MATHEMATICA MONTISNIGRI
Vol XXXII (2015) 149-157

МЕТОД РАЗДЕЛЕНИЯ ВИДЕОПОТОКОВ В ЗАДАЧАХ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

А.Г.КУШНИРЕНКО, А.В.МАЛЬЦЕВ, М.В.МИХАЙЛЮК, А.А.ПРИЛИПКО, П.Ю.ТИМОХИН, М.А.ТОРГАШЕВ

Научно-исследовательский институт системных исследований РАН
Москва, Россия
e-mail: mix@niisi.ras.ru

Ключевые слова: Дистанционное обучение, разделение видеопотоков, Kinect

Аннотация. В области дистанционного обучения одной из важных задач является передача по сети качественного видеопотока в масштабе реального времени. В данной работе для решения этой задачи предлагается метод разделения исходного потока с использованием устройства Microsoft Kinect на потоки, содержащие только доску, только лектора и маску для его выделения. Раздельное сжатие, передача и слияние этих потоков позволяют достичь меньшего битрейта, чем при передаче исходного видеопотока.

THE METHOD FOR SEPARATION OF VIDEO STREAMS IN DISTANT EDUCATION

A.G.KUSHNIRENKO, A.V. MALTSEV, M.V. MIKHAYLYUK, A.A. PRILIPKO, P.YU. TIMOKHIN, M.A. TORGASHEV

Scientific Research Institute for System Analysis of RAS
Moscow, Russia
e-mail: mix@niisi.ras.ru

Summary. One of the important issues in the field of distance education is to transmit over network a qualitative video stream in real time. In this paper to solve this task we propose a method to divide the source stream using Microsoft Kinect device into streams containing only the board, the lecturer and a mask for selecting the lecturer. Separate compression, transmission and merging of these streams into one allow us to achieve lower bitrate compared to the direct transmission of the source video.

2010 Mathematics Subject Classification: 94A08.
Key words and Phrases: Distance Education, Stream Dividing, Kinect.

1 ВВЕДЕНИЕ

Одними из основных задач реализации дистанционного обучения в масштабе реального времени являются эффективное сжатие видеопотока, его быстрая передача по информационной сети, раскодирование и вывод на экран. Примером является онлайн видеотрансляция лекции, которая читается лектором с применением стандартных средств (доски и маркеров или электронной доски). Масштаб реального времени предполагает непрерывное и равномерное воспроизведение видеопотока (без неоправданных задержек, дерганий и т.д.). Основной проблемой в этой задаче является то, что при высоком разрешении видео объем данных кадра оказывается слишком большим для передачи этого видео по сети Internet и его воспроизведения на принимающей стороне со скоростью 25 кадров в секунду. В настоящее время качество и скорости работы линий связи во многих регионах мира не позволяют осуществлять передачи с данными параметрами с высоким битрейтом. В связи с этим возникает задача формирования таких видеопотоков, которые можно было бы передать по слабым информационным линиям с сохранением приемлемого качества видеопотока.

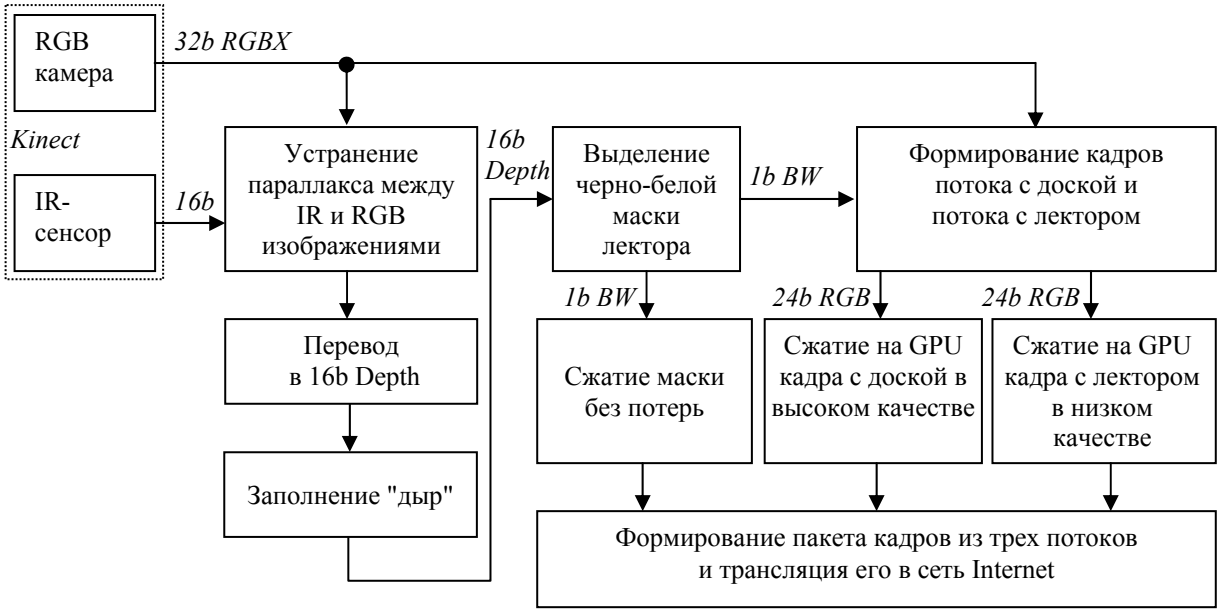

Рис. 1. Структура передающей части системы

Одним из способов решения поставленной задачи является разделение видеопотока на основе сегментации[1,2]. Например, можно формировать отдельный видеопоток для доски и отдельный – для лектора, раздельно сжимать их и передавать по сети, а для воспроизведения на принимающей стороне объединять и выдавать потребителю. При этом можно поставить вопрос об адаптивности сжатия потоков, т.е. то, что является менее информативным сжимать сильнее, понижая качество и, следовательно, уменьшая размер соответствующего видеопотока. В данной работе для реализации этой технологии предлагается использовать устройство Microsoft Kinect. На рисунке 1 показана предлагаемая структура программно-аппаратного комплекса для передающей части системы видеотрансляции, реализованная с использованием устройства Microsoft

Kinect и основанная на описанном выше принципе разделения потоков. Далее мы рассмотрим эту структуру подробно.

2 РАЗДЕЛЕНИЕ ИСХОДНОГО ВИДЕОПОТОКА

2.1 Получение данных с Kinect и их коррекция

Microsoft Kinect представляет собой мульти-компонентное устройство, подключаемое к персональному компьютеру через USB интерфейс. В его состав входят:

- инфракрасный излучатель (ИК-излучатель), испускающий ИК-лучи, которые далее отражаются от окружающих предметов;
- инфракрасный сенсор глубины (ИК-камера), который принимает отраженные ИК-лучи и на их основе вычисляет расстояния от сенсора до точек объектов;
- цветная видеокамера (RGB камера), осуществляющая захват видео с углами обзора 43° по вертикали и 57° по горизонтали.

Kinect первого поколения (Kinect-1) позволяет получить на своем выходе цветной видеопоток в формате 32-битного RGB с разрешением 640x480 и частотой 30 кадров/сек., а также карту глубины рабочего пространства устройства в градациях серого (16 бит/пиксел) с таким же разрешением и частотой смены кадров. Глубина показывает расстояние от устройства Kinect до точек объектов. Сама рабочая область прибора располагается в интервале расстояний от 0.8 до 4.0 метров от его сенсоров. Недавно появившийся в продаже Kinect второго поколения обеспечивает существенное улучшение качества цветного видеопотока. При частоте смены кадров 30 Гц разрешение изображения достигает Full HD (1920x1080).

На рис. 2 показана предлагаемая схема размещения Kinect. Устройство необходимо устанавливать так, чтобы вектор его взгляда был перпендикулярен доске, а расстояние до доски составляло около 3 метров.

Первым шагом технологии является получение синхронизованных (т.е. соответствующих одному и тому же моменту съемки) кадров изображения (RGB) и глубины (расстояний от камеры Kinect до точек объектов), см. рис. 3. Это можно сделать по запросу из устройства Kinect, использовав его SDK[3].

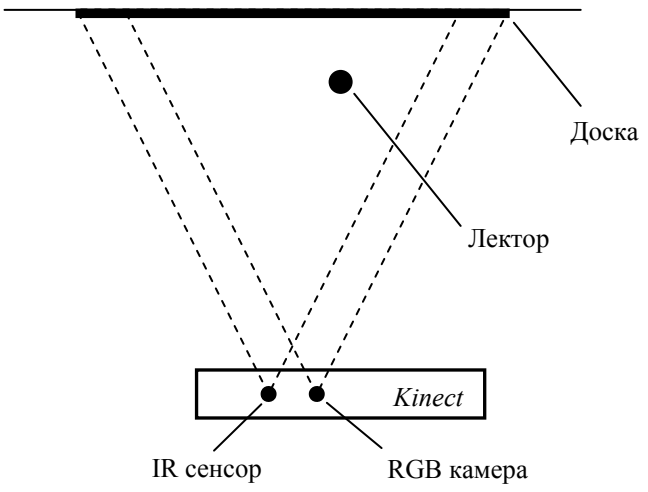

Рис. 2. Установка Kinect

Следующей задачей является устранение параллакса между этими двумя изображениями. Параллакс возникает из-за того, что в устройстве Kinect-1 RGB камера и инфракрасный сенсор глубины размещены в разных пространственных точках (рис. 2 и 3), а их векторы взгляда и углы раствора совпадают только с некоторой погрешностью. В результате этого RGB изображение и карта глубины

захватывают несколько различные области рабочего пространства перед устройством. Таким образом, один и тот же объект на обоих изображениях будет располагаться со смещением (как горизонтальным, так и вертикальным), варьирующимся в зависимости от расстояния от объекта до Kinect-1. Кроме того, существуют области, где изображение получается только из одного источника (RGB камеры или IR сенсора).

Для компенсации возникающего параллакса необходимо осуществить коррекцию карты глубины с учетом значения расстояния в каждом пикселе, результатом которой будет являться совмещение объектов на двух изображениях. Такая коррекция может быть проведена с помощью функции *MapColorFrameToDepthFrame*, включенной в Kinect SDK. Однако, при этом на результирующем изображении карты глубины появляются вертикальные (слева и справа) и горизонтальные (внизу и вверху) черные полосы, соответствующие отсутствию информации. Их следует удалить, выполнив обрезание кадра глубины на некоторые величины, которые можно задавать при калибровке системы. Отметим, что RGB изображение должно быть кадрировано аналогично.

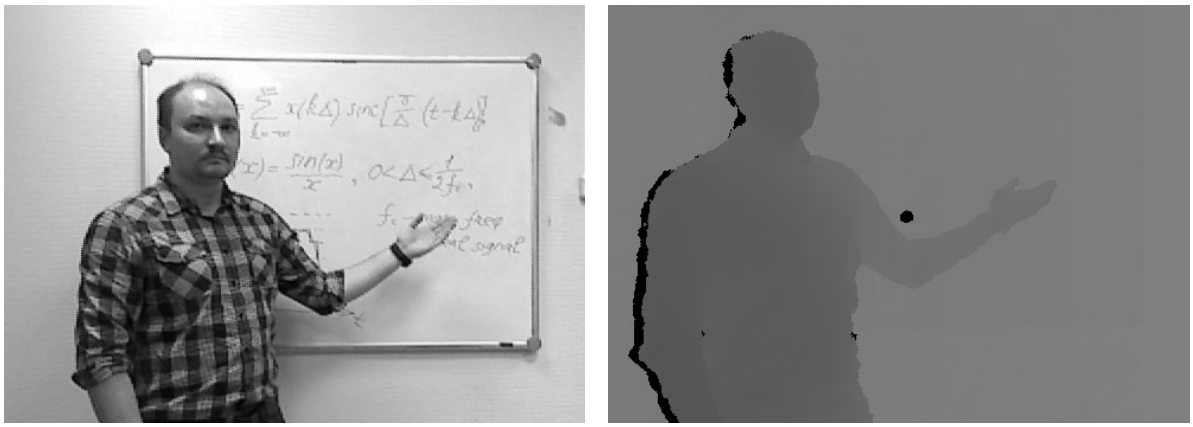

Рис. 3. RGB изображение и карта глубины, полученные из устройства Kinect

Получаемое с устройства Kinect изображение глубин содержит кроме данных о самих глубинах, представленных в виде вещественных 16-битных чисел, также дополнительную информацию, которая не представляет для нас интереса и должна быть удалена из потока глубины. Удаление можно выполнить с помощью функции *NuiDepthPixelToDepth*, входящей в пакет Kinect SDK.

Для дальнейшего использования значений глубины нам удобнее представлять их не в формате вещественных чисел, а в виде целых чисел из диапазона $[0, 2^{16}\text{-}1]$. Данное преобразование можно реализовать по формуле

$$D = (2^{16} - 1)\frac{D_f}{L},$$

где D и D_f – искомое и исходное значения глубины соответственно, L – максимальная глубина, воспринимаемая устройством с допустимой точностью, в мм (для Kinect-1 оно составляет 4000 мм).

2.2. Заполнение «дыр» в карте глубины

Полученная таким образом карта глубин содержит артефакты в виде так называемых «дыр». Под «дырами» в карте глубины мы понимаем области, в которых записан черный цвет (соответствует тому, что Kinect не смог распознать глубину). Для того чтобы быстро и качественно устранять дыры, необходимо знать, какие размеры и места их появления на изображении характерны применительно к нашей задаче. Это важно, т.к. для разных случаев лучше работают различные методы заполнения таких дыр. Рассмотрим, какие типы дыр возможны при работе с Kinect, какие из них присущи нашей задаче и выберем, какой метод лучше всего справится с их устранением.

I тип. Дыры, возникающие в результате заграждения отраженных лучей ИК-излучателя от ИК-камеры. Фактически это инфракрасные тени, образующиеся за счет разницы в расположении этих частей устройства Kinect. Они всегда появляются около границы облучаемого объекта (обычно с одной его стороны), где имеется большая разница по глубине с задним фоном. Причем, чем дальше облучаемый объект находится от фоновой отражающей поверхности, тем большей величины будут эти дыры.

II тип. Дыры, возникающие в результате сильного поглощения или рассеяния ИК-излучения поверхностями объектов. Тогда в детектор Kinect будет поступать малое количество ИК-излучения, которое недостаточно для определения глубины. Примером может служить коробка с дырой. Лучи, прошедшие через дыру под углом к стенке, рассеются внутри коробки, и лишь малая их часть выйдет через эту дыру в направлении детектора.

III тип. Дыры, возникающие от ограниченности диапазона чувствительности детектора Kinect. Воспринимаемый Kinect диапазон измеряемых глубин ограничен передней и задней отсекающими плоскостями (например, от 0,8 до 4 метров). Для всех объектов (или их частей) вне этого диапазона данные о глубине не определяются.

IV тип. Дыры, возникающие в результате интерференции ИК-излучения от различных источников ИК излучения (например, нескольких Кинектов). Такие дыры случайным образом распределены по всей карте глубин. Также к данному типу можно отнести дыры от засветки солнечным светом.

В рассматриваемой нами задаче будут встречаться только дыры **I** и **II** типа (см. рис. 3 справа). Первые – это небольшие продольные дыры вдоль одной из сторон силуэта лектора. Вторые – это небольшие дырки внутри фигуры лектора в области складок одежды и на доске в области подставки для маркеров и блика от источника освещения комнаты (или ИК-излучателя Кинекта). Появление дыр III-го типа мы избегаем ввиду расположения всей сцены (лектора и доски) непосредственно в рабочей области Kinect. Если используется одно устройство Kinect в закрытом помещении с искусственным освещением, то дыры IV-типа также исключаются.

В настоящее время задача устранения дыр в карте глубин, синтезируемой Kinect'ом, активно исследуется, и разрабатываются различные решения. Многие известные алгоритмы осуществляют восполнение данных о глубине на основе сложного анализа окрестности пикселов внутри дыр[4], используя данные из карт глубин предыдущих кадров[5], а также с привлечением информации из буфера цвета[6]. Данные алгоритмы носят достаточно универсальный характер и направлены на восстановление точного распределения глубин в дырах различного типа, ввиду чего их выполнение

сопровождается существенными вычислительными затратами, что ограничивает их применение в задачах обработки видеопотока в режиме реального времени.

В рассматриваемой задаче карта глубины необходима лишь для формирования битовой маски, по которой будет выполняться отделение силуэта лектора от фона с доской, поэтому здесь нет необходимости в точном восстановлении распределения глубин внутри дыры, а важно лишь иметь возможность различать глубины переднего и заднего плана. Исходя из этого, в данной работе предлагается разделить все дыры на те, которые принадлежат переднему плану, и те, которые относятся к заднему плану, и заполнять каждую дыру одним значением глубины, соответствующим переднему или заднему плану. Для этого в каждой дыре выполняется обход внешнего контура пикселов и подсчитывается количество пикселов, глубины в которых отличаются от минимальной глубины для всего контура на некоторую Δ, фиксированную для всей карты глубины. Если количество таких пикселов превышает 30% от числа пикселов в контуре, то вся дыра закрашивается глубиной, максимальной для всего контура дыры, в противном случае – минимальной.

Рис. 4. Исходная и скорректированная карты глубины

Эксперименты показывают, что в получаемой от Kinect карте глубины возможна такая ситуация, когда одна часть дыры будет располагаться на фигуре лектора, а другая часть – в области доски (слияние дыр I-го и II-го типа). Для обработки такого случая мы перед выделением контуров дыр применяем к карте глубины медианный фильтр с размером окна 7x7 пикселов. В результате такой фильтрации дыры в области перехода от лектора к доске разбиваются на две отдельные дыры, которые затем закрашиваются на основе описанного выше обхода их контуров. Кроме этого применение медианного фильтра позволяет уже на предварительном шаге сразу избавиться от дыр малого размера, а также получить более гладкую границу у силуэта лектора.

2.3 Получение битовой маски лектора

На основе скорректированной карты G_{IN} глубины выполняется формирование битовой маски M, необходимой для последующего выделения силуэта лектора из RGB кадра исходного видеопотока. Простой подход состоит в сравнении значений из карты G_{IN} глубины с некоторым фиксированным пороговым значением. Однако при таком

способе построения маски приходится выбирать значение порога с большим запасом во избежание ложного включения в передний план выступающих элементов фона (доски), а также учитывать разброс глубин вдоль плоскости доски, возникающий из-за неточности расположения Kinect относительно плоскости доски и погрешности работы ИК-излучателя. Это приводит к заметному ограничению рабочей зоны, в которой лектор может контактировать с доской, а также к усложнению процесса настройки системы видеотрансляции.

Рис. 5. Битовая маска

В данной работе предлагается более точный и удобный способ формирования битовой маски M, при котором каждое значение из карты G_{IN} сравнивается, по сути, с индивидуальной пороговой глубиной, записанной в специальной карте, совпадающей по размерам с картой G_{IN}, – карте пороговых глубин. Данная карта представляет собой усредненную по серии кадров карту глубины фона (без лектора), полученную с помощью Kinect на этапе калибровки системы видеотрансляции. Для каждого (i, j)-го пиксела глубина D_T в карте пороговых глубин вычисляется как

$$D_T = \frac{1}{N_F}\sum_{k=1}^{N_F} D_B^k - d\,,$$

где D_B^k - значение, записанное в (i, j)-ом пикселе карты глубины фона, снятой на k-ом кадре, N_F - количество кадров, по которым выполняется усреднение глубин D_B^k, а d - константная величина, задающая отступ по глубине от фона в сторону лектора с учетом погрешности измерения глубины Kinect. Напомним, что карты глубин фона, на основе которых формируется карта пороговых глубин, также содержат в себе дыры, которые мы устраняем после усреднения глубин с помощью метода, описанного в предыдущем разделе.

Отделение переднего плана (лектора) от заднего плана (доски) с помощью карты пороговых глубин осуществляется попиксельно следующим образом:

$$D_M = \begin{cases} 0, & \textit{если } D_{IN} \geq D_T\,, \\ 1, & D_{IN} < D_T\,, \end{cases}$$

где D_M - битовое значение (1 соответствует переднему плану, 0 - заднему), D_{IN} - значение пиксела (i, j) из карты G_{IN} глубины, D_T - значение пиксела (i, j) из карты пороговых глубин. На рис. 5 показана маска M лектора, полученная из карты глубины, изображенной на рис. 4 справа.

2.4 Получение кадров потока с лектором и потока с доской

На основе подготовленной выше маски M нам необходимо сегментировать RGB изображение, получаемое с устройства Kinect, чтобы синтезировать кадр F_B, содержащий только доску, и кадр F_L с изображением лектора. Отметим, что для повышения эффективности дальнейшего сжатия, следует произвести преобразование исходных данных из 32-битного формата RGBX (X - незначащий 8-битный канал), выдаваемого Kinect, в формат 24-битного RGB.

В нашем методе предполагается, что в момент инициализации системы видеотрансляции лектор находился вне рабочей области, т.е. кадр полностью заполняется изображением доски и записывается в память. Тогда текущий n-ый кадр F_B^n формируется следующим образом. Все пикселы, не занятые лектором (для них соответствующие пикселы маски будут содержать значение 0), будут копироваться из исходного текущего кадра. Пикселы, занятые лектором, будут копироваться из предыдущего кадра F_B^{n-1} доски. Таким образом, в кадре F_B^n на месте лектора будет записана старая информация на доске (может быть, уже неактуальная), а в остальных местах – актуальная информация.

Исходя из этого, расчет цвета каждого пиксела (i, j) синтезируемого кадра F_B^n выполним по формуле

$$C_{FB}^n = C_{IN}^n \cdot \left(1 - D_M^n\right) + C_{FB}^{n-1} \cdot D_M^n ,$$

где D_M^n – значение в пикселе (i, j) текущей маски M (может быть 0 или 1), C_{FB}^n и C_{FB}^{n-1} – цвета пиксела (i, j) текущего F_B^n и предыдущего F_B^{n-1} кадров потока с доской, C_{IN}^n – цвет пиксела (i, j) из текущего RGB кадра F_{IN}^n, получаемого с Kinect. Данный подход позволяет обеспечить достаточную статичность кадров потока с доской для их эффективного сжатия на следующем шаге нашей технологии. Отметим, что синтез кадра целесообразно производить на многоядерном графическом процессоре с поддержкой архитектуры CUDA, где цвет каждого пиксела рассчитывается на своем ядре, обеспечивая высокую степень параллелизма вычислений.

Подготавливаемый кадр F_L будет содержать изображение лектора, если тот находится в рабочей области устройства Kinect. При этом фон (значение маски M для пикселов равно 0), окружающий лектора, необходимо заполнять в зависимости от используемого для дальнейшего сжатия кодека. В ходе наших исследований установлено, что заливка фона одним цветом (например, черным) повышает степень сжатия кодека.

На рисунке 6 приведен результат разделения исходного потока на два: поток, содержащий только доску и поток, содержащий только лектора.

3 ЗАКЛЮЧЕНИЕ

В данной статье предложен метод, позволяющий с помощью устройства Microsoft Kinect разделить исходный видеопоток с лектором и доской на несколько потоков: содержащий только доску, содержащий только лектора и поток маски для выделения лектора. Экспериментальные исследования показали, что раздельное сжатие, передача

и слияние этих потоков позволяют достичь большего битрейта, чем при передаче исходного видеопотока. Предлагаемый метод может быть использован в задачах дистанционного обучения и проведения видеоконференций.

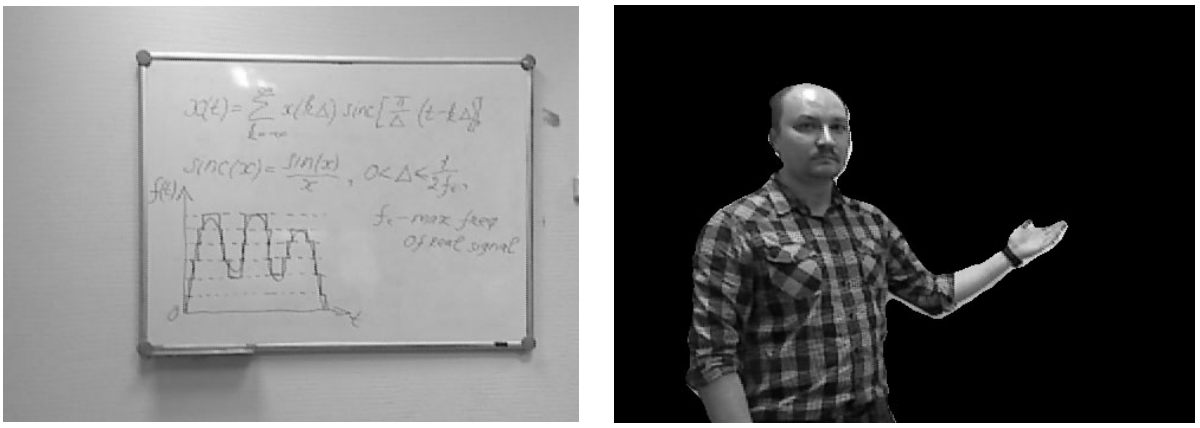

Рис. 6. Кадры потока с доской и потока с лектором

REFERENCES

[1] Komarov E.V., Losev A.S., Orlov N.S., Chistyakov D.V., “Blizhayshee budushchee nauchnoobrazovatel'nogo video”, *Sbornik trudov “Informatsionnye tekhnologii v obespechenii federal'nykh gosudarstvennykh obrazovatel'nykh standartov”,* Elets,**1**, 307-312, (2014).

[2] Wang C., Chan S. C., Ho C. H., A. L. Liu, Shum H. Y., “A real-time image-based rendering and compression system with Kinect depth camera”, *Proceedings of the 19th International Conference on Digital Signal Processing*, 626-630, (2014).

[3] *Kinect for Windows SDK*, https://msdn.microsoft.com/en-us/library/hh855347.aspx (date of review: 10.03.2015).

[4] Telea A., “An image inpainting technique based on the fast marching method”, *Journal of Graphics Tools*, **9**, 23-34, (2004).

[5] Matyunin S., Vatolin D., Berdnikov Y., Smirnov M., “Temporal filtering for depth maps generated by Kinect depth camera”, *3DTV Conference: The true vision - capture, transmission and display of 3D video (3DTV-CON),* 1-4 (2011).

[6] Kopf J., Cohen, M. F., Lischinski, D., Uyttendaele M., “Joint bilateral upsampling”, *IEEE Transactions on Graphics SIGGRAPH*, **26**, 96 -100, (2007).

Поступила в редакцию 10.01.2015.