

Оценка качества входных изображений в системах распознавания видеопотока¹

Т.С. Чернов, Н.П. Разумный, А.С. Кожаринов, Д.П. Николаев, В.В. Арлазаров

Аннотация. Системы распознавания и машинного зрения уже долгое время широко задействуются во многих дисциплинах с целью автоматизации различных процессов жизнедеятельности и промышленности. Входные изображения систем оптического распознавания могут подвергаться большому количеству различных искажений, особенно в неконтролируемых или естественных условиях съемки, что приводит к непредсказуемым результатам систем распознавания, делая невозможным оценку их достоверности. По этой причине необходимо производить контроль качества входных данных систем распознавания, чему способствует современный прогресс в области оценки качества изображений. В данной работе исследуется подход к построению систем распознавания образов со встроенными модулями оценки качества входных изображений и обратной связью, для чего введены необходимые определения и построена модель описания таких систем. Работоспособность подхода иллюстрируется на примере решения задачи выбора наилучших кадров для распознавания в видеопотоке. Приводятся экспериментальные результаты с системой распознавания документов, удостоверяющих личность, показывающие значительное увеличение точности и скорости работы системы при искусственно моделируемых условиях автоматической фокусировки камеры, приводящим к размытию кадров.

Ключевые слова: системы распознавания, оценка качества изображений, системный анализ, видеопоток, дефокус, размытие.

Введение

Системы распознавания и машинного зрения уже долгое время широко задействуются во многих дисциплинах с целью автоматизации различных процессов жизнедеятельности и промышленности [1], но построение систем распознавания с идеальной точностью на практике невозможно [2]. Тем не менее, разные задачи имеют разную цену ошибки или требуемый уровень надежности, который можно обеспечить с помощью правил определения достоверности результатов распознавания [3].

Система распознавания должна не только иметь высокую точность работы, но и стабильно выдавать отказ [4] при возникновении нештатной ситуации или при подозрении на не-

корректность результата, что является запросом на передачу контроля дальнейших действий оператору.

Входные изображения систем оптического распознавания могут подвергаться большому количеству различных искажений, особенно в неконтролируемых или естественных условиях съемки [5]: дефокусировка, смаз, цифровой шум, разладка регистратора, артефакты сжатия, блики, проективные искажения, повреждения объекта обработки и т.д. Проблема заключается в том, что при искаженных входных данных поведение систем распознавания не всегда предсказуемо [6], из-за чего невозможно корректное применение методов определения достоверности результатов распознавания. Следовательно, для построения систем распознавания заданной надежности необходима

¹ Работа выполнена при финансовой поддержке РФФИ (проекты № 16-07-01051, №15-29-06091, №17-29-03170).

разработка методов контроля качества входных изображений.

Оценка качества изображений является широкой и активно развивающейся областью, исследующей вопросы контроля качества в его различных приложениях [7]. На текущий момент разработано множество объективных (автоматических) методов, эффективно аппроксимирующих как субъективные экспертные оценки качества изображений, так и, например, точность распознавания в качестве предиктора качества при взаимодействии с изображениями документов, печатных текстов и других объектов [8, 9]. Поэтому, объективные методы оценки качества изображений представляют большой интерес при встраивании их в системы распознавания с целью контроля качества входных данных. Однако отсутствуют модели, методологии или рекомендации к созданию систем распознавания, учитывающих качество входных изображений с целью повышения точности, устойчивости и других характеристик систем, которым и посвящена данная работа.

В данной работе исследуется оригинальный подход к построению систем распознавания образов со встроенными модулями оценки качества входных изображений и обратной связью на каждом этапе обработки. Предлагаемый подход обеспечивает возможность контроля уверенности результата и отказа в дальнейшей обработке в случае низкого качества входного изображения на рассматриваемом этапе, что позволяет увеличить точность, устойчивость и скорость работы систем распознавания в неконтролируемых условиях съемки. Введены необходимые определения и построена математическая модель архитектуры систем распознавания, применимые для программной реализации подхода в проектируемых и существующих системах.

Значимость подхода иллюстрируется на примере решения задачи выбора наилучших кадров для распознавания в видеопотоке с по-

мощью функции приоритета кадра, зависящей от его качества и времени съемки. Приводятся экспериментальные результаты решения данной задачи для реализации системы распознавания удостоверяющих личность документов в искусственно моделируемых условиях автоматической фокусировки камеры, приводящим к размытию кадра. Результаты демонстрируют значительное увеличение точности распознавания с различными ограничениями на продолжительность видеопотока при выборе кадров для распознавания в соответствии с их приоритетом.

1. Строение систем распознавания

Системы распознавания и машинного зрения применяются среди многих приложений и имеют различное внутреннее устройство и архитектуру, но в их структуре можно выделить функциональные блоки, присущие большинству из них. Можно сказать, что общий подход к построению архитектур систем распознавания остается неизменным на протяжении нескольких десятков лет [1, 10]. На Рис. 1 показана обобщенная схема работы систем распознавания, состоящая из последовательных этапов. Рассмотрим их подробнее.

В любой системе оптического распознавания первым этапом является захват цифрового изображения сцены, освещенной некоторым источником света, предположительно содержащей распознаваемый объект – как со сканера, камеры мобильного устройства или стационарной камеры, так и с других типов сенсоров. Следующим этапом является предварительная обработка изображения, например, приведение его к фиксированному размеру, конвертация цветового пространства, выравнивание баланса белого, яркости, контраста, резкости и другие способы обработки, с целью приведения изображения в более удобный для последующей работы вид [11].



Рис. 1. Обобщенная схема работы систем распознавания

Затем, производится этап сегментации [12], цель которого – выделить регионы изображения, соответствующие объектам на сцене или их участкам. Результатом сегментации в общем случае являются битовые маски принадлежности регионов изображения объектам, но если объекты рассматриваются отдельно друг от друга на дальнейших этапах обработки, то возможно сопоставить каждому объекту отдельное его изображение: выделить зону интереса ограничивающим прямоугольником или выполнить проективную коррекцию для объектов прямоугольной формы. При этом производится смена координатной системы относительно полученных изображений.

Процесс сегментации имеет иерархичную структуру в случае вложенности объектов интереса друг в друга. Например, при построении систем распознавания лиц [13] до непосредственного детектирования лиц происходит сегментация участков кожи и их объединение с помощью морфологических операций. В системе распознавания документов [14] сначала определяется местоположение документа и идентифицируется его тип с помощью модификаций метода Виолы-Джонса, методов особых точек и дескрипторов, классических методов поиска границ и других. Когда документ найден, производится его проективная коррекция и выделяются фиксированные зоны, на которых, с помощью методов математической морфологии, ищутся текстовые поля. Наконец, найденные поля передаются в модуль сегментации полей, возвращающий отдельные изображения символов.

Изображения или участки найденных на этапе сегментации объектов передаются на этапы извлечения признаков и классификации: распознавания или получения дескрипторов объектов для их последующей идентификации [15]. Современные подходы, например, с использованием глубоких сверточных нейронных сетей, позволяют совместить эти два этапа в один путем отказа от "ручных" способов извлечения признаков на этапе классификации. Однако, использование последних все еще актуально во многих приложениях.

Финальным в рамках распознавания одного кадра этапом является пост-обработка резуль-

татов классификации. Примерами такой пост-обработки являются контекстная коррекция распознанного текста [16] или речи [17] в соответствии с моделью языка, проверка контрольных сумм на стандартизованных документах, а также другие обработки, использующие семантические правила предметной области. После этого этапа происходит возвращение текущего результата пользователю вместе с обратной связью системы о возможных улучшениях условий съемки: рефокусировки или изменении угла камеры, контроль освещения и т.п.

При разработке систем распознавания в видеопотоке открывается как ряд возможностей повышения качественных характеристик системы, недоступных при распознавании одного кадра, так и ряд новых задач [14]. Например, становится возможной интеграция результатов на текущем кадре с результатами распознавания предыдущих кадров на этапе пост-обработки, отслеживание положения объектов в пространстве и т.д. С другой стороны, встает вопрос о том, когда останавливать процесс распознавания.

В конце обработки каждого из этапов, ответственная за этап подсистема может выдать как корректный результат обработки вместе с скалярным уровнем своей уверенности в его корректности, так (для некоторых подсистем) и сигнал об отказе в обработке – при низком качестве рассматриваемого изображения или низкой уверенности в результате на текущем шаге. Преимуществ у отказа при заведомо невозможной дальнейшей обработке несколько [18]: экономится время, некорректные результаты не проходят дальше и не снижают качество работы системы, что наиболее актуально при интеграции результатов во время распознавания в видеопотоке.

2. Постановка задачи оценки качества изображений

Введем определения оценки качества изображений, необходимые для построения модели их внедрения в системы распознавания. Объективные методы оценки качества изображений могут быть классифицированы по доступности исходного изображения. Кроме этого, результа-

том метода может являться как скалярная оценка, так и изображение или другой объект, задающий положение и степень проявления различных искажений.

Обозначим за \mathcal{J} множество входных изображений для рассматриваемого метода оценки Q , а за \mathcal{D} – множество выходных оценок качества. Пусть имеется изображение $I \in \mathcal{J}$. Объективной функцией оценки качества в случае отсутствия информации об исходном изображении назовем функцию Q_{NR} , принимающую одиночное входное изображение:

$$Q_{NR}: \mathcal{J} \rightarrow \mathcal{D} \quad (1)$$

При наличии исходного (не подвергнутого искажению) изображения $I^* \in \mathcal{J}$, соответствующий метод оценки Q_{FR} выглядит, как:

$$Q_{FR}: \mathcal{J} \times \mathcal{J} \rightarrow \mathcal{D} \quad (2)$$

В отдельных случаях доступна лишь часть информации об исходном изображении, способствующая оценке степени его искажения. Обозначив за \mathcal{P} множество указанной вспомогательной информации, получаем вид метода Q_{PR} с частичной информацией об исходном изображении:

$$Q_{PR}: \mathcal{J} \times \mathcal{P} \rightarrow \mathcal{D} \quad (3)$$

Рассмотрим подробнее множество \mathcal{D} выходов методов. В самом распространенном для общей оценки качества случае $\mathcal{D} = \mathcal{D}_R$ представляет собой множество скалярных действительных чисел \mathbb{R} с ограничением на $[0,1]$, где 1 означает отсутствие повреждений, а 0 – наличие серьезных помех на изображении:

$$Q(I) = D \in \mathcal{D}_R \quad (4)$$

Однако в контексте применения в системах распознавания важно знать положение в пространстве и степень проявления соответствующих искажений на входном изображении. Данный эффект можно получить прямым расширением \mathcal{D} до множества выходных изображений. Эти изображения могут представлять собой бинарные маски, задавая только положение искажения в пространстве:

$$Q(I) = D \in \mathcal{D}_B, D = \{d_{x,y} \in \{0,1\} \mid x, y: I_{x,y} \in I\}, \quad (5)$$

или карты вещественных чисел, оценивающие степень проявления искажений в каждой точке $I_{x,y}$ исходного изображения:

$$Q(I) = D \in \mathcal{D}_I, D = \{d_{x,y} \in \mathbb{R}_{[0,1]} \mid x, y: I_{x,y} \in I\} \quad (6)$$

Для некоторых видов искажений информация о деградации в каждой точке исходного изображения может быть избыточной и неудобной для дальнейшей обработки, поэтому введем дополнительные типы возвращаемых значений методов оценки качества, такие, как компоненты связности для случая бинарных масок со степенью повреждения $q \in \mathcal{D}_R$ в данной компоненте:

$$Q(I) = D \in \mathcal{D}_C, D = \{(q_i, C_i)\}, C_i = \{(x, y) \mid x, y: I_{x,y} \in I\}, C_i \cap C_j = \emptyset, \quad (7)$$

или окаймляющие прямоугольники поврежденных областей, заданные координатами (x, y) верхней левой точки, шириной и высотой (w, h) :

$$Q(I) = D \in \mathcal{D}_O, D = \{(q_i, \langle x_i, y_i, w_i, h_i \rangle)\} \quad (8)$$

В дальнейшем под множеством \mathcal{D} будет подразумеваться одно из вышеперечисленных множеств:

$$\mathcal{D} \in \{\mathcal{D}_R, \mathcal{D}_B, \mathcal{D}_I, \mathcal{D}_C, \mathcal{D}_O\}, \quad (9)$$

в зависимости от конкретного приложения задающих информацию об искажениях.

3. Модель системы распознавания с модулем оценки качества изображений

Построим модель системы распознавания, содержащей подсистему оценки качества, и опишем их взаимодействие между собой.

Рассмотрим систему $G = \langle S, E \rangle$, представленную в виде ориентированного графа модулей обработки $S = \{S_0, \dots, S_N\}$ и связей между ними $E = \{(S_i, S_j) \mid S_i, S_j \in S\}$. Каждый модуль (этап, подсистема) обработки S_i может быть произвольной процедурой обработки, анализа (в том числе – распознавания), синтеза, компрессии, декомпрессии изображений и иметь своими входами и выходами как изображения, так и другие типы данных: вектора признаков, строки результатов распознавания, найденные контуры объектов и другие. Пусть S_0 – особый этап, имеющий пустой вход и заключающийся в регистрации изображения сцены, которое и возвращается на выходе этого этапа.

Выходом общего этапа S_i является упорядоченный набор объектов $r_i = (r_{i,0}, \dots, r_{i,N_i})$. Каждый отдельный объект $r_{i,j}$ содержит непосредственный набор выходных данных $v_{i,j}$, тип которых зависит от устройства S_i и может

быть, например, изображением или вектором альтернатив распознавания. Кроме того, он содержит скалярную оценку уверенности этапа в корректности своего результата $c_{i,j} \in [0,1]$ и опциональную информацию $l_{i,j}$ о местонахождении результата в глобальной системе координат – аффинное или проективное преобразование для получения позиции символа после модуля сегментации строки в координатах текущего кадра. Итого, $r_{i,j} = \langle v_{i,j}, c_{i,j}, l_{i,j} \rangle$. Данные выходы поступают на вход соединенным дугами из E этапам обработки, которые, в том числе, принимают решение о целесообразности их обработки с учетом оценок уверенности $c_{i,j}$ и учитывают их в своих последующих оценках.

Опишем процесс встраивания модулей оценки качества изображений в систему. Ограничимся модулями исходной системы S_i , и теми их выходами $r_{i,j}$, компоненты $v_{i,j}$ которых содержат данные из множества изображений, т.е. $v_{i,j} \in \mathcal{I}$.

Пусть $Q_{i,j}$ – модуль оценки качества изображений для i -го выхода подсистемы S_i , принимающий на вход $r_{i,j}$. Данный модуль реализует одну из описанных в предыдущем разделе функций $Q(I)$ и возвращает оценки качества $q_{i,j} \in \mathcal{D}_{i,j}$ ($\mathcal{D}_{i,j} \in \mathcal{D}$ по формуле 9), которые, как было обозначено ранее, являются пространственным распределением оценок качества для поступившего изображения $v_{i,j}$ в системе координат, заданной $l_{i,j}$.

Введем модуль коррекции и принятия решений о дальнейшей обработке $\Psi_{i,j}$, принимающий на вход как оценки качества $q_{i,j}$, так и результат $r_{i,j}$. Данный модуль возвращает модифицированные результаты $r_{i,j}^*$, тип которых совпадает с типом $r_{i,j}$, но компонента уверенности в корректности результата $c_{i,j}^*$ может отличаться в соответствии с логикой учета пространственного распределения оценок качества $q_{i,j}$. Важной особенностью $\Psi_{i,j}$ является возможность выдачи отказа в дальнейшей обработке за счет присвоения $c_{i,j} = 0$, когда повреждение покрывает большую часть изображения $v_{i,j}$, и передачи сообщения об этом родительской подсистеме S_i с помощью обратной связи. Другим способом модификации уверенности

при повреждениях может быть снижение значения уверенности $c_{i,j}$ путем умножения на некоторое скалярное значение $\theta(q_{i,j})$, вычисляемое как средний уровень деградации качества изображения от эталонного. Кроме того, модуль возвращает восстановленное от дефектов изображение $v_{i,j}^*$ при таковой возможности.

Последним вводимым модулем в систему является опциональный модуль внимания или интереса A_i , соответствующий подсистеме S_i . Его задача – построить карту интереса $a_i(x, y)$ подсистемы S_i в участках обрабатываемого подсистемой объекта при условии его выделения на предыдущих этапах, за счет чего возможно контролировать приоритет обработки в видеопотоке, что будет описано подробнее в следующем разделе. На вход данному модулю подаются оценки качества $q_{i,j}$ и модифицированные результаты работы подсистемы $r_{i,j}^*$, причем в модуле хранится история о результатах на предыдущих кадрах, т.е. выходная карта является интегральной по видеопотоку на текущий момент. Обновление карты интереса может происходить путем усреднения обратных показателей качества $(1 - q_{i,j})$ или уверенности в корректности работы $(1 - c_{i,j}^*)$ в областях, пересчитанных обратно в координатах исходного объекта с помощью $l_{i,j}$. Если $\Psi_{i,j}$ принимает решение об отказе дальнейшей обработки, модуль A_i все равно обновляет свою карту интереса в сторону его увеличения в поврежденных областях. Применение карты интереса особенно актуально при распознавании объектов с фиксированными формами, например, документов – низкий интерес в регионе означает, что система уже имеет высокий уровень уверенности в результатах распознавания в этом регионе, или же что в нем отсутствуют какие-либо информативные участки.

Наконец, по завершении процесса оценки качества и коррекции, модифицированные результаты $r_{i,j}^*$ передаются на следующий этап обработки в соответствии с существующими связями между подсистемами.

На Рис. 2 проиллюстрирована модель графа подсистем и их связей после добавления модулей оценки качества изображений в процесс обработки.

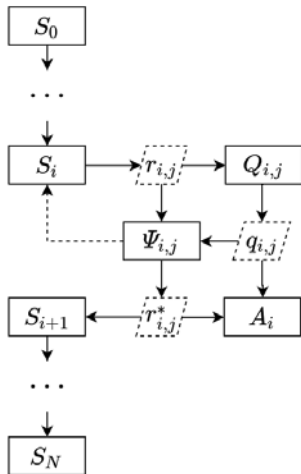


Рис. 2. Граф системы обработки с модулями оценки качества изображений

При реализации данной модели в программных системах добавляемые модули оценки качества и принятия решений могут быть многократно использованы при однотипных выходах соответствующих подсистем.

4. Выбор лучших кадров для распознавания в видеопотоке

При распознавании в видеопотоке становится актуальной задача выбора наилучших по какому-либо критерию качества кадров для распознавания из целого потока кадров. Это обусловлено несколькими причинами.

Во-первых, большинство современных камер (как мобильных устройств, так и других) имеет возможность захватывать кадры с частотой 30-60 кадров в секунду и выше, что может в разы превышать производительность обработки одного кадра системой распознавания [14].

Во-вторых, помимо распознавания на устройстве, также популярным является сценарий только съемки фотографии или видеопотока с интересующим объектом на устройстве, а затем посылка этих данных на удаленный сервер для последующего распознавания [19]. Отсылать весь захваченный видеопоток целиком неблагоприятно как при слабом канале связи, так и при необходимости обеспечить постоянную обратную связь с пользователем о процессе распознавания.

В-третьих, остаются верными предыдущие аргументы о целесообразности оценки качества в системах распознавания: передавая кадр с за-

ведомо низким качеством на распознавание существует риск снижения общей точности распознавания.

Модуль выбора кадров из избыточного видеопотока может быть встроен в соответствии с ранее приведенной схемой на Рис. 2 на основании результатов подсистемы захвата кадров S_0 и модуля предварительной оценки качества Q_0 .

Пусть Ψ_0 – модуль выбора кадров, имеющий устройство очереди с приоритетом, накапливающий входные кадры и при запросе выдающий кадр с наибольшим приоритетом. Если после добавления очередного кадра размер очереди превышает L , из нее удаляется кадр с наименьшим приоритетом. Введем дополнительный модуль оценки приоритета кадра P , зависящий от результатов оценки качества $q(x, y)$, времени t нахождения кадра в очереди, а также от карты интереса $a(x, y)$ системы распознавания в соответствующих регионах входного изображения ($0 \leq q(x, y), a(x, y) \leq 1$). На основании пространственных распределений $q(x, y)$ и $a(x, y)$ вычислим коэффициент w , показывающий пригодность данного кадра со точки зрения качества интересующих систему регионов:

$$w = \sum_{x,y} q(x, y)a(x, y) \tag{10}$$

В качестве способа вычисления итогового приоритета кадра p с учетом времени нахождения в очереди t и пригодности кадра w возьмем способ, использующий гауссово ядро [20]:

$$p(w, t) = w^\alpha e^{-\frac{\beta t^2}{\sigma^2}}, \tag{11}$$

где коэффициент α регулирует важность оценки качества, а β и σ контролирует период релевантности кадра относительно времени захвата. На Рис. 3 показаны графики функции $p(w, t)$ для различных значений w и t .

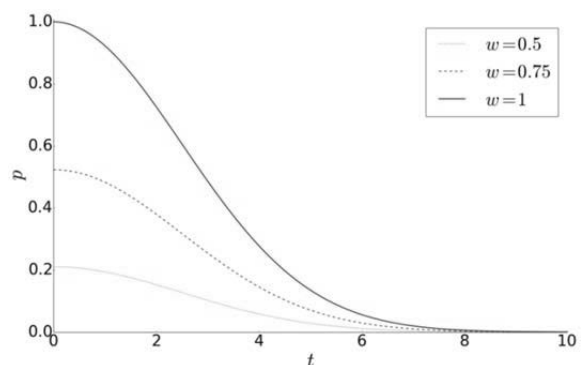


Рис. 3. Графики функции $p(w, t)$ при $\alpha = 2.25, \beta = 0.5, \sigma = 2.5$

Таким образом, с помощью вычисления приоритета $p(w, t)$ становится возможным выбор кадров видеопотока наилучшего качества для распознавания.

5. Эксперименты

Цель экспериментов – показать, что с помощью выбора кадров видеопотока в соответствии с введенной функцией приоритета можно добиться большей точности и скорости распознавания, чем при произвольном выборе кадров. В экспериментах использовалась реализация системы распознавания документов, удостоверяющих личность, в видеопотоке, описанной в [14]. Используемый в экспериментах модуль оценки качества оценивает степень размытия (при движении) и расфокусировки кадра. При этом, также проводится искусственное моделирование размытия во время процесса фокусировки камеры.

5.1. Алгоритм оценки степени размытости кадра

Для оценки размытости и расфокусировки входного кадра I в экспериментах использовался алгоритм, ранее предложенный в работе [3].

Построим карты G_1, \dots, G_4 модулей градиентов изображения I в вертикальном, горизонтальном и двух диагональных направлениях, приведя изображение I к полутоновому со значениями пикселей от 0 до 1. Для текущего изображения модуля градиента G_k вычислим его гистограмму H_k и возьмем верхний $-i$ квантиль r_k . Финальной оценкой w сфокусированности кадра будет квантиль для самого выраженного смаза, т.е. $\min(r_1, \dots, r_4)$, экспоненциально нормированная между (0,1).

Для проверки соответствия вычисленных алгоритмом значений настоящему размытию был проведен эксперимент: часть выборки (100 изображений) была подвергнута Гауссовому размытию с радиусом σ , изменяющимся с шагом в заданном интервале, в гистограмме выбирался $g = 0.05$ -й квантиль. На Рис. 4 показана зависимость усредненных значений w по выборке для фиксированных σ .

Начиная с σ_0 размытие становится заметным, а с σ_1 – текст на документе становится нечитаемым. Видно, что значения вычисленной

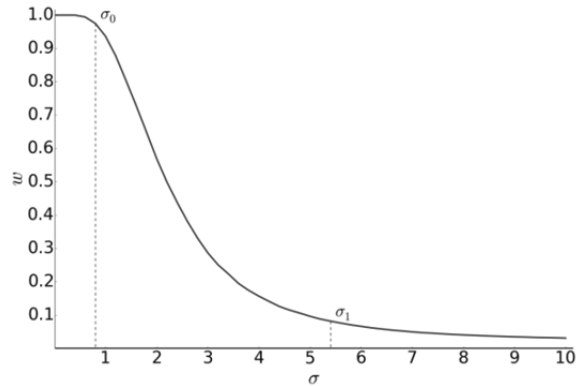


Рис. 4. График зависимости вычисленной оценки фокусировки w от радиуса σ размытия фильтром Гаусса

оценки монотонно убывают при возрастании степени размытия, что доказывает пригодность использования данного алгоритма.

5.2. Моделирование распознавания в процессе автоматической фокусировки камеры

Рассмотрим приложение систем распознавания, при котором расстояние до объекта не известно заранее, что актуально при распознавании с мобильных устройств, веб-камер и т.д. В таких случаях используются алгоритмы автоматического фокусирования камеры на объекте, сканирующие диапазон фокусного расстояния и измеряя некоторый критерий четкости кадра, в конце возвращаясь к его оптимуму [21]. Это осложняет процесс распознавания и вместе с тем является идеальным примером применения выбора наилучших кадров для распознавания, поскольку при автоматической фокусировке с камеры поступают как сильно, так и слабо размытые кадры и важно распознавать именно последние.

Для моделирования размытия в процессе автоматического фокусирования камеры возьмем функцию

$$\sigma(t) = a|\text{sinc}(b(t - t_0))| = \frac{a|\sin(b(t - t_0))|}{b|t - t_0|} \quad (12)$$

зависимости радиуса фильтра Гаусса от времени захвата кадра $t > 0$, график которой для различных коэффициентов a , b и t_0 показан на Рис. 5.

На Рис. 6 показан пример части видеопотока с искусственно подготовленным шаблоном ID-карты Германии, подвергаемый размытию в соответствии с формулой 12.

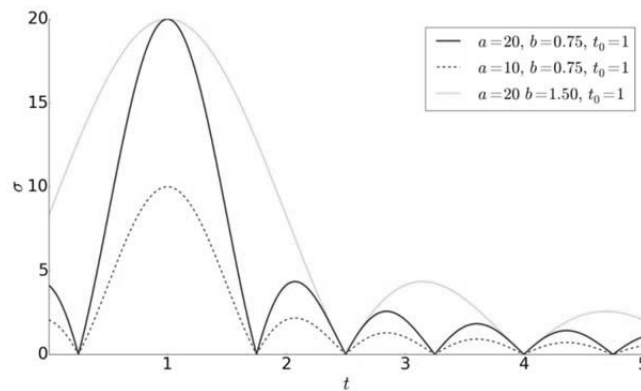


Рис. 5. График функции $\sigma(t)$ моделирования радиуса размытия при автоматической фокусировке



Рис. 6. Пример моделирования процесса автоматической фокусировки для видеопотока

5.3. Экспериментальные результаты

Для экспериментов с введенной ранее системой [14] использовались тестовые выборки из 217 и 39 видеороликов удостоверяющих личность документов (ID-карты Германии, паспорт гражданина РФ) длительностью 5 секунд каждый, снятых с использованием различных мобильных устройств. Для каждого видеоролика была подготовлена идеальная разметка полей документа: номера, имени, фамилии, национальности, даты и места рождения, даты окончания действия документа, а также полной машиночитаемой зоны для ID-карт Германии, и серии, номера и кода подразделения для паспортов РФ, поскольку они присутствуют на обеих страницах разворота. Поле считалось

правильно распознанным в случае полного совпадения возвращенного системой значения с идеальным значением на момент обработки всего видеоролика с документом. Итоговой точностью работы системы на выборке положим среднюю точность распознавания вышеперечисленных полей.

Каждый видеоролик разбивался на кадры с фиксированной частотой $F = 5$ кадров в секунду. Кадры передавались на распознавание с периодом $T = (1, \dots, 8)$ кадров (т.е. с итоговой частотой $\frac{F}{T}$ кадров в секунду), тем самым имитируя продолжительность распознавания одного кадра. Помимо этого, были введены искусственные ограничения на максимальную продолжительность видео $M = (5,3,1)$ секунд для проверки

потенциала скорости распознавания, т.е. точности при фиксированном максимальном времени обработки видеопотока. Эксперимент проводился в двух режимах: с простым выбором каждого T -го кадра ("Fixed") и выбором наилучшего кадра в каждый T -й момент времени в соответствии с предложенным способом вычисления приоритета ("Priority"). При моделировании автоматической фокусировки по формуле 12 использовались значения $a = 20$, $b = 0.75$, а время t_0 пика функции размытия в секундах был равномерно распределено на $[0,3]$.

На Рис. 7 и Рис. 8 показаны полученные результаты точности распознавания видеопотока для исходных (вверху) и размытых (снизу) выборок изображений с различными ограничениями M на максимальную продолжительность видео. По горизонтальной оси отложен период между кадрами T , а по вертикальной – средняя точность ACC распознавания всех полей. Каждый график содержит сравнение методов с фиксированным и приоритетным выбором кадра.

Графики демонстрируют, что точность распознавания видеопотока при использовании

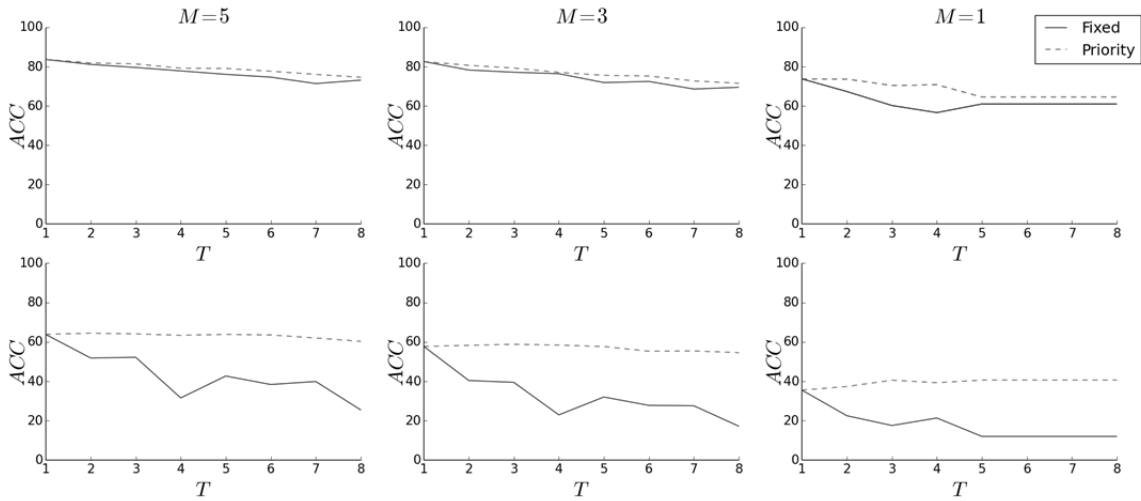


Рис. 7. Графики точности распознавания видеопотока для исходной (сверху) и размытой (снизу) выборки изображений ID карт Германии

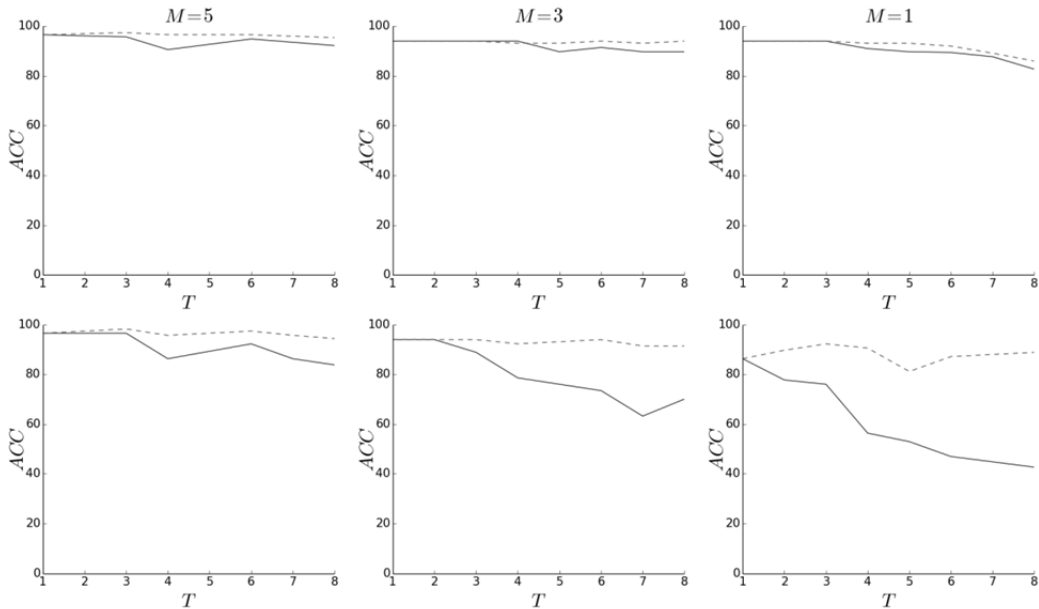


Рис. 8. Графики точности распознавания видеопотока для исходной (сверху) и размытой (снизу) выборки изображений паспортов гражданина РФ

предложенного алгоритма с вычислением приоритета значительно выше при моделировании процесса автоматической фокусировки, чем в случае выбора фиксированных кадров. При этом оно выше и на исходных выборках изображений, не подвергнутым искусственному размытию. Данный эффект наблюдается при всех выбранных ограничениях на максимальную длину видеопотока и для обоих рассматриваемых документов, имеющих разный шрифт и язык.

Таким образом, можно сделать вывод о целесообразности применения предложенного метода выбора наилучших кадров для распознавания в видеопотоке: как в процессе автоматической фокусировки камеры (или в присутствии других искажений), где он дает наибольший прирост точности и скорости работы, так и в обычных условиях.

Заключение

В данной работе предлагается оригинальный подход к построению систем оптического распознавания с встроенными модулями оценки качества входных данных, введены необходимые определения, сущности и связи между ними. В рамках данного подхода рассматривается задача выбора наилучших кадров при распознавании в видеопотоке и предлагается способ вычисления приоритета кадра с учетом его качества и времени его захвата. Экспериментально показано значительное увеличение точности и скорости работы реализации системы распознавания документов, удостоверяющих личность, в видеопотоке, при искусственно моделируемых условиях автоматической фокусировки камеры, приводящим к размытию кадров.

Литература

1. H. Golnabi and A. Asadpour, "Design and application of industrial machine vision systems," *Robotics and Computer-Integrated Manufacturing*, vol. 23, no. 6, pp. 630–637, 2007.
2. O. Russakovsky et al., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
3. K. B. Bulatov and D. V. Polevoy, "Reducing overconfidence in neural networks by dynamic variation of recognizer relevance," in *ECMS*, 2015, pp. 488–491.
4. В. Арлазаров, К. Булатов, С. Карпенко, "Метод определения надежности распознавания в задаче распознавания тисненных символов," *Труды ИСА РАН*, том. 63, №3, с. 117–122, 2013.
5. В. Арлазаров, А. Жуковский, В. Кривцов, Д. Николаев, Д. Полевой, "Анализ особенностей использования стационарных и мобильных малоразмерных цифровых видео камер для распознавания документов," *Информационные технологии и вычислительные системы*, №3, 2014.
6. S. F. Dodge and L. J. Karam, "Understanding how image quality affects deep neural networks," *CoRR*, vol. abs/1604.04004, 2016.
7. P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and objective quality assessment of image: A survey," *CoRR*, vol. abs/1406.7799, 2014.
8. A. Dutta, R. N. J. Veldhuis, and L. J. Spreeuwers, "Predicting face recognition performance using image quality," *CoRR*, vol. abs/1510.07119, 2015.
9. P. Ye and D. Doermann, "Document Image Quality Assessment: A Brief Survey," 2013 12th International Conference on Document Analysis and Recognition, pp. 723–727, 2013.
10. G. J. Awcock and R. Thomas, *Applied image processing*. McGraw-Hill, Inc., 1995.
11. S. T. Bow, *Pattern recognition and image preprocessing*. CRC press, 2002.
12. M. A. Povolotskiy, E. G. Kuznetsova, and T. M. Khanipov, "Russian license plate segmentation based on dynamic time warping," in *European conference on modelling and simulation*, 2017, proceedings., 2017, pp. 285–291.
13. C. Gürel and A. Erden, "Design of a face recognition system," in *Proc. The 15th int. Conference on machine design and production (umtik 2012)*, 2012.
14. K. Bulatov, V. Arlazarov, T. Chernov, O. Slavin, and D. Nikolaev, "Smart IDReader: Document recognition in video stream," in *7th international workshop on camera based document analysis and recognition*, 2017.
15. F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.
16. T. Manzhikov, O. Slavin, I. Faradzhev, and I. Janiszewski, "N-grams algorithm application for the correction of recognition results," *Вестник Российского фонда фундаментальных исследований*, vol. 92, no. 4, pp. 116–123, 2016.
17. J. G. Fiscus, "A post-processing system to yield reduced word error rates: Recognizer output voting error reduction (rover)," in *Automatic speech recognition and understanding*, 1997. Proceedings., 1997, pp. 347–354.
18. A. Minkina, D. Nikolaev, S. Usilin, and V. Kozyrev, "Generalization of the viola-jones method as a decision tree of strong classifiers for real-time object recognition in video stream," in *ICMV 2014, proceedings.*, 2015.
19. X. Wang et al., "Client side filtering of card ocr images." *Google Patents*, dec-2-2014.
20. E. A. Shvets and D. P. Nikolaev, "Complex approach to long-term multi-agent mapping in low dynamic environments," in *8'th international conference on machine vision*, 2015.
21. X. Xu, Y. Wang, J. Tang, X. Zhang, and X. Liu, "Robust automatic focus algorithm for low contrast images using a new contrast measure," *Sensors*, vol. 11, no. 9, pp. 8281–8294, 2011.

Чернов Тимофей Сергеевич. Младший научный сотрудник ИСА ФИЦ ИУ РАН. Окончил НИТУ "МИСиС" в 2013 году. Количество печатных работ: 15. Область научных интересов: системный анализ, системное программирование, распознавание образов, оценка качества изображений. E-mail: chernov.tim@gmail.com

Разумный Никита Павлович. Студент с 2014 года НИТУ "МИСиС". Область научных интересов: анализ изображений. E-mail: agerfirelol@gmail.com

Кожаринов Александр Сергеевич. Доцент НИТУ "МИСиС". Окончил МИСиС в 1992 году. Кандидат технических наук. Количество печатных работ: 24. Область научных интересов: искусственный интеллект, имитационное моделирование, информационные технологии. E-mail: kozharinov@misis.ru

Николаев Дмитрий Петрович. Заместитель директора ИППИ РАН. Окончил МГУ в 2000 году. Кандидат технических наук. Количество печатных работ: более 170. Область научных интересов: машинное зрение, быстрые алгоритмы обработки изображений, распознавание образов. E-mail: dimonstr@iitp.ru

Арлазаров Владимир Викторович. Заведующий лабораторией ИСА ФИЦ ИУ РАН. Окончил НИТУ "МИСиС" в 1999 году. Кандидат технических наук. Количество печатных работ: 26. Область научных интересов: распознавание образов, обработка изображений, системы массового обслуживания. E-mail: vva777@gmail.com

Image quality assessment for video stream recognition systems

T.S. Chernov, N.P. Razumnuy, A.S. Kozharinov, D.P. Nikolaev, V.V. Arlazarov

Recognition and machine vision systems have long been widely used in many disciplines to automate various processes of life and industry. Input images of optical recognition systems can be subjected to a large number of different distortions, especially in uncontrolled or natural shooting conditions, which leads to unpredictable results of recognition systems, making it impossible to assess their reliability. For this reason, it is necessary to perform quality control of the input data of recognition systems, which is facilitated by modern progress in the field of image quality evaluation. In this paper, we investigate the approach to designing optical recognition systems with built-in input image quality estimation modules and feedback, for which the necessary definitions are introduced and a model for describing such systems is constructed. The efficiency of the approach is illustrated by the example of solving the problem of selecting the best frames for recognition in a video stream. Experimental results are presented with the system of recognition of identity documents, showing a significant increase in the accuracy and speed of the system under simulated conditions of automatic camera focusing, leading to blurring of frames.

Keywords: recognition systems, image quality assessment, video stream, blur, defocus, systems analysis.

References

1. H. Golnabi and A. Asadpour, "Design and application of industrial machine vision systems," *Robotics and Computer-Integrated Manufacturing*, vol. 23, no. 6, pp. 630–637, 2007.
2. O. Russakovsky et al., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
3. K. B. Bulatov and D. V. Polevoy, "Reducing overconfidence in neural networks by dynamic variation of recognizer relevance," in *ECMS*, 2015, pp. 488–491.
4. V. Arlazarov, K. Bulatov, and S. Karpenko, "A method for estimating reliability of embossed symbols recognition," *Institute for Systems Analysis RAS Journal* 63(3), 117–122 (2013).
5. V. Arlazarov, A. Zhukovsky, V. Krivtsov, D. Nikolaev, and D. Polevoy, "Analysis of specifics of using stationary and mobile small-sized digital cameras for document recognition," *Information Technologies and Systems* (3) (2014).
6. S. F. Dodge and L. J. Karam, "Understanding how image quality affects deep neural networks," *CoRR*, vol. abs/1604.04004, 2016.
7. P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and objective quality assessment of image: A survey," *CoRR*, vol. abs/1406.7799, 2014.
8. A. Dutta, R. N. J. Veldhuis, and L. J. Spreeuwes, "Predicting face recognition performance using image quality," *CoRR*, vol. abs/1510.07119, 2015.
9. P. Ye and D. Doermann, "Document Image Quality Assessment: A Brief Survey," *2013 12th International Conference on Document Analysis and Recognition*, pp. 723–727, 2013.
10. G. J. Awcock and R. Thomas, *Applied image processing*. McGraw-Hill, Inc., 1995.

11. S. T. Bow, *Pattern recognition and image preprocessing*. CRC press, 2002.
12. M. A. Povolotskiy, E. G. Kuznetsova, and T. M. Khanipov, "Russian license plate segmentation based on dynamic time warping," in *European conference on modelling and simulation*, 2017, proceedings., 2017, pp. 285–291.
13. C. Gürel and A. Erden, "Design of a face recognition system," in *Proc. The 15th int. Conference on machine design and production (umtik 2012)*, 2012.
14. K. Bulatov, V. Arlazarov, T. Chernov, O. Slavin, and D. Nikolaev, "Smart IDReader: Document recognition in video stream," in *7th international workshop on camera based document analysis and recognition*, 2017.
15. F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.
16. T. Manzhikov, O. Slavin, I. Faradzhev, and I. Janiszewski, "N-grams algorithm application for the correction of recognition results," *Вестник Российского фонда фундаментальных исследований*, vol. 92, no. 4, pp. 116–123, 2016.
17. J. G. Fiscus, "A post-processing system to yield reduced word error rates: Recognizer output voting error reduction (rover)," in *Automatic speech recognition and understanding*, 1997. Proceedings., 1997, pp. 347–354.
18. A. Minkina, D. Nikolaev, S. Usilin, and V. Kozyrev, "Generalization of the viola-jones method as a decision tree of strong classifiers for real-time object recognition in video stream," in *ICMV 2014, proceedings.*, 2015.
19. X. Wang et al., "Client side filtering of card ocr images." *Google Patents*, dec~2-2014.
20. E. A. Shvets and D. P. Nikolaev, "Complex approach to long-term multi-agent mapping in low dynamic environments," in *8'th international conference on machine vision*, 2015.
21. X. Xu, Y. Wang, J. Tang, X. Zhang, and X. Liu, "Robust automatic focus algorithm for low contrast images using a new contrast measure," *Sensors*, vol. 11, no. 9, pp. 8281–8294, 2011.

T.S. Chernov. Federal Research Centre "Computer Science and Control Systems" The Institute for Systems Analysis of Russian Academy of Sciences, Moscow, Russia. Researcher. Graduated from NUST "MISIS" in 2013. 15 published papers. Interested in systems analysis and programming, pattern recognition and image quality assessment. E-mail: chernov.tim@gmail.com

N.P. Razumnyy. National University of Science and Technology "MISIS", Moscow, Russia. Student since 2014. Interested in image analysis. E-mail: agerfirelol@gmail.com

A.S. Kozharinov. National University of Science and Technology "MISIS", Moscow, Russia. Senior Lecturer. Graduated from MISIS in 1992. 24 published papers. Interested in artificial intelligence, imitation modeling, information technologies. E-mail: kozharinov@misis.ru

D.P. Nikolaev. Institute for Information Transmission Problems RAS, Moscow, Russia. Deputy director for research. Graduated from Moscow State University in 2000. More than 170 published papers. Interested in machine vision, fast image processing algorithms, pattern recognition. E-mail: dimonstr@iitp.ru

V.V. Arlazarov. Federal Research Centre "Computer Science and Control Systems" The Institute for Systems Analysis of Russian Academy of Sciences, Moscow, Russia. Head of laboratory. Graduated from NUST "MISIS" in 1999. 26 published papers. Interested in pattern recognition, image processing, queuing systems. E-mail: vva777@gmail.com