Distributed Computation Particle Filters on GPU Architectures for Real-Time Control Applications

Mehdi Chitchian, Andrea Simonetto, Alexander S. van Amesfoort, and Tamás Keviczky
a.simonetto@tudelft.nl

Research Question

Particle filters are an extremely effective tool in non-linear/non-Gaussian estimation. A major drawback is the required large number of particles to obtain an acceptable estimation accuracy. GPU architectures make massive parallelization of algorithms possible leading to significant improvements in execution time.

→ Can the use of GPU architectures enable particle filters to deliver accurate estimates to be implemented in high-sample-rate real-time feedback control applications?

Proposed Solution

Particle filters are suitable to be coded on parallel architectures: i.e., each computing core maintains a subset of the whole particle population. The bottleneck is the resampling procedure that involves all-to-all communication among the cores to share all the local particles. This destroys parallelism and performance.

→ Distributed computations: exchange only the $t$ particles with the highest weights with the neighboring cores.

Model, algorithm, and implementation

Discrete-time nonlinear dynamics: $x(k+1) = f(x(k), w(k))$;
Nonlinear measurement equation: $z(k) = h(x(k), \mu(k))$;
A total of $m$ particles on $N$ local filters.

Input: $(x_i(k-1))^j_{j=1...m}, z(k)$
1. Local Filter:
   for $j = 1:m$
   1.1: sampling: $x_i(k)^j \sim p(x(k)|x_i(k-1)^j)$
   1.2: weight calculation: $w_i(k)^j = p(z(k)|x_i(k)^j)$
   end
2. Sorting: sort $\{x_i(k)^j\}^j_{j=1...m}$ according to $\{w_i(k)^j\}^j_{j=1...m}$
3. Estimation: local estimation: pick $x_i(k)^j$ with maximal $w_i(k)^j$
4. Particle Exchange:
   foreach neighbor do
   send and receive $t$ particle-weight couples to/from neighbors
   end
5. Resampling: resample the $m + Nt$ particles into $m$ particles
6. Reset: set $w_i(k)^j = 1/m$ for all $j$
Output: $\{x_i(k)^j\}^j_{j=1...m}$

Conclusions

→ Feedback control based on estimation from particle filters running at 100 Hz with over 1 million particle is possible with current state-of-the-art GPU architecture, with good accuracy;
→ the accuracy is linked to the core topology and the exchanged particles $t$ (we were able to infer some analytical properties);
→ compared to available methods, our distributed solution performs better in terms of accuracy, number of particles, or allowed state dimension, by orders of magnitude.

More details