

УДК 004.93'1; 004.932

НЕКОТОРЫЕ АСПЕКТЫ ПОСТРОЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ КОМПЛЕКСОВ ВИДЕОНАБЛЮДЕНИЯ

Саниев К.Б., к.т.н., ведущий научный сотрудник ЗАО «МНИТИ» - Московский научно-исследовательский телевизионный институт, kbsaniev@concord.ru

Ключевые слова:

распознавание изображений, обнаружение, классификация и идентификация объектов по изображениям, биометрическая идентификация, унифицированный алгоритм информационного анализа изображений.

Введение

Развитие микроэлектроники и вычислительной техники привело, в частности, к широкому распространению разнообразных комплексов видеонаблюдения – от домофонов и видеорегистраторов до систем дистанционного мониторинга обстановки, средств биометрической идентификации личности и видеополиграфа.

Практика разработки и эксплуатации этих устройств показывает, что одной из основных проблем здесь является недостаточное качество автоматического распознавания изображений, в частности, таких «типовых» объектов, как люди (лица людей) и транспортные средства (автономера), особенно, в «естественных» условиях их деятельности.

Реальные трудности обеспечения приемлемого качества автоматического распознавания оптических изображений обусловлены, главным образом, заранее неизвестным разнообразием возможных изображений наблюдаемых объектов. Эта априорная неопределённость, незаметная для человека, проявляется для распознающего «интеллектуального» автомата в непредсказуемых вариациях пространственно-яркостных распределений видеосигнала, которые возникают вследствие изменчивости условий наблюдения объектов по освещённости, ракурсам и масштабам. Дополнительная неопределённость вносится разнообразием и изменчивостью фона (подстилающих поверхностей), маскирующей раскраской и собственной поведенческой активностью объектов. Для объектов со сложной структурой, в том числе биологических, таких, как лица людей, множество возможных изображений ещё более расширяется вследствие их собственного поведения, поскольку при этом возникают изменения их видимой структуры (например, мимика лица).

В этой связи по-прежнему актуальными являются исследования более эффективных методов и средств автоматического распознавания изображений [1, 2, 6]. В области цифровой обработки сигналов это означает

Обсуждаются методы и средства повышения качества автоматического распознавания изображений в комплексах видеонаблюдения. Предлагается технология «Имаджер» информационного анализа сигналов, в которой каждое изображение описывается множеством собственных структурных примитивов – яркостных пятен, соответствующих наблюдаемым объектам или их частям. При этом примитивы статического изображения содержат информацию об особенностях внешнего облика объектов, а примитивы видеоактивности – информацию об особенностях поведения объектов. В этом случае появляется возможность классификации априорно неопределённых событий и объектов по характеристикам этих примитивов, измеренным на анализируемом изображении.

поиск и реализацию таких операций обнаружения объектов, отслеживания их пространственного положения, классификации и идентификации объектов, которые бы выявляли и использовали в качестве признаков для распознавания более инвариантные, более устойчивые к вариациям параметры видеосигналов.

Технологии информационного анализа изображений

Изображениями обычно называют сигналы вида $I(x, y, z, t)$ пространственно-временных распределений, получаемые как проекции различных физических полей $E_{\Psi}(x, y, z, t)$, где I – интенсивность, а (x, y, z, t) – пространственные и временная координаты. Плоские изображения $I(x, y, t)$ или видеокadres, при дискретном представлении, могут быть определены как множества F пространственно-временных отсчетов $\{I_{i,j,k}\}$, где $i = 0, W, j = 0, H, k = 0, T$.

Распознавание изображений подразумевает выполнение информационного анализа исходного сигнала $I(x, y, t)$ или $F: \{I_{i,j,k}\}$ с целью обнаружения объектов, слежения за объектами, классификации и/или идентификации объектов. В общем виде результатом операций распознавания $RECOGN$ является описание исходного изображения списком Θ объектов S_{η} , каждый из которых характеризуется и описывается набором свойств-признаков $P: \{p_1, \dots, p_{\mu}\}$

$$I(x, y, t), \{I_{i,j,k}\} \xrightarrow{RECOGN} \Theta: \{S_1(P), \dots, S_{\eta}(P)\}. \quad (1)$$

Описания S_{η} относятся к фрагментам исходного изображения, т.е. к подмножествам f_{μ} , являющимися изображениями объектов и/или их частей и составляющими вкпе полное множество $F: \{I_{i,j,k}\} = \{f_{\mu}\}$ отсчетов исходного сигнала.

Как известно (например, из [7]), при решении задач автоматического распознавания первая возникающая проблема заключается в необходимости построения эффективного описания S_{η} , т.е. выбора информативных свойств или при-

знаков $P: \{p_1, \dots, p_\mu\}$, необходимых и достаточных для достоверного распознавания возможных изображений f_μ контролируемых классов объектов. В существующей практике автоматического распознавания основой для формирования априорных эталонных описаний S_η служат различные эвристические модели объектов, закладываемые конструкторами конкретных систем. При этом, как отмечается в [3], «решение каждого нового типа задач {расознавания видеоданных} требует проведения заново всей работы», вследствие отсутствия «для исследуемых реальных ситуаций или объектов сколько-нибудь адекватных математических моделей, на базе которых можно было бы вести расчеты и получать количественные или качественные выводы».

В этой связи одна из возможностей развития технологий автоматического распознавания состоит в использовании в качестве алфавита информационных моделей объектов собственных структурных элементов изображений $I(x, y, z, t)$, так называемых примитивов Ω , которые представляют собой яркие пятна, т.е. фрагменты $\Omega = \mathcal{G}(\bar{x}, \bar{y}, \bar{z}, \bar{t}) \in I(x, y, z, t)$, физически формируемые отражающими поверхностями реальных предметов [8]. Наиболее эффективной операцией выявления собственных примитивов изображений оказывается детектирование их как компактных групп пикселей по пространственно-яркостной связности. При этом изображения объектов f_μ будут состоять из подмножеств или групп, примитивов-пятен $\{\Omega_1, \dots, \Omega_k\}$, а свойства-признаки $P: \{p_1, \dots, p_\mu\}$ объектов будут формироваться различными комбинациями индивидуальных пространственных, яркостных и временных характеристик $F: \{f_1, \dots, f_\mu\}$ детектированных примитивов $\Omega = \mathcal{G}(\bar{x}, \bar{y}, \bar{z}, \bar{t})$

$$p_\lambda = \sum_v C_v \cdot f_v^\mu$$

Данная технология распознавания, называемая «Имаджер», по сравнению с традиционной методологией распознавания образов даёт применительно к анализу слабо формализованных сигналов, в частности изображений, следующие основные преимущества:

- в результате распознавания получают не ограниченный информационно ответ, наблюдаются или нет заранее определённые объекты, а формируется полное описание анализируемой сцены, поскольку множество обнаруженных таким образом пятен покрывает все анализируемое изображение и включает информацию о всех наблюдаемых в сцене «объектах»

$$\bigcap \Omega = \emptyset; \bigcup \Omega = F: \{I_{i,j,k}\}; \quad (2)$$

- появляется принципиальная возможность построения систем распознавания изображений с унифицированным алгоритмом [9] информационного анализа сигнала изображения, состоящего из последовательности базовых операций обнаружения собственных примитивов $\Omega = \mathcal{G}(\bar{x}, \bar{y}, \bar{z}, \bar{t}) \in I(x, y, z, t)$, измерения их характеристик $F: \{f_1, \dots, f_\mu\}$, кластеризации/(логической фильтрации) групп примитивов $\{\Omega_1, \dots, \Omega_k\}$ как изображений f_μ предполагаемых объектов и, в итоге, классификации/идентификации фрагментарных изображений f_μ объектов по диапазонам значений характеристик

$$p_\lambda = \sum_v C_v \cdot f_v^\mu \text{ составляющих их примитивов.}$$

В качестве примера функционирования такого алгоритма на рис. 1 и 2 показаны рабочие результаты функционирования комплекса видеонаблюдения «Имаджер2008» [10] по обнаружению автономеров и лиц людей

Автономера здесь обнаруживаются как объекты, состоящие из группы пятен с близкими значениями геометрических характеристик. Достоверность обнаружения номера при этом не зависит от его расположения в кадре и не зависит от структуры номера. Распознавание автономера как кода далее выполняется идентификацией отдельных пятен как символов арабских цифр, кириллицы и латиницы по параметрам их формы.

Лица людей обнаруживаются как объекты, представляющие собой группу пятен, образующих своими силуэтами заданную «маску». В качестве опорных элементов «маски» приняты реальные биометрические параметры лица – рот, нос, глаза.

Программно-аппаратный комплекс «Имаджер2008» предназначен для сбора первичных тревожных видеоданных, предварительной классификации тревожных событий и идентификации тревожных объектов. Комплекс состоит из 4-х цветных PTZ-видеокамер, устройства ввода видеосигналов по интерфейсу USB, ноутбука с процессором Core 2 Duo и операционной системой WindowsXP, программного комплекса обнаружения тревог и программного комплекса распознавания объектов классов «Лица» и «Автономера». Эти прикладные программы выполняют одновременный видеоконтроль 4-х разных видеосцен, автоматическое формирование базы данных тревожных видеок кадров, выявляемых по задаваемым уровням видеоактивности, и формирование базы данных результатов распознавания тревожных видеок кадров на предмет обнаружения в них и идентификации лиц людей и автономеров.



Рис. 1. Примеры обнаружения автономеров.

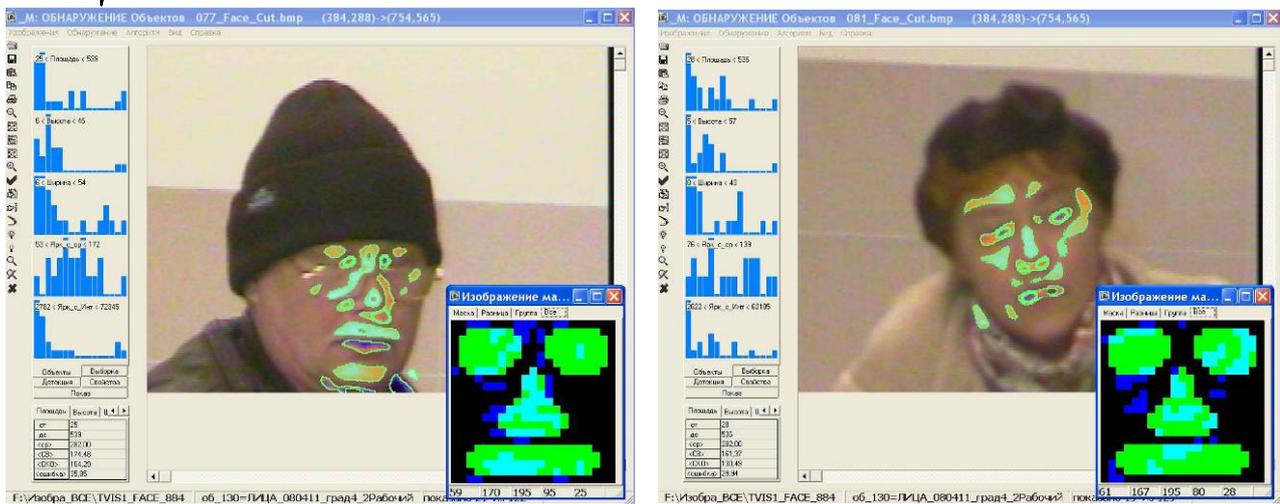


Рис. 2. Примеры обнаружения лиц.

При этом производится анализ последовательностей видеок кадров и распознавание собственно тревожных событий. Обеспечивается также возможность одновременного заполнения и редактирования баз данных контролируемых объектов (людей, автотранспортных средств), которые включают и эталонные данные, используемые при идентификации этих объектов. Имеется встроенный сервис интерактивного статистического анализа и выборки результатов распознавания с автоматическим формированием отчета с видеофото в виде документа Microsoft Word.

Развитие технологий распознавания изображений, конечно же, должно включать и анализ временной динамики изображений, которая отражает на только изменения условий наблюдения сцены, но и отражает поведение составляющих её объектов. Данную область исследований в последнее время всё чаще называют видеоаналитикой [4, 5], а их основу составляют операции выявления движений объектов, которые детектируются традиционными методами распознавания образов на основе априорных моделей фонов и объектов [1, 6].

Перспективы повышения эффективности видеоаналитики, на наш взгляд, связаны с выявлением всей видеоактивности (а не только частного случая движения) и анализом её структуры. Для плоских изображений видеоактивность $\tilde{I}(x, y, t)$ представляет собой в общем случае преобразованную (фильтрованную) разницу двух соседних во времени изображений $I_1(x, y, t_1)$ и $I_2(x, y, t_2)$, приближённо первую производную по времени от видеосигнала

$$\tilde{I}(x, y, t) = \Delta I(x, y, \Delta t) = I_2(x, y, t_2) - I_1(x, y, t_1) = \tilde{F} : \{ \tilde{I}_{i,j,k} \}. \quad (3)$$

Собственные структурные примитивы видеоактивности $\tilde{I}(x, y, t)$, называемые в дальнейшем динамическими примитивами $\tilde{Q} = \tilde{Q}(\bar{x}, \bar{y}, \bar{z}, \bar{t})$

$$\cap \tilde{Q} = \emptyset; \cup \tilde{Q} = \tilde{F} : \{ \tilde{I}_{i,j,k} \}, \quad (4)$$

всегда соответствуют изменениям состояния объектов и/или их функциональных частей, поскольку физически порождаются поведенческой активностью объектов, составляющих анализируемую сцену, и содержат максимум информации о поведении объектов.

Индивидуальное отслеживание динамических примитивов $\tilde{Q}_\tau = \tilde{Q}(\bar{x}, \bar{y}, \bar{z}, \bar{t})$ видеоактивности позволяет вычислить как локальные траекторные параметры

$$\tilde{P}^{TL} : \{ \tilde{p}_1^{TL}, \dots, \tilde{p}_\mu^{TL} \}$$

объектов: положение в поле зрения, скорость, направление движения и т.д., так и интегральные траекторные параметры $\tilde{P}^{TG} : \{ \tilde{p}_1^{TG}, \dots, \tilde{p}_\mu^{TG} \}$: начальное и конечные положения, пройденное расстояние, среднюю скорость и т.п., характеризующие поведение объектов. Измеряемые при отслеживании динамических примитивов \tilde{Q} их геометрические характеристики

$$\tilde{P}^g : \{ \tilde{p}_1^g, \dots, \tilde{p}_\mu^g \}:$$

размеры, пропорции, площадь, показатели формы, содержат сопряжённую информацию об индивидуальных особенностях внешнего вида этих объектов.

На рис. 3 показаны результаты функционирования программного макета универсального детектора видеоактивности, разрабатываемого в рамках технологии «Имаджер».

Получаемые таким образом количественные данные $\tilde{P}^{TL} : \{ \tilde{p}_1^{TL}, \dots, \tilde{p}_\mu^{TL} \}$, $\tilde{P}^{TG} : \{ \tilde{p}_1^{TG}, \dots, \tilde{p}_\mu^{TG} \}$ и $\tilde{P}^g : \{ \tilde{p}_1^g, \dots, \tilde{p}_\mu^g \}$

о характеристиках видеоактивности дают возможность существенно повысить достоверность классификации обнаруженных предположительных объектов, в том числе и независимо от априорных данных, и, кроме того, позволяют достоверно и максимально сжато интерпретировать наблюдаемую обстановку в целом, как события появления, исчезновения, активности и движения объектов. Например:

- «вертикальный объект средних размеров медленно приближается с Юго-Запада;
- транзитное движение большого объекта через 2-й квадрант на Северо-Восток;
- появился в 3-м квадранте малый объект, мерцающий с частотой 0,3 гц».

Очевидно, что такого вида «когнитивное сжатие» видеоинформации уменьшает объём необходимых для передачи данных до minimum minimumum, но обеспечивает достаточность этих данных для принятия решения по контролю обстановки.

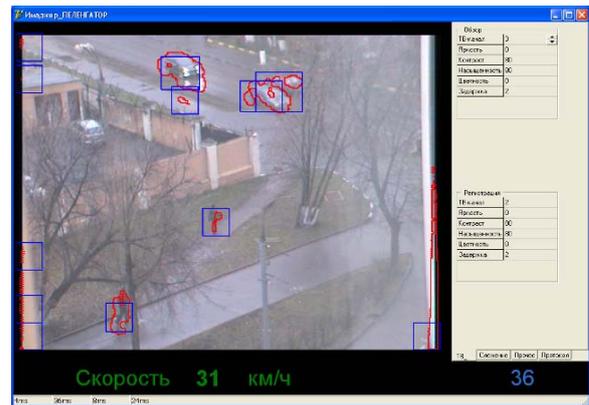
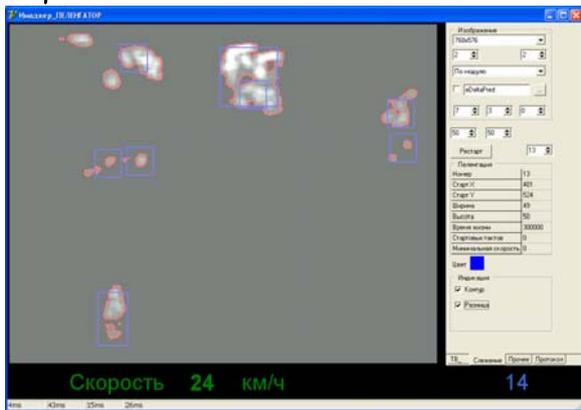


Рис. 3. Пример обработки видеоактивности.

Повышение информативности сигналов изображений

Повышение эффективности распознавания изображений «интеллектуальными» информационными автоматами традиционно начинается, как известно, с первого этапа обработки видеосигнала (фильтрация, предобработка и нормализация) и состоит в формировании видеосигнала с повышенным для объектов контролируемых классов отношением сигнал-шум. Здесь перспективные возможности повышения качества распознавания связывают с дополнительным использованием тепловизионных и стереоскопических изображений.

Тепловизионное изображение содержит дополнительный, относительно оптического изображения, признак температуры, использование которого совместно с признаками от оптического изображения позволяет, в принципе, более качественно обнаруживать и классифицировать такие разнотемпературные объекты как «Люди», «Транспортная техника», «Производственные сооружения» и т.п. Но тепловизионные датчики изображений пока обеспечивают пространственное разрешение существенно меньше оптических. Поэтому использование для обнаружения и идентификации объектов совместно оптического и тепловизионного изображений требует специальных операций сопоставления данных.

Применительно к комплексам видеонаблюдения оптимальным, по критерию «сложность реализации/информативность», представляется алгоритм замешивания тепловизионного сигнала в красную цветовую компоненту R видеосигнала RGB, с предварительным нелинейным масштабированием. Макетные исследования подобного алгоритма обработки изображений иллюстрирует рис. 4, на котором исходное «серое» тепловизионное изображение (слева) представлено также в традиционной «радужной» палитре (центр) и посредством экспериментальной «псевдоестественной RGB» палитре.

Другая возможность повышения достоверности обнаружения и идентификации объектов класса «Лица людей» и

аналогичных структурно сложных объектов связывается с результатами обработки стереоизображений – двух или более изображений одной и той же сцены, снятой с разных, но известных, ракурсов. На основе отождествления на этих разных изображениях одних и тех же «реперных областей» вычисляется псевдообъемное изображение сцены, которое представляет собой монохромное изображение, где яркость соответствует дальности до наблюдающей видеокамеры или, другими словами, отображает «высоту рельефа» наблюдаемых объектов. Метод применяется также в стереотелевидении, где вычисленное псевдообъемное изображение называют картой глубины.

Практические системы распознавания по плоским (монюкулярным) оптическим изображениям используют для обнаружения лица, по всей видимости, только те признаки, которые характеризуют область глаз, а именно, конфигурацию глазных углов. Подтверждением этого может служить тот факт, что в этих системах не происходит детекции такого информативного элемента как овал лица. Идентификация лиц выполняется, соответственно, на основе попиксельного сравнения яркостей фрагментов изображений, соответствующих лицу, в качестве которых вырезаются вокруг линии глаз прямоугольники фиксированных размеров и пропорций. Эти обстоятельства приводят к повышенной вероятности ложных обнаружений и недостаточной достоверности идентификации лиц людей.

Использование карты глубины для идентификации лиц людей может обеспечить существенное повышение качества распознавания. Во-первых, карта глубины, как эталонное описание объекта, является более эффективным по сравнению с плоским изображением, поскольку инвариантно, по определению, к яркостным вариациям, поворотам и ракурсам. Это расширяет динамический диапазон достоверных значений коэффициента подобия, по которым принимается решение об идентификации.



Рис. 4. Примеры представления тепловизионного изображения.

Во-вторых, карта глубины обеспечивает предположительно более высокое отношение сигнал-шум всего множества информативных биометрических элементов лица, включая нос, рот и овал лица.

Ещё одним перспективным приёмом повышения эффективности автоматизированного видеонаблюдения представляется использование нескольких разнофокусных видеокамер на одну контролируруемую сцену. Это позволяет реализовать оптимальные по масштабам условия наблюдения как для всей сцены, так и для отдельных её объектов. При этом одна из видеокамер – обзорная – обеспечивает наблюдение сцены целиком и обнаружение на ней одновременно нескольких, по-разному активных объектов (например, рис. 3). Это позволяет выявить интегральные особенности поведения типа «намерений» объекта на основе анализа вычисленных для него траекторных параметров.

Другие видеокамеры – регистрирующие – предназначены для обеспечения масштаба изображения, необходимого и достаточного для достоверной идентификации объекта. Условием этого является надежное обнаружение наиболее информативных элементов объектов, например, символов для автономеров (рис. 1) или элементов лица для человека (рис. 2).

Заключение

Накопленный опыт разработки систем технического зрения дает реальные предпосылки для создания универсального распознавателя визуальной обстановки на основе унифицированного алгоритма информационного анализа сигналов изображений, состоящего из последовательности базовых операций обнаружения структурных примитивов $\Omega \in I(x, y, z, t)$ изображений и динамических примитивов $\tilde{\Omega} \in \tilde{I}(x, y, z, t)$ видеоактивности, измерения их характеристик, кластеризации/ (логической фильтрации) групп примитивов как изображений f_{μ} предполагаемых объектов и, в итоге, классификации/идентификации фрагментарных изображений f_{μ} объектов по диапазонам значений характеристик

$$p_{\lambda} = \sum_v C_v \cdot f_v^{\mu}$$

составляющих их примитивов.

Целевая функция такого универсального распознавателя должна состоять в непрерывном формировании описания наблюдаемой сцены в виде списка событий, объектов и их свойств. Необходимые для этого операции информационного анализа должны состоять в следующем:

1) обнаружение на сигнале изображения динамических и статических фрагментов – предположительных объектов, вычисление и регистрация их характеристик;

2) классификация обнаруженных объектов и их мультимедийная регистрация с качеством, необходимым для идентификации;

3) интерпретация контролируемой сцены в терминах появления, исчезновения, активности и движения объектов и регистрация сформированного описания;

4) идентификация объектов по их зарегистрированным изображениям;

5) статистический анализ характеристик выявленных со-

бытий и объектов и редактирование результатов распознавания.

Ключевым звеном при реализации универсального распознавателя представляется специальный видеопроцессор-обнаружитель, выполняющий весь набор операций по обнаружению структурных примитивов, соответствующих объектам и их частям.

Характеристическим параметром видеопроцессора-обнаружителя является наличие выходного потока так называемых метаданных, которые будут представлять собой числовые оценки траекторных, пространственных и яркостных свойств обнаруженных структурных примитивов анализируемых изображений. Это множество связанных областей, получаемое для одного изображения, полностью описывает его информационную структуру и содержит данные обо всех объектах соответствующей сцены.

Развиваемая на этой основе технология распознавания изображений может служить основой для создания «интеллектуальных» систем управления движением автономных наземных роботов.

Литература

1. Advanced Video and Signal Based Surveillance, 2009. AVSS '09. Sixth IEEE International Conference on 2-4 Sept. 2009, Genova, Italy.
2. Болл Р. М., Коннел Дж. Х., Панканти Ш., Ратха Н.К., Сентор Э.У. Руководство по биометрии – Пер. с англ., ТЕХНОСФЕРА, Москва, 2007.
3. Журавлев Ю.И., Рудаков К.В., Гуров С.И., Дюкова Е.В., Кутуков Г.П. (ВЦ РАН), Матюнин С.Н. (СПП РАН), Местецкий Л.М. (Тверской ГУ). Состояние и перспективы развития исследований в области обработки и распознавания видеоинформации (аналитический обзор) №1 январь 2005. Материал взят с сайта: <http://www.techno.edu.ru> (<http://www.techno.edu.ru:16001/db/msg/22358.html>).
4. Петричкович Я.Я., Сомиков В.П., Юдинцев В.А. Анализируем системы видеоаналитики - журнал «Системы безопасности», №1, 2009.
5. Птицын Николай. Встроенная видеоаналитика: ближайшие перспективы.- Материалы с сайта <http://synesis.ru/ru/surveillance/>, © 2009 - 2010 Synesis.
6. Сб. докладов 12-й Международной конференции и выставки «ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ И ЕЁ ПРИМЕНЕНИЕ» // Труды Российского научно-технического общества радиотехники, электроники и связи имени А.С. Попова. – Серия: Цифровая обработка сигналов и её применение, выпуск: XII -2, Москва – 2010.
7. Сойфер В.А., Сергеев В.В., Попов С.Б., Мясников В.В. Теоретические основы цифровой обработки изображений: Учебное пособие/. Самарский государственный аэрокосмический университет имени академика С.П. Королёва. Самара, 2000.
8. Саниев К.Б. Распознавание сигналов изображений. - РАДИОТЕХНИКА, № 4, апрель 2007 г., с. 9-15.
9. Саниев К.Б. Способ распознавания изображений. - патент на изобретение № 2313828, РФ, госрегистрация 2007 г.
10. Саниев К.Б., Дёмин М.А., Панкратов Д.В. - ПКО «Имаджер++», свидетельство о регистрации программы для ЭВМ № 2009610765, 2009 г.