

УДК 519.6

МЕТОДЫ РЕШЕНИЯ ЗАДАЧИ СЛЕЖЕНИЯ ЗА ДИНАМИЧЕСКИМИ ОБЪЕКТАМИ

М. П. Волченков
(МГУ им. М. В. Ломоносова)

Задачи автоматического анализа изображений и сопровождения динамических объектов в видеопотоке используются в различных приложениях. Большая размерность этих задач не позволяет работать со всем изображением как множеством точек. Для понижения размерности предлагается перейти от всего изображения к набору его характерных точек. Приводятся способы выделения характерных точек, а также использование выделенных точек для слежения за динамическими объектами посредством решения задачи интерполяции и задачи о назначениях. Решается задача восстановления объекта по координатам его характерных точек, измерение которых происходит с ошибками.

Введение

Задачи автоматического анализа изображений и сопровождения динамических объектов в видеопотоке типичны для многих практических областей. В качестве динамических объектов могут рассматриваться лицо человека, облака, летательные аппараты, автомобили на шоссе и т. п. Частные случаи решения задач используются для распознавания лиц, отслеживания направления взгляда, определения выражения лица, восстановления траекторий летательных аппаратов на фоне движущихся облаков, автоматического анализа рентгеновских снимков и др.

К сожалению, огромная размерность рассматриваемых задач, обусловленная большим количеством вычислений на большом объеме данных в каждый момент времени, не позволяет полноценно работать со всем изображением в реальном масштабе времени. Поэтому на первом этапе решения задачи предлагается переходить от всего изображения к некоторому набору его точек, которые будут выбраны по определенному правилу. Такие точки назовем *характерными*. Второй этап решения представляет собой поиск траекторий каждой характерной точки и всего изображения в целом с течением времени.

Изучение некоторых программных продуктов, предназначенных для решения подобного рода задач (системы Face Lab [1], Microsoft Research [2], Stereo Tracker [3]), выявило их об-

щие недостатки: 1) необходимым условием всех используемых в программах алгоритмов слежения за характерными точками является начальная инициализация характерных точек на первом кадре, выполняемая человеком; 2) характерные точки могут быть потеряны.

В данной работе предложены эффективные методы для автоматического выделения характерных точек и восстановления их траекторий, что позволяет восстановить движение крупных объектов на изображении даже при небольших неточностях в измерении координат.

Выделение характерных точек

Дадим содержательное определение характерной точки. Рассмотрим изображение, заданное функцией от двух переменных. Пусть известно, что $f(x, y) = F(x, y) + r$, где $F(x, y)$ — достаточно гладкая функция, а r — случайная величина с нулевым средним. Пусть $F(x, y)$ имеет *ярко выраженные* локальные экстремумы. Как найти их, если шум r дает много дополнительных экстремумов функции $f(x, y)$?

Для того чтобы избавиться от помех, сглаживаем функцию $f(x, y)$, приближая ее в окрестности каждой точки поверхностями второго порядка. Поскольку шум имеет нулевое среднее, такая процедура позволяет избавиться от многих случайных экстремумов. Те экстремумы, которые

в результате указанных преобразований "не пострадали", относим к характерным точкам. Ниже будет дано определение характерных точек для конкретных задач.

Каждому кадру двумерного изображения сопоставлена матрица A размером $m \times n \times p$; ячейка $A(i, j, h)$, $i = \overline{1, m}$, $j = \overline{1, n}$, $h = \overline{1, p}$, хранит информацию о точке изображения с координатами (i, j) , имеющей цвет h . Здесь p имеет смысл числа базовых цветов. Случай $p > 1$ принципиально не отличается от случая $p = 1$, поэтому в дальнейшем будем считать, что $p = 1$ и матрица A — двумерная.

Пусть дано изображение в виде матрицы A размером $m \times n$. Выделим окно B размером $k \times k$, $k \leq \min(n, m)$, $B(i, j) = A(l + i - 1, s + j - 1)$, где $l = \overline{1, m - k + 1}$, $s = \overline{1, n - k + 1}$. Сопоставим матрице B точку с координатами центра окна.

В окне B производится аппроксимация полиномом от двух переменных. Элемент $B(i, j)$ заменяется элементом $\tilde{B}(i, j) = ai^2 + bj^2 + cij + di + ej + f$, и необходимо минимизировать функционал

$$F(a, b, c, d, e, f) = \sum_{i,j=1}^k [B(i, j) - (ai^2 + bj^2 + cij + di + ej + f)]^2,$$

что сводится к решению системы линейных уравнений

$$\frac{dF}{da} = 0; \quad \frac{dF}{db} = 0; \quad \dots; \quad \frac{dF}{df} = 0.$$

Представляет интерес квадратичная часть такого приближения — матрица

$$C = \begin{pmatrix} a & \frac{c}{2} \\ \frac{c}{2} & b \end{pmatrix}.$$

Для матриц такого вида автором была решена следующая задача.

Задача 1. Пусть известно, что два окна размерами $k \times k$ на изображении A приближаются поверхностями $m(i, j) = ai^2 + bj^2 + cij + di + ej + f$ и $\tilde{m}(i, j) = \tilde{a}i^2 + \tilde{b}j^2 + \tilde{c}ij + \tilde{d}i + \tilde{e}j + \tilde{f}$ соответственно, причем $\tilde{a} = a + \varepsilon_1$, $\tilde{b} = b + \varepsilon_2$, $\tilde{c} = c + \varepsilon_3$, $-\varepsilon \leq \varepsilon_1, \varepsilon_2, \varepsilon_3 \leq \varepsilon$ ($\varepsilon \geq 0$), $|a - b| \leq \delta_1$, $|c| \leq \delta_2$. C и \tilde{C} — квадратичные матрицы этих приближений соответственно:

$$C = \begin{pmatrix} a & \frac{c}{2} \\ \frac{c}{2} & b \end{pmatrix}; \quad \tilde{C} = \begin{pmatrix} \tilde{a} & \frac{\tilde{c}}{2} \\ \frac{\tilde{c}}{2} & \tilde{b} \end{pmatrix}.$$

Требуется оценить степень различия собственных значений матриц C и \tilde{C} .

В результате получено

Утверждение 1. Степень различия собственных значений λ_1, λ_2 матрицы C и $\tilde{\lambda}_1, \tilde{\lambda}_2$ матрицы \tilde{C} с условиями из задачи 1 оценивается неравенствами

$$\lambda_1 - \frac{3}{2}\varepsilon - \sqrt{\delta_1\varepsilon} - \sqrt{\delta_2\varepsilon} \leq \tilde{\lambda}_1 \leq \lambda_1 + \frac{3}{2}\varepsilon + \sqrt{\delta_1\varepsilon} + \sqrt{\delta_2\varepsilon};$$

$$\lambda_2 - \frac{3}{2}\varepsilon - \sqrt{\delta_1\varepsilon} - \sqrt{\delta_2\varepsilon} \leq \tilde{\lambda}_2 \leq \lambda_2 + \frac{3}{2}\varepsilon + \sqrt{\delta_1\varepsilon} + \sqrt{\delta_2\varepsilon}.$$

Методом статистических испытаний было получено, что для матриц, элементы которых отличаются не больше чем на ε , вероятность отличия собственных чисел более чем на 5ε не превышает 0,02.

Дальнейшая работа производится с собственными значениями λ_1, λ_2 матрицы C и коэффициентами аппроксимирующего полинома $m(i, j) = ai^2 + bj^2 + cij + di + ej + f$.

В первую очередь производится проверка, попадает ли критическая точка (i_0, j_0) , которая является решением системы

$$2ai + cj + d = 0;$$

$$2bj + ci + e = 0,$$

в окно размером $k \times k$, т. е. выполняется ли неравенство $1 \leq i_0, j_0 \leq k$. В случае выполнения неравенства для такой критической точки необходимо исследовать, является ли она экстремумом, проверив, остается ли неизменным знак разности $m(i, j) - m(i_0, j_0)$ в окне размером $k \times k$ [4]. В случае сохранения знака этой разности при $i = \overline{1, k}$, $j = \overline{1, k}$ называем точку $r(t)$, сопоставленную матрице B^t , *начально-характерной*.

Для начально-характерных точек рассматриваем случаи, когда произведения соответствующих им собственных чисел положительны. Пусть M — число таких случаев, S — сумма всех собственных значений для этих случаев. Пусть $\alpha = S/(2M)$, т. е. является средней величиной собственных значений для начально-характерных точек с $\lambda_1\lambda_2 > 0$. Положим $\beta = q\alpha$, где $0 < q \leq 1$. Начально-характерную точку $r(t)$ с $\lambda_1^t\lambda_2^t > 0$ назовем *предварительно-характерной*, если $\min(\lambda_1^t, \lambda_2^t) > \beta$.

Пусть \widehat{M} — число предварительно-характерных точек. Выбрав параметр δ , где $0 < \delta \leq 1$,

оставляем из них $[\delta\widehat{M}] + 1$ точек с наибольшими произведениями собственных значений. Такие точки будем называть *характерными*. Характерная точка с наименьшим произведением собственных чисел на первом кадре изображения будет определять порог χ для работы с последующими изображениями видеопотока.

Таким образом, чтобы точка на изображении являлась характерной, для нее необходимы следующие условия: $\lambda_1 > \beta$; $\lambda_2 > \beta$; $\lambda_1\lambda_2 \geq \chi$, а также для экстремума (i_0, j_0) — выполнение неравенства $1 \leq i_0, j_0 \leq k$.

Далее можно перейти от матрицы изображения A к матрице ее характерных точек.

На рис. 1 представлена иллюстрация описанного выше метода, выделяющего на изображении предварительно-характерные точки (серого цвета) и характерные точки (белого цвета).

На рис. 2 изображены два последовательных кадра видеопотока, на рис. 3 — результат совмещения их характерных точек (точки первого кадра — черного цвета, второго — белого).

Устойчивость описанного метода определения характерных точек позволяет работать с последовательностью кадров видеопотока. Небольшие изменения изображения успешно учитываются рассмотренным алгоритмом, что иллюстрирует рис. 3.

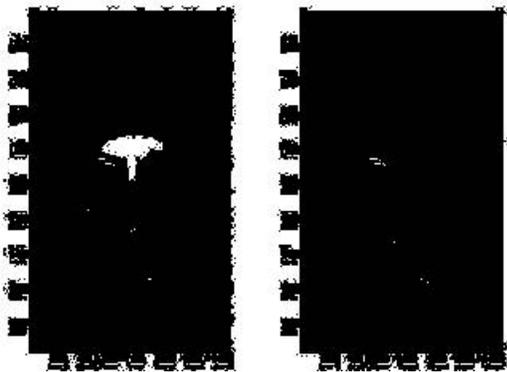


Рис. 1. Выделение характерных точек на лице человека



Рис. 2. Два соседних изображения в видеопотоке

Выделение устойчивых от кадра к кадру характерных точек можно также осуществить, используя теорию *гауссовых смесей* [5].

Под гауссовой смесью понимается случайная величина с плотностью вида суммы нормальных плотностей:

$$\omega(\mathbf{x}) = \sum_{i=1}^k p_i \varphi_i(\mathbf{x}, \boldsymbol{\mu}^i, \boldsymbol{\sigma}^i, Cov_i),$$

где $p_i \geq 0$ — веса, $\sum_{i=1}^k p_i = 1$; \mathbf{x} — вектор координат; $\boldsymbol{\mu}^i$ — векторы математических ожиданий; $\boldsymbol{\sigma}^i$ — векторы средних квадратических отклонений; Cov_i — ковариационные матрицы; φ_i — плотности нормального распределения.

Логика рассматриваемых задач предполагает двумерный случай.

Выделение характерных точек при помощи гауссовых смесей производится в трех режимах:

- 1) автоматическом (характерные точки на первом и последующих кадрах определяются без помощи оператора);
- 2) ручном (характерные точки на первом кадре видеопотока отмечает оператор);
- 3) комбинированном (характерные точки на двух первых кадрах определяются автоматически, а затем оператор вручную выделяет только те точки, которые присутствуют на обоих кадрах).

Независимо от выбранного режима после одного, двух или нескольких *обучающих* кадров накапливается массив характерных точек с соответствующими им собственными значениями λ_1



Рис. 3. Совмещение характерных точек соседних кадров

и λ_2 , посчитанными согласно описанному выше методу.

Пары собственных значений для характерных точек из обучающего массива представляют собой пары координат на плоскости и определяют набор точек с этими координатами. Создана программа, с помощью которой на основе расположения получившихся точек строится гауссова смесь с заданным числом k ; для каждого ее компонента вычисляются $p_i, \mu_1^i, \mu_2^i, \sigma_1^i, \sigma_2^i, \rho_i$, где $i = \overline{1, k}$.

Дальнейшая работа базируется на выделении характерных точек на поступающих в каждый момент времени новых изображениях.

На основе обучающего массива пар собственных значений и построенной гауссовой смеси ω для каждой точки (x_0, y_0) получаемого входного изображения производится измерение логарифма величины $\omega(\lambda_1^{(x_0, y_0)}, \lambda_2^{(x_0, y_0)})$ и сравнение его с заданным порогом, где $\lambda_1^{(x_0, y_0)}$ и $\lambda_2^{(x_0, y_0)}$ — собственные числа, соответствующие (x_0, y_0) . В случае превышения порога для рассматриваемого логарифма $\log(\omega(\lambda_1^{(x_0, y_0)}, \lambda_2^{(x_0, y_0)}))$ точка (x_0, y_0) считается характерной.

На рис. 4 приводится пример применения метода гауссовых смесей в автоматическом режиме, где в качестве потока изображений взяты кадры с изменяющимся со временем положением человеческого лица. Прослеживается следующая последовательность действий: на первом снимке дано входное изображение; на втором — изображение, полученное из первого с помощью сглаживания полиномом; на третьем снимке показаны наложенные на принятое изображение характерные точки; на четвертом — характерные точки (черного цвета) на белом фоне без привязки к изображению.

Схожая картина наблюдается в двух других режимах, однако лучшие результаты получены



Рис. 4. Автоматический режим выделения характерных точек

в комбинированном режиме. Получение гауссовой смеси на основе массива из множества точек позволяет выделять характерные точки на последующих кадрах, причем результаты выделения, как и при реализации первого алгоритма, отличаются хорошей устойчивостью (рис. 5).



Рис. 5. Характерные точки соседних кадров в комбинированном режиме

В качестве модификации описанных методов можно определить режим, когда известно, что выделенные характерные точки — разных типов (например, на лице можно выделить точки бровей, уголков глаз, губ, носа и т. п.). За каждый из типов характерных точек будет "отвечать" один из компонентов гауссовой смеси ω , и подсчет величины $\log(\varphi_i(\lambda_1^{(x_0, y_0)}, \lambda_2^{(x_0, y_0)}, \mu^i, \sigma^i, Cov_i))$ для точки изображения (x_0, y_0) будет производиться для всех i . Наибольшая из $\log(\varphi_i), \log(\varphi_s) = \max_i(\log(\varphi_i))$, сравнивается с заданным порогом, после чего делается вывод, принадлежит ли рассматриваемая точка к s -му классу или не является характерной.

Подробнее о применении гауссовых смесей можно узнать из работы [6].

Слежение за характерными точками

Несущественные изменения структуры характерных точек от кадра к кадру позволяют решать задачу слежения за характерными точками по их взаимному расположению с течением времени. Рассмотрим два подхода к решению данной задачи. В первом случае работа производится с последовательностью изображений, во втором — с двумя соседними кадрами.

Для нахождения соответствия между характерными точками последовательных кадров необходимо решить обобщенную задачу интерполяции: по множеству точек (x_i, y_i) , $i = 1, 2, \dots, nN$, найти семейство функций f_1, \dots, f_n таких, чтобы для каждого i нашлось l такое, что $f_l(x_i) = y_i$. По сути дела требуется с некоторой погрешностью "уложить" множество точек на кривые из заданного класса так, чтобы на каждой кривой лежали точки из разных кадров. Уточним постановку задачи [7].

Задача 2 (о восстановлении траекторий).

Пусть дано множество точек

$$Z_i = \left\{ \left(x_i^{(l)}, y_i^{(l)} \right), l = \overline{1, n}, i = \overline{1, N}; x_i^{(l)}, y_i^{(l)} \in \mathbb{R} \right\}.$$

Известно, что существует семейство функций f_l , $l = \overline{1, n}$, таких, что при всех i, l $\left| y_i^{(l)} - f_l(x_i^{(l)}) \right| < \delta$. Требуется эти функции найти.

Для решения задачи в случае семейства линейных и квадратичных функций был разработан рекурсивный алгоритм \mathcal{A} накопления коэффициентов при линейном и квадратичном членах.

В линейном случае, при аппроксимации зависимости прямыми вида $y = kx + b$, алгоритм состоит в следующем. Если ввести обозначения

$$\begin{aligned} S_N &= \sum_{i=1}^N \left(x_i^{(l(i))} \right)^2; & U_N &= \sum_{i=1}^N x_i^{(l(i))}; \\ V_N &= N; & R_N &= \sum_{i=1}^N y_i^{(l(i))} x_i^{(l(i))}; \\ T_N &= \sum_{i=1}^N y_i^{(l(i))}, \end{aligned}$$

то коэффициенты семейства прямых для N последовательных кадров будут определяться следующими соотношениями:

$$k^N = \frac{T_N U_N - R_N V_N}{U_N^2 - S_N V_N}; \quad b^N = \frac{R_N U_N - T_N S_N}{U_N^2 - S_N V_N}.$$

При динамическом появлении новых изображений (т. е. увеличении N , $\tilde{N} = N + 1$) пересчет коэффициентов будет происходить лишь путем добавления одного нового члена в $S_{\tilde{N}}$, $U_{\tilde{N}}$, $V_{\tilde{N}}$, $R_{\tilde{N}}$ и $T_{\tilde{N}}$. Имеет место

Теорема 1. Если $N \geq n$, то алгоритм \mathcal{A} решает задачу 2 со сложностью $n^2 N$.

Алгоритм допускает модификацию, когда учитывается не только расположение характерных точек, но и их интенсивность.

Из теоремы 1 следует, что картину движения можно отследить с определенной точностью при рассмотрении числа кадров, сопоставимого с количеством точек. Поэтому для того чтобы не рассматривать слишком большого количества кадров, целесообразно разбивать каждый снимок на области с небольшим количеством характерных точек на них.

Еще один метод представляет собой некоторое изменение задачи о назначениях. Требуется решить задачу о назначении характерных точек из последовательных кадров, где точек в разных кадрах, вообще говоря, различное количество. Математически задачу можно сформулировать следующим образом:

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^n x_{ij} d_{ij} &\rightarrow \min; \\ \sum_{i=1}^m x_{ij} &\geq 1 \quad \forall j \in \overline{1, n}; \\ \sum_{j=1}^n x_{ij} &\geq 1 \quad \forall i \in \overline{1, m}; \\ x_{ij} &\in \{0, 1\}. \end{aligned}$$

Здесь d_{ij} — расстояние между i -й точкой одного кадра и j -й точкой другого кадра; $x_{ij} = 1$, если i -я точка первого кадра назначается j -й точке второго кадра, $x_{ij} = 0$ — в противном случае (матрица назначений). Под *минимальным паросочетанием* понимается случай, когда достигается минимальная сумма расстояний между назначенными парами точек.

Анализируя задачу, получаем

Утверждение 2. Если паросочетание Σ минимально, то отрезки, соединяющие соответствующие точки разных кадров, не пересекаются.

Можно использовать метод нахождения минимального паросочетания [8, 9] для решения задачи соответствия характерных точек последовательных кадров.

Задача 3. Даны точки $(x_1^l, y_1^l), (x_2^l, y_2^l)$, $l = 1, \dots, n$, первого и второго кадров соответственно и известно, что существует подстановка $T: \{1, \dots, l\} \rightarrow \{1, \dots, l\}$, такая что

$$K_1 < \frac{y_2^{T(l)} - y_1^{(l)}}{x_2^{T(l)} - x_1^{(l)}} < K_2,$$

$$c_1 < \sqrt{\left(x_1^{(l)} - x_2^{T(l)}\right)^2 + \left(y_1^{(l)} - y_2^{T(l)}\right)^2} < c_2.$$

Восстановить подстановку T .

Теорема 2. *Решение задачи минимального паросочетания дает решение задачи 3. При этом*

$$nc_1\sqrt{1 + K_1^2} \leq \Sigma \leq nc_2\sqrt{1 + K_2^2}.$$

На основе предложенных алгоритмов была написана программа, которая позволяет находить минимальные паросочетания и проверять выполнение теоремы 2. Численные исследования показывают, что существенные всплески значений величины Σ возникают при появлении точек, движение которых не описывается ранее найденным семейством параллельных прямых.

Заключение

Определив движение характерных точек от кадра к кадру, с некоторой точностью можно установить перемещение любой точки изображения. Взяв характерные точки в качестве узлов, нужно решить задачу построения оптимальной триангуляции [9].

После этого для любой точки изображения существует близкое к ней множество характерных точек с определенными векторами перемещения. Зная эти векторы и руководствуясь некоторыми выбранными принципами, для каждой точки изображения можно определить вектор ее смещения от одного кадра к другому. В результате строится векторное поле в каждой точке изображения, которое определяет движения точек с течением времени в парах соседних кадров.

Описанные принципы работы с характерными точками на основе вычисления собственных значений и рассмотрения гауссовых смесей успешно зарекомендовали себя в практических задачах.

Понижение размерности задач в результате перехода от всех точек изображения к набору характерных точек, определяющих поведение всего изображения, позволяет решать задачи подобного рода в многочисленных приложениях.

Список литературы

1. *Matsumoto Y., Zelinsky A.* An Algorithm for Real-time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement. Australian National University, Nara Institute of Science and Technology, 2000.
2. *Yang R., Zhang Z.* Model-Based Head Pose Tracking with Stereovision. <http://research.microsoft.com/~zhang/Papers/TR01-102.pdf>
3. *Gorodnichy D. O., Malik S., Roth G.* Affordable 3D face tracking using projective vision // Vision Interface. Calgary, Alberta, Canada. May 2002. P. 383–390.
4. *Выгодский М. Я.* Справочник по высшей математике // Гос. изд-во тех.-теор. лит. М.: 1957. С. 657.
5. *Феллер В.* Введение в теорию вероятностей и ее приложения. Т. 2. М.: Мир, 1984. С. 69–90.
6. *Волченков М. П.* Решение задачи слежения методом выделения характерных точек // Молодежь в науке. Сб. докл. IV науч.-тех. конф. Саров: РФЯЦ-ВНИИЭФ, 2006. С. 23–30.
7. *Волченков М. П., Самоненко И. Ю.* Об автоматическом распознавании лиц // Интеллектуальные системы. 2005. Т. 9. Вып. 1–4. С. 135–156.
8. *Сердюков А.* К задаче о покрытии // Управляемые системы. Институт математики им. С. Л. Соболева СО АН СССР, 1975. Вып. 14.
9. *Черчмен У., Акоф Р., Арноф Л.* Введение в исследование операций. М.: Наука, 1968.
10. *Кормен Т., Лейзерсон Ч., Ривест Р.* Алгоритмы. Построение и анализ. М.: Московский центр непрерывного математического образования, 2004.