称号及び氏名	博士（工学）　Fairuz Safwan Mahad
学位授与の日付	2023 年 3 月 31 日
論　　文　　名	「Engineering Features for 3D Reconstruction （三次元復元のための特徴操作）」
論文審査委員	主査　　黄瀬　浩一 副査　　森　直樹 副査　　能島　裕介 副査　　岩村　雅一

論文要旨

3D reconstruction has been a popular topic among computer vision researchers. 3D reconstruction is referred to as the process of capturing the shape of an object or scene. 3D reconstruction can be realized either by traditional means or by means of deep learning-based methods. Traditional 3D reconstruction methods were the norm prior to the advancement of deep neural networks. To date, traditional 3D reconstruction methods are still commonly used in these industries due to the accuracy and robustness. Traditional methods are mostly accurate but completeness is a huge challenge. The advancement of deep neural networks in 3D reconstruction saw impressive results. In most cases, the reconstructed 3D model is complete. However, the shape of the reconstructed 3D model is not faithful to the shape of the original object. Industries such as medical and automobile cannot afford to compromise with accuracy, hence deep neural network-based 3D reconstruction methods are impractical to be used as these industries involve people's lives. To improve the accuracy of the traditional and deep neural network methods, this thesis pursues ways to engineer features for the purpose of 3D reconstruction. To improve the accuracy of the traditional and deep neural network methods, this thesis pursue ways to engineer features for the purpose of 3D reconstruction..

In the traditional 3D reconstruction methods, lack of completeness is a huge challenge. Although the 3D shapes of successfully reconstructed parts of the 3D model are accurate, gaps or holes in the reconstructed 3D models are caused mainly on the surfaces with less texture. It is because the traditional 3D reconstruction methods rely on feature points extracted in the 2D images. The number of extracted feature points heavily depends on the surfaces, and less feature points tend to be extracted on a smooth surface, which is with less texture. To address this issue, we propose a method that makes a smooth surface complex. The proposed method uses a set of randomized kernels. This allows feature point extractors to extract more unique and complex features, thus increasing the number of feature correspondences. We used a plant object dataset consisting of 27 images taken from different viewpoints. The leaves and stems of the plant have a smooth texture, which means it lacks texture. Hence, this makes it a suitable dataset to show the effectiveness of our proposed pipeline. Our proposed pipeline shows that the gaps and holes in the reconstructed 3D model were reduced. More parts of the leaves were reconstructed. The stems of the plant were also denser. Furthermore, the outlines of some leaves were successfully reconstructed, which entirely failed in the traditional pipeline. The results show that our randomized kernels were considerably effective in improving the quality of the reconstructed 3D model.

In the deep neural network-based 3D reconstruction methods, reconstructing accurate 3D models, which have a shape identical to the original object, is a huge challenge. Taking the practical implementation of 3D reconstruction to the next level is quite important but this issue has been overlooked in the computer vision community. We are the first to spend much effort on this issue and took the lead of the community. To address it, we propose a pyramidal hierarchical network. The network can extract features of multiple levels and unify them thanks to two pathways: the bottom-up and the top-down pathways. The bottom-up pathway downsamples the feature maps by half the size, while the top-down pathway upsamples the feature maps twice the size. Due to the processes, features obtained in the bottom-up pathway are with high localization accuracy but semantically weaker, and those in the top-down pathway are the opposite. Our proposed pyramidal hierarchical network aims to combine these features and obtain balanced features with high localization accuracy and semantically stronger. This allows the network to learn both the detailed parts and the overall shape of the object. Quantitative and qualitative results show that our pyramidal hierarchical network outperforms the state-of-the-art. Results show that our pyramidal hierarchical network is able to reconstruct the detailed parts of the object.

To address the same issue mentioned in the previous paragraph, we propose another method called multi-branch network. The concept of our proposed multi-branch network is to separate the feature extraction and learning process into three different components called Low Net, Mid Net, and Global Net. The Low Net is used to learn local features. The Global Net is used to learn generic features, while the Mid Net is used to learn intermediate features. These three components are independent of each other except for the last layer of all three components. Therefore, the network is expected to learn the shapes of 3D models in different levels: the Low Net learns detailed parts, the Global Net does the overall shape, and the Mid Net does the intermediate level of the shape. The qualitative results show that both our proposed methods improved the completeness of the reconstructed 3D models. Qualitative evaluations confirmed that our multi-branch network is better at reconstructing the detailed parts of the object than the state-of-the-art. Quantitative results also show that our proposed method outperformed the traditional method in the majority of the object categories.

Since two methods are proposed in the same problem, we compared both proposed methods. As a result, our pyramidal hierarchical network is more stable than our multi-branch network.Our pyramidal hierarchical network produced more consistent results, whereas our multi-branch network might surpass our pyramidal hierarchical network or even underperform under certain circumstances. In addition, our multi-branch network excelled more in reconstructing the detailed parts of the object than our pyramidal hierarchical network. Our results show that both our proposed networks improved the quality and accuracy of the reconstructed 3D models.

This thesis is aimed at engineering features for 3D reconstruction. Three methods to engineer features were proposed in this thesis: one method is proposed in a traditional approach, and two methods are deep neural network-based. Our proposed method in the traditional approach can improve the quality of 3D reconstruction used in real applications, such as the industry. Our proposed methods in the deep neural network-based approach can enhance the chance that deep learning-based 3D reconstruction methods are used in the industry. All the three methods contribute not only to the computer vision community but related industries.

審査結果の要旨

本論文は，物体やシーンの複数枚の写真から三次元形状を推定する三次元復元に関する研究についてまとめたものであり，以下の成果を得ている．

(1)　深層学習を用いない従来の三次元復元手法では、複数枚の写真から特徴点を抽出し，その対応に基づいて三次元形状を復元する．しかし，抽出される特徴点数は物体表面の滑らかさ等に依存しており，例えば滑らかな物体表面からはほとんど特徴点が抽出されない．このような場合には，特徴点の対応も求まらないため，復元された三次元形状に穴が開くという問題が起こる．この問題に対して，乱数に基づいて作成する画像パッチ（カーネル）を畳み込んでから特徴点を抽出する手法を提案した．これにより，滑らかな物体表面かも特徴点が抽出でき，復元された三次元形状に穴を埋める効果が確認できた．

(2)　深層ニューラルネットワークに基づく三次元復元手法では，従来の三次元復元手法のように復元した三次元形状に穴が開くことは無いが，復元した三次元形状に本来の形状からの「ずれ」が生じることが課題であった．このずれは，細い部位など，特に細部の形状に現れる．しかし，この問題はこれまで大きく注目されることは無く，見過ごされてきた．本研究では，三次元復元に用いるニューラルネットワークが三次元形状を限られた粒度の特徴で表現しようとすることにこの原因があると考え，特徴表現の粒度を増やす手法を提案した．具体的には，ピラミッド型の階層構造を持つ特徴抽出器を用いて複数の粒度を加味した特徴を抽出した．これにより，提案手法は従来手法より良い三次元形状が復元できることを定量的ならびに定性的な評価により確認した．

(3)　前項で述べた深層ニューラルネットワークに基づく三次元復元手法の課題に対して，別のアプローチの解決方法を提案した．提案手法では，局所的な特徴を学習する Low Net，中間的な特徴を学習する Mid Net，概略の特徴を学習する Global Net と呼ぶ 3 つのサブネットワークを特徴抽出の際に用いる．これらの特徴を組み合わせることにより，三次元形状の概形のみならず，細部の形状の復元が可能になる．定量的ならびに定性的な評価により，提案手法は従来手法より良い三次元形状が復元できることを確認した．

以上の研究成果は，深層学習を用いない従来の三次元復元手法ならびに深層学習に基づく三次元復元手法の両方において，特徴操作による三次元復元の改良に関して重要な知見を与えるとともに，本分野の学術的・産業的な発展に寄与するところが大である．また，申請者が自立して研究活動を行うに必要な能力と学識を有することを証したものである．学位論文審査委員会は，本論文の審査および最終試験の結果から，博士（工学）の学位を授与することを適当と認める．