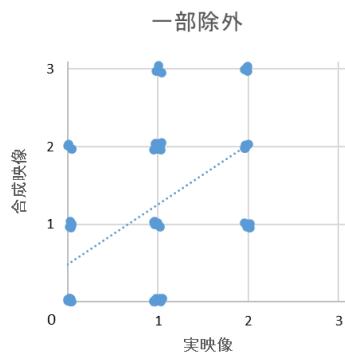


図 10 ベテラン群による評価(一部除外)

のような評価がなされたのか確認を行った。

4.3.1. 評価難易度の分析

その結果を表 1 に示す。まず、表 1 の黄色に塗られた膝 ID1 と 6 への評価に着目する。これらは、元映像であっても評価者間で最大 2 の差が出ておりベテラン医師の間でも評価が分かれる難しい例といえる。

次に、緑色で塗られた膝 ID7 と 8 はすべての医師が 0 と回答しており、ラテラルスラストなしと判断して問題ないと考えられる。また、水色で塗られた膝 ID2, 3, 5, 9 では 0 か 1 の評価がされている。つまりボーダーでラテラルスラストありとするかなしとするか微妙な膝といえる。最後に塗られていない膝 ID4 と 10 で回答は 1 か 2 となっており、全員の医師がラテラルスラストありと判断している例である。これらのクラス分けによりラテラルスラストは単純な数値データでのレベル分けは難しく、対象となる膝ごとに診断の難易度が異なることが確認できた。

4.3.2. 評価難易度を考慮した結果の分析

専門家でも評価が割れて判断が難しかった黄色の膝を除いて再度相関を調べた。散布図を図 10 に示す。相関係数は 0.548 であり除外前の 0.456 からかなり相関が高まったことがわかる。

表 2 ベテラン群によるラテラルスラスト有無レベルでの評価割合

実映像に対する評価(ベテラン)	合成映像に対する評価	
	なし	あり
一 ボーダー	0.867	0.133
	0.383	0.617
あり	0.067	0.933

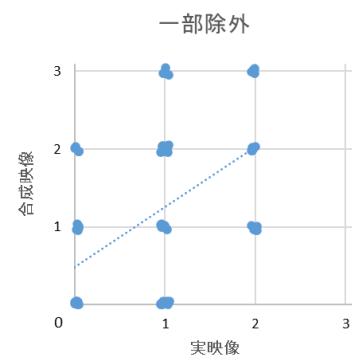


図 11 若手群による評価(一部除外)

次に、評価の詳細度を下げた検証を行った。前述の通り、0 から 3 の 4 段階での評価ではベテラン群の判断は大きく割れている。そこで、実映像においては回答がすべて 0 なものをラテラルスラストなし、0 か 1 だったものをボーダー、1 か 2 であったものをラテラルスラストありとした。これに対して合成映像においては 0 であればなし、1 から 3 であればありとして評価を丸めて再度確認を行った。その結果を割合でまとめたものを表 2 に示す。表 2 を見ると実映像でなしと判断されたものについては約 9 割でなしと判断されている。また、ありとされたものについては 9 割以上でありとされている。また、ボーダーとされたものにおいては評価が割れるという結果となった。このことから、実映像での評価傾向がかなり正しく合成映像に反映されていることがわかる。

4.4. 若手群による評価結果及び比較

ベテラン群と同様に実動画への評価と合成動画への評価の相関を調べた。散布図を図 11 に示す。なお、この散布図はベテラン群の実映像への評価で評価が割れ、判断が難しい膝(表 1 の黄色の例)を除いたものである。相関係数は 0.193 となり相関ありとはいえない水準である上、ベテラン群と比べ低下した結果となった。

また、詳細度を下げた検証では表 3 のようになり、ラテラルスラストなしへの評価の一一致率は約 9 割と高いがラテラルスラストありにおける一致率は 6.5 割とベ

表 3 若手群によるラテラルスラスト
有無レベルでの評価割合

		合成映像に対する 評価	
		なし	あり
実映像に 対する評 価(ベテラ ン)	なし	0.875	0.125
	ボーダー	0.662	0.338
	あり	0.35	0.65

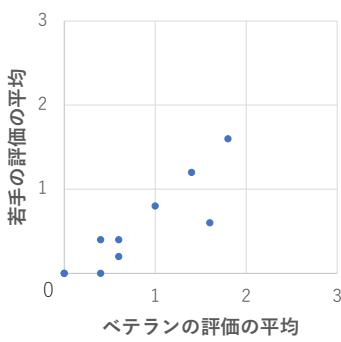


図 12 実映像への評価の平均

テラン群と比べ低くなっている。ボーダーへの評価ではベテラン群と同様に評価が割れている。

また、実映像を対象として各膝へのベテラン群及び若手群の評価の平均を散布図で表したものと図 12 は、合成映像を対象としたものを図 13 に示す。図 12 では全体の傾きは 1 に近く実映像においては両グループの評価はおおむね等しいと考えられる。一方、図 13 では傾きが緩やかになっており、若手群は合成映像においては判断精度が低下し、ラテラルスラストなしの方向に寄った評価をする傾向が見て取れる。

この原因のひとつであると考えられるのが合成による画質の低下である。合成映像は実映像に比べノイズがあり、判断の難易度が上昇していると推測される。ベテラン群は画質がある程度低下しても判断精度を維持できるが、若手群はうまく判断できなくなると考えられる。このことからベテラン群には合成動画でも見抜けるが、若手群には見抜けないラテラルスラストの判断における着眼点があると推測できる。それを合成動画によるトレーニングで獲得できれば臨床の場においてこの手法を役立てられると期待できる。

5. おわりに

本研究では、ラテラルスラストの診断および若手臨床家の育成における課題の解決策として Motion Transfer によるラテラルスラスト症例の Data Augmentation 手法を提案した。また、学習モデルへの

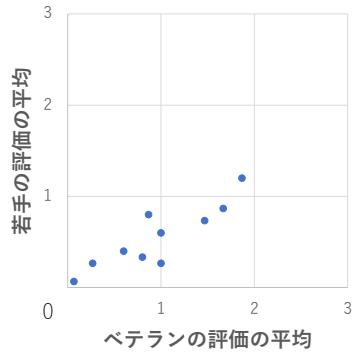


図 13 合成映像への評価の平均

入力用画像にスケーリング処理を行うことで、学習データの拡張なしに奥行変化に対する頑健性を実現した。

実験では、ベテラン群による評価により実映像に対する評価と合成映像への評価に相関がみられたことに加え、合成元となった人物の各膝の評価傾向が反映されていたことから、ラテラルスラストが適切に合成できていることが確認できた。ベテラン群と若手群を比較すると、若手群は実映像にはベテラン群と同程度の精度で判断できているが、合成映像においては精度が落ちることがわかった。

今後の展望としては、若手医師のトレーニングにこの手法で作成した合成動画を使用し、その効果を検証していく必要がある。

文 献

- [1] 立花陽明，“変形性膝関節症の診断と治療”，理学療法化学，Vol 20, no 3, 2005
- [2] 井野拓実ら，“ラテラルスラストの動態，一変形性膝関節症例の歩行分析一”，第 49 回日本理学療法学会大会抄録集, Vol 41, 2014
- [3] C. Chan et al., “Everybody Dance Now,” Proc. ICCV, 2019.
- [4] P. Isola et al., “Image-to-Image Translation with Conditional Adversarial Nets,” Proc. CVPR, 2017.
- [5] Z. Cao et al., “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” IEEE T. PAMI, 2019.