

УДК: 004.931; 004.932

Симметрии дифференциальных уравнений в задачах компьютерного зрения

Д. А. Вражнов^{1,а}, А. В. Шаповалов^{2,3}, В. В. Николаев²

¹ ООО «Томсклаб», Россия, 634055, Томск, проспект Академический, 8/8

² Томский государственный университет, Россия, 634050, Томск, проспект Ленина, 36

³ Томский политехнический университет, Россия, 634050, Томск, проспект Ленина, 30

E-mail: ^а vrazhnov@tomsklabs.com

Получено 25 марта 2010 г.

В данной работе приводится обобщение подхода к построению инвариантных векторов признаков изображений в задачах распознавания образов. Базовым элементом предлагаемого алгоритма является замена обычно применяемого гауссова фильтра исходного изображения сверткой функции изображения с функцией Грина эволюционного оператора, наследующей свойства симметрий этого оператора. Применение обобщенной фильтрации позволяет выделять дополнительные характеристики инвариантных векторов признаков.

Ключевые слова: компьютерное зрение, распознавание образов, фильтрация, симметрии дифференциальных уравнений, функция Грина

Symmetries of differential equations in computer vision applications.

D. A. Vrazhnov¹, A. V. Shapovalov^{2,3}, V. V. Nikolaev²

¹ Tomsklabs LLC, Akademicheskii av., 8/8, 634055, Tomsk, Russia

² Tomsk State University, Lenin av., 36, 634050, Tomsk, Russia

³ Tomsk Polytechnic University, Lenin av., 30, 634050, Tomsk, Russia

Abstract. – In our work we present generalization of well-known approach for construction of invariant feature vectors of images in computer vision applications. Basic feature of the suggested algorithm is replacement of commonly used Gaussian filter by convolution of image function with Green's function of evolution operator, which inherits symmetries of this operator. The use of general filtration allows to obtain additional characteristics of invariant feature vectors.

Keywords: computer vision, pattern recognition, filtering, symmetries of differential equations, Green Function

Citation: *Computer Research and Modeling*, 2010, vol. 2, no. 4, pp. 369–376 (Russian).

Работа выполнена при частичной финансовой поддержке АВЦП Министерства образования и науки РФ № 2.1.1/3436; ФЦП «Научно-педагогические кадры инновационной России» на 2009-2012 годы.

© 2010 Денис Александрович Вражнов, Александр Васильевич Шаповалов, Виктор Владимирович Николаев

Введение

Машинное зрение является совокупностью подходов и методов обработки видеoinформации в приложениях к робототехнике, биофизике, медицине и других. Решения большинства задач в данной области основаны на применении теории распознавания образов [Гонсалес и Вудз, 2006; Форсайт и Понс, 2004]. Базовой схемой является построение множеств векторов признаков на основе анализа особенностей исходного изображения и их дальнейшая классификация с целью обнаружения объектов и их взаимосвязей. Под объектами, как правило, понимается объединение векторов признаков в классы [Шапиро и Стокман, 2006].

Эффективность использования данной схемы существенно зависит от методов выделения особенностей изображения и построения векторов признаков. Одной из проблем обработки видеoinформации является большая вариативность параметров съемки, таких, например, как: освещенность, местоположение камеры, альбеда объекта, применяемая цветовая схема. Это приводит к тому, что один и тот же объект, в зависимости от условий съемки будет характеризоваться различными векторами признаками. Общеизвестным подходом к решению этой проблемы является применение инвариантных векторов признаков. В настоящее время предложены методы построения векторов признаков, инвариантных относительно поворотов, масштабных преобразований и положения объекта, а также с частичной инвариантностью к изменениям освещенности [Mikolajczyk, 2004]. Значительный вклад в развитие теории построения инвариантных векторов признаков внесли работы Линдберга по конструированию пространства масштабов [Lindeberg, 1998]. Главной идеей предложенного им подхода является свертка функции яркости изображения с гауссианом. Дальнейшее развитие этой идеи привело к пониманию, что гауссиан можно рассматривать как функцию Грина для решения линейного уравнения диффузии, в котором эволюционная переменная связана с дисперсией функции Грина.

Заметим, что кроме выделения особенностей (ключевых точек) с заданными свойствами, фильтрация при помощи функций Грина применяется для решения задач, связанных в вычислением оптического потока [Obinata G. and Dutta A., 2007].

Построение векторов признаков изображения

Математическое представление некоторого объекта Ω подразумевает существование набора $\{D_1, \dots, D_N\}$ операторов (детекторов, измерений), которые ставят в соответствие объекту вещественный вектор (вектор признаков) $X = (d_1, \dots, d_N)$, где $d_i = D_i(\Omega)$. В данной работе используются векторы признаков, построенные по методу Bag-of-features (BoF). Подход Bag-of-features представляет собой простой, но эффективный способ визуальной классификации объектов по категориям [Csurka, 2004; Civic, 2005].

Краткое описание алгоритма BoF

1. На первом этапе из исследуемого набора изображений извлекаются особенности и строятся дескрипторы, содержащие координаты характерных точек изображений и параметры малых окрестностей этих точек, используя, например, алгоритм Scale Invariant Feature Transform (SIFT) [Lowe, 2004], описанный ниже. Назовем полученные дескрипторы «буквами».
2. Далее, кластеризуем множество дескрипторов на заданное количество классов. Назовем это разбиение «алфавитом».
3. Для каждого изображения строим гистограмму распределения «букв» из «алфавита». Эта гистограмма и будет итоговым вектором признаков изображения.
4. Обучаем (подбираем параметры) классификатор, например, Support Vector Machine (SVM) [Steinwart I. and Christmann A. 2008]. Он разделяет подаваемые изображения на определенные пользователем категории, соответствующие типам объектов, например, «велосипеды», «машины», «лица».

Рассмотрим применение BoF в задачах распознавания образов. Первые три этапа алгоритма BoF служат для создания «словаря» обучающей базы изображений. Для тестирования произвольного изображения добавим следующие пункты к вышеизложенному алгоритму.

5. Применяем первые три этапа к тестовому изображению и находим его вектор признаков.
6. С помощью обученного классификатора определяем принадлежность тестового изображения к известному классу (объекту).

Основная проблема метода BoF – отсутствие геометрической информации в формируемых векторах признаков. Один из способов учета геометрических свойств дает метод Spatial pyramid match kernel [Labeznik S., Shmid C., Ponce J., 2006]. Суть метода – в последовательности разбиений изображения равномерной сеткой с уменьшающимся шагом. Затем строятся гистограммы векторов–признаков, полученных при помощи SIFT, с весами, зависящими от геометрической связанности этих векторов.

Заметим, что дополнение первичных векторов признаков геометрическими связями в общем случае нарушает свойства инвариантности полученных векторов относительно поворота и местоположения объекта на кадре. Ниже приводится алгоритм SIFT и внесенные нами изменения.

Алгоритм SIFT

1. Масштабно-пространственный детектор: первая стадия вычисления ищет ключевые точки по всем масштабам изображения и пространству.
2. Фильтрация обнаруженных ключевых точек для выделения стабильных точек.
3. Назначение одной или нескольких ориентаций для каждой ключевой точки.
4. Вычисление дескриптора каждой ключевой точки

Алгоритм обнаружения ключевых точек в пространстве масштабов

Рассмотрим произвольный непрерывный сигнал $f: R^D \rightarrow R$. Назовем представлением L сигнала f свертку $L(\cdot; t) = G(\cdot; t) * f(\cdot)$ функции f с функцией Грина $G(\cdot; t)$ линейного уравнения диффузии

$$\partial_t U = \frac{1}{2} \nabla^2 U = \frac{1}{2} \sum_{i=1}^D \partial_{x_i x_i} U.$$

Функция Грина имеет следующий вид:

$$G: R^D \times R_+ \rightarrow R, \quad x = (x_1, \dots, x_D)^T, \quad G(x; t) = \frac{1}{(2\pi t)^{N/2}} e^{-(x_1^2 + \dots + x_D^2)/(2t)}. \quad (1)$$

Представление L сигнала f (с начальным условием $L(\cdot; 0) = f(\cdot)$) принадлежит линейному пространству решений уравнения диффузии, называемому пространством масштабов, поскольку эволюционный параметр t в уравнении диффузии и, следовательно, в функции Грина имеет смысл параметра масштаба. Далее определим ключевую точку или особенность изображения, как точку (x_j, y_k) изображения, в которой представление функции L изображения f в пространство масштабов имеет локальный экстремум. Из необходимых условий экстремума свертки получаем:

$$\begin{aligned} \frac{\partial}{\partial t} (G(\cdot; t) * f(\cdot)) &= \frac{\partial}{\partial t} (G(\cdot; t)) * f(\cdot) = 0, \\ \frac{\partial}{\partial x_i} (G(\cdot; t) * f(\cdot)) &= \frac{\partial}{\partial x_i} (G(\cdot; t)) * f(\cdot) = 0. \end{aligned}$$

То есть задача поиска ключевых точек в пространстве масштабов сводится к решению экстремальной задачи $\frac{\partial}{\partial t}(G(\cdot; t)) * f(\cdot) = \frac{\partial}{\partial x_i}(G(\cdot; t)) * f(\cdot) = 0$. Данная задача имеет конечно-разностное представление и допускает эффективное численное решение.

В работах [Koenderink, 1984; Florack et al., 1994] показано, что в классе линейных преобразований гауссиан является уникальной функцией для создания пространства масштабов (Lindeberg, 1998). Главными требованиями при этом являются линейность и инвариантность представлений изображений относительно сдвига. Однако следует отметить, что линейность пространства представлений следует из структуры свертки, а инвариантность относительно сдвига есть следствие соответствующей инвариантности уравнения диффузии, поскольку функция Грина есть инвариантно решение уравнения диффузии [Ибрагимов, 1983]. С ростом масштабного параметра t , степень сглаженности представления L изображения f увеличивается, что приводит к уменьшению значений пространственных производных функции L . Для компенсации сглаживания вводится специальная нормализация пространственных производных [Lindeberg, 1998].

Численная реализация описанного выше алгоритма основана на аппроксимации производной слоев по масштабу их конечной разностью и называется Difference of Gaussians (DoG). Слои группируются в октавы, соответствующие удвоенному значению параметра t .

Как только DoG слоев были получены, ключевые точки идентифицируются как локальные минимумы или максимумы DoG слоев. Это выполняется путем сравнения каждого пикселя в слое с его восемью соседними пикселями того же слоя и девятью пикселями на каждом из соседних слоев. Если выбранный пиксель максимален или минимален по сравнению со своими соседями, то он выбирается как кандидат, который может стать ключевой точкой на изображении.

Рассмотренный масштабно-пространственный детектор выделяет слишком много кандидатов, некоторые из которых являются неустойчивыми. Следующий шаг заключается в вычислении интерполированного положения ключевой точки. Интерполирование основано на разложении в ряд Тейлора до второго порядка включительно DoG представления слоев с точкой, подозрительной на экстремум в качестве начальной. Эта процедура позволяет исключить точки, обладающие низким контрастом (и поэтому чувствительны к шуму).

Этот детектор суть один из методов, которые развивал Линдберг, рассматривая экстремум нормализованного Лапласиана. DoG является конечно-разностной аппроксимацией оператора Лапласа, применяемого к слоям.

Алгоритм выбора доминирующей ориентации.

Наиболее распространенный метод для выделения векторов признаков, инвариантных относительно поворота исходного изображения на произвольный угол – построение гистограмм ориентированных градиентов (Histogram of oriented Gradients, HoG). Суть подхода состоит в построении гистограммы градиентов функции яркости в окрестности ключевой точки по заданным направлениям. Количество направлений определяется желаемой степенью детализации описания ключевой точки (задаваемой точностью). Следующим шагом является присвоение ключевой точке доминирующего направления.

Рассмотрим способ построения HoG для SIFT. Возьмем набор масштабов σ_i , порождающих представление L_i исходного изображения I , называемого слоем. Тогда для каждого пикселя из окрестности заданного радиуса ключевой точки с координатами (x_j, y_k) , обнаруженной на слое L_i , величина градиента $m(x_j, y_k)$ и ориентация $\theta(x_j, y_k)$ вычисляются по формулам:

$$m(x_j, y_k) = \sqrt{(L_i(x_j + 1, y_k) - L_i(x_j - 1, y_k))^2 + (L_i(x_j, y_k + 1) - L_i(x_j, y_k - 1))^2} \times \theta(x_j, y_k) = \tan^{-1} \left(\frac{L_i(x_j, y_k + 1) - L_i(x_j, y_k - 1)}{L_i(x_j + 1, y_k) - L_i(x_j - 1, y_k)} \right). \quad (1)$$

Гистограмма ориентированных градиентов состоит из 36 ячеек, каждая из которых покрывает сектор в 10 градусов окрестности ключевой точки. Пики в этой гистограмме соответствуют доминирующим направлениям. Каждое доминирующее направление умножается на вес, зависящий от σ_i .

Алгоритм вычисления дескриптора ключевых точек

С помощью предыдущих шагов были найдены ключевые точки на изображении I в различных масштабах и назначена им ориентация. Это обеспечило инвариантность ключевых точек изображения относительно поворота, масштаба и сдвига. Выделения ключевых точек с ориентациями недостаточно для распознавания объекта на изображении. В вектор признаков объекта необходимо включить дополнительную информацию (дескриптор).

Перейдем к описанию алгоритма построения дескрипторов. Сначала создается массив размерности 4×4 гистограмм с 8-ю ячейками в каждой. Эти гистограммы строятся на основе доминирующих направлений, вычисляемых в подобластях размерами 4×4 пикселя в окрестности ключевой точки, размером 16×16 пикселей. Величины доминирующих ориентаций далее умножаются на коэффициент, пропорциональный t_i . Итоговый дескриптор ключевой точки строится в виде вектора на основе имеющихся гистограмм. Поскольку мы имеем $4 \times 4 = 16$ гистограмм, каждая из которых имеет по 8 ячеек, то вектор имеет 128 элементов. Этот вектор нормализуется на единицу длины в целях повышения инвариантности относительно аффинного изменения в освещенности.

Чтобы уменьшить эффекты нелинейного освещения, применяется порог 0.2, и вектор снова нормализуется.

Алгоритм Speed Up Robust Features (SURF)

Алгоритм SURF впервые предложен Гербертом Бэем [Bay, 2008]. Основное преимущество данного метода в скорости работы детектора ключевых точек. Это достигается использованием интегральных изображений и вычислением взвешенного определителя матрицы Гесса. При построении дескриптора применяются двумерные вейвлеты Хаара. Главным преимуществом SURF авторы называют скорость работы детектора ключевых точек. Кроме того, SURF инвариантен к масштабу и вращению изображения, небольшим изменениям освещения. На рисунке 1 показано сравнение SURF с SIFT-128 и GLOH [Mikolajczyk, Schmid, 2005].

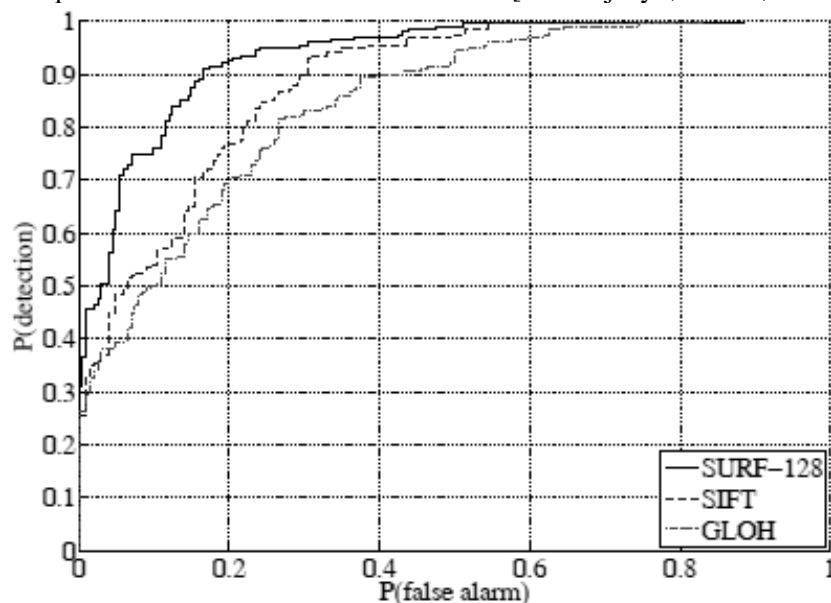


Рис. 1. Сравнение различных детекторов и дескрипторов в рамках BoF подхода

Расширение схемы построения детектора

Идея предлагаемого подхода заключается в том, чтобы построить детектор, выделяющий векторы признаков, инвариантные относительно заданных симметрий (свойств инвариантности изображения). Изменяя набор симметрий, мы получаем для одного и того же изображения вектора признаков, отражающие различные свойства изображения, повышая тем самым, информативность представления изображения. Как отмечалось выше, в алгоритме SIFT используется функция Гаусса. Заменяя ее на функцию Грина для уравнения, описывающего процесс Орнштейна–Уленбека мы получим векторы признаков, также обладающих симметриями [Лагно, Спичак, Стогний, 2004], но отражающих другие свойства изображения. Преимуществом данного подхода является использование дополнительного параметра в функции Грина, что приводит к повышению устойчивости работы детектора. Процесс Орнштейна–Уленбека описывается уравнением

$$\partial_{\tau} U(y, \tau) = \partial_y (\lambda y U(y, \tau)) + \frac{D}{2} \partial_{yy} U(y, \tau),$$

функция Грина которого имеет вид:

$$U(y, \tau | y_0, \tau_0) = \frac{\sqrt{\lambda}}{\sqrt{D\pi(1 - e^{-2\lambda(\tau - \tau_0)})}} \times \exp \left[-\frac{\lambda}{D\pi(1 - e^{-2\lambda(\tau - \tau_0)})} (y - y_0 e^{-\lambda(\tau - \tau_0)})^2 \right]. \quad (2)$$

Результаты сравнения стандартной и расширенной схем

Проведено сравнение стандартной реализации алгоритма SIFT на основе библиотеки Andrea Vidaldi и реализации этого же алгоритма (дескрипторы остались неизменными) с функцией Грина вида (2). Результат проиллюстрирован на рисунках 2–4. В работе [Shmid et al., 1998] рассматривается «стабильность» как критерий качества работы детектора. Под «стабильностью» подразумевается отношение $St = p_{inv} / P$, где p_{inv} – количество обнаруженных детектором инвариантных точек, а P – общее число обнаруженных точек.

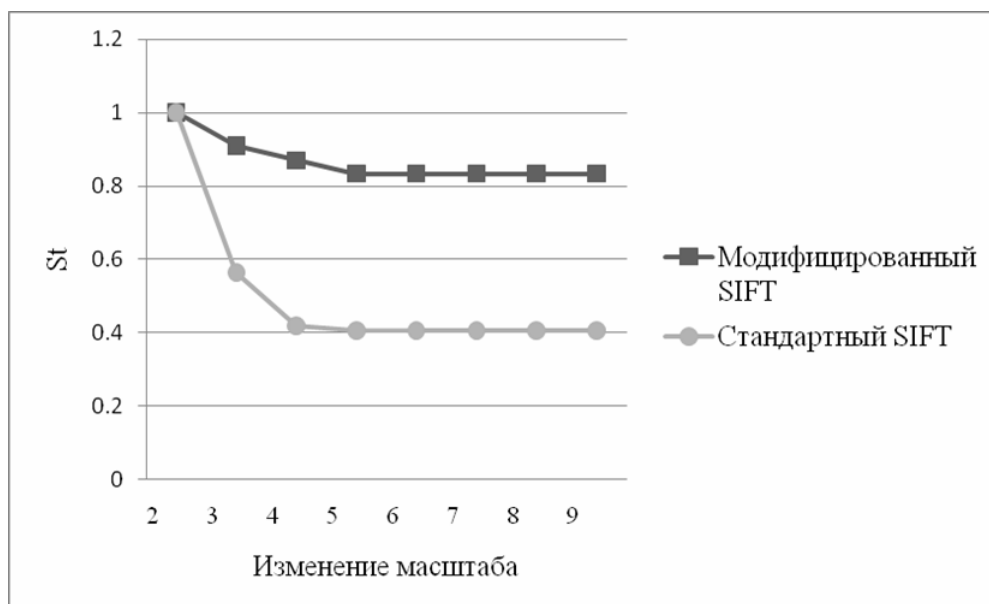


Рис. 2. Зависимость стабильности St работы детектора при изменении масштаба изображения

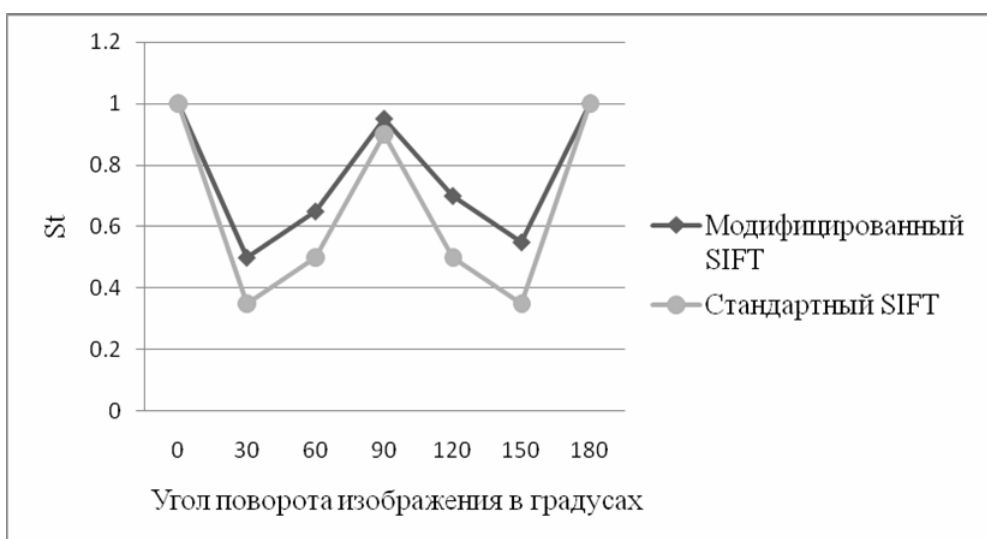


Рис. 3. Зависимость стабильности St работы детектора при повороте изображения



Рис. 4. Зависимость стабильности St работы детектора при изменении яркости изображения

Заключение

Предлагаемый подход в совокупности с методом Bag of Features позволит эффективно решать современные задачи распознавания объектов. Программный комплекс и тестирование производилось в ООО «Томсклаб».

Таким образом, в работе показаны те структурные элементы стандартных методов анализа изображений, которые могут быть модифицированы применением функции Грина уравнений с заданным набором симметрий, частным случаем которого является уравнение диффузии.

Список литературы

- Bay H., Ess A., Tuytelaars T., Gool L.V. Speeded Up Robust Features (SURF) // Computer vision and image understanding. – 2008. – Vol. 110(3). – Pp. 346–359.
- Csurka G., Dance C., Fan L., Willamowski J., Bray C. Visual Categorization with bags of keypoints // In ECCV Workshop on Statistical Learning in Computer Vision, – 2004.
- Florack L. M. J., Haar Romeny B. M. ter., Koenderink J. J., Viergever M. A. Linear scale-space // J. of Mathematical Imaging and Vision. – 1994. – Vol. 4(4). – Pp. 325–351.
- Koenderink J.J. The structure of images // Biological Cybernetics. – 1984. – Vol. 50. – Pp. 363–370.
- Labeznik S., Schmid C., Ponce J., Beyond Bags of Features: spatial pyramid matching for recognizing natural scene categories // In CVPR. – 2006. – Pp. 2169–2178.
- Lindeberg T. Feature Detection with Automatic Scale Selection // IJCV(30). – 1998. – № 2. – Pp. 79–116.
- Lowe D. G. Distinctive image features from scale-invariant keypoints // International Journal of Computer Vision. – 2004. – Vol. 60. – № 2. – Pp. 91–110.
- Mikolajczyk K., Schmid C. A performance evaluation of local descriptors // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2005. – Vol. 10, – Tome 27. – Pp. 1615–1630.
- Mikolajczyk K., Tuytelaars T., Schmid C., Zisserman A., Matas J.G., Schaffalitzky F., Kadir T., Van Gool L. J. A Comparison of Affine Region Detectors // IJCV(65). – 2005. – № 1–2. – Pp. 43–72.
- Obinata G., Dutta A. (eds.) Vision systems. Segmentation and pattern recognition. – I-Tech. – 2007.
- Schmid C., Mohr R., Bauckhage C. Comparing and Evaluating Interest Points // Proceedings of the 6-th International Conference on Computer Vision, Bombay, India. – 1998.
- Sivic J., Russell B. C., Efros A. A., Zisserman A., Freeman W.T. Discovering Objects and their Localization in Images // ICCV. – 2005. – Pp. 370–377.
- Steinwart I., Christmann A. Support Vector Machines. – N.Y.: Springer, 2008. – 602 p.
- Гонсалес Р., Вудз Р. Цифровая обработка изображений. – Техносфера, 2000. – 1072 с.
- Ибрагимов Н. Х. Группы преобразований в математической физике. – М.: Наука, 1983. – 280 с.
- Лагно В. И., Спичак С. В., Стогний В. И. Симметрий анализ уравнений эволюционного типа. – Москва–Ижевск: Институт компьютерных исследований. – 2004. – 392 с.
- Форсайт Д., Понс Ж. Компьютерное зрение. Современный подход. – Вильямс, 2004. – 928 с.
- Шапиро Л., Стокман Дж. Компьютерное зрение. – Бином. Лаборатория знаний, 2006. – 752 с.