

This research strives for natural language moment retrieval in long, untrimmed video streams. The problem is not trivial especially when a video contains multiple