

УДК 004.932.4

ОБНАРУЖЕНИЕ ИГРОВЫХ ЭПИЗОДОВ В ВИДЕОПОТОКЕ

*Петрова К.Ю., к.т.н., руководитель проектов, Исследовательский Центр Самсунг, Москва,
e-mail: xenyu.petrova@samsung.com;*

*Рычагов М.Н., д.ф.-м.н., проф., начальник управления, Исследовательский Центр Самсунг, Москва,
e-mail: michael.rychagov@samsung.com;*

*Седунов С.М., старший инженер, Научно-Исследовательский Центр Хуавей, Москва,
e-mail: Sedunov.Sergey@huawei.com.*

FIELD GAME EPISODE DETECTION IN VIDEO SEQUENCES

Petrova X.Y., Rychagov M.N., Sedunov S.M.

Description of classifier implementing field game episode detection in video sequences in real time is provided. Method is based on visual cues, described by one and multi-dimensional distributions of color and texture features. Classifier has form of directed acyclic graph with threshold transforms, linear sub-classifiers and Boolean logical functions in the nodes. Solution can be used in TV receivers for automatic selection of settings, providing comfortable viewing of sport programs and as a pre-processing classification cascade in off-line solutions for high precision video classification.

Keywords: video classification, real time, sport, soccer, visual cue, acyclic directed graph, color, texture.

Ключевые слова: классификация видео, реальное время, спорт, футбол, визуальный ключ, ациклический направленный граф, цвет, текстура.

Введение

В современных телевизионных приёмниках принято использовать различные режимы обработки видеосигнала, соответствующего производственным жанрам видео, таким как спорт, кино, музыка и т.п. Например, для обеспечения лучшего качества изображения при приеме спортивных передач обычно рекомендуется выбирать средний контраст, небольшой коэффициент улучшения деталей и низкий коэффициент шумоподавления, в то время как при просмотре фильмов и музыкальных передач эти коэффициенты могут принимать повышенные значения. Таким образом, детектирование игровых эпизодов в реальном времени может быть полезно для автоматической подстройки коэффициентов, обеспечивающей наиболее комфортный просмотр спортивных телепередач.

Основной целью исследований, результаты которых представлены в статье, является разработка алгоритма обнаружения эпизодов, содержащих игровые виды спорта в видеопоследовательности в реальном времени с

Классификация как статических, так и видеоизображений является бурно развивающейся областью и обладает большим потенциалом коммерциализации. В настоящей работе решается задача определения игровых эпизодов в видеопотоке на основе применения т.н. визуальных ключей, характеризующихся одномерными и двумерными распределениями цветовых и текстурных признаков. Реализована технология обработки видео, обеспечивающая требуемую классификацию с точностью до одного кадра в реальном времени и позволяющая простую аппаратную реализацию.

целью последующей автоматической подстройки настроек изображения в ТВ-приемнике (рис. 1).

Обзор работ по классификации видеопоследовательностей

Известные работы по классификации видеопоследовательностей можно разделить по следующим признакам: цель классификации (например, определение жанра или детектирование определенных объектов), используемые модальности (такие как видео, аудио, субтитры), выбор признаков и тип классификатора. Наиболее близкими по целевому признаку являются широко представленные работы по автоматическому определению жанра [1-15].



Рис. 1. Распознавание жанра видео в конвейере видеобработки телевизионного приемника



Определение жанра может производиться с использованием различных модальностей, например, субтитров [1], звука [2-4], или видео (как во всех прочих источниках), или одновременно нескольких модальностей [5].

В [1] в качестве признаков использованы субтитры и коэффициенты ДКТ разложения видеок кадров. Авторы получили высокую точность детектирования, однако отметили, что субтитры отсутствуют во многих телевизионных трансляциях. Хотя в некоторых странах, как, например, в США законодатель обязывает вещателей сопровождать теле-трансляции субтитрами. При этом, субтитры не являются описанием того, что демонстрируется на экране, и не генерируются для сцен, в которых отсутствуют диалоги. Наконец, обучение и классификация по такому признаку могут сопровождаться большой вычислительной сложностью, поскольку вектор признаков может состоять из многих тысяч элементов.

В [2] видеопоследовательности, содержащие запись спортивных соревнований, проиндексированы с помощью SVM-классификатора на основании признаков аудиопотока.

Подходы, основанные на анализе аудио потока или субтитров, неприменимы в телевизионном приемнике для автоматической подстройки коэффициентов видеопроботки, поскольку эти признаки не всегда совпадают по времени с соответствующим видеорядом.

Далее будут рассматриваться методы, основанные исключительно на анализе видео-потока.

В [6] производится определение следующих жанров: мультфильмы, реклама, новости и спорт. В качестве базового алгоритма выбран метод опорных векторов, использующий ориентированный ациклический граф (DAGSVM). При этом выделялись 15 визуальных признаков четырех типов (монтаж, цвет, текстура и движение). С точки зрения монтажа оценивалась частота смены эпизодов и соотношение между количеством резких и плавных переходов между сценами. В качестве цветовых признаков выступали гистограмма, средняя яркость и насыщенность, а также процент пикселей с яркостью и насыщенностью выше предопределенного порога. Текстурные признаки соответствовали статистическим свойствам полутонной матрицы смежности: контрасту, однородности, энергии, энтропии и корреляции. Признаки движения определялись как среднее изменение яркости и средняя разность в цветовом пространстве RGB между соседними кадрами, а также доля кадров, которые отличаются незначительно от предыдущего (доля медленных и статических сцен).

В работе [7] проводится классификация кинофильмов по жанрам: боевики, драмы и триллеры. Использовались шесть глобальных признаков: средняя длина эпизода, вариация цвета, движение, освещение (наличие высвечивающего фонаря), статистика типов переходов между сценами. Классификатор реализован в виде дерева принятия решений.

В работе [8] описано применение свойств границ и классификатора K средних для того, чтобы отличить бадминтон, баскетбол, футбол, коньки и теннис. В более ранней работе тех же авторов свойства границ использовались для того, чтобы выделить кадры, содержащие общий план зрительских трибун и рекламу вне игрового поля.

В [9] анализировались траектории лиц и блоков текста для того, чтобы отличить рекламу, новости, комедийные сериалы и мыльные оперы. Классификация была произведена путем нахождения максимальной проекции распределения траекторий классифицируемого фрагмента на множество траекторий обучающего множества.

В [10] кадры видеозаписи лекций разделены на 4 класса. Для этого предлагается квази-оптимальная процедура выбора признаков (из 300 исходных признаков).

В [11-12] посредством анализа движения камеры производилась классификация пяти видов спорта (сумо, теннис, бейсбол, футбол и американский футбол). Для каждого из этих видов спорта характерен особый способ монтажа. Был проведен анализ использования съемки с проводкой, наездов камеры, эпизодов со статичной и трясущейся камерой, а также типов переходов от одного вида движения камеры к другому. Поскольку метод не опирается на цветовую информацию, он может быть очень эффективен при классификации трансляций игр разных спортивных лиг (например, на открытом чемпионате Великобритании по теннису корты – зеленые, а на открытом чемпионате Франции – красные).

В [13] предложен новый низкоуровневый признак – поток энергии активности. Он подавался на вход сети радиальных базисных функций для определения спорта, кино, ландшафтных съемок и новостей. Поток энергии активности измерялся в пределах одного эпизода и был основан на статистике макро-блоков сжатого видео, включая количество I -блоков, блоков с надежно предсказанным движением и блоков с ненадежно предсказанным движением (например, в случае однородного фона при движущейся камере).

В [14] для определения видов спорта (теннис, трековые виды спорта, плавание, яхтинг и велогонки) использовалась концепция «визуальных ключей». Эти ключи придавали семантическое значение низкоуровневым признакам кадров видеопоследовательности. В общей сложности выделялось 17 типов ключей: трек для занятий легкой атлетикой, боксерский ринг, крытый велотрек, океан, лыжный трамплин, бассейн, теннисная сетка, трава, синее небо, открытый велотрек, океан (средний план), беговая дорожка (дальний план), беговая дорожка (средний план), беговая дорожка (крупный план), бассейн (крупный план), теннисный корт. Первые 9 детекторов были обученными нейронными сетями, а последние 8 использовали т.н. текстурные коды. Результаты детектирования визуальных ключей подавались на вход классификатора K средних, который определял, к какому виду спорта относится эпизод. В [15] набор признаков был расширен посредством добавления к нему мультимодальных элементов для детектирования визуальных ключей. При этом применялись скрытые марковские модели для анализа временных зависимостей.

В [16] рассматривалась задача выделения 7 видов спорта (альпинизм, баскетбол, автогонки, гольф, прыжки с трамплина, футбол и мотогомки) и 5 типов сюжетов, имеющих отношение к спорту: финальные титры, комментаторы в студии, начальные титры, трейлер и турнирная таблица. Решение принималось на основе вычисления векторов цветовой когерентности [17].

В [18] произведена классификация 7 популярных видео жанров: мультфильмы, реклама, документальные фильмы, кино, музыкальные клипы, новости и спорт. Было использовано три категории дескрипторов – на уровне цвета, динамики и структуры. Цветовые свойства описывались глобально посредством статистики распределения различных цветов, элементарных цветов, цветовых свойств (яркость, насыщенность) и отношений между ними. С точки зрения динамики, оценивался ритм видео, статистика движений и процентная доля плавных переходов. Структурная информация извлекалась на уровне кадра путем построения гистограмм контуров и выявления отношений между ними.

Инструментарий, используемый для классификации изображений и видео весьма разнообразен. Ранее были продемонстрированы эффективность метода решающие деревья [7], анализа основных компонентов [19], метода опорных векторов [2, 3], нейронных сетей [5, 14], сетей Кохонена [20], скрытых марковских моделей [21]. В [22] использован классификатор по принципу ближайшего соседа. Некоторые исследователи используют случайный лес или байесовский подход [23].

В [20] сделана интересная попытка установить соответствие между семантической концепцией и списком синонимов этой концепции, найденных в WordNet. Эта идея может послужить основой для быстрого и эффективного синтеза визуального классификатора на основе вербального описания, сделанного экспертом. Похожая идея описана в [24], где банк классификаторов визуальных объектов использует набор примитивных классификаторов, которые можно использовать в качестве входов более сложных классификаторов.

С точки зрения решаемой задачи интерес представляют также методы определения смены сцены, однако мы ограничились простейшим случаем – прямым переходом, поскольку это преобладающий тип перехода в спортивных трансляциях.

Требования к алгоритму классификации видеок кадров в реальном времени

Будем предъявлять следующие требования к алгоритму классификации:

1. Для управления настройками конвейера видеобработки результат классификации должен быть получен для каждого отдельного кадра и не иметь резких скачков за исключением случая смены сцены. В связи с этим должен использоваться лишь один проход по видеоккадру.

2. Алгоритм должен быть универсальным, т.е. игровые эпизоды должны быть дифференцированы от всех прочих типов сцен, включая кино, новости, мультфильмы, компьютерную графику, концерты и т.д.

3. Алгоритм должен быть нечувствителен к качеству видео потока (поддерживать различное разрешение, как стандартное, так и высокое, быть нечувствительным к различным методам и качеству сжатия).

4. Сложность алгоритма должна быть невысокой, допускающей аппаратную реализацию в виде части микросхемы видеобработки на чипе. Поэтому желательно, чтобы при анализе пикселей использовалась ограниченная локальная окрестность.

5. Алгоритм должен иметь модульную организацию для обеспечения его взаимодействия с другими подсистемами видеобработки.

Описание алгоритма

Синтез алгоритма сопровождался формированием базы данных, состоящей из видеофрагментов различного содержания, разрешения и качества сжатия. Анализ этих данных позволяет сделать следующие выводы:

1. Важность получения правильного результата не одинакова для различных кадров: более важно получить стабильный результат детектирования для кадров, содержащих зеленое поле с движущимися по нему игроками, чем для кадров со зрительскими трибунами или комментатором, или даже для игрока, демонстрируемого крупным планом.

2. Не все ошибки классификации имеют одинаковый вес. Так, вполне допустимо перепутать футбол с бейсболом, но совершенно недопустимо перепутать футбол с музыкальным клипом.

3. Видеоданные (на уровне отдельных кадров) по своей природе обладают очень большим смещением, так как распределение больше всего зависит от длительности эпизодов, содержащих кадры с близкими характеристиками. Т.е. частота довольно слабо связана с важностью.

Чтобы формализовать эти требования, вводится дополнительная под-классификация в виде простой иерархии и весовых коэффициентов для матрицы несоответствий. Матрица несоответствий – это популярный инструмент для оценки качества алгоритмов классификации [25]. Категорию «игровые виды спорта» мы подразделили на три подкатегории: С1 (игровые эпизоды, футбол), С2 (игровые эпизоды, игры на поле кроме футбола), С3 (спорт, неигровые эпизоды). В рамках этой классификации в категорию С4 попали все остальные типы кадров. Весовые коэффициенты для матрицы несоответствий показаны в табл. 1.

Для упрощения синтеза алгоритма, класс С1 был подразделен на три класса: дальний план, общий план и крупный план. Для каждого из классов строился отдельный классификатор. На первом шаге рассматривался наиболее простой и очевидный случай: для того, чтобы отличить кадры дальнего и общего планов демонстрации футбольного матча от кадров прочих жанров, в качестве признака использовалось процентное соотношение зеленых пикселей. Дальнейший синтез проводился с использованием усеченной базы данных, состоящей только из кадров, на которых классификатор, синтезированный на первом шаге, выдает ошибку первого рода.

Таблица 1.

		Спорт			Не-спорт
		C1	C2	C3	
Спорт	C1	0	1/100	1/10	1
	C2	1/100	0	1/10	1
	C3	1/10	1/10	0	1/10
Не спорт	C4	1	1	1/100	0

За основу берется формула $Gr0 = G > RB_{\max}$, где RB_{\max} – максимум значений красного и синего каналов.

тат работы детектора кожи. Средняя яркость вычисляет-

$$F_3 = \frac{\sum_{i=1..w, j=1..h} Y(i, j)}{w \cdot h}$$

ся как $F_3 = \frac{\sum_{i=1..w, j=1..h} Y(i, j)}{w \cdot h}$, где $Y(i, j)$ – яркость пик-

$$F_4 = \frac{\sum_{i=1..w, j=1..h} |D_Y(i, j)| \cdot \delta(G_r(i, j))}{\sum_{i=1..w, j=1..h} \delta(G_r(i, j))}$$

селья. Среднее значение градиента вычисляется как $F_4 = \frac{\sum_{i=1..w, j=1..h} |D_Y(i, j)| \cdot \delta(G_r(i, j))}{\sum_{i=1..w, j=1..h} \delta(G_r(i, j))}$, где горизонтальная производная яркости D_Y получена путем свертки яр-

костной компоненты Y с линейным фильтром $K_{grad} = [0 \ 0 \ 0 \ 1 \ 0 \ 0 \ -1]$.

Доля ярких и насыщенных пикселей вычисляется по формуле $F_5 = \frac{\sum_{i=1..w, j=1..h} \delta(B_s(i, j))}{w \cdot h}$, где $B_s(i, j)$ – ре-

зультат детектирования ярких и насыщенных пикселей. Средняя насыщенность зеленых пикселей определяется $F_6 = \frac{\sum_{i=1..w, j=1..h} S(i, j) \cdot \delta(G_r(i, j))}{\sum_{i=1..w, j=1..h} \delta(G_r(i, j))}$, где $S(i, j)$ –

насыщенность. Доля белых пикселей вычисляется как $F_7 = \frac{\sum_{i=1..w, j=1..h} \delta(W(i, j))}{w \cdot h}$, где $W(i, j)$ резуль-

тат детектирования белых пикселей. Средняя яркость зеленых пикселей вычисляется как $F_8 = \frac{\sum_{i=1..w, j=1..h} Y(i, j) \cdot \delta(G_r(i, j))}{\sum_{i=1..w, j=1..h} \delta(G_r(i, j))}$. Среднее значение

синего цветового канала зеленых пикселей вычисляется по формуле: $F_9 = \frac{\sum_{i=1..w, j=1..h} B(i, j) \cdot \delta(G_r(i, j))}{\sum_{i=1..w, j=1..h} \delta(G_r(i, j))}$. Степень

компактности гистограммы яркости зеленых пикселей F_{10} вычисляется в несколько шагов:

Строится гистограмма H_{YGr} значений яркости Y пикселей, принадлежащих зеленому областям;

Вычисляется ширина гистограммы D как расстояние между ее правым и левым ненулевыми элементами;

Признак F_{10} вычисляется как пропорция гистограммы, лежащая не дальше восьмой части ее ширины от ее

середины $F_{10} = \frac{\sum_{i=P_8-D/8}^{P_8+D/8} H_{YGr}(i)}{\sum_{i=0}^{255} H_{YGr}(i)}$.

Вычисляется ширина гистограммы D как расстояние между ее правым и левым ненулевыми элементами;

Признак F_{10} вычисляется как пропорция гистограммы, лежащая не дальше восьмой части ее ширины от ее

$$F_{10} = \frac{\sum_{i=P_8-D/8}^{P_8+D/8} H_{YGr}(i)}{\sum_{i=0}^{255} H_{YGr}(i)}$$

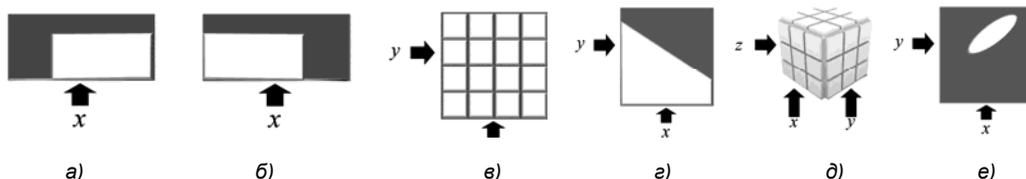


Рис. 3. Элементарные классификаторы

Результирующий классификатор синтезируется в форме ориентированного ациклического графа с элементарными классификаторами в узлах.

Для визуального синтеза классификаторов разработано программное обеспечение, реализующее элементарные классификаторы следующих типов: одномерные пороговые преобразования $\theta_i^{HI}(x)$ и $\theta_i^{LO}(x)$, двумерные пороговые преобразования $\theta_i^{2D}(x, y)$, линейные классификаторы $\theta_i^L(x, y)$, трехмерные пороговые преобразования $\theta_i^{3D}(x, y, z)$ и двумерные эллиптические классификаторы $\theta_i^e(x, y)$.

Одномерные пороговые преобразования имеют один входной аргумент и описываются формулой $\theta_i^{HI}(x) = \begin{cases} TRUE, & x > T_i \\ FALSE, & otherwise \end{cases}$ (рис. 3.а)

или $\theta_i^{LO}(x) = \begin{cases} TRUE, & x < T_i \\ FALSE, & otherwise \end{cases}$ (рис. 3.б). Параметры

двумерных пороговых преобразований (или табличных классификаторов) представлены на рис. 3.в. Они описываются двумя векторами порогов:

$V_T^1 = [T_0^1 \ T_1^1 \ \dots \ T_N^1]$, где $T_0^1 < T_1^1 < T_2^1 < \dots < T_N^1$, и $V_T^2 = [T_0^2 \ T_1^2 \ \dots \ T_M^2]$, где $T_0^2 < T_1^2 < T_2^2 < \dots < T_M^2$.

Количественные значения порогов для каждого измерения могут не совпадать: $M \neq N$. Выход табличного классификатора имеет форму двумерного вектора размерности $M \times N$:

$\theta^{2D}(x, y) = [y_{11}(x, y), y_{12}(x, y), \dots, y_{1N}(x, y), y_{21}(x, y), \dots, y_{M1}(x, y), y_{M2}(x, y), \dots, y_{MN}(x, y)]$, где каждый элемент является логической величиной $y_{ij}(x, y) =$

$= x \geq T_{i-1}^1 \wedge x \leq T_i^1 \wedge y \geq T_{j-1}^2 \wedge y \leq T_j^2$. Выход линейного классификатора вычисляется на основании следующего выражения: $\theta_i^L(x, y) = K_1 \cdot x + K_2 \cdot y + B > 0$, где K_1, K_2 и B – предопределенные константы (рис. 3.г). Параметры трехмерных табличных классификаторов (рис. 3.д) описываются тремя векторами порогов:

$V_T^1 = [T_0^1 \ T_1^1 \ \dots \ T_N^1]$, где $T_0^1 < T_1^1 < T_2^1 < \dots < T_N^1$; $V_T^2 = [T_0^2 \ T_1^2 \ \dots \ T_M^2]$, где $T_0^2 < T_1^2 < T_2^2 < \dots < T_M^2$ и $V_T^3 = [T_0^3 \ T_1^3 \ \dots \ T_L^3]$, где $T_0^3 < T_1^3 < T_2^3 < \dots < T_L^3$.

Выход табличного классификатора имеет вид трехмерного вектора размерности $M \times N \times L$: $\theta^{2D}(x, y) = [y_{111}(x, y, z), y_{112}(x, y, z), \dots, y_{MNL}(x, y, z)]$.

Выход табличного классификатора имеет вид трехмерного вектора размерности $M \times N \times L$: $\theta^{2D}(x, y) = [y_{111}(x, y, z), y_{112}(x, y, z), \dots, y_{MNL}(x, y, z)]$.

Выход табличного классификатора имеет вид трехмерного вектора размерности $M \times N \times L$: $\theta^{2D}(x, y) = [y_{111}(x, y, z), y_{112}(x, y, z), \dots, y_{MNL}(x, y, z)]$.

Выход табличного классификатора имеет вид трехмерного вектора размерности $M \times N \times L$: $\theta^{2D}(x, y) = [y_{111}(x, y, z), y_{112}(x, y, z), \dots, y_{MNL}(x, y, z)]$.

Выход табличного классификатора имеет вид трехмерного вектора размерности $M \times N \times L$: $\theta^{2D}(x, y) = [y_{111}(x, y, z), y_{112}(x, y, z), \dots, y_{MNL}(x, y, z)]$.

Выход табличного классификатора имеет вид трехмерного вектора размерности $M \times N \times L$: $\theta^{2D}(x, y) = [y_{111}(x, y, z), y_{112}(x, y, z), \dots, y_{MNL}(x, y, z)]$.

Выход табличного классификатора имеет вид трехмерного вектора размерности $M \times N \times L$: $\theta^{2D}(x, y) = [y_{111}(x, y, z), y_{112}(x, y, z), \dots, y_{MNL}(x, y, z)]$.

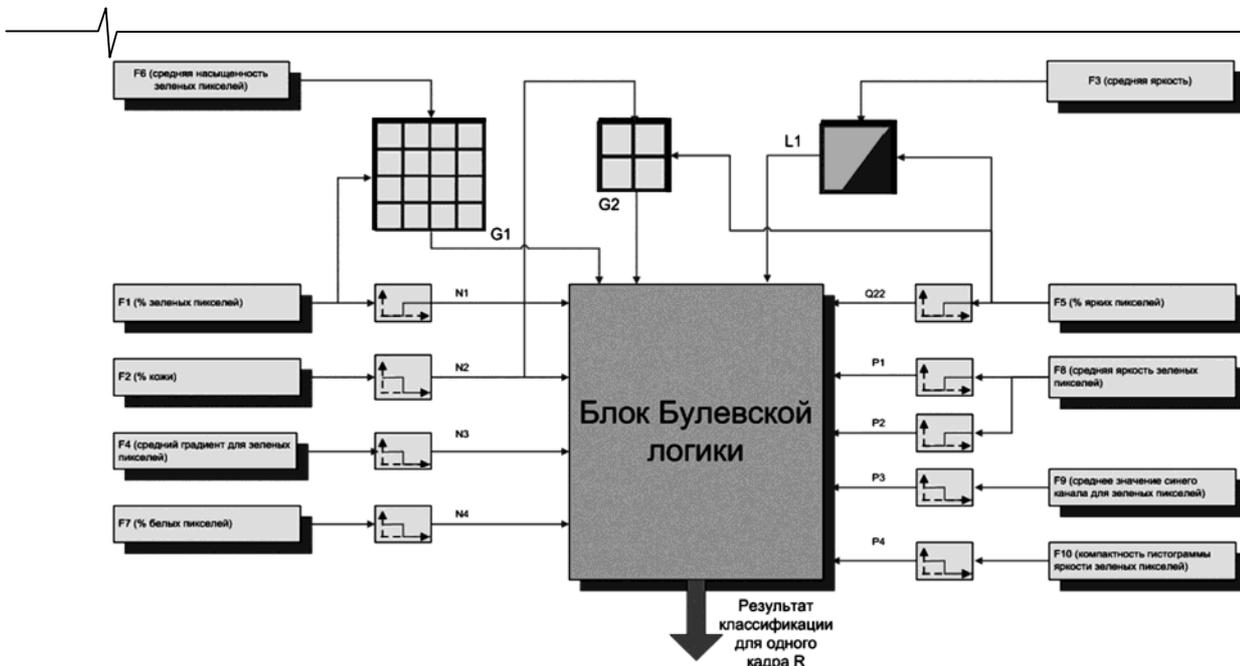


Рис. 4. Структура классификатора для одного кадра

Выход эллиптического классификатора (рис. 3 е) вычисляется как $\theta_i^e(x, y) = \frac{(x' - X_C)^2}{a^2} + \frac{(y' - Y_C)^2}{b^2} < 1$, где $x' = x \cdot \cos \alpha - y \cdot \sin \alpha$, $y' = x \cdot \sin \alpha + y \cdot \cos \alpha$ а X_C, Y_C, a, b и α – предопределенные константы.

В процессе исследований выяснено, что во многих случаях вполне достаточно ограничиться простейшим вариантом классификатора, включающего одномерные и двумерные пороговые преобразования, а также набор линейных классификаторов (рис. 4).

Сформулируем основные правила, на основе которых делается окончательный вывод о том, относится ли текущий кадр к игровым эпизодам. При этом будем считать, что пороговые величины T^1, \dots, T^{13} , $\tilde{T}_i^1, \tilde{T}_i^2, i = 1..4$ и $\hat{T}_i^1, \hat{T}_i^2, i = 1..2$ являются действительными числами, определенными на интервале $[0, 1]$.

Первое правило реализовано в виде табличного классификатора на основе признаков F_1 и F_6 и ориентировано на вычисление вектора

$$G_1 = |y_{11}, y_{12}, y_{13}, y_{14}, y_{21}, y_{22}, y_{23}, y_{24}, y_{31}, y_{32}, y_{33}, y_{34}, y_{41}, y_{42}, y_{43}, y_{44}|.$$

Цель этого правила состоит в том, чтобы выделить случаи дальнего, общего и крупного планов, которым соответствуют ненулевые значения y_{ij} (очевидно, что нулевая норма вектора G_1 равна единице). Результат выполнения этого правила используется в дальнейшем для выбора порогов. При этом если насыщенность зеленого мала, то имеет смысл применить более строгие правила, нежели в случае, когда она высока или принимает средние значения. Второе правило, реализуемое в виде табличного классификатора на основе признаков F_2 и F_5 , ориентировано на определение вектора $G_2 = |z_{11}, z_{12}, z_{21}, z_{22}|$. Применение этого правила обеспечивает в дальнейшем выделение кадров с низкой до-

лей зеленых пикселей.

Случай с очень низкой долей зеленых пикселей рассмотрен отдельно и реализован в виде «негативного» правила: $N_1 = F_1 > T^3$. Случай с очень высокой долей пикселей цвета кожи также контролируется «негативным» правилом: $N_2 = F_2 < T^4$. Текстурные свойства зеленых пикселей контролируются еще одним «негативным» правилом $N_3 = F_4 < T^5$. Игровые эпизоды содержат также определенную (впрочем, сравнительно небольшую) долю белых пикселей: $N_4 = F_7 > T^6$.

Поскольку вспомогательный детектор ярких и насыщенных пикселей основан на фиксированных порогах, очевидно, что результат детектирования будет зависеть от общей яркости изображения.

После применения простой гамма-коррекции изображения доля детектированных ярких и насыщенных пикселей может измениться в разы. Для решения этой проблемы используется линейный классификатор $Q_1 = K_1 \cdot F_3 + K_2 \cdot F_5 + B > 0$, где K_1, K_2 и B – предопределенные константы.

Количество ярких и насыщенных пикселей контролируется правилом $Q_2 = F_5 > T^9$. Средняя яркость зеленых пикселей разбита на два диапазона, для каждого из которых в дальнейшем используется различная эмпирическая логика, например, $P_1 = F_8 > T^{10}$ и $P_2 = F_8 > T^{11}$. Среднее значение синей компоненты для зеленых пикселей контролируется правилом $P_3 = F_9 < T^{12}$, а степень компактности гистограммы яркости зеленых пикселей – $P_4 = F_{10} < T^{13}$. Окончательный результат детектирования R вычисляется по формуле $R = (\neg V_1) \wedge P_1 \wedge P_4 \wedge V_2 \wedge (P_2 \vee (\neg P_2 \wedge P_3))$, где $V_1 = N_1 \vee N_2 \vee N_3 \vee N_4 \vee z_{11}$ и $V_2 = (y_{22} \wedge Q_2) \vee y_{23} \vee (y_{24} \wedge Q_1) \vee y_{32} \vee y_{33} \vee y_{34} \vee y_{42} \vee y_{43} \vee y_{44}$.

Для обеспечения согласованных результатов детек-

тирования в пределах одной сцены, результат детектирования, полученный для первого кадра, распространяется на всю сцену (рис. 5).



Рис. 5. Структура классификатора для видеопоследовательности

Для детектирования момента смены сцены предложен следующий алгоритм (рис. 6) Значения цветовых каналов RGB текущего видеокadra и центры кластеров K_C с выхода блока задержки подаются на вход блока кластеризации. Центры кластеров представляют собой матрицу размером $K \times 3$, где K – количество класте-

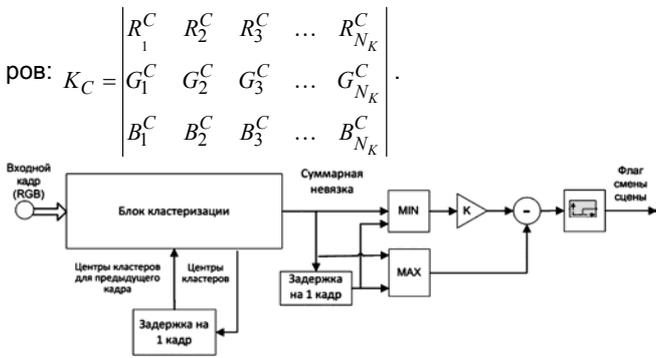


Рис. 6. Структура детектора смены сцены

В настоящем исследовании K принималось равным 8. Кластеризация представляет собой один итерационный шаг метода K средних, когда каждый пиксель $P(i, j) = |R(i, j) G(i, j) B(i, j)|$ с координатами i, j относят к кластеру $K(i, j) = \arg \min_{k=1..N_k} D(P(i, j), C_k)$, где

$C_k = \begin{bmatrix} R_k^C & G_k^C & B_k^C \end{bmatrix}$ – центр k -го кластера и $D(x, y) = \|x - y\|$ – какая-нибудь векторная норма. Суммарная невязка вычисляется при этом как $E = \sum_{i=1..w, j=1..h} D(P(i, j), C_{K(i, j)})$. Центры кластеров обновля-

ются по формуле $\tilde{K}_C = \begin{bmatrix} \tilde{R}_1^C & \tilde{R}_2^C & \tilde{R}_3^C & \dots & \tilde{R}_{N_k}^C \\ \tilde{G}_1^C & \tilde{G}_2^C & \tilde{G}_3^C & \dots & \tilde{G}_{N_k}^C \\ \tilde{B}_1^C & \tilde{B}_2^C & \tilde{B}_3^C & \dots & \tilde{B}_{N_k}^C \end{bmatrix}$, где

$$\tilde{R}_k^C = \frac{\sum_{i=1..w, j=1..h} \delta(K(i, j) = k) \cdot R(i, j)}{\sum_{i=1..w, j=1..h} \delta(K(i, j) = k)}, \quad \tilde{G}_k^C = \frac{\sum_{i=1..w, j=1..h} \delta(K(i, j) = k) \cdot G(i, j)}{\sum_{i=1..w, j=1..h} \delta(K(i, j) = k)},$$

$$\tilde{B}_k^C = \frac{\sum_{i=1..w, j=1..h} \delta(K(i, j) = k) \cdot B(i, j)}{\sum_{i=1..w, j=1..h} \delta(K(i, j) = k)}. \text{ Флаг смены сцены вы-}$$

числяется по формуле $\frac{\max(E_{cur}, E_{prev})}{\min(E_{cur}, E_{prev})} > T_{break}$, где

E_{cur} – суммарная невязка в текущем кадре, E_{prev} – суммарная невязка в предыдущем кадре а $T_{break} = 8$ – заданный порог. Такой детектор работает более надежно, чем детектор, основанный на гистограмме, и при этом является более дешевым с точки зрения его аппаратной реализации.

Результаты

Для демонстрации работы алгоритма реализована компьютерная программа на языках C++/C#.NET. Значения признаков $F_1, F_2, F_3, F_4, F_5, F_6$ выводятся в программном интерфейсе в форме отдельных графиков, а признаки $F_1, F_2, F_3, F_4, F_5, F_7$ в форме столбиковых эпюр (рис. 7). Классификатор протестирован не менее чем на 20 часах видео в реальном масштабе времени (табл. 2). Ошибки первого рода оценивались путем усреднения выхода классификатора на видеопоследовательностях класса С4. В процессе классификации достигнут 95%-ый порог точности. Для оценки ошибок второго рода выбирались случайные кадры общим числом $N_{total} = 220$ из различных последовательностей классов С1-С3 и точность классификации вычислялась как $\frac{N_{C1}^+ + N_{C2}^+}{N_{total}} \cdot 100\%$, где N_{C1}^+ число кадров класса С1,

классифицированных как «спорт» и N_{C2}^+ число кадров класса С2, классифицированных как «спорт». Точность классификации составила 96.5%. Вычисления с использованием коэффициентов из табл. 1 для кадров класса С3 показали приемлемую результирующую точность (выше 95%), однако эти измерения не имеют особой ценности, поскольку мнения экспертов расходятся по поводу того, как классифицировать многие из изображений класса С3.

Производительность алгоритма на компьютере с 2GHz процессором и 2Mb памяти достигала 30 кадров в секунду. Описанный алгоритм может быть реализован с использованием только сдвигов и сложений, что делает его привлекательным для аппаратной реализации.

Обработка сложных случаев

Несмотря на высокое качество классификации, продемонстрированное выше, требуется дальнейшая работа по устранению замеченных погрешностей классификации (рис. 8). Так, на рис. 8.а фрагмент из фильма о природе был принят за дальний план игры в футбол, а гусеница на рис. 8.б ошибочно принята за крупный план игрока. В следующей версии алгоритма таких ошибок можно будет избежать за счет применения усовершенствованного детектора кожи, добавления детектора белой разметки для дальних планов и добавления кросс-доменного признака, сочетающего текстуру и цвет для области ярких и насыщенных цветов. Ошибка классификации на рис. 8.в вызвана слишком широким толкованием понятия «зеленого» цвета. Для решения этой проблемы целесообразно применить адаптацию детекторов

#	классы	Название	Продолжительность, мин.	Разрешение
1	C1,C3	Футбол: 2006 FIFA World Cup Semifinal: Italy vs. Germany	147	960x544
2	C1,C3	Футбол: Liverpool- Manchester United, BBC broadcast	83	592x336
3	C1,C3	Футбол: Milan – Liverpool	51	640x352
4	C1,C3	Футбол: Manchester United – Villareal (Sky Sports2)	72	624x336
5	C2,C3	Амер. футбол: Penn State - Ohio State	141	352x288
6	C2,C3	Амер. футбол: Sugarbowl	168	352x288
7	C4	Wild South America	100	720x576
8	C4	Greatest places	24	1280x720
9	C4	«Люди в черном»	98	720x480
10	C4	Файлы из youtube, всего 84	421	720x528
		Всего	1305	

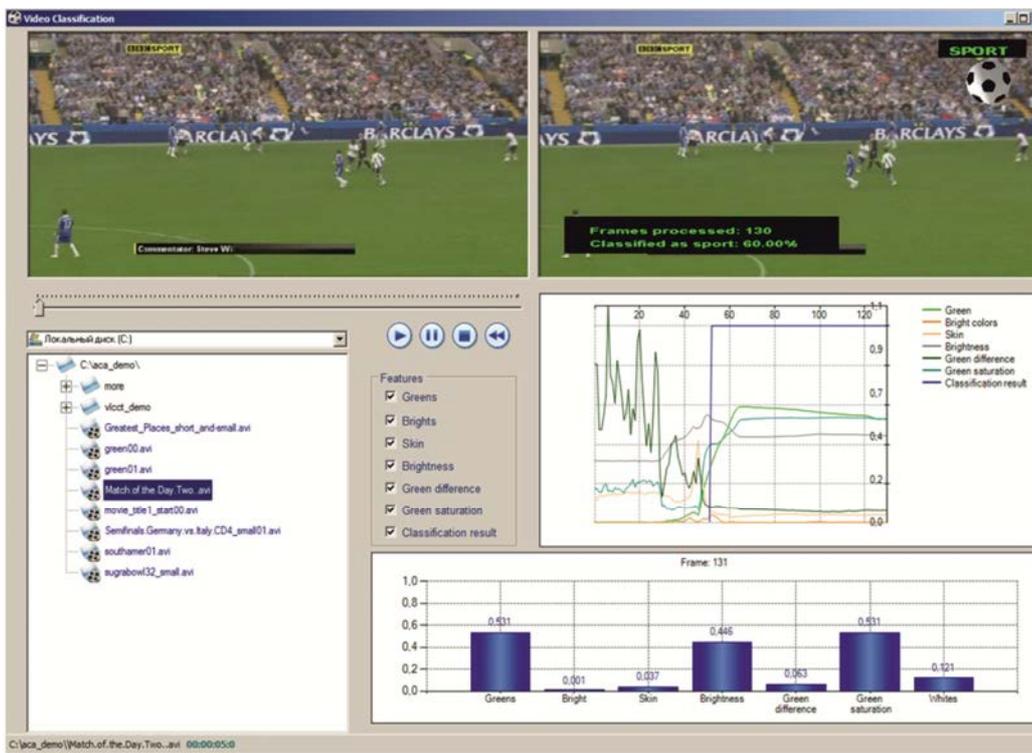
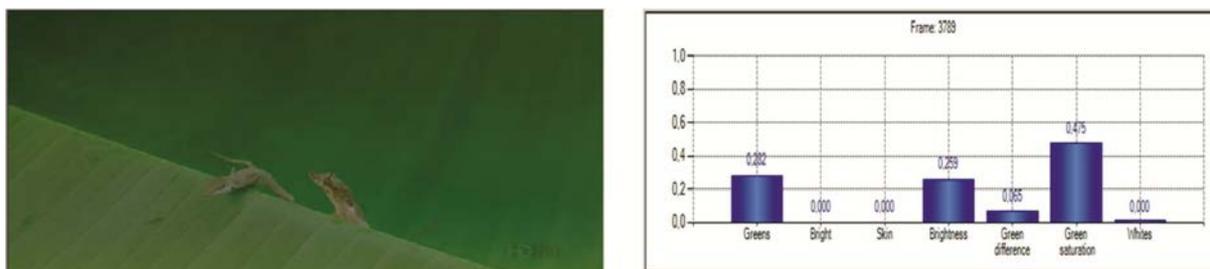
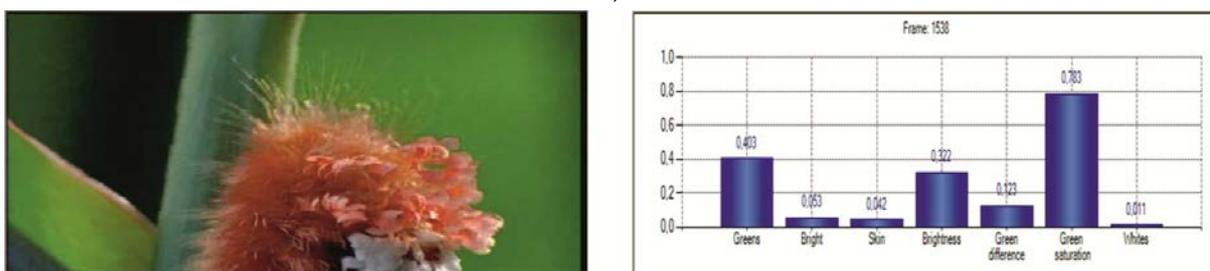


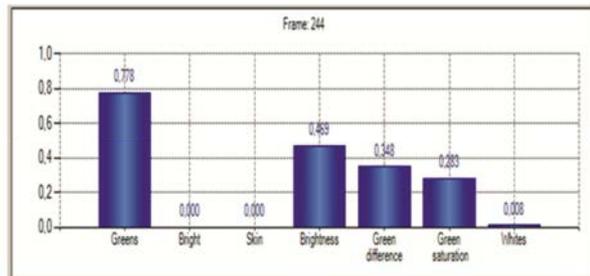
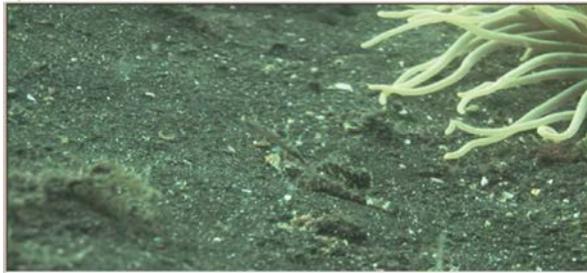
Рис.7. Интерфейс демонстрационной программы



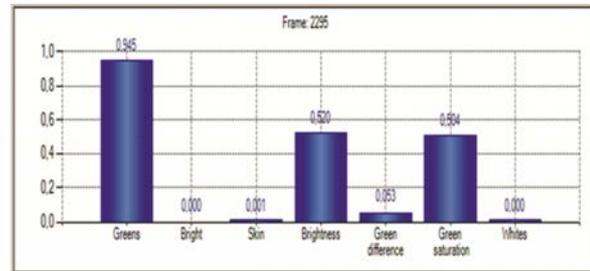
a)



б)



е)



е)

Рис. 8. Ошибки детектирования

цвета к глобальному освещению сцены. Для исправления ошибки, показанной на рис. 8.г придется разработать классификатор силуэтов, что представляет собой значительно более сложную задачу, для которой, скорее всего, не удастся найти приемлемое по производительности решение.

Заключение

Разработан быстродействующий интуитивно понятный алгоритм выделения игровых эпизодов в видеопотоке. Алгоритм дает результат в реальном времени по одному кадру, при этом гарантируется его полноценная реализация на стороне телевизионного приемника в виде блока микросхемы видеобработки.

Литература

1. Brezeale D., Cook D.J., Using Closed Captions and Visual Features to Classify Movies by Genre, Poster session of the Seventh International Workshop on Multimedia Data Mining (MDM/KDD2006), 2006.
2. Liang Bai; Song-Yang Lao; Hu-Xiong Liao; Jian-Yun Chen, Audio Classification and Segmentation for Sports Video Structure Extraction using Support Vector Machine. International Conference on Machine Learning and Cybernetics, Aug. 2006 pp. 3303-3307
3. Dinh P.Q., Dorai C., Venkatesh S. Video genre categorization using audio wavelet coefficients. In 5th Asian Conference on Computer Vision, Melbourne, Australia, Jan 23-25 2002.
4. Roach M., Mason J.. Classification of video genre using audio. Eurospeech, 4:2693-2696, 2001.
5. Subashini K., Palanivel S., Ramalingam V., Audio-Video based Classification using SVM and AANN. International Journal of Computer Applications Volume 44- No.6, April 2012. Pp. 33-39
6. Jiang X., Sun T., Chen B., A Novel Video Content Classification Algorithm Based on Combined Visual Features Model. 2nd International Congress on Image and Sig-

nal Processing, 2009. CISP '09. 17-19 Oct. 2009, pp. 1-6.

7. Huang H.Y., Shih W.S., Hsu W.H., A Film Classifier Based on Low-level Visual Features, Journal of Multimedia, vol. 3, no. 3, July 2008

8. Yuan Y., Wan C., The application of edge feature in automatic sports genre classification. IEEE Conference on Cybernetics and Intelligent Systems, 2004, Vol. 2, pp. 1133-1136.

9. Wei G., Agnihotri L., Dimitrova N., Tv Program Classification Based On Face And Text Processing. IEEE International Conference on Multimedia and Expo, 2000. ICME 2000. Vol. 3, pp. 1345-1348.

10. Y. Liu; J.R. Kender, Video frame categorization using sort-merge feature selection Proceedings. Workshop on Motion and Video Computing, 2002. Volume, Issue, 5-6 Dec. 2002 pp: 72 – 77.

11. Takagi S., Hattori S.M., Yokoyama, K.; Kodate, A.; Tominaga, Sports video categorizing method using camera motion parameters H. International Conference on Multimedia and Expo, 2003. ICME apos;03. Proceedings. 2003 Volume 2, Issue , 6-9 July 2003 Page(s): II - 461-4 vol.2

12. Statistical analyzing method of camera motion parameters for categorizing sports video Takagi, S.; Hattori, S.; Yokoyama, K.; Kodate, A.; Tominaga, H. International Conference on Visual Information Engineering, 2003. VIE 2003. 7-9 July 2003 pp. 222-225

13. Gillespie, W.J.; Nguyen, D.T., Classification of video shots using activity power flow First IEEE Consumer Communications and Networking Conference, 2004. Volume 5-8 Jan. 2004, Page(s): 336-340

14. Kittler J., Messer K., Christmas W., Levinaise-Obada, Kourbaroulis D. Generation of Semantic Cues for Sports Video Annotation. In ICIP, pp. 26-29, 2001. 7new11. Jaser, E. Kittler, J. Christmas, W., Hierarchical decision making scheme for sports video categorisation with temporal post-processing IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. 27 June-2 July 2004, Vol. 2, pp. II-908- 913



15. Jaser, E. Kittler, J. Christmas, W., Hierarchical decision making scheme for sports video categorisation with temporal post-processing IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. 27 June-2 July 2004, Vol. 2, pp. II-908- 913
16. Choroś K., Pawlaczyk P., Content-Based Scene Detection and Analysis Method for Automatic Classification of TV Sports News. Rough Sets and Current Trends in Computing Lecture Notes in Computer Science Volume 6086, 2010, pp. 120-129
17. Pass G., Zabih R., Miller J. Comparing images using color coherence vectors, In Proc. Proceedings of the fourth ACM international conference on Multimedia (1996)
18. Ionescu B. E., Rasche C., Vertan C., Lambert P., A Contour-Color-Action Approach to Automatic Classification of Several Common Video Genres. Adaptive Multimedia Retrieval. Context, Exploration, and Fusion. Lecture Notes in Computer Science Volume 6817, 2012, pp 74-88
19. Vaswani N., Chellappa R., Principal Components Null Space Analysis for Image and Video Classification IEEE Transactions on Image Processing, Vol. 15, no. 7, July 2006, pp. 1816-1830
20. Koskela M., Sjöberg M., Laaksonen J. Improving Automatic Video Retrieval with Semantic Concept Detection. Lecture Notes in Computer Science Volume 5575, 2009, – pp 480-489
21. Truong B.T., Venkatesh S., Dprai C., Automatic Genre Identification for Content-Based Video Categorization. 15th International Conference on Pattern Recognition (ICPR'00) – Volume 4, p. 4230.22. SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. (1997). Neural Computation, 9(4), 777-804. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9161022>
23. Machajdik J., Hanbury A., Affective Image Classification using Features Inspired by Psychology and Art Theory Proceedings of the international conference on Multimedia, ACM New York, NY, USA 2010. pp. 83-92
24. Li-Jia Li, Hao Su, Eric. P. Xing, Li Fei-Fei. Object Bank: A High-Level Image Representation for Scene Classification and Semantic Feature Sparsification. Proceedings of the Neural Information Processing Systems (NIPS). 2010.
25. Godbole S. Exploiting confusion matrices for automatic generation of topic hierarchies and scaling up multi-way classifiers. Indian Institute of Technology – Bombay. Annual Progress Report. January 2002. <http://www.it.iitb.ac.in/~shantanu/work/aps2002.pdf>
26. Gomez G., Sanchez M., Sucar L.E., On selecting an appropriate color space for skin detection, Springer-Verlag: Lecture Notes in Artificial Intelligence, vol. 2313, 2002, pp. 70–79.

У в а ж а е м ы е а в т о р ы !

Редакция научно-технического журнала «Цифровая обработка сигналов» просит Вас соблюдать следующие требования к материалам, направляемым на публикацию:

1) Рукопись статьи сопровождается:

- краткой аннотацией на русском и английском языках;
- номером УДК;
- сведениями об авторах (Ф.И.О., организация, должность, ученая степень, телефоны, электронная почта);
- ключевыми словами на русском и английском языках;
- актом экспертизы (при наличии в вашей организации экспертной комиссии).

2) Требования к текстовым материалам и сопроводительным документам:

- текст – текстовый редактор Microsoft Word;
- таблицы и рисунки должны быть пронумерованы. На все рисунки, таблицы и библиографические данные указываются ссылки в тексте статьи;
- объем статьи до 12 стр. (шрифт 12). Для заказных обзорных работ объем может быть увеличен до 20 стр;
- название статьи на русском и английском языках.

3) Требования к иллюстрациям:

- векторные (схемы, графики) – желательно использование графических редакторов Adobe Illustrator или Corel DRAW;
- растровые (фотографии, рисунки) – М 1:1, разрешение не менее 300 dpi, формат tiff.