REVIEW REPORT

for the doctoral thesis of Martin Kruliš

"Employing Parallel Architectures in Similarity Search"

submitted to

Charles University in Prague, Faculty of Mathematics and Physics

May 8, 2013

This thesis addresses the issue of the similarity search in the context of large image databases. The author has investigated the use of modern heterogeneous many-core architectures to speed up the similarity search process. Major outcomes of the thesis include: a task scheduling approach, and enablement of efficient computation of Signature Quadratic Form Distance and feature extraction for heterogeneous computing systems that use GPUs as accelerators. Proposed solutions in this thesis have been evaluated empirically using real-world image data sets, such as ALOI (72000 images), CoPhiR (951532 images), Profimedia (17.5 million images) or TWIC (11555 images).

Basically this thesis contributes to better understanding of the programmability of heterogeneous many-core architectures, which currently is one of the major challenges in computing systems research. The task scheduling approach developed in this thesis is generic and may be used to speed up not only the similarity search in image databases but virtually any application that runs on GPU-accelerated computing systems. The proposed scheduling approach provides mechanisms to handle tasks that involve disk I/O or network communication with the aim at improving the overall application performance. The introduced mechanism of feeding threads addresses the issue of data movement between the CPU and GPU with the aim at increasing the GPU utilization.

The thesis is well structured. After a short introduction that highlights the objectives and contributions of this thesis, the author provides an excellent overview of parallel architectures. Thereafter, the major contributions of the thesis are described in Chapters 3 – 5. The thesis concludes with a summary of achievements and proposals for future research. The writing style of the author and the quality of figures are very good. With the thesis is provided a DVD that includes the source code of the software that the author has developed, LaTeX source code of the thesis, image data sets used for performance evaluation and measurements results.

The thesis proves that the author is able to: study the state of the art, identify research problems and propose solutions that go beyond the state of the art, implement proposed solutions and study the obtained results. Key results of this thesis have been presented at conferences (CIKM 2011, MMM 2013) or appeared as a journal publication (Distributed and Parallel Databases, 2012, Springer).

Conclusion: *the author has demonstrated the ability for creative scientific work and the thesis fulfills the criteria for the doctoral thesis in Computer Science.*

Sabri Pllana, PhD
Associate Professor
Linnaeus University
Department of Computer Science
SE-351 95 Växjö, Sweden

E sabri.pllana@lnu.se
T +46 (0) 470 76 7429
F +46 (0) 470 832 17