Data cleaning:

Stage 1. Cleaning for raw data:

(1) Removing data files that don’t contain any mouse data or eye data or only contain few data points (or few data lines). If a mouse data file is deleted, the corresponding eye data file will also be deleted, vice versa. (Implemented in C# program DataFileChecking)

Stage 2. Cleaning for aligned data:

(1) Remove lines with invalid value. (e.g. 28.2.31) (in program dataAlign, function dataValidCheck(…))

(2) Remove lines with negative, zero or empty value. (in program dataAlign, function dataValidCheck(…))

(3) Removing data files that don’t contain any mouse data or eye data or only contain few mouse or eye data points (or few data lines) before start aligning mouse data and eye data. (In program dataAlign, function GenerateFile (…))

Stage 3. Cleaning for extracting features

(1) In the movement path of a data file, after eye arrived end button, ignore the rest of eye data points.

**Algorithm**: check from the beginning of moving path, find the first data point that is inside the ending range. The range is decide by the position of ending button, which is a constant value that obtained from the experiment machine. After that, remove the data points that obtained after this data point.

(2) In the movement path of a data file, after mouse arrived end button, ignore the rest of mouse data points.

**Algorithm**: The same with the above one.

(3) Removing aligned data files that don’t contain any mouse data or eye data or only contain few mouse or eye data points (or few data lines). (Implemented in program FeatureExtract)

(4) Removing aligned data files that having a very short moving path; all data points cluster in a small area. (Implemented in program FeatureExtract)

**Algorithm**: Check the moving distance of eye and mouse, if any one of them is less than the threshold, remove this data point. To obtain moving distance, the program first obtained start point and end point of eye path and mouse path, then calculate distance between start point and end point for eye and mouse.

Notes:

Removing data files that don’t contain any data or only contain few data (either mouse or eye data) is included in those three different stages of data cleaning.

In first stage, the raw data file may contain few data points or less, so these data files need to be removed.

In the second stage, at the beginning, the program need to remove invalid value or negative values, this may remove all data or most of data in a file (in case of most of data is bad). This may result in empty or few-data files, so these file need to be removed before doing alignment.

In the third stage, the first step may remove some data points that locate behind the ending point (a time point) in a data file, this may result the data file become an empty or few-data file. So the removal also need to be done in this stage.