

(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 2024/056078 A1

(43) 国际公布日
2024 年 3 月 21 日 (21.03.2024)

- (51) 国际专利分类号:
G10L 21/10 (2013.01) *G06T 15/00* (2011.01)
G06T 13/40 (2011.01) *H04N 21/439* (2011.01)
G10L 25/03 (2013.01)
- (21) 国际申请号: PCT/CN2023/119140
- (22) 国际申请日: 2023 年 9 月 15 日 (15.09.2023)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202211127727.7 2022年9月16日 (16.09.2022) 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, Guangdong 518057 (CN)。
- (72) 发明人: 苏贵楠(SU, Guinan); 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, Guangdong 518057 (CN)。 吴高(WU, Gao); 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, Guangdong 518057 (CN)。 李志锋(LI, Zhifeng); 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, Guangdong 518057 (CN)。 刘威(LIU, Wei); 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, Guangdong 518057 (CN)。 崔语诗(CUI, Yushi); 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, Guangdong 518057 (CN)。
- (74) 代理人: 深圳市深佳知识产权代理事务所(普通合伙) (SHENPAT INTELLECTUAL PROPERTY

(54) Title: VIDEO GENERATION METHOD AND APPARATUS AND COMPUTER-READABLE STORAGE MEDIUM

(54) 发明名称: 视频生成方法、装置和计算机可读存储介质

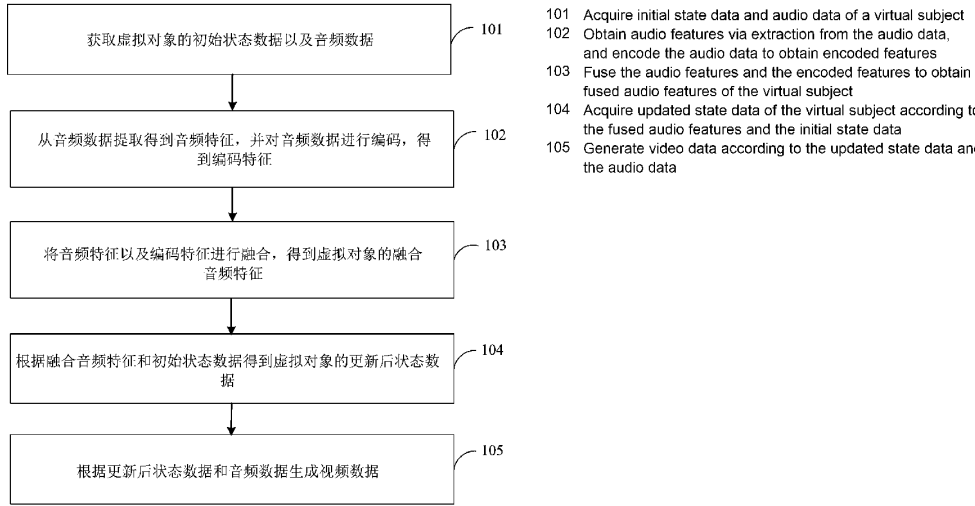


图 2

(57) Abstract: The embodiments of the present application disclose a video generation method and apparatus and a computer-readable storage medium. The method comprises: acquiring initial state data and audio data of a virtual subject; extracting audio features from the audio data, and encoding the audio data to obtain encoded features; fusing the audio features and the encoded features to obtain fused audio features of the virtual subject; acquiring updated state data of the virtual subject according to the fused audio features and the initial state data; and generating video data according to the updated state data and the audio data. Thus, matching accuracy of a

[见续页]

AGENCY); 中国广东省深圳市罗湖区南湖街道春风路庐山大厦B座18C2、18D、18E、18E2, Guangdong 518001 (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, 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))。

target part in the video data and the audio data is improved, thereby enhancing the video generation accuracy of the virtual subject, and further increasing the video generation efficiency of the virtual subject.

(57) 摘要: 本申请实施例公开了一种视频生成方法、装置和计算机可读存储介质; 通过获取虚拟对象的初始状态数据以及音频数据; 从音频数据中提取音频特征, 并对音频数据进行编码, 得到编码特征; 将音频特征以及编码特征进行融合, 得到虚拟对象的融合音频特征; 根据融合音频特征和初始状态数据得到虚拟对象的更新后状态数据; 根据更新后状态数据和音频数据生成视频数据。以此, 提高视频数据中目标部位与音频数据匹配的准确性, 进而提高虚拟对象的视频生成准确性, 进一步提升虚拟对象的视频生成效率。