DEIM2022 E33-4(day3 p25)

貼り付け元画像の位置関係とサイズ関係を用いたオブジェクト画像合成

相場　築† 服部　峻††

†,†† 室蘭工業大学 ウェブ知能時空間研究室　〒 050–8585 北海道室蘭市水元町 27–1
E-mail: †21043001@mmm.muroran-it.ac.jp, ††hattori@csse.muroran-it.ac.jp

あらまし　近年，画像分類や画像処理，敵対的生成ネットワークである GAN [1] による画像生成など，画像に関する研究が盛んに行われている．特に，GAN による画像生成の研究分野は急成長を遂げており，テキスト入力によって生成物の指定が出来る text-to-image GAN [2] や，ラベルによって生成物を制御する cGAN [3] など，派生する研究は多岐に亘る．しかし，機械学習の不安定さや学習データを大量に必要とする条件から，実際に他のコンテンツとして利用出来るような生成物を得るのは，単純な構造の生成物でなければ現状難しい段階にあり，複雑な構造の生成物の画像を自動的に合成するには，複数の問題が挙げられる．そこで，本論文では，貼り付け元の画像と単純な構造の生成画像とを合成する自動画像合成器を構築し，より自然に見えるオブジェクト画像合成手法について提案する．画像生成（合成）に関する問題点の中から特に「位置関係」と「サイズ関係」に焦点を当て，これらの関係に基づいて元の画像の空間構造を推定し，合成画像のサイズを自動的に調節することで，自然に見えるオブジェクト画像の合成手法について提案する．

キーワード　画像合成, 空間構造推定, 敵対的生成ネットワーク (GAN), 画像生成, 画像融合

1　ま え が き

近年，画像分類や画像生成など画像についての研究が盛んに行われている [4]．特に，画像生成に特化した GAN は「鳥だけ」や「椅子だけ」といった単純な構造の画像であれば，学習するデータセットに依存するが，生成対象物が人によって「鳥」や「椅子」などと認識出来る程度の画像が生成出来る場合が多い．しかし，生成物の条件が指定出来るような派生形の GAN を利用し,「雪山でスキーを滑る人」や「会議室で会議をする人々」など，複雑な構造の生成物になると，何が写っているのかが判断出来ない生成画像がしばしば表示される．そこで，本稿では画像生成だけでなく画像「合成」技術も活用し,「人だけ」「廊下だけ」「ポスターだけ」といった単純な構造の生成画像を上手く合成することで,「廊下でポスターを見る人」のような複雑な構造の生成画像を得られる手法の開発が目的である [1]．そのためには，合成した画像が自然に見えるように様々な問題点を克服しなければならず，例として「位置」「サイズ」「光と影」「色合い」等，思いつく限りでも多数挙げられる．また GAN は入力に対し出力をすぐ得られる点から，合成方法についても自動化し，生成画像を得るために求められるユーザの労力を出来る限り削減するような仕組みが求められる．

本稿では,「位置」「サイズ」に対する問題点に対して，2 つの関係に基づいて貼り付け元の画像における立体空間構造を推定し，貼り付け位置に対するオブジェクト画像のサイズを調節して貼り付けるというアプローチによって，人間が自然に見えるような複雑な構造の画像を生成することを目的とした手法について提案する．

具体的には，まず貼り付け元の画像に含まれているオブジェクトを認識し [2]，次にその複数のオブジェクトの座標および認識されたオブジェクトのバウンディングボックスのサイズから,「この種類のオブジェクトはこの位置ならばこのサイズが適切である」といったサイズを調節するための回帰直線を求める．最後に，この回帰直線を用いて貼り付ける素材のオブジェクト画像のサイズを決定する [3]．

例として，単純な構造の生成物として「廊下」「ポスターを見る人」を用意し，2 つを画像合成する場合，どの位置に貼れば見栄えが良くなり，その位置にどんなサイズであれば人間が見た時に違和感を持たないかなどが課題として挙げられる．さらに,「ポスターを見る人」に限らず,「走る人」であったり,「廊下」でない場所でも「光と影」や「位置」と「サイズ」などを自動的に調節し，なるべく違和感を持たれないような画像合成を行うことが出来るようにするといったことが必要である．その中で，本稿では，最低限,「位置関係」と「サイズ関係」について，画像合成された画像が違和感が無く自然なものになるように，貼り付け元の画像中の手動で指定した任意の位置に対して，適切なサイズになるようにオブジェクト画像を調節して貼り付ける手法について提案を行う．

2　関 連 研 究

画像の空間構造を推定する手法はいくつか研究されているが，その推定のためには，同じ空間について多角的に撮影された複

1：4 章の評価実験では，貼り付け元の画像には，後述するオブジェクトが認識出来ている条件を満たすために「廊下に椅子が複数ある」画像を用い，単純な画像として人だけの画像を利用し，背景を透過したものを使用する

2：オブジェクト認識の条件は 3 章で具体的に記述する

3：サイズの決定方法の詳細は 3.1.3 項で記述する

数枚の写真または，空間構造が推定出来るだけの，人間が作成した図などが必要であり，簡単に利用することは難しい．石川ら [5] は同じ建物を撮影した多角的な写真を用いて，被験者に作図させることで空間把握を行うプロセスについて提案している．また，水野ら [6] はオブジェクトのスケッチ画像を「正面断面図」及び「サイドビュー」の 2 枚を用いて，3 次元データを生成するシステムについて開発している．いずれの研究も用意すべき画像が 2 枚以上であったり，空間把握（3 次元オブジェクトの生成）のために特殊な図を用いるため，未だ空間構造を推定する手法については不十分であると考える．

一方で，複雑な構造の生成物の画像を合成することを目的としているが，1 章でも述べたように，現状の技術レベルで「雪山でスキーを滑る人」などを指定すると，人間の目で見て何が写っているのか判別出来ない画像が生成されてしまう．Qiao らが発表した text-to-image GAN である「MirrorGAN [7]」では，上記のような複雑な指定を行うと上手く生成出来ないことが判明している．しかし,「鳥」だけ「人」だけのような単純な画像は生成出来るため，これらの生成画像を用いることで複雑な画像を合成する目的で，手法の提案を行う．

3 提 案 手 法

本章で提案する「位置関係」と「サイズ関係」に着目した画像合成手法には，貼り付け元と貼り付ける素材の 2 つの画像が必要になる．貼り付け元の画像には何らかのオブジェクトが複数存在するという前提条件を課し，一方，貼り付ける素材の画像の前提条件は後述する提案手法の詳細に依って変化する．以上の前提条件を満たした 2 つの画像の合成の自動化において，サイズの観点で合成の違和感を減らす手法について提案する．まず，提案手法で必要な 2 つの画像の前提条件について詳述する．

4 章の評価実験では，人間の目から見て消失点（奥行）の場所が判断出来る貼り付け元の画像を用意し，貼り付ける素材は同じ場所で撮影した人を切り抜いたものを利用する．

前提として，実験の貼り付け元の画像からは回帰直線を求めるために使う全てのオブジェクト認識が出来ているとし，本稿の評価実験であれば画像中に含まれる「椅子」という種類（クラス）のオブジェクト（インスタンス）の全てをオブジェクト認識出来ているものとする．それらのオブジェクトは人間が一般的にそのカテゴリであると認識出来るもの（椅子というカテゴリ）で且つ，3 つ以上含まれていることが必要である．この場合，回帰直線を求めるのに利用するオブジェクト以外が認識されない場合はラベルは必要無い．(3.2 節の手法には必要)

もし，オブジェクト認識がより高度なもの（壁から離れている孤立した物体だけでなく，壁にかかっているようなポスターも認識されるようなもの）であれば，認識されるオブジェクトが複数種類出てくるので，その場合は回帰直線のために用いるオブジェクトを区別するためのラベルも必要である．さらに，オブジェクト認識で求められる認識範囲はオブジェクトを含んだ長方形のバウンディングボックスが必要である．また，貼り付ける素材の画像については切り抜いて，背景を透過したものを利用する．貼り付ける素材を自動的に取得するのであれば，オブジェクト認識がオブジェクトの輪郭を認識出来ている必要があり，背景とオブジェクトを区別出来なければならないが，本稿では考慮しない [4]．

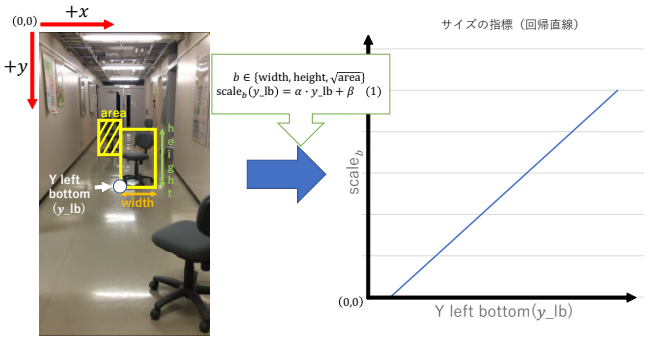

図 1　提案手法の概要（立体空間構造の推定）

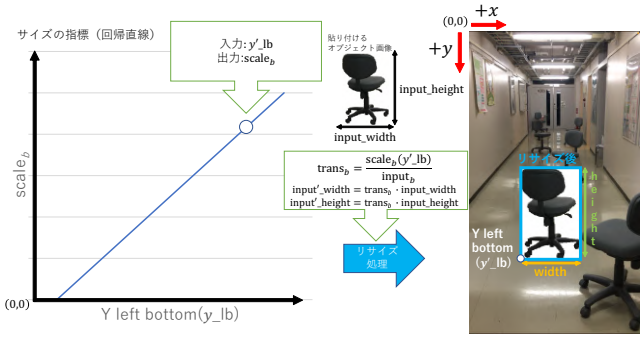

図 2　提案手法の概要（回帰直線に基づくリサイズ）

上述した前提では，同じ種類のオブジェクトで複数存在する画像を想定しているが，実際の画像では異なる種類のオブジェクトが複数存在しているので，全てのオブジェクトの位置とサイズを加味して，追加するオブジェクトのサイズを調節出来るようになるのが今後の研究課題である．3.1 節では，貼り付けるオブジェクト画像のクラスと同一のオブジェクトを貼る際の手法および概要について述べ，その内の 3.1.1 項から 3.1.3 項では，貼り付け元の画像から切り抜いたオブジェクト（背景透過を施したもの，以降「オブジェクト画像」と呼称）を用いて，再び貼り付け元の画像に貼り付ける際に,「位置」と「サイズ」について調節する手法について述べる．また，3.2 節では，回帰直線の算出に用いたオブジェクト群のクラスと異なるオブジェクト画像を用いて，貼り付け元の画像に「位置」と「サイズ」に関して調節して貼り付ける手法について詳述する．

3.1 概　　要

提案手法の概観については図 1 及び図 2 の通りである．

本節では，提案する手法の基本としている処理について述べる．基本の処理手順は大まかに分けて 3 つの処理に分けられる．次項から，これら 3 つの処理について順に説明していく．

4：合成についての提案なので，貼り付ける素材はユーザが用意するものとした

図 3　重回帰式による合成結果 1

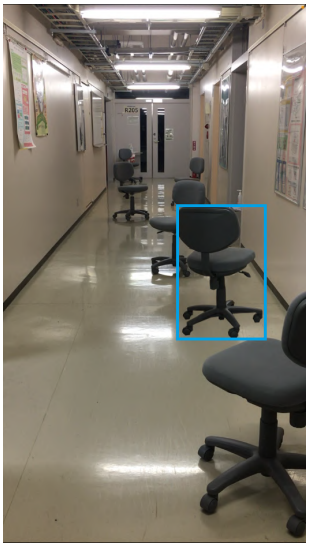

図 4　重回帰式による合成結果 2

Step 1. 貼り付け元のオブジェクト認識
Step 2. サイズの指標となる回帰直線の算出
Step 3. 回帰式に基づく貼り付けオブジェクトのサイズ調節

3.1.1　Step 1：貼り付け元のオブジェクト認識

貼り付け元の画像に含まれているオブジェクトを認識出来るツール（YOLOv3 [8] 等）を用いてオブジェクトのバウンディングボックスを取得する．この認識結果を用いて 3.1.2 項で回帰直線を求める．本稿の実験では，貼り付け元の回帰直線を求めるのに利用するオブジェクトは認識出来ているものと仮定して実験を行う．

3.1.2　Step 2：サイズの指標となる回帰直線の算出

3.1.1 項で求めた貼り付け元のオブジェクト群の認識結果からサイズの指標となる回帰直線を算出する．具体的に，バウンディングボックスの左下の y 座標（y_lb（y left bottom）と呼称する）を説明変数として代入し，バウンディングボックスの「width」,「height」,「$\sqrt{\mathrm{area}}$」（横幅，高さ，$\sqrt{\text{面積}}$）のいずれか 1 つを目的変数として回帰直線の計算を行う（4 章の評価実験では，Scikit-learn [9] を用いる）．

ここで，y 座標のみを利用する理由としては，(x, y) の両方を利用した場合，同じ y 座標に合成する際にサイズが変化し，意図しない挙動 [5] をすることが事前実験でわかっているためである．図 3 及び図 4 は (x, y) の両方を用いて重回帰分析し，重回帰式を用いて合成した結果（水色で囲われた椅子が合成されたもの）である．図 3 は回帰式算出の際に利用した座標であるため，人間の目では適したサイズに見えるが，この位置から右に座標をずらして合成した時，現実ではサイズは変化しないはずが，図 4 では微小に大きく変化していることがわかる．以上の理由から，x 座標を用いず，y_lb のみを用いる．

さらに，バウンディングボックスの左下座標 y_lb を用いる理由としては，実験中に上側の座標を基準に回帰分析を行った際，基準点が画像外の座標を指し示し，合成出来ない状況が発生したためである [6]．

以下の数式（1）は貼り付け元の画像中で認識されたオブジェクト群の「width」,「height」,「$\sqrt{\mathrm{area}}$」のいずれか 1 つを目的変数とした回帰直線を表し，$b \in \{\mathrm{width}, \mathrm{height}, \sqrt{\mathrm{area}}\}$（横幅，高さ，$\sqrt{\text{面積}}$）とする．

$$\mathrm{scale}_b(y\_\mathrm{lb}) = \alpha \cdot y\_\mathrm{lb} + \beta \tag{1}$$

以降，簡潔のために「width」を目的変数として回帰直線を求めるものを手法（1）とし，同様に「height」を目的変数としたものを手法（2），「$\sqrt{\mathrm{area}}$」を目的変数としたものを手法（3）と表す．

3.1.3　Step 3：回帰式に基づく貼り付けオブジェクトのサイズ調節

図 1 のように 3.1.2 項で得られた回帰直線の数式（1）を元に実際に貼り付けるオブジェクト画像のサイズを調節する．具体的には，貼り付け元の画像の任意の座標 (x', y') に貼り付けるオブジェクト画像の横幅と高さを数式（3，4）で算出する．

$$\mathrm{trans}_b = \frac{\mathrm{scale}_b(y'\_\mathrm{lb})}{\mathrm{input}_b} \tag{2}$$

$$\mathrm{input}'\_\mathrm{width} = \mathrm{trans}_b \times \mathrm{input}\_\mathrm{width} \tag{3}$$

$$\mathrm{input}'\_\mathrm{height} = \mathrm{trans}_b \times \mathrm{input}\_\mathrm{height} \tag{4}$$

繰り返し詳述すると，$b \in \{\mathrm{width}, \mathrm{height}, \sqrt{\mathrm{area}}\}$（横幅，高さ，$\sqrt{\text{面積}}$）のいずれかであり，$\mathrm{scale}_b(y'\_\mathrm{lb})$ を用いて，実際に貼り付けるオブジェクト画像のサイズを調節する．input_b は貼り付ける画像の b [7] を，trans_b は b に基づくリサイズ比率を表している．4 章の評価実験では数式（1）について様々に変化させ，その中で最も「サイズ関係」について，自然な画像合成の結果になるものを模索する．

3.2　貼り付けるオブジェクト画像のクラスと異なるオブジェクト群の回帰式に基づく画像合成

本節では，3.1.3 項で記述した手法を基に，貼り付け元に含まれるオブジェクトを用いて算出された「scale」を利用し，貼り付け元と異なるクラスのオブジェクトを貼り付け，「位置」と「サイズ」の観点から出来る限り自然に見えることを目的として手法を提案する．異なるオブジェクトを合成する際，貼り付け元に含まれるオブジェクトによって算出された「scale」をそのままの手順で利用すると，貼り付ける素材のサイズの比率が変わるため，不自然な合成が起きることが考えられる．そこで，異なるオブジェクトで合成する場合に出来る限り自然に見えるような工夫を行う必要があり，その方法について述べていく．提案手法には 2 つの手法が存在する．

（1）　実際の物体データから比率を乗算
（2）　オブジェクト認識のサイズの比率を乗算

それぞれの提案手法に依って，用意しなければならないデータ（実際のオブジェクトの寸法や倍率データ）が異なるため詳し

5：同じ y 座標（横）に配置した際，現実的にはサイズが変化しないことを意図して作成している

6：合成したい位置や貼り付けるオブジェクト画像に依るため，基準点の位置を変えた時に変化があるか確認が必要であることは今後の課題である

7：$\mathrm{input}\_\mathrm{width}$ は b が width，$\mathrm{input}\_\mathrm{height}$ は b が height の時を示す

い条件については，各手法の項で記述する．4 章の評価実験では 2 つの手法を試みて，サイズの観点から自然に見える度合いとデータを揃える必要性の 2 つの点から各手法の比較と考察を行う．

3.2.1 実際の物体データから比率を乗算

この手法では，現実の寸法データ（椅子であれば，高さ 76~86cm，横幅 53~55cm）から「scale」算出の時に用いたオブジェクトと貼り付けるオブジェクトの比率を計算し，サイズを調節する．「人」や「背もたれが伸び縮みする椅子」のような場合，可変長なので，その長さが変化する範囲から一様分布乱数を用いて比率として使う寸法を決めるため結果が実行毎に変化する．

具体的に，数式（5）の $\mathrm{inputsize}_b$ は 4 章の評価実験では人の寸法（身長 150~182cm[8]，横幅 20~34cm と設定）から 1cm 刻みでランダムな値，bgsize_b は椅子の寸法（高さ 76~86cm（5cm 刻み），横幅 53~55cm（1cm 刻み）と設定）からランダムな値のどれかが入力される．サイズの調節については，実データの寸法から比率を用いるため，結果が実行毎に変化するもののある程度自然に見える画像が合成されることが期待出来る．揃えるデータの必要性については，「scale」を算出する際に用いたオブジェクトと同じクラスの寸法と，貼り付けるオブジェクト画像のクラスの 2 つの実データの寸法が必要である．また，「scale」を算出する際に用いたクラスと異なるクラスの貼り付けるオブジェクト画像（「異クラスオブジェクト画像」と呼称）が必要である．数式（7，8）によって，異クラスオブジェクト画像の「横幅」と「高さ」を算出する．

$$\mathrm{scale}'_b(y\_\mathrm{lb}) = \mathrm{scale}_b(y\_\mathrm{lb}) \times \frac{\mathrm{inputsize}_b}{\mathrm{bgsize}_b} \tag{5}$$

$$\mathrm{trans}_b = \frac{\mathrm{scale}'_b(y'\_\mathrm{lb})}{\mathrm{input2}_b} \tag{6}$$

$$\mathrm{input2}'\_\mathrm{width} = \mathrm{trans}_b \times \mathrm{input2\_width} \tag{7}$$

$$\mathrm{input2}'\_\mathrm{height} = \mathrm{trans}_b \times \mathrm{input2\_height} \tag{8}$$

但し，$b \in \{\mathrm{width}, \mathrm{height}, \sqrt{\mathrm{area}}\}$（横幅，高さ，$\sqrt{\text{面積}}$）のいずれかであり，「input2_width」及び「input2_height」は貼り付ける異クラスオブジェクト画像の「横幅」と「高さ」を示している．また，「$\mathrm{input2}'$_width」及び「$\mathrm{input2}'$_height」は貼り付ける任意の座標 (x', y') に異クラスオブジェクト画像を位置に依ってサイズを調節した後の「横幅」と「高さ」を示しており，このサイズを用いて，実際に，貼り付け元の画像へ異クラスオブジェクト画像を貼り付ける．

3.2.2 オブジェクト認識のサイズの比率を乗算

この手法では，「scale」を算出した際に用いたクラス（実験では「椅子」）と貼り付けたいクラス（実験では「人」）が共に写っている画像の集合を用いてオブジェクト認識を行い，認識された 2 つのクラスのバウンディングボックスの比率（ϵ と呼称）を，画像の枚数分算出し平均したものを乗算することで「scale」を異クラス用に変換し，サイズを調整する手法である．

8：4 章の評価実験では，正解のデータが 171cm の人物の画像しか用意出来なかったため，身長は 171cm で固定されている

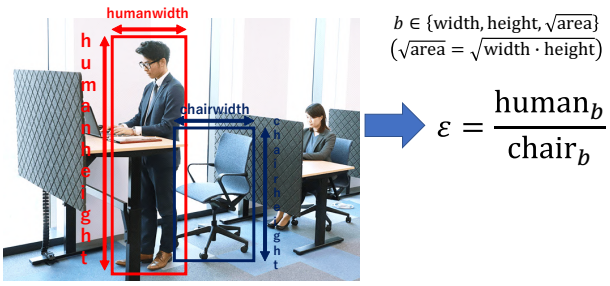

図 5　比率（ϵ）算出方法（画像引用 [10]）

比率 ϵ の算出方法を図 5 に示す．

異クラスオブジェクト画像の「横幅」と「高さ」を算出する流れは数式（7，8）を算出するものと変わらないが，異クラス用の回帰式「scale′」の算出の際に「scale」に乗算する比率が 3.2.1 項と異なる．

$$\mathrm{scale}'_b(y\_\mathrm{lb}) = \mathrm{scale}_b(y\_\mathrm{lb}) \times \epsilon \tag{9}$$

それぞれの変数が表すものは 3.2.1 項と変わらず，異なっているのは数式（5）の代わりに数式（9）が利用される部分である．サイズの調節については，集められる画像の量や，質 [9] に依って左右されるものの，適切に処理が出来れば十分に適切なサイズの調節が可能であると考えられる．揃える画像の必要性については，2 クラスが同時に撮影されている画像を用意しなければならず，より精度の高い比率を求めるためには，集めた画像にノイズが含まれている事を念頭に置くと 10 枚以上用意出来る事が望ましい．本稿の実験では，人力で適する画像を集めたものの，自動化のためには収集作業も自動でなければならず，それにより多くのノイズを含む可能性が高いため，質の高い画像を汲み取る方法についても今後の課題である．つまり，一般的には思いつかないような 2 クラスの組み合わせであると，思惑通りに動作しない可能性がある．これらのことから 3.2.1 項よりも揃えるデータの条件が厳しくなっていると考えられる．

4　評価実験

本章では，3.1 節で提案したサイズの指標となる回帰直線「scale」を算出する際の目的変数である「width」「height」「$\sqrt{\mathrm{area}}$」（横幅，高さ，$\sqrt{\text{面積}}$）に依って，数式（3，4）の貼り付けるオブジェクトの横幅と高さにどのような影響を及ぼすのかを検証する．具体的に，まず図 6（椅子が 4 つ [10]）の画像を人間がオブジェクト認識を行い，その画像に写っているオブジェクトのクラスとバウンディングボックスを設定する [11]．次に，設定したバウンディングボックスの情報から目的変数となる「width」「height」「$\sqrt{\mathrm{area}}$」を用いて，サイズの指標である

9：2 クラスがはっきりと写っており不自然なサイズになっていないもの

10：最も手前にある椅子は見切れており，正確な「width」及び「$\sqrt{\mathrm{area}}$」の値を取得出来ないため無視する

11：オブジェクト認識が 3 章で説明したような前提条件を満たす必要があるため

図 6　椅 子 4 個

図 7　椅 子 8 個

「scale」を算出する．その指標を用いて 3.1.3 項で解説した手順で，貼り付けるオブジェクトの横幅と高さを算出する．次に，椅子が 4 つの画像（図 6[10]）に椅子を 1 つ合成した場合を評価するために，同じ場所で撮影された同じ画角の画像を用意し，図 7（椅子が 8 つの画像 [10]）を同様に人間がオブジェクト認識して，バウンディングボックス情報を設定する．4.1 節では，図 7 を正解の画像として用意し，図 6 に写っている椅子 4 つ [10] のバウンディングボックス情報を用いて，その画像に存在していない位置に椅子を配置した時の，実際の位置に対するサイズを比較した結果を載せる．同様に，4.2 節では，3.2 節で解説した貼り付けるオブジェクト画像のクラスと異なるオブジェクト群を貼り付ける実験として，図 6 の椅子 4 つ [10] のバウンディングボックス情報と，人の画像を用いて，3.2.1 項や 3.2.2 項で解説したサイズの指標である「scale」に，人用の比率を乗算する．つまり，椅子の「位置」に対する「サイズ」の情報から人を貼り付けた時の結果を載せる．人の画像の正解データとして，例を 3 つ載せる（図 8 から図 10）．

加えて，4.3 節では，「位置関係」に対する適切なサイズを調節するタスクを被験者に行わせ，各位置 [12] にサイズを調節した時の被験者全体の横幅及び高さを用いて，提案手法に依る横幅と高さの調節結果と比較した時の評価値を載せる．

4.1　貼り付けるオブジェクト画像のクラスと同一のオブジェクト群の回帰式に基づく画像合成の結果

本節では，貼り付け元の画像に含まれる椅子（図 6 の椅子 4 つ [10] のバウンディングボックス情報）を用いてサイズの指標となる回帰式を計算し，これに基づいて，同じクラスである椅子のオブジェクト画像を合成した時の結果例と評価値を載せる．具体的に，3.1.3 項で解説した，サイズの指標を算出する際に用いる目的変数が異なる 3 手法に依って得られた結果から以下に記述する平均絶対誤差（横幅と高さの差（数式（10））），及び比率（数式（11））を計算し載せる．

12：椅子であれば 4ヶ所，人であれば 12ヶ所

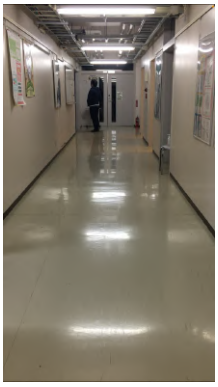

図 8　人の正解データ 1

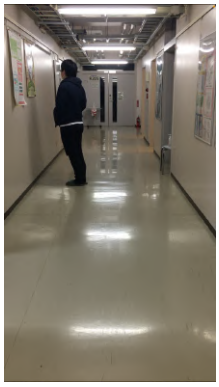

図 9　人の正解データ 2

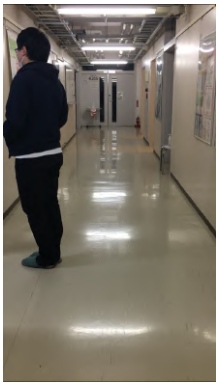

図 10　人の正解データ 3

表 1　椅子を合成した時の評価値

	MAE_width	MAE_height	MAE_rat_width	MAE_rat_height
手法（1）	**3.25000**	3.50000	0.96917	0.95949
手法（2）	6.50000	7.25000	1.06287	1.04768
手法（3）	3.50000	**3.25000**	**1.01012**	**1.00255**

表 2　被験者 11 名の感覚と椅子を合成した時の比較による評価値

	MAE_width	MAE_height	MAE_rat_width	MAE_rat_height
手法（1）	6.18939	9.37121	0.91618	0.91714
手法（2）	**2.00000**	**3.10606**	**1.00470**	**1.00138**
手法（3）	2.93939	4.12121	0.95455	0.95823

ここで，数式（10）の $x_b(i)$ は b に依ってリサイズされた横幅または高さを示しており，$b \in \{\text{width}, \text{height}\}$ である．つまり，MAE は横幅と高さの 2 つ存在する．

$$\text{MAE}_b = \frac{1}{n} \cdot \sum_{i=1}^{n} |x_b(i) - c_b(i)| \tag{10}$$

$$\text{MAE\_rat}_b = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{x_b(i)}{c_b(i)} \tag{11}$$

また，$c_b(i)$ は，筆者が図 7 からバウンディングボックスを設定して取り出した i 番目の椅子の横幅または高さを示す．数式（11）についても，b は同様に 2 種類存在し，リサイズ後の各値（$x_b(i)$）と，正解とした値（$c_b(i)$）との比率を示している．以上の数式から，図 7 には椅子が存在しており，図 6 には椅子が存在していない位置 i（合計 4ヶ所，$n = 4$）に貼り付けた時の評価値として算出する．図 11 から図 16 は「width」,「height」,「$\sqrt{\text{area}}$」（手法（1），手法（2），手法（3））を目的変数とした回帰直線を用いて実際に画像を合成した結果の例である（水色で囲われた椅子が合成されたもの）．

また，表 1 は貼り付けるオブジェクト画像のクラスと同一のオブジェクト群の回帰式に基づいて画像を合成した時の数式（10）と数式（11）の評価値である．

4.2　貼り付けるオブジェクト画像のクラスと異なるオブジェクト群の回帰式に基づく画像合成の結果

本節では，貼り付け元の画像に含まれる椅子（図 6 の椅子 4 つ [10] のバウンディングボックス情報）を用いてサイズの指標となる回帰式を計算し，これに基づいて，オブジェクト画像を合成する．4.1 節の結果との違いは，回帰式に基づいて合成する

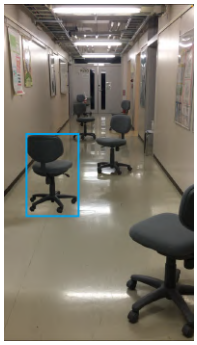

図 11　手法（1）　手前

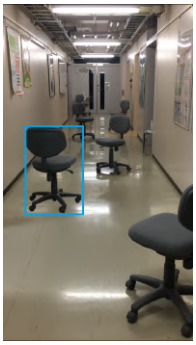

図 12　手法（2）　手前

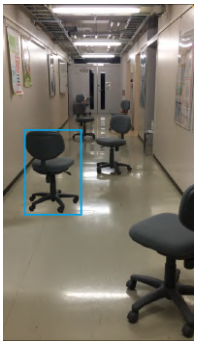

図 13　手法（3）　手前

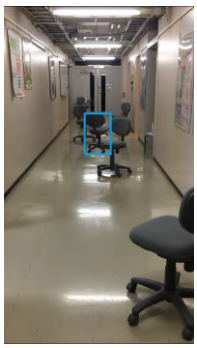

図 14　手法（1）　奥

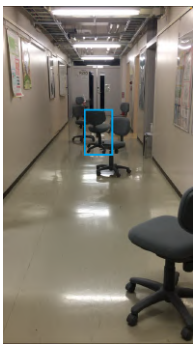

図 15　手法（2）　奥

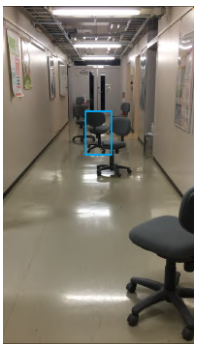

図 16　手法（3）　奥

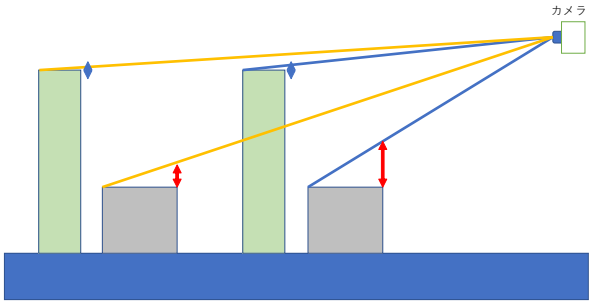

図 17　カメラの目線の高さとオブジェクトの高さに依る誤差の変化

オブジェクト画像のクラスが回帰式の算出に用いたオブジェクトのクラスと異なる点である．本稿の実験では，上述の通り回帰式算出には椅子を使用し，合成する異クラスオブジェクト画像として人間を合成する．評価値の結果について，サイズの指標を算出する際に用いる目的変数が異なる 3 手法と，3.2 節で解説した，異クラスオブジェクトを合成する際に行う 2 手法の組み合わせ（$3 \times 2 = 6$ 手法）に依って得られた結果から同様に平均絶対誤差（横幅と高さの差（数式（10）），及び比率（数式（11））を計算し載せる．尚，人間の横幅と高さを評価するために正解の画像として 12 枚（図 8〜 図 10 は正解の一部），同人物で位置が違うものを用意し，筆者がバウンディングボックスを設定した（合計 12ヶ所，$n = 12$）．図 18 から図 23 は全ての組み合わせの手法に依って得られた結果の一例である．

また，表 3 は貼り付けるオブジェクト画像のクラスと異なるオブジェクト群の回帰式に基づいて画像を合成した時の数式（10）と数式（11）の評価値である．

4.3　人間の感覚に適したサイズ調整との比較結果

本節では,「位置関係」に対する適切なサイズを調節するアンケートについて椅子 4ヶ所分の位置，人 12ヶ所分の位置を被験者に行わせ全ての調節結果を各位置で平均した時と，提案手法に依って得られた横幅及び高さの平均絶対誤差（横幅と高さの差（数式（10）），及び比率（数式（11））を算出し結果を載せる．表 2 は椅子 4ヶ所の位置を被験者 11 名で調節した時の評価値であり，表 4 は人 12ヶ所の位置を被験者 11 名で調節した時の評価値である．また，被験者がサイズの調節を行った際にかかった所要時間は，椅子の場合には平均で約 15.7 秒，人の場合には平均で約 13.6 秒であり，一方，提案手法でサイズ調節を行った時の所要時間は，いずれも平均で約 0.25 秒であった．提案手法で調節する方が紛れもなく早いことがわかる．

4.4　考　　察

表 1 の椅子を各手法に依って配置した時の評価値を見ると，総合的に見て最も誤差が少ないのは手法（3）の $\sqrt{\text{面積}}$ を基準とした時である．対して，最も誤差が多いのは手法（2）の高さを基準とした時である．次に，表 3 の人を各手法に依って配置した時の評価値を見ると，総合的に見て誤差が最も少ないのは手法（2）であり，対して，誤差が最も大きいのは目に見えて手法（1）の横幅を基準とした時である．これらの情報から考察すると，配置するオブジェクト画像の縦横比率に依って，自然に見えるサイズに調節するタスクを行う上では，適した手法が変化するということが考えられる．これは，撮影したカメラの目線の高さに依って傾きが生じ，画像として認識されるオブジェクトの高さが低いほど，この傾きに依って自然なサイズとの乖離が起きてしまい，その結果椅子を配置する時は，高さを

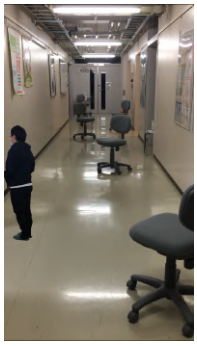

図 18　手法（1） 実寸法比率

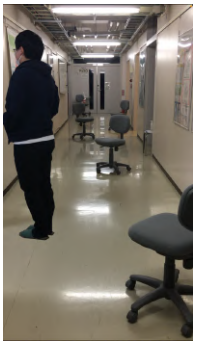

図 19　手法（2）　実寸法比率

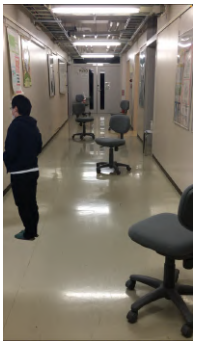

図 20　手法（3）　実寸法比率

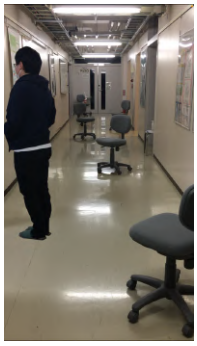

図 21　手法（1）
オブジェクト認識サイズ比率

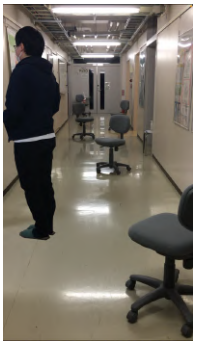

図 22　手法（2）
オブジェクト認識サイズ比率

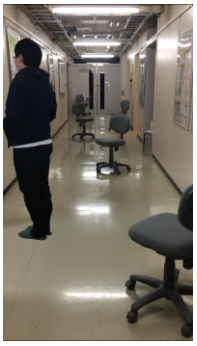

図 23　手法（3）
オブジェクト認識サイズ比率

表 3　人を合成した時の評価値

	MAE_width	MAE_height	MAE_rat_width	MAE_rat_height
手法（1）　実寸法比率	29.00000	118.50000	0.56596	0.52892
手法（2）　実寸法比率	4.25000	9.50000	1.05438	**0.98569**
手法（3）　実寸法比率	12.75000	57.41667	0.85485	0.79725
手法（1）　オブジェクト認識サイズ比率	4.16667	28.66667	0.94529	0.88243
手法（2）　オブジェクト認識サイズ比率	2.91667	**4.83333**	1.04115	0.97318
手法（3）　オブジェクト認識サイズ比率	**2.66667**	19.33333	**0.98125**	0.91765

表 4　被験者 11 名の感覚と人を合成した時の比較による評価値

	MAE_width	MAE_height	MAE_rat_width	MAE_rat_height
手法（1）　実寸法比率	36.62879	131.40152	0.48071	0.48206
手法（2）　実寸法比率	4.87879	17.47727	1.03012	1.03249
手法（3）　実寸法比率	12.29545	44.73485	0.82263	0.81979
手法（1）　オブジェクト認識サイズ比率	5.46212	19.81818	0.90269	0.90193
手法（2）　オブジェクト認識サイズ比率	**2.92424**	**10.66667**	**0.99428**	**0.99471**
手法（3）　オブジェクト認識サイズ比率	3.05303	11.09091	0.93715	0.93796

基準にすると誤差が大きくなり，人を配置すると横幅が露骨に誤差を多く含み，高さを基準にすると誤差が少なくなったと考えられる．この考察については，図 17 を見ると誤差が増減する要因がよくわかる．縦長のオブジェクトに対して，正方形のオブジェクトの方が図の矢印部分で示された幅の変化が大きく変わることがわかる．これが誤差を発生させる要因であると考えられる．但し、表 2 を見ると，表 1 と比較すると，手法（1）と手法（2）の有用性が逆転している．これは，評価に用いた平均値に依る影響が出ていると考えられる [13]．そのため，人を合成した時の結果である表 3 と表 4 を比較すると，誤差の大きい順位は変化しておらず，有用性も変化していない．以上を踏まえると，本稿の実験画像においては，貼り付けるオブジェクト画像の縦横比率が縦に長いほど，高さを基準にした手法（2）が有用であり，縦横比率が正方形に近いほど，$\sqrt{\text{面積}}$ を基準にした手法（3）が有用であり，縦横比率が横に長いほど，横幅を基準にした手法（1）が有用であると考えられる．

また，表 3 の実寸法手法とオブジェクト認識比率を用いた手法との比較では，総合的にオブジェクト認識比率を用いた手法の方が誤差が減る傾向にあるということがわかる．これは，本稿の実験において，人間がオブジェクトを認識し，バウンディ

13：誤差がお互いに微小であるため，僅かな誤差で有用性が逆転する

ングボックスを設定していることが強く影響していると考えられる．上記の考察で述べた，人のような，縦に伸びているオブジェクト画像は手法（2）が適切な手法であると仮定した場合，手法（2）同士の比較では，実寸法手法も，オブジェクト認識比率を用いる手法も，評価値が大きく変化している訳ではないため，どちらの手法も不自然さを払拭する上では有用であると考えられる．しかし，データの集め易さを考慮した場合，本稿の実験結果は，オブジェクトの認識比率を求める際に，画像の収集及び，オブジェクト認識そのものを人力で行なっている上での評価値であるため，これらのタスクを自動化するのであれば，タスクを処理する AI の性能に依存し，現状の技術レベルで本稿の実験結果を再現することは難しいと考える．以上のことから，実寸法手法の方が有用であると考えられる．

尚，手法（1）及び手法（3）においては，オブジェクト画像の種類に依っては，画像に写り込む角度に依って横幅が変わってしまう（ベッドなどであれば，写る向きに依って横幅と定義するものが大きく変化する）ため，誤差が激増する可能性があり，実用化するには多くの課題が残されていると考える．

5　まとめと今後の課題

本稿では，奥行きがわかり易い画像を対象に，貼り付け元に含まれているオブジェクトのバウンディングボックス情報を用いて，位置に対するサイズを調節する手法について，貼り付けるオブジェクトのクラスと回帰式算出の際に用いるクラスが同じ場合，或いは異なる場合について提案した．次に，実験において，それらの手法を貼り付けるオブジェクトのクラスと同じ場合と異なる場合の両方を検証し，実際にその位置に配置した時のサイズと，実際のオブジェクトのサイズ及びアンケートによる被験者が適切であると判断したサイズと比較して，横幅と高さの誤差について評価を行なった．その結果，貼り付けるオブジェクトの縦横比率に依って，適した提案手法が変化する可能性があることがわかった．しかし，横幅を基準にすることの有用性について確認が取れていないため実験の必要がある．加えて，横幅や $\sqrt{\text{面積}}$ を基準にして，適したサイズを求める場合，オブジェクトの回転に依って横幅や $\sqrt{\text{面積}}$ とする値が大きく変わってしまうことが課題として挙げられる．例えば，シングルサイズのベッドを貼り付ける場合，見る角度に依って短辺と長辺が入れ替わる．つまり，オブジェクト認識を行う際，対象の物体がどの程度回転しているのかを認識し，その回転に応じた寸法や他の画像から算出した比率を用いるといった工夫が必要であると考えられる．

また，自動化については，実寸法を用いる手法であれば，文字通り実際の寸法データを自動的に収集することや，あらかじめ様々なオブジェクトの寸法を記録した大規模なデータを用意するといったことが必要である．一方，他の画像からオブジェクト認識のサイズによって比率を算出する方法では，回帰式を求めるのに利用したクラスと，貼り付けるクラスの両方が写っており，出来る限りオブジェクト同士が並列に写っているといった厳しい前提条件を満たす画像を，複数枚自動で収集出来る必要がある．加えて，手法全体として，オブジェクト認識が現状より高精度（少なくとも壁や他のオブジェクトと隣接していないような状況の場合は，正しくサイズやクラスを推測出来る程度）でなければならないため，その点についても課題が残されている．さらに，現状では奥行きが人間の目で明確に認識出来るような（消失点がわかり易い）画像であり，加えて，貼り付け元の画像に何らかの同一のクラスのオブジェクトが存在している場合に限り，提案手法で位置に合ったサイズを調節出来るため，より幅広い条件の画像に適用出来るようになるためには，貼り付け元に存在するオブジェクトのクラスが違う場合でも，サイズの指標となる回帰式が適切に求められるような手法や，セマンティックセグメンテーション（意味的領域分割）によって，床や壁の領域を特定し，その領域の奥行きに対する収縮拡張具合をもとに回帰式を計算するなど多くの改善方法が考えられる．まとめると，回帰式の計算方法及び，横幅や $\sqrt{\text{面積}}$ を基準とした計算手法，完全自動化など，複数の課題が残されており，本稿の提案手法は，多くの改善の余地を含んでいる．

文　献

[1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, “Generative Adversarial Networks”, Machine Learning (stat.ML), Machine Learning (cs.LG), pp.1–9 (2014). https://arxiv.org/abs/1406.2661

[2] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, Honglak Lee, “Generative Adversarial Text to Image Synthesis”, Neural and Evolutionary Computing (cs.NE), Computer Vision and Pattern Recognition (cs.CV), pp.1–10 (2016). https://arxiv.org/abs/1605.05396

[3] Mehdi Mirza, Simon Osindero, “Conditional Generative Adversarial Nets”, Machine Learning (cs.LG), Artificial Intelligence (cs.AI), Computer Vision and Pattern Recognition (cs.CV), Machine Learning (stat.ML), pp.1–7 (2014). https://arxiv.org/abs/1411.1784

[4] 相場 築, 荒澤 孔明, 服部 峻, “GAN 出力画像をフィルタリングするための複数の CNN を用いた画像評価”, 電子情報通信学会情報ネットワーク研究会 (SIG-IN), 信学技報, Vol.120, No.311, IN2020-53, pp.55–60 (2021).

[5] 石川 恵悟, 岩田 伸一郎, “空間写真から空間把握を行うプロセスに関する研究”, 一般社団法人 日本建築学会, pp.91–96 (2008). https://www.aij.or.jp/paper/detail.html?productId=641056

[6] 水野 将成, 松平 真義, 興膳 生二郎, “2 次元画像から 3 次元データを生成するシステムの開発と応用（1）”, Japanese Society for the Science of Design, pp.1–2 (2004). https://www.jstage.jst.go.jp/article/jssd/51/0/51_0_C19/_article

[7] Tingting Qiao, Jing Zhang, Duanqing Xu, Dacheng Tao, “MirrorGAN: Learning Text-to-image Generation by Redescription”, Computation and Language (cs.CL), Computer Vision and Pattern Recognition (cs.CV), Machine Learning (cs.LG), pp.1–10 (2019). https://arxiv.org/abs/1903.05854

[8] Joseph Redmon, Ali Farhadi, “YOLOv3: An Incremental Improvement”, Computer Vision and Pattern Recognition (cs.CV), pp.1–6 (2018). https://arxiv.org/abs/1804.02767

[9] Scikit-learn ライブラリ, https://scikit-learn.org/stable/ (2022).

[10] 人と椅子の画像例, https://www.irisohyama.co.jp/led/houjin/tokyo-antenna-office/ (2022).