

Застосування 3D-відео камер та інфрачервоних датчиків руху дозволяє виконувати детектування людини у просторі та виконувати стеження за його рухами і жестами. Авторами запропоновано розширення системи застосування технології OpenNI у сфері обробки звуку, внесення коректив в систему озвучування з урахуванням координат переміщення людини у просторі

Ключові слова: OpenNI, інтерактивні технології, Kinect, VST-плагін, OpenCV, 3D-камера, синтез хвильового поля

Применение 3D-видео камер и инфракрасных датчиков движения позволяет выполнять детектирование человека в пространстве и выполнять слежение за его движениями и жестами. Авторами предложено расширение системы применения технологии OpenNI в сфере обработки звука, внесения корректировок в систему озвучивания с учетом координат перемещения человека в пространстве

Ключевые слова: OpenNI, интерактивные технологии, Kinect, VST-плагин, OpenCV, 3D-камера, синтез волнового поля

МЕТОД КОРРЕКЦИИ АКУСТИЧЕСКОГО СИГНАЛА В СООТВЕТСТВИИ С ПЕРЕМЕЩЕНИЕМ ЗОНЫ ПРОСЛУШИВАНИЯ

С. М. Порошин

Доктор технических наук, профессор*

И. С. Беликов

Аспирант*

E-mail: igorajon@yandex.ua

*Кафедра мультимедийных информационных технологий и систем

Национальный технический университет «Харьковский политехнический институт» ул. Фрунзе 21, г. Харьков, Украина, 61002

1. Введение

В современном обществе всё более востребованным стало интерактивное участие человека в управлении графическим интерфейсом. В скором будущем такие органы управления, как клавиатура и манипулятор «мышь» постепенно отойдут от массового использования, уступив место более интуитивным методам передачи указаний от человека к машине, основанных на жестах, мимике и движениях человека [1 – 3].

2. Анализ литературных данных и постановка проблемы

На данный момент существует две разработки, позволяющие в реальном времени детектировать присутствие человека в пространстве, а также следить за его жестами, перемещением, мимикой лица.

Одной из таких революционных разработок является технология OpenCV. Суть данной технологии заключается в детектировании черт лица человека в изображении, слежении за перемещением и захватом движения рук, определением жестов и невербальных команд человека (рис. 1).

Алгоритм `opencv_video` позволяет преобразовывать отклонение человека от центрального местоположения в набор трехмерных координат, данные которых могут быть направлены на внесение корректировки амплитуды и задержки акустического сигнала, поступающего на один из пяти каналов звуковой системы.

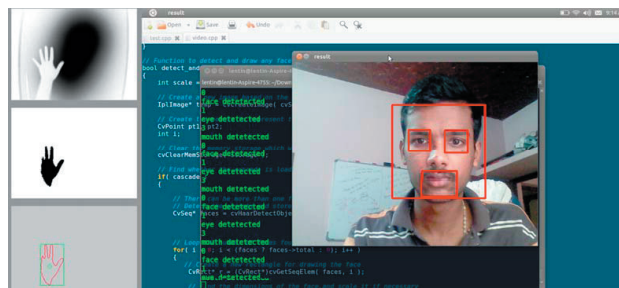


Рис. 1. Идентификация жестов и мимики человека в системе OpenCV

Недостатком системы OpenCV является, некорректная работа со многими моделями веб-камер, ошибки детектирования черт лица в условиях недостаточной освещенности и при отклонении лица от строго перпендикулярного положения относительно оси камеры.

Параллельно с технологией OpenCV, компания PrimeSense начала разработку технологии (OpenNI) для реализации определения движения объектов в трехмерном пространстве при помощи 3D видео камеры и инфракрасного сканирующего датчика [4].

Технология OpenNI направлена на взаимодействие человека и компьютера без каких либо манипуляторов. Специально разработанная камера Microsoft Kinect производит сканирование пространства при помощи инфракрасного лазера, определяет контуры тела человека и разновидности его жестов (рис. 2) выполняется 3D видео камерой, снимающей с частотой 30 кадров в

секунду с разрешением каждого кадра 640*480 пикселей.

Присутствие лазерного сканирования пространства позволяет отделять объект от препятствий, окружающих человека. Также, отличием от алгоритма OpenCV является успешное детектирование черт лица и движения человека, который отклоняется от перпендикулярного расположения относительно оси камеры и свободно перемещается в пространстве от 0,8 м до 3,5 м в радиусе действия камеры.

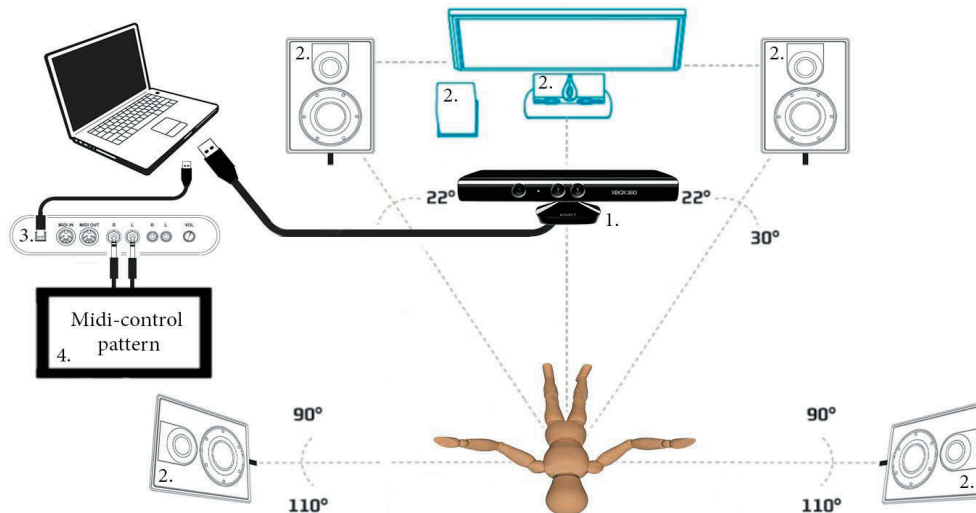


Рис. 3. Структурная схема аппаратно-программного комплекса : 1 - 3D камера Microsoft Kinect; 2 - акустическая система 5.1; 3 - звуковая карта компьютера; 4 - управляющие акустическим сигналом миди-сообщения

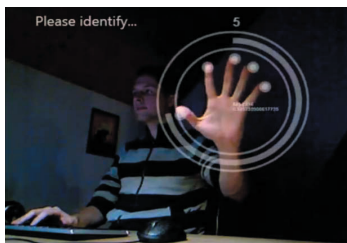


Рис. 2. Определение жестов человека в системе OpenNI

3. Цель и задачи исследования

Целью данной работы является объединение разработок в области управления акустическим сигналом в системе громкоговорителей 5.1 и системы распознавания местоположения человека в пространстве. Таким образом, согласовывая координаты заранее установленных сателлитов акустической системы 5.1 и отслеживая перемещения человека в пространстве, окруженном этими громкоговорителями, можно более детально восстанавливать эффект присутствия во время прослушивания звуковых композиций [1].

Подразумевается два варианта использования данной системы:

- сохранение оптимальной зоны прослушивания, при отклонении местоположения человека от центральной точки внутри пространства акустической системы 5.1, путём внесения задержек и регулирования амплитудно-частотной характеристики в каналах акустического тракта;

- акцентирование внимания на кажущемся источнике звука, который перемещается последовательно за перемещением слушателя в пространстве между сателлитами акустической системы 5.1. Таким образом, человек становится полноправным участником виртуального акустического пространства, что сейчас активно применяется в разработке компьютерных игровых симуляторов. Структурная схема аппаратно-программного комплекса представлена на рис. 3.

4. Основная часть

Целью исследования является разработка алгоритма преобразования координат перемещения оптимальной зоны прослушивания в пространстве во вторичный управляемый импульс, который в дальнейшем модулирует исходный акустический сигнал, поступающий на микширование в акустический тракт. Расширение возможностей многоканальной акустической системы возможно при помощи технологии синтеза волнового поля (WFS). Разработки над данной системой звуковоспроизведения ведутся с 1988 года и основаны на принципе Гюйгенса, согласно которому, кажущийся источник звука моделируется путем генерирования фронта волны, состоящей из нескольких сферических волн, синхронно поступающих из массива громкоговорителей [2 – 4].

Синтез волнового поля является одной из ключевых технологий для пространственного воспроизведения звука. На основе чисто физического описания акустических волновых полей, он имеет потенциал для точного воспроизведения желаемого звука в пространстве. Существует возможность создания в помещении эффекта объемного звучания, виртуального перемещения кажущегося источника звука в пространстве, создание дополнительных эффектов присутствия для слушателя. Эффект виртуальных источников звука воссоздается непрерывным распределением монополюсных и дипольных источников на замкнутой поверхности вокруг зоны прослушивания (рис. 4).

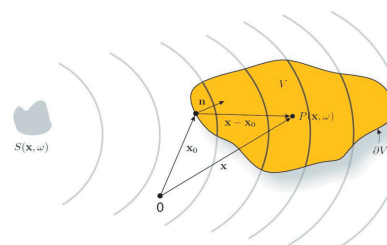


Рис. 4. Синтез оптимальной зоны прослушивания в пространстве

Реализация технологии WFS основана на частично линейных, или круговых акустических массивах громкоговорителей [5 – 7].

В данной работе выбран подход интерпретации акустических волновых полей как многомерных сигналов. Воспроизведение звукового поля, с измененным кажущимся местоположением источников звука, возможно путем обобщенной пространственно-временной свертки (фильтрации) источника волнового поля с управляющим сигналом, параметры которого задаются исходя из требуемых координат пространства.

Внесение необходимых корректировок в акустический тракт и последующее микширование аудио материала в системе 5.1 выполняется при помощи специально спроектированного плагина стандарта VST. Данный стандарт повсеместно поддерживается любыми существующими звуковыми картами компьютера и программными аудио секвенсорами.

Отладка и подключение таких приборов обработки звука как: дилэй, энхансер и компрессор, необходимые для корректировки временных и амплитудно-частотных характеристик, выполняется в программной конструкторской оболочке Native Instrument Reaktor (рис. 5). Данное программное обеспечение работает внутри аудио секвенсора непосредственно через аудио драйвер ASIO, что минимизирует задержки при обработке звука до 10 мс и менее, что не так заметно на человеческий слух.

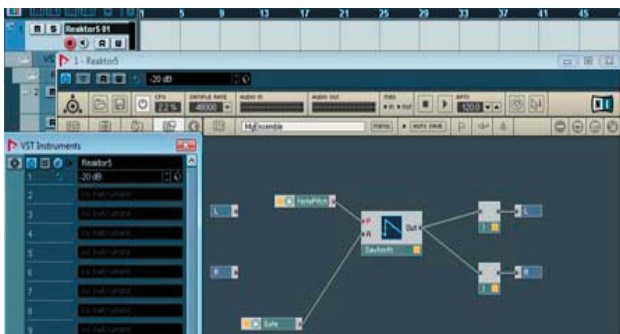


Рис. 5. Проектирование плагина стандарта VST в конструкторе Native Instruments Reaktor

Управление параметрами задержек, вносимых в аудио канал секвенсора, регулируется миди-сообщениями, поступающих на вход звуковой карты. Данные миди-сообщения несут в себе информацию о трехмерном расположении человека в заданном пространстве между сателлитами акустической системы, или за её пределами. Задачей данного исследования является построение алгоритма соответствия получаемых данных о координатах перемещения объекта и величин

вносимых искажений в акустический сигнал каждого из каналов акустической системы [8 – 10].

6. Выводы

Из двух существующих систем детектирования и трекинга движений человека, была принята OpenNI. Авторами проводится изучение явлений влияния задержек в аудиоканалы акустической системы типа 5.1 на восприятие объемного звука человеком:

- измерения и компенсации возникающих аппаратных и программных задержек при обработке сигнала;
- использование VST-конструкторов, таких как Native Instruments Reaktor, обрабатывающий акустический сигнал в реальном времени.

Литература

1. Miles, R. Start Here! Learn Microsoft Kinect API [Текст] / Rob Miles; O'Reilly Media, Inc. // Gravenstein Highway North Sebastopol, California 95472. – 2012. – с. 272.
2. Spors, S. Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis. [Текст] / S. Spors, R. Rabenstein // Deutsche Telekom Laboratories, Ernst-Reuter Platz 7, 10587 Berlin, Germany. – 2006. – с. 14.
3. Wittek, H. Perceptual difference between wavefield synthesis and stereophony [Текст] / H. Wittek // Department of music and sound recording school of Arts, Communication and Humanities, University of Surrey. – 2007. – с. 228.
4. Jana, A. Kinect for Windows SDK Programming Guide. [Текст] / A. Jana; Published by Packt Publishing Ltd. // Livery Place, 35 Livery Street, Birmingham B3 2PB, UK. ISBN 978-1-84969-238-0. – 2012. – с. 392.
5. Catuhe, D. Programming with the Kinect for Windows. [Текст] / D. Catuhe // Microsoft Press, Redmond, Washington 98052-6399. – 2012. – с. 224.
6. Bradski, G. Learning OpenCV. [Текст] / G. Bradski, A. Kahl; O'Reilly Media, Inc., // 1005 Gravenstein Highway North, Sebastopol, CA 95472. – 2008. – с. 571.
7. Kinect for Windows. Human Interface Guidelines v1.7 [Текст] / Microsoft Corporation. – 2013. – с. 135.
8. Кононович, Л. М. Стереофоническое восприятие звука. [Текст] / Л.М. Кононович, Ю.А. Ковалгин. – М.: Радио и связь, 1981. – 184 с., ил.
9. Крылов, В. В. Основы теории излучения и рассеяния звука. [Текст] – М.: Издательство Моск. ун-та, 1989. – с.118.
10. OpenAL 1.1 Specification and Reference [Текст] / – 2005. – с. 62.