Multi-dimensional feature and data mining

Georgiou, T.

1 Introduction

1.1 Background

Computer vision is a very broad field with numerous applications in scientific, medical and ”everyday life” domains. Human understanding of the world through visual queues extends to three as well as four dimensions, e.g. understanding an action such as falling, hugging etc. In recent years people have been generating and consuming a plethora of multimedia content which has higher dimensionality than the classic two dimensional image, i.e. three and four dimensions. This usage has given rise to many exciting and important applications, such as autonomous vehicles and automated medical 3D & 4D image analysis. These applications both have potentially high impact in our society, but also introduce complex challenges, making the process of such high dimensional data the leading edge of modern computer vision.

In the meantime, the increasing computational capacity of modern hardware, has led to an unprecedented capability of simulating complex fluid dynamics systems, i.e. computational fluid dynamics (CFD) simulations, such as the air around objects, the mix and tumble of air and gas in an internal combustion engine et cetera. These simulations can produce an enormous amount of data, in multiple dimensions, e.g. 4, and many modalities, i.e. a plethora of information for every physical location. The increasing availability of such data has given rise to new applications. For example, is it possible to retrieve engineering designs from a large database based on the flow similarities? Is it possible to utilize these large flow fields to automatically optimize the designs? Such applications might help engineers and scientists understand better the correlation of design principles to the flow and the flow patterns to performance.

For the purpose of machine learning and more specifically computer vision, CFD simulation output shares many similarities to visual data. They both represent the real world in a similar manner. The physical (2-4D) space is separated by a grid, and each individual cell is assigned values representing the real world. In the case of images, these values are the RGB colors, for depth images they are (usually) the distance to the sensor and in the case of CFD simulation output properties of the flow, such as the velocity vector, pressure etc. This similarity between the representation of the data deems possible the adaptation of computer vision techniques to processing CFD simulation output, as it is a much more mature research field.

1.1.1 Feature extraction of CFD simulation output

There exist a plethora of feature extraction methods for flow fields, the vast majority of which are focused on visualization. Good overviews of the flow field feature visualization are [83, 229, 293, 360, 273, 390]. The steps towards feature visualization can be divided into feature definition, decomposition, extraction and visualization. All of these steps can be categorized into two main categories, steady and unsteady (i.e. time dependent) flow field. According to our research, most of unsteady flow feature visualization focuses on tracking steady flow features in time. There are a few exceptions to this rule, such as the path lines, which are specific unsteady flow features.

There are two main categories of steady flow features, local and global features. Local features have specific local behavior and are mathematically defined. Although this is the case, there are many algorithms that try to extract them, all with their limitations and advantages. These features are defined around points in the flow where the flow “vanishes”, i.e. the magnitude of the vectors of the vector field becomes zero. These features can be categorized according to the behavior of the flow around the “vanishing” point. The categorization was introduced by Helman and Hesselink in 1989 for 2D [155] and 1991 for 3D [156]. The field of extracting and visualizing these features was created in the same work and called Vector Field Topology (VFT). VFT is now one of the most established ways of visualizing and analyzing flow field behavior.

Global features, unlike local usually have vague definitions and their detection depends on the specific implementation and application. Some examples are the vortex, flow separation and shock waves. For example, a vortex refers to the swirling motion of a fluid around a specific point [229]. An example definition of the vortex is given by Robinson [294]:

“A vortex exists when instantaneous streamlines mapped onto a plane normal to the vortex core exhibit a roughly circular or spiral pattern, when viewed from a reference frame moving with the center of the vortex core.”

This swirling motion is understood through visual observations and is hard to be mathematically defined. For example how much ”swirliness” is enough to categorize a vortex? How are the boundaries of the vortex defined? These issues make the detection of such features difficult and the implementation usually varies according to the needs of the application.

The limitations described above make the use of flow visualization defined flow field features very hard to use for the purpose of machine learning, where the known algorithms need specific definitions to operate.

1.1.2 Computer vision

Computer vision has been pushing the limits of automated image understanding for decades. The main focus of it is to extract high-level information from visual content such as images, with applications varying from image classification [184] and object detection [100], to scene understanding [334], localization and mapping [351, 352] and many more. Throughout the years, the approaches followed by computer vision researchers have changed significantly. Popular approaches have been global features and description of images such as textures, and color histograms [357, 120]. These features were very computationally efficient which was very attractive for the computational capacity of hardware at that time. As years progressed though, these approaches showed their limits as they are very sensitive to occlusion and clutter. Moreover, local information such as object shapes are disregarded making the distinction between similar objects (e.g. red car vs red motorbike) infeasible.

To deal with such issues, local features are introduced. These approaches follow several steps. First detection of more informative points and regions in an image, then description of such areas and finally feature matching, or aggregation for global description. Some popular examples of local features are the SIFT [225, 226], SURF [19], FREAK [7] and the ORB [299]. These features encode local information, such as histograms of image gradients in a neighborhood, or pixel differences for different point patterns in a neighborhood. These features have enabled applications such as object or scene matching, using algorithms such as the RANSAC. Meanwhile, using feature aggregation to create image global descriptions has enabled applications such as image classification, and content based image retrieval [341].

In recent years the focus has moved to artificial neural networks (ANN) and more specifically deep learning and convolutional neural networks (CNN). Deep learning models boosted the performance of computer vision systems by a large margin [184, 368]. The modern availability of large scale datasets as well as the high computational capabilities of modern GPUs has rendered the training of huge models feasible. Deep learning models are hierarchical models that, in contrast to the more traditional local features description, learn representations solely from the data that they are trained on. Very important stepping stones, that made the success of these models possible is early research done on optimization algorithms, such as the gradient descent and back-propagation [301, 202]. The back-propagation algorithm enables the propagation of error from one layer to the next (or previous, hence “back”) in hierarchical models and thus updating the parameters of these layers. Later, the introduction of convolutional layers made possible the reusability of parameters, minimizing the total number of free parameters and thus making that training of huge models possible [199, 201]. Today, deep neural networks are the most popular approach for high performance tasks, varying from object classification [445], image retrieval [74], scene semantic segmentation [111] and even generating new content like style transfer approaches [221] and swapping faces [181].

With the increase of available higher dimensional content, such as video [340], RGB-D images [344] and CAD models [419], the need to extract high-semantical information from them became prevalent. As we will see in Chapter 2, the same trends with the approaches applied on two dimensional images, are followed on the higher dimensions.

The vast amount of research done in extracting high level information from a data source so similar to CFD simulation output motivates us to explore the adaptation of the core ideas to the high dimensionality of CFD simulation output.

1.2 Research questions

In this thesis we focus on high dimensional computer vision (i.e. higher than the two dimensional image), and more specifically on extracting meaningful features from CFD simulations output. To achieve our goal, we focus on the following research questions:

RQ1: How are computer vision approaches being generalized to deal with higher dimensionality problems?

Computer vision methodologies are not applied only on the traditional two dimensional image. A lot of research areas and applications concentrate on higher than two dimensional data, such as video processing or RGB-D images. These areas can may seem to be disconnected from areas focusing on two dimensional data. We want to investigate, to what extent methodologies are extended from the two dimensional case to the higher dimensions, as well as common practices and pitfalls these higher dimensionality methodologies share.

RQ2: Can deep learning techniques represent flow fields, in a meaningful manner?

Deep learning techniques are the focus of modern computer vision. As the dimensionality of the datasets increases, the complexity of the patterns needed to be identified is increasing as well. From our previous research question we already got a glimpse of how deep learning approaches are able to handle the increase of the number of dimensions. Nonetheless, the output of CFD simulations has one of the highest dimensionalities we have seen so far. Our question then becomes, how can we best apply deep learning approaches in such complicated data towards maximizing performance?

RQ3: Can local feature based approaches represent flow fields in a meaningful way?

Deep learning approaches, although very powerful, require a vast amount of data to be trained on. Most hand crafted local feature based approaches though were proposed and evolved before the modern huge datasets were available. One of the main differences to deep learning is that the low level features, or encodings are not learned from the data as with deep learning, but are predefined by scientists. Thus, in theory, one only requires enough data to learn high level correlations. Meanwhile, CFD simulation complexity can vary. As the simulation complexity increases, the amount of time required to produce the simulation output increases as well. In some applications, like the flow in a cylinder of a internal combustion engine, it can even take a month to compute on high core count clusters. Thus, acquiring many examples of such high complexity simulations to train deep neural networks is infeasible. Therefore, we want to investigate whether the, more traditional, hand crafted features are capable of representing the flow fields in a meaningful manner and how they compare to deep learning approaches, especially when the number of training data is limited.

RQ4: Can we take advantage of the vector field representation to construct a more efficient convolutional operator?

A large part of the CFD simulation output is the velocity vector field. Convolutional neural networks perform scalar convolutions regardless of the input. Can we take advantage of the fact that the input to the convolution is a vector field? What extra information can be extracted? Can we utilize any extra information in a deep learning framework?

RQ5: How can we better regularize deep neural networks, to reduce overfitting and increase the convergence speed?

There is a variety of limiting factors on the performance of artificial neural networks. Many of them are related to optimization inefficiencies. Some examples are the covariant shift, the exploding and vanishing gradients as well as the scaling-based weight space symmetry. All existing approaches have their limitations. Usually, while trying to solve an issue we are introducing another. For example, by applying orthogonalization, the learning capacity of each layer is limited. Therefore we are interested in investigating whether we can efficiently regularize the weight learning such that performance is maximized.

1.3 Dissertation outline

This dissertation is structured according to the research questions defined in Section 1.2. In Chapter 2 a wide literature study is conducted on how high dimensional data is used in computer vision. Moreover, the approaches are clustered according to (i) the data they are applied to, (ii) whether they are deep learning approaches or traditional hand crafted feature based approaches and (iii) what kind of increase of dimensionality they are tackling, i.e. increase of physical dimensions or increase of amount of information per physical point. Finally, we identify the most popular datasets and benchmarks concentrating on higher than two dimensional data and describe the most studied research areas and discuss the respective state of the art approaches.

In Chapter 3 we construct a large scale 3D CFD simulation dataset, which focuses on the air flow around a passenger car. Using this dataset as a benchmark, a number of deep learning approaches, tailored to the specific high dimensional data are proposed and evaluated, tackling RQ2. Chapter 4 then tackles RQ3, evaluating hand crafted based approaches and comparing them to deep learning approaches. Since in computer vision there is a much larger variety and more generic two dimensional feature detectors and descriptors, whilst two dimensional data require less computational time, we decided to first evaluate using two dimensional data and potentially move to three. Thus, we constructed another dataset which consists of 2D flows of air around an airfoil.

In Chapter 5 we investigate whether its possible to take advantage of vector field representation to gain more information than the response of scalar convolution. We define a new operator that takes advantage of the vector field representation and show that its applicable to more standard computer vision problems as well. In Chapter 6 we proposed a new weight regularization method that tackles some of the issues mentioned in RQ5 and test it on popular benchmarks and architectures. Finally, in Chapter 7 the conclusions of this dissertation are presented and potential future directions discussed.

1.4 Dissertation contributions

The main contributions of the author of this dissertation are the following:

Theodoros Georgiou, Yu Liu, Wei Chen, and Michael Lew. A survey of traditional and deep learning-based feature descriptors for high dimensional data in computer vision. *International Journal of Multimedia Information Retrieval (IJMIR)*, pages 1–36, 2019

Theodoros Georgiou, Sebastian Schmitt, Markus Olhofer, Yu Liu, Thomas Bäck, and Michael Lew. Learning fluid flows. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018

Theodoros Georgiou, Sebastian Schmitt, Nan Pu, Wei Chen, Thomas Bäck, and Michael Lew. Comparison of deep learning and hand crafted features for mining simulation data. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*. IEEE

Theodoros Georgiou, Sebastian Schmitt, Thomas Bäck, and Michael Lew. Orientational equivariant neural networks using clifford convolutions
(Submitted for publication at Neurocomputing, Elsevier)

Theodoros Georgiou, Sebastian Schmitt, Wei Chen, Thomas Bäck, and Michael Lew. Norm loss: An efficient yet effective regularization method for deep neural networks. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*. IEEE

1.5 Other work by the author

Yu Liu, Yanming Guo, **Theodoros Georgiou**, and Michael Lew. Fusion that matters: convolutional fusion networks for visual recognition. *Multimedia Tools and Applications*, 77:1–28, 2018

Umut Özaydın, **Theodoros Georgiou**, and Michael Lew. A comparison of cnn and classic features for image retrieval. In *International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–4, 2019

Yanming Guo, Yu Liu, **Theodoros Georgiou**, and Michael Lew. A review of semantic segmentation using deep neural networks. *International Journal of Multimedia Information Retrieval (IJMIR)*, 7:87–93, 2018

Nan Pu, **Theodoros Georgiou**, Erwin M Bakker, and Michael Lew. Learning a domain-invariant embedding for unsupervised person re-identification. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019