



**Модуль «Бастион-Паспорт»
Распознавание личных документов**

Руководство оператора

Версия 2.0

Содержание

1	Общие сведения о модуле распознавания.....	3
1.1	Назначение и характеристики.....	3
1.2	Технические требования к рабочей станции	3
1.3	Технические требования к устройству сканирования.....	4
2	Алгоритм работы АПК «Бастион» с применением модуля распознавания.....	4
3	Установка и описание компонентов модуля распознавания.....	5
3.1	Установка модуля распознавания	5
3.2	Установка драйверов Guardant	6
4	Описание интерфейса модуля распознавания	7
4.1	Использование модуля в бюро пропусков АПК «Бастион».....	7
4.2	Панель инструментов	9
4.3	Определение настроек сканирования.....	10
4.4	Подготовка к сканированию и распознаванию	11
4.5	Сканирование и распознавание документа.....	12
4.6	Обработка и верификация результатов	13
4.7	Просмотр протокола работы	15
5	Настройка прав пользователей АПК «Бастион»	15
6	Внутренние настройки модуля распознавания	16
6.1	Внутренняя система распознавания Scanify API.....	16
6.2	Версии системы Scanify API.....	16
6.3	Файл контекстных настроек.....	17

1 Общие сведения о модуле распознавания

1.1 Назначение и характеристики

Модуль «Бастион-Паспорт» представляет собой систему автоматического ввода документов со сканера и их *распознавания* с целью получения значений полей документа. На основании распознанных полей можно создать новую заявку на пропуск и выдать ее, привязав к карте доступа. Система используется в модуле «Бюро пропусков» АПК «Бастион».

Модуль «Бастион-Паспорт» (в дальнейшем, модуль распознавания) используется для ускорения ввода в АПК «Бастион» новых заявок на пропуски на предприятиях:

- с большой интенсивностью входного потока персонала на точках контроля,
- с большим количеством выдаваемых временных и разовых пропусков.

Модуль выполняет распознавание следующих *типов личных документов*:

1. Паспорт гражданина РФ (далее, паспорт).
2. Водительское удостоверение гражданина РФ (далее, права).
3. Заграничный паспорт гражданина РФ (далее, загранпаспорт). Паспорт может быть как образца с биометрическими данными, так и ранее выдаваемого образца без них.
4. Паспорт иностранного гражданина (далее, паспорт ино).

Модуль выполняет следующие *функции*:

- сканирование личного документа с TWAIN-совместимого сканера или фотоаппарата;
- определение типа полученного личного документа (паспорт, права и т.п.);
- распознавание отсканированного документа;
- предоставление пользователю возможности верификации распознанного документа;
- передачу распознанных полей в модуль «Бюро пропусков»;
- создание новой заявки, либо поиск уже имеющегося пропуска по распознанным полям;
- экспорт сканированного образа в файловый архив.

1.2 Технические требования к рабочей станции

Для работы модуля необходимо:

1. Наличие лицензии на модуль «Бастион-Паспорт» в ключе аппаратной защиты Hasp на рабочей станции, где предполагается автоматический ввод документов.
2. Наличие аппаратного ключа защиты Guardant с лицензией на работу внутренней системы распознавания. Разделение на модуль «Бастион-Паспорт» и «Бастион-Паспорт-Про» устранено в версии 2.0 модуля «Бастион-Паспорт».
3. Установленный АПК «Бастион» версии не ниже 1.6.0.1.

4. Установленное на рабочем месте TWAIN-устройство ввода информации:
 1. Сканер (через USB или LPT-интерфейс)
 2. Цифровой фотоаппарат.

1.3 Технические требования к устройству сканирования

Устройство ввода информации – сканер или фотоаппарат – должно:

1. Поддерживать работу в одной из операционных систем Microsoft Windows 98, ME, 2000, XP, 2003 – той, которая установлена на рабочей станции.
2. Формат документа – не меньше А5 (148x210 мм)
3. Разрешение сканирования – 300 dpi (точек на дюйм)
4. Режим сканирования – 24 или 8 бит (цветное или серое изображение)
5. Установленный, настроенный и проверенный на тестовом документе драйвер.

2 Алгоритм работы АПК «Бастион» с применением модуля распознавания

Без использования распознавания алгоритм работы оператора бюро пропусков имеет вид:

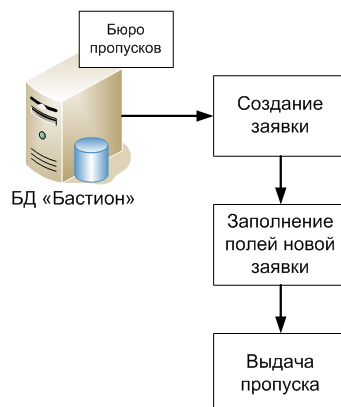


Рис. 1 – Работа оператора бюро пропусков без применения распознавания

При использовании модуля распознавания большинство полей заявки заполняется в автоматическом режиме, при этом в работу оператора добавляется три новых простых операции:

1. Подготовка изображения.
2. Сканирование документа.
3. Верификация полей.

Эти операции призваны:

1. Снизить время заполнения полей новой заявки.
2. Сократить ошибки при печатном вводе большого числа поступающих новых заявок.
3. Вести архив личных документов сотрудников.

С применением распознавания алгоритм работы оператора принимает вид, представленный на Рис. 2. Новые операции показаны пунктирными линиями.

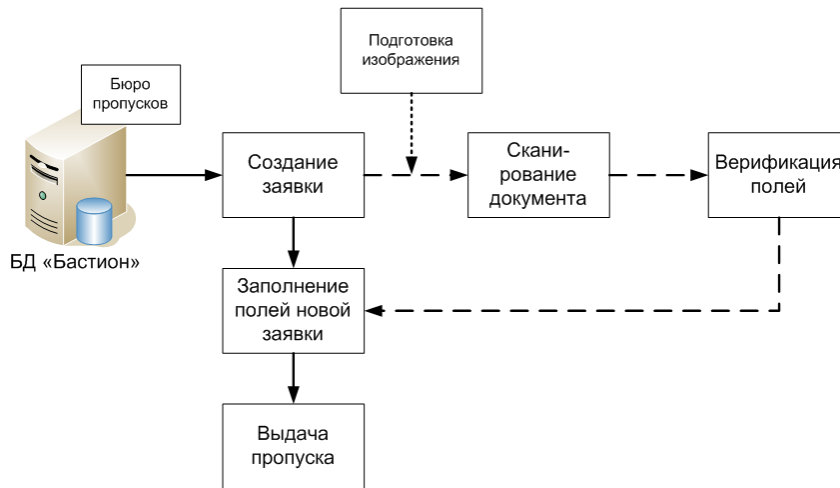


Рис. 2 - Работа оператора бюро пропусков с применением распознавания

3 Установка и описание компонентов модуля распознавания

3.1 Установка модуля распознавания

Для установки модуля распознавания требуется запустить инсталлятор VPassport_setup.exe с диска АПК «Бастион» из папки Passport.

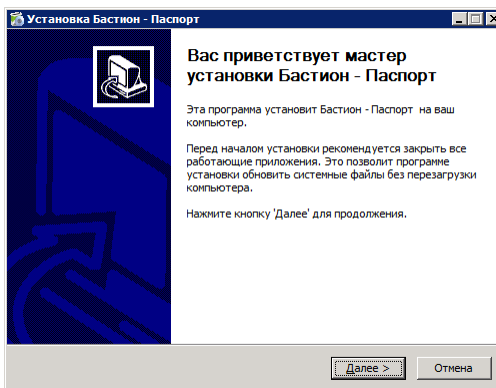


Рис. 3 – Запуск инсталляции модуля распознавания

Далее, необходимо следовать инструкциям инсталлятора и подтвердить установку. После установки будет создана программная группа «Распознавание личных документов» в меню «Пуск \ Программы \ Бастион». Она имеет следующий вид:

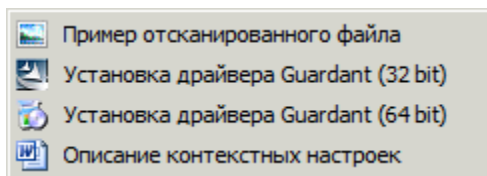


Рис. 4 – Программная группа модуля «Распознавание личных документов»

Сканирование и распознавание документов запускается с помощью модуля «Бюро пропусков» АПК «Бастион» с панели инструментов.

Пункт «Пример сканированного документа» открывает для просмотра файл, на котором можно опробовать работу распознавания. Сделать это можно, открыв этот файл через бюро пропусков в модуле «Распознавание личных документов».

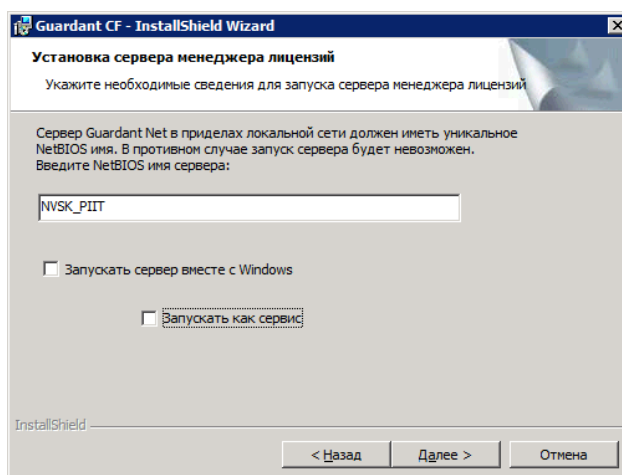
Пункт «Описание контекстных настроек» вызывает руководство, согласно которому для различных видов сканеров можно настроить расположение мест на планшете, куда оператору следует прикладывать документы.

Пункт «Установить драйвера Guardant» является обязательным для выполнения, см. следующий раздел.

3.2 Установка драйверов Guardant

Модуль распознавания работает только с подключенным к компьютеру ЗЕЛЕНЫМ ключом-брелком защиты Guardant Stealth II, который поставляется отдельно от ключа защиты HASP АПК «Бастион». Для опознавания ключа Guardant в системе необходимо:

1. **Извлечь** все ключи из портов USB перед установкой.
2. Запустить ярлык «Пуск \ Программы \ Бастион \ Распознавание личных документов \ Установка драйвера Guardant 32 bit» (или 64 bit в требуемом случае).
3. Установить драйвера Guardant БЕЗ поддержки сетевого варианта, сняв две соответствующие галочки. В противном случае при запуске рабочей станции будет выводиться окно проверки состояния, которое может визуальнo помешать работе АПК «Бастион» и бюро пропусков.



4. **Вставить** ключ в порт USB
5. Разрешить и выполнить установку драйвера Guardant стандартным мастером системы Windows с автоматическим поиском драйвера, который будет отображаться в области системного лотка около часов.

В дальнейшем, при запуске модуля будет отображаться следующее окно:

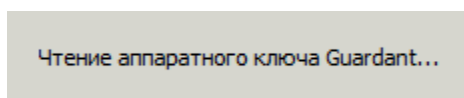


Рис. 5 - Запуск модуля "Бастион-Паспорт"

Следует отметить, что проверка лицензии при чтении аппаратного ключа Guardant занимает некоторое время (порядка 5 секунд), ввиду чего могут возникнуть неудобства в работе. В связи с этим требуется держать модуль распознавания постоянно запущенным в бюро пропусков при большом потоке посетителей.

4 Описание интерфейса модуля распознавания


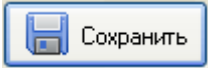


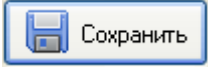
4.1 Использование модуля в бюро пропусков АПК «Бастион»

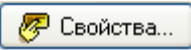

Модуль распознавания может использоваться в двух режимах:


1. **Режим создания заявки.** В этом режиме сначала создается новая заявка, затем сканируется, распознается и верифицируется документ, после чего результаты сохраняются в новую заявку.
2. **Режим создания заявки с поиском.** Оператор сначала сканирует документ, распознает и верифицирует его. Затем, по данным распознавания, АПК «Бастион» выполняет глобальный поиск пропусков. Если не найдены пропуска с распознанными личными данными, то создается новая заявка, и оператор сохраняет ее.

Если пропуска были найдены (сотрудник уже внесен в базу данных), предлагается либо изменить данные какого-либо найденного пропуска, подставив туда распознанные, либо выполнить обычное позиционирование на найденном результате (например, для сдачи пропуска).

Модуль распознавания работает только из АПК «Бастион» и вызывается двумя способами:

1. Для работы в режиме создания заявки:
 - a. Открыть окно «Бюро пропусков».
 - b. Создать новую заявку.
 - c. Нажать на форме «Свойства пропуска» кнопку , которая находится справа от фотографии. При этом откроется форма распознавания.
 - d. Выполнить сканирование и верификацию документа.
 - e. Сохранить документ в АПК «Бастион» кнопкой  «Сохранить».
2. Для работы в режиме создания заявки с поиском:
 - a. Открыть окно «Бюро пропусков».
 - b. Нажать кнопку  из главной панели инструментов. Она располагается рядом с кнопкой  создания новой заявки. Откроется форма распознавания.
 - c. Выполнить сканирование и верификацию документа.
 - d. Сохранить документ в АПК «Бастион» кнопкой  «Сохранить».
 - e. Если пропуска с распознанными личными данными были найдены, откроется форма «Результаты поиска», где можно либо изменить найденный пропуск, согласно

распознанным данным кнопкой , либо перейти к найденному пропуску кнопкой .

Модуль распознавания целесообразно НЕ ЗАКРЫВАТЬ при потоковом использовании, чтобы сократить время работы оператора. Он будет автоматически сворачиваться в панель задач при сохранении в АПК «Бастион». Каждое последующее нажатие кнопки  будет разворачивать модуль, запускать новое сканирование, после чего нужно верифицировать поля документа и снова выполнить сохранение в АПК «Бастион».

В режиме создания заявки с поиском алгоритм действий оператора примет вид на Рис. 6:

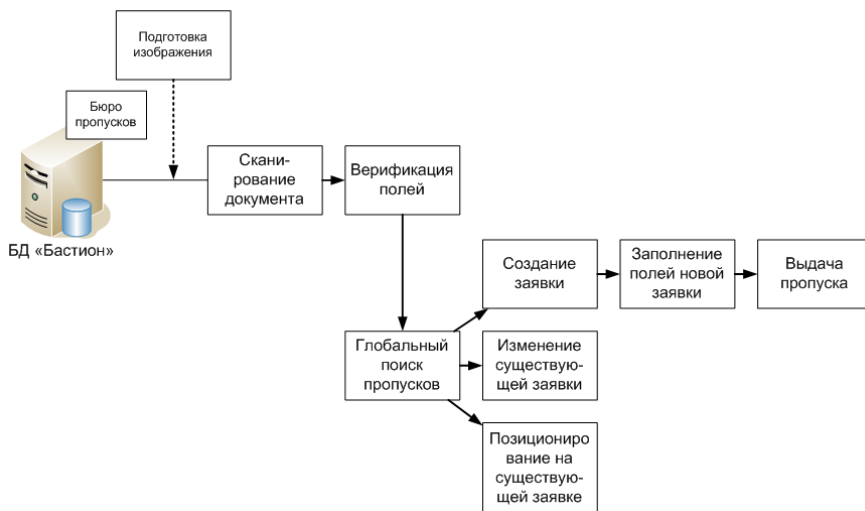


Рис. 6 - Работа оператора бюро пропусков в режиме добавления заявки с поиском

Внешний вид АПК «Бастион» с вызванной формой распознавания приведен на Рис. 7.

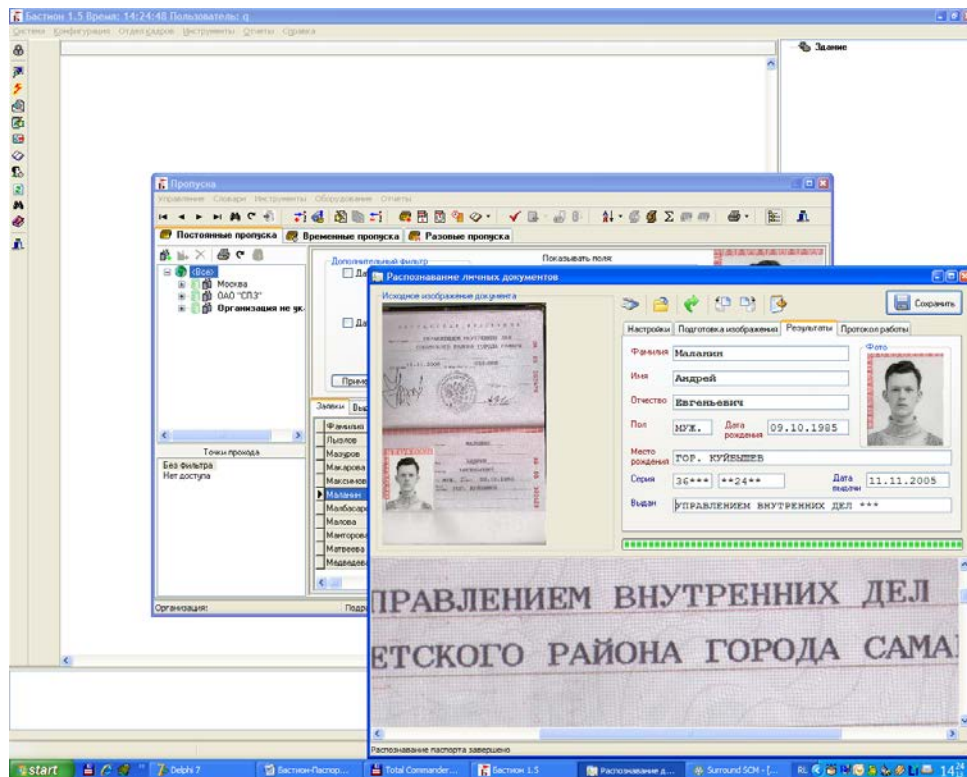


Рис. 7 - Пример вызова модуля распознавания документов в АПК "Бастион"

Общий вид формы распознавания документов приведен на Рис. 8:

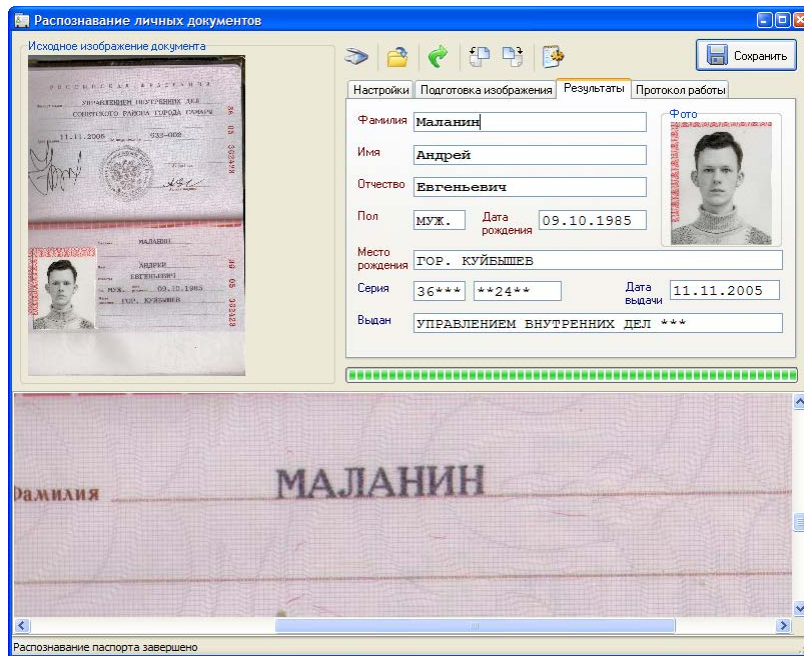


Рис. 8 – Общий вид экранной формы распознавания документов





Экранная форма состоит из четырех основных областей:



1. Панель инструментов (вверху).
2. Изображение отсканированного документа общего вида (слева).
3. Изображение увеличенного сканированного документа для верификации (внизу).
4. Набор из четырех вкладок с возможными действиями пользователя (справа):
 1. Настройки сканирования.
 2. Подготовка изображения к распознаванию.
 3. Результаты распознавания.
 4. Протокол работы.

4.2 Панель инструментов

Панель инструментов экранной формы содержит шесть пунктов.

Таблица 1 - Описание пунктов панели инструментов

Изображение	Наименование	Описание пункта
	Сканировать и распознать (F9)	Немедленно сканирует документ, либо открывает диалог настроек сканера. Затем распознает изображение.
	Открыть и распознать (F5)	Открывает изображение документа из файла, затем автоматически распознает его. Оно должно иметь такую глубину цвета, которая указана на странице «Подготовка изображения». Если это не так, следует изменить настройку глубины цвета и открыть изображение снова.
	Распознать повторно... (Ctrl+R)	Открывает изображение последнего документа, затем повторно распознает его. Этот пункт полезен при несоответствии документа и предварительных настроек.
	Повернуть против	Поворачивает активное изображение против часовой

	часовой стрелки (Ctrl+1)	стрелки
	Повернуть по часовой стрелке (Ctrl+2)	Поворачивает активное изображение по часовой стрелке
	Настройки положения документов на странице	Открывает ini-файл, в котором указываются настройки препроцессинга, расположение и размеры форм документов. Доступен только с соответствующими правами доступа АПК «Бастион»

4.3 Определение настроек сканирования

Вкладка настроек сканирования приведена на Рис. 9:

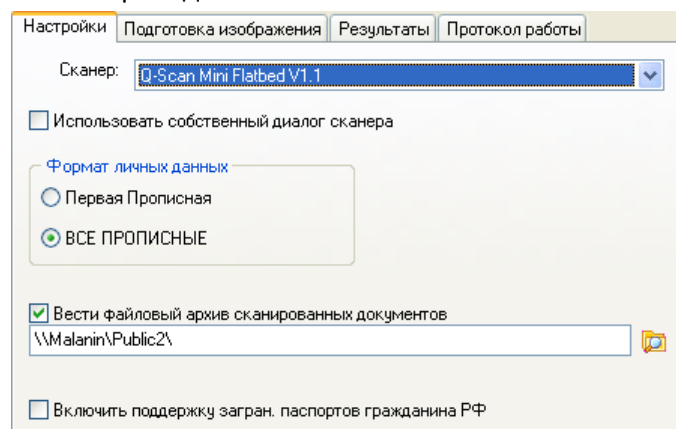


Рис. 9 - Вкладка настроек сканирования

В поле «Сканер» оператор должен выбрать TWAIN-источник, который будет использоваться для сканирования документов.

При потоковом сканировании документов в бюро пропусков оператору может понадобиться отобразить собственный диалог настроек сканера, например, чтобы вручную указать область сканирования. Это может возникнуть, если личный документ вставлен в бумажник человека, труднодоступен, находится на ксерокопии и т.п. Чтобы выполнить это, оператор должен отметить флаг «Использовать собственный диалог сканера».

В группе «Формат личных данных» администратор может указать формат регистра, в котором данные из модуля распознавания будут передаваться в «Бастион». Возможны два варианта, которые необходимо выбрать согласно внутреннему регламенту организации:

- Маланин Андрей Евгеньевич,
- МАЛАНИН АНДРЕЙ ЕВГЕНЬЕВИЧ.

Администратор системы может указать, вести ли файловый архив документов, т.е. сохранять ли все изображения документов согласно распознанным личным данным в формате jpg в определенной папке на диске.

Также, здесь указывается, включена ли поддержка распознавания заграничных паспортов в аппаратном ключе защиты Guardant, а также отображать ли в списке типов документов вариант заграничного паспорта.

4.4 Подготовка к сканированию и распознаванию

Вкладка подготовки к сканированию приведена на Рис. 10:

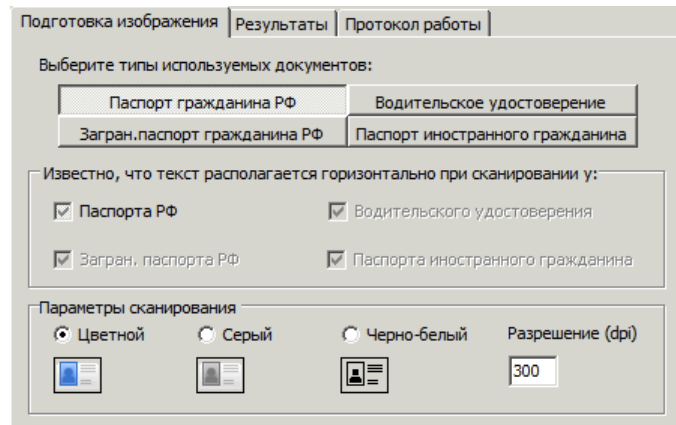


Рис. 10 - Вкладка подготовки изображения

На этой вкладке необходимо указать тип распознаваемого документа, либо установить его автоматическое определение.

Нажатые кнопки в группе типов определяют документ, который может быть распознан.

Если оператор сканирует только паспорта, целесообразно нажать только первую кнопку.

Если оператор сканирует несколько видов документов, которые могут встречаться с одинаковой частотой, то рекомендуется держать соответствующие кнопки нажатыми.

Если оператор располагает временем для принятия решения о виде документа, который ему ПРЕДСТОИТ отсканировать и распознать, ему целесообразно держать нажатой только ту кнопку, которая соответствует конкретному виду документа.

Оператору ПЕРЕД началом сканирования целесообразно определить, как располагается документ в устройстве, с целью экономии времени распознавания.

Он может быть повернут на 90 градусов (система автоматически определяет – по или против часовой стрелки) и может быть развернут на 180 градусов.

Для **паспорта, загранпаспорта и паспорта иностранного гражданина** нормальным считается положение, когда текст читается вертикально. При этом длинная сторона разворота паспорта (18 см) лежит по ширине формата А4, а узкая сторона (13 см) - по высоте формата. Такое положение уменьшает время одного прохода сканера и является предпочтительным. Однако все возможные положения паспорта допустимы. Система автоматически выполнит поворот после успешного распознавания, если потребуется.

Однако, если ПЕРЕД сканированием заведомо известно, как лежит паспорт в сканере, то можно установить или снять флаг «Текст располагается горизонтально у ...». Это может дополнительно сэкономить время распознавания.

- Если флаг установлен, то текст на документе должен читаться горизонтально, слева направо, система не будет пытаться его повернуть на 90 градусов.
- Если флаг снят, то текст на документе должен читаться вертикально, снизу вверх или сверху вниз, система не будет пытаться его повернуть на 90 градусов.
- Если флаг имеет неопределенное, серое или заштрихованное значение, то система будет выполнять автоматический поворот, что может немного замедлить распознавание.

Для **водительского удостоверения** нормальным считается положение, когда текст читается горизонтально. При этом длинная сторона карточки удостоверения должна лежать по ширине формата А4. Определение направления текста осуществляется аналогично.

Также здесь можно указать разрешение и глубину цвета, в которых изображение документа будет получено с устройства. Для устойчивой работы и повышения вероятности распознавания всех типов документов рекомендуется использовать разрешение *300 dpi цветного изображения*.

4.5 Сканирование и распознавание документа

Для распознавания документа необходимо нажать соответствующую кнопку:

- сканировать документ или
- открыть существующий скан из файла.

После того, как изображение будет получено, начнется процесс распознавания. Если система сможет определить форму, которой соответствует документ, распознавание завершится так или иначе с различной долей распознанных полей, в зависимости от качества документа и настроек сканирования. Если форма не будет определена, модуль выдаст сообщение о невозможности распознавания (Рис. 11):

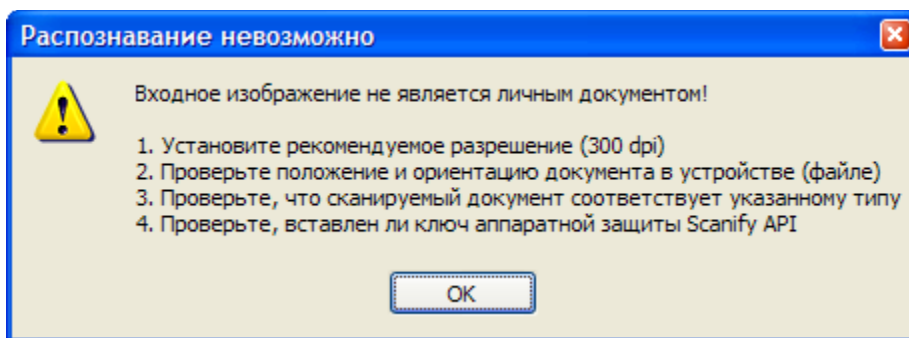


Рис. 11 - Сообщение о невозможности распознавания полученного изображения

Это может произойти в случаях, если:

- Изображение имеет плохое качество:
 1. документ заполнен органами выдачи от руки;
 2. документ имеет защитную пленку, вложен в бумажник;
 3. документ ламинирован и имеет блики;
 4. документ лежит не под прямым углом в устройстве;
 5. часть полей документа (например, номер и серия паспорта) закрыты корочками, визитницей или краями сканера.

Решение: попытаться улучшить качество документа – извлечь из бумажника, положить ровно в устройство, отогнуть загораживающие детали. Если это не помогает или если документ заполнен от руки, заполнить поля заявки вручную, отказавшись от распознавания.

- Установлено разрешение сканирования, отличное от рекомендуемого – 300 dpi. *Решение:* установить разрешение 300 dpi.

- Ориентация документа в устройстве не соответствует флагам на странице «Подготовка изображения». *Решение:* установить флаги поворота (см. п.0) либо развернуть документ в устройстве и повторить сканирование.
- Тип документа указан явно на панели инструментов и не соответствует документу в устройстве. *Решение:* установить верный тип документа из списка или выбрать автоматическое определение.
- Недоступен, отсутствует или был извлечен ключ аппаратной защиты Guardant. *Решение:* проверить его доступность службой диагностики Guardant, затем перезапустить форму распознавания и повторно выполнить сканирование.

4.6 Обработка и верификация результатов

Результаты распознавания документов отображаются на вкладке «Результаты». Для различных документов она приведена на Рис. 12 и Рис. 13.

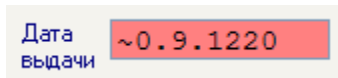
Рис. 12 – Вкладка результатов распознавания паспорта

Рис. 13 – Вкладка результатов распознавания водительского удостоверения

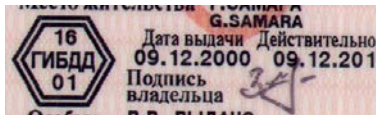
Из рисунка видно, что во втором случае часть полей отмечена *красным* цветом. Это поля, распознанные с ошибкой. Если поля отмечены *желтым* цветом – они распознаны с сомнением.

При этом оператор должен провести верификацию полей. Для этого ему необходимо:

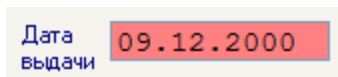
1. Переместить курсор в ошибочно или сомнительно распознанное поле, например:



2. Найти фрагмент сканированного изображения для указанного поля в окне верификации внизу экранной формы. Этот фрагмент автоматически откроется после перемещения курсора в поле ввода, например:



3. Ввести в поле ввода правильное значение поля на основании оригинального изображения, например:



4. Выполнить пункты 1-3 для всех ошибочно и неуверенно распознанных полей.



5. Подтвердить итоговые результаты распознавания кнопкой

Возможна ситуация, когда поле будет распознано с ошибкой, но не будет помечено, как ошибочное или сомнительное. Поэтому, для достижения лучших результатов следует проверять правильность всех полей.

Если все поля были распознаны корректно, оператору только требуется подтвердить результаты распознавания.

После верификации результаты распознавания автоматически появятся в окне новой заявки в АПК «Бастион», как показано на Рис. 14 и Рис. 15.

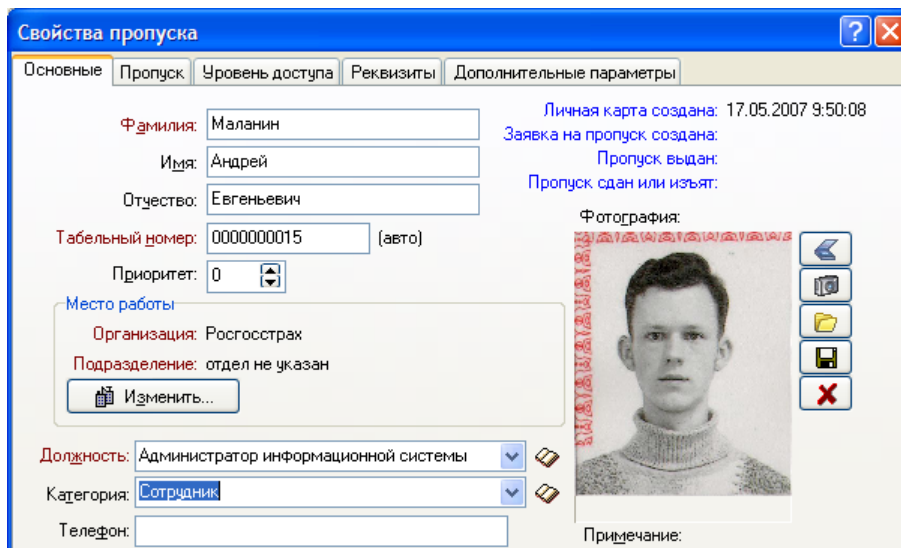


Рис. 14 - Результаты распознавания личных данных в окне новой заявки

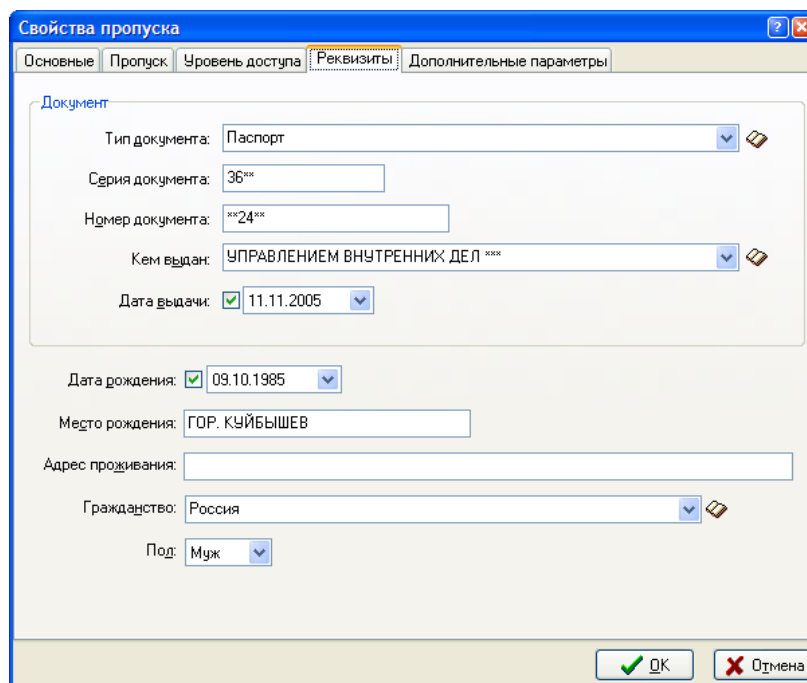


Рис. 15 - Результаты распознавания реквизитов в окне новой заявки

Создание заявки можно подтвердить нажатием кнопки ОК.

4.7 Просмотр протокола работы

В ходе потокового ввода личных документов системой ведется протокол работы распознавания. Он доступен на вкладке «Протокол работы», как на Рис. 15 (в нормальном режиме работы).

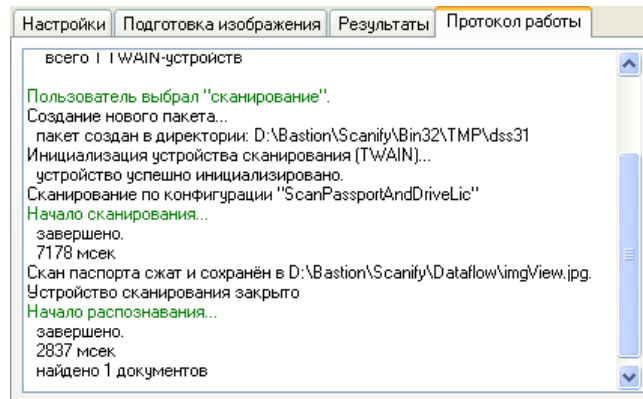


Рис. 16 - Вкладка и пример протокола работы

Если возникают ошибки в модуле распознавания, эта вкладка автоматически открывается и сигнализирует о возникшей проблеме:

1. Недоступность сканера. Сообщение об ошибке имеет вид: «*Ошибка инициализации устройства TWAIN*».
2. Несоответствие формата открываемого изображения указанной глубине цвета. Сообщение об ошибке имеет вид: «*Отсканировано 0 изображений*».

5 Настройка прав пользователей АПК «Бастион»

В модуле предусмотрены возможности разграничения прав пользователей к функциям настроек сканирования и опций форм документов. Для настройки прав пользователей следует выбрать пункт меню «Конфигурация \ Полномочия пользователей» и перейти на страницу «Бюро пропусков» (Рис. 17).

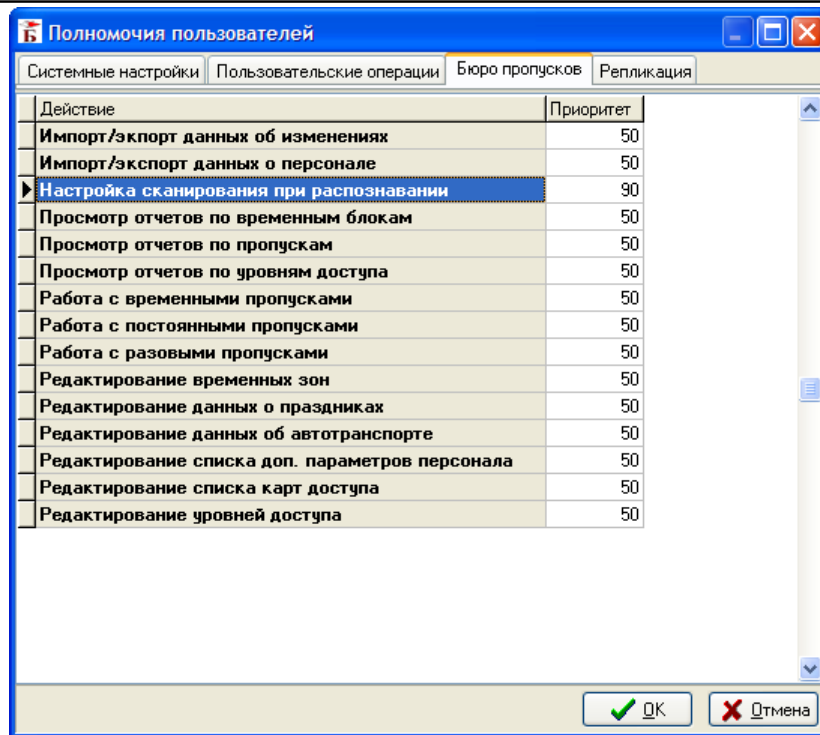


Рис. 17. Настройка полномочий пользователей для системы репликации

Настройка сканирования при распознавании определяет, может ли пользователь с заданным уровнем полномочий работать с вкладкой «Настройки» формы распознавания:

- Выбирать сканер, используемый для получения изображения.
- Отображать ли каждый раз собственный диалог настроек сканера.
- Определять, вести ли файловый архив сканированных документов.
- Изменять конфигурацию форм документов и настройки препроцессинга.

6 Внутренние настройки модуля распознавания

6.1 Внутренняя система распознавания Scanify API

Подсистема распознавания модуля «Бастион-Паспорт» является разработкой компании Cognitive и защищена законом об авторском праве. Ее исполняемые и служебные файлы находятся в подпапке Scanify папки, куда установлен АПК «Бастион» (Bastion). Если папки Scanify не существует, значит, в системе отсутствует возможность использования системы распознавания.

6.2 Версии системы Scanify API

1. Версия 1.4 Scanify API была стартовой при разработке модуля.
2. Версия 1.8 была ускорена на 10-15%.
3. Версия 2.0 ускорена на 40-50%. Кроме этого:
 - Реализовано снятие помех ламинатной пленки и красной рамки по периметру российского паспорта. Раньше такие паспорта вообще не воспринимались системой Scanify как личный документ.

- Реализовано снятие голографических помех на фотографии российского паспорта. Помехи не мешают работе системы, но остаются на фотографии после распознавания. Также периодически могут возникать неточно распознанные участки в местах появления помех-«орлов», если они располагаются на областях текста.
- Добавлена возможность распознавания биометрических загранпаспортов. При этом используется перекрестная проверка MRZ (Machine Readable Zone) и данные их визуальных полей.
- Установщик системы теперь представляет собой один исполняемый файл `bpassport_install.exe`. Это было сделано для реализации возможности установки модуля на разных версиях ОС Microsoft Windows.

6.3 Файл контекстных настроек

Файл контекстных настроек `scapi.ini` располагается в папке `Bastion\Scanify\Bin32`.

Секции параметров в `ini`-файле контекстных настроек условно разделены на 3 категории, соответствующие этапам обработки документа (документов) в процессе выполнения:

- *сканирование изображения*: получение изображения со сканера + разрезание полученного изображения на геометрические зоны документа (прямоугольники);
- *препроцессинг зоны документа*: сюда относятся графические фильтры, применяемые к зонам изображения, полученным на предыдущем этапе, такие как цветоподавление, бинаризация и т.п.;
- *распознавание зоны документа*: на этом этапе Scanify пытается приложить формы документа из набора имеющихся (набор задаётся в параметрах зоны) к зоне, иначе говоря, привязать форму к изображению; в случае успеха происходит собственно распознавание значений полей в соответствующих прямоугольниках формы.

В процессе работы Scanify каждую вырезанную зону помещает в отдельный графический файл, доступ к которому можно получить через описатель и соответствующие функции, а также через временную папку распознавания.

Прямоугольник зоны сканирования задаётся четвёркой чисел, разделённых пробелом, которые последовательно, слева направо обозначают *x*-координату левого угла, *y*-координату левого угла, ширину и высоту. При этом каждое число справа снабжено единицей измерения. В качестве единиц измерения допускаются миллиметры (mm), дюймы (inch) и пиксели (pix).

Далее идут атрибуты зон вырезаемых документов. Сюда относятся прямоугольник зоны и тип зоны для каждого документа. Атрибуты зон вырезания определяют разветвление алгоритма с обработки одного изображения на обработку нескольких документов. Помимо этого имя зоны вырезания играет роль ссылки на соответствующую секцию препроцессинга.

Секция распознавания содержит перечень путей к формам документа (если их несколько, то Scanify будет пытаться их привязать в порядке их описания, пока не найдётся та, которая привяжется).

Более детальное описание приведено в файле описания контекстных настроек системы Scanify.