

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования



**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Направление подготовки/профиль 09.06.01 Информатика и вычислительная техника
Школа Инженерная школа информационных технологий и робототехники
Отделение Информационных технологий

**Научный доклад об основных результатах подготовленной
научно-квалификационной работы**

Тема научного доклада
Алгоритмы процедурной генерации реалистичных лицевых анимаций для антропоморфных виртуальных агентов с применением методов машинного обучения

УДК 004.928:004.946:004.588

Аспирант

Группа	ФИО	Подпись	Дата
А7-39	Коровкин Виталий Александрович		

Руководитель профиля подготовки

Должность	ФИО	Ученая степень, звание	Подпись	Дата
профессор	Спицын Владимир Григорьевич	д.т.н., профессор		

Руководитель отделения

Должность	ФИО	Ученая степень, звание	Подпись	Дата
заведующий кафедрой - руководитель отделения информационных технологий на правах кафедры	Шерстнев Владислав Станиславович	к.т.н., доцент		

Научный руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
профессор	Спицын Владимир Григорьевич	д.т.н., профессор		

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования.

Человеческая коммуникация представляет из себя сложный процесс, включающий множество компонент: мимику, эмоции, позы тела, интонации, взгляд, движения глаз, жесты, жестикуляции и т.д. Одними только невербальными методами передается и неосознанно усваивается более 80 процентов информации согласно исследованиям А. Пиза. Поэтому для достижения более приемлемых результатов в решении различных задач человеко-машинного взаимодействия и коммуникации стараются использовать несколько различных модальностей вместе, главными из которых являются визуальная и звуковая.

Совместное использование визуальных (например, текущие лицевая мимика и эмоция) и звуковых (например, интонация в голосе) данных позволяет разрабатывать новые решения в областях анимации, человеко-машинной коммуникации, робототехники, биометрии и других мультимедийных сферах. Согласно ряду проектов и исследований, выражение лица является наиболее выразительным средством, с помощью которого люди демонстрируют свои эмоции.

Количество областей применения компьютерной анимации (в том числе и лицевой) за последнюю декаду лет значительно выросло. Сейчас они представлены не только анимационными и VFX-проектами в киноиндустрии, но и видеоигровыми, различными цифровыми ассистентами (помощниками), симуляторами, тренажерами, системами удаленного присутствия и виртуальной реальностью. Кроме этого, сильно возросли и требования к качеству контента, поставляемого пользователям.

Стоит сразу отметить, что применение традиционных методов захвата движений и мимики (motion capture system) позволяет получать стабильно высокий уровень детализации и реалистичности в анимации. Главным его недостатком является дороговизна эксплуатации: он требует специальной аппаратуры, маркеров, определенных камер, датчиков, уровня освещенности и площади помещения для записи. Кроме этого, получаемая анимация является уникальной и для ее изменения необходимо произвести заново захват движений актёра. Поэтому применение данного метода проблематично в проектах с большим числом уникальных виртуальных агентов.

С технической точки зрения создание правдоподобной анимации человеческих лиц в целом и мимики в частности даже в современных специализированных пакетах (например, Autodesk Maya, HoudiniFX, Modo и др.) представляет сложную трудоемкую задачу в компьютерной графике, для решения которой необходим специалист высокого уровня. Выполнение данной работы требует большого количества ручных операций. Анимации 3D-модели, построенная только на основе чистой геометрии не допускает микро-деформации из-за небольшого вектора параметров, что не позволяет создавать динамическую реалистическую визуализацию в сложных областях

человеческого лица таких как рот, складки глаз и т.д. Существует также методы анимации, основанные на использовании динамических текстур. Они могут захватывать детали и небольшие движения, которые не требуют изменения самой геометрии модели. Главными недостатками использования таких методов являются высокие требования к памяти и ограниченная гибкость получаемой анимации (динамические последовательности текстур должны быть легко объединены, что в реальности не всегда возможно и часто вызывает появление визуальных артефактов), что снижает их применимость в приложениях реального времени.

Альтернативой является создание и использование систем, основанных на методах процедурной генерации. Они могут использовать одну или несколько модальностей в качестве входных данных: аудиопоток, текст, набор лицевых эмоций. Использование одновременно нескольких модальностей позволяет повысить реалистичность сгенерированной анимации.

Чтобы виртуальные агенты выглядели реалистичными и правдоподобными для анимации мимики необходимо учитывать не только макро-выражения, но и малозаметные движения или «микро-выражения», такие как саккады и сокращения мимических мышц.

Большинство современных автоматизированных систем синтеза лиц или их анимации выполняется через поиск соответствующих местоположений признаков на них.

Кроме возрастающего практического интереса применения технологий процедурной генерации анимации, актуальность темы подтверждается резким увеличением количества крупных исследований и докладов на международных конференциях по компьютерной графике и проблемам компьютерного зрения за последнюю пару лет. К ярким примерам можно отнести доклады с «SIGGRAPH», опубликованные разработки Nvidia, Microsoft Research, «Графикон», «IEEE Computer Vision and Pattern Recognition», «ACM Transactions on Intelligent Systems and Technology» и т.д. Много исследований от ведущих специалистов в данной области можно найти на arXiv.org