

А. А. Андреев, А. Р. Гасанов, К. Э. Гришкевич, Ю. А. Егоров

Тюменский государственный университет, г. Тюмень

УДК 004.85

РАЗРАБОТКА ПРИЛОЖЕНИЯ ДЛЯ ОТДЕЛЕНИЯ КАДРОВ В ВИДЕОПОТОКЕ ПО НАЛИЧИЮ КЛЮЧЕВЫХ ОБЪЕКТОВ

Аннотация. В статье рассмотрено применение методов машинного обучения для отделения кадров с интересующим объектом из видеопотока от кадров, не содержащих изображение объекта.

Ключевые слова: компьютерное зрение, машинное обучение, видеопоток, метод опорных векторов, дескрипторы изображений

Введение

В данной статье рассмотрена разработка приложения для отделения из видеопотока кадров, содержащих изображение человека, от кадров без людей.

В текущее время все большую популярность обретает использование систем видеонаблюдения в различных сферах деятельности человека. Видеокамеры используют в сельском хозяйстве, системах охраны, в местах массового скопления людей. Системы видеонаблюдения предоставляют возможность хранения происходящих событий для дальнейшего визуального анализа. Самостоятельный просмотр видеозаписей затратный по времени процесс. Также при ручном анализе видеоряда имеет место человеческий фактор.

Возникает задача автоматизированного извлечения фрагментов видеозаписи, полезных для определённых нужд, будь то поиск кадров с преступником на записи, или отслеживание подозрительных действий на едином государственном экзамене с целью выявления нарушений. Необходимость фильтрации кадров обусловлена тем, что зачастую можно

избежать обработку заведомо пустых кадров. Было решено автоматизировать отделение кадров, содержащих изображение человека, из общего потока.

Приведем описание задачи фильтрации кадров. Дана видеозапись, для которой необходимо провести бинарную классификацию кадров. К первому классу относятся кадры, на которых присутствуют ключевые объекты (в данном случае такими объектами являются люди), ко второму – кадры, на которых отсутствуют ключевые объекты.

Целью данного проекта было узнать, можно ли при помощи метода опорных векторов, обученного на дескрипторах изображений, реализовать фильтрацию кадров видеозаписи.

Обзор литературы

Были изучены работы, описывающие решение поставленной проблемы.

В книге [1] предложен детектор людей на основе дескриптора НОГ и классификатора SVM. Детектирование объектов является схожей задачей, так как в ходе её выполнения определяется наличие объекта на изображении.

Построение описания локальных окрестностей ключевых точек и применение их для формирования мешка слов рассмотрено в статье [2]. В данном способе обучение классификатора происходит на построенном описании каждого изображения из выборки. Описание представляет собой гистограмму встречаемости «слов».

Описание обучающей выборки

Обучающая выборка была сформирована из кадров видеозаписей с единого государственного экзамена. Исходные видео были разбиты по кадрам, после чего изображения были преобразованы к формату **160 × 80** пикселей. Случайным образом были выбраны кадры с аудиториями с различной планировкой и освещением, таким образом, чтобы присутствовали в равном количестве кадры, на которых есть люди и кадры, на которых их нет.

Обучение классификатора происходило на цельных кадрах из видео. Величина выборки составила 2000 кадров. Люди, изображенные на кадрах обучающей выборки, принимают разное положение в пространстве, могут

быть частично скрыты, могут находиться как лицом к камере, так и располагаться спиной или в профиль.

Имеющиеся видеозаписи различаются по степени освещенности помещения. Данный фактор влияет на насыщенность цвета пикселей. Это сказывается на восприятии изображения классификатором, и даже пустая аудитория при разном освещении может выдавать различные результаты после обработки классификатором.

Планировка в каждом кабинете также разная. Различие локаций позволяет проверить работу классификатора в различных условиях.

Методы решения

Были построены две модели. Первая модель обучена на дескрипторах изображений. Вторая модель работает на необработанных изображениях.

Рассмотрим принцип работы первой модели. На каждом кадре, поступающем на вход, происходит детектирование особых точек на основании интенсивности цвета, после чего формируется описание локальных окрестностей данных точек в виде дескриптора SIFT. Данный дескриптор инвариантен по отношению к аффинным изменениям освещенности.

После определения описаний ключевых точек осуществляется кластеризация дескрипторов при помощи алгоритма k-means. Формируется словарь, где в качестве «слов» выступают центроиды полученных кластеров. Строится описание каждого отдельно взятого кадра: для каждого кластера подсчитывается число ключевых точек данного изображения, отнесенных к нему. Таким образом формируется признаковое описание кадров.

На основе полученного описания кадров обучается классификатор SVM.

Для оценки точности работы готовой модели применены метрика Ассигасу и метод перекрестной проверки с разбиением выборки на 5 частей. Точность классификации данной модели составила 85%.

Также была построена вторая модель. В качестве описания изображений выступают векторы, сформированные из значений насыщенности цвета каждого пикселя. Классификатор SVM обучается на образованных векторах. Такая модель в результате кросс-проверки показала точность в 71%.

Модель, построенная на основе мешка слов с применением дескриптора SIFT показала большую точность по сравнению со второй моделью.

Большая точность первого классификатора обусловлена использованием дескриптора, так как кадры с различным освещением при обработке первым классификатором получали правильную метку в результате классификации чаще, чем те же кадры при обработке вторым классификатором, так как дескриптор инвариантен к аффинным изменениям освещенности.

Также на точность распознавания модели влияет качество обучающей выборки. При схожих по дескрипторам наборах особых точек возможны ситуации, когда точки находятся в разных конфигурациях, при этом описания объектов совпадают. В результате наблюдается уменьшение точности классификации.

В настоящей работе использованы следующие инструменты: язык программирования Python 3, фреймворк OpenCV (написание модели), NumPy.

Заключение

Точность распознавания наличия человека в кадре при кросс-валидации на выборке из 2000 изображений составила 85%.

Несмотря на довольно высокую точность распознавания, следует отметить, что построенная модель не учитывает пространственную информацию о распределении ключевых точек на изображении, что может приводить к ухудшению результатов классификации.

Предполагается, что учет пространственной информации о распределении ключевых точек изображений позволит улучшить качество классификации кадров по признаку наличия-отсутствия на нем интересующих объектов.

В перспективе, обученную модель можно применить при разработке приложений для выделения наиболее важных фрагментов из видеопотока, что может упростить анализ видеозаписей и уменьшить человеческий фактор.

СПИСОК ЛИТЕРАТУРЫ

1. Kaehler A. Learning OpenCV 3 [Текст] / A Kaehler, G Bradski. – Sebastopol: O'Reilly, 2016. – 992 с.
2. Jun Yang. Evaluating bag-of-visual-words representations in scene classification [Текст] / Jun Yang, Yu-Gang Jiang, Alexander G. Hauptmann, Chong-Wah Ngo // Proceedings of the international workshop on Workshop on multimedia information retrieval: сб. статей. – Аугсбург, 2007. – С. 197-206.