Ми представляємо семантико-ймовірнісну модель опису поведінки людини і її дій, що об'єднує в собі елементи графічної ймовірнісної моделі — Байєсівської мережі, а також семантичної моделі на основі ієрархічної онтології предметної області, для визначення контексту відеопотоку в інтелектуальних системах відеоспостереження

Ключові слова: Байєсова мережа, онтологія, відеоспостереження, семантична модель, ймовірнісна модель, людська поведінка

Мы представляем семантико-вероятностную модель описания поведения человека и его действий, объединяющую в себе элементы графической вероятностной модели — Байесовской сети, а также семантической модели на основе иерархической онтологии предметной области, для определения контекста видеопотока в интеллектуальных системах видеонаблюдения

Ключевые слова: Байесовская сеть, онтология, видеонаблюдение, семантическая модель, вероятностная модель, человеческое поведение

УДК 004.932.72'1

СЕМАНТИКО-ВЕРОЯТНОСТНАЯ СЕТЬ ДЛЯ ОПРЕДЕЛЕНИЯ КОНТЕКСТА ВИДЕОПОТОКА В СИСТЕМАХ ВИДЕОНАБЛЮДЕНИЯ

Н. В. Коваленко
Аспирант*
E-mail: kov.nikit@gmail.com
С. Г. Антощук
Доктор технических наук, профессор*
E-mail: svetlana_onpu@mail.ru
*Кафедра информационных систем
Одесский национальный политехнический университет
пр. Шевченко, 1, г. Одесса, Украина, 65044

1. Введение

В последние годы наблюдается бурное развитие систем видеонаблюдения. Это привело к необходимости совершенствования существующих аппаратных и программных средств обработки и анализа видеоинформации, подходов к ее хранению и определению контекста и его поиску по запросу.

Распознавание человеческого поведения сегодня является одной из наиболее важных задач, решаемых системами видеонаблюдения. Необходимость распознавать сложные действия человека и определять его поведение в видеопотоке возникает во многих важных приложениях, начиная от мониторинга состояния пациентов в медицинских учреждениях и заканчивая обеспечением безопасности и предотвращением преступлений. Еще больший интерес представляют системы, способные не только распознавать поведение отдельных индивидуумов в видеопотоке, но и оценивать характер взаимодействия между ними.

Подавляющее большинство существующих систем видеонаблюдения (СВН) для мониторинга предполагают человека-оператора. Поскольку современные СВН могут состоять из сотен видеокамер, то эффективность таких систем обусловлена не технологическими возможностями или расположением и количеством камер, а внимательностью и численностью персонала, обслуживающего систему. Более того, даже хорошо тренированные работники не в состоянии поддерживать высокую степень внимания в течение продолжительного периода времени. В связи с этим возникает необходимость в автоматизированной обработке видеоряда, полученного системой наблюдения, с целью определения «нестандартных событий» и обеспечения круглосуточной эффективной работы СВН.

2. Анализ литературных данных и постановка проблемы

Распознаванию поведения человека и его действий в видеопотоке в последние годы было посвящено большое число исследований [1]. Существующие методы анализа и распознавания человеческого поведения можно классифицировать на две категории: однослойные подходы и иерархические подходы.

Однослойные подходы определяют и распознают действия человека «напрямую», путем обработки и оценки низкоуровневых характеристик по последовательности изображений. Однослойные подходы можно разделить в зависимости от того, как они моделируют человеческое поведение, на две категории: пространственно-временные и последовательностные подходы [2 — 5]. Однослойные подходы хорошо подходят для распознавания жестов и действий с последовательностными характеристиками.

Иерархические подходы позволяют проводить определение на каждом слое набора простых действий, называемых обычно под-событиями, и получать высокоуровневые характеристики для оценки действий поведения на видеоряде. Построение систем, состоящих из нескольких слоёв, позволяет проводить анализ сложных событий и действий.

Иерархические подходы классифицируются в зависимости от используемых методологий распознавания: статистические подходы [6], синтаксические

подходы [7] и подходы, основанные на описаниях (дескрипторные, описательные) [8].

Статистические и синтаксические подходы обеспечивают вероятностные фреймворки для надежного распознавания в условии шумов. Однако, они имеют сложности с представлением и распознаванием действий, имеющих пересекающиеся под-события.

Дескрипторные подходы представляют человеческие действия путем описания под-событий и их временных, пространственных и логических структур. К таким подходам относятся семантические модели, позволяющие проводить семантический анализ взаимодействий между людьми или объектами, а также сложных групповых взаимодействий. Преимуществом таких подходов является способность обходиться меньшим объемом тренировочных данных и использовать априорные знания о предметной области, а недостатком - неспособность компенсировать ошибки, возникающие на более низком уровне (например, ошибка при распознавании жеста). Среди статистических подходов больше всего выделяются графические вероятностные модели (сети Байеса, скрытые модели Маркова), которые часто используются при решении задачи распознавания поведения объектов.

Подход, основанный на скрытых моделях Маркова, применяется для решения различных задач, например, при поиске и распознавании лиц на изображениях. Однако природа визуальной информации носит достаточно сложный характер, что при обработке изображений и видеоинформации снижает эффективность таких механизмов, как скрытые модели Маркова и стохастический синтаксический анализ, популярные в области распознавания и обработки речи.

Поэтому для решения проблем компьютерного зрения при обработке видеоинформации с учетом с ее сложности и неопределенности лучше подходят Байесовские сети [9].

3. Цель и задачи исследования

Основной недостаток графических вероятностных моделей, в частности Байесовской сети, — отсутствие четкой структуризации и формализации данных, в результате чего построение структуры Байесовской сети может быть достаточно сложной задачей, особенно в таких сложных системах, как системы распознавания человеческого поведения. Высокая сложность этой задачи приводит к повышению сложности анализа, и, как следствие, к необходимости ограничивать фокус разрабатываемой системы, либо экспоненциально увеличивать размер и сложность Байесовской сети.

Для формализации процесса построения структуры Байесовской сети и обеспечения эффективного логического вывода и интерпретации вероятностных выходов предлагается использование структурированных знаний о предметной области и обучающих выборок, аннотированных дескрипторами спецификации, которые соответствуют проводимому анализу. Для такого формального описания предложено использовать семантическую модель — онтологию, которая предоставляет иерархически структурированное априорное описание предметной области в терминах сущностей, состояний, событий и отношений, которые представляют интерес.

Таким образом, целью данной работы является разработка семантико-вероятностной модели на основе Байесовской сети и онтологического аппарата для определения контекста видеопотока в интеллектуальных системах видеонаблюдения.

4. Построение онтологии предметной области

Онтологии содержат реляционную структуру концептов, которые могут быть использованы для описания и рассуждения об аспектах мира. Содержимое сцены организуется в виде иерархической онтологии (рис. 1), путем его декомпозиции на несколько уровней абстракции: общие сценарии, ситуации, в которых оказываются объекты, выполняемые ими роли и их состояния (атрибуты).

Рис. 1. Онтология предметной области

На самом нижнем уровне абстракции находятся визуальные дескрипторы, вычисляемые из самого видеопотока: скорость объекта, трек, состояние перекрытия и некоторые другие.

Состояния (атрибуты) объектов представляют собой свойства динамики объектов, и показывают, в каком состоянии или состояниях в данных момент находится объект. Допустим, человек может одновременно находиться в состояниях "Активен", "Бежит" и "Машет руками".

Роли объектов представляют собой общую линию (модель) поведения человека в кадре, выводимую из состояний, в которых он находится. Таким образом, роль может складываться из нескольких состояний. Например, состояния "Активен" и "Бежит" означают роль "Бегущего", в то время как состояния "Бежит" и "Машет руками" может означать роль "Паникующего". Ситуации, в которых находятся действующие лица, зависят от их ролей и их состояний в сцене и показывают как объекты взаимодействуют с ней и друг с другом. Например, ситуация "драка" может состоять из двух объектов, пребывающих на близком расстоянии и движущихся по направлению друг к другу, при этом выполняющих роли "Бегущих" и имеющих состояния "машущих руками".

Наконец на самом высоком уровне абстракции находятся сценарии – общий контекст, общая обстановка в кадре, складывающаяся из набора ситуаций.

Приведенную систематизацию предлагается учитывать при построении сетей Байесовского вывода при анализе видеопотока.

5. Низкоуровневые визуальные дескрипторы

При распознавании поведения человека для СВН необходимо сначала вычислить низкоуровневые признаки видеопотока для последующего преобразования их в высокоуровневые и использования при построении сети Байесовского вывода. Для этой цели был выбран ряд низкоуровневых признаков, которые, используя методы компьютерного зрения, можно определить из видеоряда:

cvSpeed: текущая скорость объекта – вычисляется как перемещение ограничивающей рамки объекта, и выражается в пикс/сек (вычисляется для каждого кадра и нормализуется используя фреймрейт камеры);

cvFlow: трек объекта – недавняя история его движения

$$fl_t = \gamma\left(\left|cx_t - cx_{t-1}\right| + \left|cy_t - cy_{t-1}\right|\right) + (1-\gamma) fl_{t-1}, \qquad (1)$$

где $\gamma = 1 - e^{-1/3}$ и (cx_t, cy_t) – центр объекта в момент времени t.

cvLifetime: оставшееся время жизни показатель того, является ли объект новым или же скоро будет удален по причине того, что больше не детектируется на изображении;

cvOccstat: показатель перекрытия – оценка того, является ли объект перекрыт, неперекрыт или исчезнувший;

cvDistance: вектор нормализованных относительных расстояний $\widehat{D}_i = \{\widehat{D}_i^j\}$, i,j=1..N от каждого объекта i до каждого другого объекта j в кадре,

$$\forall i,j : D_i^j = \sqrt{\left(CV_x^i - CV_x^j\right)^2 + \left(CV_y^i - CV_y^j\right)^2}, \qquad (2)$$

где CV_x^i и CV_y^i – координаты объекта i в кадре;

CV_x^j и CV_y^j – координаты объекта j в кадре;

cvOrientation: вектор значений $R_i = \left\{R_i^j\right\}$, i,j=1..N, показывающих относительное направление между объектами i и j:

$$\forall i,j : R_i^j = \frac{180 - \left|CR_i - CR_j\right|}{180}, \qquad (3)$$

где CR_i и CR_j – углы направления векторов движения объектов i и j соответственно;

cvHistdist: межкадровая вариация формы объекта – взвешенная сумма EMD-измерения гистограммы и правдоподобности модели смеси Гауссиан.

Используя наборы аннотированных эталонных обучающих видеопоследовательностей, проводится обучение системы, для всех последовательностей вычисляются вышеперечисленные низкоуровневые признаки. Пространство признаков затем кластеризуется с использованием алгоритма k-средних, в результате чего из набора визуальных дескрипторов получаем набор состояний, характеризующих свойства динамики объектов в видеопотоке (например, "Активен", "Идет", "Бежит", "Стоит", "Сидит" и т.д.).

Таким образом, осуществляется переход от низкоуровневых признаков к высокоуровневым семантическим: состояниям объектов, ролям, ситуациям и сценариям.

Состояние наблюдаемого человека может зависеть напрямую от его текущей роли, его текущая роль — от ситуации, в которой он находится, а ситуация, в свою очередь, — от сценария, в котором он участвует. Выраженные в таком виде иерархические отношения используются как структурная основа для построения и обучение Байесовской сети.

6. Построение и тренировка Байесовской сети

Байесовская сеть представляет направленный ациклический граф G, матрица связности которого определяет условные отношения (не)зависимости между его узлами X, и, следовательно, определяет вид таблиц условных вероятностей [10].

Построение сети (определение структуры) требует средств для перебора пространства всех возможных направленных ациклических графов на наборе узлов X и функции оценки полученной структуры на обучающей выборке D.

В качестве метода определения (обучения) структуры сети был выбран алгоритм K2 [10] – алгоритм жадного поиска, который начинается с пустой сети, но с начальным упорядочиванием узлов. Байесовская сеть составляется итеративно, путём добавления направленных дуг к текущему узлу от того родительского узла, добавление которого максимально увеличивает оценку результирующей структуры графа. Процесс оканчивается в том случае, если ни одно из возможных добавлений не увеличивает оценку. В предлагаемом методе определения структуры сети начальное состояние сети учитывает состояние предметной области (рис. 1) на основе априорной структурной информации, содержащейся в полученной на предыдущем этапе онтологии с последующим доопределением на основе алгоритма K2.

Для вычисления оценки сети-кандидата на обучающей выборке, чтобы избежать переобучения, использовался Байесовский информационный критерий [11], который аппроксимирует маргинальное правдоподобие, используя подход дескриптора минимальной длины:

$$BIC(D,G) = \sum_i \left[\sum_{jk} N_{ijk} \log \theta_{ijk} - \frac{d_i}{2} \log M \right], \qquad (4)$$

где d_i – число параметров в таблице условных вероятностей, связанных с узлом X_i;

M – число образцов в обучающей выборке D;

θ_{ijk} – дискретные условные вероятности узлов графа;

N_{ijk} – число вхождений (X_i=k, $Pa(X_i)$=j) в данные;

$Pa(X_i)$ – родители узла X_i.

Построенная и обученная Байесовская сеть позволяет учитывать преимущества как вероятностного так семантического подходов при распознавании событий в видеопотоке. Такую сеть предложено называть семантико-вероятностной сетью. Она может быть предназначена для распознавания поведения людей в видеопотоке в реальном времени. Визуальные дескрипторы, вычисляемые из видео методами компьютерного зрения, выступают в качестве начальных переменных для сети. На основании обученных параметров сети вычисляется наиболее вероятные состояния объектов, ситуации, в которых они находятся и их роли в этих ситуациях, и, наконец, общий контекст происходящего (рис. 2).

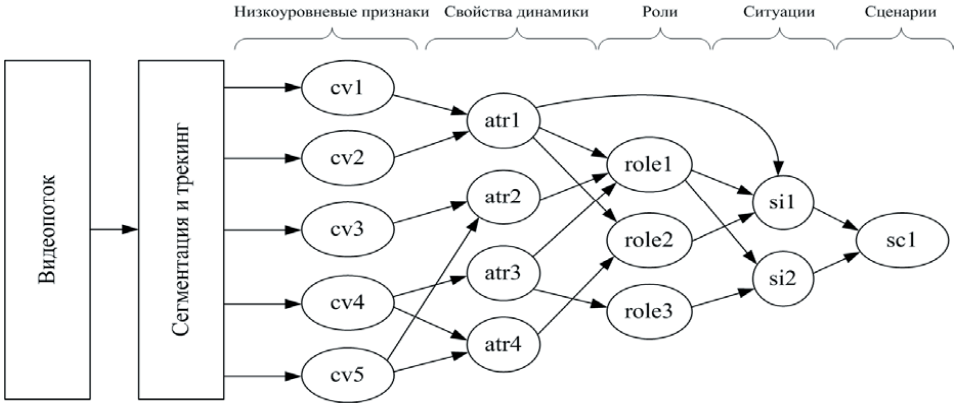

Рис. 2. Структура распознавания контекста в видеопотоке

В данной работе для тестирования разработанной системы использовались тестовые наборы видеопоследовательностей HMDB [12], представляющие собой нарезки из простых действий (бега, ходьбы и проч.), а также аннотированная обучающая выборка CAVIAR [13], содержащая набор сценариев и ситуаций с участием одного или нескольких человек, для построения онтологии предметной области и тестирования производительности системы.

7. Выводы

В данной работе было предложено создание семантико-вероятностной сети для определения контекста видеопотока, что позволило использовать преимущества семантических (дескрипторных) и графических вероятностных моделей для более эффективного распознавания поведения человека в системах интеллектуального видеонаблюдения.

Было установлено, что используя эталонные аннотированные обучающие видеопоследовательности, систему трекинга, а также формальное описание поведения человека на основе онтологии, учитывающее формализованные иерархические отношения, можно построить Байесовскую сеть для высокоуровневого интеллектуального анализа человеческого поведения и оценки взаимодействий между людьми в видеопотоке.

Тестирование разработанной семантико-вероятностной сети проводилось с использованием тестовой выборки CAVIAR.

Тестирование показало, что использование онтологии позволило улучшить достоверность распознавания человеческого поведения и взаимодействий между людьми в видеопотоке в среднем на 8%, по сравнению с традиционной Байесовской сетью, построенной без использования онтологий.

Литература

1. Aggarwal, J. Human activity analysis: A review [Текст] / J. Aggarwal, M. Ryoo // ACM Computing Surveys. – 2011. – Vol. 43, Issue 3. – C. 43.
2. Aggarwal, J. Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities [Текст] / J. Aggarwal, M. Ryoo // In IEEE International Conference on Computer Vision (ICCV). – 2009. – C. 1593–1600.
3. Niebles, J. C. Unsupervised learning of human action categories using spatial-temporal words [Текст] / J.C. Niebles, H. Wang, L. Fei-Fei // International Journal of Computer Vision (IJCV). – 2008. – Vol. 79, Issue 3. – C. 299–318.
4. Gupta, A. Objects in action: An approach for combining action understanding and object perception [Текст] / A. Gupta, L.S. Davis // In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2007. – C. 1–8.
5. Filipovych, R. Recognizing primitive interactions by exploring actor-object states [Текст] / R. Filipovych, E. Ribeiro // IEEE Conference on Computer Vision and Pattern Recognition. – 2008. – C. 1–7.
6. Damen, D. Recognizing linked events: Searching the space of feasible explanations [Текст] / D. Damen, D. Hogg // IEEE Conference on Computer Vision and Pattern Recognition. – 2009. – C. 927–934.
7. Joo, S.-W. Attribute Grammar-Based Event Recognition and Anomaly Detection [Текст] / S.-W. Joo, R. Chelappa // Conference on Computer Vision and Pattern Recognition Workshop. – 2006. – C. 107–107.
8. Gupta, A. Understanding videos, constructing plots learning a visually grounded storyline model from annotated videos [Текст] / A. Gupta, P. Srinivasan, J. Shi, L.S. Davis // In IEEE Conference on Computer Vision and Pattern Recognition. – 2009. – C. 2012–2019.
9. Malgireddy, M.R. A temporal Bayesian model for classifying, detecting and localizing activities in video sequences [Текст] / M.R. Malgireddy, I. Inwogu, V. Govindaraju // Computer Vision and Pattern Recognition Workshops. – 2012. – C. 43–48.
10. Cooper, G. A Bayesian method for the induction of probabilistic networks from data [Текст] / G. Cooper, E. Herskovits // Machine Learning. – 1992. – Vol. 9. – C. 309–347.
11. Murphy, K. Dynamic Bayesian Networks: Representation, Inference and Learning [Текст] / K. Murphy // PhD thesis, University of California at Berkley. – 2002.
12. HMDB: A Large Video Database for Human Motion Recognition [электронный ресурс] / SERRE LAB. A Brown University Research Group. – Режим доступа : \www/ URL: http://serre-lab.clps.brown.edu/resources/HMDB/ – 2011. – Загл. с экрана.
13. CAVIAR Test Case Scenarios [Электронный ресурс] / INRIA Labs at Grenoble, France. – Режим доступа : \www/ URL: http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/ – 2004. – Загл. с экрана.