



**М.В. Руцков**

Генеральный директор  
компании MegaPixel Ltd., к.т.н.

**Т**оварищи дорогие! Разрешите всех вас и самого себя, естественно, поздравить с первым микроюбилеем — пятилетием моей дружбы с издательством "Гротек"! Именно в 2003-м году накатал первую статью: "Видеодетекторы — взгляд изнутри". Пять лет — это много или мало? А как посмотреть. В разрезе эволюции, цивилизации и эмансипации, а также мировой революции — тьфу, песчинка! Однако если сосредоточиться на научно-техническом прогрессе, особенно на информационных технологиях — период гигантский! Попробуйте вспомнить, какими были компьютеры пять лет назад. А флешки? Так-то вот. О нашей микроотрасли — "Охранное видеонаблюдение" — уж и не говорю. Всё нарастало со страшной силой: скорости контроля, число каналов, емкость архива и т.д. и т.п.

И всё бы хорошо, только вот с его Величеством Видеодетектором — облом получился! Каким он был (в общей массе) — таким остался. А почему? Да потому, что, видимо, нужен определённый уровень знаний, не обладая которым, невозможно перейти от количества к качеству. С другой стороны — всем уже до лампочки. Главное продавать, а что — не важно! И понеслась — научные разработки заменили PR-мероприятиями со всеми вытекающими последствиями. Короче, как сказал бы великий Аркадий Райкин: "Весь пар ушёл в гудок!!!". Такого надудели — уши завяли у заказчиков от видеоаналитического угара. Журналы псевдонаучные выпускаются, статьи идиотские пишутся..., эх!

Что ж вы, ребята, делаете — техно-похмелье не за горами! Или вы рассчитываете как в кино "Женитьба Бальзаминова": "С деньгами, маменька, мы и без ума проживем"! Купим профессоров с доцентами на корню — они ж такие умные, в Интернете тусуются, всё знают, любые алгоритмы заделают легко, например, чтение мыслей по походке! Пардон — извините, просто нервы ни к чёрту, наболело! А давайте по-простому как-то повышать уровень знаний! Школ, увы, нет, да и при таком подходе не предвидятся. Ну, тогда давайте внести скромный вклад,

# Видеодетекторы: пять лет спустя

## Часть 1

### Статья печатается в авторской редакции

исходя из своего жизненного опыта. Это что-то типа лекций на тему изложения личного видения проблемы. И так, поехали — на второй виток видеоанализа!

Для начала зададимся основным вопросом: "А зачем нам все эти видеодетекторы нужны?". Будем жать да писать! Вот она тупая мечта обольщенного пользователя: "Хочу всё видеть, а иногда и отмотать, чтоб гол забитый смаковать!". А толку-то? Поэтому снова приведу ключевую фразу известного забугорного профи-аналитика Крэйга Дональда.

**С**истемы, стоящие много миллионов долларов, по-прежнему основываются на способности операторов обнаруживать происшествя и способности менеджеров организовать такое обнаружение должным образом. Система видеонаблюдения без эффективного механизма обнаружения подобна автомобилю, двигатель которого включен, но сам он никуда не движется. Все работает, но ничего не достигается!

Вот она, глубочайшая мысль — "обнаружение"!!! Таким образом, видеодетектирование — это и не "интеллектуальный" кураж, и не пиарная блажь, а насущная необходимость! Ну, тогда давайте начнём от обратного: что для реализации алгоритмов видеодетектирования делать не надо, вот так мне кажется. А потом уж плавно перейдём и к другим вариациям на тему.

Итак, первым делом — противоположностью работать по индивидуальному пикселам, типа точечного детектора. Их амплитудные значения сильно "гуляют" по причине флуктуаций освещения, квантовых шумов и просто помех. Необходимо осуществлять обработку определённой локальной области — апертуры, чтобы понять, принадлежит ли конкретный пиксел, например, к контуру. И это надо делать для всех точек изображения — в скользящем режиме, а не интегрально для квадратно-гнездовых областей фрейма. Поверьте — это азы.

Второе — ни в коем случае не вычислять межкадровую разность! После этой процедуры полностью теряется индивидуальность кадров. А кроме того, начинаются проблемы с медленными движениями — разность становится крохотной. Именно на медленных тестах и завалилось

большинство систем — посмотрите соответствующие материалы тестов на выставке "Интерполитех"<sup>2</sup>. Или просто включите какую-нибудь камеру со встроенным детектором. Пока не взмахнёшь рукой, как саблей, — красный квадрат тревоги не появится! А так можно хоть слона медленно внедрять!

Но почему так происходит? Да потому, что межкадровая разность (при одной и той же скорости объекта) становится всё менее заметной с увеличением темпа контроля. При 25-ти кадр/с можно вообще ничего не поймать, а на 3-х кадр/с вроде что-то появится. Тогда некоторые алгоритмисты поступают следующим образом — не обновляют текущее эталонное значение до тех пор, пока не сработает детектор, после чего перезаписывают опорный фрейм. Однако так можно дожидаться и глюка (не путать с известным композитором) от изменения самого фона (светлеет, темнеет, облака, ла-лу-лаЕ). А кроме того, возникает другая проблема. Если в поле зрения — несколько объектов, то всегда будет детектироваться только самый быстрый из них (естественного в угловых скоростях). Тогда начинают мудрить ещё круче — организуют сразу несколько каналов вычисления межкадровой разности с разными дельтами по времени. Короче, всё разрастается, как снежный ком. И возникает дилемма — что считать фоном, а что объектом. Угу — только распознавания ещё и не хватало! Но самое интересное состоит в том, что некоторые объекты, остановившись, по логике вещей должны сами превратиться в фон. Всё это мы перепробовали — толку никакого!!! Вот тогда и пришлось слегка поизучать нейрофизиологию, поскольку самые совершенные видеодетекторы придумала - Матушка Природа!!! И так, начнём потихоньку — с глаза. Сразу привожу фразу из собственного произведения: "Видеодетекторы — взгляд изнутри. Грани интеллекта"<sup>3</sup>:

**Думаю, для большей ясности надо немного рассказать об устройстве глаза, а именно о его видеосенсоре — сетчатке. Опустим физиологические подробности, главное — это геометрия распределения светочувствительных элементов, коими являются палочки (про колбочки, воспринимающие цвет, говорить не будем, поскольку их всего 6 миллионов, что в 20**

<sup>1</sup> [http://www.mpixel.ru/public\\_htm/3.htm](http://www.mpixel.ru/public_htm/3.htm)

<sup>2</sup> [http://www.mpixel.ru/public/test\\_prost.pdf](http://www.mpixel.ru/public/test_prost.pdf)

<sup>3</sup> [http://www.mpixel.ru/public\\_htm/5.htm](http://www.mpixel.ru/public_htm/5.htm)

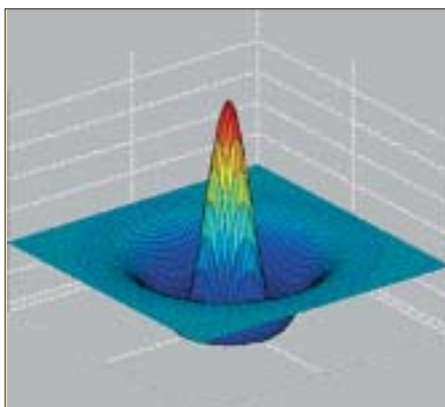
раз меньше). Распределены они крайне неравномерно, причём наиболее густо представлены, в так называемой "центральной ямке". Чтобы было понятно - перевод этой зоны высокого разрешения на рассматриваемый объект, равносильно "наезду" трансфокатора с 30–35-кратным увеличением.

Все эти данные позаимствованы у великого нейрофизиолога, лауреата Нобелевской Премии Дэвида Хьюбела из его работы "Глаз, Мозг, Зрение" — фантастическая книга с картинками. Кому интересно — почитайте, очень увлекательно. Ну а если времени нет, приведу ещё пару фактов. Во-первых, сетчатка — это часть мозга, вынесенная на периферию, причём уже на её уровне осуществляется неслабая обработка. Ещё лет 50–60 назад считали, что зрительный нерв просто передаёт информацию с рецепторов (палочек и колбочек) прямо в мозг! Цепляли микроэлектроды и тихо обалдевали — ни прямая фоновая засветка, ни даже магниевая вспышка не давали никакого эффекта. Это вам не фотодиоды с проводами. И лишь потом установили, что нейронная активность начинается при предъявлении специфических световых стимулов — светлых или тёмных пятнышек!

Во-вторых, выяснилось: аксонов (выходов с ганглиозных клеток), составляющих зрительный нерв, всего-то около миллиона, притом, что самих рецепторов около 125 миллионов. Куда чего подевалось? И так, что же "видит" каждая ганглиозная клетка, т.е. какая информация передаётся по аксонам зрительного нерва? Появляется понятие — "рецептивное поле". Снова отсылаю к вышеназванной книге (стр. 50):

**Термин "Рецептивное поле" в узком смысле означает просто совокупность рецепторов, посылающих данному нейрону сигналы через один или большее число синапсов. В зрительной системе это всего лишь некоторая область сетчатки...**

Ну а дальше — сами почитайте. Короче, в сетчатке рецептивные поля имеют форму круга с противоположенными по знаку воздействиями в центре и периферии. Причём это не означает, что все сигналы от рецепторов в центре просто суммируются с одним знаком, а от периферийных — с другим. Нет, они ещё и перемножаются на коэффициенты со знаком (положительным или отрицательным). Если построить их 3D-график, то получится что-то типа "мексиканской шляпы". Сумма всех коэффициентов равна нулю!



Вот почему фоновая засветка и магниевая вспышка не давали никаких результатов. Далее, эти рецептивные поля не напиханы бок о бок друг другу встык, а расположены с сильным перекрытием по всему полю сетчатки. Более того, они ещё отличаются и по размерам. В одной из моделей зрения человека (построенной на базе нейрофизиологических исследований) предполагается наличие 4-х типоразмеров рецептивных полей. Они увеличиваются линейно, примерно с коэффициентом 2. Таким образом, наше зрение многоканально — мы как бы "видим" одновременно несколько картинок с разным пространственным разрешением. Такие вот чудеса! Сейчас покажу.

Буду, как врач-вирусолог, измываться над самим собой — беру собственное фото из своей же статьи про практическую плоскость в видеодетектировании. Вот — пальцем шевелю (красненьким — детектор сработал):



Фото 1. Исходное изображение

А теперь пропускаю сие через три "мексиканские шляпы" с приблизительным наружным диаметром (размер рецептивного поля) — 10, 20 и 40 пиксел соответственно:



Фото 2. Рецептивное поле 10 пиксел



Фото 3. Рецептивное поле 20 пиксел



Фото 4. Рецептивное поле 40 пиксел

Так вот, сия процедура называется полосовой фильтрацией. Что-то типа эквалайзера, но не звукового, а для изображений, которые так же можно разложить на частоты, но уже пространственные. Чем они выше — тем за более мелкие объекты в ответе, и не только. Высокие частоты присутствуют ещё и на границах крупных образований, где скачкообразно меняется яркость. Теперь о "мексиканской шляпе". По-научному она называется "Лапласиан от Гауссиана". Этот оператор обладает сразу двумя свойствами — интегрирующим и дифференцирующим. Первое помогает давить шумы и помехи типа "снега", второе — подчёркивает контуры. А они нам как раз и нужны! Если контуры стоят на месте, значит движения нет.

Обратите внимание — полосовая фильтрация ещё уничтожает постоянную составляющую! Ну это когда меняется освещённость — солнце за облако уехало, фонари моргают и т.д. А как же всё это работает? Над изображением выполняется процедура свёртки — convolution (не путать со сворачиванием плащ-палатки). Например, в нашей системе берётся ядро размером 13x13 пиксел (и это лишь для высокочастотного канала), в котором заданы коэффициенты "шляпы". Далее оно ездит по всем без исключения местам входного изображения. Встали на точку, да вычислили сумму произведений коэффициентов на значения окрестных пикселов, записали в выходной фрейм (в одноимённую точку), сместились, снова вычислили и так до упора. Естественно, возникает вопрос — а у процессора морда не треснет с перенапрягу? Вот в следующий раз и проанализируем. ■

Ваше мнение и вопросы по статье направляйте на [ss@groteck.ru](mailto:ss@groteck.ru)

\* <http://nehudlit.ru/books/detail7111.html>