	そん　かいてん
氏　　名	**孫　海天 (HAITIAN SUN)**
本籍（国籍）	中国
学位の種類	博士(工学)
学位記番号	工博　第３１６号
学位授与年月日	令和２年３月２３日
学位授与の要件	学位規則第５条第１項該当　課程博士
研究科及び専攻	工学研究科デザイン・メディア工学専攻
学位論文　題目	**Image Retrieval and Object Detection based on Multiple Categories of Queries** **(複数類のクエリに基づいた画像検索及び目標検出)**
学位審査委員	主査　教授　　藤本　忠博 副査　准教授　　田中　隆充 副査　准教授　　明石　卓也

論　文　内　容　の　要　旨

Nowadays, images and videos are increasingly popular and appear in people's daily life frequently. Institutes carry a surprisingly large number of images, the number of which is still growing fast. It often occurs that people intend to find their desired elements and factors in the massive images. In this case, the approaches to retrieving the images that contain various visual information and detecting the objects in such images accurately and fast are necessary. Therefore, image retrieval (IR) and object detection (OD) has been studied for decades. Content-based image retrieval (CBIR), also known as content-based visual information retrieval (CBVIR), has achieved noticeable successes in the last decades. The term "content" represents numerous visual information that can be possibly extracted from images, such as color, texture, and shape, rather than semantic meta-data such as tags or descriptions. The CBIR system requires a query and measures the similarity between the query and each database image to rank the database images. OD aims at locating known instances (objects) in images or image sequences. Common OD detects semantic objects of classes with the preknowledge. One extreme situation of OD is when each semantic training class has only one training image. Such a problem is defined as one-shot OD or template matching, where the image for training is also named as the template. Another extreme situation is named zero-shot OD, meaning there is no image for direct training. The classifier for zero-shot OD is usually trained by the relationship with other known classes.

The query for IR is an input image, representing visual information and treated

as a precise request. The query for IR can be an RGB image, depth image, sketch image, contour image, etc., involving various categories of objects. On the other hand, for OD, the semantic objects of classes are considered as queries in this research.

In order to systematize and find the commonalities of the multiple categories of queries, in this thesis, three of them are discussed, including the two for IR and the one for OD.

1) The image containing one whole-body human for IR. Instead of visual similarity defined by colors, shapes, or textures, this research aims to retrieve images with respect to the visual similarity defined by the human pose. In this framework, all the poses are derived from images, which is inspired by the recent development of 3D human pose reconstruction. Furthermore, to make the retrieval more robust against reconstruction error, a recurrent bidirectional similarity measure named recurrent best-buddies similarity (RBBS) is proposed. Both of the qualitative and quantitative results show the usefulness of this framework, especially the quantitative results evaluated by mean average precision (MAP or mAP) exhibit RBBS is improved by 14.13% compared to the most competitive alternative methods.

2) The simplified drawing with only strokes, named sketch for IR. Sketch-based image retrieval (SBIR) is a popular research field that is to rank database images by comparing the similarity between query sketch and database images. This thesis proposes to compress binary line drawing (sketch) by approximation automatically, considering that it can be applied to SBIR. Specifically, a sketch contains several strokes, each of which can be segmented into several segments by extracting breakpoints according to the curvature. The approximation of the segments is recorded for the compression. The experiment reveals that the relationship between a certain pair of segments can be represented by some geometrical functions approximately in a rather low dimension. The proposed compressed representation is not only invariant with respect to rotation, scaling, and translation but can also filter out the noise of wobbly lines in some cases if applying it to SBIR.

3) Product images for OD. Product recognition performs a significant role because of its benefits to the compliant arrangements of stores, which further affects the commercial contracts, customer satisfaction, and sale achievement. Automatic recognition systems have been proposed owing to the high cost of the manual inspection by clerks currently. Because of the difficult collection of product images, the systems are commonly in one-shot cases, in which the training data is template product images actually. However, despite the development of one-shot recognition, the systems rarely utilize special characteristics of products on retail store shelves, and the frequent updating of templates is still challenging. Furthermore, it is considered that the product detection can be the basis of product

recognition. In this research, instead of the present workflow, a novel product detection system, named TemplateFree is proposed, which combines product segmentation and zero-shot learning. It detects products on retail store shelves by single store shelf images, i.e., corresponding template product images are not necessary. TemplateFree concentrates on the characteristic that a store shelf can be segmented horizontally into layers then vertically into products so that each product can be detected according to the segmentation. Double zero-shot deep learning frameworks are employed to improve the segmentation. In experiments, TemplateFree achieves better results than the present method.

論 文 審 査 結 果 の 要 旨

本論文は，複数種類のクエリを用いた画像検索及び目標物体検出アルゴリズムを提案している．近年，ハードウェアの発展によって日常生活において，静止画像だけでなく動画像を扱う場面が多い．これらのデータを有効に利用するためには，大量の画像や動画の中から目標となる画像やシーンを探索する方法が必要不可欠である．画像検索ではクエリ画像とデータベース画像の近似度を計算し，近似度によってデータベース内の画像をランク付けする．画像検索におけるクエリは入力画像である．一方，目標物体検出では画像または画像シーケンスの中の既知の目標物体の位置を推定することを目的とし，テンプレートなどの事前知識によって検出する．事前知識であるテンプレートを 1 枚しか準備できない場合の物体検出手法を one-shot 物体検出と呼ぶ．これに対して，直接的な事前知識を用いない zero-shot 物体検出がある．この zero-shot 物体検出は目標物体のクラスに関する情報は無いが，他のクラスとの関係を用いて学習する．あらゆるクラスを扱うシステムを構築するためにはこのようなクラス間の情報を用いる手法が必要である．本論文では，クラスをクエリと捉え，3 種類のクエリを討論している．最初のふたつは，人間の全身姿勢を含む画像検索とスケッチ画像検索を対象としている．3 つ目はクエリに商品画像を用いている．それぞれの提案手法は既存手法と比較して優れており，従来の研究では行われておらず，新規性の高いものである．

本論文の構成は以下の通りである．

第 1 章は序論であり，この研究の背景や取り組んでいる内容，これまでの画像検索と目標物体検出手法のまとめ，クエリの問題点，討論するクエリの紹介，提案手法のアプローチといった概要が記述されている．最後に，討論するクエリに関する提案手法について記述されている．

第 2 章では，画像検索の人間姿勢画像というクエリに関する提案手法について述べられている．色や形状やテクスチャによって定義される視覚的な類似性の代わりに，人間姿勢によって定義される視覚的な類似性に関する画像を取得することを目的とする．すべての姿勢は撮影された 1 枚の画像から推定される．さらに，2D 姿勢から 3D 姿勢を再構築する際に発生するエラーに対し，リカレントベストバ

ディ類似性（RBBS）という再発双方向類似性測度を提案している．定性的および定量的な評価により提案手法の有用性を示している．特に，平均精度（MAP または mAP）で評価される定量的結果において，RBBS が既存手法と比較して 14.13%向上している．

第 3 章では，画像検索のスケッチというクエリに関する提案手法について述べられている．スケッチによる画像検索（SBIR）は，クエリスケッチとデータベース画像の類似性を比較し，データベース画像をランク付けする研究である．SBIR への適用を考慮し，近似によって自動的にバイナリ線画（スケッチ）を圧縮することを提案している．具体的には，スケッチには複数のバイナリ線画が含まれており，これらの線画は曲率に基づいて複数のセグメントに分割される．セグメントの近似値が圧縮のために記録される．提案された圧縮手法は，SBIR に適用する場合，回転とスケーリングと移動に関して不変であるだけでなく，不安定な線のノイズを除去することも可能である点が優位である．

第 4 章では，一般的な商品画像をクエリとした目標物体検出手法について述べられている．現在，人による手動検査のコストが高く，自動認識システムが提案されている．商品画像の収集が難しいため，一般的に one-shot 物体検出が用いられ，トレーニングデータがテンプレート商品画像となる．しかし，one-shot 物体検出はテンプレートの頻繁な更新が必要となる．この章では，TemplateFree という新しい物体検出システムが提案されている．これは，物体のセグメンテーションと zero-shot 学習を組み合わせたものであり，テンプレート画像を必要としない．TemplateFree は，商品棚を水平方向にレイヤーに基づいて分割し，次に垂直方向に商品領域を分割するという特性を利用し，各商品は分割に従って検出される．分割を改善するため，zero-shot 深層学習フレームワークが利用されている．実験では，TemplateFree と既存手法を比較し，優位な結果を達成している．

第 5 章は結論と今後の課題について述べている．討論したクエリに関する提案手法の有効性と有用性を示し，これらのクエリを組み合わせた手法の将来性を検討，分析している．

以上，複数種類のクエリを用いた画像検索及び目標検出アルゴリズムを新規に提案し，その有効性と有用性を示したものであり，メディア工学分野やコンピュータビジョンの発展に寄与するところが大である．

よって，本論文は博士（工学）の学位論文として合格と認める．

原著論文名（３編を記載．ただし，単位取得満期退学後１年以内の申請の場合は，１編を記載）

Haitian Sun, Jing Zhang, Takuya Akashi: TemplateFree: Product Detection on Retail Store Shelves, IEEJ Transactions on Electrical and Electronic Engineering, Vol.15, No.2, pp.242-251, 2020