1 Introduction

2 Mathematical Description

Algorithm 1 k-Means Estimation

Require: k the number of clusters
Require: S a set of samples arranged in $m \times n$ matrix
  where m is the number of attributes,
  n is the number of samples.
  Set means matrix Q such that Q is $m \times k$
  Select k columns from S at random and insert them into Q
  repeat
    Create Q' s.t. $Q = 0$ and Q' is $m \times k$.
    Create row vector $\vec{r}$ so that $\vec{r}$ is $1 \times k$
    for each $\vec{s}_j \in S$, where $\vec{s}_j$ is the jth column of S **do**
      Measure distance w/ the columns in Q
      Identify the smallest distance column, t.
      Add $\vec{s}_j$ to $\vec{q'}_t$, and insert into Q' at column t
      Increments r_t
    end for
    Scale divide: Q' with $\vec{r}$
    $\epsilon = |Q - Q'|$
    $Q \leftarrow Q'$
  until ϵ is satisfied
Ensure: Q

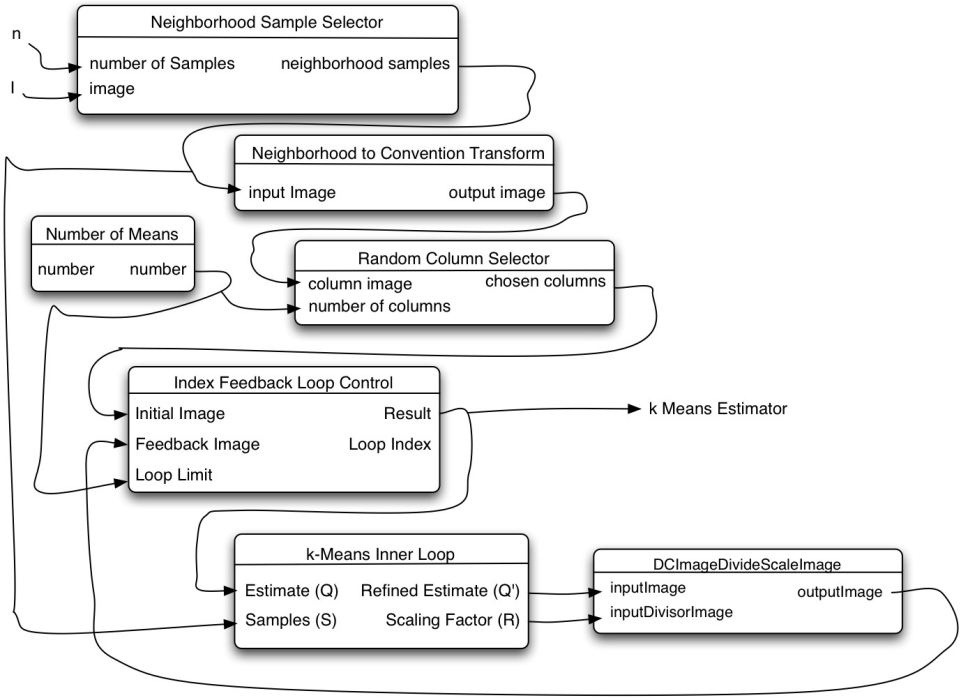

Figure 1: k-Means Outer Loop

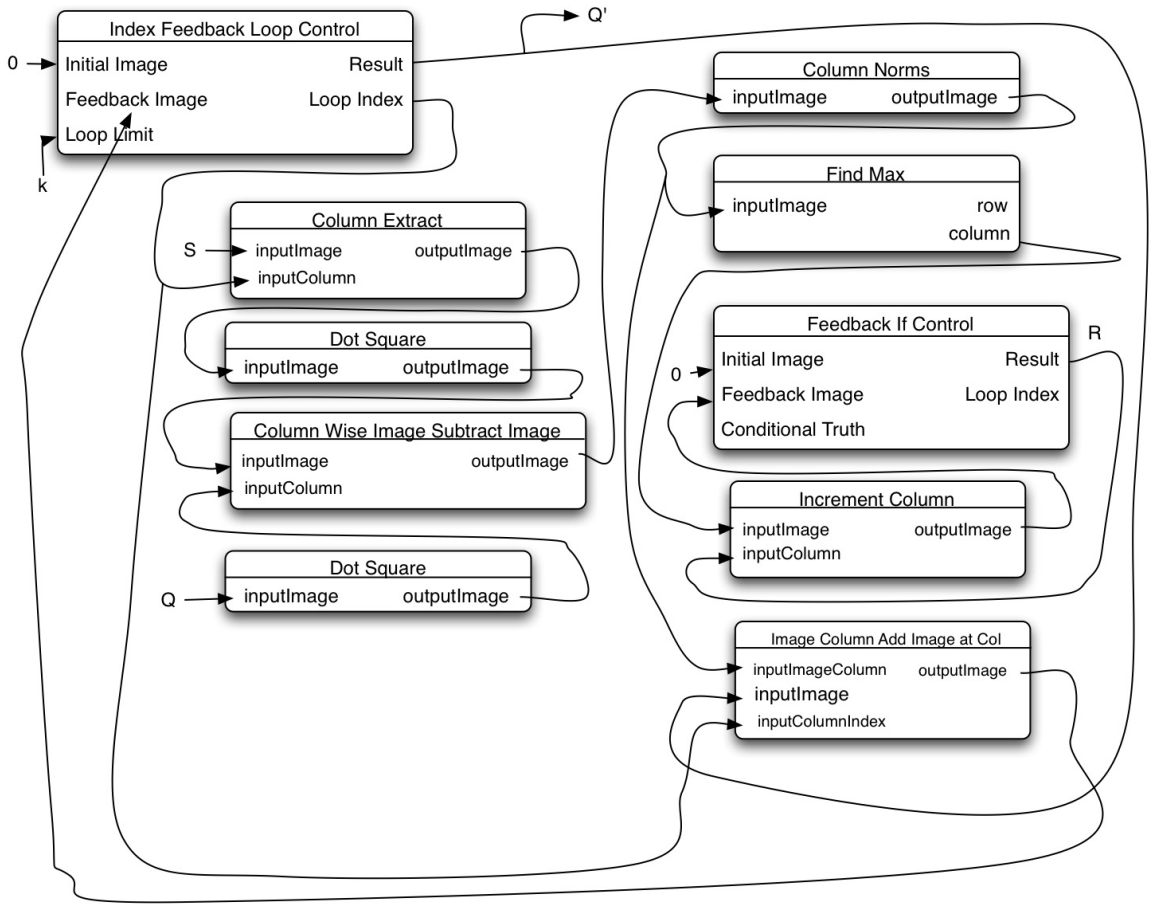

Figure 2: k-Means Inner Loop