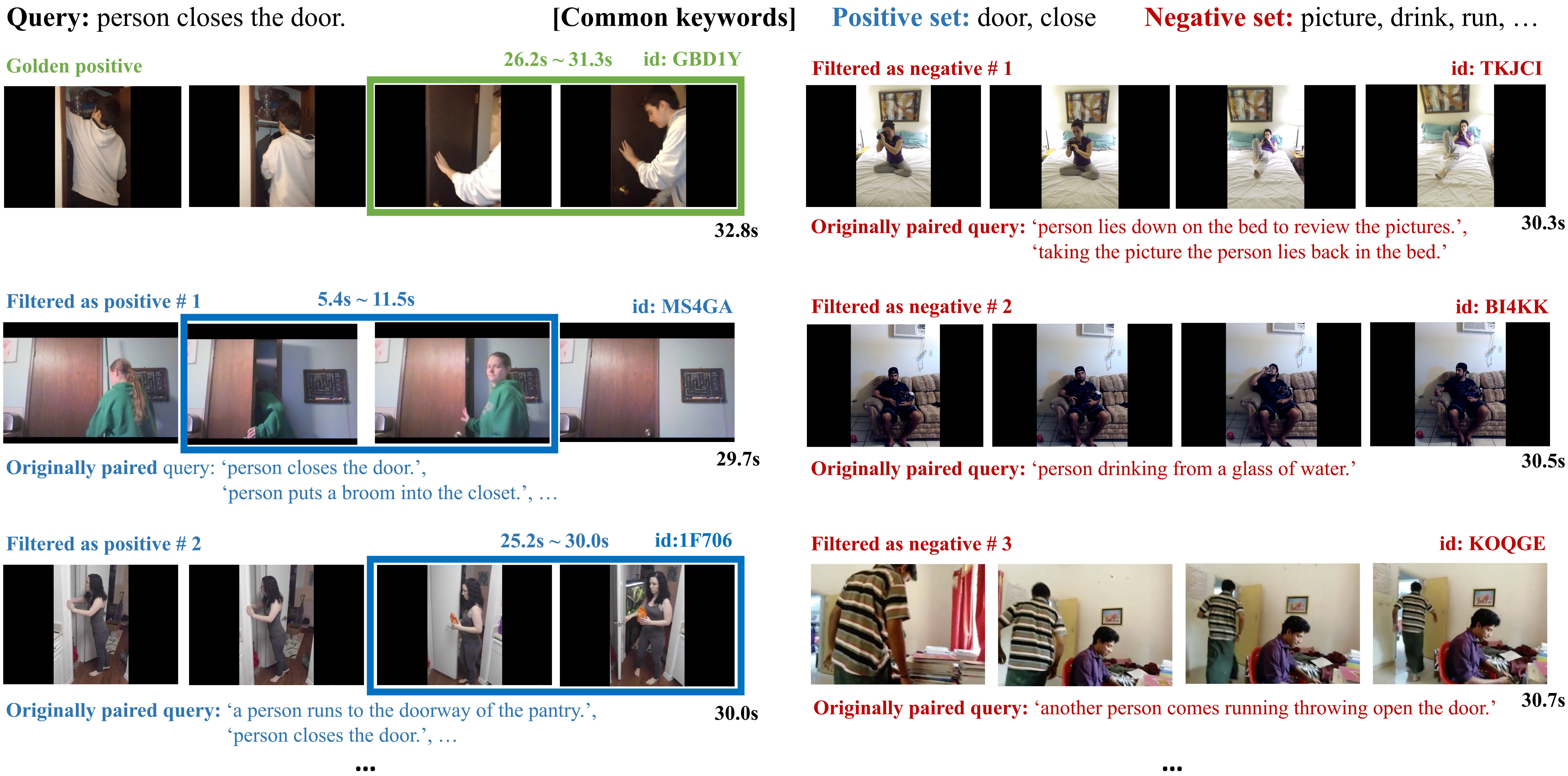
Query: person closes the door.
[Common keywords]
Positive set: door, close
Negative set: picture, drink, run, …
Golden positive
26.2s ~ 31.3s
id: GBD1Y
32.8s
Filtered as positive # 1
5.4s ~ 11.5s
id: MS4GA
29.7s
Originally paired query: 'person closes the door.', 'person puts a broom into the closet.', …
Filtered as positive # 2
25.2s ~ 30.0s
id:1F706
30.0s
Originally paired query: 'a person runs to the doorway of the pantry.', 'person closes the door.', …
…
Filtered as negative # 1
id: TKJCI
30.3s
Originally paired query: 'person lies down on the bed to review the pictures.', 'taking the picture the person lies back in the bed.'
Filtered as negative # 2
id: BI4KK
30.5s
Originally paired query: 'person drinking from a glass of water.'
Filtered as negative # 3
id: KOQGE
30.7s
Originally paired query: 'another person comes running throwing open the door.'
…

Figure. Video-query Alignment Qualitative Analysis for MVMR Charades-STA dataset. We conducted a qualitative case analysis on the MVMR Charades-STA dataset we constructed. In the figure, 'Golden positive' cases are highlighted in green, 'Positive filtered' in blue, and 'Negative filtered' in red. We marked the matching moments (label) for the filtered positive videos with a solid line. From this analysis, we observed that videos categorized as 'Positive' showed high coherence and relevance to the target query. Conversely, those classified as 'Negative' lacked relation to the target query and had no significant coherence. This distinction was further substantiated by analyzing the top count keywords for each query set within the positive and negative clusters.