

氏名	李 宏林
博士の専攻分野の名称	博士（情報科学）
学位記番号	医工博甲第437号
学位授与年月日	平成30年9月27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
学位論文題目	Portrait feature detection and its applications to clothes matching and caricature synthesis (肖像写真からの特徴検出とその服装検索及び似顔絵生成への応用)
論文審査委員	主査 教授 茅 暁陽 教授 福本 文代 准教授 安藤 英俊 准教授 服部 元信 准教授 豊浦 正広 東京工科大学メディア学部教授 近藤 邦雄

学位論文内容の要旨

Portrait photos are widely used in social daily life for various applications. Most of standard portrait photos are made up of hair, face and the half upper part of body, which are used in formal occasions, such as applying for job and various certificates. Recently, with the development of Internet and mobile phone, taking selffile, making various styles of corresponding caricatures and displaying caricatures in social or commercial applications are more and more popular. In addition, how to collocate upper clothes (especially for collars) with corresponding faces and hair styles, and how to find desired upper clothes with appropriate collars are interesting applications.

As e-commerce develops quickly, more and more consumers are searching online shops for clothes items. Collar is a crucial part of the clothes due to its horizontal perspective position closest to the observers' eyes and serving as the frame for one's face in portrait photos. Searching for desired clothes with appropriate collars is very important for recommending clothes collocation. Many sites support keyword-based searches, but items in online shops often lack specific design-related tags and include technical names that few shoppers are familiar with. One image search approach that researchers have proposed as visual query-based alternatives to keyword-driven searches is the Content-Based Image Retrieval (CBIR) method, which involves expressing image content in feature vectors and then comparing the similarities between various images. Although CBIR methods eliminate the need to build queries out of keywords or other linguistic information and allow users to search for visual information with visual information input, their efficiencies largely depend on the quality of the feature vectors, and it remains to be challenged to extract the image features capturing the design of clothes well. Some researchers have attempted to use color and texture, but few researchers have delved into the possibilities of developing feature vectors to capture the designs of clothing in detail. The first contribution of the

thesis focuses on capturing the image features representing details of the collar designs. Three types of collar features are designed, which are Sift feature, Saliency feature and Saliency-Sift feature. In addition, a prototype of clothes image retrieval system based on Relevance Feedback approach and Optimum Forest algorithm is also developed to improve the query results iteratively. A series of experiments are conducted to test the qualities of the three types of designed features and validate the effectiveness and efficiency of the RF-OPF prototype from multiple aspects. The evaluation scores of initial query results are used to test the qualities of the feature extraction methods. The average scores of all RF steps, the average numbers of RF iterations took before achieving desired results and the score transition of RF iterations are used to validate the effectiveness and efficiency of the proposed RF-OPF prototype.

Portrait caricatures have been serving as icons, avatars and so on in real life and internet. Most caricature synthesis systems are based on photo-transformed or example-based techniques. Recently, style-transfer-based and deep-learning-based techniques have also been developing rapidly. While photo-transformed systems are difficult to be generalized for synthesizing different caricatures of expressive styles, traditional example-based systems require photo-caricature pairs and suffer from the difficulty of obtaining sufficiently large databases of images for use as teaching pairs. Style-transfer-based systems mainly transfer colors and textures and cannot transfer geometric information from the example caricatures. Developed deep-learning-based systems also require huge quantities of paired samples for training. The second contribution of the thesis is a new caricature synthesis system under example-based framework based on the feature deviation matching method, a cross-modal distance metric, which employs the deviation from average features rather than the values of features themselves to search for similar components from caricature examples. The contributions of this proposed framework are: 1. The newly proposed cross-modal distance metric called feature deviation matching technique makes it possible to generate various styles of caricatures under the conventional example-based framework without requiring paired photo-caricature training sets. 2. By focusing only on the perceptually prominent features, the designed feature vectors are robust and effective for capturing the visual features of input portrait photos. 3. The proposed system enables users to control the exaggeration of individual facial components. Various combinations of individual facial and hairstyle components, based on different exaggeration coefficients and similarity rankings, can provide users with different candidates to satisfy their particular preferences; this has not been achieved in most existing style-transfer-based and deep-learning-based approaches. Extensive experiments are conducted to evaluate the results: 1. Similarity of the three types of caricatures (expressive, photo-realistic and drawing) with input portrait photos. 2. Comparison with the paired-example method. 3. Effectiveness of the designed hair and facial component features.

Retrieving similar or desired clothes images according to users' input clothes images is popular for online shopping. Collocation recommendation or evaluation systems are advanced applications developed from retrieval modules. Most existing recommendation systems are designed for clothes component or clothes-situation collocations without considering personality. In other words, they do not consider the factors of personal faces, hairstyles, body sizes and so on. Different people with different hairstyles and different clothes will also give us different impressions. It is necessary to recommend or evaluate clothes collocation according to personal faces, hairstyles and body sizes, which are very important for clothes collocation. Based on the above considerations, the third contribution of the thesis includes three preliminary frameworks based on deep learning techniques for evaluating portrait photos and recommending appropriate upper clothes according to input face photos.

In summary, this thesis focuses on the portrait photos and designed a set of hand-crafted as well deep features and developed the state-of-art learning frame works for retrieving desired clothes with appropriate collars, generating various types of extractive caricatures and evaluating portrait photo and upper clothes collocation. The effectiveness of the proposed features and algorithms have been evaluated through carefully designed subject experiment and the limitations are discussed in the thesis.

論文審査結果の要旨

肖像写真は、様々な用途のために日常生活において広く使用されている。標準的な肖像写真のほとんどは、髪、顔、身体の上半分の部分で構成され、就職や各種証明書の申請などの公式な用途にも使用されている。本学位論文では肖像写真画像からの特徴抽出技術とそれらの特徴を利用した①好みの襟の衣服画像検索、②似顔絵自動生成、③似合う服装の推薦の3つの応用技術を提案した。

電子商取引が急速に発展するにつれて、より多くの消費者がオンラインショップで衣料品を探している。襟は、水平位置が観察者の目に最も近いことと、顔の枠として機能するため、服のもっとも重要なデザイン要素である。多くのサイトではキーワードベースの検索がサポートされているが、より詳細なデザインを指定することは不可能か、オフショールダーやタートルなど、購入者が慣れていない専門的なファッション用語を知っている必要がある。キーワード駆動型検索に対して、視覚的なクエリーに基づく検索アプローチとして画像間の類似性を比較するコンテンツベース画像検索 (CBIR) 方式が知られている。CBIR 法の効率性は主に画像間の比較に使われる特徴ベクトルの設計に依存し、特徴ベクトルは服のデザインをうまくキャプチャする必要がある。これまで衣服画像の特徴ベクトルとして色とテキストを提案した研究者はいるが、デザインを詳細に把握できる特徴ベクトルはまだ提案されていなかった。本学位論文の1番目の貢献は、襟のデザインの詳細を表す画像特徴の設計と検出に焦点を当てている。Sift、Saliency、Saliency-Sift の3種類の特徴を設計し、画像からこれらの特徴を検出するアルゴリズムを開発した。さらに、検索結果を繰り返し改善するために、ユーザの評価に基づき結果を Refine できる Relevance Feedback (RF) アプローチと Optimum Forest(OPF)学習アルゴリズムに基づく衣服画像検索システムのプロトタイプも開発した。3つの特徴と RF-OPF プロトタイプの有効性と効率性を複数の側面から検証する一連の実験を行い、すべての RF ステップの平均スコア、所望の結果を達成するまでの RF 反復の平均回数、RF 反復のスコア遷移を使用して、提案された RF-OPF プロトタイプの有効性および効率性を検証した。

一方、似顔絵は、実生活やインターネットでアイコン、アバターなどとして使われてきた。ほとんどの既存似顔絵合成システムは、画像フィルタリングまたは事例に基づく技術を用いている。近年、スタイル転写に基づく技術と深層学習に基づく技術も急速に発展している。画像フィルタリング技術はスタイルに特化したアルゴリズムを使用するため、汎用性に欠ける問題がある。一方、伝統的な事例ベースシステムは写真と似顔絵のペアを必要とし、教示ペアとして使用するのに十分な大きさの画像データベースを得ることが困難である。スタイル転写ベースのシステムは主に色とテキストを転写し、似顔絵から顔や目の形状のような幾何情報を転写することはできない。近年注目されている深層学習に基づくシステムも学習に大量のペアのサンプルを必要とする。この論文の第2の貢献は、特徴の値そのものよりもむしろ特徴の平均値からの偏差を採用するというクロスモーダル新しい距離メトリックの提案と、この偏差マッチング法に基づく事例ベースフレームワークの下での新しい似顔絵合成システムの実装である。提案されたクロスモーダル距離メトリックは、ペアとなっている写真と似顔絵トレーニングセットを必要とせずに、従来の事例ベースフレームワークの下で様々なスタイルの似顔絵を生成することを可能である。また、知覚的に顕著な特徴にのみ焦点を当てることにより、設計された特徴ベクトルは、頑強でありながら、顔の視覚的特徴を効果的に捕捉することができる。また、提案されたシステムにおいて、ユーザが個々の顔構成要素の強調を制御することも可能である。

本学位論文の第3の貢献は、顔の特徴と衣服の特徴とのマッチングに着目した似合う衣服の推薦に関する研究である。ユーザの入力衣服画像に応じて類似または所望の衣服画像を検索することは、オンラインショッピングにおいて一般的になりつつある。一方、コーディネート推奨または評価システムは、検索モジュールから開発された高度なアプ

リケーションである。しかし、既存の推薦システムのほとんどは、パーソナリティを考慮せずに、服装の構成要素のコーディネートを行っている。言い換えれば、既存システムでは個人の顔、髪型、体の形状などの要因を考慮していない。異なる髪型と異なる服装をもつ人は見る人に異なる印象を与えるはずである。したがって、衣服のコーディネートのためには個人の顔、髪型、および身体の特徴を考慮する必要がある。上記の考えに基づき、本学位論文は、肖像写真を評価し、似合う上着を推奨するための深層学習に基づく3つの枠組みを提案し、予備的な実験により枠組みの有用性と今後の課題を示した。

以上のように、本学位論文は、肖像写真に焦点を当て、衣服の検索とコーディネート及び似顔生成に必要な特徴抽出技術とその応用技術を提案・実装した。特徴検出という画像処理の基礎的な課題の解決への貢献のみでなく、ソーシャルメディアやオンラインショッピングへの応用をはじめ、社会的な意義も大きい。

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、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。