УДК: 004.932

К вопросу о качестве работы алгоритмов слежения за объектами

А. А. Шапошников, Е. В. Шапошникова, А. И. Шапошников

НИ Томский государственный университет,
Россия, 634050, г. Томск, проспект Ленина, д. 36.

E-mail: AlbertElena@mail.ru

*Получено 04 февраля 2014 г.,
после доработки 10 августа 2014 г.*

Разобран алгоритм трекинга, позволяющий в процессе слежения учитывать независимые изменения вертикального и горизонтального размеров и ориентации объекта слежения. Показано, что в процессе слежения определяющим является учет всех характеристик области слежения, второстепенным — предсказание положения объекта.

Ключевые слова: слежение, ядро области слежения, трекинг

About quality of Kernel based object tracking

A. A. Shaposhnikov, E. V. Shaposhnikova, A. I. Shaposhnikov

Tomsk state university, 36 Lenin avenue, Tomsk, 634050, Russia

Abstract. — The kernel based object tracking algorithms were described that take in account the independent changes of the 4 and 5 out of 5 parameters of the elliptic tracking region. It is shown that in tracking this conditions are sufficient and attempts of prediction are not necessary.

Keywords: automatic tracking, kernel based object tracking

Citation: *Computer Research and Modeling*, 2014, vol. 6, no. 4, pp. 495–502 (Russian).

Введение

Задача слежения за объектами (трекинг объектов) — одна из важных задач современного анализа видеоизображений. В различных приложениях часто используются частные ее решения. В последние 20 лет обнажились затруднения при слежении за объектами (втомобилями, людьми, животными) размеры которых достаточно велики, чтобы различать их форму, цветовое разнообразие, отдельные элементы, а движения слабо прогнозируемы или вовсе непредсказуемы. В то же время, человек, наблюдающий видео, легко решает задачу слежения именно за такими объектами — крупными, хорошо различимыми и при этом движущимися относительно медленно и не однообразно (регулярно). Человек без проблем справляется с ситуациями, когда объект слежения внезапно меняет направление движения, поворачивается, наклоняется, снимает шапку, размахивает руками или распахивает пиджак — ситуациями, для которых в компьютерном слежении вошел в обиход даже специальный термин — окклюзии (occlusions) [Maggio, Cavallaro, 2010], их обнаружению посвящен ряд работ [Вражнов, Шаповалов, Николаев, 2012]. С другой стороны, использование модификаций тех механизмов видеоаналитики, которые уже разработаны [Comaniciu, Ramesh, Meer, 2003; Калайда, Елизаров, 2012] их дополнений до завершеного вида, часто приводят к положительным результатам и необходимы с теоретической точки зрения.

Задача слежения может быть естественным образом разбита на захват объекта — выделение объекта в первом кадре последовательности некоторым окном и смещение с трансформацией этого окна на последующих кадрах. В данной статье решается вторая часть задачи — выделение подходящих для слежения параметров окна и алгоритм подгонки этих параметров в процессе смены кадров. Представлена модификация известного метода сдвига среднего [Comaniciu, Ramesh, Meer, 2003], позволяющая проводить трекинг объекта в условиях, когда объект движется с многочисленными коллизиями — бывает частично или полностью загорожен другими объектами, смещается, будучи загороженным. Идея метода — представить трекинг как свертку двух функций, описывающих изображение и объект соответственно. При этом все «плохие» свойства функции изображения остаются при ней, а для описания объекта подбирается достаточно удобная функция, имеющая эллиптический носитель. Идея подбора именно такой функции и приемы дифференцирования именно такой свертки взята из [Дерюгин, Суворов, 2012]. Для апробации использованы, в частности, материалы из этого же журнала [Вражнов, Шаповалов, Николаев, 2012], где теоретически обоснована сложность этих материалов для обработки. Результаты апробации выложены здесь [Шапошников, 2013а; Шапошников, 2013b; Шапошников, 2013c; Шапошников, 2014а; Шапошников, 2014b]. Программа составлена и работала в интерпретаторе Octave, где время счета составило около 80 миллисекунд на кадр. Количество операций при обработке одного кадра пропорционально количеству цветов в палитре и размеру объекта слежения и не зависит от номера кадра (как при таксономии) и размера кадра. Алгоритм слежения зарегистрирован в ФИПС [Шапошникова, 2013] и ОФЕРНИО [Шапошников, Шапошникова, 2012].

Описание объекта цветовой гистограммой и построение функции схожести

Одно изображение — кадр из видео или фотографию описываем функцией $f: P \rightarrow B$, где $P = [0, d_1] \times [0, d_2]$ — прямоугольник, B — конечное множество. Обозначим через $\chi_{f^{-1}(b)}(x)$, где $x \in P$, характеристическую функцию полного прообраза $f^{-1}(b)$ элемента $b \in B$, то есть $\chi_{f^{-1}(b)}(x) = 1$, если $f(x) = b$, и $\chi_{f^{-1}(b)}(x) = 0$ в противном случае. Конкретно в нашей задаче элементы $b \in B$ суть интенсивности трех цветовых каналов R, G и B, то есть трехмерные векторы.

Для каждого $b \in B$ определим величину $H_b = \iint_P \chi_{f^{-1}(b)}(x) dx_1 dx_2$. Можно предполагать, что этот интеграл существует (это выполнено, если цветовые пятна плотные и непрерывные, то есть множество точек разрыва функции $\chi_{f^{-1}(b)}(x)$, а все они первого рода, имеет меру нуль). Множество $\{H_b | b \in B\}$ называется гистограммой кадра.

Введем в рассмотрение семейство функций $g(x) = \begin{cases} 0, & \text{если } 1 - \left(\frac{x_1 - y_1}{l_1}\right)^2 - \left(\frac{x_2 - y_2}{l_2}\right)^2 < 0; \\ 1 - \left(\frac{x_1 - y_1}{l_1}\right)^2 - \left(\frac{x_2 - y_2}{l_2}\right)^2 & \text{в противном случае,} \end{cases}$ где $x = (x_1, x_2) \in P$, x — переменная; $y = (y_1, y_2) \in P$, $l_1, l_2 \in R$, y_1, y_2, l_1, l_2 — параметры.

Пусть теперь $G \subset P$, G — эллиптическая область с центром в точке $y = (y_1, y_2)$ и полуосями l_1, l_2 . Выбор эллиптической области обусловлен тем, что объект в кадре определяется своим центром $y = (y_1, y_2)$ и размерами $l = (l_1, l_2)$, то есть эллиптическая область хорошо подходит для выделения объекта в кадре.

Для определения гистограммы эллиптической области G рассмотрим функции $H_b(y, l) = \iint_P \chi_{f^{-1}(b)}(x) g(x) dx_1 dx_2$, где $y = (y_1, y_2)$, $l = (l_1, l_2)$. Величина $H_b(y_0, l_0)$ есть вес той части области G с центром в точке y_0 и размерами l_0 , которая имеет данный цвет b , — вес цвета b . Цвет, который в центре области, вносит больший вклад в гистограмму. Заметим, что $\sum_{b \in B} H_b$ есть величина, характеризующая область G . Будем называть ее весом области. Тогда множество чисел $H = \{H_b | b \in B\}$ будет гистограммой области. Ясно, что гистограмма H изменяется при изменении положения центра $y = (y_1, y_2)$ и размеров $l = (l_1, l_2)$ области.

Однако для построения функции схожести сами гистограммы $H = \{H_b | b \in B\}$ непригодны, так как H_b существенно зависят от размеров l области. А при пропорциональном увеличении или уменьшении (масштабировании) изображения объекта функция схожести должна идентифицировать получившиеся изображения как изображения одно и того же объекта. Поэтому рассмотрим $h_b = \frac{H_b}{\sum_{b \in B} H_b}$ — отношение веса цвета b к весу всей области, то есть относительный вес цвета.

Тогда $h = \{h_b | b \in B\}$ — нормированная гистограмма области G . Ясно, что нормированная гистограмма h изменяется при изменении $y = (y_1, y_2)$ и $l = (l_1, l_2)$, то есть является функцией переменных y_1, y_2, l_1, l_2 . Для двух нормированных гистограмм h_0 и h определим, следуя [Comaniciu, Ramesh, Meer, 2003], функцию схожести $\rho = \sum_{b \in B} \sqrt{h_{0b} \cdot h_b}$, где $h_0 = \{h_{0b} | b \in B\}$ — некоторая гистограмма, с которой производим сравнение.

Тогда ρ есть функция аргументов $y = (y_1, y_2)$ и $l = (l_1, l_2)$, зависящая от параметра h_0 . Для этой функции $\rho = \rho(y, l)$ получаем $\rho: R^3 \rightarrow R$, если размеры объекта l_1 и l_2 связаны каким-либо соотношением, например, пропорциональны как при приближении одноцветного шара. Именно такой случай рассмотрен в историческом цикле работ Петера Меера и Дорина Коменечи [Comaniciu, Ramesh, Meer, 2003; Comaniciu, Meer, 2002]. А именно: для определения параметра h («bandwidth») предложено проводить серию расчетов с различными значениями и выбирать больший, то есть находить его перебором?

Или $\rho: R^4 \rightarrow R$, если l_1 и l_2 не зависят друг от друга,

или $\rho: R^5 \rightarrow R$, если ввести еще один параметр α , который обеспечивает изменение угла поворота области. При введении в рассмотрение угла поворота α , $\alpha \in [-90^\circ, 90^\circ]$, в равенстве $H_b(y, l) = \iint_P \chi_{f^{-1}(b)}(x) g(x) dx_1 dx_2$ в качестве функции $g(x)$ следует использовать

$$g(x) = \begin{cases} 0, & \text{если } 1 - \left(\frac{(x_1 - y_1) \cos \alpha - (x_2 - y_2) \sin \alpha}{l_1}\right)^2 - \left(\frac{(x_1 - y_1) \sin \alpha + (x_2 - y_2) \cos \alpha}{l_2}\right)^2 < 0; \\ 1 - \left(\frac{(x_1 - y_1) \cos \alpha - (x_2 - y_2) \sin \alpha}{l_1}\right)^2 - \left(\frac{(x_1 - y_1) \sin \alpha + (x_2 - y_2) \cos \alpha}{l_2}\right)^2 & \text{в противном случае.} \end{cases}$$

Градиентный спуск

Найдем градиент функции схожести $\rho: R^4 \rightarrow R$ $\text{grad} \rho = \left(\frac{\partial \rho}{\partial y_1}, \frac{\partial \rho}{\partial y_2}, \frac{\partial \rho}{\partial l_1}, \frac{\partial \rho}{\partial l_2} \right)$. Для этого запишем $\rho = \sum_{b \in B} \rho_b$, где $\rho_b = \sqrt{h_{0b} \cdot h_b}$.

Теперь находим

$$\frac{\partial g}{\partial y_i} = \frac{2\chi_{\text{supp}(g)}(x_i - y_i)}{l_i^2}, \quad \frac{\partial g}{\partial l_i} = \frac{2\chi_{\text{supp}(g)}(x_i - y_i)^2}{l_i^3}, \quad i = 1, 2,$$

где $\text{supp}(g)$ — носитель функции g , то есть множество точек, в которых функция принимает ненулевые значения. Отсюда получаем

$$\begin{aligned} \frac{\partial H_b}{\partial y_i} &= \iint_P \chi_{f^{-1}(b)}(x) \frac{\partial g}{\partial y_i} dx_1 dx_2, & \frac{\partial H_b}{\partial l_i} &= \iint_P \chi_{f^{-1}(b)}(x) \frac{\partial g}{\partial l_i} dx_1 dx_2, \\ \frac{\partial h_b}{\partial y_i} &= \frac{\frac{\partial H_b}{\partial y_i} - h_b \sum_{b \in B} \frac{\partial H_b}{\partial y_i}}{\sum_{b \in B} H_b}, & \frac{\partial h_b}{\partial l_i} &= \frac{\frac{\partial H_b}{\partial l_i} - h_b \sum_{b \in B} \frac{\partial H_b}{\partial l_i}}{\sum_{b \in B} H_b}, \quad i = 1, 2. \end{aligned}$$

Тогда имеем

$$\frac{\partial \rho}{\partial y_i} = \sum_{b \in B} \frac{1}{2h_b} \frac{\partial h_b}{\partial y_i} \rho_b, \quad \frac{\partial \rho}{\partial l_i} = \sum_{b \in B} \frac{1}{2h_b} \frac{\partial h_b}{\partial l_i} \rho_b, \quad i = 1, 2.$$

Для функции схожести ρ (в предположении ее дифференцируемости) имеем

$$\rho \approx \rho(y_0, l_0) + \text{grad} \rho|_{(y_0, l_0)} \cdot (\Delta y_1, \Delta y_2, \Delta l_1, \Delta l_2) = \rho(y_0, l_0) + t |\text{grad} \rho|_{(y_0, l_0)}|^2,$$

где $(\Delta y_1, \Delta y_2, \Delta l_1, \Delta l_2) = t \text{grad} \rho|_{(y_0, l_0)}$. То есть вектор приращений выбираем в направлении градиента, так как в каждой точке вектор градиента по численному значению и по направлению характеризует наибольшую скорость возрастания величины, в данном случае схожести. Вообще говоря, знание градиента в точке дает лишь направление шага, а для определения величины шага этого недостаточно. Однако знание максимального значения функции схожести ρ дает приблизительную оценку для $t \approx \frac{1 - \rho(y_0, l_0)}{|\text{grad} \rho|_{(y_0, l_0)}|^2}$. Откуда для величины шага имеем $(\Delta y_1, \Delta y_2, \Delta l_1, \Delta l_2) \approx \frac{1 - \rho(y_0, l_0)}{|\text{grad} \rho|_{(y_0, l_0)}|^2} \text{grad} \rho|_{(y_0, l_0)}$.

Численные эксперименты

Для эксперимента были использованы ролики из [Вражнов, Шаповалов, Николаев, 2012] со свободно прохаживающимися людьми и ролик с автомобильным движением на перекрестке. Целью эксперимента является обводить красной (или другой) рамкой объект, который в процессе своего движения создавал бы следующие трудности для наблюдения: неожиданное изменение траектории и скорости движения, наклоны и повороты, уменьшение-увеличение одного из размеров или непропорциональное изменение обоих, частичное или полное загораживание. Результаты трекинга с ноября 2013 находятся в свободном доступе в интернете [Шапошников, 2013а; Шапошников, 2013б; Шапошников, 2013с; Шапошников, 2014а; Шапошников, 2014б]. В рамках этой статьи отметим, что удалось вести человека в розовой рубашке до самого конца, несмотря на многочисленные загораживания. Такая техника слежения является перспективной при исключении ошибок соскальзывания окна на сторонние объекты: «маршруты полетов малайзийского

Рис. 1. Первый кадр — тот, с которого был записан признак. Далее кадры 291, 295, 308

Boeing и Су-25 ... совпали, слившись ... система сопровождения автоматически перестраивается на ведение той, которая имеет наибольшую площадь» [РИА Новости].

На рисунке 1 представлен кадр, на котором с объекта был снят признак, — описанная выше взвешенная цветовая гистограмма. Далее — одна из коллизий, заслонение на 290–308 кадрах, начиная с того, на котором снимали признак. Число возле рамки — коэффициент схожести.

Рис. 2. Кадры 688, 762, 782, 865, 1300, 1330, 1360, 1376, 1388, 1418, 1440 и последний 1492-й

На рисунке 2 представлена серия коллизий в конце всего видео, завершающим — последний кадр. Всего 1492 кадра, начиная с того, на котором снимали признак. Объект все время находится в рамке.

В процессе обсуждения многие коллеги связывали успешный трекинг человека в розовой рубашке с видимой уникальностью рубашки. Однако сам принцип построения признака делает таким же уникальным практически любой объект этого видео. Этот факт отражен в нарезке на рисунке 3. Вектор признаков человека в голубой рубашке снят на первом кадре нарезки. Далее видно, что овал рамки не съезжает ни на голубую стену, ни на других людей в голубом даже при коллизиях. Объект все время находится в рамке.

Второй пример из [Вражнов, Шаповалов, Николаев, 2012] (из двух) с удаляющимся человеком с частичным заслонением и изменяющимися независимо размерами. Трекер ведет объект до конца [Шапошников, 2014a]. Рисунок 4.

В одном из приложений съемка велась камерой, подвешенной над перекрестком. Проезжающие перекресток автомобили обрабатывались с помощью MeanShift [Comaniciu, Ramesh, Meer, 2003] и обводились наклонными эллипсами. Для поворачивающихся автомобилей эллипсы должны были менять и ориентацию и наклон, из-за чего MeanShift терял объекты. Нарезка из этого видео [Шапошников, 2013b] с проблемным объектом представлена на рисунке 5. Теперь трекер ведет объект до конца.

Следующий рисунок 6 демонстрирует пример применения алгоритма [Шапошникова, 2013] для отыскания объектов по образцу. Снимок взят из базы [PETS, 2007]. С объекта, помеченного как «Sample» на рисунке слева, взят образец взвешенной гистограммы (вектор признаков). Затем весь кадр разбит на участки, и на каждом участке запущена программа поиска максимума схожести. На рисунке слева показаны результаты, удовлетворяющие порогу по схожести и допустимым размерам. Всего их получилось 5. Самый нижний сотрудник в униформе попал в два конечных результата.

Рис. 3. Первый кадр — тот, с которого был записан признак. Далее кадры 225, 632, 659, 716, 724, 920, 1043

Рис. 4. Первый кадр — тот, с которого был записан признак. Далее кадры 190, 162, 163, 210, 384, 393, 394

Список литературы

Вражнов Д. А., Шаповалов А. В., Николаев В. В. О качестве работы алгоритмов слежения за объектами на видео // Компьютерные исследования и моделирование. — 2012. — Т. 4, № 2. — С. 303–313.

Рис. 5. Первый кадр — тот, с которого был записан признак. Далее кадры 139, 229, 309

Рис. 6. Слева — образец, справа — результаты. Нижний красный прямоугольник — слияние двух прямоугольников.

Дерюгин Е. Е., Суворов Б. И. Стадийность «истинных» диаграмм нагружения // Вопросы атомной науки и техники, — 2011. — № 4. — С. 94–99.

Калайда В. Т., Елизаров А. И. Метод опорной гиперповерхности для идентификации изображения лица человека // Вычислительные технологии. — 2012. — Т. 17, № 5. — С. 65–70.

РИА Новости URL: <http://www.rg.ru/2014/07/25/boeing-site-anons.html> (дата обращения 04.08.2014).

- Шапошников А. И., Шапошникова Е. В.* Алгоритм «Локализация объекта в кадре» // Хроники объединенного фонда электронных ресурсов «НАУКА И ОБРАЗОВАНИЕ». — 2012. № 10(41). Октябрь 2012. URL: <http://ofernio.ru/portal/newspaper.php> (дата обращения 3.02.2014).
- Шапошникова Е. В.* Программа слежения за подвижной областью в видеопоследовательности. Свидетельство № 2013660663 в реестре программ для ЭВМ Федеральной службы по интеллектуальной собственности. Заявитель и правообладатель НИ Томский государственный университет. — Заявл. № 2013617286 от 13.08.2013. Зарег. 14.11.2014.
- Шапошников А. И.* Kernel Based Object Tracking. The MeanShift's modification. 2013. URL: <http://www.youtube.com/watch?v=aCUeJI0C-5k> (дата обращения 3.02.2014).
- Шапошников А. И.* Kernel based object tracking of the rotating and receding car. 2013. URL: <http://www.youtube.com/watch?v=oQKtpG524MU> (дата обращения 3.02.2014).
- Шапошников А. И.* Kernel Based Object Tracking: the receding man. 2013. URL: http://www.youtube.com/watch?v=pUr1PBA2_8k (дата обращения 3.02.2014).
- Шапошников А. И.* Kernel based object tracking with rotating (on 5 parameters). 2014. URL: <http://www.youtube.com/watch?v=bK4woix0ml8> (дата обращения 3.02.2014).
- Шапошников А. И.* Find objects in the image which are similar to the sample. 2014. URL: http://my.mail.ru/mail/albertelena/photo?album_id=Mywork (дата обращения 3.02.2014).
- Comaniciu D., Meer P.* Mean Shift: A robust approach toward feature space analysis // Pattern Analysis and Machine Intelligence, IEEE Transactions. — May, 2002. — V. 24. — P. 603–519.
- Comaniciu D., Ramesh V., Meer P.* Kernel-based object tracking // Pattern Analysis and Machine Intelligence, IEEE Transactions. — May, 2003. — V. 25, № 5. — P. 564–577.
- Maggio E., Cavallaro A.* Video Tracking: Theory and Practice. — Wiley, 2010. — 266 p.
- The 11th IEEE International Conference on Computer Vision 2007. PETS 2007. Benchmark Data. <http://pets2007.net/> (дата обращения 3.02.2014).