Let $S = \{x_i, y_i\}$ for $i = 1, 2, 3, \ldots, n$ be a set of $n$ points $(p_1, p_2, \ldots, p_n)$ in two-dimensional (2D) space, and each point is defined by its color $c[i]$. The color (category) $c[i]$ belongs to $\sigma$ is associated with every point in 2D. An orthogonal range reporting query $Q = [a, b] \times [c, d]$ on set $S$ is sent to report the set of points in $S$ which interact with the query rectangle. The reported output is a set of $(x, y)$ in $S$ where $x$ belongs to $[a, b]$ and $y$ belongs to $[c, d]$ of points and this query time is dependent on the output size. A wavelet tree data structure over $S$ can answer such queries in $O(\log n)$ time per output-point and its space occupancy is $O(n)$ words. Wavelet tree data structure can be considered as a grid of points, reordering or sequence. Wavelet tree grid is space adjusting and number of queries can be applied on it. The categorical range reporting query applied is an extension of orthogonal range reporting, where each point $(x_i, y_i)$ in $S$ is associated with a color $c[i]$ belongs to $\sigma$ and the task is to report the set of distinct colors within the query region $[a, b] \times [c, d]$. The well-known data structures consume $O(n \log n)$ space and $O(\log n + k)$ query time, where $k$ is the output size. In this paper, we present a new solution for this problem using wavelet trees. Our structure takes $O(n \log \sigma)$ space and query time is $O(k \log n \log \sigma)$. Notice that the new result is more efficient in space when $\log \sigma = O(\log n)$.

Major: Computer Engineering

Educational Career:
Bachelor's of Electronics and Communication Engineering, BS, 2015, Visvesvaraya Technological University

Committee in Charge:
Sharma Thankachan, Chair, Computer Science
Kalpathy Sundaram, EECS
Sumit Kumar Jha, Computer Science

Approved for distribution by Sharma Thankachan, Committee Chair, on May 17, 2018.

The public is welcome to attend.