

РАЗРАБОТКА МЕТОДИКИ АВТОМАТИЧЕСКОГО ТЕСТИРОВАНИЯ АЛГОРИТМОВ ПОИСКА ЧЕЛОВЕКА ПО ПРИМЕТАМ

Андрей Дураков

Аннотация: В статье рассматривается проблема автоматического тестирования алгоритмов поиска человека по приметам. Особое внимание уделяется испытаниям этих алгоритмов в реальных условиях. Сформулированы ключевые показатели для оценки производительности и качества работы алгоритмов с учетом пользовательского опыта. Для оценки производительности предлагается измерять два параметра: скорость построения одного индекса человека и скорость сравнения двух индексов. Измерения производительности проводятся относительно эталонного алгоритма, который длительное время испытывался в реальных условиях. Качественными показателями работы алгоритмов являются доля выраженных ошибок в результатах поиска и доля камер, на которых искомый объект был найден. Именно эти два показателя сложнее всего измерить автоматически ввиду их субъективной природы. Требование репрезентативности тестовой выборки, т.е. соответствия ее реальным условиям, дополнительно усложняет задачу измерения. В качестве тестовой выборки предложено использовать большое количество видеороликов, снятых в различных условиях освещенности и на камеры различных моделей и производителей. Для измерения качественных показателей предложена методика, основанная на предварительной разметке тестовых видеороликов человеком и использовании модуля трекинга людей в рамках одной камеры. В разметку закладываются человеческие субъективные оценки степени похожести людей на видеороликах. Разметка видеороликов состоит из двух частей: отметка временных границ движения отдельных людей и заполнение матрицы сравнения людей между собой. Модуль трекинга необходим для сопоставления людей в видеороликах и людей, проиндексированных в системе видеонаблюдения. После разметки вся тестовая выборка подается на вход системы видеонаблюдения, где она индексируется. После этого происходит сопоставление людей в размеченных видеороликах и архиве, а на основе сопоставления автоматически рассчитываются качественные показатели.

Ключевые слова: видеоанализ, поиск человека по приметам, автоматическое тестирование.

ACM Classification Keywords: D.2 SOFTWARE ENGINEERING: D.2.5 Testing and Debugging – Testing tool; D.2.8 Metrics – Performance measures. I.5 PATTERN RECOGNITION: I.5.4 Applications – Computer vision.

Введение

В современном мире системы видеонаблюдения прочно вошли в нашу жизнь. Трудно себе представить крупное предприятие, банк или торговый центр, не использующие видеонаблюдение в своей системе безопасности. В крупных городах видеокамеры нередко встречаются в жилых домах и даже в отдельных квартирах, не говоря уже об офисных и административных зданиях.

В последние десятилетия главным трендом в развитии систем видеонаблюдения является повышение их интеллектуального уровня [Дее, 2008]. Уже сейчас на рынке систем видеонаблюдения представлено большое количество программных продуктов, выполняющих не только хранение и отображение видеоданных, но и различного рода анализ. Такие продукты включают в себя специальные

интеллектуальные модули, которые позволяют решать задачи обнаружения и распознавания лиц, перехвата объектов, интерактивного поиска, распознавания номеров и т.д.

Одной из самых востребованных функций на данный момент является поиск человека по приметам, который работал бы приемлемо в различных условиях освещенности, на массовых камерах низкого разрешения и не требовал бы больших вычислительных мощностей.

Если качество видеокамер недостаточно хорошее, то методы распознавания людей по лицу не работают из-за невозможности четко выделить на лице отдельного человека ключевые признаки, на основе которых и происходит сравнение [Viola, 2004]. В связи с этим необходимо использовать другие характерные особенности внешнего вида людей и различать людей по ним. Такими особенностями могут быть, например, основные цвета одежды и их взаимное расположение, а также паттерн, то есть узор одежды.

В этом случае задача поиска человека по приметам предполагает поиск не одного заданного человека, а всех похожих на него людей, включая и искомого. Такое смягчение условий задачи является вынужденным следствием требования об использовании массовых IP-камер, не обеспечивающих возможности точно идентифицировать человека по лицу. Важно отметить, что образец может быть задан самыми разными способами: кадр из архива системы видеонаблюдения, фотография, сделанная на мобильный телефон или фотоаппарат, шаблон, заданный в виде рисунка, и даже описание на естественном языке.

Таким образом, ставится задача разработки программного комплекса видеонаблюдения, который должен по заданному образцу найти похожих людей во множестве видеопотоков с нескольких камер, выделить их и показать охраннику только те кадры, где присутствует эти люди.

Эта функция является крайне востребованной, т.к. ее решение значительно повышает мобильность службы безопасности, увеличивая скорость обработки большого объема видеоданных и, тем самым, уменьшая время реакции охраны с нескольких часов или даже дней до нескольких минут.

Несмотря на большую значимость этой функции и большое количество исследований по всему миру эта задача до сих пор не решена в полной мере. Одной из причин, препятствующих ее решению, является сложность оперативной оценки качества, а главное, применимости разрабатываемых учеными алгоритмов в реальных условиях. Под реальными условиями будем понимать типичные условия, в которых работают системы видеонаблюдения, а именно:

- камеры бывают различного качества;
- камеры установлены под различными углами и на различном расстоянии относительно охраняемой территории (отчего сильно варьируется размер и ракурс людей на изображении);
- камеры размещены как в помещениях, так и на улице;
- камеры находятся в разных условиях освещенности;
- освещенность периодически меняется (рассвет, день, сумерки);
- изображение с камер подвержено большому количеству внешних помех: дождь, снег, листопад, пыль, ветер, трясущий камеру, блики от окружающих предметов и т.д.

Для того чтобы учесть все особенности реальных условий функционирования систем видеонаблюдения при разработке новых алгоритмов, необходимо иметь не только обширную тестовую выборку и способ измерения качества работы алгоритма, но и инструменты, автоматизирующие процесс оценки качества. Автоматизация процесса тестирования является отдельной сложной задачей [Greiffenhagen, 2001].

В данной статье пойдет речь о создании методики тестирования комплекса алгоритмов видеонаблюдения, предназначенных для решения задачи поиска человека по приметам.

Измеряемые показатели

Прежде всего, необходимо подчеркнуть, что предметом тестирования будет являться целый комплекс алгоритмов видеоанализа, которые работают над решением одной задачи и выдают единый для всех результат. Тестируемый комплекс алгоритмов содержит двух ключевых функций:

- индексирование человека, которое включает в себя следующие этапы:
- детектирование движения и обнаружение теней;
- фильтрация шумов;
- цветовая коррекция изображения;
- выделение и сохранение ключевых признаков;
- сравнение двух индексов.

Таким образом, предметом оценки и будет являться качество работы двух указанных функций.

Для выделения ключевых показателей, которые необходимо измерять, будем исходить из потребностей пользователей системы видеонаблюдения, где и должны работать все эти алгоритмы. Если абстрагироваться от требований к удобному визуальному интерфейсу поиска, который не является предметом исследования, то остаются два ключевых требования:

- требование к производительности системы;
- требование к качеству поиска.

Требование к производительности системы означает, что работа алгоритмов индексирования людей не должна значительно ухудшать производительность системы видеонаблюдения в целом. А также, что время поиска человека по приметам должно быть достаточно мало, чтобы обеспечивать возможность службе безопасности оперативно реагировать на обнаружение искомого человека.

Требование к качеству поиска означает, что среди результатов поиска человека по приметам обязательно должен присутствовать искомый человек на абсолютном большинстве камер, но при этом процент ошибок (людей, абсолютно не похожих на искомого визуально) не должен превышать некоторого порога, после которого пользователь теряет доверие к функции поиска.

Формализуя требование к качеству поиска, удалось выделить следующие четыре категории объектов, встречающихся в результатах поиска:

1. Искомый объект.
2. Объект, визуально похожий на искомый.
3. Объект, визуально не похожий на искомый, но при этом имеются видимые основания, которые позволяют объяснить пользователю причину ошибки. Например, объект находится в толпе людей, в кадре присутствуют помехи, в маску движения объекта попала его тень или часть фона. Будем называть такие объекты «объяснимо не похожими».
4. Объект, визуально не похожий на искомый, и при этом нет никаких оснований, которые бы позволили объяснить пользователю причину ошибки. Будем называть такие объекты «абсолютно не похожими».

В процессе формализации требований удалось также выявить два условия, при которых уровень доверия пользователей к результатам функции поиска человека по приметам будет достаточным для ее использования вместо поиска, осуществляемого человеком-оператором системы видеонаблюдения. Во-первых, доля камер, на которых был обнаружен искомый объект, среди всех, на которых он попал в поле зрения, должна составлять не менее 80%. Во-вторых, доля абсолютно не похожих на искомый объектов в результатах поиска не должно превышать 15%.

Таким образом, можно выделить следующие измеряемые показатели:

- $T1$ – скорость построения индекса одного объекта;
- $T2$ – скорость сравнения двух индексов объектов между собой;
- $K1$ – процент камер, на которых был обнаружен искомый объект, среди всех камер, где он попал в поле зрения;
- $K2$ – процент ошибочно найденных объектов.

Методология измерения производительности

Показатели $T1$ и $T2$ следует измерять таким образом, чтобы измерения, сделанные в различное время и на различных вычислительных системах, были сравнимы между собой. Этого можно добиться, если измерять показатели не в абсолютных величинах (время выполнения), а в относительных. В качестве такого эталона может выступать одна из реализаций всего комплекса алгоритмов, решающих задачу поиска человека по приметам. Причем, это должна быть реализация, которая прошла проверку в реальных условиях и удовлетворяет требованиям производительности, сформулированным ранее. Так как все алгоритмы проходят проверку в системе видеонаблюдения Macroscop, имеющей открытую архитектуру, то за эталон была взята реализация функции поиска человека по приметам из версии Macroscop 1.2, которая работает на большом количестве объектов в реальных условиях уже больше года.

Пусть $t1_{\text{эталон}}$ и $t2_{\text{эталон}}$ – время индексирования человека и сравнения индексов эталонным алгоритмом соответственно; $t1_{\text{тестируемый}}$ и $t2_{\text{тестируемый}}$ – время индексирования человека и сравнения индексов тестируемым алгоритмом соответственно. Тогда: $T1 = \frac{t1_{\text{эталон}}}{t1}$, а $T2 = \frac{t2_{\text{эталон}}}{t2}$.

Все перечисленные замеры должны производиться на единой тестовой базе, которая может со временем пополняться. Так, в процессе тестирования выяснилось, что скорость индексирования алгоритмов значительно зависит от размера движущихся объектов, в частности, большой вклад вносят различные фильтры изображений. Поэтому измерения показателей $T1$ и $T2$ производились на различных тестовых выборках (рис.1), которые были ранжированы по размерам движущихся объектов. Тестовые выборки состоят из кадров и масок движения, полученных с помощью системы Macroscop в реальных условиях, а также из некоторого количества (10-20%) искусственно созданных образцов.

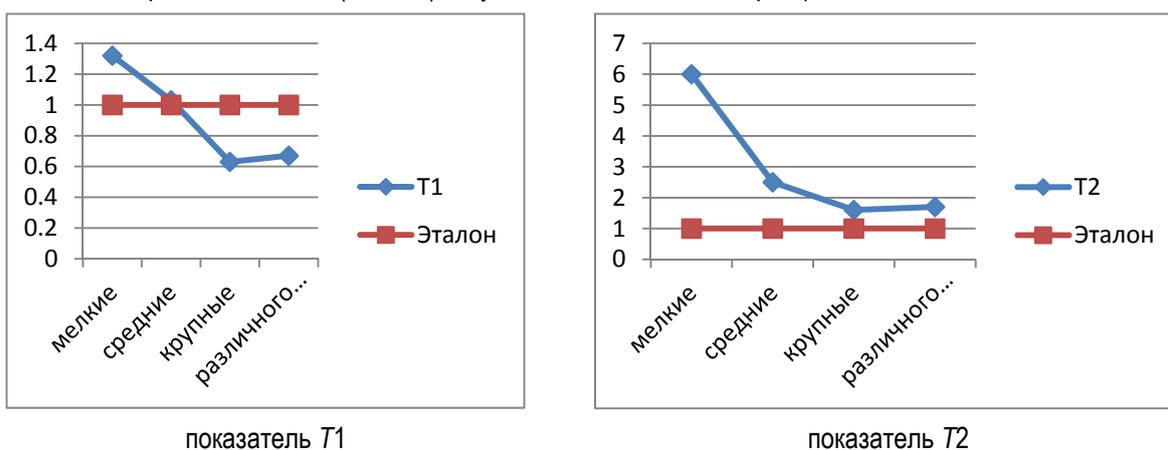


Рисунок 1. Пример сравнения показателей производительности текущей версии алгоритма относительно эталона

Для повышения точности измерений необходимо множество раз запускать тестируемые функции и усреднять значения измерений. В примере (рис.1) измерения проводились 100 раз.

Методология измерения качества работы

На данный момент накоплен некоторый опыт измерения качества работы алгоритмов поиска людей по приметам вручную. Например, измерение показателей в системе Macroscop на выборке из 15 людей, заснятых на 2 камеры (всего 30 роликов), заняло 3 часа рабочего времени у двух человек (6 человеко-часов). Такие показатели скорости измерений не являются приемлемыми для больших выборок и оперативной оценки результатов изменений алгоритмов. Тем более что тестовая выборка должна быть близка к реальным условиям, описанным ранее, а это неминуемо ведет к значительному увеличению количества тестовых роликов и невозможности за разумные сроки измерить показатели алгоритмов.

Определим состав тестовой выборки. Прежде всего, это видеоматериал, который записан на видеокамерах, работающих в реальных условиях. Половина тестовой выборки должна состоять из видеороликов, на которых присутствуют одни и те же люди. Например, один человек проходит в поле зрения 10 камер, находящихся в разных условиях освещенности. Это необходимо для расчета показателя $K1$. Вторая же часть выборки предназначена для моделирования разнообразия людей, попадающих в объектив, и должна содержать видеоролики большого количества людей: от похожих до колоритных (явно выделяющихся из толпы). Причём каждый из этих людей должен попадать в объектив одной или двух камер, но не более. Кроме того, на одном видеоролике могут встречаться несколько людей. Это значительно усложняет задачу тестирования, но является принципиально важным условием, которое приближает тестовую выборку к реальности. Вторая часть выборки необходима для более точного измерения показателя $K2$.

На данном этапе разработки методологии автоматического тестирования алгоритмов можно предположить характеристики репрезентативной тестовой выборки:

- все видеоролики должны быть получены с 10 камер, находящихся в различных условиях освещенности и установленных как в помещениях, так и на улице;
- 10 абсолютно не похожих человек должны быть отсняты на 5 случайных камерах каждый;
- ещё 50 роликов с участием разнообразно одетых людей должно быть отснято на тех же камерах.

Всего в репрезентативной тестовой выборке должно быть порядка 100 роликов. Отметим, что типичное время присутствия человека в поле зрения одной камеры составляет от 5 до 20 секунд. Это означает, что типичное количество индексируемых кадров для одного человека составляет от 50 до 200 (при темпе записи 10 кадров в секунду). Тогда для всей тестовой выборки получим от 5 000 до 20 000 индексируемых кадров.

После того как отсняты видеоролики для тестовой выборки, необходимо установить отношения похожести между людьми в видеороликах, т.е. отнести их друг относительно друга в одну из четырех выявленных ранее категорий. Эти отношения необходимы для подсчета показателей $K1$ и $K2$. Для этого сначала необходимо выделить людей в видеороликах. Выделение производится вручную и заключается в отметке времени, когда объект появился в кадре, и когда он покинул кадр. Пример разметки приведён на рис. 2.

На рис. 2 видно, что объекты Чел1А и Чел3А – это один и тот же человек. Таким образом, необходимо учитывать, что человек может появляться в видеоролике несколько раз и создавать несколько интервалов для него. Для этих интервалов необходимо задать отношение похожести из группы 1.

Для 100 объектов необходимо заполнить матрицу отношений похожести 100×100 из 10 000 ячеек, а в силу ее симметрии относительно главной диагонали необходимо задать 4 950 отношений похожести, что

является довольно трудоемким процессом. Тем не менее, такие усилия оправданы, т.к. эта процедура производится всего один раз.

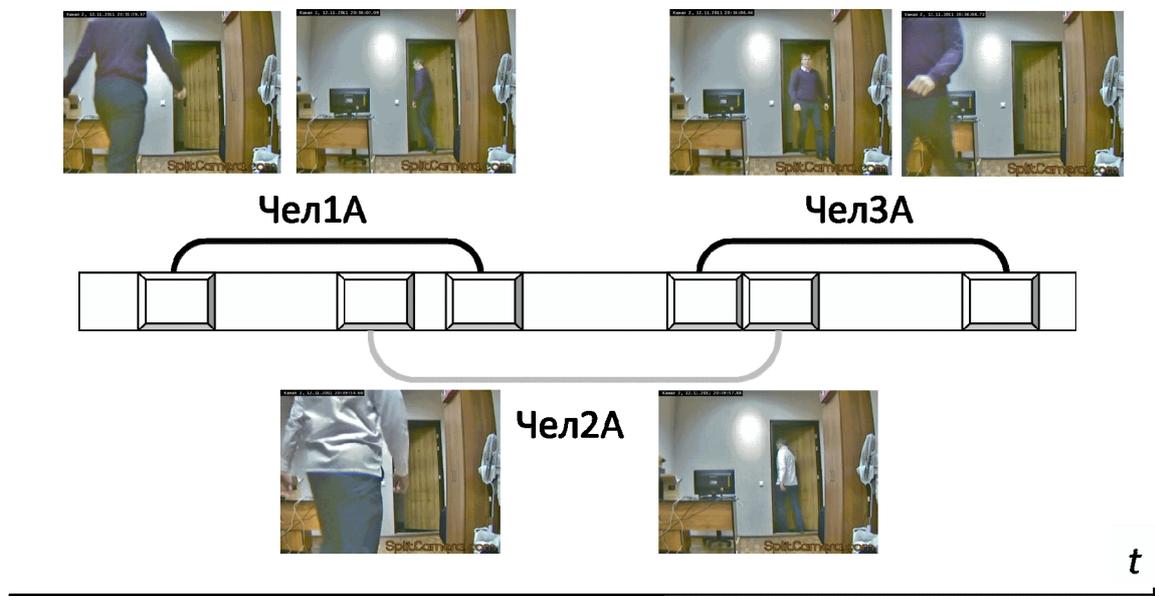


Рисунок 2. Предварительная разметка видеоролика

Разметка видеороликов и матрица соотношений похожести сохраняется в специальном файле метаданных тестовой выборки. Далее подготовленные видеоролики разбиваются на кадры и подаются на вход системы видеонаблюдения Macroscop через специальный плагин-камеру SampleReader (рис. 3). Каждому ролику соответствует свой плагин, работающий в отдельном потоке.

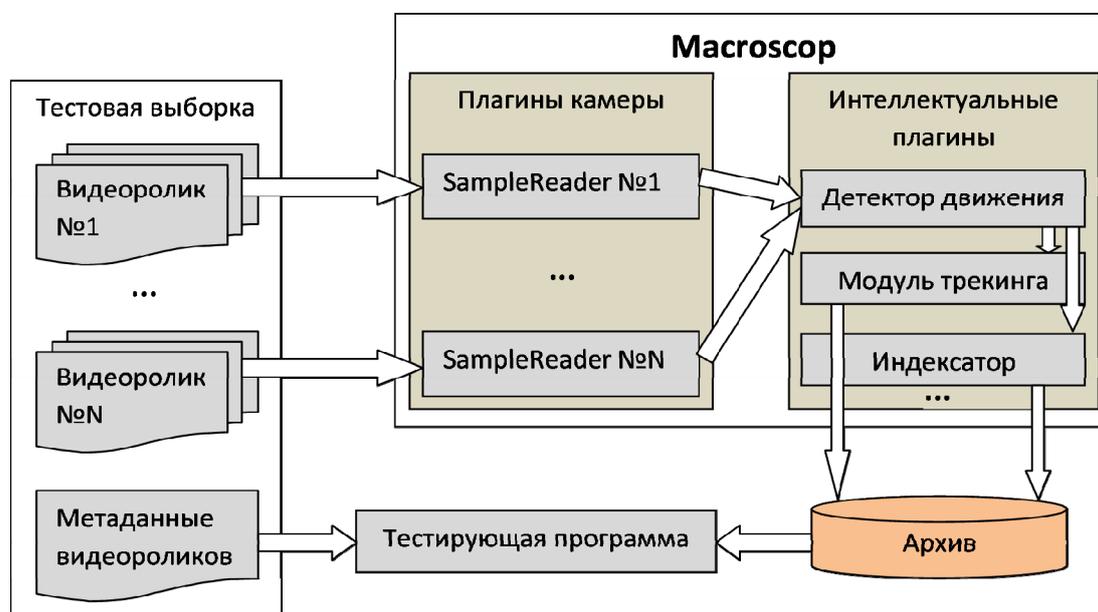


Рисунок 3. Схема работы механизма автоматического тестирования алгоритмов поиска человека по приметам

SampleReader эмулирует работу камеры и подаёт подготовленные кадры на вход интеллектуальных модулей, качество работы которых и предстоит оценить. Результаты работы интеллектуальных модулей поступают в архив.

После того как все видеоролики будут проиндексированы, тестирующая программа на основе файла метаданных тестовой выборки и результатов работы модуля трекинга производит сопоставление движущихся объектов. Таким образом, для каждого индекса в архиве становится известно, какому объекту в видеороликах он соответствует. Это означает, что любые два индекса можно сравнить, а затем, используя матрицу похожести, оценить правильность работы алгоритма сравнения и рассчитать показатели $K1$ и $K2$.

Заключение

В статье подробно рассмотрена проблема автоматического тестирования и оценки производительности и качества работы алгоритмов поиска человека по приметам. На основе требований к функции поиска, выдвигаемых пользователями, разработаны показатели $T1$, $T2$, $K1$ и $K2$, которые необходимо измерять. Формализованы требования пользователей к этим показателям.

Разработана методика измерения показателей производительности ($T1$ и $T2$) алгоритмов поиска человека по приметам на основе эталонного алгоритма. Реализовано вспомогательное программное обеспечение, реализующее предложенный подход. В процессе тестирования оно позволило выявить некоторые зависимости. В частности, зависимость относительной скорости работы различных алгоритмов от размера движущихся объектов.

Разработана методика автоматического измерения качественных показателей ($K1$ и $K2$) работы алгоритмов поиска человека по приметам. Предложен механизм формирования тестовой выборки видеороликов.

Данная статья была подготовлена при участии сотрудников компании ООО «Сателлит» – разработчика системы видеонаблюдения для IP-камер Macroscop.

Библиографический список

[Dee, 2008] Hannah M. Dee and Sergio A. Velastin. How close are we to solving the problem of automated visual surveillance?: A review of real-world surveillance, scientific progress and evaluative mechanisms. Mach. Vision Appl. 19, 5-6 (September 2008), 329-343. DOI=10.1007/s00138-007-0077-z <http://dx.doi.org/10.1007/s00138-007-0077-z>

[Greiffenhagen, 2001] M. Greiffenhagen, D. Comaniciu, H. Niemann, and V. Ramesh. Design, analysis and engineering of video monitoring systems: An approach and case study. Proc. IEEE, vol. 89, no. 10, pp. 1498-1517.

[Viola, 2004] Paul Viola and Michael J. Jones. Robust Real-Time Face Detection. Int. J. Comput. Vision 57, 2 (May 2004), pp. 137-154. DOI=10.1023/B:VISI.0000013087.49260.fb <http://dx.doi.org/10.1023/B:VISI.0000013087.49260.fb>

Сведения об авторах



Андрей Дураков – Заведующий лабораторией инструментальных средств разработки программного обеспечения кафедры математического обеспечения вычислительных систем ПГНИУ; 614990, Россия, г. Пермь, ул. Букирева, 15; e-mail: adurakov@gmail.com.

Сфера научных интересов: Распознавание образов, Интеллектуальный видеонализ, Высокопроизводительные вычисления, Дополненная реальность.