

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(10) 国际公布号
WO 2026/011984 A1

(43) 国际公布日
2026 年 1 月 15 日 (15.01.2026)

(51) 国际专利分类号:
G06V 20/40 (2022.01)

(21) 国际申请号: PCT/CN2025/096755

(22) 国际申请日: 2025 年 5 月 23 日 (23.05.2025)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202410929056.9 2024 年 7 月 11 日 (11.07.2024) CN

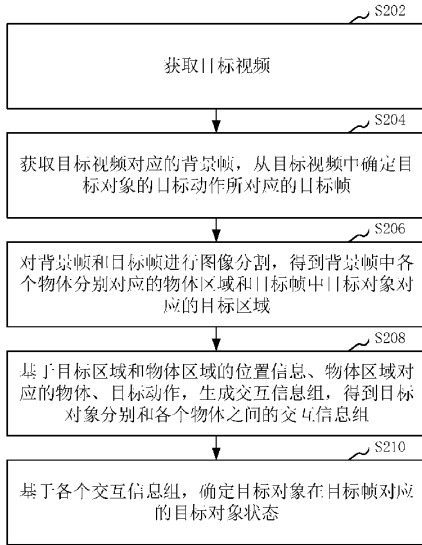
(71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层 518057 (CN)。

(72) 发明人: 黄湘琦(HUANG, Xiangqi); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层 518057 (CN)。高杰(GAO, Jie); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层 518057 (CN)。王楠(WANG, Nan); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层 518057 (CN)。莫小帅(MO, Xiaoshuai); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层 518057 (CN)。杨博(YANG, Bo); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层 518057 (CN)。窦瑞刚(DOU, Ruigang); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层 518057 (CN)。

(74) 代理人: 华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区珠江东路6号4501房

(54) Title: OBJECT STATE RECOGNITION METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 对象状态识别方法、装置、计算机设备和存储介质



S202 Acquire a target video
S204 Acquire a background frame corresponding to the target video, and determine, from the target video, a target frame corresponding to a target action of a target object
S206 Perform image segmentation on the background frame and the target frame to obtain object regions respectively corresponding to objects in the background frame and a target region corresponding to the target object in the target frame
S208 Generate interaction information groups on the basis of position information of the target region and the object regions, the objects corresponding to the object regions, and the target action, so as to obtain an interaction information group between the target object and each object
S210 On the basis of each interaction information group, determine a target object state of the target object corresponding to the target frame

图 2

(57) Abstract: The present application relates to an object state recognition method and apparatus, a computer device, a storage medium, and a computer program product. The method comprises: acquiring a target video (step S202); acquiring a background frame corresponding to the target video, and determining, from the target video, a target frame corresponding to a target action of a target object (step S204); performing image segmentation on the background frame and the target frame to obtain object regions respectively corresponding to objects in the background frame and a target region corresponding to the target object in the target frame (step S206);

[续页]

(部 位: 自 编 01-03 和 08-12 单 元) (仅 限
办公用途) 510623 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家
保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG,
BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,
CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ,
IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ,
LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN,
MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA,
PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,
SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UY, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚
(AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE,
BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR,
HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO,
PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN,
TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

generating interaction information groups on the basis of position information of the target region and the object regions, the objects corresponding to the object regions, and the target action, so as to obtain an interaction information group between the target object and each object (step S208); and on the basis of each interaction information group, determining a target object state of the target object corresponding to the target frame (step S210).

(57) 摘要: 本申请涉及一种对象状态识别方法、装置、计算机设备、存储介质和计算机程序产品。所述方法包括: 获取目标视频(步骤S202); 获取目标视频对应的背景帧, 从目标视频中确定目标对象的目标动作所对应的目标帧(步骤S204); 对背景帧和目标帧进行图像分割, 得到背景帧中各个物体分别对应的物体区域和目标帧中目标对象对应的目标区域(步骤S206); 基于目标区域和物体区域的位置信息、物体区域对应的物体、目标动作, 生成交互信息组, 得到目标对象分别和各个物体之间的交互信息组(步骤S208); 基于各个交互信息组, 确定目标对象在目标帧对应的目标对象状态(步骤S210)。