



Абрахам Харольд Маслоу

Для оптимального выбора конфигурации программного обеспечения мы предлагаем пользователям присылать примеры имеющихся сканированных документов или привозить документы к нам в офис, чтобы продемонстрировать возможности программ и вместе с ними выбрать необходимую конфигурацию программного обеспечения. Как правило, такое предварительное общение помогает самим потенциальным пользователям лучше понять цели, которые им нужно достичь. Забавно, но практически 90% пользователей присылают для подобного анализа "самый плохой" с их точки зрения документ. При этом ставится задача: "Перевести все в векторный вид и желательно с минимальным нажатием кнопок, а лучше в автоматическом режиме". Эта цель понятна, и если бы все было так просто, то в программе осталась бы лишь одна кнопка — *Растр в векторы*.

Прежде чем приступать к процессу сканирования и обработки, необходимо четко представлять себе цель: что в конечном счете должно получиться, к какому результату необходимо прийти. Естественно, эту цель следует соизме-

В этой статье нам бы хотелось рассказать не столько о технологиях, заложенных в программах Raster Arts, сколько о подходе к решению задач обработки сканированных документов. Надеемся, что статья поможет нашим пользователям правильно подойти к решению задач сканирования и обработки изображений, сформулировать цели, которых необходимо достичь, и разработать правильные алгоритмы.

Сканирование — самый важный этап для последующей обработки документа. От того, насколько правильно подобран режим сканирования и заданы параметры, зависит успех всего процесса. Естественно, качество зависит и от возможностей самого сканера. Выбор аппарата для сканирования должен определяться не только показателями скорости сканирования, но и качеством, которое можно получить с помощью этого устройства (оптическое разрешение, цветовые режимы, возможности программного обеспечения).

Пользователей можно разделить на два типа: те, у которых есть только "бумага" и им ее нужно отсканировать, и те, кому сканированные растры достались "по наследству" и у них нет возможности сканировать оригинал заново. С первыми работать гораздо проще, их нужно только правильно сориентировать, проработать технологию и добиться оптимального результата. Ну а тем, у кого есть то, что есть, придется рассчитывать только на программные средства повышения качества изображения. Благо, в Spotlight и RasterDesk имеются широкие возможности восстановления документов плохого качества.

Перечислим основные ошибки при сканировании документов инженерной графики, с которыми приходится сталкиваться в процессе общения с клиентами.

1. Изображение содержит ограниченное количество цветов (цветной чертеж, карта, схема). При этом пользователь выбирает полноцветный режим сканирования RGB. В результате получается тяжелый файл, содержащий избыточную информацию (24 или 48 бит на современных сканерах). Правильнее в этом случае выбрать индексированный режим сканирования, а еще лучше использовать при сканировании подготовленный файл палитры с определенными заранее цветами. Документ при этом должен получиться оптимального размера с сохранением всей исходной информации.
2. Цветное изображение после сканирования сохраняется в формат *.jpg (*.jpeg) для уменьшения размера файла. Мы уже неоднократно говорили об этой проблеме. Данный формат в большинстве случаев не подходит для корректной работы со сканированной инженерной графикой. Он оптимально работает со сканированными фотографиями или изображе-

ниями, содержащими градиентную заливку. При использовании же этого формата для работы с инженерной графикой происходит значительная потеря точности и цвета, образуются цветовые ореолы на границах линий, что существенно усложняет дальнейшую обработку и векторизацию изображений. Оптимальным вариантом для сохранения цветных документов является формат *.tif (*.tiff). Файл этого формата будет больше, чем файл *.jpg, но при правильном задании группы компрессии его размер можно оптимизировать. При этом пользователю гарантировано сохранение цветовой гаммы и отсутствие потери информации.

3. Сканирование монохромного документа в цветном режиме. Пользователь (оператор сканера) может делать это не специально. По умолчанию в параметрах сканирования стоит цветной режим, вот он и используется автоматически для всего. При открытии документа в программе для

4. Сканирование монохромного документа плохого качества на аппарате среднего класса в черно-белом режиме. В этом случае сложно добиться хороших результатов. Гораздо проще отсканировать такой документ в режиме *Градации серого*, а затем программным способом перевести его в монохромный вид.
5. Сканирование документа с цветным фоном в черно-белом режиме, например, "синьки". При этом существенно усложняется процесс обработки. Фон и полезная информация сливаются. Дальнейшая чистка с удалением мусора приводит к значительной потере информации в документе. Как и в предыдущем случае, необходимо отсканировать документ в режиме *Градации серого* и программным способом перевести его в монохромный режим.

"Золотой молоток", или Волшебная "красная кнопка"

Технологии, которые используются для работы с растровой графикой, отличаются от методов и технологий векторного проектирования. Если я, к примеру, хочу выбрать векторный отрезок, чтобы изменить его геометрию или свойства, мне достаточно указать на него курсором мыши. Система изначально "знает", что это за объект и какие у него свойства. При выборе растрового отрезка все гораздо сложнее. Растровое изображение не имеет объектной структуры, для системы загруженная картинка пока всего-навсего набор точек. Все, что пока известно, — общее количество точек, их плотность (dpi) и глубина цвета. Объектами (растровыми отрезками, дугами, окружностями и т.д.) элементы становятся лишь при использовании сложных алгоритмов распознавания, реализованных в программах Spotlight и RasterDesk. В предыдущих статьях мы неоднократно описывали преимущества уникальных гибридных технологий наших программ. Но все эти интеллектуальные методы распознавания могут не помочь, если исходное качество изображения не позволяет эффективно его обработать. Одна из "проблем", с которой сталкиваются начинающие пользователи, заключается в поспешности применения функций без знания структуры изображения и принципов обработки. Это приводит к тому, что полученный результат не всегда удовлетворяет пользователя, и он принимает ошибочное решение, что



обработки он выглядит как монохромный, но, по сути, является полноцветным, и те команды, которые нужно применять для чистки мусора и монохромной фильтрации, на нем попросту не работают. Конечно, есть возможность перевести документ в монохромный вид при помощи команд бинаризации, но качество в результате такой обработки может быть несколько потеряно. Лучше было бы изначально подобрать для монохромного документа правильный черно-белый режим сканирования. Тем более что современные сканеры отлично обрабатывают такого рода документы, используя встроенные аппаратные алгоритмы чистки и подавления растрового мусора.

эта программа ему не подходит. Загружая цветное растровое изображение, пользователи применяют монохромные фильтры (удаление мусора и т.п.), которые не работают с данным типом растра. Или другой пример — применение автоматической векторизации без предварительной настройки системы и подготовки изображения. Пользователь получает далеко не ту картинку, которую ожидал увидеть при векторизации.

Другая "проблема" — шаблонность методов обработки изображений, так называемый принцип "золотого молотка". Если на одном типе растра пользователю с помощью применения определенной последовательности команд удалось получить хороший результат, то при обработке другого растра результат может быть противоположным. Это не означает, что нельзя автоматизировать обработку изображений. Более того, в Spotlight Pro есть очень удобные инструменты автоматизации: Мастер командных файлов, позволяющий записать скрипт из последовательности команд обработки, и Мастер пакетных заданий, позволяющий в пакетном режиме обрабатывать большое количество файлов. В пользовательской настройке интерфейса командный файл можно назначить на кнопку на панели инструментов и автоматически применить целый набор команд одним нажатием этой кнопки. При использовании такого рода автоматизации необходимо понимать, какого результата вы ждете от такой обработки. Например, при пакетном режиме удаления мусора не исключен вариант, что при заданных параметрах размеры элементов полезной информации (знаки препинания в тексте, мелкие объекты чертежа) могут оказаться меньше размера фракций растрового мусора. И в результате обработки будет потеряна полезная информация. А вот такие команды, как *Автоматическое устранение перекоса*, *Поворот изображений*, *Автоматическая обрезка* и многие другие, при использовании в автоматизации обработки существенно сокращают время на обработку большого количества файлов. Применять автоматизацию можно и нужно, это экономит время. Но применять ее нужно с умом. Постарайтесь классифицировать изображения перед автоматической обработкой, подобрать группы по цветности и типам дефектов. Например, если есть цветные или "серые" изображения и вам нужно их в даль-

нейшем векторизовать, то для одного типового растра удобно подобрать параметры бинаризации (перевод в монохромный вид), записать параметры команды во внешний файл и в пакетном режиме перевести в "монохром" все документы данного типа. При этом важно, чтобы подобранные параметры подходили ко всем обрабатываемым растрам.

Нужно ли чистить растр?

Этот вопрос часто возникает при обработке документов. Естественно, с эстетической точки зрения пользователю хочется получить чистое изображение, с ним приятнее и удобнее работать. Но не всегда эта цель оправдывает затраченные средства. Если нужно обработать небольшое количество документов и вы располагаете временем для дополнительной ручной или полуавтоматической чистки растров, то почему бы и не привести их в



порядок. Другое дело, если стоит глобальная задача сканирования большого числа документов и мусор на изображениях хоть и присутствует, но не мешает "читать чертеж". В этом случае вопросы чистки и приведения документа в порядок можно отложить и на потом. Главное для оператора сканера — полностью сохранить информацию с исходного документа. Да и обработку чертежа непрофессиональному пользователю, возможно, и не стоит поручать. Качественней эту работу может сделать не оператор сканера, а конечный пользователь, которому в дальнейшем предстоит работать с этим документом.

Прежде чем векторизовать...

Возвращаясь к вопросу о целях пользователей. "Все отсканировать и векторизовать", — так потенциальные пользователи формулируют нам основную задачу. Либо пользователь придумал такой план действий сам, либо такая задача поставлена перед ним руководством. Когда задаешь наводящий вопрос: "А зачем вам нужно все перевести в векторный вид?", в большинстве случаев получаешь ответ: "Чтобы вносить изменения в документ". И тут самое время продемонстрировать уникальные возможности программ Raster Arts для гибридного редактирования и интеллектуального распознавания объектов. Демонстрация выбора растровых объектов щелчком мыши и их изменения из геометрии и свойств практически снимает все вопросы о редактировании растра без векторизации. Преимуществ у этой технологии несколько.

1. Самое главное — вы не тратите значительное время на векторизацию и последующее приведение полученных векторов в нужный вид, вносите изменения в растровый документ привычными инструментами векторного редактирования.
2. Не происходит потеря данных, возможная при переводе изображения при векторизации, если исходное изображение было не очень хорошего качества.
3. Сохраненный гибридный документ открывается обычным редактором, поддерживающим формат *.dwg и загрузку растровых изображений, или любым вьюером. Не требуется специализированного растрового редактора для дальнейшей работы с документом.

Значительная часть наших пользователей выбрала именно этот метод внесения изменений в отсканированную документацию. Изменяется и векторизуется только нужная часть документа. Средства привязки и объектного отслеживания, работающие с растровыми примитивами так же, как и с векторными, позволяют быстро внести изменения, удалив часть растровых объектов, и привязать новые векторы к имеющейся графике.

Векторизация

Если все же необходимо векторизовать изображение, то следует определить задачу векторизации. Задач может быть несколько, например:

1. Перевести все изображения в векторный вид и передать документ в даль-

нейшую работу именно в векторном виде. Это наиболее кропотливая работа, требующая много времени даже при относительно хорошем качестве исходного материала.

2. Векторизация отдельных элементов и передача векторной графики в сторонние приложения для дальнейшей работы (оцифровка горизонталей на отсканированных планшетах и передача графики в специализированное ПО для построения поверхностей, векторизация контуров для передачи геометрии в специализированное ПО для технологической обработки и т.д.). Задача решается с помощью автоматической векторизации и автоматической векторной коррекции при хорошем качестве материала или полуавтоматической векторизации в выбранном режиме для документов среднего качества.

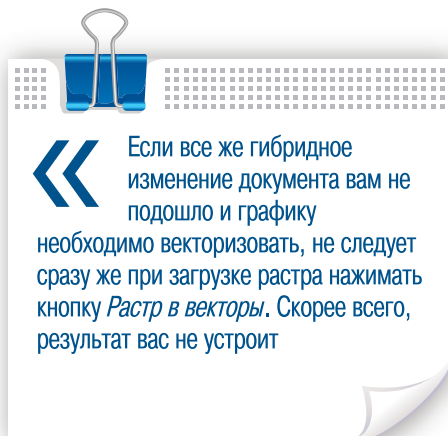
Если все же гибридное изменение документа вам не подошло и графику необходимо векторизовать, не следует сразу же при загрузке раstra нажимать кнопку *Растр в векторы*. Скорее всего, результат вас не устроит. Вот несколько пунктов, которые нужно выполнить перед нажатием этой кнопки:

- устранить геометрические искажения в документе. При векторизации "кривого раstra" вы получите "кривые" векторы, а исправить их на векторном чертеже на порядок сложнее, чем устранить искажения в растровом изображении;
- подавить имеющийся фон, если монохромный чертеж был отсканирован в цветном режиме или в градациях серого, и получить качественное монохромное изображение;
- корректно настроить параметры векторизации (команда *Параметры конверсии*): как минимум, необходимо указать точность, максимальную толщину линий, максимальный размер и по необходимости остальные настройки;
- желательно "расслоить" изображение по тематическим объектам с помощью специализированных команд: *Выделить текст*, *Выделить штриховку*, *Выделить линейные объекты*. При этом объекты каждого типа переносятся на отдельные слои и для каждого типа графики (текст, линейные объекты, полилинии) настраиваются свои параметры. На сложном перегруженном растре векторизовать текст и графику отдельно гораздо проще, чем одновременно.

Цели и задачи пилотного проекта

Перед тем как приступать к реализации большого проекта по сканированию бумажного архива, желательно выполнить пилотный проект на ограниченном количестве документов. Для сканирования необходимо подобрать документы разного качества и размера, провести цикл сканирования и подбор оптимальных параметров для каждого типа документов. Следует отработать технологию внесения документов в систему электронного архива и оценить трудозатраты на повышение качества и обработку сканированных изображений. Анализ пилотного проекта позволит вам:

- оптимизировать процесс, определить необходимые технологические показатели;
- согласовать работу оператора сканера и специалиста, занимающегося



- наполнением электронного архива;
- оценить временные затраты на реализацию всего проекта по переводу документации в электронный вид.

Итак, подводя итог сказанному, хотелось бы еще раз акцентировать внимание на важных аспектах, учет которых позволит вам эффективно работать со сканированными документами любой сложности.

1. Изучите качество вашего материала, определите, какой тип растрового изображения вам нужен для дальнейшей работы.
2. Для сканирования подберите оборудование, класс которого позволяет не только быстро сканировать документ, но и получить нужное качество электронного документа.
3. Правильно отсканированный растр — залог дальнейшей успешной работы с документом в электронном виде. Не

определяйте в качестве приоритета для себя размер полученного изображения, а постарайтесь добиться баланса между необходимым и достаточным качеством и размером полученного документа.

4. Продумайте весь цикл работы с отсканированным изображением. От того, насколько правильно вы сформулируете цель вначале, зависит эффективность реализации всего проекта.
5. Досконально изучите возможности программного обеспечения, которое вы будете использовать для обработки и векторизации отсканированных документов. Как правило, большинство пользователей ограничиваются 20-30% возможностей имеющейся у них программы. Знание функций и методов в широком объеме позволит вам максимально быстро и хорошо обрабатывать документы разного качества, решать специализированные прикладные задачи с использованием данных из отсканированного документа.

Со своей стороны мы готовы оказать посильную помощь пользователям при работе с отсканированными документами в нашем программном обеспечении серии Raster Arts. Мы предоставляем пользователям возможность пройти полный курс обучения работе с программным обеспечением серии Raster Arts: Spotlight или RasterDesk. Под руководством преподавателя на практических примерах пользователь освоит не только полный функционал программного обеспечения, но и приобретет ценные навыки обработки документов разной степени сложности. Любые вопросы, связанные с обработкой изображений в нашем программном обеспечении, можно решить с помощью специалистов технической поддержки ГК CSoft, которые помогут начинающему пользователю с выбором ПО, необходимого для решения его задач, а постоянным пользователям ответят на вопросы, связанные с функционированием программ.

Более подробно ознакомиться с программным обеспечением серии Raster Arts можно на сайте программного продукта www.rasterarts.ru.

Илья Шустиков

CSoft

Тел.: (495) 913-2222

E-mail: shustikov@csoft.ru