

**CHENNAI – PONDICHERRY**

**A Unified Framework for Tracking Based Text Detection and Recognition from Web Videos**

**Abstract**

Video text extraction plays an important role for multimedia understanding and retrieval. Most previous research efforts are conducted within individual frames. A few of recent methods, which pay attention to text tracking using multiple frames, however, do not effectively mine the relations among text detection, tracking and recognition. In this paper, we propose a generic Bayesian-based framework of Tracking based Text Detection And Recognition (T2DAR) from web videos for embedded captions, which is composed of three major components, i.e., text tracking, tracking based text detection, and tracking based text recognition. In this unified framework, text tracking is first conducted by tracking-by-detection. Tracking trajectories are then revised and refined with detection or recognition results. Text detection or recognition is finally improved with multi-frame integration. Moreover, a challenging video text (embedded caption text) database (USTB-VidTEXT) is constructed and publicly available. A variety of experiments on this dataset verify that our proposed approach largely improves the performance of text detection and recognition from web videos.