

Распределенная система видеомониторинга для интеллектуального пространства

Ронжин Ал.Л., Прищепа М.В., Будков В.Ю., Карпов А.А., Ронжин А.Л.

Лаборатория речевых и многомодальных интерфейсов

Санкт-Петербургский институт информатики и автоматизации РАН (СПИИРАН), Санкт-Петербург, Россия
{ronzhinal, prischepa, budkov, karpov, ronzhin}@ias.spb.su

Аннотация

В статье представлена система видеомониторинга, основанная на наборе распределённых видеокамер, и предназначенная для автоматической регистрации участников различных научных мероприятий, проводимых в интеллектуальном пространстве (smart space). Проблема определения местоположений и позиций активных участников в разработанном интеллектуальном зале, а также детектирования их лиц, решается с применением широкоугольных, стационарных видеокамер и видеооборудования с функциями наклона, поворота (pan/tilt/zoom - PTZ) и масштабирования, установленных на стенах и потолке зала. В ходе экспериментов производилась оценка качества автоматического фотодокументирования участников лекций, а также точность наведение PTZ-камер на лица людей. Смещение лиц относительно центра кадра (ошибка наведения) в среднем составило около 8%, в среднем лицо занимает не менее 7% изображения видеокadra. Эксперименты показали, что среднее время автоматической многомодальной регистрации участников мероприятия возрастает с увеличением числа участников и уменьшением уровня освещённости. Дальнейшая разработка системы направлена на развитие системы поддержки проведения мероприятия, включая поддержку выступлений, дискуссии, голосования, а также иных ключевых частей мероприятий.

Ключевые слова: видеомониторинг, компьютерное зрение, интеллектуальное пространство, отслеживание

1. ВВЕДЕНИЕ

В связи с быстрым развитием широкоэшелательных сетей, достижениями в обработке и передаче мультимедийной информации, появлением Интернет приложений для телеконференций и дистанционного образования, всё более популярными в бизнесе, академических институтах и университетах при проведении распределённых мероприятий становятся так называемые системы E-meeting и E-lecture. Такие системы позволяют экономить на транспортных расходах, выбирать индивидуальный способ обучения, а также предоставляют удобные средства поиска и доступа к информации разной природы. Технологические достижения в области записи, анализа, поиска и передачи мультимедийных данных уже сейчас используются в ряде исследовательских систем для проведения телемероприятий (распределённых совещаний, лекций, конференций и т.д.). Далее попытаемся проанализировать круг проблем, сдерживающих развитие средств дистанционного обучения и совместной работы, а также методы их решения, реализованные в исследовательских телесистемах.

Распределённые мероприятия, организованные посредством специализированных веб-приложений, становятся реальной альтернативой традиционным совещаниям и лекциям, где участники находятся “лицом к лицу”. С развитием

мультимедийных технологий телемероприятия выполняют не только функции ввода/вывода аудио- и видеоданных, но также используют более сложные технологии, такие как: интеллектуальный анализ, распознавание и структурирование мультимедийных данных. Все это, несомненно, расширяет возможности участников и приводит к появлению новых способов доступа к мероприятиям в режиме реального времени, а также возможностей для обработки архивных мультимедийных записей [1]. Значительная часть мировых исследований, проводимых в данной области, посвящена развитию программно-аппаратных средств аудио- и видеообработки. Сейчас активно используются панорамные камеры, планшеты для рукописных записей, электронные доски (whiteboard), средства для мультимедийных презентаций, технологии синхронизации и объединения данных, а также интеллектуального анализа поведения и способов взаимодействия участников мероприятий.

Автоматический анализ аудиовидеозаписей совещаний или лекций осложняется тем, что такие мероприятия обычно проводятся в больших залах с большим числом участников, которые произвольно меняют свое положение. Чтобы обеспечить захват и слежение за наибольшим числом участников, применяются массивы микрофонов, панорамные камеры, PTZ-камеры с функциями наклона, поворота и масштабирования, распределённые системы видеокамер.

Так, в системе Cornell Lecture Browser [2] для видеозаписи используются два сенсора с последующей синхронизацией и объединением видеопотоков. В проекте eClass, также известном как Classroom 2000 [3], видеозаписи лекций комбинируются с рукописными записями на электронной доске. Например, система AutoAuditorium [4] обеспечивает слежение за диктором с помощью PTZ-камеры. Кроме того, система производит также автоматический выбор видеопотока от одного из трёх сенсоров, установленных в аудитории. В работе [5] предлагается также использовать пространственную локализацию источников звука (речь участников) для первичного наведения PTZ-камер и слежения за выступающими и слушателями, а также определён список правил и рекомендаций, применяющихся на практике профессиональными операторами, соблюдение которых помогает выбрать оптимальное расположение сенсоров в зале. В работе [6] управление восьмью PTZ-камерами осуществляется посредством интеллектуального анализа видеопотока, определения положения источника звука и данных от ультразвуковых датчиков движения. В системе FlySpec [7] используются одновременно PTZ- и всенаправленные камеры, управление которыми может осуществляться одновременно несколькими удалёнными участниками. При этом автоматическая система управления корректирует углы наведения PTZ-камер, чтобы удовлетворить командам максимального числа пользователей. Применение панорамных камер также позволяет вести съёмку всех событий, происходящих в зале, и отслеживать местоположение каждого участника [8].

2. СИСТЕМА ВИДЕОМОНИТОРИНГА УЧАСТНИКОВ МЕРОПРИЯТИЙ

В 2008-2010 годах в институте разрабатывается зал (натурная модель интеллектуального пространства), в котором изучаются вопросы человеко-машинного взаимодействия в различных ситуациях. Разработанный интеллектуальный зал представляет собой распределенную систему, которая содержит сеть интеллектуальных агентов (аппаратно-программных модулей), активационных устройств, мультимедийных средств и аудиовизуальных сенсоров. Основная задача зала — обеспечение участников мероприятий необходимыми сервисами на основе автоматического анализа текущей ситуации в зале и окружающей обстановки[9].

Наличие многомодального пользовательского интерфейса является отличительной характеристикой созданного интеллектуального пространства. Разработанные технологии для автоматической обработки аудиовизуальных данных были успешно внедрены в зале для обеспечения естественного взаимодействия с оборудованием. Среди наиболее важных технологий, применённых в интеллектуальном пространстве, следует отметить: автоматическое распознавание русской речи, голосовую

идентификацию диктора, локализацию источников звука, определение положения и слежение за двигающимися объектами и лицами людей, определение позы человека.

На рисунке 1 представлена технологическая сеть интеллектуального зала. Совместная работа различных технологий снабжает систему управления зала данными о текущей обстановке в помещении, о поведении пользователей, а также обеспечивает робастное распознавание голосовых команд за счет анализа пространственно-временной, ситуативной информации и предпочтений пользователей.

Для проведения телеконференций выполнена модификация расположения оборудования внутри интеллектуального зала. При этом учитывались эргономические аспекты использования мультимедийного оборудования, а также выбрано такое расположение аудио- и видеозаписывающих устройств, которое обеспечивает захват наибольшего пространства. При выборе количества камер, их месторасположения, разрешающей способности, угла обзора объективов и иных параметров системы видеомониторинга, учитывались размеры зоны охвата, максимальное число объектов, требующих одновременного слежения, степень детализации анализируемых объектов, уровень освещенности и т.д.

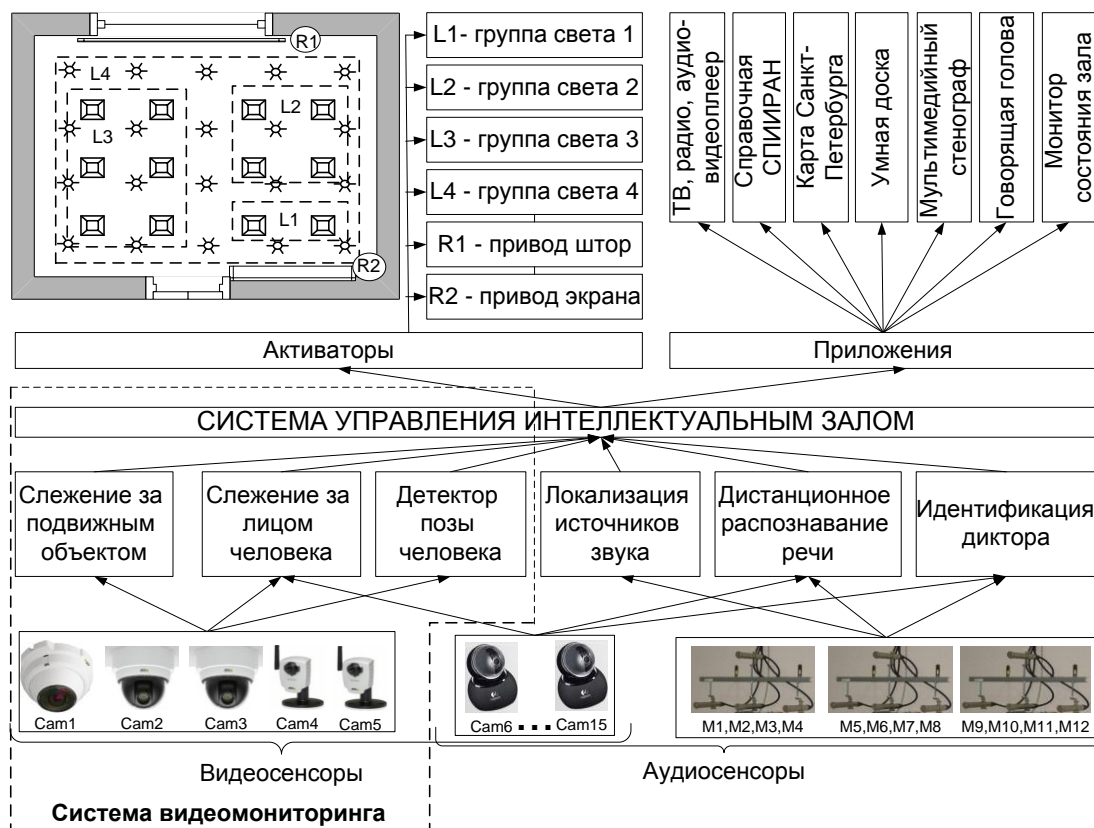


Рисунок 1: Технологическая сеть интеллектуального пространства с распределенной системой видеомониторинга

На рисунке 2 представлена схема расположения посадочных мест в зале; мультимедийного оборудования (ТВ, проектор), интернет-камер AXIS (PTZ-камеры Cam3, Cam5; беспроводные WiFi сенсоры Cam2, Cam4; видеокамера с

широкоугольным объективом Cam1, установленная на потолке по центру зала); 10-ти персональных веб-камер Logitech AF Sphere, установленных на конференц-стол;

четырёх групп освещения. Основное внимание в данной статье уделяется описанию разработанной и внедренной системы видеомониторинга, поэтому на рисунке не показаны четыре Т-образных массива микрофонов, расположенных на каждой из стен. Модельно-алгоритмическое обеспечение для многоканальной обработки аудиопотоков, использованное при разработке интеллектуального пространства представлено в работе [9].

Для размещения участников небольших совещаний (круглых столов до 10 человек) в левой части зала расположен конференц-стол с установленными на нем персональными

видеокамерами. В правой части зала расположены ряды кресел, на которых могут разместиться до 32 участников конференций или лекций, слежение за которыми реализовано посредством распределенной системы видеокамер и массивов микрофонов. С учетом указанного расположения оборудования, в зале выделены три непересекающиеся зоны (см. области, показанные на рисунке 2 синим цветом), требующие постоянного видеослежения: (1) зона перемещения основного докладчика; (2) зона участников, сидящих за конференц-столом; (3) зона участников, сидящих в рядах кресел.

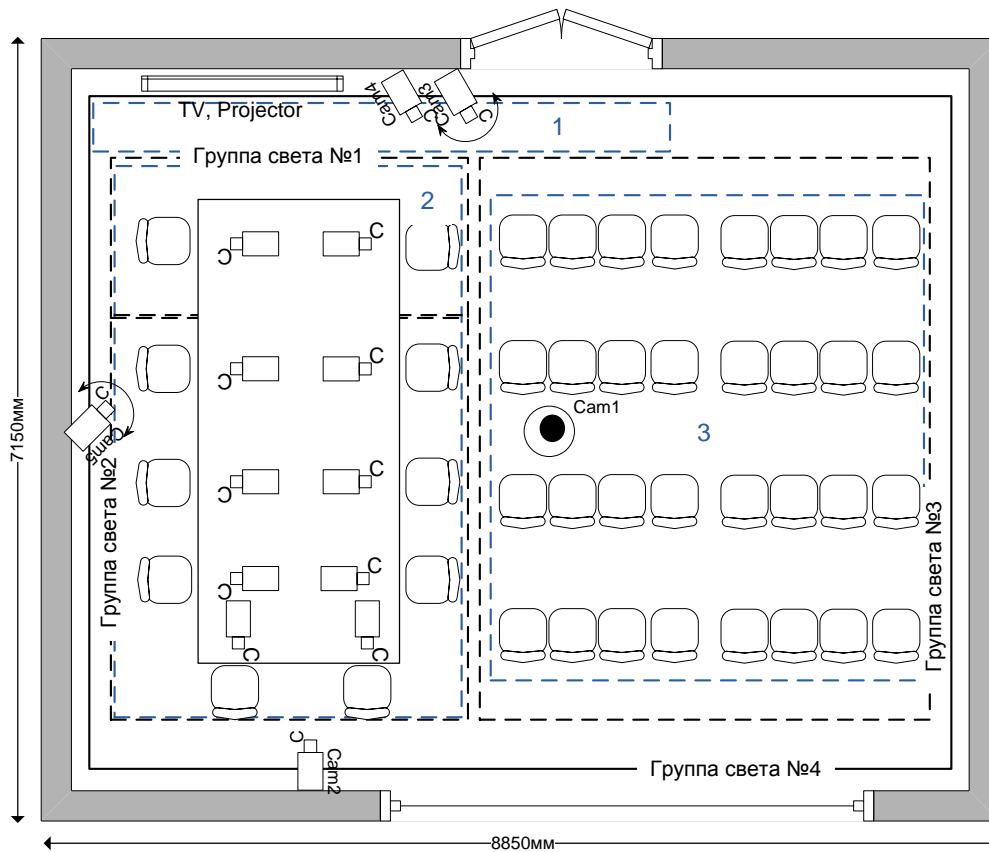


Рисунок 2: План-схема интеллектуального зала с расположением видеокамер.

Видеомониторинг всех трех зон осуществляется с применением видеокамеры Cam1, обладающей углом обзора 140 градусов и разрешающей способностью 2048x1536 пикселей. Стационарные видеокамеры Cam2 и Cam4 имеют угол обзора 74 градусов и разрешающую способность 1280x1024 пикселей. Две камеры Cam3 и Cam5 с углами обзора 52 градуса и разрешающей способностью 704x576 пикселей, оснащенные функциями наклона, поворота и масштабирования, обеспечивают наведение на любую точку пространства в зале.

Видеокамера Cam2 установлена на высоте 2 метров на противоположной от проекционного экрана стене и используется для аудиовидеозаписи хода мероприятия. Cam1 имеет широкоугольный объектив, и расположена на потолке в центре зала, поэтому она одновременно охватывает все пространство зала и позволяет определить координаты всех участников в горизонтальной плоскости. Стационарная камера Cam4 установлена слева от входной двери и

направлена таким образом, чтобы охватывать сразу всех пользователей, сидящих в зоне кресел. PTZ-камера Cam3 и стационарная видеокамера Cam2 установлены рядом и в большинстве случаев работают синхронизированно. Анализируя видеокadres, поступающие от Cam2, определяется положение наблюдаемого объекта и формируется управляющая команда для наведения Cam3 в направлении, необходимом для захвата лица определенного участника, находящегося в зоне кресел. PTZ-камера Cam5 установлена по центру левой стены зала на высоте 2 метра от уровня пола и служит, в основном, для захвата видеоизображения выступающего докладчика.

Разработанная многофункциональная система видеомониторинга интеллектуального зала на основе системы IP-видеокамер AXIS позволяет осуществлять видеозапись мероприятия, а также автоматизировать такие ключевые этапы совещания, как регистрация участников, съемка выступающего и активных участников в аудитории во

время дискуссии и т.д. Исходя из основных этапов предполагаемых мероприятий, были составлены различные режимы работы распределенной системы видеомониторинга.

В таблице 1 показаны функции каждого видеосенсора в пяти основных режимах: наблюдение, слежение за участниками, регистрация участников, выступление докладчика, дискуссия.

Камера (Cam)	Режимы работы системы				
	Наблюдение	Слежение за участниками	Регистрация участников	Сопровождение мероприятия	
				Выступление докладчика	Дискуссия
1	Слежение за изменением состояния буферной зоны	Слежение за перемещениями участников и определение сидящих людей		Слежение за перемещением выступающего и определение участников, сидящих в зоне кресел	
4	-	Поиск лиц сидящих участников			
3	-	Фотографирование лиц сидящих участников		Запись выступлений людей, сидящих в зале	
5	-	Наведение и запись перемещающегося участника		Наведение и запись выступления основного докладчика	
2	-	Запись общего вида на аудиторию			

Таблица 1: Функции видеокамер в различных режимах работы системы видеомониторинга интеллектуального зала.

В режиме «Наблюдение» выполняется обработка буферной зоны в видеокадре от Cam1 для детектирования входящих в зал людей. Как только обнаружен хотя бы один участник, автоматически запускается режим «Слежение за участниками», в котором производится обработка видеокадров от Cam1 для определения координат данного участника в осях зала и вычисления углов наклона и поворота камеры Cam5, которая ведет запись перемещения участника. Также осуществляется мониторинг зоны кресел для определения сидящих участников. Как только хотя бы один из участников сядет в любое кресло, система видеомониторинга переходит в режим «Регистрация участников», в котором задействованы сразу три видеокамеры: Cam1 осуществляет определение сидящих участников; Cam4 дополнительно проверяет наличие сидящих людей путем детекции лиц в зонах, закрепленных за номерами кресел, где выявлены участники; Cam3 фотографирует крупным планом лица обнаруженных сидящих участников.

Для основной части мероприятия определены два режима работы системы: «Выступление докладчика» и «Дискуссия». В первом режиме выполняется видеослежение за перемещением выступающего и детекция участников, сидящих в зоне кресел. При этом наведение и запись выступления основного докладчика выполняется посредством Cam5. В режиме «Дискуссия» используется Cam3 для записи выступающих, сидящих в зале. Во всех режимах работы, кроме «Наблюдение», Cam2 осуществляет запись общего вида интеллектуального зала. В таблице 1 не представлен режим «Ожидание», в котором камеры находятся в спящем состоянии, до тех пор, пока не будет включена, по крайней мере, одна группа света. Об изменении состояния активаторов, отвечающих за работу осветительного оборудования, система видеомониторинга

узнает из файла-журнала, в котором хранятся параметры, описывающие текущую ситуацию в интеллектуальном зале.

На рисунке 3 приведена укрупненная блок-схема алгоритма выбора текущего режима системы видеомониторинга. Вначале считывается файл-журнал, содержащий данные о состоянии групп света в зале, и если свет выключен, устанавливается режим «Ожидание». Напротив, при включенном свете проверяется наличие участников в зале, и если они отсутствуют, то система переходит в режим «Наблюдение». В случае нахождения участников, выполняется проверка, есть ли сидящие в зоне кресел, и если таковых нет, запускается режим «Слежение за участниками». В противном случае проверяется наличие человека в зоне выступления, и если он не обнаружен, то анализируется, в какой фазе находится мероприятие в данный момент. Если основная часть мероприятия (режимы «Выступление докладчика» и «Дискуссия») уже начата, то включается режим «Пассивная регистрация». Если же идет подготовительная стадия, то система переходит в режим «Активная регистрация». При наличии основного докладчика в зоне выступлений проверяется посредством метода аудиолокализации и массива микрофонов, есть ли говорящие среди зарегистрированных сидящих участников, и если нет, то устанавливается режим «Выступление», в обратном случае система переходит в режим «Дискуссия».

Для удобства наблюдения за режимами работы вся информация передается системе управления интеллектуальным залом, в диалоговом окне которой отображаются обнаруженные в процессе видеомониторинга объекты, в том числе участники, сидящие за конференц-столом, в зоне кресел, а также перемещающиеся в настоящий момент. Фотографии зарегистрированных участников отображаются в диалоговом окне в соответствии с занимаемыми местами в креслах.

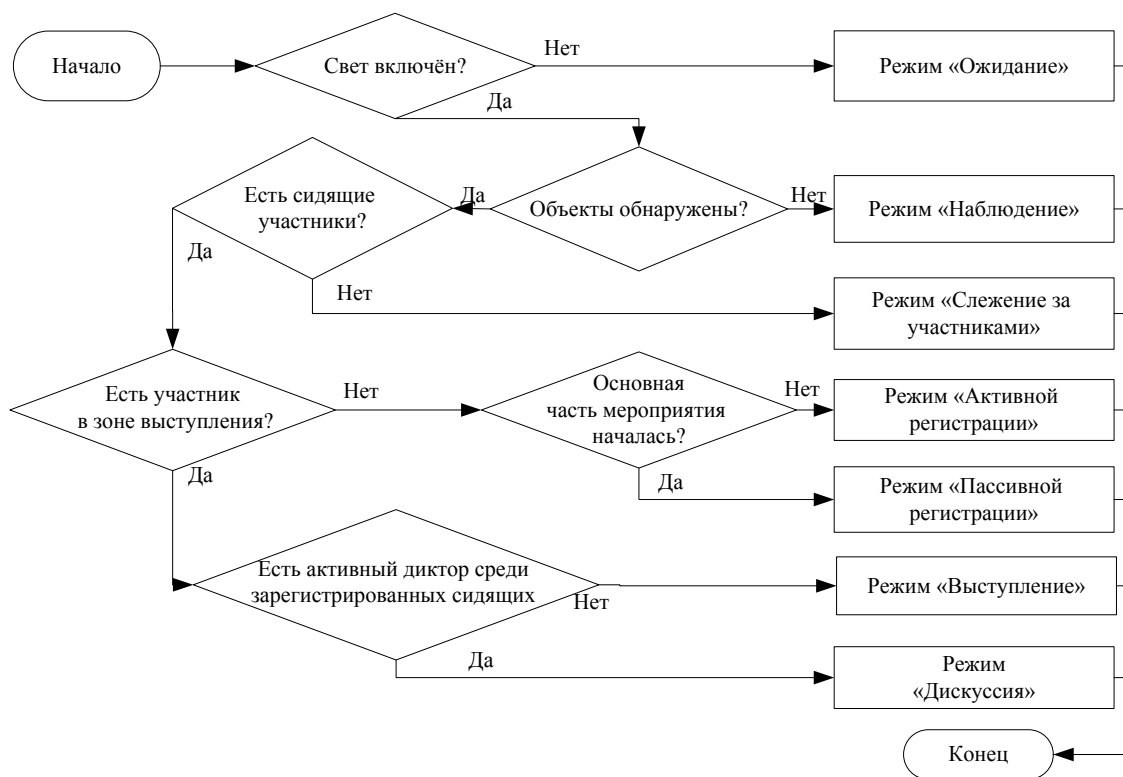


Рисунок 3: Блок-схема алгоритма выбора текущего режима работы системы видеомониторинга

3. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

В ходе серии предварительных экспериментов было организовано восемь небольших мероприятий и проведена регистрация участников, всего было зарегистрировано 106 участников. После ручной проверки было обнаружено 6 фотографий, на которых лица были детектированы неверно. Ошибки связаны, в основном, с изменениями положения головы участника в процессе фотографирования. Остальные 100 фотографий были использованы для подсчета скорости (времени) и качества фотодокументирования системы в режиме «Регистрация».

На рисунке 4 показан пример работы системы в режиме «Регистрация», где (1), (2) и (3) – кадры, выполненные видеокамерами Cam1, Cam4, Cam3, соответственно. С помощью Cam1 алгоритмом детекции движения [10] найдено 6 участников, сидящих в зоне кресел (один ложный объект был определен как участник). Номера этих кресел далее используются системой для анализа соответствующих

интересующих регионов на видеокдрах от Cam4. В данном примере анализ кадра (используется детектор лиц методом AdaBoost [11]) показал наличие четырех лиц в шести зонах, закрепленных за вышеуказанными креслами, в двух зонах лица не были найдены в связи с тем, что участники не следили за ходом мероприятия и не смотрели на лектора. Результирующий список номеров кресел затем передается на Cam3, которая последовательно устанавливается таким образом, чтобы крупным планом сфотографировать лицо каждого участника с необходимым качеством. Так как участники произвольно изменяют положения своего тела и головы, то при фотографировании производится запись ограниченной серия видеок кадров и проверяется наличие лица в нем методом [11]. Снимок сохраняется в базе данных регистрации только после обнаружения на нем графического объекта с параметрами, удовлетворяющими модели лица.



Рисунок 4: Пример работы системы в режиме «Регистрация».

Качество фотографий (кадры участников мероприятия крупным планом), сделанных системой в режиме «Регистрация», зависит, во многом, от точности наведения PTZ-камеры Cam3 и может оцениваться по двум критериям: (1) точность наведения видеокамеры на лицо; (2) отношение размера обнаруженного лица к размеру фотографии. Первый критерий показывает, насколько точно были выбраны углы поворота и наклона PTZ-камеры, а второй отвечает за правильность расчета коэффициента масштабирования. На каждой стадии обработки вычислялось относительное отклонение D центра лица от центра кадра (по двум осям). Для того чтобы избежать влияния изменения масштаба,

величина смещения определяется в процентах относительно полного размера видеокadra.

Относительный размер R обнаруженного в кадре лица вычисляется как отношение диагонали лица к диагонали кадра. В случае, если лицо в кадре не было обнаружено системой, то величина отклонения D выставляется равной 100%, а относительный размер лица - 0%. Кроме относительных размеров лица и отклонения лица, также вычислялась относительная и абсолютная площадь лица, S и A, соответственно. Кроме того, было оценено среднее время автоматической регистрации участника мероприятия. В таблице 2 приведены результаты оценки качества полученных системой фотографий.

R, %			D, %			S, %			A, Mpixels			T, c		
Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
24	47	35	0	28	11	6	23	13	0.03	0.10	0.06	0.03	11.59	1.08

Таблица 2: Количественная оценка качества фотографий участников мероприятий, сделанных системой в режиме «Регистрация»

4. ЗАКЛЮЧЕНИЕ

Слежение за подвижными объектами является сложной и важной задачей, решение которой позволит, в частности, автоматизировать проведение совещаний, телеконференций и других мероприятий. В разработанной системе видеомониторинга применены и исследуются технологии поиска и слежения за человеком, детекции позы и положения лица, позволяющие обеспечить информационное и техническое сопровождение научно-образовательных мероприятий, в том числе дистанционных. Разработанная многофункциональная система видеомониторинга работает в четырех основных режимах: (1) наблюдение, (2) слежение за участниками, (3) регистрация, (4) сопровождение мероприятия. На данный момент все режимы работы были протестированы по качественным показателям в ходе лекционных занятий и совещаний в интеллектуальном зале. Особое внимание уделено режиму автоматической регистрации участников. Было проанализировано влияние уровня освещения, числа участников и их местоположения на точность, время поиска лиц и последующее фотографирование участников. На следующем этапе исследования будут проанализированы математические и программные средства распознавания простейших жестов, выполняемых руками, и на их основе предполагается разработка подсистемы автоматического подсчета количества голосов, в мероприятиях, когда голосуют поднятием рук.

5. БЛАГОДАРНОСТИ

Данное исследование поддержано Министерством образования и науки РФ в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы (госконтракт № П2360), а также фондом РФФИ (гранты № 10-08-00199-а и № 09-07-91220-СТ_а).

6. СПИСОК ЛИТЕРАТУРЫ

- [1] V. Erol, Y., Li, An overview of technologies for e-meeting and e-lecture. Proc. *IEEE International Conference on Multimedia and Expo*, pp. 6–12, 2005.
- [2] S. Mukhopadhyay, B. Smith, Passive capture and structuring of lectures. *ACM Multimedia*, pp. 477–487, 1999.

- [3] J.A. Brotherton, J.R. Bhalodia, G.D. Abowd, Automated capture, integration, and visualization of multiple media streams. *IEEE Multimedia*, pp. 54–63, 1998.
- [4] M. Bianchi, AutoAuditorium: A fully automatic, multicamera system to televise auditorium presentations. *Proc. Joint DARPA/NIST Smart Spaces Tech. Workshop*, 1998.
- [5] Y. Rui, A., Gupta, J. Grudin, L. He, Automating lecture capture and broadcast: Technology and videography. *ACM Multimedia Systems Journal*, pp. 3–15, 2004.
- [6] Y. Kameda, S. Nishiguchi, M. Minoh, CARMUL: Concurrent automatic recording for multimedia lecture. *Proc. ICME*, 2003.
- [7] Q. Liu, D. Kimber, J. Foote, L. Wylcox, J. Boreczky, FLYSPEC: A multi-user video camera system with hybrid human and automatic control. *ACM Multimedia*, pp. 484–492, 2002.
- [8] D.S. Lee, B.J. Erol, Graham, J.J. Hull, N. Murata, Portable meeting recorder. *ACM Multimedia*, pp. 493–502, 2002.
- [9] P.M. Юсупов, А.Л. Ронжин. От умных приборов к интеллектуальному пространству // *Вестник Российской Академии Наук: научный и общественно-политический журнал*, Том 80, Вып. 1, С. 45–51, 2010.
- [10] J. Davis and G. Bradski, Real-time motion template gradients using Intel CVLib, Proc. *ICCV Workshop on Frame-Rate Vision*, Corfu, Greece, 1999.
- [11] P. Viola and M. J. Jones, Rapid Object Detection Using a Boosted Cascade of Simple Features, Proc. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, HI, USA, 2001.

ОБ АВТОРАХ

Ронжин Александр Леонидович – студент Санкт-Петербургского государственного университета аэрокосмического приборостроения (СПбГУАП), программист Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН), ronzhinal@iias.spb.su

Прищепа Мария Викторовна – студент Санкт-Петербургского государственного университета аэрокосмического приборостроения (СПбГУАП), программист Санкт-

Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН), prischepa@ias.spb.su
Будков Виктор Юрьевич – студент Санкт-Петербургского государственного университета аэрокосмического приборостроения (СПбГУАП), программист Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН), budkov@ias.spb.su
Карпов Алексей Анатольевич – к.т.н., с.н.с. лаборатории речевых и многомодальных интерфейсов Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН), karpov@ias.spb.su
Ронжин Андрей Леонидович – к.т.н., зав. лабораторией речевых и многомодальных интерфейсов Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН), ronzhin@ias.spb.su

A DISTRIBUTED VIDEO SURVEILLANCE SYSTEM FOR THE SMART SPACE

Abstract

This paper presents the video-monitoring system based on a distributed camera system, which is intended for automation of participants registration of scientific meetings carried out in a smart space model. The problem of detecting positions of participants in the developed smart room and their faces is solved by usage of a complex of wide-angle, omni-directional and PTZ video-cameras located on the walls and the ceiling of the room, as well as application of the face detection technology for photographing zones of probable locations of participants. The quality of photographing was evaluated using the accuracy measure of pointing of PTZ video-cameras and size of a detected face in the captured image. Deviation of the face region relatively to the frame center was estimated as 8% in average and the face region occupies at least 7% of the frame area. It was observed also that mean time of the registration increases, when the number of participants increases and the illumination level decreases. Further development of the system is aimed for automatic detection of an active speaker, discussion, voting and other key phases of the scientific events.

Keywords: video surveillance, computer vision, smart space, human detection and tracking