УДК 004.96’14

А.Д. Соколова, А.В. Савченко

КЛАСТЕРИЗАЦИЯ ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЕЙ В СИСТЕМАХ ВИДЕОНАБЛЮДЕНИЯ НА ОСНОВЕ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

Национальный исследовательский университет Высшая школа экономики – Нижний Новгород

Рассматривается задача структурирования информации в программных системах видеонаблюдения с помощью группирования видеоданных, в которых присутствуют идентичные лица. Сделан акцент на эффективную кластеризацию видеопоследовательностей с использованием сверточных нейронных сетей для извлечения характерных признаков. Разработан новый алгоритм кластеризации фрагментов видео на основе технологий глубокого обучения и статистического подхода. Приведены предварительные результаты экспериментального исследования точности и быстродействия предложенного подхода.

Ключевые слова: глубокое обучение, кластеризация, сверточные нейронные сети, распознавание лиц, проверка статистических гипотез об однородности.

Введение

В последние десятилетия в связи с наблюдаемым спросом на технологии видеоконтроля и видеонаблюдения все большее внимание привлекает задача автоматического распознавания объектов по видеоизображению. Среди таких технологий особенно востребованным в сфере обеспечения общественной безопасности являются системы распознавания лиц по видео [1, 2]. Увеличение объема накопленных видеоданных привело к необходимости решения задачи их упорядочивания [3]. Практическую ценность разработанные алгоритмы имеют для систем видеонаблюдения, которые анализируют полученную ими информацию. Например, можно построить систему подсчета различных людей, попадающих в поле зрения камеры за определенный период времени. Таким образом, можно говорить о построении системы наблюдения с возможностью автоматической кластеризации событий [4].

В связи с этим в настоящей работе рассматривается задача автоматической группировки видеоданных, на которых присутствует один человек, на основе методов иерархической агломеративной кластеризации [5]. Основной акцент сделан на использование глубоких сверточных нейронных сетей, которые показали высокую точность при решении многих сложных задач распознавания изображений [6]. На данный момент наблюдается тенденция в создании новых, более глубоких и широких, видов архитектур сверточных нейронных сетей [7, 8].

Постановка задачи

Пусть задана последовательность видеокадров $\{x(t)\}, t=1,2,\ldots,T$, где T – общее число кадров, а $x(t)$ – вектор признаков фиксированной размерности для изображения лица на t-м кадре. Без потери общности можно считать, что лица были предварительно детектированы (например, с помощью метода Виолы-Джонса [9]), и последовательность кадров содержит только их изображение. Также для упрощения дальнейших выкладок предполагается, что последовательность кадров состоит из изображений ровно одного человека. Задача состоит в том, чтобы разбить заданную последовательность на подмножества, а затем похожие подпоследовательности объединять в кластер, далее продолжать объединение до тех пор,

пока объекты не будут составлять один кластер. Разбив на кластеры, должно получиться так, чтобы объекты разных кластеров существенно различались друг от друга, а объекты внутри одного кластера были достаточно близки.

Для реализации данных методов необходимо решать задачу проверки, что на двух видеопоследовательностях изображен один и тот же человек. Это можно свести к проверке гипотез о статистической однородности. Для этого обычно проверяется гипотеза о равенстве математических ожиданий двух выборок с помощью, например, t-критерия Стьюдента[10]. Такой подход рассчитан на числовые независимые выборки, поэтому в случае обработки последовательностей кадров требуется выполнить его модификацию с помощью методов нечисловой статистики и кластеризации на основе использования некоторой меры близости объектов $\rho(x, x`)$ [11].

Один из вариантов решения задачи состоит в выполнении периодизации двух уровней [12]. В начале выполняется периодизация первого уровня, то есть в заданных видеоданных выделяется множество из последовательных однородных сегментов. Алгоритм сводится к тому, что проверяется условие однородности распределений вектора отсчетов $x(t)$ текущего фрейма и вектора отсчетов текущего однородного сегмента X_l. Текущий накопленный сегмент x_l и t-й фрейм следует объединить, если выполняется следующее условие [13]:

$$r(x(t), X_l) < r_0. \quad (1)$$

В противном случае количество однородных сегментов L увеличивается и полагаем X_{l+1}= $\{x(t)\}$. Для инициализации проверки (1) вначале полагают X_1= $\{x(1)\}$. Такая процедура повторяется для всех последующих фреймов, в результате получается множество из L выявленных однородных сегментов (треков), на которых отображается один и тот же человек. Далее происходит периодизация второго уровня, а именно выделенные однородные сегменты объединяются в кластер, то есть для всех $l = \overline{1, L^*}$ проверяется условие $\exists j \in \{1,...,l_r\} \rho(x_l^*, x_j^{(r)}) < \rho_2 = const$, $r = \overline{1, R}$ об однородности распределений вектора отсчетов текущего однородного участка x_l^* и вектора отсчетов из списка $\{X_r^*\}$. Здесь ρ_2 - допустимый уровень рассогласования, а кластер X_r^* представляет множество объема l_r близких однородных сегментов.

Альтернативный способ осуществления периодизации второго уровня связан с вычислением среднего расстояния между элементами и дисперсии расстояния внутри трека, а также между элементами двух разных треков [4].

$$\overline{\rho}_{in} = \frac{2}{N_{1q}(N_{1q}-1)} \sum_{i=1}^{N_{1q}} \sum_{j=i+1}^{N_{1q}} \rho_{ij}, \quad (2)$$

$$D(\rho_{in}) = \frac{2}{N_{1q}(N_{1q}-1)} \sum_{i=1}^{N_{1q}} \sum_{j=i+1}^{N_{1q}} \rho_{ij}^2 - \overline{\rho}_{in}^2, \quad (3)$$

$$\overline{\rho}_{ex} = \frac{2}{N_{1q}(N_{2q}-1)} \sum_{i=1}^{N_{1q}} \sum_{j=i+1}^{N_{2q}} \rho_{ij}, \quad (4)$$

$$D(\rho_{ex}) = \frac{2}{N_{1q}(N_{2q}-1)} \sum_{i=1}^{N_{1q}} \sum_{j=i+1}^{N_{2q}} \rho_{ij}^2 - \overline{\rho}_{ex}^2, \quad (5)$$

где ρ_{ij} – евклидово расстояние между векторами признаков i-го и j-го кадров, N_{1q} и N_{2q}– количество элементов первого и второго однородного сегмента.

Считается, что два трека будут содержать изображение одного и того же человека, т.е. их можно объединить в один кластер, если выполняется:

$$\begin{cases} \overline{\rho}_{ex} - \overline{\rho}_{in} < K\sqrt{D(\rho_{in})} \\ \overline{\rho}_{ex} - \overline{\rho}_{in} < K\sqrt{D(\rho_{ex})} \end{cases}, \quad (6)$$

где K – параметр, регулирующий жесткость условия.

К сожалению, такой способ кластеризации является слишком вычислительно сложным, так как в выражениях (2)-(4) вычисляются рассогласования между всеми кадрами всех сегментов. Для преодоления указанного недостатка в настоящей работе предложен следующий вариант периодизации второго уровня. В критерии Стьюдента вместо средних величин будем сопоставлять их медоиды x_l^* – кадр, соответствующий минимальной сумме расстояний до остальных кадров однородного сегмента X_1. Вместо разности средних значений в традиционном t-критерии используется расстояние между двумя медоидами кадров, а вместо дисперсий элементов выборок для каждого видео вычисляются оценки дисперсий расстояний D_1 (сумма квадратов расстояний между каждым кадром сегмента и его медоидом).

$$\frac{r(x_i^*, x_j^*)}{\sqrt{\frac{D_i}{n_i} + \frac{D_j}{n_j}}} < t = const, \quad (7)$$

где n_i – количество фреймов в i-м однородном сегменте (треке). В таком случае для каждого трека достаточно вычислить только его медоид, а далее для объединения двух треков в один кластер сравниваются только медоиды, а не все кадры, что приведет к существенному повышению вычислительной эффективности по сравнению с известным подходом (2)-(6) [4].

На рис. 1 представлена предлагаемая структурная схема системы видеонаблюдения с автоматической кластеризацией.

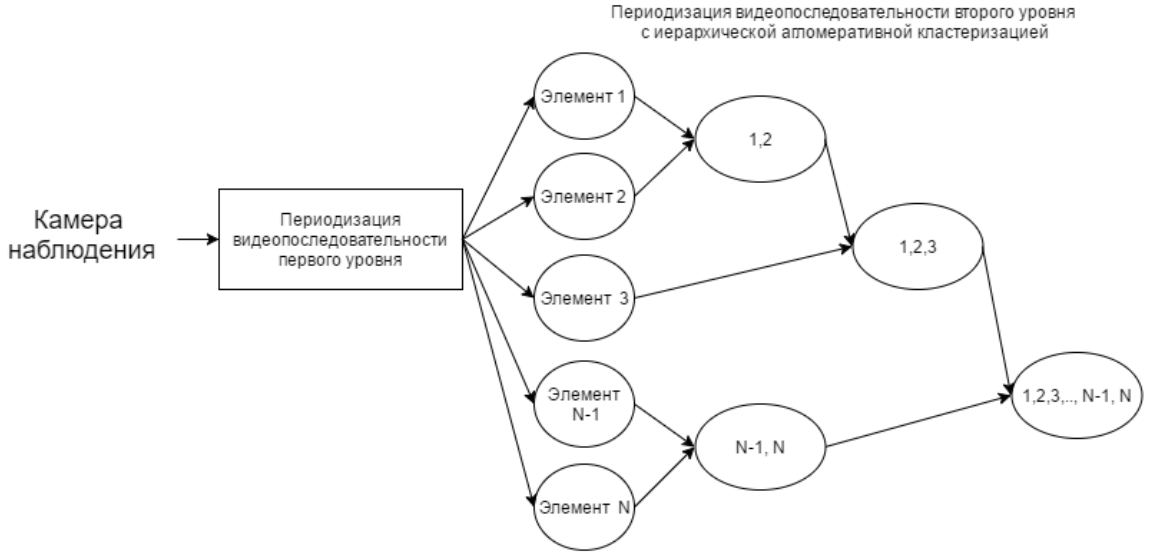

Рис. 1. Схема работы системы видеонаблюдения с автоматической кластеризацией

Обработка входной видеопоследовательности начинается с поиска идущих последовательных кадров одного человека (периодизация первого уровня (1)), с помощью которого получается совокупность треков (однородных сегментов). Далее осуществляется поиск схожих сегментов (периодизация второго уровня) с помощью методов агломеративной кластеризации: схожие объекты последовательно объединяются в группы, создав в конце итоговый кластер. Для осознания того, разные ли люди на изображении или нет, используется основанные на t-критерии Стьюдента выражения (7). Результатом работы алгоритма является совокупность выделенных кластеров однородных сегментов (треков).

Практическая реализация системы

Разработка и тестирование предложенной системы (рис. 1) проводилось на языке C++ с помощью MS Visual Studio 2015 с использованием библиотеки OpenCV [14]. Среднее время распознавания одного кадра на ПК Lenovo ideapad 310, 64-разрядной операционной системе Windows 10 составляет 150 мс. Изображение каждого видеокадра переводится в полутона, после чего с помощью метода Виолы-Джонса и каскадов Хаара детектируется область лица (рис. 2). Для повышения точности детектирования на всех обнаруженных лицах осуществлялся поиск области глаз.

Рис. 2. Детектирование лица и перевод в полутона

Для извлечения 256 признаков из каждого изображения лица использовалась глубокая сверточная нейронная сеть Lightened CNN C [8], преимуществами которой является скорость нахождения вектора признаков детектированного объекта (в среднем 60 мс), а также высокая точность верификации лиц для нескольких современных наборов данных [8]. На данный момент реализована периодизация первого уровня (1). Вектор признаков вычисляется у кадра, соответствующего минимальной сумме расстояний до остальных кадров, то есть медианы, затем сравнивается с векторами других изображений. Для определения того, один ли человек на кадрах или это разные люди, экспериментально было подобрано пороговое значение $r_0 = 100$, который сравнивался с евклидовым расстоянием между векторами признаков. Если евклидово расстояние меньше 100, то можно утверждать, что на последовательных кадрах присутствует один и тот же человек, в противном случае – обратное. При этом процедура проверки (1) выполнялась только после того, как на нескольких последовательных кадрах лицо не было обнаружено.

Выводы

Таким образом, в настоящей работе предложен подход к группировке видеоданных и осуществлена программная реализация периодизации первого уровня (1) на основе признаков, выделенных с помощью сверточных нейронных сетей. Следующий этап работы состоит в реализации предложенного подхода (7) к периодизации второго уровня с иерархической кластеризацией. Кроме того, в дальнейшем необходимо провести ряд экспериментальных исследований предлагаемого алгоритма. Например, с помощью базы данных YouTube Faces можно оценить конкретный порог определения различия людей на изображениях, а также проверить точность реализуемого алгоритма. Также возможно применение такой известной архитектуры для нейронных сетей как VGGNet [15] и сравнить полученные результаты с моделью Lightened CNN C [8].

В итоге подразумевается получить систему, которая сможет распознавать людей по разным видеоданным и автоматически выполнять их кластеризацию. Преимуществами предлагаемой системы является как низкая вычислительная сложность, так и ожидаемая высокая точность распознавания на основе модификации традиционного t-критерия Стьюдента.

Статья подготовлена в результате проведения исследования (№ 17-05-0007) в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2017 г. и в рамках государственной поддержки ведущих университетов Российской Федерации "5-100".

Библиографический список

1. **Chellappa, R.** Face Tracking and Recognition in Video./ Chellappa R., Du M., Turaga P., Zhou S.K. // Handbook of Face Recognition. 2011. P. 323-351
2. **Shan, C.** Face Recognition and Retrieval in Video. Video Search and Mining, Studies in Computational Intelligence. 2010. V. 287. P. 235-260
3. **Savchenko, A.V.** Search Techniques in Intelligent Classification Systems. Springer International Publishing, 2016.
4. **Шемагина, О.В**. Системы обнаружения, сопровождения и кластеризации объектов на основе нейроноподобного кодирования./ О.В.Шемагина, Н.С.Беллюстин, Ю.Д.Калафати, А.В.Ковальчук, А.А.Тельных, В.Г.Яхно // Информационно-Измерительные И Управляющие Системы. - 2010. - Т. 8. - No 2.- С. 29–34.
5. **Вовк, О.Л.** Иерархический агломеративный алгоритм кластеризации для выделения регионов изображений //Труды XIV Международной конференции по компьютерной графике и зрению “GrapiCon. – 2004. – С. 245-248.
6. **Хайкин, С.** Нейронные сети: полный курс, 2-е издание. – Издательский дом Вильямс, 2008.
7. **Jia, Y**. et al. Caffe: Convolutional architecture for fast feature embedding //Proceedings of the 22nd ACM international conference on Multimedia. – ACM, 2014. – С. 675-678
8. **Wu, X.** A Lightened CNN for Deep Face Representation./ Wu X., He R., Sun Z. //arXiv preprint arXiv:1511.02683 (2015)
9. **Акимов, А.В.** Разработка и исследование алгоритмов распознавания изображений на основе метода Виолы-Джонса с использованием технологии вычислений на графических процессорах CUDA./ А.В. Акимов, А.А. Сирота //Вестник ВГУ, Серия: Системный анализ и информационные технологии. – 2014. – №. 3. – С. 100-108.
10. **Орлов, А.И**. Проверка статистической гипотезы однородности математических ожиданий двух независимых выборок: критерий Крамера-Уэлча вместо критерия Стьюдента //Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2015. – №. 110.
11. **Новиков, Д.А.** Математические методы классификации./ Д.А. Новиков, А.И. Орлов // - Журнал «Заводская лаборатория». 2012. Т.78. No.4. С.3-3.
12. **Савченко, А.В**. Об одном подходе к разработке автоматизированной системы дистанционного обучения произношению слов на основе вероятностной нейронной сети с проверкой однородности / А.В. Савченко // в сб. «Нелинейная динамика в когнитивных исследованиях-2013». Институт прикладной физики РАН, Нижний Новгород.– 2013.– С.144-147
13. **Savchenko, A.V**. Probabilistic neural network with homogeneity testing in recognition of discrete patterns set / A.V. Savchenko // Neural Networks. 2013. Vol. 46. P. 227–241.
14. **OpenCV**: библиотека алгоритмов компьютерного зрения [Электронный ресурс]. – Режим доступа: http://opencv.org/, свободный. – Загл. с экрана.
15. **Parkhi, O.M.** Deep Face Recognition./ O.M. Parkhi, A. Vedaldi, A. Zisserman //BMVC. – 2015. – Т. 1. – №. 3. – С. 6.