

# Основные этапы перевода архивных документов в электронный вид в НИИ "Полюс" им. М.Ф. Стельмаха



**Ф**едеральное государственное унитарное предприятие научно-исследовательский институт "Полюс" им. М.Ф. Стельмаха (далее НИИ "Полюс") было основано в 1962 г.

НИИ "Полюс" специализируется в следующих направлениях квантовой электроники:

- твердотельные лазеры на активированных кристаллах и приборы на их основе;
- полупроводниковые лазеры всех типов;
- лазерные гироскопы на газовых лазерах и навигационные приборы на их основе;
- активные, электрооптические и нелинейные кристаллы для лазеров;
- полупроводниковые структуры для полупроводниковых лазеров и фотоприемных устройств;
- лазерные медицинские и технологические установки.

В настоящее время НИИ "Полюс" является крупнейшим в России научно-производственным центром в области квантовой электроники. Его сотрудники создали конструкторскую и технологическую документацию на уникальные изделия. В настоящее время практически вся эта техническая документация находится в архиве НИИ "Полюс" на бумажных носителях.

Для решения задач перевода архивной технической документации с бумажных носителей в электронный вид в октябре 2008 г. в НИИ "Полюс" была создана специальная лаборатория электронной документации.

## Выбор аппаратного и программного обеспечения

На российском рынке аппаратного и программного обеспечения для решения задач перевода технической документации с бумажных носителей в электронный вид и создания электронного архива представлено более пяти компаний, реализующих в этих вопросах комплексный подход. После анализа ряда предложений специалисты службы главного технолога НИИ "Полюс" выбрали ЗАО "СиСофт".

В качестве аппаратного обеспечения для сканирования, кроме обычных планшетных, был приобретен широкоформатный сканер Contex Toucan G25<sup>®</sup> – этот аппарат полностью удовлетворяет потребности в сканировании вплоть до формата A0. Кроме того, было приобретено программное обеспечение, разработанное компанией CSoft Development – программы серии Raster Arts (RasterID и Spotlight), а сотрудники НИИ "Полюс" прошли курс обучения работе с ними в ЗАО "СиСофт".

Хотелось бы отдельно отметить эффективность этого обучения. Преподаватели ЗАО "СиСофт" обладают высокой квалификацией и, несомненно, являются мастерами своего дела. В результате обучения сотрудники НИИ "Полюс" приобрели знания, необходимые для самостоятельной работы.

## Сканирование документов

Перевод документов с бумажного носителя в электронный вид начинается со сканирования. Сканирование в программах серии Raster Arts производится с помощью модуля WiseScan, который обеспечивает прямую поддержку сканеров компании Contex, что позволяет в открываемом окне диалога (рис. 1) выполнять все необходимые настройки сканирования.

Качество получаемых растровых изображений зависит от типа (бумага различного качества, калька, ватман, синька, пленка) и состояния бумажного носителя, аппаратного обеспечения (профессиональное или общего назначения) и настроек параметров сканирования.

## Настройка параметров сканирования

Перед сканированием необходимо определиться со значениями трех основных параметров для получаемого растрового изображения – глубины цвета (черно-белый, серый или цветной), физического размера и разрешения.

При сканировании в черно-белом режиме информация отображается всего одним цветом, а линии на чертеже могут быть различной яркости. Вследствие этого, в процессе предварительного сканирования в окне диалога (рис. 2) необходимо выполнить настройку дополнительного параметра – порога сканирования. В результате сканирования точки на

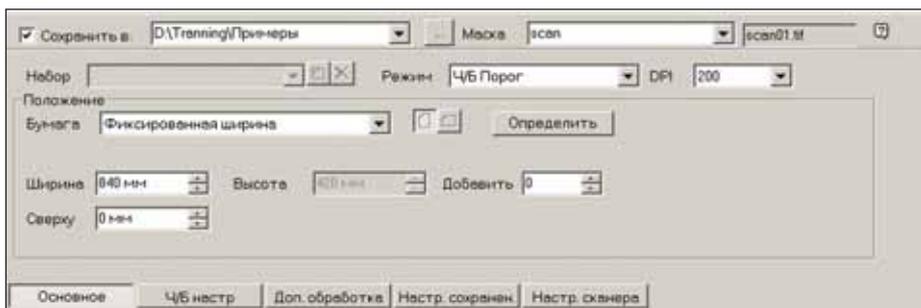


Рис. 1. Окно диалога модуля WiseScan

чертеже, имеющие значения яркости ниже заданного, преобразуются на полученном растровом изображении в линии, текст и другие объекты информации, а выше заданного – образуют фон изображения (рис. 3).

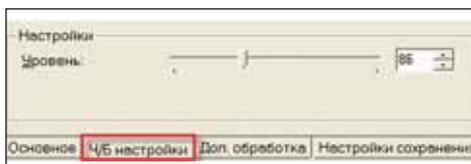


Рис. 2. Настройки порога сканирования

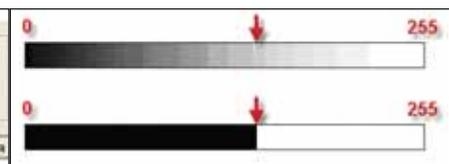


Рис. 3. Схема задания порога сканирования

### Преобразование цветного или полутонового изображения в монохромное

При сканировании документов плохого качества в черно-белом режиме иногда не удается подобрать порог сканирования, чтобы линии были четкими и сплошными: либо часть информации на изображении теряется (рис. 4), либо линии скрываются под полученным растровым мусором (рис. 5).

В тех случаях, когда трудно подобрать порог сканирования в черно-белом режиме, документ можно сканировать в полутоновом или цветном режиме (рис. 6), а затем, используя различные команды перевода цветного изображения в монохромное (бинаризации по порогу и диапазону, по яркости и цветности, адаптивную бинаризацию, разделение по цветам), добиться более качественного результата (рис. 7).

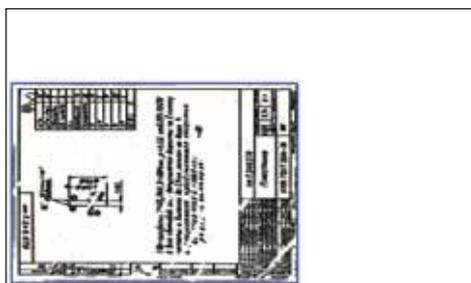


Рис. 8. Задание размера при сканировании



Рис. 9. Документ после обработки

### Задание размера растра

Физический размер изображения, получаемого в результате сканирования, должен быть задан так, чтобы сканируемый документ полностью умещался на созданном растре. Если подобранный комплект состоит из чертежей разного формата, то в качестве параметра можно задать автоматический подбор формата растрового изображения или выбрать из списка стандартный формат, размер ко-

торого больше формата самого большого сканируемого документа из комплекта (рис. 8).

Повернуть, обрезать растровое изображение, получить небольшие поля, необходимые для дальнейшей обработки (рис. 9), можно, сформировав командный файл, который позволит выполнить эти процедуры одновременно либо в процессе сканирования, либо позже в пакетном режиме.

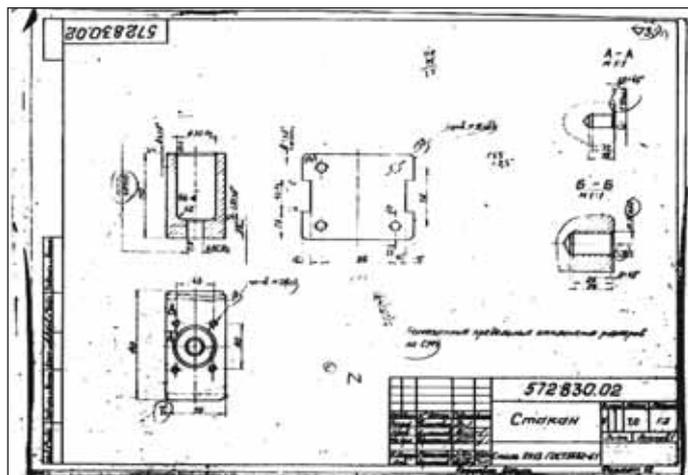


Рис. 4. Порог сканирования низкий

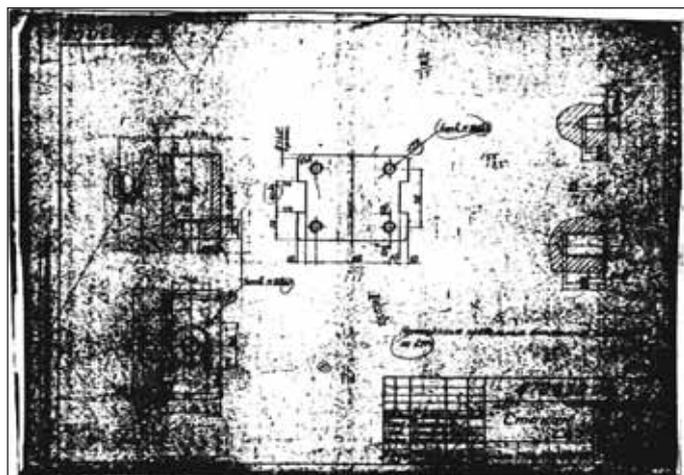


Рис. 5. Порог сканирования высокий

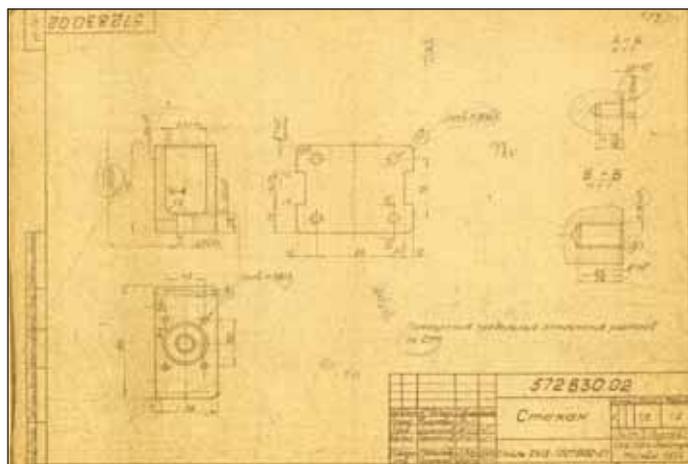


Рис. 6. Чертеж сканирован в цветном режиме

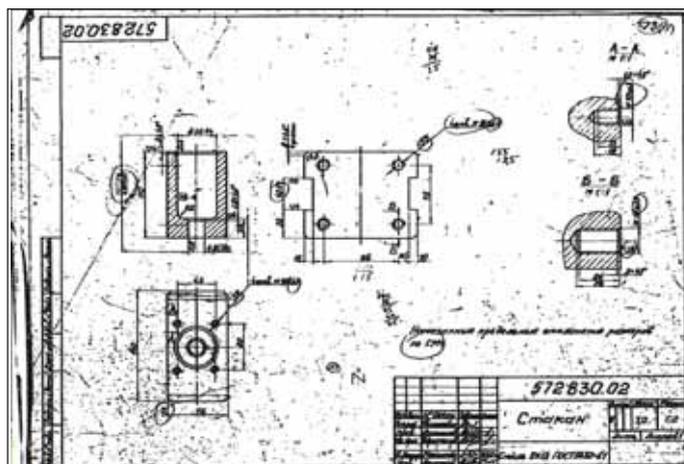


Рис. 7. Чертеж после адаптивной бинаризации

Монохромный без компрессии	2127 кб
Монохромный с компрессией	64 кб
Цветной без компрессии	50991 кб
Цветной с компрессией	774 кб

Таблица 1. Чертеж формата A4 (синька)

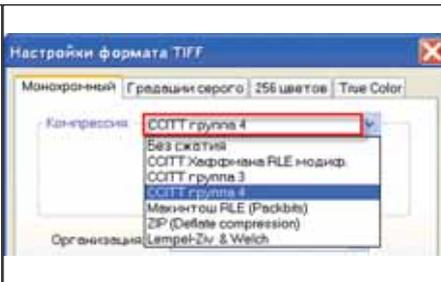


Рис. 10. Настройки параметров формата TIFF

### Устранение геометрических искажений растрового изображения

При сканировании, печати и тиражировании документов происходит искажение изображений, в том числе и их произвольное масштабирование. Геометрические искажения растровых изображений можно устранить с помощью коррекции по 4 точкам или калибровки. В результате выполнения этих команд устраняются различные виды искажений изображения, а размеры элементов на документе становятся равными заданным.

Если истинные размеры чертежа больше, чем размер сканированного документа, то часть изображения за пределами растра после коррекции пропадает (рис. 11). Чтобы этого не произошло, необходимо увеличить размер сканированного изображения, то есть добавить поля, как это было показано в предыдущем примере. Тогда после коррекции по 4 точкам все элементы изображения будут размещены на растре и информация не потеряется (рис. 12).

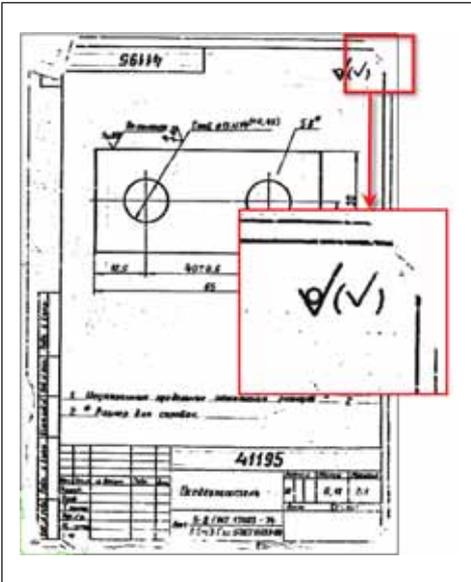


Рис. 13. Углы рамок отсутствуют

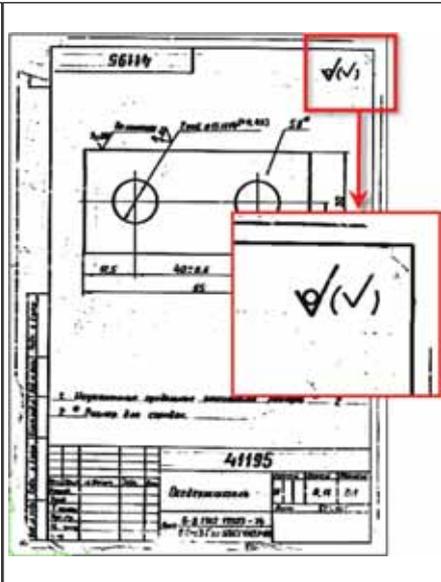


Рис. 14. Углы внутренней рамки восстановлены

### Разрешение растрового изображения

Для дальнейшей корректной обработки изображения рекомендуется сканировать документы с разрешением 300-400 dpi. Задание разрешения меньшего значения приводит к ухудшению качества изображения, а увеличение разрешения значительно увеличивает объем файла, в котором будет сохранен документ.

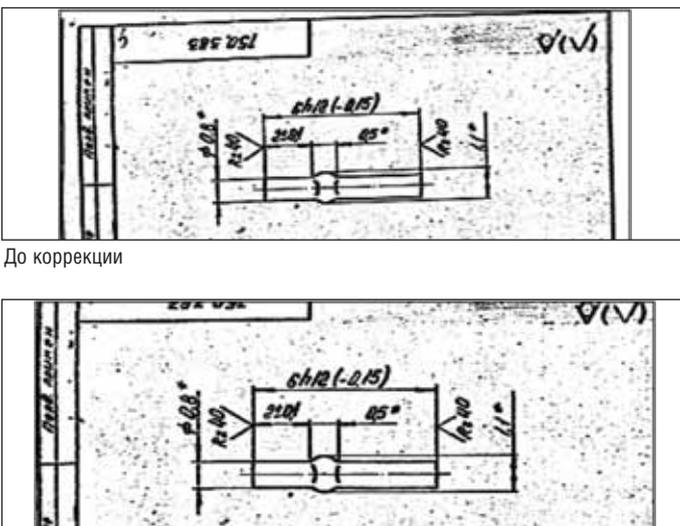
Уменьшить объем файла и при этом сохранить качество растрового изображения можно, используя различные способы компрессии (см. табл. 1), которые задаются в окне диалога (рис. 10). Для сохранения монохромных изображений рекомендуется использовать формат TIFF с компрессией CCITT группа 4, что значительно уменьшает объем файла без потери информации.

### Восстановление углов рамок

У архивных документов чаще всего страдают углы рамок (рис. 13). Если нужно выполнить коррекцию изображения по 4 углам рамки, то прежде следует восстановить углы внешней или внутренней рамки, например с помощью команды *Продлить до пересечения* (рис. 14), а затем выполнить коррекцию по полученным точкам.

### Калибровка изображения

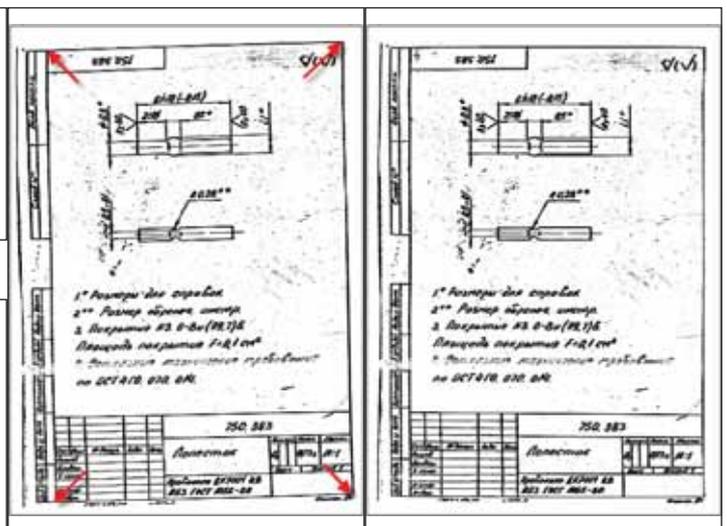
Если линии рамок и отрезков на чертеже не прямые, коррекция по 4 точкам не помогает. Чертеж остается искаженным (рис. 15). В подобных случаях гео-



До коррекции

После коррекции

Рис. 11. Растровое изображение без полей



До коррекции

После коррекции

Рис. 12. Растровое изображение с полями

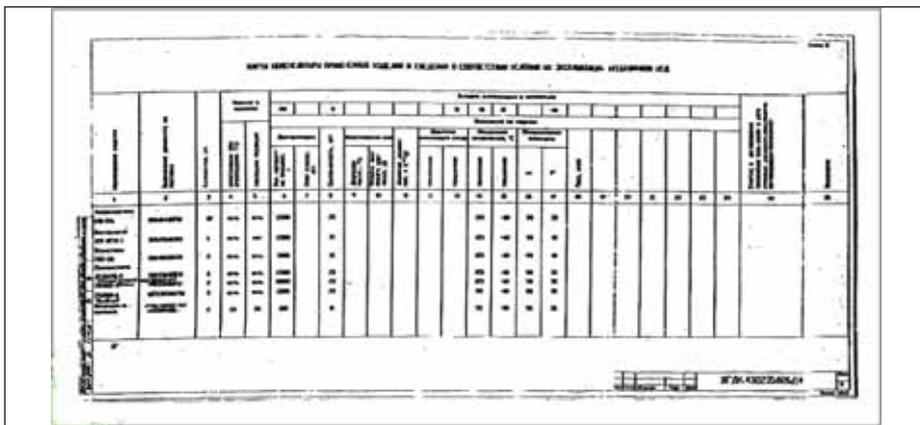


Рис. 15. Лист после коррекции по 4 точкам

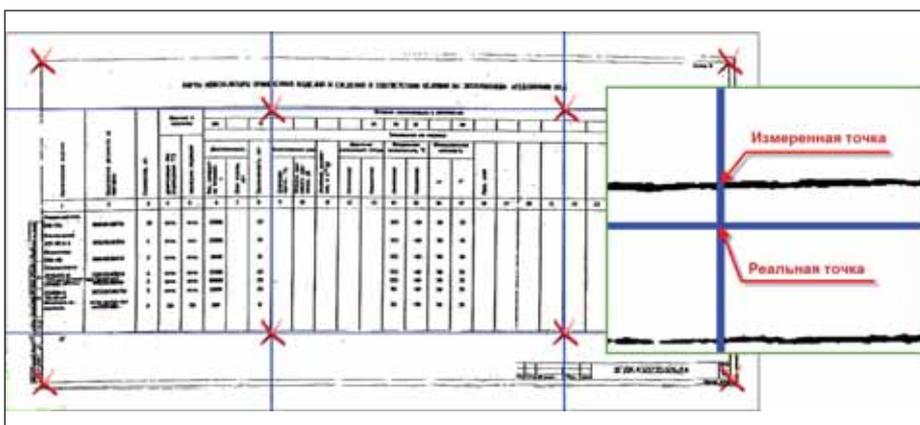


Рис. 16. Задание реальных точек

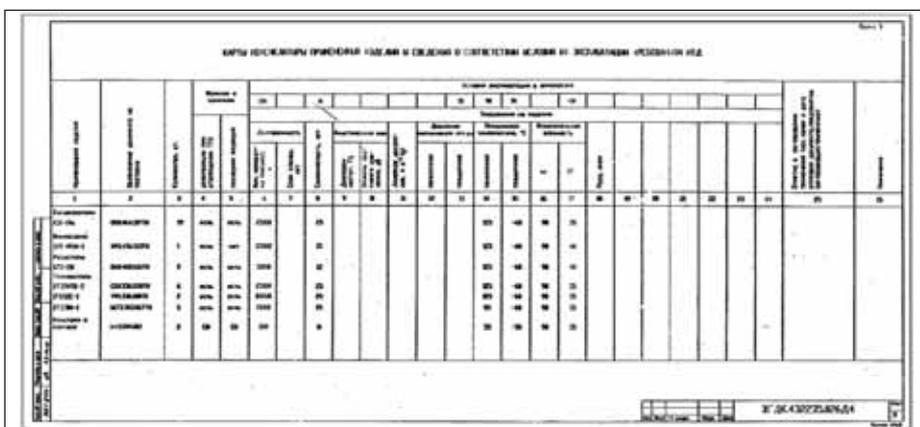


Рис. 17. Документ после окончательной обработки

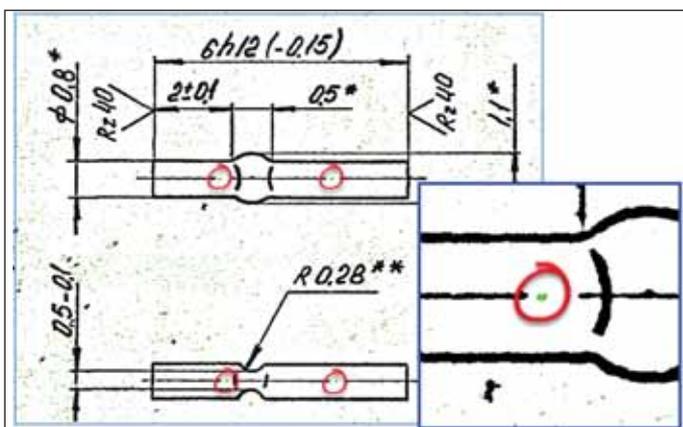


Рис. 18. Разделение объектов по размеру

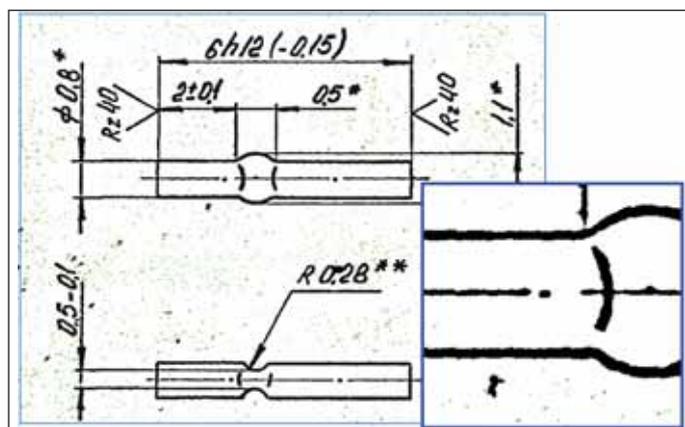


Рис. 19. Возврат элементов на основной чертеж

метрические искажения устраняются командой *Калибровать*.

Перед запуском команды нужно создать (или выбрать) шаблон, по которому будет калиброваться растровое изображение. В качестве шаблона можно использовать несколько линий, точки пересечения которых имеют положение с известными координатами. Такие точки называются "реальными" (рис. 16).

Реальным точкам можно найти на растровом изображении соответствующие точки, которые называются "измеренными". В качестве параметров задаются значения калибровочных пар, то есть координаты реальных и измеренных точек. Координаты можно задавать с помощью числовых значений или указывать на изображении. В процессе калибровки выполняется перерасчет всех точек изображения, в результате которого измеренные точки должны совпасть с реальными, а элементы документа – выровняться (рис. 17).

### Удаление растрового мусора

Удаление мусора и ненужной информации выполняется с помощью одноименного фильтра, разделения объектов по размеру и различными способами выбора.

При удалении растрового мусора фильтром вместе с мусором часто удаляются и некоторые элементы чертежа, такие как точки в осевых линиях, точки в текстах технических указаний и полях штампов, следы некачественных линий. Если элементы чертежа соразмерны с растровым мусором и очистка фильтром не дает желаемого результата, для удаления растрового мусора лучше воспользоваться разделением объектов по размеру. В этом случае объекты, имеющие размеры в заданном диапазоне, не удаляются, а переносятся на другой слой (рис. 18). Элементы чертежа, ошибочно перенесенные вместе с мусором после разделения, можно выбрать и вернуть на основной чертеж (рис. 19).

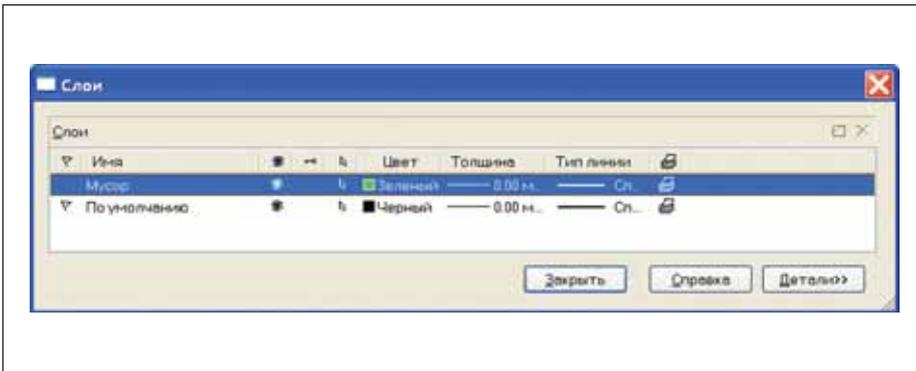


Рис. 20. Отображение слоев после разделения по размеру

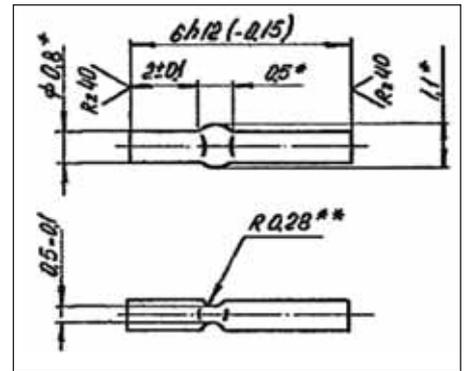


Рис. 21. Результат очистки фона

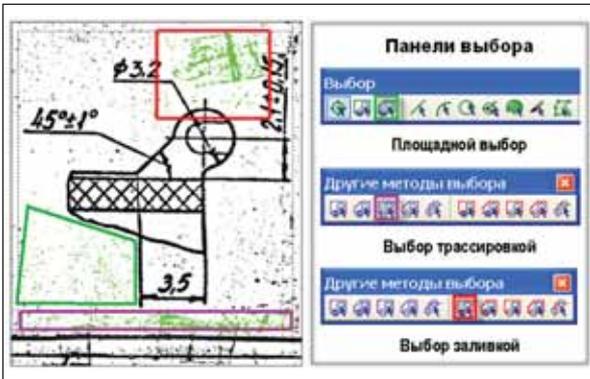


Рис. 22. Составление набора выбора

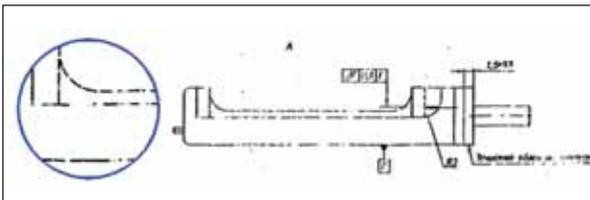


Рис. 23. Фрагмент сканированного изображения

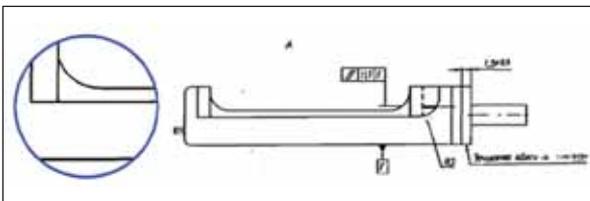


Рис. 24. Фрагмент после редактирования

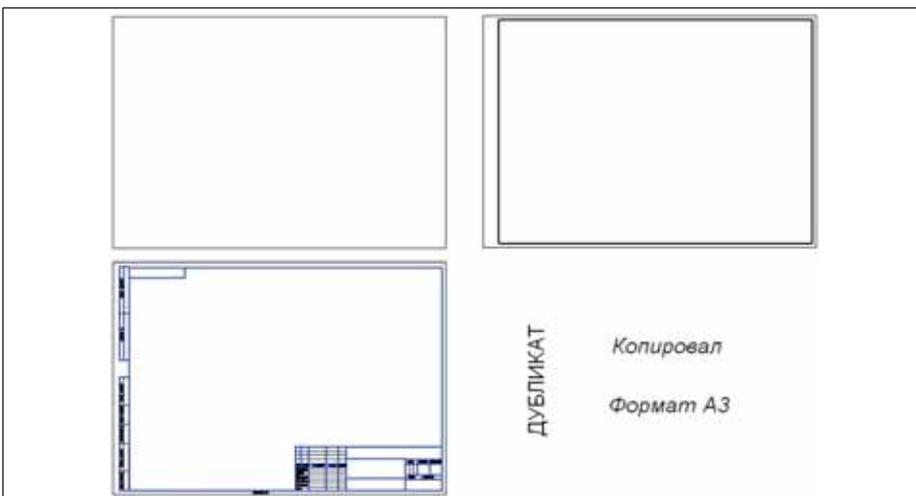


Рис. 25. Стандартные элементы документов



Рис. 26. Библиотека фрагментов

Слой с мусором отображается в окне *Слои* (рис. 20). В процессе дальнейшей обработки изображения слой с мусором рекомендуется временно погасить, а удалить растр с мусором можно после завершения обработки документа (рис. 21).

### Удаление объектов выбором

Нестандартный мусор, крупные объекты, следы сгибов документа, следы плохо стертых на бумаге элементов удаляются с помощью различных способов выбора (рис. 22). Например, с помощью площадного способа выбираются все точки, попавшие внутрь заданного полигона, способом трассировки секущей рамкой удобно выбирать ненужные объекты, примыкающие к линиям изображения, заливка с заданием обычной рамки позволяет выбрать изолированные объекты.

### Восстановление линий на чертеже

Качество растровых линий, тонких шероховатых, имеющих небольшие разрывы, можно повысить, используя монохромные фильтры (сгладить, удалить дырки, утолщить, утоньшить).

Для восстановления частично утраченных линий (рис. 23) лучше применить полуавтоматическую векторизацию, трассировку со сглаживанием или различные команды редактирования и коррекции (рис. 24).

Восстановить на чертеже полностью потерянные элементы можно с помощью инструментов рисования.

### Библиотека стандартных элементов чертежа

В архивных документах информация чаще всего теряется на краях изображения. В основном это рамки, штампы, тексты на полях документов.

Стандартные элементы документа (рамки, штампы, тексты и т.д.) можно начертить один раз (рис. 25) и создать из них библиотеку фрагментов. В библиотеку фрагментов могут входить элементы различных типов – векторные объекты, растровые фрагменты, гибридные изображения.

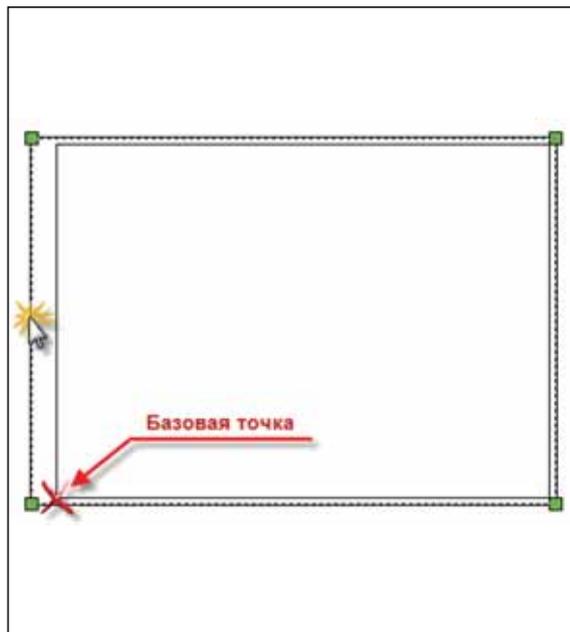


Рис. 27. Создание блока

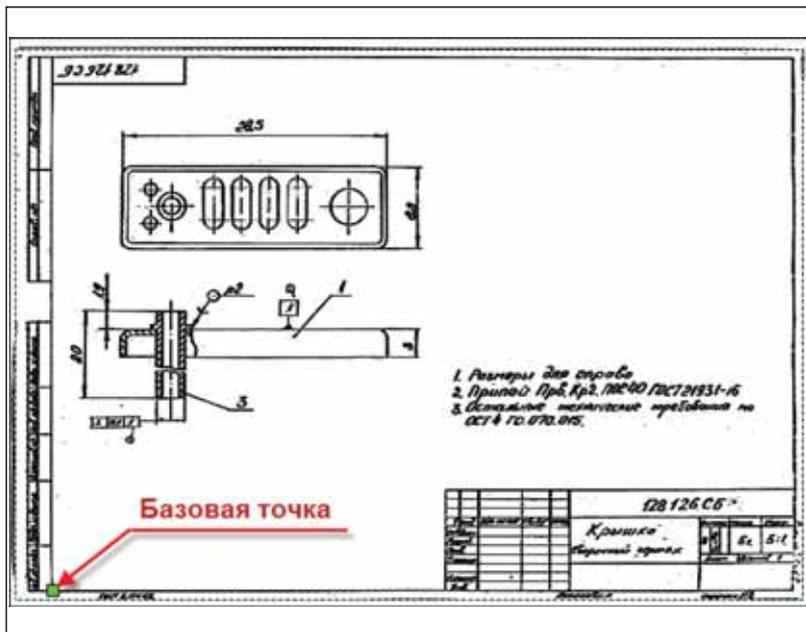


Рис. 28. Вставка внешней рамки



Рис. 29. Штамп в блоке



Рис. 30. Поля в окне Инспектор

При обработке каждого документа элементы из библиотеки фрагментов (рис. 26) можно с помощью мыши перетащить на рабочее поле и разместить в нужном месте на растровом изображении. Перед вставкой стандартных элементов, содержащих рамки, растровое изображение необходимо откорректировать по 4 углам одной из рамок или откалибровать по шаблону.

**Создание и использование блока**

При вставке внешних рамок в документ нужные углы, к которым можно привязать рамку, чаще всего отсутствуют. Но даже на самых плохих документах обычно сохраняется нижний левый угол внутренней рамки, и к нему удобно привязывать вставляемые рамки и рамки со штампами. На вставляемой внешней рамке соответствующая точка, которая должна быть базовой, находится вне объектов, создающих эту рамку.

Задать базовую точку как на объекте, так и вне его поможет создание блока из

выбранных объектов (рис. 27). При вставке блока из библиотеки фрагментов базовую точку блока с помощью мыши можно совместить с нужной точкой на растровом изображении (рис. 28). Вставленный фрагмент займет нужное положение.

**Создание стандартных элементов с атрибутами**

Стандартные элементы чертежа, например рамки и штампы, могут содержать и растровые компоненты, например логотип организации, который удобно вставлять вместе с рамками и штампом. Из растровых фрагментов нужно создать растр, а затем вместе с остальными объектами сформировать в блок.

Если в стандартный элемент чертежа входит штамп, для удобства заполнения его полей можно создать блок с атрибутами (рис. 29). В этом случае, чтобы заполнить поля штампа, нужно с клавиатуры ввести необходимую информацию в соответствующие поля окна *Инспектор* (рис. 30).

Файлы с обработанными документами размещаются в электронном архиве, а информационные данные из штампа документа заносятся в базу данных вручную или автоматически, используя возможности программного обеспечения RasterID. О технологии занесения данных в электронный архив, используя в качестве иллюстрации примеры из нашей работы, мы расскажем в следующей статье, посвященной вопросам создания электронного архива в НИИ "Полус".

*Авторы выражают искреннюю признательность сотрудникам ЗАО "Си-Софт" Илье Шустикову и Валентине Хлебниковой за помощь в написании этой статьи.*

**Виктор Янчук,  
Дмитрий Панков  
НИИ "Полус"  
E-mail: YanchukVG@yandex.ru**