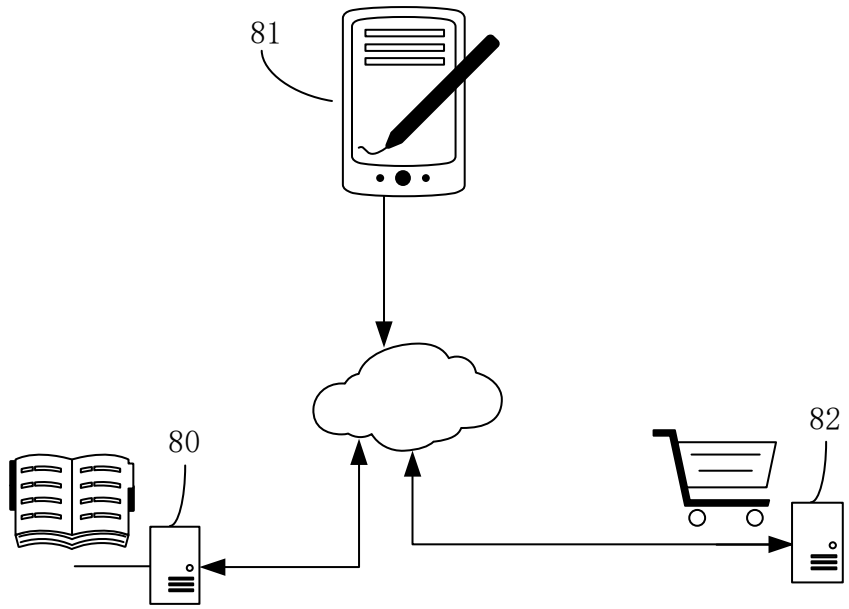
81
80
82

FIG. 1

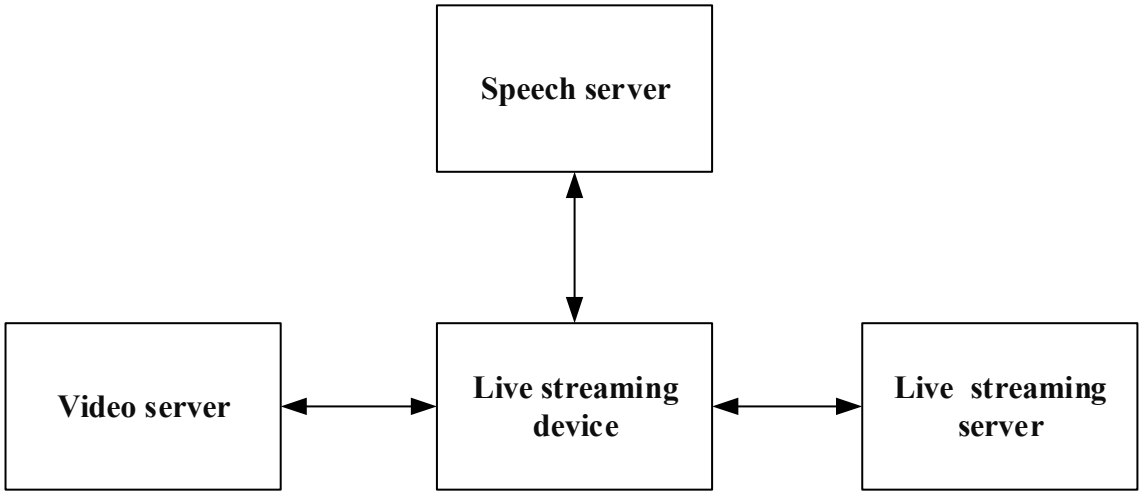
Speech server
Video server
Live streaming device
Live streaming server

FIG. 2

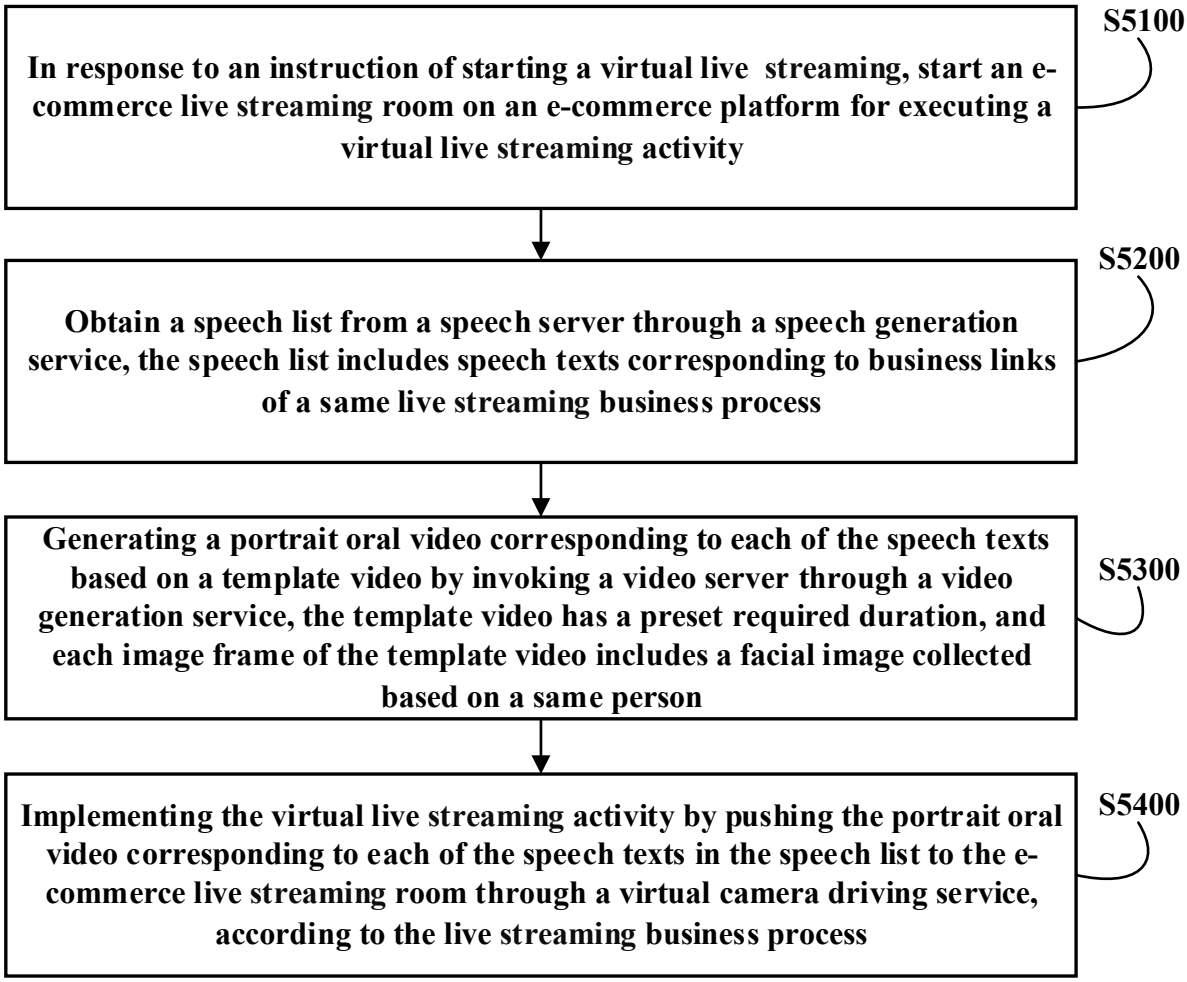
S5100
In response to an instruction of starting a virtual live streaming, start an e-commerce live streaming room on an e-commerce platform for executing a virtual live streaming activity
S5200
Obtain a speech list from a speech server through a speech generation service, the speech list includes speech texts corresponding to business links of a same live streaming business process
S5300
Generating a portrait oral video corresponding to each of the speech texts based on a template video by invoking a video server through a video generation service, the template video has a preset required duration, and each image frame of the template video includes a facial image collected based on a same person
S5400
Implementing the virtual live streaming activity by pushing the portrait oral video corresponding to each of the speech texts in the speech list to the e-commerce live streaming room through a virtual camera driving service, according to the live streaming business process

FIG. 3

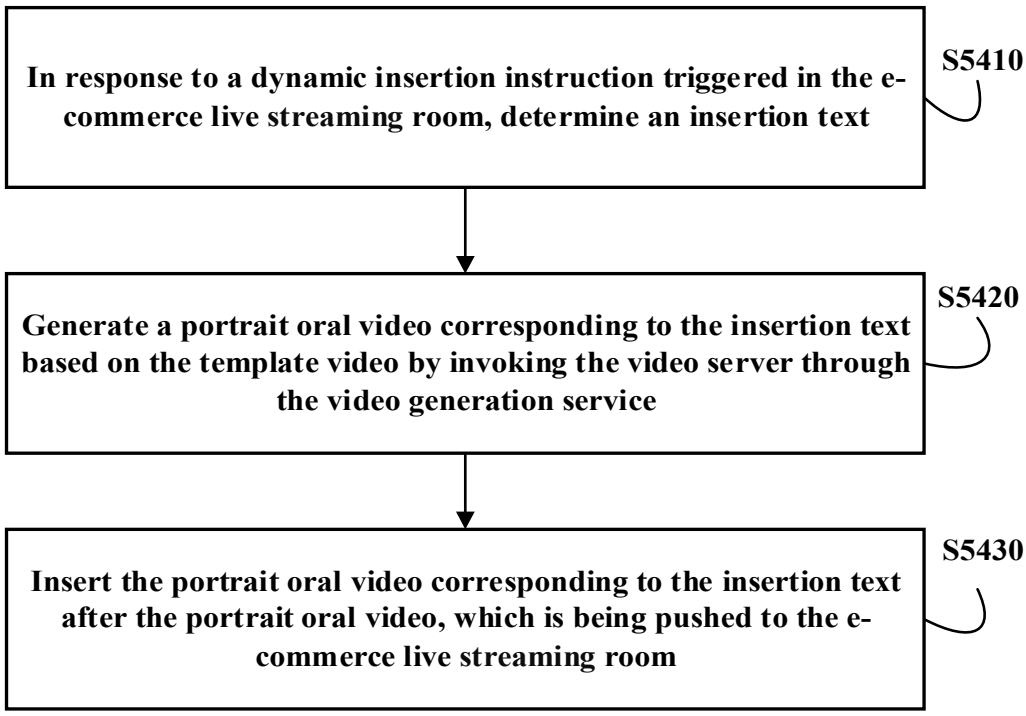
S5410
In response to a dynamic insertion instruction triggered in the e-commerce live streaming room, determine an insertion text
S5420
Generate a portrait oral video corresponding to the insertion text based on the template video by invoking the video server through the video generation service
S5430
Insert the portrait oral video corresponding to the insertion text after the portrait oral video, which is being pushed to the e-commerce live streaming room

FIG. 4

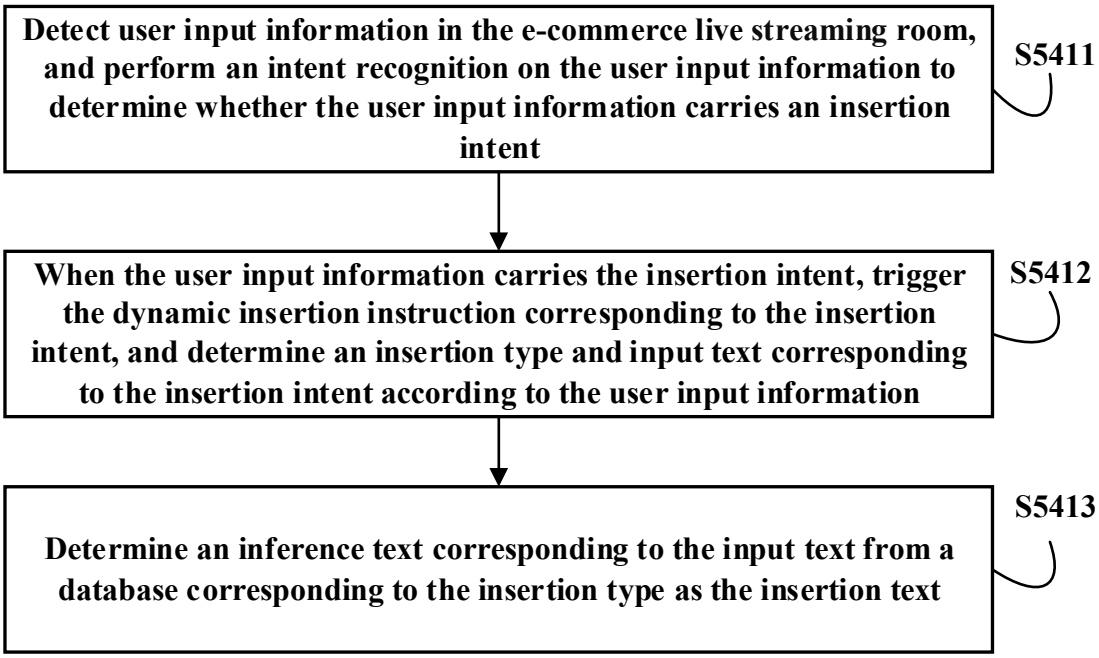
Detect user input information in the e-commerce live streaming room, and perform an intent recognition on the user input information to determine whether the user input information carries an insertion intent
S5411
When the user input information carries the insertion intent, trigger the dynamic insertion instruction corresponding to the insertion intent, and determine an insertion type and input text corresponding to the insertion intent according to the user input information
S5412
Determine an inference text corresponding to the input text from a database corresponding to the insertion type as the insertion text
S5413

FIG. 5

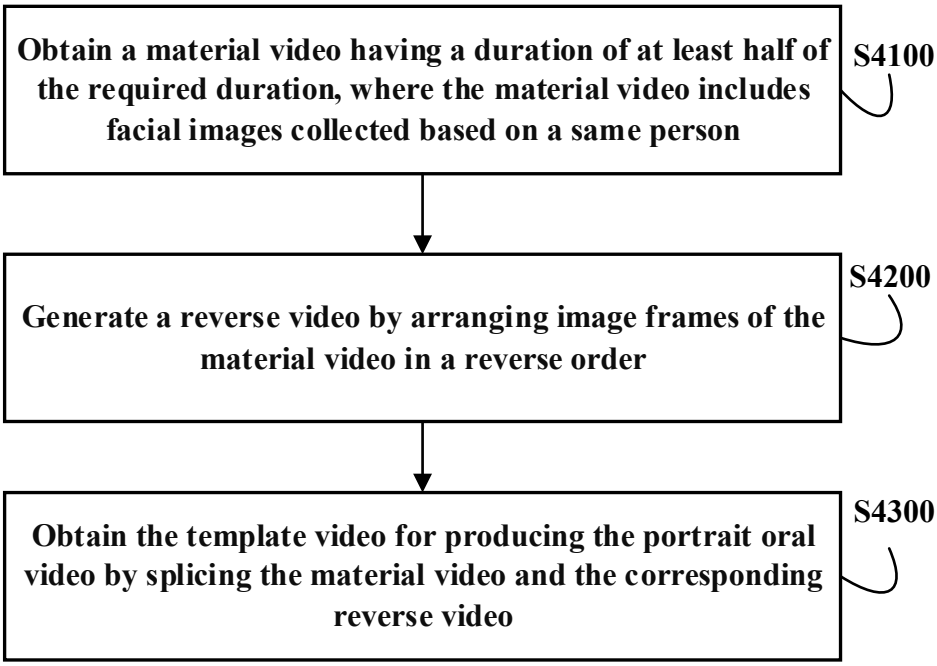
Obtain a material video having a duration of at least half of the required duration, where the material video includes facial images collected based on a same person
S4100
Generate a reverse video by arranging image frames of the material video in a reverse order
S4200
Obtain the template video for producing the portrait oral video by splicing the material video and the corresponding reverse video
S4300

FIG. 6

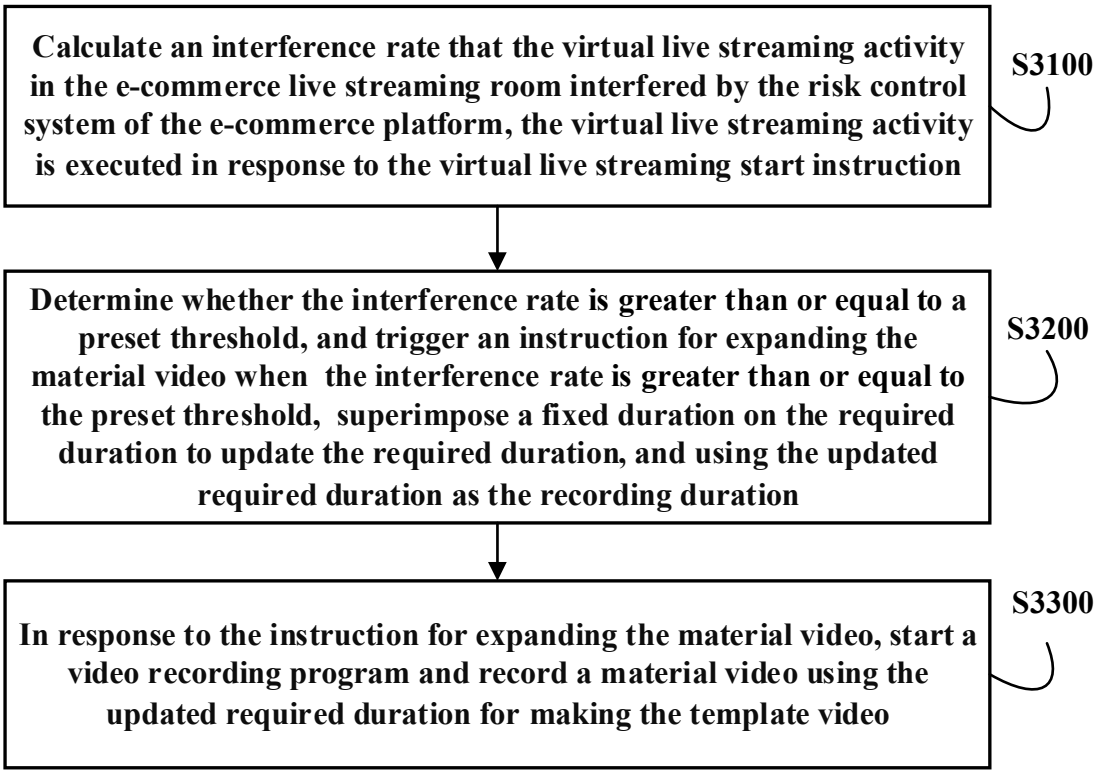
Calculate an interference rate that the virtual live streaming activity in the e-commerce live streaming room interfered by the risk control system of the e-commerce platform, the virtual live streaming activity is executed in response to the virtual live streaming start instruction
S3100
Determine whether the interference rate is greater than or equal to a preset threshold, and trigger an instruction for expanding the material video when the interference rate is greater than or equal to the preset threshold, superimpose a fixed duration on the required duration to update the required duration, and using the updated required duration as the recording duration
S3200
In response to the instruction for expanding the material video, start a video recording program and record a material video using the updated required duration for making the template video
S3300

FIG. 7

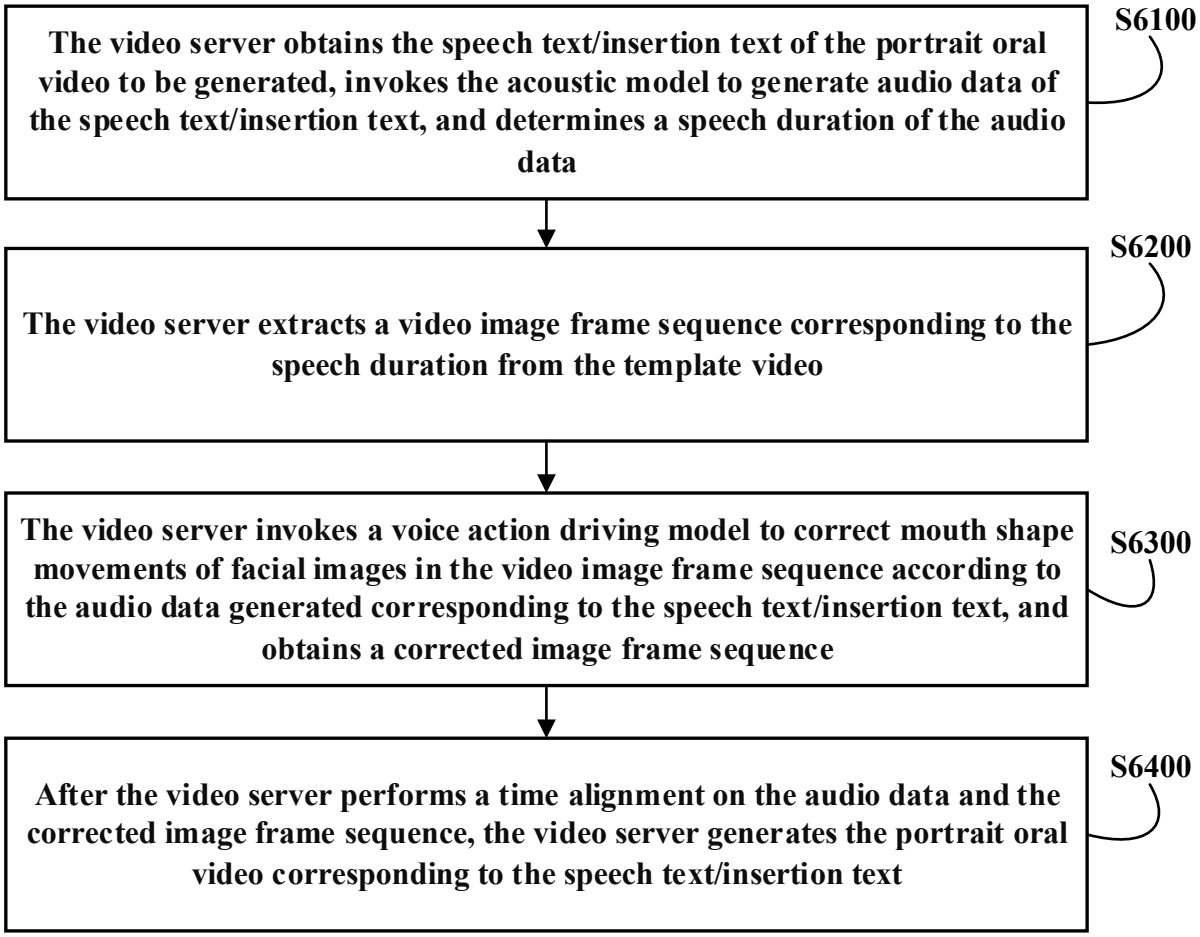
The video server obtains the speech text/insertion text of the portrait oral video to be generated, invokes the acoustic model to generate audio data of the speech text/insertion text, and determines a speech duration of the audio data
S6100
The video server extracts a video image frame sequence corresponding to the speech duration from the template video
S6200
The video server invokes a voice action driving model to correct mouth shape movements of facial images in the video image frame sequence according to the audio data generated corresponding to the speech text/insertion text, and obtains a corrected image frame sequence
S6300
After the video server performs a time alignment on the audio data and the corrected image frame sequence, the video server generates the portrait oral video corresponding to the speech text/insertion text
S6400

FIG. 8

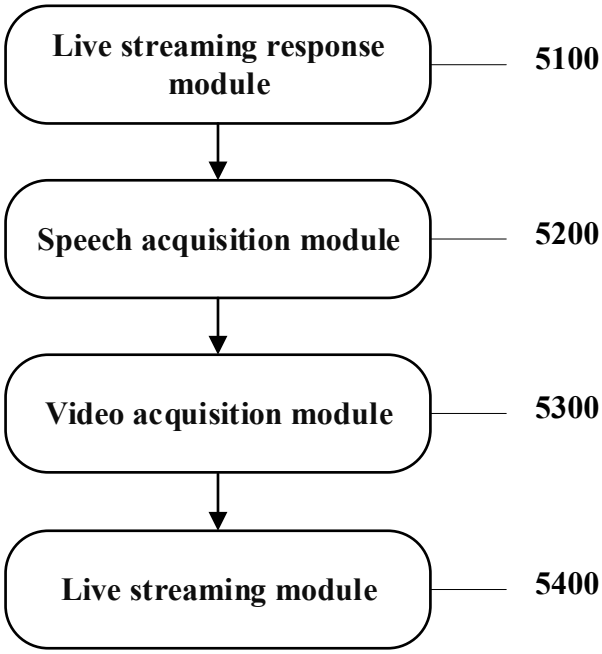
Live streaming response module
5100
Speech acquisition module
5200
Video acquisition module
5300
Live streaming module
5400

FIG. 9

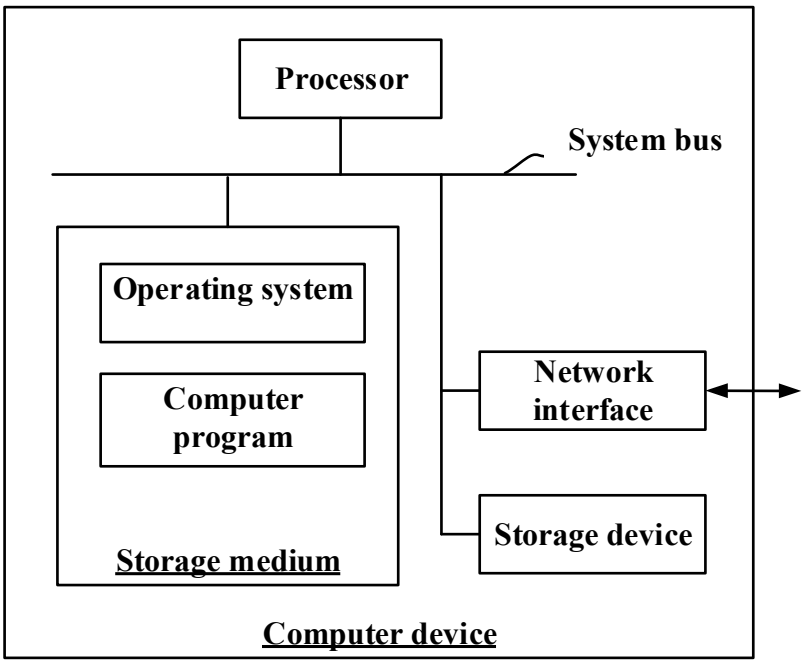
Processor
System bus
Operating system
Computer program
Storage medium
Network interface
Storage device
Computer device

FIG. 10