

Распознавание объектов: интеллектуальные системы видеонаблюдения, мониторинг изображений в Интернете, автоматический ввод графической информации.

Ю.Д. Калафати

к.ф.-м.н.

Генеральный директор ООО «ССТ - Технология хаоса»

В настоящее время распознавание объектов на цифровых изображениях и видеопотоке из стадии лабораторных исследований перешло в стадию внедрения в различные коммерческие проекты. Ведутся также работы, которые сделают возможным определить не только отдельные объекты, но и целиком сценарий (отношения между объектами), представленный на изображении. Этому способствовали значительные финансовые как государственные так и частные вложения в Европе и США в технологии систем безопасности. В результате масштабных инвестиций были созданы эффективные алгоритмы и методы идентификации личности. Современные фотоаппараты автоматически фокусируются при помощи алгоритмов поиска лиц на изображении и даже анализируют эмоциональное состояние этих лиц (улыбка, возраст и т.п.). Все это говорит о значительных теоретических и прикладных достижениях в области распознавания двумерных образов, которые произошли за последние годы.

В компании «ССТ – Технология хаоса» работы по применению технологий распознавания образов ведутся в нескольких направлениях. Разрабатываются алгоритмы по поиску различных объектов на цифровых фотографиях и в видеопотоке. К таким объектам относятся лица людей, автомобили, различные символы, логотипы компаний. При детектировании лица на изображении определяется также пол человека, возраст (взрослый – ребенок) и другие характеристики (наличие очков, волос и т.п.). Разрабатывается детектор, автоматически классифицирующий формулы органических соединений. Созданные детекторы затем используются для разработки интеллектуальных систем видеонаблюдения, для разработки систем мониторинга изображений в Интернете (экспертных систем) и автоматического ввода графической информации в базы данных.

В настоящем докладе, который основан на материалах [1], подробно рассказывается о классификации сценариев в видеопотоке. Процесс нахождения и анализа объектов на видеокадрах имеет существенные отличия от анализа цифровых статических изображений. Основной проблемой при создании архива (журнала) появлений объекта, например, лица человека, в поле зрения видеокамеры состоит, во-первых, в сильной избыточности информации, а во-вторых, в необходимости определения объекта на каждом кадре. При этом в качестве единицы хранения следует выделять не каждый кадр, а последовательность кадров (трек) на которых присутствует один и тот же объект. Для этого необходимо применять целый набор алгоритмов анализа изображений. В числе этих алгоритмов:

- 1) алгоритм детектирования объекта (например, лица) на каждом кадре,
- 2) алгоритм детектирования движения,
- 3) алгоритм сопровождения объекта,
- 4) алгоритм сравнения (верификации) объектов (лиц).

Соединение этих алгоритмов в единое целое для интеллектуального описания и архивирования видеопотока представляет собой серьезную технологическую задачу даже при существовании разработанных алгоритмах анализа изображения.

Важным элементом для разработки интеллектуальных систем видеонаблюдения является сегментация видеозаписи.

Под сегментацией видеозаписи мы понимаем процедуру разбиения видеозаписи на части, содержащие последовательности кадров с выделенным на них объектом одного

типа. В случае временного исчезновения объекта из кадра таких последовательностей получается несколько, и их совокупность образует кластер. В данной работе описывается алгоритм сегментации видеозаписи (разбиение видеозаписи на кластеры), показавший свою эффективность для решения практических задач. Система сегментации видеопотока необходима для предварительного анализа и сортировки данных с систем видеонаблюдения или передающихся по телевизионным каналам. Система сегментации может применяться и самостоятельно, например, с ее помощью рекламодатель может осуществлять контроль эфирного времени. Важной особенностью задачи сегментации видеопотока является то, что большинство приложений требует работы в реальном времени, т.е. необходимо обрабатывать десятки кадров в секунду. В качестве интересующего нас объекта было выбрано лицо человека. Для детектирования лица человека на избранных кадрах видеоклипа использовался созданный в компании ССТ – Технология хаоса высококачественный детектор с ошибкой ложного обнаружения (FAR) $10^{-6} - 10^{-7}$ на фрагмент, что позволило им создать достаточно надежную систему сегментации.

В разработанной нами системе процесс сегментации происходит в три этапа.

Этап выделения объектов

Входными данными для первого этапа служит видеоклип. В процессе просмотра видеоролика разработанный в компании ССТ – Технология хаоса детектор лица человека [2] анализирует каждый кадр видеопотока, выделяет лица, и формируется файл отчета, содержащий информацию о каждом обнаруженном объекте (номер кадра в видеопотоке, местоположение объектов на данном кадре, идентификатор каждого объекта). На первом этапе детектор лица работает только в режиме обнаружения: сканирование проводится по всему изображению, детектор в бинарном режиме определяет лицо или не лицо на каждом фрагменте, с последующей кластеризацией накладывающихся друг на друга фрагментов с объектом [2]. Области с найденным объектом «передаются» системе сопровождения, позволяющей спрогнозировать положение объектов на последующих кадрах. В этих областях детектор работает не в бинарном режиме (лицо – не лицо), а в количественном режиме.

Результатом работы детектора в этом режиме является значение «степени уверенности», вычисляемое по промежуточным результатам обработки - суммируются рейтинговые веса всех слабых классификаторов [3] во всех каскадах детектора, проголосовавшие за рассматриваемый фрагмент. Получаемая оценка степени уверенности сравнивается с регулируемым порогом для принятия окончательного решения. Начиная с того кадра, где был найден объект, детектор работает одновременно в двух режимах: количественном в областях с прогнозируемым местоположением лица и в бинарном для всех остальных фрагментов изображения, что необходимо для поиска новых лиц (Рис.1).



Рис.1 Результат работы детектора лиц

Этап формирования треков

На этапе формирования треков собираются в трек последовательные кадры видеоклипа, содержащие один и тот же объект, и все объекты одного трека получают один и тот же идентификатор. Трек представляет собой непрерывную последовательность кадров с лицом одного и того же человека – если на каком-то кадре лицо в искомой области не обнаруживается, то трек прерывается (Рис.2). Завершение трека может произойти по двум причинам – или из-за того, что лицо действительно вышло из кадра, или из-за ошибки пропуска лица детектором. Второе обстоятельство делает особенно важным следующий, третий, этап - объединение в один кластер всех треков с лицом одного и того же человека. Такая процедура кроме других задач поможет уменьшить влияние на конечный результат ошибок пропуска лица. По окончании второго этапа мы получаем набор N треков, содержащих k_i , $i = 1, \dots, N$ кадров с выделенной на них областью лица.



Рис.2 Последовательность кадров с выделенной на них областью лица – трек

Этап формирования кластеров

Задачей третьего этапа является объединение треков, содержащих один и тот же объект (в нашем случае лицо одного и того же человека), в кластер. При этом вводится понятие расстояния между двумя кластерами. Сначала области лица вырезаются на каждом кадре трека и приводятся к стандартизованному виду. В нашем случае изображения приводятся к размеру 32x32 пикселя, нормируются по освещенности и поворачиваются так, чтобы глаза были расположены горизонтально (Рис.3).



Рис.3 Пример стандартизованного изображения

С использованием метода главных компонент [3] и [5] все изображения проецируются на ранее созданное подпространство невысокой размерности L – после чего каждое лицо характеризуется совокупностью L чисел. Таким образом, теперь каждый трек представляет собой k_i последовательностей L чисел.

Далее полагаем, что на первом этапе каждый кластер состоит из одного трека, и начинаем процесс объединения кластеров. В качестве меры близости используем минимальное расстояние между парами треков, входящих в разные кластера.

При сравнении двух треков длины k_i в первом и k_j во втором, каждая проекция изображения лица в одном треке сравнивается с проекцией изображения лица во втором треке. В качестве меры используется евклидово расстояние. Расстояние между треками определяется как минимальное расстояние между элементами этих треков.

В итоге мы имеем матрицу расстояний между кластерами размерности $N \cdot N$.

Ищется минимальный r_{\min} и максимальный r_{\max} элемент этой матрицы, после чего вычисляется порог $Thr = r_{\min} + K \cdot (r_{\max} - r_{\min})$, где параметр K – регулировочный. Далее кластера с минимальным расстоянием между собой объединяются, для укрупненного кластера пересчитываются элементы матрицы расстояний, и процедура повторяется до тех пор, пока минимальный элемент матрицы расстояний меньше этого порога – если это

условие не выполнено, то процесс прекращается. Результатом выполнения третьего этапа является список кластеров, каждый из элементов которого содержит список принадлежащих ему треков, состоящих, в свою очередь, из фрагментов видеоролика, на которых находится лицо предположительно одного и того же человека.

Ошибка работы данного алгоритма складывается из трех составляющих. Первая составляющая – это ошибка детектирования лица на изображении. Она невелика и составляет для нашего детектора $<10\%$.

Вторая составляющая связана с ошибкой системы сопровождения, которая приводит к попаданию в трек «артефактов» - кадров с выделенными областями, не содержащими лица. Управление параметром, позволяющим регулировать уровень этой ошибки, выведен в пользовательский интерфейс.

И третья составляющая – ошибка алгоритма распознавания, основанного на использовании метода главных компонент (примерно 20%).

Заключение

Разработанная система хорошо себя показала на тестовых примерах и, при подборе оптимальных параметров системы слежения, ошибка сводится к ошибке алгоритма распознавания, основанного на методе главных компонент. Поэтому дальнейшие исследования будут связаны с поиском оптимального пространства признаков.

Представленная система также предполагается для использования в человеко-машинных интерфейсах

Список литературы.

1. Telnykh A. A., Bellustin N.S., Shemagina O.V., Kalafati Yu.D., Yakhno V.G. The training algorithm of a neuron-like object detector and variants of its implementations. Алгоритм сегментации видеоклипа и его реализация., 9 International Conference on Pattern Recognition and Image Analysis: New Information Technologies (PRIA-9-2008) Conference Proceedings V.2, pp. 208 – 214., Nizhni Novgorod, Russian Federation, 2008.
2. Беллюстин Н.С., Калафати Ю.Д., Ковальчук А.В., Тельных А.А., Шемагина О.В., Яхно В.Г., Нейроподобный детектор лица. Технические особенности реализации и обучения, X Всероссийская научно-техническая конференция "Нейроинформатика-2008", Сборник научных трудов, часть 2 МИФИ, Москва, янв. 2008, с. 123-132
3. Viola P., Jones M. Rapid Object Detection using a Boosted Cascade of Simple Feature. Conference on Computer Vision and Pattern Recognition 2001
4. С. Хайкин, Нейронные сети: полный курс, М., ООО «И.Д. Вильямс», 2006, с.509-572
5. J.- Y.Bouguet, Pyramidal Implementation of the LucasKanadeFeature Tracker: Description of the algorithm, Technical Report, Intel Corporation: Microprocessor Research Labs, 2000, OpenCV documentation.
6. M. Turk and A. Pentland. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3:71-86, 1991