

УДК 004 932.72'1

С. Г. Антощук, д-р техн. наук,
Н. А. Годовиченко

ОЦЕНКА ЭФФЕКТИВНОСТИ МОДЕЛЕЙ СОБЫТИЙ В ЗАДАЧАХ СЕМАНТИЧЕСКОГО АНАЛИЗА ВИДЕОПОТОКА

Аннотация. Представлена система, которая позволяет оценить эффективность моделей событий. Была предложена структура системы, выбраны показатели оценки эффективности работы модели, а также предложено измерять временную сложность модели с фильтром частиц. Были представлены используемые наборы видеоданных, проведено экспериментальное тестирование разработанной системы.

Ключевые слова: модель событий, метрика эффективности, ROC-кривая, ETISEO, CAVIAR, BEHAVE, фильтр частиц, сеть Петри, семантический анализ видеопотока

S.G. Antoshchuk, ScD.,
N.A. Godovichenko

EVENT MODELS PERFORMANCE EVALUATION IN THE VIDEOSTREAM SEMANTIC ANALYSIS

Abstract. The system, which allows to evaluate event models performance was presented. The structure of the system, indicators for evaluating the models performance were proposed and measuring time complexity of model with particle filter was proposed. Videodata were presented. Experimental testing was held.

Keywords: event model, performance metric, ROC-curve, ETISEO, CAVIAR, BEHAVE, particle filter, Petri nets, semantic analysis of video

С. Г. Антощук, д-р техн. наук,
М. А. Годовиченко

ОЦІНКА ЕФЕКТИВНОСТІ МОДЕЛЕЙ ПОДІЙ В ЗАДАЧАХ СЕМАНТИЧНОГО АНАЛІЗУ ВІДЕОПОТОКУ

Анотація. Представлено систему, яка дозволяє оцінити ефективність моделей подій. Було запропоновано структуру системи, обрані показники ефективності роботи моделі, а також запропоновано вимірювати часову складність моделі з фільтром часток. Були представлені набори відеоданих, що використовуються, проведено експериментальне тестування розробленої системи.

Ключові слова: модель подій, метрика ефективності, ROC-крива, ETISEO, CAVIAR, BEHAVE, фільтр частинок, мережа Петрі, семантичний аналіз відеопотоку

Введение. Важным направлением в развитии систем искусственного интеллекта является разработка систем семантического анализа видеопотока, предназначенных для извлечения и распознавания значимых событий и для принятия организационных решений, управления и др. Подобная задача возникает при создании охранных систем нового поколения, систем интеллектуального видеонаблюдения, робототехнических систем и т.д.

Будем считать, что видеопоток представляет собой визуальное отображение последовательности сцен:

$$S_i \Rightarrow FS \Rightarrow \dots \Rightarrow S_{i+k},$$

где S_i , FS , S_{i+k} – некоторые сцены, представленные в видеопотоке, каждая из кото-

рых представляет набор некоторых подсобытий; i – номер кадра; k – некоторое количество кадров.

Семантически значимое событие (факт, инцидент, ситуация) FS на видео представляет собой высокоуровневый семантический концепт, составленный из набора некоторых подсобытий. Семантически значимое событие воспринимается и выделяется как значимое для наблюдателя или системы семантического анализа видеопотока с точки зрения осуществления ими дальнейших действий.

Разработка систем распознавания событий связана с решением двух задач [7]:

1) задача извлечения данных из видеопоследовательности – выделение из видеопотока семантически значимых данных;

2) задача представления и распознавания значимых событий – семантически зна-

© Антощук С.Г., Годовиченко Н.А., 2014

чимых данных с учетом выбранной модели событий.

Выбор и разработка модели событий является определяющим этапом для решения задачи распознавания сцен в целом [5, 6, 9]. Подходящая модель должна иметь достаточную выразительную мощность для моделирования сложных иерархических отношений, обладать механизмом коррекции ошибок, который позволяет модели успешно работать в условиях неопределенности данных, которые возникают на этапе извлечения данных, а также быть реализуемой на практике.

Поскольку модель распознавания событий определяет успешность работы системы семантического анализа видеопотока, то актуальным является вопрос выбора подходящей модели, который должен базироваться на оценке ее эффективности и возможности сравнения с существующими аналогами. Решение этого вопроса является нетривиальной задачей, т.к. не существует программных продуктов, предназначенных для моделирования событий и оценки эффективности моделей.

Целью данной работы является разработка системы оценки эффективности моделей событий с целью выбора надлежащей модели событий для ее последующего использования в системах семантического анализа видеопотока.

Система оценки эффективности модели событий. Авторами предложена структура программной системы для оценки эффективности модели событий, которая состоит из нескольких компонентов (рис. 1). Необработанная видеопоследовательность (*Вход 1*) подается на вход модуля детектирования и трекинга объектов (*Модуль 1*), который генерирует список объектов и их местоположение на каждом кадре

Результаты работы алгоритмов трекинга подаются на вход модуля распознавания подсобытий (*Модуль 2*). Также, на вход этого модуля подается спецификация сцены (*Вход 2*).

Распознавание подсобытий происходит согласно простым правилам, представленным в работе [12]. На выходе данного модуля формируются распознанные подсобытия с указанием достоверности распознавания.

Распознанные подсобытия, а также модель событий (*Вход 3*) подаются на вход модуля распознавания событий (*Модуль 3*).

Результаты работы модуля распознавания событий, а также размеченные данные (*Вход 4*) передаются на вход модуля оценки эффективности модели событий (*Модуль 4*). Результатом работы данного модуля являются количественные значения эффективности модели событий с учетом используемых показателей.

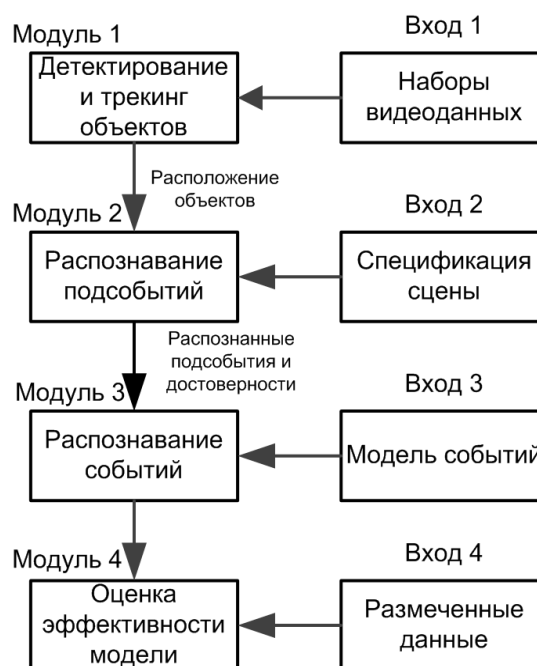


Рис. 1. Структура системы оценки эффективности модели событий

Наборы видеоданных. Для оценки эффективности модели событий подобраны наборы входных видеопоследовательностей, которые использованы в качестве тестовых данных.

К используемым видеоданным предъявляются следующие требования: в них должны присутствовать сложные высокоуровневые события, одно и то же событие должно быть представлено в как можно большем количестве роликов. Еще одним требованием является наличие размеченных данных, которые используются для оценки эффективности работы моделей.

В качестве используемых наборов видеороликов, использовались наборы видеоданных ETISEO, известный набор видеоданных CAVIAR, а также набор BEHAVE.

Набор видеоданных ETISEO (рис. 2) состоит из 86 видеороликов, которые представляют различные ситуации, возникающие при осуществлении видеонаблюдения. Ролики сочетают в себе сцены внутри и вне помещений, съемки в коридоре, на улице, в метро.

Данный набор обладает большим объемом размеченных данных, которые включают в себя информацию о наличии объекта, его местоположении, классе и траектории движения.

Набор видеоданных CAVIAR (рис. 3) является одним из самых часто используемых наборов при разработке систем видеонаблюдения. Набор насчитывает несколько сотен видеороликов, которые можно использовать для распознавания событий, для тестирования алгоритмов трекинга, распознавания групповых действий и т.д. Также набор сопровождается большим количеством размеченных данных.



Рис. 2. Кадр из набора ETISEO с размеченными данными

Набор ВЕНАВЕ (рис. 4) содержит видеоданные, связанные в основном с взаимодействиями между людьми. Набор насчитывает десять основных событий: люди находятся в группе, люди приближаются друг к другу, люди идут вместе, люди разделяются, один человек игнорирует другого, один человек следует за другим, один человек преследует другого, два человека дерутся, два человека бегут вместе, два человека встречаются. Набор сопровождается размеченными данными.



Рис. 3. Кадр из набора CAVIAR с размеченными данными

Кроме наборов реальных видеороликов, также был разработан набор синтетических видеороликов, которые представляют собой анимационные ролики, выполненные с помощью технологии Adobe Flash, которые представляют некоторые события (относительное перемещение простых геометрических фигур).

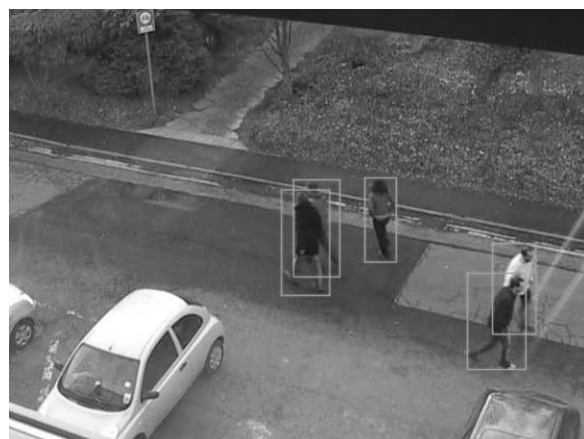


Рис. 4. Кадр из набора ВЕНАВЕ с размеченными данными

Детектирование и трекинг объектов.

Модуль решает задачу извлечения данных из видеопоследовательности. Данные извлекаются в виде списка распознанных объектов, с указанием их траектории движения, вектора движения, а также физических параметров. В качестве алгоритмов трекинга использовались размеченные данные (для моделирования ситуации «идеального» трекинга), алгоритм трекинга основанный на фильтре частиц [8] и алгоритм, основанный на адаптированном варианте алгоритма усиления классификаторов AdaBoost [4]. В данной си-

стеме используется полуавтоматический подход к задаче трекинга: если трек объекта был потерян, тогда он заново реинициализируется с помощью размеченных данных.

Спецификация сцены. Спецификация сцены представляет собой данные о каждом видеоролике в виде координат зон интереса, данных о калибровке камеры и другой информацией, которая используется для распознавания подсобытий.

Распознавание подсобытий. Данный модуль осуществляет распознавание простых подсобытий, каждое из которых связано с определенным объектом, например: *появление*, *исчезновение*, *объект_в_зоне*, *объект_движется*, *объект_остановился*, *два_объекта_рядом* и т.д. На выходе данного модуля формируются распознанные подсобытия с указанием достоверности распознавания.

Модель событий. Для использования модели событий в предложенной системе оценки эффективности модели событий необходимо представить ее в определенном программном формате, для дальнейшего использования в модуле распознавания событий. Авторами была произведена программно-алгоритмическая реализация трех моделей событий. Выбор моделей обусловлен наличием полной информации о модели и возможности ее реализации программными средствами.

Модель *The Store Totally Recognized Scenarios* (STRS) является примером детерминистического подхода к задаче распознавания событий [12]. В данном подходе, информация полученная на этапе извлечения данных (*Модуль 1*) обрабатывалась пороговым методом: на вход модели подавались данные выше определенного порога.

Модель *Propagation Net* (P-Net) представлена в работе [10] и является примером вероятностного подхода к решению задачи распознавания событий. Данная модель является простой надстройкой над динамической байесовской сетью.

В качестве третьей модели событий используется предложенная авторами комбинированная модель на основе детерминистической модели на основе сети Петри и вероятностной модели на основе байесов-

ского рекурсивного фильтра, обладающей механизмом коррекции ошибок [1–3].

Распознавание событий. Модуль распознавания событий осуществляет непосредственно распознавание. В качестве входных данных используется информация о распознанных подсобытиях, с указанием достоверности распознавания, а также информация об объектах, полученная на этапе детектирования и трекинга объектов. На выходе данного модуля формируется xml-файл с указанием распознанных событий с параметрами.

Размеченные данные. Размеченные данные представляют собой составленную вручную информацию о событиях в видеоролике с указанием параметров. Эти данные считаются эталонными, и используются для сравнения с результатами работы модели.

Выбор показателей эффективности модели событий. Для сравнительной оценки эффективности моделей, необходимо выбрать метрику, которая бы позволяла одинаково оценивать эффективность различных моделей.

Для оценки использовалось сравнение распознанных событий с размеченными данными (в данном случае размеченные данные представляют собой распознанные человеком события). Это позволяет оценить количество истинно-положительных, истинно-отрицательных, ложно-положительных и ложно-отрицательных оценок [11].

В силу того, что на используемых видеороликах происходило относительно небольшое количество событий, количество положительных оценок гораздо меньше, чем количество отрицательных оценок. Это справедливо и для реальных систем распознавания событий. По этой причине, модель которая давала бы постоянно отрицательные оценки, формально имела бы хорошую эффективность. Исходя из этого, оценка эффективности производилась только с учетом уровня истинно-положительных и ложно-положительных оценок.

Результаты оценки эффективности моделей событий представляются в виде кривой ошибок (ROC-кривая).

Следует отметить, что в разработанной авторами комбинированной модели эффективность работы зависит от количества частиц в рекурсивном фильтре для представления и оценки апостериорного распределения. Поэтому как еще один важный показатель комбинированной модели в работе оценивалась эффективность модели как функция от ее временной сложности, напрямую зависящей от количества используемых частиц.

Количество частиц должно увеличиваться с увеличением сложности пространства состояний. Таким образом, уменьшение размерности пространства состояний означает уменьшение временной сложности и увеличение временной эффективности модели.

Чтобы измерить улучшение эффективности, предлагается использовать показатель временной сложности. Временная сложность представляет собой зависимость успешности распознавания (которая измеряется с помощью F-меры) в зависимости от количества используемых частиц [11]. Если успешность распознавания остается неизменной при уменьшении количества частиц, то можно предполагать что модель хорошо масштабируется при увеличении размерности пространства состояний.

Также данная оценка позволяет подобрать оптимальное количество используемых частиц.

Экспериментальное тестирование моделей распознавания событий. Проверка работоспособности и тестирование разработанной системы оценки моделей событий проводилась с использованием набора видеороликов «Вход в здание», которая взята из набора данных ETISEO.

Проверка проводилась с использованием моделей, указанных выше.

Тестирование проводилось в двух режимах: в режиме «идеального» трекинга (рис. 5), когда в качестве результатов работы алгоритмов трекинга использовались размеченные данные, а также в режиме «реального» трекинга (рис. 6), при котором данные извлекались с помощью алгоритмов трекинга.

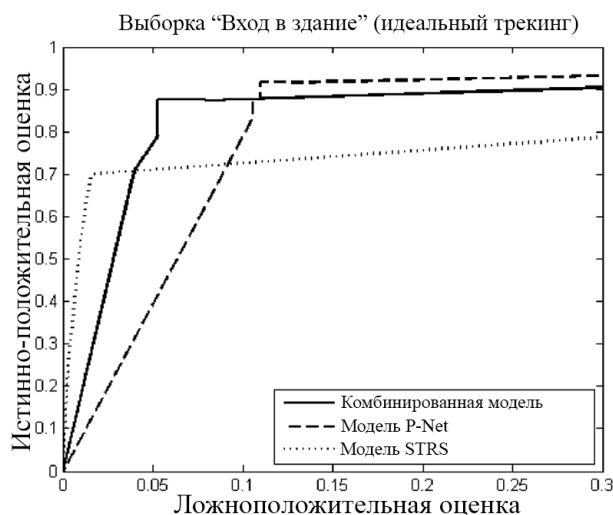


Рис. 5. Результаты работы в режиме «идеального» трекинга

Результаты тестирования представлены в виде кривой ошибок (ROC-кривой).

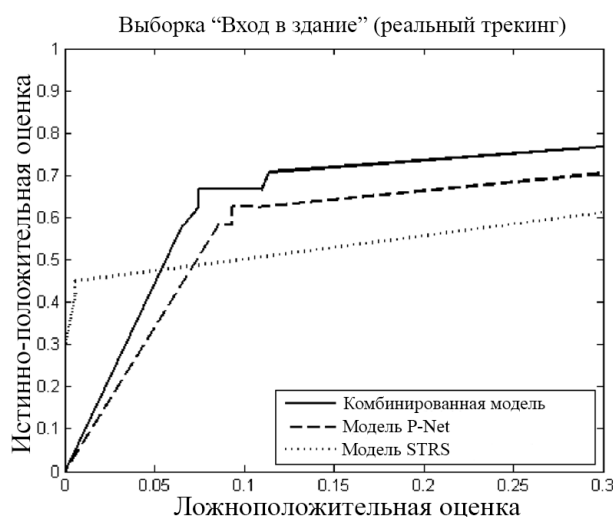


Рис. 6. Результаты работы в режиме «реального» трекинга

Количественную интерпретацию полученных результатов можно получить путем вычисления площади под ROC-кривой (чем выше показатель, тем выше эффективность модели). Количественная оценка представлена в табл. 1.

Результаты экспериментального тестирования показали, что в режиме «идеального» трекинга сеть Петри и предложенная модель показывают похожие результаты. Однако, в режиме «реального» трекинга, эффективность сети Петри снижается.

1. Количественная интерпретация полученных ROC-кривых

Модель	Идеальный трекинг	Реальный трекинг
Комбинированная модель	0,91021	0,81021
Модель P-Net	0,88542	0,75326
Модель STRS	0,84241	0,72190

Для комбинированной модели также была проведена оценка временной эффективности модели (рис. 7).

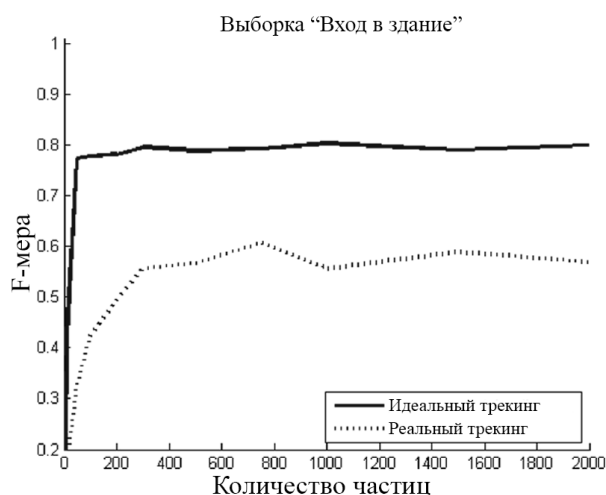


Рис. 7. Оценка временной эффективности комбинированной модели

Как можно видеть из приведенных зависимостей, успешность распознавания события при реальном трекинге остается неизменной, когда количество частиц выше 250. Следовательно, такое количество частиц является достаточным, модель хорошо масштабируется.

Выводы. В данной работе представлена система оценки эффективности работы моделей событий для системы семантического анализа видеопотока. Предложена структура системы, которая состоит из четырех модулей, в качестве входных данных принимает видеопоследовательность, спецификацию сцены, модель событий, а также размеченные данные. Подобраны наборы видеоданных, которые позволяют обеспечить систему входными видеороликами и размеченными данными. Предложены показатели оценки эффективности моделей событий с учетом особенности задачи распознавания событий, для комбинированной модели с фильтром частиц предложено оценивать оптимальное

количество частиц для оптимизации временной сложности модели. Проведено экспериментальное тестирование системы оценки эффективности моделей событий, получены количественные показатели эффективности моделей событий.

Основные модули разработанной системы оценки эффективности работы моделей событий могут быть использованы для систем семантического анализа видеопотока. Результаты статьи будут интересны специалистам, занимающимся разработкой разных прикладных приложений систем искусственного интеллекта на основе интеллектуального анализа динамической видеоинформации, таких как охранные системы, системы обеспечения правопорядка, системы управления дорожным движением, системы интеллектуального видеонаблюдения, робототехнические системы и др.

Список использованной литературы

1. Антощук С. Г. Модели представления событий при анализе видеопотока / С. Г. Антощук, Н. А. Годовиченко // *Электротехнические и компьютерные системы*. – К. : – 2013. – №11. – С. 142 – 149.
2. Антощук С. Г. Моделирование событий в видеопотоке с помощью стохастических сетей Петри / С. Г. Антощук, Н. А. Годовиченко // *Оптико-електронні інформаційно-енергетичні технології*. – Винница : – 2013. – № 1. – С. 5 – 12.
3. Антощук С. Г. Учет неопределенности данных при моделировании событий с помощью сети Петри / С. Г. Антощук, Н. А. Годовиченко // *Электротехнические и компьютерные системы*. – К. : – 2013. – № 12. – С. 138 – 146.
4. Babenko B., Yang M., and Belongie S. Visual Tracking with Online Multiple Instance Learning, (2011), *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 102 – 110, doi: 10.1.1.192.9575, url: http://vision.ucsd.edu/~bbabenko/data/miltrack_cvpr09.pdf.
5. Cohn A. G., Magee D. R., Galata A., Hogg D., and Hazarika S. M. Towards an Architecture for Cognitive Vision Using Spatial-temporal Representations and Abduction, (2003), *In Spatial Cognition*, No. 2, pp. 232 – 248, doi: 10.1.1.15.1821.

6. Hongeng S., and Nevatia R. Multi-agent Event Recognition, (2001), *International Conference on Computer Vision*, pp. 84 – 93, doi:10.1109/ICCV.2001.937608, url: <http://www.jianmeauto.com/uploadfile/file/2013020222243062401.pdf>.

7. Hu W., Tan T., Wang L., and Maybank S. A Survey on Visual Surveillance of Object Motion and Behaviors, (2004), *Systems, Man and Cybernetics*, Part C., No. 4, pp. 334 – 352, doi: 10.1109/TSMCC.2004.829274, url: <http://vc.cs.nthu.edu.tw/home/paper/codfiles/yzchen/200812041421/A%20Survey%20on%20Visual%20Surveillance%20of%20Object%20Motion%20and%20Behaviors.pdf>.

8. Isard M., and Blake A. Condensation – Conditional Density Propagation for Visual Tracking, (1998), *International Journal of Computer Vision*, pp. 5 – 28, doi: 10.1.1.36.8357, url: http://www.stat.ucla.edu/~sczhu/courses/ucla/Stat_202C/lecture_note/Particle_filtering_Isard_1998.pdf

9. Medioni G.G., Cohen I., Bremond F., Hongeng S., and Nevatia R. Event Detection and Analysis from Video Streams, (2001), *IEEE Transactions on Pattern Analysis and Machine Intelligence*, No. 8, pp. 873 – 889, doi:10.1109/34.946990, url: <http://iris.usc.edu/outlines/papers/1998/iuwevent98.pdf>.

10. Shi Y., Huang Y., Minnen D., Bobick A., and Essa I. Propagation Networks for Recognition of Partially Ordered Sequential Action, (2004), *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 862 – 869, doi: 10.1.1.92.4938, url: http://www.researchgate.net/publication/221364362_Propagation_Networks_for_Recognition_of_Partially_Ordered_Sequential_Action/file/9c960514cab7b204cc.pdf

11. Van Rijsbergen C. J. Information Retrieval, *Oxford*, (1979), Butterworth-Heinemann, 224 p.

12. Vu V., Bremond F., and Thonnat M. Automatic Video Interpretation: a Novel Algorithm for Temporal Scenario Recognition, (2003), *International Joint Conference on Artificial Intelligence*, No. 1, pp. 1295 – 1300, url: <http://www.sop.inria.fr/orion/Publications/Articles/ijcai2003.pdf>.

Получено 26.02.2014

References

1. Antoshchuk S.G., and Godovychenko N.A. Modeli predstavleniya sobytiiy pri analize videopotoka [Events Representation Models for Videostream Analysis], (2013), *Elektrotechnic and Computer Systems Publ.*, Kiev, Ukraine, No. 11, pp. 142 – 149 (In Russian).

2. Antoshchuk S.G., and Godovychenko N.A. Modelirovanie sobytiiy v videopotoke s pomoschyu stohasticheskikh setey Petri [Modeling Events in Videostream With Stochastic Petri Nets], (2013), *Optoelectronic Information-energy Technologies Publ.*, Vinnitsa, Ukraine, No. 1, pp. 5 – 12 (In Russian).

3. Antoshchuk S.G., and Godovychenko N.A. Uchet neopredelennosti dannyh pri modeli-rovanii sobytij s pomoshh'ju seti Petri [Accounting Data Uncertainty in Modeling Events With Petri Nets], (2013), *Elektrotechnic and Computer Systems Publ.*, Kiev, Ukraine, No. 12 pp. 138 – 146 (In Russian).

4. Babenko B., Yang M. and Belongie S. Visual Tracking with Online Multiple Instance Learning, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2011), pp. 102 – 110 (In English), doi: 10.1.1.192.9575, url: http://vision.ucsd.edu/~bbabenko/data/miltrack_cvpr09.pdf

5. Cohn A. G., Magee D. R., Galata A., and others. Towards an Architecture for Cognitive Vision Using Spatial-temporal Representations and Abduction, (2003), *In Spatial Cognition*, No. 2, pp. 232 – 248 (In English), doi: 10.1.1.15.1821.

6. Hongeng S., and Nevatia R. Multi-agent Event Recognition, (2001), *International Conference on Computer Vision*, pp. 84 – 93 (In English), doi:10.1109/ICCV.2001.937608, url: <http://www.jianmeauto.com/uploadfile/file/2013020222243062401.pdf>.

7. Hu W., Tan T., Wang L., and others. A Survey on Visual Surveillance of Object Motion and Behaviors, *Systems, Man and Cybernetics*, Part C., (2004), No. 4, pp. 334 – 352 (In English), doi: 10.1109/TSMCC.2004.829274, url: <http://vc.cs.nthu.edu.tw/home/paper/codfiles/yzchen/200812041421/A%20Survey%20on%20V>

isual%20Surveillance%20of%20Object%20Motion%20and%20Behaviors.pdf.

8. Isard M., and Blake A. Condensation – Conditional Density Propagation for Visual Tracking, (1998), *International Journal of Computer Vision Publ.*, pp. 5 – 28 (In English), doi: 10.1.1.36.8357

url: http://www.stat.ucla.edu/~sczhu/courses/ucla/Stat_202C/lecture_note/Particle_filtering_Isard_1998.pdf

9. Medioni G.G., Cohen I., Bremond F., and others. Event detection and analysis from video streams, (2001), *IEEE Transactions on Pattern Analysis and Machine Intelligence*, No. 8, pp. 873 – 889 (In English), doi:10.1109/34.946990,
url:<http://iris.usc.edu/outlines/papers/1998/iuwent98.pdf>.

10. Shi Y., Huang Y., Minnen D., and others. Propagation Networks for Recognition of Partially Ordered Sequential Action, (2004), *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 862 – 869 (In English), doi: 10.1.1.92.4938,
url: http://www.researchgate.net/publication/21364362_Propagation_Networks_for_Recognition_of_Partially_Ordered_Sequential_Action/file/9c960514cab7b204cc.pdf.

11. Van Rijsbergen C. J. Information Retrieval, Oxford, (1979), Butterworth-Heinemann, 224 P.

12. Vu V., Bremond F., and Thonnat M. Automatic Video Interpretation: a Novel Algorithm for Temporal Scenario Recognition, (2003), *International Joint Conference on Artificial Intelligence*, No. 1, pp. 1295 – 1300 (In English),
url:<http://www.sop.inria.fr/orion/Publications/Articles/ijcai2003.pdf>.



Антощук Светлана Григорьевна,
д-р техн. наук, проф.,
директор ин-та компьютерных систем, зав. каф. информационных систем Одесского нац. политехн. ун-та,
тел.: +38(048)7348-584
E-mail: asg@ics.opu.ua



Годовиченко Николай Анатольевич,
аспирант Одесского нац. политехн. ун-та,
м/т: +3(068)2619923