A New Parallel Algorithm for Two-Pass Connected Component Labeling

***Abstract*—Connected Component Labeling(CCL) is one of the most important step in pattern recognition and image processing. Connected component labeling assigns labels to a pixel such that adjacent pixels sharing the same features are assigned the same label. Typically, CCL requires several passes over the data. For example, in a two-pass technique, the first pass, each pixel is given a provisional label and label equivalence information is stored. In the second pass,an actual label is given to each pixel. Suzuki et al have proposed two algorithms for CCL with two-pass technique called Link by Rank and Path Compression(LRPC), and ARun. The $LRPC$ algorithm uses a decision tree to assign provisional labels and an array-based union-find datastructure to store label equivalence information. The $ARun$ algorithm employs a special scan order over the data and three linear arrays instead of the conventional union-find datastructure. To the best of our knowledge, there has not been any effort yet on parallelizing two-pass CCL for shared memory architecture.**

We present a scalable parallel two-pass CCL algorithm called $PAREMSP$, which employs scan strategy of $ARun$ algorithm and the best union-find technique called $REMSP$ for storing label equivalence information of pixels in a 2-D image. In the first step, we divide the image into chunks of equal size among all threads and each thread run the scan strategy of $ARun$ algorithm along with $REMSP$ simultaneously. As $REMSP$ is easily parallizable, we use the parallel version of $REMSP$ for merging the pixels on the boundary of each chunk after the first step. Our experiment on image of size 11411×10192 pixels show the scalability of $PAREMSP$ achieving speedups up to 19.5 using 24 cores on shared memory architecture. Also, the parallel algorithm does not make use of any hardware specific routines, and thus is highly portable