

МЕТОДИКА РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ НА ОСНОВЕ РАНДОМНЫХ ДЕРЕВЬЕВ В СИСТЕМАХ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ РАСШИРЕННОЙ РЕАЛЬНОСТИ

Финогеев А. Г., Четвергова М. В.

*ФГБОУ ВПО Пензенский государственный университет, Пенза
Пенза, Россия (440027, г. Пенза, ул. Красная, 40) alfin@mail.ru, marysha85@inbox.ru*

В работе рассматривается методика распознавания объектов на базе рандомных деревьев в системах автоматизированного проектирования расширенной реальности. Системы расширенной реальности позволяют дополнять окружающий физический мир искусственными объектами виртуального пространства. Для подобного рода систем одной из основных задач является задача выделения, распознавания и сопоставления объектов реального и виртуального пространств. Предложен алгоритм построения и обучения подсистемы распознавания на основе леса рандомных деревьев. Предложена принципиально новая структура леса рандомных деревьев: базовая и расширенная. Проведены сравнительные исследования доступных методов описания изображений по ключевым точкам, таких как SURF, SIFT и предлагаемых вариантов рандомных деревьев. Результаты эксперимента подтверждают эффективность предложенного метода.

Ключевые слова: расширенная реальность, распознавание изображений, особые точки, детектирование, дескриптор, рандомные деревья.

TECHNIQUE OF IMAGE RECOGNITION BASED ON RANDOM TREES IN AUTOMATED DESIGN ENGINEERING SYSTEMS OF AUGMENTED REALITY

Finogeev A. G., Chetvergova M. V.

*Penza State University, Penza
Penza, Russia (440027, Penza, Krasnaya st, 40) alfin@mail.ru, marysha85@inbox.ru*

This paper describe the technique of recognition of objects on the base of random trees for systems of augmented reality. Systems of augmented reality allow to add the environmental physical world artificial objects of the virtual space. One of the main task in this systems is the task of separation, recognition and comparison of objects of real and virtual space. Proposed algorithm of creation and training of a subsystem of recognition based of random trees. Essentially new structure of the wood of random trees is offered: basic and expanded. Comparative researches of available methods of the description of images on key points, such as by SURF, SIFT and offered options of random trees are carried out. Experimental results confirm efficiency of proposed method.

Key words: augmented reality, image recognition, point feature, detection, descriptor, random trees.

На современном этапе развития информационно-вычислительных технологий особое место занимают программно-инструментальные средства автоматизированного создания систем виртуальной и расширенной реальности. Если компоненты системы виртуальной реальности предназначены для «погружения» пользователей в виртуальную среду, создаваемую компьютером, то компоненты расширенной реальности «искусственно» изменяют окружающий мир посредством композиции реальных и виртуально синтезированных виртуальных объектов.

Появление и распространение мобильных коммуникационных устройств с достаточно большой вычислительной мощностью процессоров и видеокамерами высокого разрешения позволило перейти к технологиям получения и обработки видеопотока, с одновременным добавлением виртуальных текстовых и графических объектов на трехмерные сцены,

фактически в режиме реального времени. Это расширяет сферу применения методов искусственного интеллекта, технологий трехмерного моделирования в направлении создания новых принципов и интерфейсов человеко-машинного взаимодействия на основе технологии дополненной или расширенной реальности (augmented reality). Пользователь с ноутбуком, смартфоном или планшетом может непосредственно участвовать в процессе синтеза и визуализации виртуальных объектов расширенной реальности, с возможностью встраивания в сцены реального мира и управления ими (рис. 1).

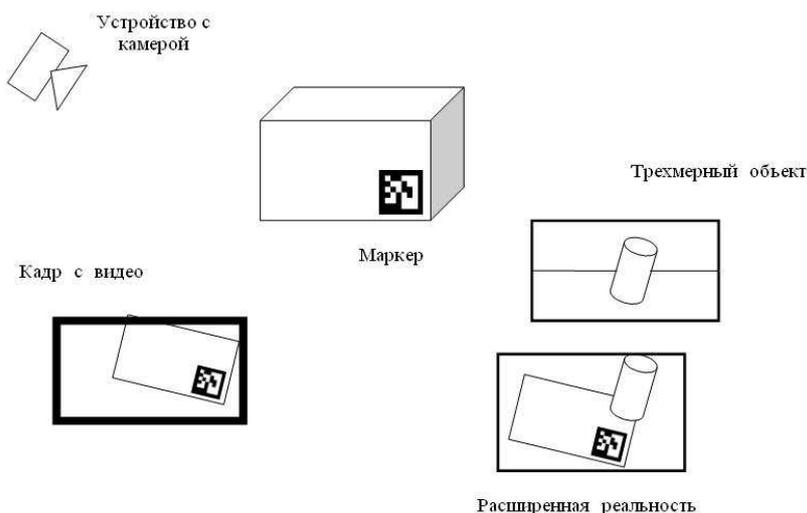


Рис. 1. Схема системы расширенной реальности

Развитие аппаратных средств реализации данных технологий выдвигает на первый план разработку новых моделей, методик и алгоритмов автоматизированного проектирования виртуальных компонент для систем расширенной реальности, способов их привязки к объектам реального мира и технологий интерактивного манипулирования с использованием возможностей беспроводной мобильной телекоммуникационной связи.

В общем случае для создания системы расширенной реальности требуются следующие компоненты: аппаратное, информационное, алгоритмическое и программное обеспечения, а также маркер. В качестве маркера может выступать искусственно создаваемые планарные изображения или изображения реальных предметов, которые служат для привязки 3D моделей.

К аппаратному обеспечению необходимо отнести или видео, или web-камеру, или мобильное устройство с камерой (в дальнейшем будем называть видеоустройство), или планшетный компьютер.

Информационное обеспечение включает несколько «баз данных»: с заранее подготовленным битональным представлением планарных маркеров, базу данных

информационных ресурсов, представленную видео-файлами, трехмерными объектами виртуального пространства или иного рода данными.

К программному обеспечению относятся программы, реализующие следующие алгоритмы: получение и обработка входного видеосигнала с его последующей идентификацией, распознавание элемента расширенной реальности, сопоставление маркера информационного объекта, определение положения камеры в пространстве, оценка общего пространственного расположения наблюдаемых объектов и т.д.

Как видно первоочередной и сложной задачей в процессе автоматизированного проектирования виртуальных компонент расширенной реальности с их привязкой к объектам реальной сцены является задача выделения и распознавания в видеопотоке с камеры определенных частей объектов реальной сцены. Рассмотрим данную задачу подробнее.

Распознавание элементов изображений реальных объектов в видеопотоке

Для соединения элементов виртуального и реального мира на экране дисплея одной из актуальных научно-исследовательских проблем является решение задачи выделения и распознавания в видеопотоке элементов изображения. Данные задачи сводятся к определению точек привязки виртуальных объектов к объектам реального окружения по их единичным кадрам – изображениям, получаемым с камеры. При практической реализации сложной задачи распознавания в режиме реального времени решаются вопросы формирования, получения, хранения, передачи и воспроизведения компонентов изображений техническими средствами с соответствующим математическим и программным обеспечением [1]. Вычислительные и коммуникационные возможности беспроводных мобильных средств связи позволяют использовать более совершенные программные средства анализа изображений, что способствует развитию новых математических и алгоритмических методов решения задачи обработки и распознавания в режиме реального времени. Несмотря на то, что создано достаточно большое количество методик и алгоритмов распознавания, потребность в их совершенствовании остается, поскольку распознавание реальных объектов в видеопотоке зависит от многих факторов: освещенность сцены, сложность текстур поверхности объектов распознавания, зашумленность, параметры видеокамеры и т.д.

Эмпирически известно, что любое изображение содержит сложную и трудно извлекаемую структурированную информацию о наблюдаемой сцене. В нашем случае требуется метод, который позволяет извлекать из потока видеоданных, получаемого в

режиме реального времени с камеры мобильного устройства, структурированную информацию об особых частях объектов или о нанесенных на них маркерах для их идентификации и распознавания.

В процессе распознавания объектов сцены одной из наиболее значимых проблем остается задача сопоставления выделенного из фотоснимка или видеокadra изображения объекта с его эталонными образцами, хранящимися в базе данных. Существующие средства распознавания изображений используют различные способы сопоставления объектов распознавания и их образов в специализированных базах изображений. Однако основной методикой решения задачи является установление соответствия между особыми или опорными точками исходного изображения и его образами. Особая точка представляет собой наиболее простой геометрический элемент дискретного представления математической функции описания объекта распознавания.

В общем случае, методика сопоставления изображения (xL_I, yL_I) к его эталонному изображению (xR_I, yR_I) включает следующие шаги:

- 1) Детектирование особых (опорных) точек изображения (xL_I, yL_I) ;
- 2) Составление описания опорных точек или определение дескриптора;
- 3) Обучение и заполнение «эталонной» базы данных об опорных точках изображения;
- 4) Составление описания объектов на изображении (xR_I, yR_I) ;
- 5) Поиск связей или расстояния между дескрипторами особых точек на сравниваемых изображениях. Предполагается, что если между точками в паре расстояние по дескрипторам маленькое, то пара более «качественная» и, в первую очередь, метод проверяет подобные пары;
- б) Синтез модели преобразования изображений, с помощью которой из одного изображения можно получить другое.

Особая точка должна иметь ряд признаков, существенно отличающих её от множества соседних с ней точек изображения объекта сцены. В частности, совокупность особых точек может представлять некоторые характерные области на изображениях, например, вершины геометрических фигур, края плоскостей, углы, небольшие окружности и круги, перепады яркости или контрастности, резкие цветовые переходы и т.п.

Определение особых точек на изображениях реальных объектов

Для определения таких точек будем использовать понятие окрестности. Тогда особой (опорной) точкой изображения будем называть точку p_i , окрестность которой $O(p_i)$ можно отличить от окрестности $O'(p_j)$ любой другой особой точки изображения p_j . Процесс

выделения этой точки на кадре называется *детектированием*, а программа, реализующая данную функцию, – *детектором*. В настоящее время создано большое количество детекторов особых точек изображений с разными подходами к формированию пиксельной оценочной функции поиска и выделения точек.

Для реализации нашей методики был выбран детектор Ши-Томаси [2], который в свою очередь базируется на угловом детекторе Харриса [3] и учитывает аффинные искажения изображения окрестности особой точки.

Для особой точки, найденной с помощью данного детектора, необходимо рассчитать *вектор-дескриптор*, описывающий структуру окрестности точки, выделяющей её из остального множества особых точек. При этом дескриптор рассчитывается таким образом, чтобы его значение было инвариантным по отношению к требуемым аффинным преобразованиям изображения.

В нашем случае дескриптор определяется как:

$$d_n = (f_{n,1}, \dots, f_{n,j}), \quad (1)$$

где $f_{n,j}$ – параметры, отвечающие следующим условиям: специфичность, локальность, устойчивость, простота в вычислении и т.д.

Множество дескрипторов определяет модель изображения, которая в дальнейшем используется для сопоставления изображений с эталонными образцами из базы данных. Решение о том, являются ли сопоставляемые изображения аналогами, принимается на основе анализа векторов дескрипторов исходного объекта и векторов дескрипторов его образов в базе данных.

Сопоставление векторов осуществляется в два этапа. На первом этапе устанавливаются пары наиболее близких дескрипторов. При этом степень близости рассчитывается как расстояние в пространстве изображений:

$$M = \sqrt{\sum (f(xL_1, yL_1) - f(xR_1, yR_1))^2}, \quad (2)$$

где (xL_1, yL_1) , (xR_1, yR_1) – сравниваемые изображения.

На втором этапе выбираются пары наиболее близких векторов в пространстве изображений и принимается решение о соответствии исходного изображения его образу, путем сравнения расстояний относительно некоторого порогового значения.

В настоящее время существует ряд методик описания изображений по особым точкам. К наиболее известным относятся:

1. Алгоритм SIFT (Scale Invariant Feature Transform) [4]. Один из наиболее часто используемых алгоритмов описания особых точек. Точки, полученные с помощью

алгоритма, инвариантны к масштабированию и поворотам изображения, устойчивы к изменениям освещения, шумам и изменениям позиции наблюдателя.

2. Алгоритм SURF (Speeded Up Robust Features) [5]. Методика основана на поиске особых точек и создание дескрипторов, инвариантных к масштабированию и вращению с помощью матрицы Гессе. При этом для каждой точки считается градиент максимального изменения яркости и коэффициент масштабирования по матрице Гессе.

3. Метод RIFF (Rotation Invariant Fast Features) [6]. В основу метода положено радиальное и тангенциальное разложение гистограмм градиента и последующая обработка по кольцам. Дескриптор также инвариантен к масштабированию, вращению и изменению освещенности.

Проведенные экспериментальные исследования данных подходов при анализе и идентификации элементов изображений, получаемых от 5 мегапиксельной камеры смартфона при движении наблюдателя и условиях изменения освещенности в связи с погодными условиями были определены основные недостатки данных методик. В частности, основными недостатками методов SIFT и SURF являются нечеткое выделение объекта относительно фона и низкий процент правильного распознавания элементов изображений объектов без ярко выраженной текстуры. Метод RIFF показал плохие результаты, вызываемые размытием изображения вследствие движения наблюдателя с камерой относительно неподвижных объектов.

Поэтому для выделения особых точек и идентификации реального объекта с целью последующей привязки к нему синтезированного виртуального текстового, графического или звукового объекта в нашем случае был выбран подход с использованием рандомных деревьев (Random Forest – RF) для обучения системы распознавания.

Методика распознавания объектов на базе рандомных деревьев

Основная идея заключается в обучении системы распознавания на базе собранной статистики распределения решений путем построения леса рандомных деревьев. На этапе детектирования предполагаемая особая точка как бы «перебрасывается» по дереву вывода (рис. 1), которое строится для каждого изображения распознаваемого объекта, снимаемого с разных ракурсов и при различных условиях. В каждом узле рассчитывается значение вероятности того, что данная особая точка относится к одной из уже известных, определенных ранее.

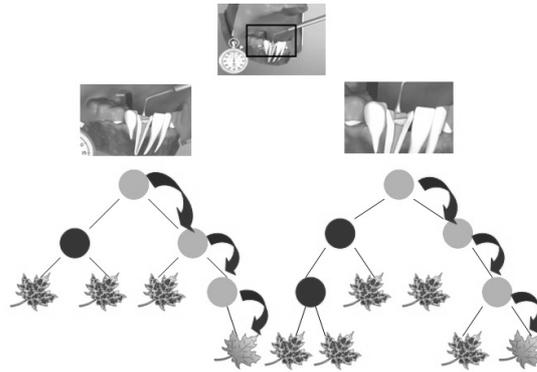


Рис. 2. Пример случайного дерева

Процесс выполняется для множества деревьев, причем итоговая вероятность накапливается по каждой предполагаемой гипотезе о принадлежности особой точки к множеству известных точек распознаваемого объекта. Наибольшая вероятность покажет связь изображения исходного объекта с эталонными образцами.

Рассмотрим подробнее предлагаемую методику. Она относится к классу алгоритмов выделения окрестности вокруг особой точки для ее фиксации на изображении объекта. Пусть (x, y) – координаты некоторой точки p_i на изображении объекта, при этом $p_i \in \{P\}$, где P – множество особых точек. В основе методики лежит идея классификатора Байеса.

Пусть $I(x_i, y_i)$ – цветное значение пикселя в произвольной точке $p_i(x_i, y_i)$, причем $0 \leq x_i < W_i$ и $0 \leq y_i < H_i$, где W_i и H_i , соответственно, ширина и высота изображения в пикселях.

Для снижения размерности изображение первоначально приводится к монотонному на базе шкалы градаций серого. Для того чтобы определить и зафиксировать окрестность особой точки, необходимо работать по направлениям градиентов яркости вокруг нее.

В [7] показано, что именно случайный выбор разностей градиентов яркости для произвольного окружения точки дает устойчивый результат.

Величина градиента вычисляется по формуле:

$$G(x, y) = \sqrt{(I(x+1, y) - I(x-1, y))^2 + (I(x, y+1) - I(x, y-1))^2} \quad (3)$$

Перед выбором разности необходимо определить множество векторов D_{xy} , каждый из которых представляет значения разностей для каждой точки $p_i \in \{P\}$.

Пусть $T(D_{xy})$ – множество тестов для определения разности градиента на множестве точек P . Каждый тест возвращает битовое значение 0 или 1 при выполнении условия в первой p_1 и второй p_2 точках окрестности в каждом направлении:

$$T(D_{xy}) = \begin{cases} 0, & \text{если } I(D_{xy}^n(p_1)) < I(D_{xy}^n(p_2)) \\ 1, & \text{если } I(D_{xy}^n(p_1)) > I(D_{xy}^n(p_2)) \end{cases}, \quad \text{где } n \in [0, D_{xy}) \quad (4)$$

Результатом выполнения множества тестов будет битональная маска направления градиентов вокруг особой точки.

Маска представляет собой бинарное дерево решений, в котором каждый элемент множества D_{xy} рассматривается как узел дерева. Условие перехода «вправо» или «влево» определяется тем, какое значение возвращает тест $T(D_{xy})$ при выполнении условия (1) в первой и второй точке окрестности $p_i(x_i, y_i)$.

Каждая вершина дерева является счетчиком. Переход по дереву решений происходит с увеличением счетчика на 1 согласно случайным разностям градиентов D_{xy} . В итоге, после определения множества образов окрестности одной и той же особой точки для разных изображений одного объекта, на множестве вершин дерева получим распределение вероятности соответствия исходного изображения его образу.

Перебор производится для всех деревьев в лесу, соответствующих множеству эталонных образов исходного объекта, а итоговая вероятность накапливается по каждому распознаванию. Максимум вероятности для множества деревьев показывает связь изображения контура исходного объекта и множества его образов, полученных на этапе обучения при различных условия съемки.

Таким образом, каждое дерево решений строится для сравнения исходного объекта с конкретным эталонным образом из базы данных и представляет собой лес решений для множества образов. В общем случае результатом построения леса является множество F – деревьев, с массивом P вершин размерностью $2^{|D_{xy}|}$, где хранятся вероятности распределения решений для конкретного дерева. Для снижения размерности и получения независимых результатов при распознавании можно выбрать случайное подмножество деревьев F_i , которое будет включать в себя подмножество случайно выбранных D_{xy} .

Обучение подсистемы распознавания на базе рандомных деревьев

На этапе обучения для каждой гипотетической точки p^* найдем индекс вершины для каждого вектора $D_{xy} \in F_i$. Тогда P_a представляет подмножество особых точек, для которых рассчитывается итоговая вероятность по вершинам дерева для каждой новой точки из обучаемой коллекции. В итоге получим вероятность того, что выбранная точка p^* соответствует какой либо точке $p_i \in \{P\}$, где P_a^i – вероятность, что $p^* = p^i$. Пороговое число вероятностей соответствия близких к 100 % для множества особых точек изображения исходного объекта $p_i^* = p_j^i$ подтвердит гипотезу, что данное изображение объекта соответствует некоторому эталонному образу, который есть в базе данных и, следовательно, это изображение может считаться также образом конкретного объекта. Таким образом, после каждого правильного распознавания и в случае, если итоговая вероятность соответствия всех особых точек исходного объекта конкретному образу меньше 100 %, распознанный

объект считается эталонным образом, полученным при новых условиях съемки, и может быть добавлен в базу данных, что и является обучением системы распознавания. При следующем распознавании для него строится новое дерево решений.

Однако при таком подходе всегда существует вероятность того, что множество особых точек изображения исходного объекта всегда будет показывать их соответствие множеству особых точек образов данного объекта из обучающей выборки. Поэтому необходимо выполнить соответствующую фильтрацию по заданному априори пороговому значению вероятности соответствия. Для определения гомографии между двумя множествами особых точек изображений, соответствующих одной и той же точке трехмерного объекта, можно использовать метод RANSAC [8].

Поскольку на начальном этапе обучения отсутствует большое число образов окрестностей особых точек для множества изображений объекта, то можно синтезировать множество синтетических изображений конкретного объекта, применяя различные аффинные преобразования (масштабирование, перенос, поворот) и случайный набор шумов для искажения изображения.

Экспериментальное исследование методик распознавания

В результате экспериментального исследования данного подхода было установлено, что качество распознавания изображений объекта зависит от числа выбранных особых точек на изображениях (xL_l, yL_l) и (xR_l, yR_l) и множества вершин (глубины) рандомного дерева. Было проведено исследование предложенной методики для числа особых точек изображения 200, 500 и 1000 и для двух вариантов лесов рандомных деревьев:

1. Вариант RF(базовый) – для леса из 30 деревьев решений с глубиной 8 вершин.
2. Вариант RF(расширенный) – для леса из 30 деревьев решений с глубиной 12 вершин.

Также было проведено сравнение результатов распознавания с методами SURF и SIFT. Результаты аналитического сравнения подходов приведены в таблице 1.

Таблица 1. Аналитическое сравнение результатов распознавания

Метод	Кол-во точек	Качество распознавания, %	Объем требуемой памяти, Мб	Время сопоставления, сек
SURF	200	74	0,05	~ 0,5-1
SIFT		63	0,98	~ 0,5-1
RF (base)		68	6,1	~ 0,4-0,8

RF (extended)		88	18,3	~0,8-1,2
SURF	500	76	0,86	~0,8-1,3
SIFT		78	0,93	~0,8-1,3
RF (base)		80	12,5	~1-1,5
RF (extended)		92	26,7	~1-2
SURF	1000	75	1,2	~1-1,2
SIFT		78	1,2	~1-1,3
RF (base)		82	25,6	~2
RF (extended)		94	56,9	~2

Таким образом, результаты сравнения наиболее часто используемых алгоритмов SIFT и SURF с предложенным подходом показывают, что использование методики распознавания на базе леса случайных деревьев дают лучшие оценки качества распознавания при более высоких требованиях к оперативной памяти и большем, но вполне приемлемом времени работы системы в зависимости от размерности дерева. Одним из основных отличий предложенной структуры системы распознавания на базе случайных деревьев является то, что методы SIFT и SURF являются запатентованными.

Заключение

Результаты проведенных экспериментальных исследований предложенных вариантов случайных деревьев демонстрируют вполне допустимое качество работы даже на объектах с небольшим количеством особых точек (около 200). При этом на изображениях с очень большим количеством особых точек качество и время распознавания являются достаточно высокими, что позволяет использовать данную методику для систем, работающих в режиме реального времени и на мобильных устройствах. Однако большой объем оперативной памяти, затрачиваемой при работе алгоритма, является критичным для реализации алгоритма на мобильные устройства.

В дальнейшем планируется оптимизировать предлагаемую имплементацию леса случайных деревьев для использования на мобильных платформах различных версий.

Список литературы

1. Васильев В. Н. Математические методы и алгоритмическое обеспечение анализа и распознавания изображений в информационно-телекоммуникационных системах / В. Н. Васильев, И. П. Гуров, А. С. Потапов // Всероссийский конкурс обзорно-аналитических статей по приоритетному направлению «Информационно-телекоммуникационные системы», Режим доступа: <http://www.itc.edu.ru/itkonkurs2008/>

2. Shi J. Good features to track / J. Shi, T. Kanade // IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'94), IEEE Computer Society, Sietl – 1994.
3. Harris C. A combine corner and edge detector / C. Harris, M. Stephens // In Fourth Alvey Vision Conference, Manchester, UK, pp. 147- 151.
4. Lowe D. Object Recognition from Local Scale-Invariant Features / David G. Lowe // *Proceedings of the International Conference on Computer Vision*. 2. 1999, pp. 1150–1157.
5. Herbert B. SURF: Speeded Up Robust Features / B. Herbert, Andreas Ess, Tinne Tuytelaars, Luc Van Gool // *Computer Vision and Image Understanding (CVIU)*, Vol. 110, No. 3, 2008, pp. 346–359.
6. Takacs G. Unified Real-Time Tracking and Recognition with Rotation-Invariant Fast Features / G. Takacs, V. Chandrasekhar, S. Tsai, D. Chen, R. Grzeszczuk, B. Girod // [Электронный ресурс] CVPR2010_RIFF.pdf
7. Calonder M. BRIEF: Binary Robust Independent Elementary Features / M. Calonder, V. Lepetit, P. Fua // [Электронный ресурс]: <http://cvlab.epfl.ch/~lepetit/>
8. Hartley R. Multiple View Geometry In Computer Vision / R. Hartley, A. Zisserman. – 2nd edition: Cambridge University Press, 2003. p. 670.

Рецензенты:

Бершадский Александр Моисеевич, д.т.н., профессор, зав. кафедрой САПР, ФГОУ ВПО «Пензенский государственный университет», г. Пенза.

Бождай Александр Сергеевич, д.т.н., профессор кафедры САПР, ФГОУ ВПО «Пензенский государственный университет», г. Пенза.