VFR Technical Report 2026-03 · Dataset paper · 19 May 2026

HETKI: A Multilingual Benchmark for Natural-Language Moment Retrieval in Licensed Stock Video

Mei Takahashi, Lotta Haapanen, Aino Lehtovaara, Jussi Pelkonen
VideoFinder Research — Helsinki · Kadrowa Labs Nordics Oy, 00180 Helsinki

Abstract

We present HETKI, a benchmark of 40,212 natural-language queries grounded to time ranges in 9,346 licensed clips, written by professional editors in nine languages including Finnish, Polish, Japanese and Arabic. Unlike benchmarks built from movie captions or instructional videos, HETKI queries describe visual properties that editors search for — motion, colour, composition and mood — and every query carries boundary annotations from three independent annotators. We release the annotations under CC BY 4.0, an evaluation toolkit, and a soft-IoU metric that accounts for annotator disagreement. Baselines show that current video–language models handle object-centric queries well but fall 20 points behind on motion and mood.

About this document

This is the public extended abstract. The full report, including appendices, ablations and evaluation scripts, is available on request from research@videofinder.link. Code and data referenced in this work are listed at https://helsinki.videofinder.link/open-source.html and /datasets.html.

Cite as

```
@techreport{takahashi2026hetki,
  title = {HETKI: A Multilingual Benchmark for Natural-Language Moment
  Retrieval in Licensed Stock Video},
  institution = {VideoFinder Research, Kadrowa Labs Nordics Oy},
  year = {2026}, number = {VFR-TR-2026-03}
}
```