氏　　名	許　彩娥
博士の専攻分野の名称	博士（情報科学）
学位記番号	医工農博甲第３７号
学位授与年月日	令和２年９月２９日
学位授与の要件	学位規則第４条第１項該当
専攻名	人間環境医工学専攻　生命情報システム学コース
学位論文題目	Semi-Automatic Generation of Users' Desired Face Image （ユーザが望む顔画像の半自動生成）
論文審査委員	主査　教授　茅　暁陽 教授　福本　文代 教授　大山　勲 教授　服部　元信 准教授　木下　雄一朗 准教授　豊浦　正広

学位論文内容の要旨

Face image synthesis has many potential applications including public safety, such as video surveillance and law enforcement. For example, one important application is in assisting the police to create the face image of suspects based on the memories of witnesses or victims. However, drawing an image based on descriptions of what is in one's mind is not an easy task for the majority of people. Although the montage approach to face image synthesis allows users to create face images by selecting face components, it involves the time-consuming task of choosing the right parts from a wide array of options. It is known that the composition of face parts is a more important factor in the perception of a face than the individual parts. It can be very difficult to adjust the positions of individual parts to achieve a desired composition. Several methods have been developed for synthesising face images according to sketches. Such methods, however, require the user to provide a sketch, which is not always a possibility. Motivated by these potential applications and the limitation of existing face image synthesis technologies, this thesis introduces two methods for allowing a user to generate her desired face image in a semi-automatic way. While the first project uses traditional hand -crafted feature, the second project employ the state-of-the-art deep neurol network technique.

In the first project, a user-friendly system that can create a facial image based on the user's feedback. Unlike most of the existing methods, which require a sketch as input or the tedious work of selecting similar facial components from an example database, our method can synthesize a user desired face image without questioning the user on the explicit features of the face in his or her mind. Through a dialogic approach based on a relevance feedback strategy to

translate facial features into input, the user only needs to look at several candidate facial images and judge whether each image resembles the face that he or she is imagining. Specifically, the proposed system includes three major components: extracting primary features, training an OPF classifier based on relevance feedback and synthesizing face images that do not already exist in the database. We used 1,000 images of Asian faces from the CAS-PEAL database and the Cartoon Face database with 96 × 128 resolution. We extract face feature from pixel-level and use the 80 dimensions with the highest eigenvalues based on a Primary Component Analysis (PCA) as our feature space. The Optimum-Path Forest (OPF) classifier is employed for classifying whether a face image resembles the face in user's mind and it is trained based on user's feedbacks. The purpose of classifying is to synthesize a new image resembling the face in the user's mind rather than to classify or retrieve. A set of sample face images that are based on users' feedbacks are used to dynamically train the Optimum-Path Forest to classify the relevance of face images. Based on the trained Optimum-Path Forest classifier, the top K candidate face images that best match the user's desire are retrieved and interpolated to synthesize new face images the user had imagined. The experimental results show that the proposed technique succeeded in generating images resembling a face a user had imagined or memorized. However, there are some disadvantages under this method. The result sometimes is blurring and doesn't look like the desired face exactly. The proposed method can't synthesize color face image. Such drawback is mainly due to the feature representation. We employed a global feature space based on PCA which fail to capture the personal detail well, causing the generated face quite similar to the average face. Another problem of this technique is that the results are synthesized from the linear interpolation of the top K of user favored face images and it fails to generate face image of completing new features result from.

Recently, with the rapid development of deep learning technology, various research areas and applications, such as computer vision, robotics, big data analysis, and pilotless automobiles, have achieved major advancements. The field of face image generation and synthesis is no exception, as it has also undergone significant developments. In particular, the emergence of the Generative Adversarial Network (GAN), which is a type of neural network architecture for the generative model first proposed by Goodfellow et al. in 2014, brought about a major breakthrough in the field of face image generation. The second method using a novel approach combining GAN with relevance feedback to compensates for the low image quality of the first method.

GAN consists of two networks: the generator that creates as realistic data as possible and the discriminator that attempts to distinguish fake samples from real ones. The two networks compete with each other during the training process, resulting in a generator that can produce realistic data. Since the very first GAN model is proposed in 2014, there are a large variety of GAN have been developed. Conditional generative adversarial network (CGAN) is proposed to gain some control over the generated results by allowing inputting a condition to the model in addition to the noise. Various method based on CGAN model has been proposed for generating face images with specific attributes, such as race, age, and emotion. Although these varieties of GAN have improved the generated results from different angles, however, there are no detailed control over the generated results. To the best of our knowledge, none of the existing GAN models can provide users with easy control over detailed facial features, such as the shapes and positions of individual parts of the face.

The proposed method combines GP-GAN, a CGAN model which can generate face image from the feature vectors representing the geometry information of face, with OPF for allowing the users to generate their desired face images in a semi-automatic way. It consists of two parts: (1) a relevance feedback framework for users to generate the landmarks of new candidate faces by evaluating the sample face images and (2) the face image generator using GP-GAN with the

new landmarks resulting from the relevance feedback process. The relevance feedback framework consists of three steps: constructing the feature space, training the OPF classifier, and exploring the candidate feature vectors. First, the feature space based on the extracted landmark features that are used to localize and represent salient regions of the face, such as: eyes, eyebrows, nose, mouth, and jawline, is constructed. Second, an OPF classifier is trained for the feature space based on the user's feedback so that the user can quickly retrieve the training images that are most similar to the target face. Third, the new candidate feature vectors for synthesizing the target face image is created. At the third step, unlike the algorithm in the first method, n new candidate landmarks of the desired face are created by moving the landmark the use favour the most toward a direction which can further improve the similarity to the user's desired face. In this way, the proposed method can take full advantage of the high image quality while compensating for the lack of user intervention of state-of-the-art GAN technology. Three types of experiments are conducted to validate the effectiveness of the proposed method. The first experiment invited participants to create face images, and the second experiment had another group of participants evaluate the generated results. The third experiment aimed to compare the results of two projects. The experiment results demonstrated that the proposed method can be used to generate not only a face image resembling the target face but also a face image in the user's memory or imagination.

In summary, this thesis proposed a semi-automatic approach to face synthesis. The first method features with the idea of generating face image in the user's mind based on relevance feedback using OPF. Through a dialogic way, the user only needs to look at several candidate face image and judge them according to similarity. The second method succeeded in overcoming the disadvantage in the first project and improved the image quality by combining OPF with the state-of-the-art deep learning technology.

論文審査結果の要旨

１５００字前後（目安：４４字×３５行＝１５４０字程度）

顔画像合成には，公共安全を含む多くの用途がある．たとえば，重要なアプリケーションの１つは，目撃者または被害者の記憶に基づいて警察が容疑者の顔画像を作成するのを支援することである．しかし，記憶に基づいて画像を作ることは簡単な作業ではない．顔画像合成へのモンタージュアプローチでは，ユーザは顔のコンポーネントを選択して顔画像を作成できるが，さまざまなオプションから適切なパーツを選択するのに時間がかかる．また，顔のパーツの構成は，個々のパーツよりも顔の知覚においてより重要な要素であることが知られている．個々のパーツの位置を調整して，目的の構成を実現することは非常に困難である．スケッチに従って顔画像を合成するためのいくつかの方法が開発されているが，入力としてユーザがスケッチを提供する必要があり，これは常に可能であるとは限らない．このようなニーズと現状を踏まえ，本学位論文は，ユーザが自分の望む顔画像を半自動で生成できる２つの方法を提案した．最初の方法は従来型の画像特徴を利用することに対して，二つ目の方法は最先端のディープニューラルネットワーク技術を採用した．

最初の方法では，relevance feedback 戦略に基づく対話的なアプローチを通じて，ユーザはいくつかの候補の顔画像を見て，各画像が自分の想像している顔に似ているかどうかを判断するだけで済む．実装したシステムは特徴の抽出，relevance feedback

に基づく分類器の学習，新しい顔画像の合成の 3 つの主要コンポーネントから構成される．特徴抽出では入力画像の１ピクセルを１次元とする画像特徴について，主成分分析（PCA）を行い， 80 次元の特徴空間を得る．顔画像がユーザの望む顔に似ているかどうかを分類するために Optimum-Path Forest（OPF）分類器を使用し，ユーザのフィードバックに基づいて学習する．分類の目的は，顔画像の分類や検索ではなく，ユーザが望む顔の特徴空間における位置を特定するものである．学習した OPF 分類器に基づいて，ユーザの希望に最も一致する上位候補の顔画像の特徴が取得され，補間されることでユーザが想像した顔画像が合成される．実験結果は，提案手法が，ユーザが想像または記憶した顔に似た画像を生成可能であることを示した．ただし，この方法には．結果の画像がグルースケールであったり，ぼやけていたり，望む顔と正確に一致しないなどの欠点がある．これらの欠点は，主に用いた特徴によるものであった． PCA に基づく画像特徴は，個人の詳細をうまく捉えることができず，生成された顔は平均的な顔とよく似ている．また，別の問題として，ユーザの望む顔画像は OPF から検索した上位 K 個の画像の特徴を線形補間して生成した結果であり，完全に新しい顔画像は生成できないことである．

本学位論文が提案する二つ目の方法は，近年急速な発展を遂げているディープラーニング技術を用いて上述の一つ目の方法の問題を解決した．ディープラーニングはコンピュータビジョン，ロボット工学，ビッグデータ分析，自動運転などのさまざまな研究分野やアプリケーションに応用され大きな成功を収めている．顔画像生成分野も例外ではなく，重要な発展を遂げている．特に，2014 年に Goodfellow らによって最初に提案された生成モデルのニューラルネットワークアーキテクチャの一種である生成敵対ネットワーク（GAN）の出現が顔画像生成の分野に大きな進歩をもたらした． 2 つ目の方法は，GAN を relevance feedback と組み合わせて，最初の方法の低画質の問題を解決する．

GAN は 2 つのネットワークで構成されている．可能な限り現実的なデータを作成する生成器と，偽のサンプルを実際のサンプルから区別しようとする弁別器である． 2 つのネットワークは，学習プロセス中に互いに競合し，現実的なデータを生成できる生成器を生成する． 2014 年に最初の GAN モデルが提案されて以来，さまざまな GAN が開発されている．条件付き生成敵対的ネットワーク（CGAN）は，ノイズに加えてモデルに条件を入力できるようにすることで，例えば，種，年齢，感情など，生成された結果をある程度制御することが可能である．しかし，既存の GAN モデルのいずれも，顔の個々の部分の形状や位置などの詳細な特徴を簡単に制御することはできない．

提案手法は，顔の輪郭と各パーツの幾何学的特徴を捉えられる Landmark と呼ばれる制御点から顔画像を合成できる GP-GAN と呼ばれる CGAN モデルを OPF と組み合わせて，ユーザが半自動で目的の顔画像を生成できるようにした．これは，(1) ユーザがサンプルの顔画像を評価して新しい候補顔の Landmark を生成するための relevance feedback フレームワーク，および (2) GP-GAN を使用して relevance feedback から定義される新しい Landmark から顔画像を生成するディープニューラルネットワークの 2 つの部分で構成される．relevance feedback フレームワークは，①特徴空間の構築，②OPF 分類器の学習，③候補特徴ベクトルの探索の 3 つのステップで構成される．③のステップでは，一つ目の方法とは異なり，OPF から得られる特徴ベクトルをユーザの望む方向にさらに移動することにより，n 個の新しい landmark 候補を生成する．このようにして，提案方法は，最先端の GAN 技術のユーザ介入の欠如を補償しながら，高画質を十分に活用できる．提案方法の有効性を検証するために，3 種類の実験が行われた．最初の実験では参加者に顔の画像を作成してもらい，2 番目の実験では別の参加者グループに生成された結果を評価させる． 3 番目の実験は，方法１と方法２の結果を比較した．実験

結果は，提案方法がターゲットの顔に似た顔画像を生成できるだけでなく，ユーザの記憶または想像の顔画像の生成にも使用できることを示した．

以上のように，本学位論文はユーザに大きな負担をかけることなく，それぞれが望む顔を生成できる新しい技術を確立した．特に，二つ目の方法は，世界に先駆けて，処理がブラックボックス化されている生成型ディープニューラルネットワークモデルに Relevance Feedback の仕組みを導入することに成功し，当該分野において様々な後続研究を触発するきっかけとなる可能性が高い．

本論文の研究内容に関し，博士論文審査要綱に基づき最終試験を実施した．提出された博士論文および公聴会における発表の内容に関連し，研究背景，概念規定，アルゴリズムデザイン，評価実験の妥当性と信頼性，論文構成，情報学的価値などに関する質疑を行い，論文提出者の見識を問うた．その結果，試問の内容において妥当な解答が得られたこと，並びに発表論文の基準を満たすものであったことから，博士論文審査委員会は博士に相応しい学力と見識を有するものとして認め，最終試験を合格とした．