Multimodal Visual Understanding（Text Version）

Before running the function, you need to close the App and large programs. For the closing method, refer to [4. Preparation] - [1. Manage APP control services].

1. Function Description

After the program runs, you can input visual-related questions or descriptions of the current scene through the terminal. The program will take an image of the current environment and upload it to the Alibaba Cloud platform. The visual model will then analyze the image based on the question and provide an answer.

2. Startup

Jetson-Nano board users need to enter the docker container and then enter the following command. Orin board users can directly open a terminal and enter the following command:

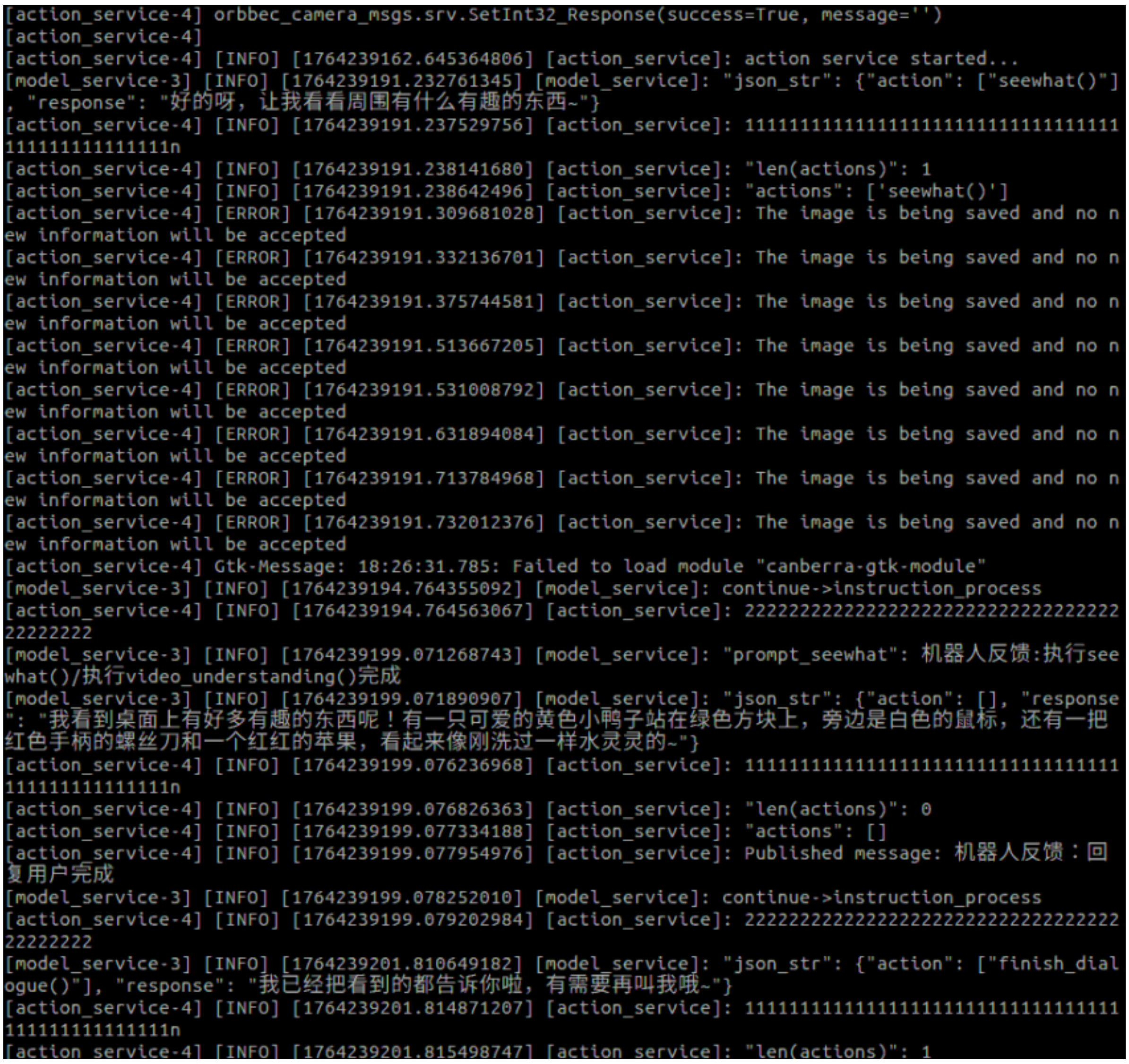
```
ros2 launch largemodel largemodel_control.launch.py text_chat_mode:=True
```

Then open a second terminal and enter the following command:

```
ros2 run text_chat text_chat
```

Then, in the text_chat terminal, enter your visual understanding-related questions. You can refer to the following examples:

Describe the scene you see.



```
[action_service-4] orbbec_camera_msgs.srv.SetInt32_Response(success=True, message='')
[action_service-4]
[action_service-4] [INFO] [1764239162.645364806] [action_service]: action service started...
[model_service-3] [INFO] [1764239191.232761345] [model_service]: "json_str": {"action": ["seewhat()"]
, "response": "好的呀，让我看看周围有什么有趣的东西~"}
[action_service-4] [INFO] [1764239191.237529756] [action_service]: 1111111111111111111111111111111111
111111111111111n
[action_service-4] [INFO] [1764239191.238141680] [action_service]: "len(actions)": 1
[action_service-4] [INFO] [1764239191.238642496] [action_service]: "actions": ['seewhat()']
[action_service-4] [ERROR] [1764239191.309681028] [action_service]: The image is being saved and no n
ew information will be accepted
[action_service-4] [ERROR] [1764239191.332136701] [action_service]: The image is being saved and no n
ew information will be accepted
[action_service-4] [ERROR] [1764239191.375744581] [action_service]: The image is being saved and no n
ew information will be accepted
[action_service-4] [ERROR] [1764239191.513667205] [action_service]: The image is being saved and no n
ew information will be accepted
[action_service-4] [ERROR] [1764239191.531008792] [action_service]: The image is being saved and no n
ew information will be accepted
[action_service-4] [ERROR] [1764239191.631894084] [action_service]: The image is being saved and no n
ew information will be accepted
[action_service-4] [ERROR] [1764239191.713784968] [action_service]: The image is being saved and no n
ew information will be accepted
[action_service-4] [ERROR] [1764239191.732012376] [action_service]: The image is being saved and no n
ew information will be accepted
[action_service-4] Gtk-Message: 18:26:31.785: Failed to load module "canberra-gtk-module"
[model_service-3] [INFO] [1764239194.764355092] [model_service]: continue->instruction_process
[action_service-4] [INFO] [1764239194.764563067] [action_service]: 2222222222222222222222222222222222
22222222
[model_service-3] [INFO] [1764239199.071268743] [model_service]: "prompt_seewhat": 机器人反馈:执行see
what()/执行video_understanding()完成
[model_service-3] [INFO] [1764239199.071890907] [model_service]: "json_str": {"action": [], "response
": "我看到桌面上有好多有趣的东西呢！有一只可爱的黄色小鸭子站在绿色方块上，旁边是白色的鼠标，还有一把
红色手柄的螺丝刀和一个红红的苹果，看起来像刚洗过一样水灵灵的~"}
[action_service-4] [INFO] [1764239199.076236968] [action_service]: 1111111111111111111111111111111111
111111111111111n
[action_service-4] [INFO] [1764239199.076826363] [action_service]: "len(actions)": 0
[action_service-4] [INFO] [1764239199.077334188] [action_service]: "actions": []
[action_service-4] [INFO] [1764239199.077954976] [action_service]: Published message: 机器人反馈：回
复用户完成
[model_service-3] [INFO] [1764239199.078252010] [model_service]: continue->instruction_process
[action_service-4] [INFO] [1764239199.079202984] [action_service]: 2222222222222222222222222222222222
22222222
[model_service-3] [INFO] [1764239201.810649182] [model_service]: "json_str": {"action": ["finish_dial
ogue()"], "response": "我已经把看到的都告诉你啦，有需要再叫我哦~"}
[action_service-4] [INFO] [1764239201.814871207] [action_service]: 1111111111111111111111111111111111
111111111111111n
[action_service-4] [INFO] [1764239201.815498747] [action_service]: "len(actions)": 1
```

The saved photo is located at:

`LargeModel_ws/install/largemodel/share/largemodel/resources_file/image.png`

3. Core Code Analysis: seewhat

The program calls the `seewhat` function to take photos. The source code path of this function is: `LargeModel_ws/src/largemodel/largemodel/action_service.py`

```
def seewhat(self):
    #Save image
    self.save_single_image()
    time.sleep(3.0)
    msg = String(data="seewhat")
    self.seewhat_handle_pub.publish(
        msg
    )  # Normalization, publishing the 'seewhat' topic, and calling the large
language model via model_service.

def save_single_image(self):
    """
    保存一张图片 / Save a single image
    """
    self.IS_SAVING=True
    time.sleep(0.5)
    if self.image_msg is None:
        self.get_logger().warning("No image received yet.")  # Image not yet
received...
        return
    try:
        # 将ROS图像消息转换为OpenCV图像 / Convert ROS image message to OpenCV image
        cv_image = self.bridge.imgmsg_to_cv2(self.image_msg, "bgr8")
        # 保存图片 / Save the image
        cv2.imwrite(self.image_save_path, cv_image)
        display_thread = threading.Thread(target=self.display_saved_image)
        display_thread.start()

    except Exception as e:
        self.get_logger().error(f"Error saving image: {e}")  # Error saving
image...
    self.IS_SAVING=False
```