

# Алгоритм кластеризации последовательности изображений с формированием кластеров на основе опорных триплетов

В.А. Микин

В данной работе предложен алгоритм кластеризации последовательностей изображений, идея которого заключается в формировании кластеров на основе минимальных трёхточечных симплексов, образованных классифицируемыми данными в многомерном пространстве признаков. Рассмотрены варианты кластеризации в отложенном и псевдореальном масштабе времени.

**Ключевые слова:** кластеризация изображений, определение смены сюжета, анализ видео, опорные кадры, кластеризация в реальном времени.

## Введение

Одной из известных проблем в обработке последовательностей растровых изображений (далее - видеопоследовательностей), является проблема выделения опорных (ключевых) кадров [1, 2, 3]. Аналогичная задача возникает в рамках автоматизированного анализа различных графических материалов презентационного характера, циркулирующих в сети Интернет, причём в этом случае, вследствие значительных объёмов анализируемых данных, особую актуальность приобретает сокращение избыточности, свойственной как отдельным изображениям, так и их последовательностям [4].

Как правило, выбор опорных кадров сводится к определению моментов смены сюжета (существенного изменения структурного содержания двух соседних изображений) в видеопоследовательности [1]. В то же время, известны подходы к обнаружению опорных кадров на основе алгоритмов кластеризации [5, 6]. Если рассматривать произвольную видеопоследовательность как выборку однотипных наблюдений-прецедентов, представляемых составляющими её кадрами, то выделение опорных кадров является задачей кластеризации, в контексте которой количество наблюдений существенно превышает количество классов, в то время как явная информация о количестве и структуре классов отсутствует, и единственным априорным условием является некоторое допустимое расстояние между прецедентами в пространстве признаков.

Пусть видеопоследовательность  $G$  – совокупность исследуемых прецедентов (изображений),  $P$  – множество соответствующих им образов – векторов признаков размерности  $\nu$ , получаемых некоторым преобразованием  $T : G \rightarrow P$ ;  $p = T(g)$ ,  $p = p(x, y)$ ;  $x = 1, \dots, w$ ;  $y = 1, \dots, h$ ;  $\nu = wh$ . В векторном виде двумерные образы  $p$  представляются в соответствии со следующим правилом развёртки:  $p(i) = p(x, y)$ ;  $i = 1, \dots, \nu$ ;  $i = yw + x$ .

В рассматриваемом случае целью формирования множества  $P$  является снижение размерности данных, по сравнению с исходными изображениями, то есть  $T(g)$  – масштабирующее отображение:

$$p(x, y) = \frac{1}{4R^2} \sum_{k=1}^{2R} \sum_{l=1}^{2R} g(2R(x-1) + k, 2R(y-1) + l), \quad (1)$$

где  $R$  выбирается численно равным средней оценке радиуса корреляции (расстояния, на котором значения яркости пикселей изображения могут считаться статистически независимыми [7]) для исследуемого класса изображений. Заметим также, что при построении образов яркость пикселей определяется линейной комбинацией цветовых составляющих исходного изображения [1], то есть графическая структура представляется в градациях серого цвета.

Ограничим диапазон значений, принимаемых элементами векторов  $p$ , интервалом  $[0, 1]$ . Таким образом, любой вектор  $p_k$  представляется точкой внутри единичного  $\nu$ -куба  $P \in H^\nu \subset \mathbb{R}^\nu$ . Максимальное евклидово расстояние между двумя точками, принадлежащими такому гиперкубу, будет равным длине его диагонали

–  $d_{max}(p_j, p_k) = \sqrt{\nu}$ . Это позволяет задать пороговую величину  $\mu_{th} = \alpha d_{max}$ ,  $\alpha \in [0, 1]$ , характеризующую предельно допустимое расстояние между образами  $p_j$  и  $p_k$ , при котором они считаются совпадающими.

Будем полагать, что вероятностное распределение образов близко к нормальному, и форма образуемых кластеров хорошо аппроксимируется сферой (подобное допущение зачастую оказывается приемлемым на практике, в частности, если отличия промежуточных кадров видеопоследовательности от образующих её сюжетных составляющих обусловлены не столько структурными изменениями, сколько присутствием шума на изображениях, как, например, в системах видеонаблюдения, используемых вне помещений). Таким образом, в рамках рассматриваемой задачи определим кластер как множество точек  $\{p_k\}$ , удалённых от некоторого центра  $p_0$  на расстояние не превышающее  $r = \frac{1}{2}\mu_{th}$ :

$$K_l = \{p_0, (p_1, \dots, p_{m_l})\} : \forall k \in [1, m_l] \ d(p_0, p_k) \leq r. \quad (2)$$

Заметим, что в рамках определения (2) любой набор точек  $(p_1, \dots, p_m) : \forall i, j \in [1, m] \ d(p_i, p_j) \leq \mu_{th}$  формально является кластером с не указанным явно центром  $p_0$ . Данный факт отражает важность и влияние корректности изначального выбора потенциальных центров кластеров на результат кластеризации [8].

В принятой вероятностной модели, если точки некоторого кластера  $K_l$  распределены нормально, то вероятность выбрать  $M$  точек, удалённых друг от друга не более чем на  $\mu_{th}$  максимизируется в  $r$ -окрестности математического ожидания (то есть центра кластера  $p_0$ ). Тем самым, если в  $H^\nu$  найдётся минимальный правильный симплекс

$$S_l = (p_1, \dots, p_{M_l}), \ M_l \leq \nu + 1 : \\ \forall i, j \in [1, M_l] \ i \neq j \ d(p_i, p_j) \xrightarrow{H^\nu} \min, \quad (3)$$

то  $r$ -окрестность его центра может быть приписана одному кластеру. Таким образом, возможно построение процедуры кластеризации в виде гибридной схемы, включающей в себя выбор начального приближения кластера  $K_l$  в виде опорного симплекса  $S_l$  и его дополнение точками из  $r$ -окрестности центра  $S_l$ .

При этом, чтобы не исключать возможности существования кластеров, радиусом меньших  $r$ , целесообразно рассматривать описанную вокруг  $S_l$  гиперсферу как границу кластера  $K_l$ :  $r_l = R(S_l) \leq r$ .

Дальнейшее объединение кластеров в более крупные структуры возможно по мере итераций процесса кластеризации.

Выбор опорного симплекса  $S_l$  для каждого кластера – оптимизационная задача, характерная для иерархических способов кластеризации. Снизить её комбинаторную сложность можно, снизив соответственно размерность  $S_l$ . Наименьшая размерность симплекса (выбираемого по критерию (3) минимизации расстояния между его вершинами), при которой в описанную вокруг него гиперсферу могут попасть другие точки, кроме вершин симплекса, равна двум.

Очевидно, сколь-нибудь плотное заполнение  $H^{\nu}$  точками исключено в большинстве практических ситуаций, и вероятность найти среди этих точек правильные симплексы даже малой размерности близка к нулю. Но, в действительности, основное условие при выборе опорного симплекса  $S_l$  заключается лишь в том, что точки, попавшие в сформированный на его основе кластер  $K_l$ , должны быть удалены друг от друга на расстояние, не превышающее  $2r$ , то есть  $r_l \leq r$ . Таким образом, если  $S_l$  представляет собой тройку (триплет) точек  $S_l = (p_1, p_2, p_3)$ , то треугольник  $p_1p_2p_3$  должен быть остроугольным:

$$\forall i, j, k \in [1, 3], i \neq j \neq k : d^2(p_i, p_j) \leq d^2(p_j, p_k) + d^2(p_k, p_i). \quad (4)$$

С учётом определения (2), критерия (3) и ограничения (4) можно привести следующее словесное описание алгоритма кластеризации:

- 1) Поиск в  $P$  триплета  $S_l$ , образующего минимальный треугольник  $p_1p_2p_3$ , длина стороны которого не превосходит  $\mu_{th}$ , и для которого выполняется условие (4). Если триплет  $S_l$  не найден, то кластеризация завершена.
- 2) Формирование кластера  $K_l$  на основе опорного триплета  $S_l$  и всех точек из  $P$ , попадающих в  $r_l$ -окрестность центра  $S_l$ .
- 3) Исключение всех точек кластера  $K_l$  из  $P$ , включение в  $P$  точки, соответствующей центру  $K_l$ .
- 4) Переход к п. 1.

Применение предложенного алгоритма имеет смысл для задач, в которых количество классов мало по сравнению с количеством представителей этих классов, поскольку шаг 1 алгоритма требует перебора  $C_N^3$  комбинаций точек, и шаг 2 должен компенсировать эту вычислительную сложность.



Рис. 1: Тестовый фрагмент видеопоследовательности.

На рис. 1 представлен фрагмент тестовой видеопоследовательности (пример обучающего видео с сайта [www.freevidelectures.org](http://www.freevidelectures.org), содержащий несколько смен графического сюжета. Кадры, соответствующие сменам сюжета, выделены цветом и обозначены пунктирной рамкой.

Результат обнаружения опорных кадров по предложенному алгоритму и визуализация соответствующих изображениям векторов признаков приведены на рис. 2. Отметим, что в представленном примере видеопоследовательности встречается возврат к предшествующему графическому сюжету (изображение, появившееся в результате смены сюжета в верхнем ряду кадров, вновь появляется в третьем ряду). Возможность отслеживания подобных ситуаций является важной особенностью подхода к выбору опорных кадров на основе кластеризации.

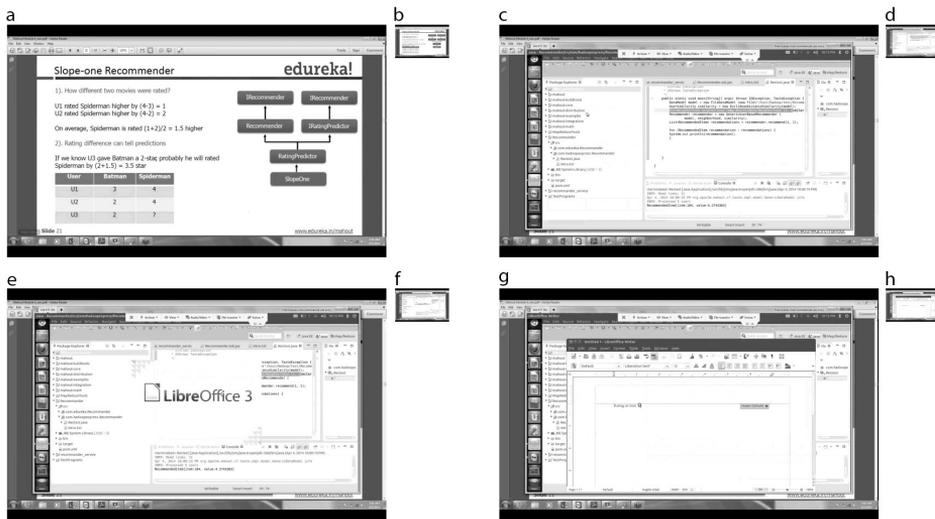


Рис. 2: Опорные кадры, извлечённые из тестового фрагмента видеопоследовательности (а, с, е, г), и соответствующие им образы (b, d, f, h).

На практике анализ видеопоследовательностей в отложенном режиме представляет собой весьма ресурсоёмкий процесс, ввиду значительных объёмов хранимой графической информации. В связи с этим представляет интерес возможность выделения опорных кадров непосредственно в процессе формирования видеопоследовательности. Рассмотрим вариант предлагаемого алгоритма кластеризации в псевдореальном масштабе времени. Пусть  $A$  – пополняемое множество образов, соответствующее множеству  $P$  в отдельные моменты времени. Кроме того, введем множества  $B$  – временный буфер образов, и  $T$  – набор построенных кластеров. В программной реализации множества  $A$ ,  $B$  и  $T$  представляются списками-очередями.

Зададим следующие базовые параметры алгоритма:

- Пороговое значение евклидова расстояния между образами  $\mu_{th}$ .
- Глубина буфера  $D_B$ , определяющая максимально допустимое количество образов в буфере  $B$ .
- Время ожидания кластера  $\Delta t_c$ . Данный параметр задает ин-

тервал времени, в течение которого к кластеру из  $T$  не добавляются новые образы.

- Общий таймаут  $\Delta t$ , определяющий максимальное время ожидания поступления данных в  $A$ .

Алгоритм кластеризации представляет собой пошаговую процедуру, выполняющуюся до тех пор, пока во входное множество  $A$  поступают новые образы. На каждом шаге выполняется следующая последовательность действий:

- 1) Из входного набора  $A$  в буфер  $B$  перемещается некоторое количество образов  $n$ , меньшее глубины буфера  $D_B$ .
- 2) Если множество  $T$  непустое, то предпринимается попытка определить принадлежность образов в буфере  $B$  существующим кластерам из  $T$  по критерию удалённости образа от центра кластера не более чем на  $\frac{1}{2}\mu_{th}$  (причём образ приписывается к тому кластеру, к центру которого он наиболее близок).
- 3) Из оставшихся в буфере образов выбирается триплет  $(b_1, b_2, b_3)$ , для которого минимизируется  $\tilde{d} = \max(d_{12}, d_{23}, d_{31})$ , причём  $\tilde{d} \leq \mu_{th}$ . Кроме того, должно выполняться ограничение (4). Если такой триплет существует, то он образует кластер, помещаемый в  $T$ .
- 4) Для всех кластеров из множества  $T$  проверяется превышение интервала  $\Delta t_c$ . Если время ожидания превышено, выполняется финализация кластера. Под финализацией подразумевается некоторая процедура, включающая в себя сохранение прецедентов, приписанных данному кластеру, выделение центра кластера в отдельный образ и его передачу в очередь  $A$  для дальнейшей кластеризации.
- 5) Если количество образов в буфере  $B$  превышает  $D_B$ , то излишек образов удаляется из буфера в соответствии с принципом FIFO, и эти образы исключаются из дальнейшего рассмотрения.
- 6) Если все кластеры из  $T$  финализированы, и множество  $A$  пусто в течение интервала времени  $\Delta t$ , то производится очистка буфера  $B$ , процесс кластеризации завершается.

## ЗАКЛЮЧЕНИЕ

В работе предложен алгоритм кластеризации видеопоследовательностей, основанный на поиске минимальных трёхточечных симплексов в пространстве признаков и дальнейшем объединении в кластер точек, оказавшихся внутри гиперсфер, описанных вокруг этих симплексов.

Показано, что применение предложенного алгоритма имеет смысл для данных, характеризующихся высокой избыточностью.

Рассмотрен вариант алгоритма триплет-кластеризации в псевдореальном масштабе времени, позволяющий проводить анализ видеопоследовательностей в режиме «онлайн».

## Список литературы

- [1] Дворкович В.П., Дворкович А.В. Цифровые видеоинформационные системы (теория и практика). - М: Техносфера, 2012, 1008 с.
- [2] V. Vijayakumar and R. Nedunchezian. A study on video data mining. International journal of multimedia information retrieval, vol. 1 (3), p. 153-172, 2012.
- [3] Singh, S.; Mandal, A.S.; Shekhar, C.; Vohra, A. Real-time Implementation of Change Detection for Automated Video Surveillance System. ISRN Electronics Volume 2013, 2013.
- [4] Красильников Н. Н. Цифровая обработка 2D- и 3D-изображений. - СПб.: «БХВ-Петербург», 2011. - 608 с.
- [5] Vasileios T. Chasanis, Aristidis C. Likas, Nikolaos P. Galatsanos, Scene detection in videos using shot clustering and sequence alignment, IEEE Transactions on Multimedia, v. 11 n. 1, p. 89-100, 2009.
- [6] Nitin J. Janwe, Kishor K. Bhojar. Video Key-Frame Extraction using Unsupervised Clustering and Mutual Comparison. International Journal of Image Processing, vol. 10 (2), p. 73-84, 2016.

- [7] Современные информационные технологии в задачах навигации и наведения беспилотных маневренных летательных аппаратов / Под ред. М.Н. Красильщикова, Г.Г. Себрякова. – М.: ФИЗМАТЛИТ, 2009, 556 с.
- [8] Мерков А.Б. Распознавание образов: Введение в методы статистического обучения. М.: Едиториал УРСС, 2011, 256 с.