

AnimE-Lise: 日本アニメを模した人物誇張映像のライブ合成

樋爪 真子^{†, a)} 苗村 健^{†, b)}

概要: 日本のアニメでは、手前のものを実際よりも大きく描くことにより、人物の遠近感を誇張する表現が使われ、視聴者により強い印象を与えるという効果がある。この誇張を 3D CG で実現した手法に E-IMPACT があり、手描きアニメだけでなく、CG でも人物の遠近感を誇張することが好まれることが分かっている。しかし、実写映像における 3 次元的な透視変換を考慮した誇張については、まだ検討されていない。そこで本稿では、E-IMPACT の考え方を実写に適用する際の問題点を明らかにし、実写映像において人物の遠近感をリアルタイムに誇張するシステムを提案する。Microsoft 社製の Kinect for Windows v2 を使い、深度情報や人物の骨格情報をもとに人物の骨格構造を把握した。そして、ある関節を基準として手前の部分に擬似カメラを設置する。2 つのカメラからの画像を合成することにより、リアルタイムに誇張された映像を実現した。このシステムを展示会で公開し、300 人の人に体験してもらった。体験者が誇張された映像を写真に収めたり、身体だけでなく手に持っているものまでも誇張させようとしていたりする姿が多く見受けられ、デジタルサイネージへの有用性が示唆された。

AnimE-Lise: Anime-like Exaggeration of Live-action Image using Skeleton Structure

MAKO HIZUME^{†, a)} TAKESHI NAEMURA^{†, b)}

Abstract: In Japanese hand-drawn animations, perspective is often exaggerated by drawing close objects larger than the actual scales. This exaggeration produces impressive pictures for human. Recently, a method that achieves this exaggeration effect in 3D CG, called E-IMPACT, has been proposed and it has shown that exaggerated perspective is effective and preferred even for CG animations. However, this method can be used only for CG animations, and exaggeration for live-captured videos has not been examined. In this paper, we propose a system of perspective exaggeration for live-action videos in real-time. Our system uses human skeleton structures and depth information by Kinect for Windows v2. Using this 3D model information, we synthesize an image from a pseudo-camera in a closer position and combine it into the original color image. Our system was evaluated by 300 users in public exhibitions and we found a potential application of our system for digital signage.

1. はじめに

日本のアニメでは、実世界にはないような人物の誇張された表現が多く使われている。それは例えば、目が実際より大きく描かれていたり、動きが大げさであったりすることである。こういった誇張は、実際の生活ではあり得ることのない大きさ、あるいは動きであるのだが、視聴者としては、より印象的な映像として見ることができる。本研究では、これら誇張表現の中でも、人物の遠近感を誇張した表現に着目する。この表現は、迫力のあるシーンで特に使われることが多い。遠近感を誇張することで、物理的には正しくない構図になるが、前に出ているものが実際より大きく描かれるので、人の注目を集め、より迫力を与えることができる。

この遠近感の誇張を 3D CG で実現した方法に E-IMPACT がある[1][2]。それまで CG では、正確な透視投影に限られる傾向にあったが、あえてその正しさを崩し、アニメのように遠近感を誇張することで、CG でも手描きのように柔

軟な表現をすることを可能にした。好ましいと思う誇張の度合いを調べるユーザーテストにおいては、実際の大きさよりも誇張した方が好まれる傾向にあることが示されている[2]。以上より、これまで手描きや CG では、人物の遠近感を誇張した構図が使われ、好まれてきた。しかし、実写映像における誇張については、プリント倶楽部での目の拡大のような 2 次元画像処理のレベルに留まっており、特に動画像に対して 3 次元的な透視変換を考慮した手法はまだ検討されていない。

本稿では、自分自身のパンチやキックといった動作が、インタラクティブに誇張されるシステムについて検討する。近年インタラクティブなデジタルサイネージが増えてきている。インタラクショナルができることによって、自分ですべて楽しむ、あるいは他の人がやっているのを見て楽しむという付加価値が付き、広告に対する注目度を高めることもできるだろう。

そこで、E-IMPACT の考え方を実写に適用する際の問題点を明らかにし、実写映像において人物の遠近感をリアルタイムに誇張するシステムを提案する。この提案システムを、Anime-like Exaggeration of Live-action Image using Skeleton Structure から AnimE-Lise と呼ぶ。

[†] 東京大学
The University of Tokyo
a) hizume@nae-lab.org
b) naemura@nae-lab.org

2. 関連研究

2.1 実写を漫画風に加工する研究

実写画像を漫画風に加工する研究はこれまでにいくつかなされてきた。奈良らによる瞬刊少年マルマル[3]では、ストーリーを持つ漫画に対して自らが登場人物となり、漫画の中に入り込めるインタラクティブシステムを提案している。また、スマートフォンのカメラで撮影した画像が漫画のように加工される漫画カメラ[4]というスマートフォンアプリがあり、人々は実写を漫画風の表現にすることを手軽に楽しむことができるようになっている。しかしこれらは静止画を撮影し、その画像を加工して漫画のように白黒にするものであり、リアルタイムにインタラクションできるシステムではない。Umeda ら[5]は、Kinect で取得した人物の骨格の動きや音声と、漫画でよく用いられる表現であるスピード線や集中線を対応させ、映像中に漫画的な効果をリアルタイムに加えることを実現した。

以上の研究は、どれも実写を漫画に入り込ませるものであって、実写を構図から変え、実写でありながらアニメのようにするというものはこれまでになかった。

2.2 E-IMPACT

E-IMPACT は、漫画やアニメにおける遠近感の誇張を 3D CG で実現した手法である。関節構造を持った人物 3D モデルに対し、関節から関節までの部位ごとに擬似カメラを設置し、手前にある部位を実際よりも近くから、遠くにある部位を実際より遠くから撮影する。各擬似カメラで撮った各部位の画像を元の画像に重ね合わせることで、誇張した画像を作り出している。カメラを動かす際、ただ位置を動かすだけではカメラ間の境界となる関節の部分が画像上で一致しない。そこで、それぞれのカメラの位置や、位置に応じた画角を自動で調節することにより、カメラ間の境界となる関節部分を画像上で繋がるようにしている。調節の仕方は、次の通りである(図 1)。まず、ある関節(図中の Elbow)を基準としてそこから 1 つ手前の関節(図中の Wrist)までの部位を擬似カメラが写すとする。擬似カメラは、元のカメラと、基準となる関節の 1 つ手前の関節を結ぶ線上に設置する。その際、基準とした関節がある平面において、元のカメラと擬似カメラが同じ範囲を写すように擬似カメラの画角を設定する。そうすることで、基準とした関節の画像上での位置を、元のカメラと擬似カメラで一致させることができる。1 つの擬似カメラを設置するためには、基準となる関節とその 1 つ手前の関節の合計 2 つの関節の 3 次元位置が分かれば良いことが分かる。

AnimE-Lise では、E-IMPACT の誇張の考え方を適用して、実写における人物の遠近感の誇張を実現するが、CG とは違い、実写では人物のモデル化が難しいという問題がある。そこで本研究では、Kinect for Windows v2 (以降 Kinect v2) [6]を用いることで、深度情報や人物領域、骨格情報を取得し、人物の骨格を把握し、誇張を実現する。

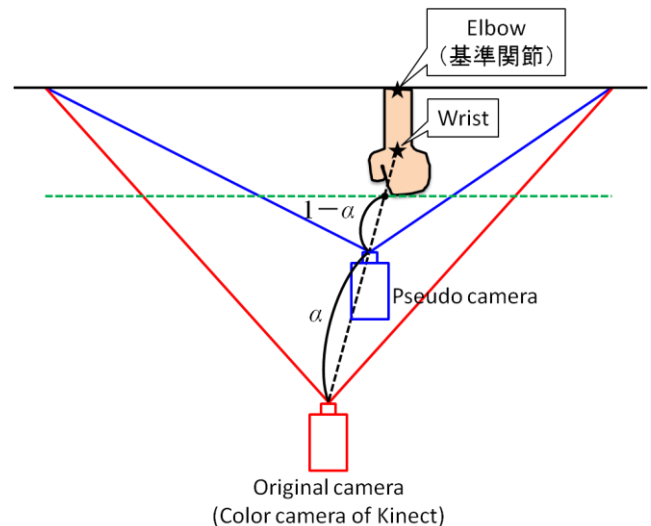


図 1 擬似カメラの設置

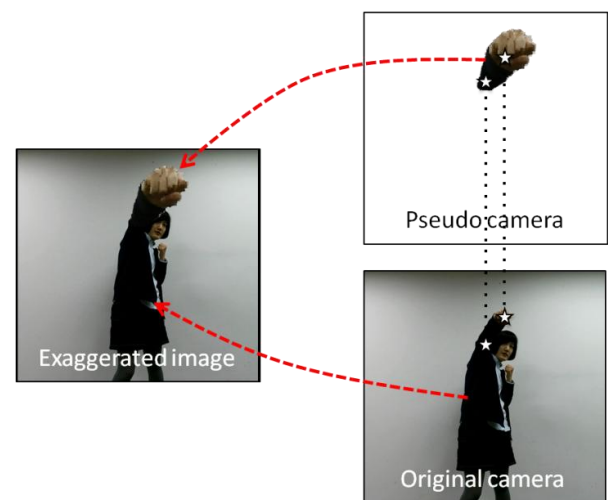


図 2 合成の模式図

3. 提案システム

3.1 システムの概要

AnimE-Lise では、E-IMPACT と同様、人物の骨格構造を用い、ある関節を基準として手前の部分に擬似カメラを設置する。2 つのカメラからの画像を合成することにより、誇張された映像を得る(図 2)。

現段階で同時に 1 人、かつ右腕、左腕、右脚、左脚のうち 1 箇所誇張が可能である。ここでは、腕は肘を基準として手側の部分、脚は膝を基準として足側の部分を指す。

3.2 システムの構成

Kinect v2 でデータを取得してから誇張画像を得るまでの手順を図 3 に示す。合計 7 ステップで 1 フレームを構成する。以下では、ある関節箇所 (joint type) の z 座標を $z(\text{joint type})$ で表すこととする。

(0) 事前準備

Kinect v2 のカラーカメラとデプスセンサーのキャリブ

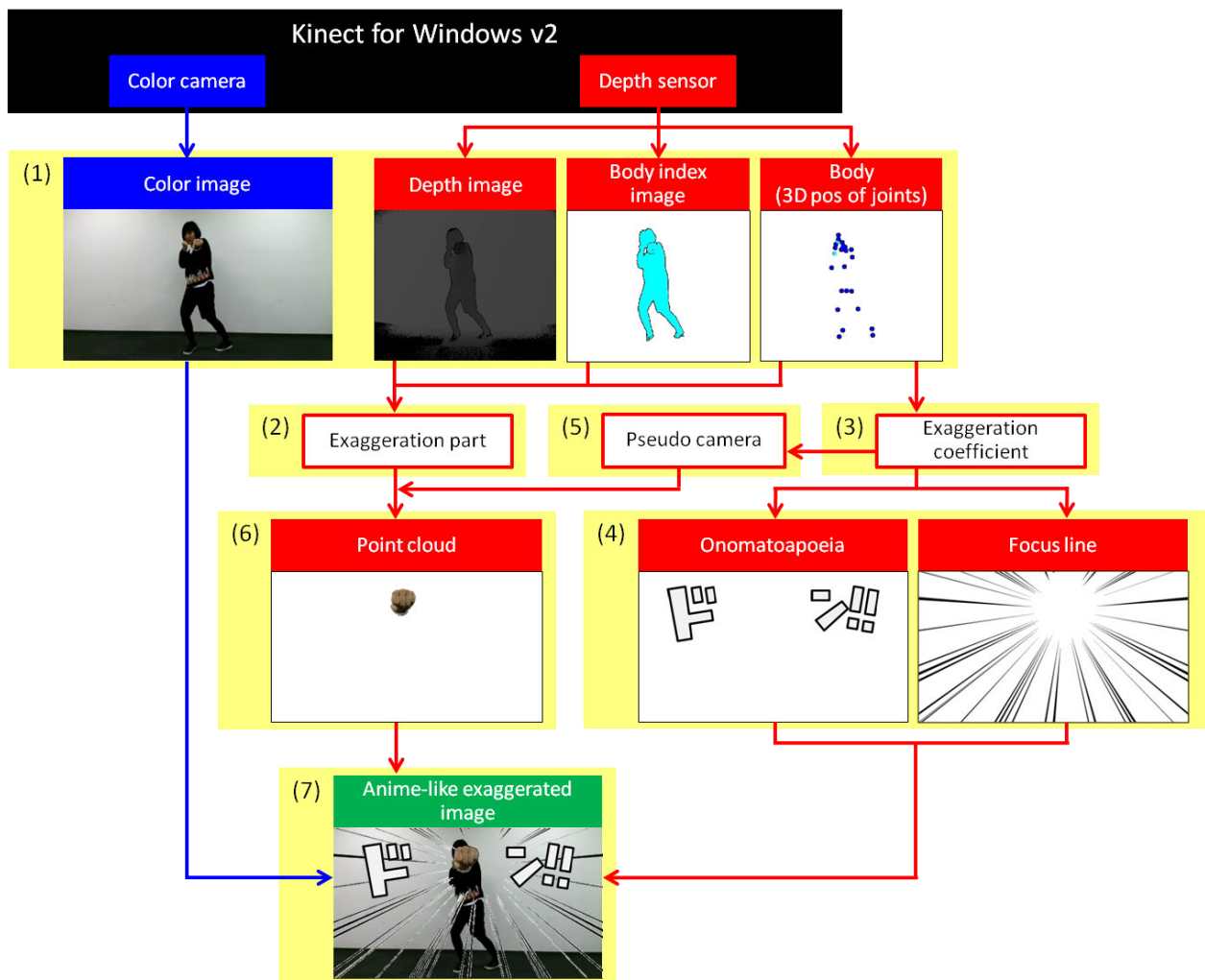


図 3 システム構成

レーションをし、デプスセンサーからカラーカメラへの変換行列を求めておく。

(1) Kinect v2 からデータを取得する

カラー画像、深度画像、人物領域（画像上のある 2 次元座標において、その画素が人物領域であるか否か、またそれが何人目であるか）、骨格情報（1 人あたり 25 箇所の関節の 3 次元座標）を取得する。ただし、人が複数いる場合は、 $z(\text{Base of the spine})$ が最も小さい人、すなわちカメラから見て、一番手前にいる人に関してのみ骨格情報を保存する。

(2) 誇張部分の抽出

右肘、左肘、右膝、左膝のうち、最も手前にある関節を求め（以後、この関節を基準関節と呼ぶ）、基準関節より手前にあり、かつ人物領域である部分を抽出する。しかし、基準関節より前にその部位以外のものが存在する場合、他の部位まで抽出してしまうという問題がある。そこで更に次のような手順をとる。基準関節より手前にある人物領域をマスクとして保持する。マスク画像上において、連結成分をグループ化し、基準関節と最も近いグループのみを残す。そして残った部分の各ピクセルにおいて、Kinect v2 から取

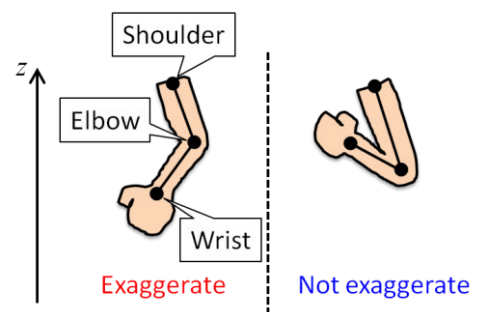


図 4 誇張の判定

得した 25 個の全ての関節との 3 次元上での距離を求め、一番近い関節が誇張箇所へ属する関節である場合、そこは誇張箇所として残す。最終的に残った部分が誇張箇所となる。

ここで同時に、誇張するかどうかの判定も行ふ。本研究では、胴より手前にあるものを誇張する。そのため、胴より手前に腕や脚がない場合は誇張する必要がない。誇張するかどうかの判定方法は次の通りである。基準関節が肘の場合は、 $z(\text{Wrist}) < z(\text{Elbow}) < z(\text{Shoulder})$ となるとき誇張する（図 4）。基準関節が膝の場合は、 $z(\text{Ankle}) < z(\text{Knee}) < z(\text{Hip})$ となるときに誇張する。そしてそれ以外

のときは誇張しない。誇張する場合は(3)へ進み、誇張しない場合は(1)に戻る。

(3) 誇張係数 α を定める

α は、擬似カメラの位置を決める値である。本研究では、E-IMPACT と同様、図 1 のように、元のカメラと基準関節の 1 つ手前の関節までを結ぶ線上に擬似カメラを置く。ただし、基準関節の 1 つ手前の関節まで擬似カメラを前に出してしまうと、それより手前の部分が写らなくなってしまうので、身体の 1 番手前 (図の緑点線部分) までしか擬似カメラが前に行かないように制限した。(元のカメラから擬似カメラまでの距離) : (擬似カメラから身体の 1 番手前までの距離) が $\alpha : (1 - \alpha)$ となるようにする。そのため、 α は $0 \leq \alpha < 1.0$ の範囲を動く数であり、 α が大きくなるほど誇張の度合いも大きくなる (図 5)。 α が大きくなると、それだけ画素が粗くなるため、今回は α の範囲を $0.0 \leq \alpha \leq 0.95$ とした。

今回は、パンチやキックの速さによって誇張係数が変化するように設定した。その方法は次の通りである。右手、左手、右足、左足の関節の 3 次元座標のうち、z 座標のみを保持し、各フレーム間での動きの量を Δz とする。誇張係数 α は、 Δz の定数 a 倍で求める (図 6)。 a が大きいほどグラフの傾きが大きくなり、大きく誇張されやすいことを表す。今回は、 $a = 3.0$ とした。求めた α を、前のフレームで計算した α と比較し、今のフレームの方が大きければ更新、前のフレームの方が大きければ前のフレームの α を用いる。そして、 $\alpha > 0.95$ のときはすべて 0.95 として処理した。また、手や足が腰より後ろにくると、誇張係数を 0 にリセットする。こうして、パンチやキックの勢いが強ければ強いほど大きく誇張されるようにした。

(4) エフェクトをつける

誇張を分かりやすくするために、上記で求めた誇張係数 α を元に、 $\alpha \geq 0.5$ のとき「ドン!!」というオノマトペを、 $\alpha \geq 0.8$ になるとオノマトペに加えて集中線が出るようにした (図 6, 図 7)。集中線は、誇張する箇所を中心にして広がるようにし、線の長さや太さ、間隔をランダムにした。

(5) 擬似カメラの設置、パラメータ設定

元のカメラの座標を原点、手や足の座標を (x, y, z) とすると、擬似カメラの 3 次元座標は $(\alpha x, \alpha y, \alpha z)$ で表される。擬似カメラの位置を決めた後は、擬似カメラの画角を、基準関節を含む xy 平面において元のカメラと同じ範囲を写すように設定する。図の赤い線、青い線がそれぞれ Kinect v2 のカラーカメラと擬似カメラの写す範囲を表しており、横からの図も同様である。こうすることで、元のカラ画像と擬似カメラからの画像をそのまま重ねるだけで、誇張する部分としない部分との境界である基準関節部分がつながった画像を実写でも得ることができる。

(6) ポイントクラウドによる誇張部分の描画

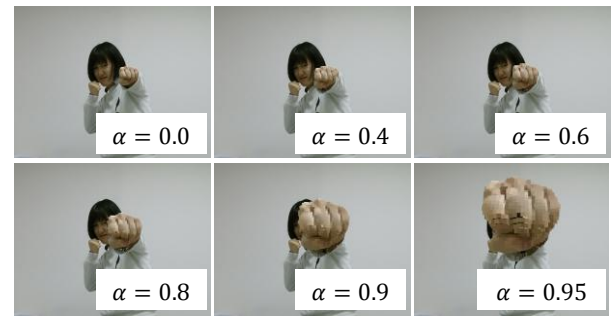


図 5 誇張係数 α の変化による誇張の変化

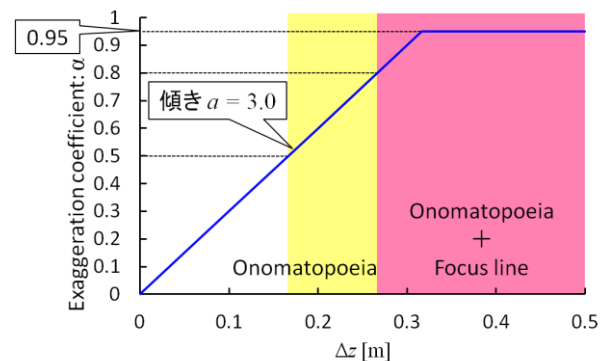
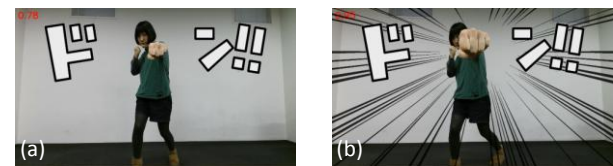


図 6 フレーム間の速度と誇張係数の関係



(a) オノマトペ (b) オノマトペ+集中線

図 7 誇張係数によるエフェクトの変化

(2)で求めた誇張部分を 3 次元上に点群として配置する。この際、Kinect v2 から取得した 3 次元座標はデプスセンサーを原点とした座標なので、(0)で事前求めた変換行列をかけることによって、カラーカメラからの 3 次元座標に変換する。ここで、点群の各点の大きさについて考える。誇張係数が大きくなると、より大きく拡大されるということなので、誇張係数が大きいほど各点の大きさも大きくする必要がある。ここで、簡単のために、腕がカメラの真正面にあるときを考え、Kinect v2 から肘までの z 方向の距離を z_0 [m]、肘から拳の先までの z 方向の距離を l [m] とする。このとき、ある $z (< z_0)$ にある物体に対して、擬似カメラが元のカメラより撮影範囲が狭くなっているため、元のカメラの撮影範囲に占める割合より、擬似カメラの撮影範囲に占める割合の方が高くなり、それだけ画素を拡大する必要がある。最も拡大が必要な拳の先について考えると、元のカメラが写す範囲 (図中の緑線の長さ) は $2(z_0 - l) \tan 42^\circ$ であり、擬似カメラが写す範囲は $2(1 - \alpha)(z_0 - l)z_0 \tan 42^\circ / \{(1 - \alpha)z_0 + \alpha l\}$ である。ここで 42° は、

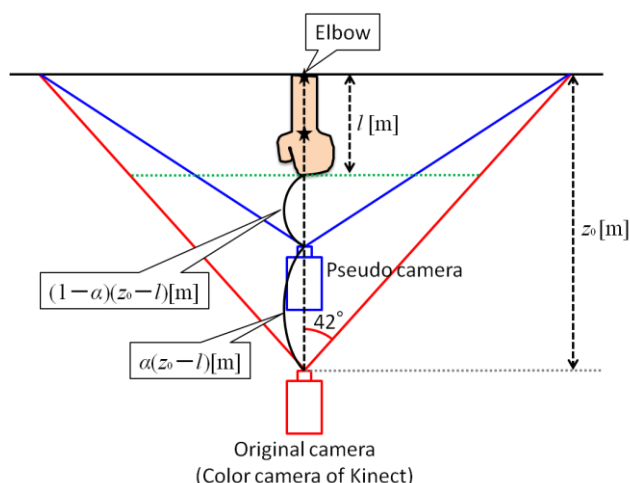


図 8 拡大率を計算するための設定

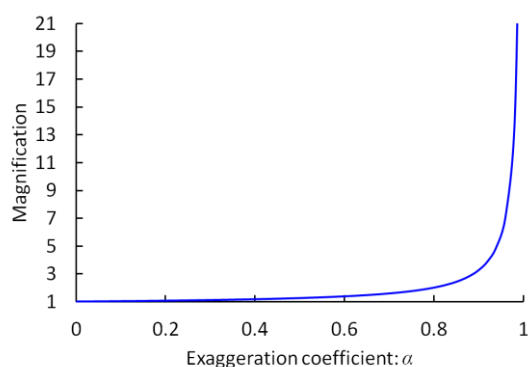


図 9 誇張係数 α による拡大率の変化

Kinect v2 の画角の半分である．拡大率は長さの比の逆比で表すことができるので，以下で計算することができる．

$$\frac{2(z_0 - l) \tan 42^\circ}{2(1 - \alpha)(z_0 - l)z_0 \tan 42^\circ} = 1 + \frac{\alpha l}{(1 - \alpha)z_0}$$

$z_0 = 1.36$, $l = 0.34$ としたとき (図 5) のグラフを図 9 に示す．画素の粗さが気になるときは， α のとり得る値に制限をつければよい．

(7) 合成

カラー画像とポイントクラウドとエフェクトを重ねることにより，誇張された映像を完成させる．

4. 展示

2014 年 11 月 13 日から 17 日に行われた第 16 回東京大学制作展[7]において，AnimE-Lise を展示し，一般の方に体験してもらった (図 10)．そのうち，11 月 13 日，14 日，16 日，17 日の 4 日間をビデオ記録した．ビデオから，4 日間で合計 304 人の人が体験したことが分かった．

4.1 設置方法

制作展では，Kinect v2 で取得したものを PC で処理し，誇張した後の映像をリアルタイムに短焦点プロジェクタで

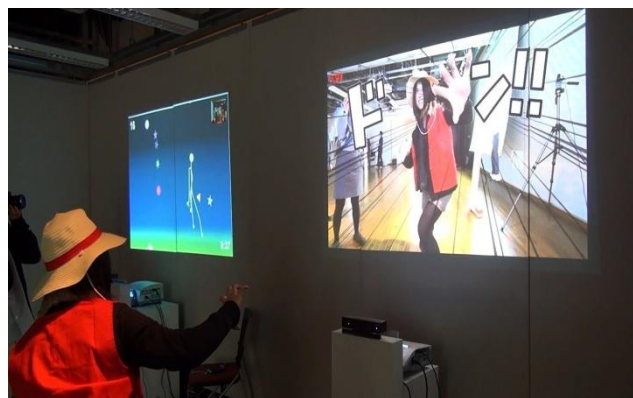


図 10 制作展展示の様子

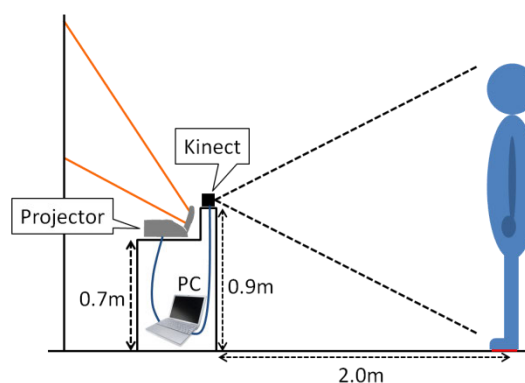


図 11 展示の空間設計

壁に投影した．各機材の位置関係は図 11 のようである．ユーザーの適切な立ち位置，すなわち全身が Kinect v2 の画角内に収まるような位置にテープを貼り，かつスポットライトを当てた．

4.2 展示結果

展示をした結果，体験した人から「すごい」や「迫力ある」といった驚きのコメントを多くもらった．また，体験者に特徴的な行動が見られたので，以下に報告する．

• 写真を撮る

304 人の体験者のうち，67 人，つまり 4.5 人に 1 人が自分が誇張された映像を自分で写真に撮る，あるいは一緒に来た人に撮られていた．誇張されたものを写真に撮りたくなるのは，本システムが体験者自身が写るコンテンツであり，体験者自身が変形していることに対して驚きや面白さを感じ，記録に残しておきたいと思ってもらえたからなのではないかと考えている．また，写真に撮るということは，後で見返したり，他の人に見せたりする可能性が高いとも考えられる．

• 手に持っているものを誇張させる

Kinect v2 は，身体に接しているものも人物認識することが多く，例えば手に缶を持ち，その手を前に出すと，手に持った缶まで誇張されるといった場面が見受けられた．

以上 2 つの行動は，デジタルサイネージに用いることで商業的にも役に立つのではないかと考えられる．

- 体験に適切な位置より前に行ってしまう

4.1 で述べたように、体験する適切な位置に印をつけておいた。しかし、体験者の中には適切な位置より前に行ってしまう人がいた。これは特に子どもに見られた行動であり、パンチやキックという攻撃的な行動をしていたために、前に出て行ってしまったのではないかと考えられる。しかし、前に行きすぎると Kinect v2 の画角外になってしまい、誇張箇所が写らなくなるという問題がある。実際に写らなくなってしまう場面もあったのだが、体験者は後ろに下がろうとはしていなかった (図 12)。原因として考えられることは、Kinect v2 が体験者の顔を下から写しているのに対し、投影された映像を見るときは自分の目の高さより上を見る必要があり、投影された映像中の自分と映像を見ている自分との間に視線のずれが生じていることである。投影された映像、つまり自分の視線より上を見ながらパンチをするとどうしてもパンチが上に行ってしまうのである。そのため、体験者が見る映像と、Kinect v2 が写す映像の高さに生じるずれを小さくする対策をする必要がある。例えば、今回は Kinect v2 を台に置いていたが、投影している映像の上に取り付けければ、Kinect v2 と投影された映像のどちらも目の高さより上になるので、このような現象が緩和されるのではないかと考えられる。

- パンチやキック以外の行動をする

今回、パンチやキックを想定して右腕、左腕、右脚、左脚のうち同時に 1 箇所を誇張できるようにした。しかし、体験者の中には、両手を同時に前に出したり、顔を前に出してみたりと、様々なポーズで誇張を試みる人がいた。体験者が、狙った箇所以外のところも誇張したくなるという行動によって、誇張することに興味を持っていたことが分かる。しかし今回は、両手を同時に前に出しても片方の腕しか誇張されず、顔の誇張にも対応できていなかった。両腕を同時に誇張するためには、擬似カメラを 2 つに増やす必要がある。顔を誇張することは 1 つの擬似カメラで可能なのだが、誇張すると画質が粗くなる。腕や脚は同じような色で構成されるためあまり目立たなかったが、顔は、目や鼻、口など様々なパーツで構成され、複雑であるため、画質が粗くなると目立ってしまうと考えられる。

5. おわりに

本研究では、漫画やアニメで使われる人物の遠近感の誇張表現を実写に適用し、リアルタイムに人物を誇張できるシステム AnimE-Lise を制作した。これを第 16 回東京大学制作展で展示し、一般の方に体験してもらった。そこで、写真を撮る人が多いことや、身体以外の部分も誇張させようとする人がいたことから、デジタルサイネージに使うことへの可能性が示された。一方、適切な体験位置より前に出てしまい、誇張箇所がカメラの画角から外れてしまうことや、意図していた行動 (パンチやキック) 以外の行動を



図 12 適切な体験位置より前に出てしまった例

する体験者もいたため、システムの改善点が明らかになった。

参考文献

- 1) 宇都木契, 苗村健: E-IMPACT ～日本のアニメーションにおける誇張表現を模した 3DCG～, 映像情報メディア学会誌 Vol. 66, No. 2, pp. 102-105 (2012).
- 2) Utsugi, K., Naemura, T., Koike, T., Oikawa, M.: E-IMPACT: Exaggerated Illustrations using Multi-perspective Animation Control Tree Structure, Proc. ACE 2011, 16-4 (2011).
- 3) 奈良優斗, 小出雄空明, 藤村航, 白井暁彦: 瞬刊少年マルマル, 映像情報メディア学会技術報告, Vol. 37, No. 17, pp. 125-128 (2013)
- 4) 漫画カメラ, SUPERSOFTWARE
<http://tokyo.supersoftware.co.jp/mangacamera/>
- 5) Umeda, D., Moriya, T. and Takahashi, T.: Real-time Manga-Like Depiction Based on Interpretation of Bodily Movements by Using Kinect, Proceedings of ACM SIGGRAPH Asia 2012 Technical Briefs, 28, 1-4 (2012).
- 6) Kinect for Windows v2, Microsoft
<http://www.microsoft.com/en-us/kinectforwindows/>
- 7) 東京大学制作展
<http://www.iiiexhibition.com/>