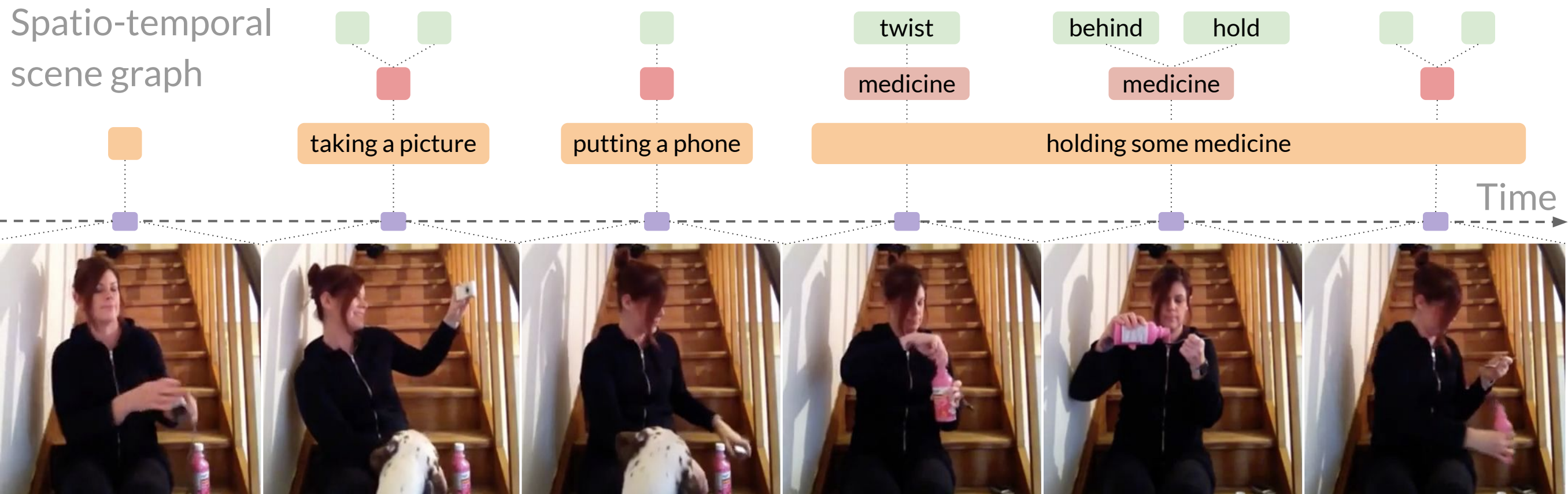


Spatio-temporal scene graph



Compositional spatio-temporal question answers

Q: What did the person **hold** after **putting a phone** somewhere?

A: **medicine**

Q: Were they **taking a picture** or **holding some medicine** for longer?

A: **holding** some **medicine**

Q: Did they **take a picture** before or after they did the **longest action**?

A: **before**

Q: Did the person **go behind** the **object** they **twisted** before **taking a picture**?

A: no

Legend: ■ objects ■ relationships ■ actions ■ time