Document Images Analysis Recognition And Retrieval (DIARR) has never been a more exciting technical field. After many years of theoretical foundation and the gradual establishment of some core technical products such as machine print OCR systems, a much broader range of applications of our research are now emerging. Pen-based PDAs and tablet PCs; digital libraries and archive conversion; business document databases; web documents, text recognition in natural scenes; online and offline handwriting recognition products; and biometrics, are all rapidly developing and beginning to find widespread application.

The framework of DIARR:

My Works:

I focus largely on the online document images layout analysis and feature extraction, and then retrieve the document image based on these features.

The Result of DLA:

The document image retrieval system: