

УДК 519.218.82

Н. И. Орешко

Восстановление закона изменения гетероскедастического шума при траекторных измерениях на основе вейвлет-технологии

Рассматривается проблема очистки от шума траекторных и телеметрических данных при наличии гетероскедастического шума. В отличие от более ранних статей [1]–[4], в которых был рассмотрен специальный класс модели изменения шума от времени, а именно кусочно-постоянная модель, здесь рассматривается общий случай изменения дисперсии шума во времени по неизвестному закону при естественных ограничениях на гладкость такой модели. Задача решается восстановлением закона изменения среднеквадратической погрешности нестационарного шума.

Очистка от шума, гетероскедастический шум, вейвлет-преобразование

Постановка задачи. Предположим, имеется набор некоторых измеренных данных y_i ($i=1, \dots, n$), выбранных на равномерной временной сетке t_i ($i=1, \dots, n$). Рассматривается аддитивная модель вида

$$y_i = f(t_i) + \sigma_i \xi_i \quad (i=1, \dots, n),$$

где $f(t_i)$ – истинные значения наблюдаемого процесса в моменты времени t_i ($i=1, \dots, n$), а случайные погрешности ξ_i независимы и одинаково распределены по нормальному закону $\xi_i \sim N(0,1)$, но дисперсия $\text{var}(\varepsilon_i) = \text{var}(\sigma_i \xi_i) = \sigma_i^2$, соответствующая случайной погрешности $\varepsilon_i = \sigma_i \xi_i$, изменяется во времени. Предполагается также, что закон изменения $\sigma_i = \sigma(t_i)$ от времени описывается, возможно, сложной, но гладкой функцией времени. Необходимо оценить $f(t_i)$ по результатам измерений y_i .

В такой постановке в статистике используется специализированный термин, а именно гетероскедастичность, означающий неоднородность наблюдений и выражающийся в неодинаковой (непостоянной) дисперсии случайной ошибки регрессионной модели. Гетероскедастичность противоположна *гомоскедастичности*, означающей однородность наблюдений, т. е. постоянство дисперсии случайных ошибок модели.

При рассмотрении классической регрессионной модели метод наименьших квадратов (МНК) дает наилучшие линейные несмещенные оценки лишь при выполнении ряда предпосылок, одной из которых является постоянство дисперсии отклонений (гомоскедастичность) для всех наблюдений. При невыполнении данной предпосылки (гетероскедастичности) последствия применения МНК будут следующими:

1. Оценки остаются несмещенными и линейными.
2. Оценки не будут эффективными.
3. Дисперсии оценок будут рассчитываться со смещением.
4. Статистические выводы, полученные при стандартных проверках качества оценок, могут быть ошибочными.

Для проверки наличия гетероскедастичности применяют, например, статистические тесты Уайта, Голдфелда–Куандта, Бройша–Пагана, Парка, Глейзера, Спирмена. В данной статье использован критерий Спирмена [5]. Для проверки гипотезы на основе данного критерия вначале производится выделение случайной погрешности ε_i ($i=1, \dots, n$) с помощью специализированного вида избыточного вейвлет-преобразования, а именно максимально накладывающегося дискретного вейвлет-преобразования (МНДВП). Выделенная случайная погрешность на всем интервале наблюдения и подвергается проверке с использованием данного критерия.

Тест ранговой корреляции Спирмена – непараметрический статистический тест, позволяющий проверить гетероскедастичность случайных ошибок. Особенность теста заключается в том, что не конкретизируется форма возможной зависимости дисперсии случайных ошибок модели от той или иной переменной.

Для проверки гипотезы гетероскедастичности выделенного случайного процесса погрешностей определяется коэффициент ранговой корреляции Спирмена:

$$\hat{\rho} = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)},$$

где d_i – разность рангов переменных ε_i и n (n соответствует рангу временной последовательности). При справедливости нулевой гипотезы (отсутствии гетероскедастичности, т. е. в данном случае – равенство нулю истинного значения коэффициента ранговой корреляции Спирмена ρ) статистика $\hat{\rho}\sqrt{n-1}$ асимптотически (т. е. при достаточно большом n) имеет стандартное нормальное распределение $N(0,1)$. Соответственно, если значение указанной статистики больше критического значения этого распределения (при данном уровне значимости), то гетероскедастичность признается значимой. В противном случае гетероскедастичность незначима.

Если коэффициент корреляции $\hat{\rho}$ для генеральной совокупности равен нулю, то статистика

$$t = \frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}}$$

имеет распределение Стьюдента с числом степеней свободы $\nu = n - 2$. Следовательно, если наблюдаемое значение t -статистики превышает $t_{кр} = t_{\frac{\alpha}{2}, n-2}$ (определяемое по таблице критических точек распределения Стьюдента), то необходимо отклонить гипотезу о равенстве нулю коэффициента корреляции $\hat{\rho}$, а следовательно, и об отсутствии гетероскедастичности.

В противном случае гипотеза об отсутствии гетероскедастичности принимается.

Обзор существующих подходов. Стандартным подходом при работе с гетероскедастичным шумом является предположение о виде его зависимости от времени (например, в виде некоторой простой параметрической зависимости –

чаще всего линейной) с последующей оценкой параметров предложенной модели. Если же модель изменения неизвестна, то используют методы вычисления дисперсии погрешностей на скользящем интервале. Для оценки σ могут применяться также различные методы сглаживания (ядерные методы, сглаживание сплайнами). Таким образом, в данном случае очистка от гетероскедастического шума основана на применении различной ширины полосы в непараметрической ядерной регрессии, что приводит к большему сглаживанию участков сигнала с высокой дисперсией. В [6] дисперсионная функция оценивается с использованием локального полиномиального сглаживания квадратов остатков. Другой подход, приведенный в [7], оценивает $\log(\sigma^2(t))$, используя сглаживание любым непараметрическим регрессионным методом логарифма квадрата остатков $\log(y(i) - \hat{f}(i))^2$. Далее дисперсия определяется из равенства $\hat{\sigma}_t^2(t) = e^{\hat{q}(t)}$, где $\hat{q}(t)$ – сглаживающая оценка $\log(\sigma^2(t))$. Недостаток этого подхода заключается в том, что логарифм очень малых остатков приводит к большим выбросам. Адаптивное оценивание дисперсионной функции с использованием вейвлет-пороговой обработки описано в [8], где процедура вейвлет-пороговой обработки применяется к квадрату разностей исходных наблюдений. В этом случае задача оценивания дисперсионной функции сводится к оцениванию среднего.

Описанные в литературе методы в основном используют квадраты остатков для оценивания закона изменения шумовой составляющей. Однако они довольно чувствительны даже к небольшим выбросам и точность восстановления модели невысокая.

Алгоритм удаления нестационарного шума, дисперсия которого изменяется по неизвестному гладкому закону. Предлагаемый алгоритм восстанавливает закон изменения среднеквадратичного отклонения шумовой составляющей, используя абсолютные значения вейвлет-коэффициентов, и осуществляет пороговую обработку с учетом восстановленного закона.

Алгоритм включает в себя следующие этапы:

1. Выделение шумовой составляющей с использованием МНДВП.

2. Вычисление модуля выделенной шумовой составляющей во временной области.

3. Определение тренда модуля шума.

4. Пороговая обработка вейвлет-коэффициентов с учетом определенного закона для шумовой составляющей.

5. Обратное вейвлет-преобразование.

Выделение шумовой составляющей. Шумовая составляющая сигнала определяется за счет разделяющей способности вейвлет-преобразования. Шумовые вейвлет-коэффициенты сосредоточены на самых тонких (высокочастотных) уровнях вейвлет-разложения. Мультиразрешающий анализ на основе МНДВП позволяет выделить высокочастотные составляющие во временной области. Эта процедура подобна использованию фильтра с нулевой фазой, поэтому закон изменения СКО шума не искажается за счет несимметричной формы используемого при разложении вейвлет-фильтра. Таким образом, выделенную шумовую составляющую можно представить в виде

$$\tilde{\sigma} = \underbrace{B_1^T W_1}_{D_1} + \underbrace{A_1^T B_2^T W_2}_{D_2} + \dots + \underbrace{A_1^T \dots A_{J-1}^T B_J^T W_J}_{D_J},$$

где W_1, W_2, \dots, W_J – вейвлет-коэффициенты первых J уровней разложения; D_1, D_2, \dots, D_J – проекции на детализирующее пространство соответствующих частотных полос; A, B – матрицы преобразований.

Выделенная составляющая содержит в основном шум, если истинный сигнал является гладким. Для сигналов с особенностями (резкими изменениями, пиками) составляющая содержит кроме шума небольшое число значительно отличающихся по амплитуде вейвлет-коэффициентов, соответствующих этим особенностям. В этом случае для меньшего искажения особенностей при пороговой обработке рекомендуется использовать вейвлет с малой длиной фильтра, т. е. с меньшим носителем.

Определение закона изменения СКО шума.

Шумовая составляющая может изменяться по какому-либо неизвестному закону, однако предполагается, что этот закон описывается некоторой гладкой функцией. Более точно предполагается, что существует по крайней мере вторая непрерывная производная для этой функции.

Закон изменения СКО шума определяется по абсолютному значению выделенной на предыдущем этапе шумовой составляющей с использованием разных подходов, например ядерного сглаживания, разложения по эмпирическим модам, вейвлет-разложения. В данном исследовании высокую эффективность и гибкость показал метод робастной сплайн-аппроксимации [9].

Для окончательного получения оценки закона изменения шума $\hat{\sigma}(t)$ вычисленный тренд нормируется на величину $\sqrt{2/\pi}$, поскольку для среднего значения распределения модуля случайной величины, распределенной по нормальному закону, справедливо соотношение

$$E(\tilde{\xi}) = \sigma\sqrt{2/\pi} \text{ при } \mu = 0,$$

где $\xi \in N(\mu, \sigma)$ и $\tilde{\xi} = |\xi|$ [5].

Восстановленный закон изменения СКО шума может использоваться для очистки от шума исходного сигнала с использованием различных методов. Если применять вейвлет-подход, то вейвлет-преобразование должно применяться и для исходного сигнала, и для восстановленного закона изменения СКО шума. В этом случае порог рассчитывается с учетом вейвлет-преобразования закона СКО шумовой составляющей.

Результаты. Рассмотрим результаты работы предложенного подхода для двух экспериментов. Для первого эксперимента (рис. 1–4) в качестве основной задачи исследовалась возможность восстановления сложного поведения дисперсии шумовой составляющей. Закон изменения шумовой компоненты моделировался в виде синусоиды (рис. 3, а), а на рис. 3, б приведена сама реализация шумового процесса. Сам же сигнал (светлым цветом) и его очистка от шума (темная жирная линия) на основе стандартного избыточного вейвлет-преобразования изображены на рис. 1. Сигнал (светлым цветом) и его очистка от шума (темная жирная линия) на основе предложенной технологии изображены на рис. 2. На рис. 4 приведены этапы работы предложенного алгоритма. На рис. 4, в приведена нормированная восстановленная реализация, на рис. 4, б – аппроксимация (светлым цветом) абсолютного значения выделенного шума (темным цветом), а на рис. 4, а – аппроксимация (светлым цветом) выделенной компоненты (темным цветом).

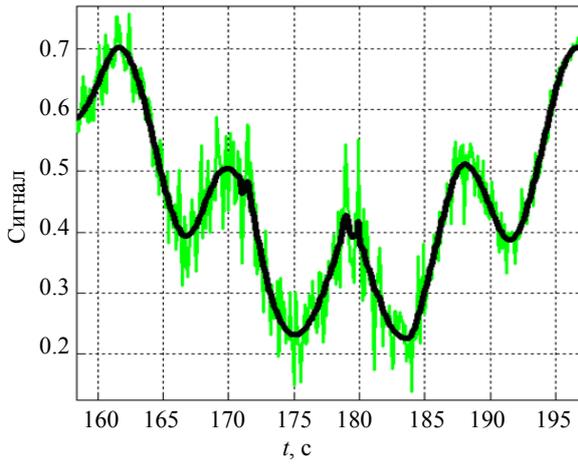


Рис. 1

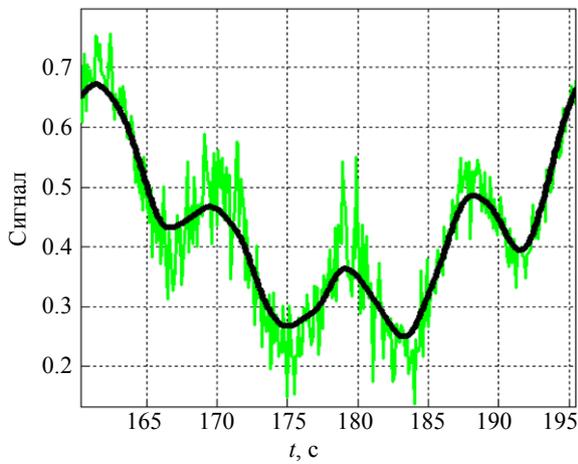
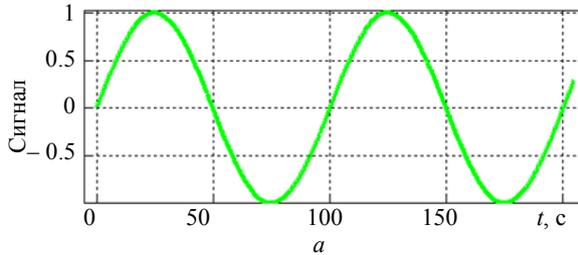


Рис. 2



а

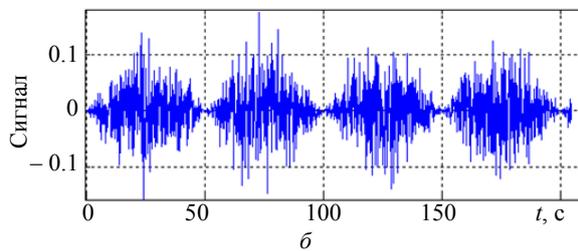
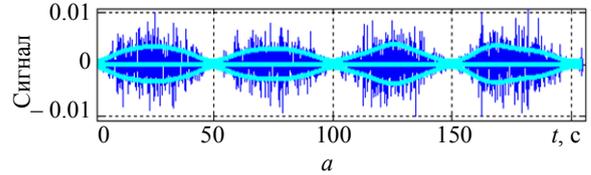


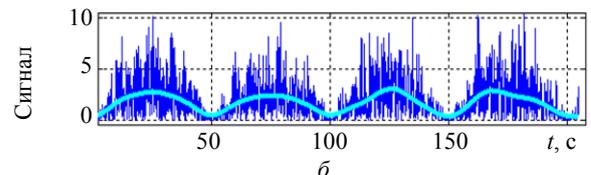
Рис. 3 б

Основная цель второго эксперимента (рис. 5–10) – показать эффективность работы предложенного алгоритма в задаче очистки сложного сигнала при наличии гетероскедастического шума. Сигнал моделировался в виде волнообразного процесса (рис. 5), а шум – в виде процесса с посто-

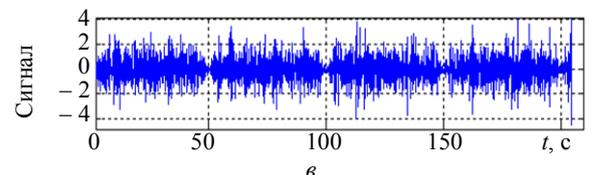
янной единичной дисперсией, существенно возрастающей на конечном интервале наблюдения (рис. 9). На рис. 6 показаны исходный (светлая линия) и очищенный (темная линия) с использованием стандартной вейвлет-технологии сигналы, а на рис. 7 – исходный и очищенный с использованием предложенной вейвлет-технологии сигналы.



а



б



в

Рис. 4

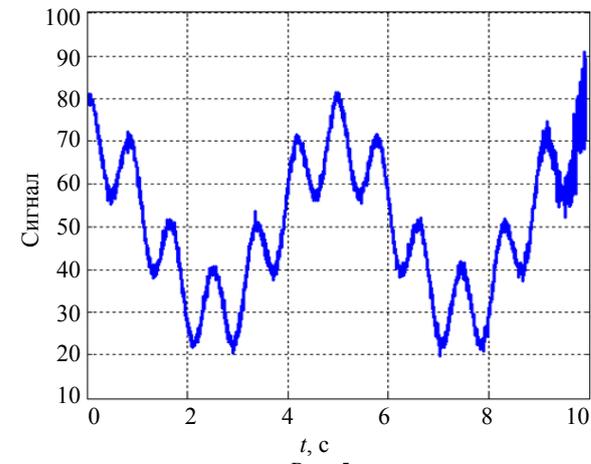


Рис. 5

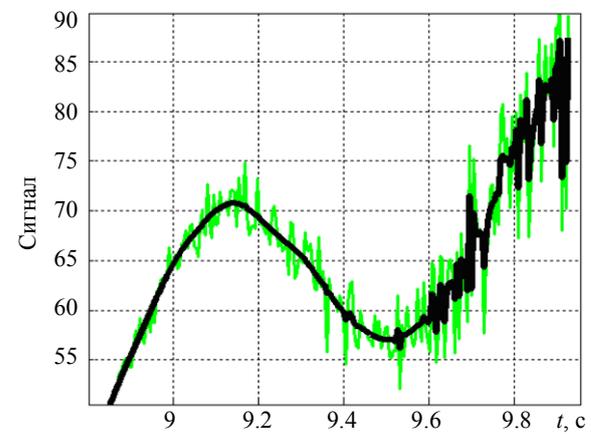


Рис. 6

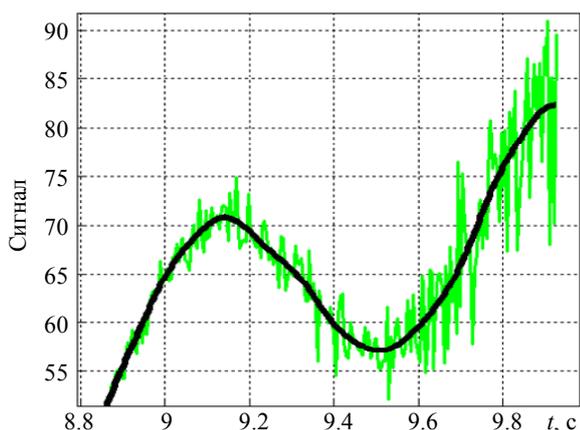


Рис. 7

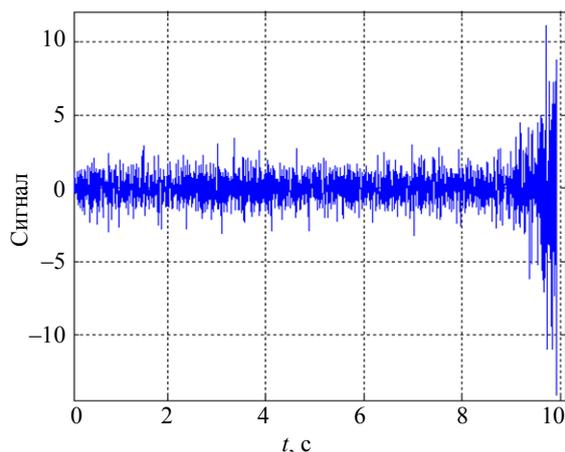
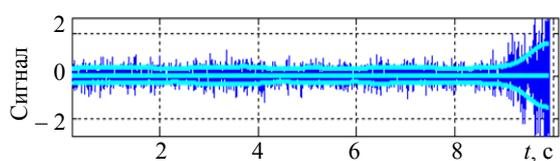
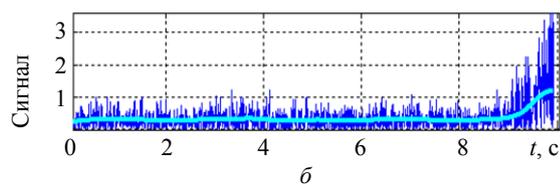


Рис. 9



а



б

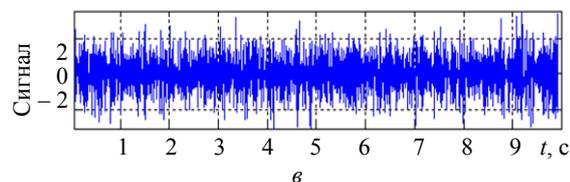


Рис. 8

На рис. 8 приведены этапы работы предложенного алгоритма. На рис. 8, в приведена нормированная восстановленная реализация, на рис. 8, б – аппроксимация (светлым цветом) абсолютного значения выделенного шума (темным цветом), а на рис. 8, а – аппроксимация (светлым цветом) выделенной компоненты (темным цветом).

На рис. 9 приведен выделенный алгоритмом шумовой процесс.

Из сравнения рис. 6 и 7, где приведены в увеличенном масштабе конечные части исходного и очищенного от шума сигналов, видно радикальное улучшение качества очистки от шума на основе предложенной технологии.

Таким образом, алгоритм проявляет высокую эффективность при наличии гетероскедастического шума с неизвестным законом изменения для сигналов со сложным поведением.

СПИСОК ЛИТЕРАТУРЫ

1. Князева Т. Н., Новиков Л. В., Орешко Н. И. Удаление нестационарного шума из экспериментальных данных // Науч. приборостроение. 2008. Т. 18, № 2. С. 61–65.
2. Knyazeva T. N., Oreshko N. I. Wavelet Denoising of Experimental Data with Non-stationary Noise // Тр. Междунар. конф. "Wavelets and Applications". 2009. P. 28–30.
3. Орешко Н. И., Князева Т. Н. Вейвлет-технология анализа и очистки от шума сигналов // Цифровая обработка сигналов. 2008. № 3. С. 21–25.
4. Орешко Н. И., Князева Т. Н. Очистка от шума траекторных данных при наличии гетероскедастических погрешностей // Изв. СПбГЭТУ "ЛЭТИ". 2007. Вып. 2. С. 31–40.
5. Кобзарь А. И. Прикладная математическая статистика. М.: ФИЗМАТЛИТ, 2006.
6. Fan J., Yao Q. Efficient estimation of conditional variance functions in stochastic regression // Biometrika. 1998. № 85. P. 645–660.
7. Yu K., Jones M. Likelihood-based local linear estimation of the conditional variance function // J. of the American Statistical association. 2004. № 99. P. 528–555.
8. Cai T., Wang L. Adaptive variance function estimation in heteroscedastic nonparametric regression // Ann. Statist. Vol. 36, № 5. P. 2025–2054.
9. Garcia D. Robust smoothing of gridded data in one and higher dimensions with missing values // Computational Statistics & Data Analysis. 2010.

N. I. Oreshko

ESTIMATING LAW OF VARIATION OF HETEROSCEDASTIC NOISE IN TRAJECTORY MEASUREMENTS ON THE BASIS OF WAVELETS

In this paper we describe a two-step method of denoising of complex data in heteroscedastic nonparametric regression. At the first step we extract using wavelets a noise component and estimate for its absolute values using robust spline fitting the model for the variance function. At the second step we employ the estimated variance function for wavelet coefficients thresholding on the basis of overlap discrete wavelet transform.

Denoising, heteroscedastic noise, wavelet transform

УДК 621.397.01

С. С. Голяк, С. А. Гордеев, И. А. Горошков, Г. В. Разумовский

Комплексная система видеонаблюдения за технологическими процессами

Описывается архитектура комплексной системы видеонаблюдения, рассматриваются вопросы ее взаимодействия с различными информационными системами промышленного предприятия и использования для контроля технологических процессов.

Система видеонаблюдения, детектор видеоаналитики, сервер обработки событий, контроллер видеосигналов

Необходимость слежения и контроля за технологическими процессами обуславливается многими факторами, основные из которых – безопасность производства и предотвращение хищений товарно-материальных ценностей. Технологический процесс можно рассматривать как последовательность некоторых технологических событий, каждое из которых может контролироваться отдельными средствами и методами. Полностью контролировать технологический процесс одним средством практически невозможно, поэтому для контроля технологических событий используются различные датчики, сканеры, магнитные метки и считыватели, а также целые автоматизированные системы управления. Относительно недавно к этому списку добавились и видеокамеры, которые в совокупности объединены в систему видеонаблюдения.

Специализированное программное обеспечение способно анализировать получаемую с видеокамеры картинку и при необходимости генерировать соответствующие события. Видеокамеры выгодно отличаются от других средств контроля тем, что их можно установить почти в любом месте, даже там, где нельзя поставить датчики, либо

датчики в этом месте будут плохо работать. Также часто предпочтение отдается видеокамерам ввиду их относительной дешевизны по сравнению с дорогостоящими датчиками, метками и сканерами.

Система видеонаблюдения вместе со специализированным программным обеспечением анализа изображения составляют систему видеоаналитики. Система видеоаналитики предоставляет следующие возможности:

- идентификация объекта, т. е. принятие решения о том, что в кадре присутствует искомый объект;
- определение направления движения объекта;
- распознавание пересечения объектом заданной линии;
- распознавание текста, в частности, номеров автомобилей и железнодорожных составов.

Таким образом, система видеоаналитики совместно с другими датчиками и средствами контроля располагают исходными данными для отслеживания технологического процесса комплексной системой видеонаблюдения. Слежение происходит сопоставлением полученных событий от системы видеоаналитики и других средств контроля с некоторыми правилами, заданными в системе заранее.