Form 3

Abstract of Thesis

Name (Marcus Walldén)	
Title	Accelerating On-the-Fly Visualization and Analysis of Large-Scale Scientific Simulations (大規模科学シミュレーションのオンザフライ可視化および分析の高速化)

Data generated by large-scale scientific simulations is expected to increase by orders of magnitude in the future as we approach exascale computing. Input/output constraints of supercomputers have increased the use of *co-processing* approaches, i.e., visualizing and analyzing scientific simulations on the fly. Co-processing tasks consume valuable simulation time, thus affecting the fidelity and scale of phenomena that can be simulated in a given time frame. As the complexity and scale of both simulations and co-processing tasks are bound to increase in the future, there is a need for new techniques to accelerate challenging co-processing tasks.

In this dissertation, we identify and address three challenges facing the large-scale co-processing of simulation data. The first challenge is how to expeditiously determine what constitutes important data. Analysis and visualization tasks can be focused on the most essential data, thus accelerating the co-processing. We present a method that evaluates the importance of different regions of simulation data and a data-driven approach that uses this information to accelerate the in-transit co-processing of large-scale simulations. We use the importance metrics to simultaneously employ multiple compression methods on different data regions to accelerate the in-transit co-processing. Our approach strives to adaptively compress data on the fly and uses load balancing to counteract memory imbalances. We demonstrate the method's efficiency through a fluid mechanics application, a Richtmyer-Meshkov instability simulation, showing how to accelerate the in-transit co-processing of simulations. The results show that the proposed method can identify regions of interest expeditiously, even when using multiple metrics. Our approach achieved a speedup of 1.29× in a lossless scenario. The data decompression time was sped up by 2× compared to using a single compression method uniformly.

The second challenge concerns load balancing simulation data. In large computing clusters, dynamically load balancing data can lead to significant inter-process memory imbalances; as a result, data is typically statically distributed among processes. We propose a novel compositing pipeline and a dynamic load balancing technique for volume rendering that utilizes a two-layered structure to achieve effective and scalable load balancing. The technique enables each process to render data from non-contiguous regions of the volume with minimal impact on the total rendering time. We demonstrate the effectiveness of the proposed technique by performing a set of experiments on a computing cluster. The experiments show that using the technique results in up to a 35.7% lower worst-case memory usage as compared to a dynamic *k*-d tree load balancing technique, whilst simultaneously achieving similar or lower rendering times. The proposed technique was also able to lower the amount of transferred data during the load balancing stage by up to 72.2%. The technique has the potential to be used in many scenarios where other dynamic load balancing techniques have proved to be inadequate, such as in large-scale visualization.

The third challenge relates to image batch visualization. In many cases, rather than saving large quantities of simulation data, thousands to millions of images of different variables and viewpoints can be rendered on the fly and saved to permanent storage. This image generation process can be accelerated by rendering and compositing images in batches. Specifically, images can be combined into larger *multi-images*, which results in less synchronization and communication overhead during the image

compositing stage. We present a technique to accelerate such batch processing, called dynamic image resolutions. The dynamic image resolutions technique maps regions of blank pixels in each image and uses this information to dynamically restructure the multi-images to reduce the total image size with no loss of detail. An evaluation of the technique demonstrates a 2.02× speedup of the compositing stage as compared to traditional image compositing and a 1.82× speedup compared to existing multi-image techniques.

Our work successfully addresses three important challenges facing on-the-fly coprocessing of large-scale simulation data. The first application shows how essential data can be identified, and how this information can be used to accelerate co-processing tasks of large-scale data sets. The second application significantly reduces memory imbalances resulting from dynamic load balancing, thus making load balancing feasible in large-scale environments. The third application shows that batch image visualization can be sped up significantly by taking advantage of the underlying image data.

様式７

論文審査の結果の要旨及び担当者

氏　名　（ Marcus Walldén ）		
論文審査担当者	（職）	氏　名
	主査　教授	伊野　文彦
	副査　教授	井上　克郎
	副査　教授	増澤　利光

論文審査の結果の要旨

本学位論文は，並列計算機上で大規模科学シミュレーションを可視化および分析するための高速化手法をまとめたものであり，シミュレーションと可視化を同時に進めるオンザフライ処理方式に関して，以下の３つの研究成果を得ている．

１．計算領域の重要性を迅速に評価する手法の提案
学位論文の２章では，オンザフライ処理方式の一つであるイントランジット型処理において，計算領域の重要性を迅速に評価するための手法を提案している．さらに，提案手法をデータ圧縮手法に適用し，複数のデータ圧縮手法を低いオーバヘッドで併用することで，高速なオンザフライ処理方式を実現する．提案手法を，リヒトマイヤー・メシュコフ不安定性を調べる流体シミュレーションに適用した結果，複数の指標に基づいて重要な計算領域を迅速に特定できただけでなく，無圧縮の場合と比べて1.29倍の高速化を得た．また，単一の圧縮手法と比べて，圧縮時間を半減できた．

２．並列計算機においてメモリ使用量を均一化する動的負荷分散手法の提案
３章では，シミュレーションデータの動的負荷分散手法を提案している．提案手法は，不連続なシミュレーション領域を計算ノードに割り当てながらも，通信パターンを簡略化できるグループ分けを実現することにより，ノードあたりのメモリ使用量に関して均一化を図る．既存手法の一つであるkd木と比べ，提案手法は同様の可視化時間を実現しながらメモリ使用量を35.7%削減できた．また，負荷分散に伴う通信量を72.2%削減できることを示した．

３．バッチ実行による高速な並列可視化手法の提案
４章では，並列可視化システムのための高速化手法を提案している．シミュレーション結果を理解するための可視化は，着目すべき視点やデータを変えながら数百万枚もの画像を生成することが多い．そこで，複数の視点あるいはデータをまとめて処理（バッチ実行）することで，少ない同期回数や通信量を実現する並列可視化手法を提案した．提案手法は，複数の画像を一枚のマルチ画像として可視化する．さらに，画質を損なうことなく画像の解像度を動的に削減し，各計算機上で独立に生成されるマルチ画像を一枚の最終画像に合成する．単一画像に対する既存手法と比較して，提案手法は2.02倍の高速化を達成した．また，既存の1.82倍の高速化を達成した．

以上のように，本学位論文で得られた研究成果は，多次元データに基づく大規模シミュレーション結果を迅速に理解する際に役立ち，シミュレーションの大規模化に寄与する．よって，本論文は博士（情報科学）の学位論文として価値のあるものと認める．