

Condition

Overview

Remember that our goal for using Grooper is to automate tasks that we would otherwise have to perform manually.

Imagine we have a pile of papers in our inbox. We can think of this as the batch we just made, so they arrived there via the `Acquire` phase.

If we had to work through these documents by hand, we'd probably have to go through them and make sure they're all facing the same way, remove any sticky notes on them, take out the staples, and potentially white out smudges and marks.

This is the `Condition` phase, and we can automate these tasks in Grooper.

In this phase, we will:

- create and test an Image Processing Profile,
- use the new profile to clean up our batch, and
- use OCR to obtain text from our images that we'll use later.

Let the conditioning commence!

But first

For the next few phases, we'll be testing the results of adding a single Batch Process step at a time. Let's take a deeper look at our batch so that when we start adding steps, the process makes a little bit more sense.

Step 1

In Grooper Design Studio, navigate to `(root) > Batch Processing > Batches > Production > Invoices Process`.

`(root)`

└ Batch Processing

 | └ Batches

 | | └ Production

 | | | └ Invoices Process

```

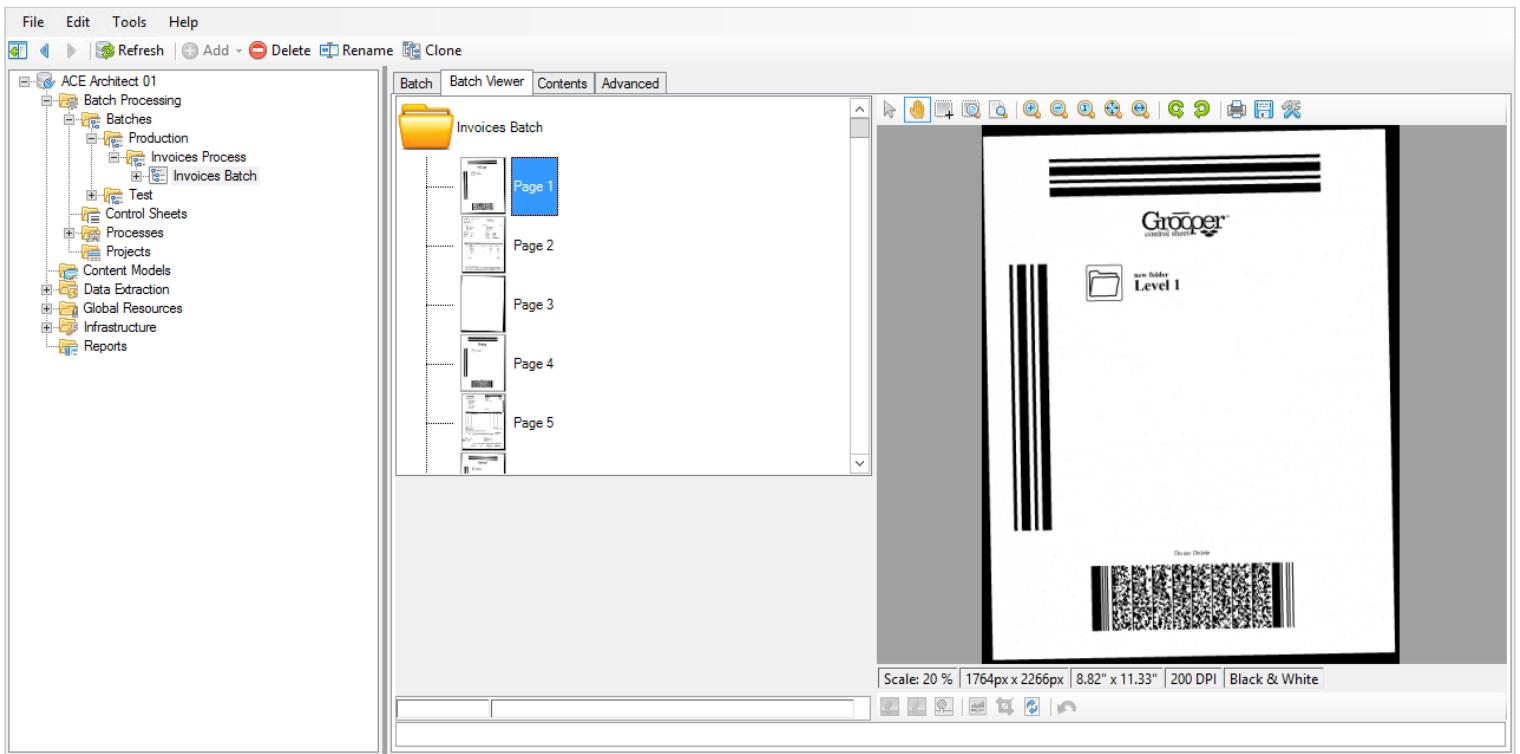
| | └ Test
| └ Control Sheets
| └ Processes
| └ Projects
└ Content Models
└ Data Extraction
└ ...

```

The screenshot shows the ACE Architect 01 application interface. The left sidebar contains a navigation tree with categories like Batch Processing, Control Sheets, Processes, Projects, Content Models, Data Extraction, Global Resources, Infrastructure, and Reports. Under Batch Processing, there are Batches, Production, and Test environments. In the Test environment, there is an 'Invoices Process' which contains an 'Invoices Batch'. The main workspace is titled 'Batch' and displays detailed information about this batch. The 'General' tab shows Priority (3), Current Step (3), and Description. The 'Batch Information' tab shows details such as Created (8/6/2018 4:59 PM), Created By (BIS\vdelk), Current Step No (-1), and Current Step Name ((none)). Below this is a 'Batch' section with a description: 'A Batch is a hierarchy of folders and pages which serves as the fundamental unit of capture in Grooper.' There is also a 'Remarks' section. At the bottom of the workspace, there are tabs for Processing Status, Statistics, Tasks, Batch Viewer, Processing History, Event Log, and Details. A large chart titled 'Task Status' shows a single task named 'Scan (1 / 1)' with a progress bar from 0% to 100%, where the entire bar is filled blue. A legend indicates four states: Pending (grey), In Progress (green), Completed (blue), and Errors (red).

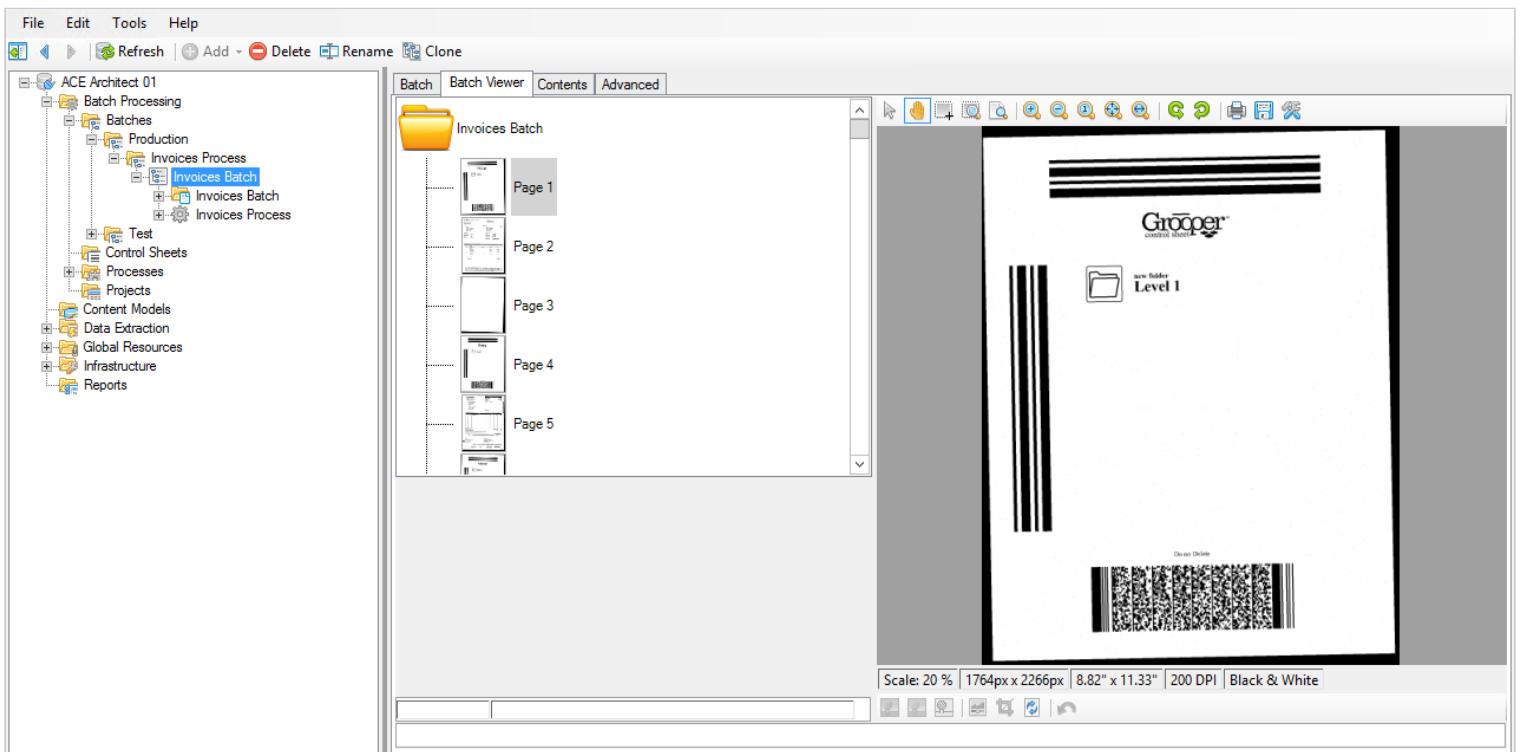
Any batches that are created will also be visible in the node tree. Viewing the batch from this location gives us access to more information about the batch itself.

You can simply view the batch by clicking on the **Batch Viewer** tab.



Step 2

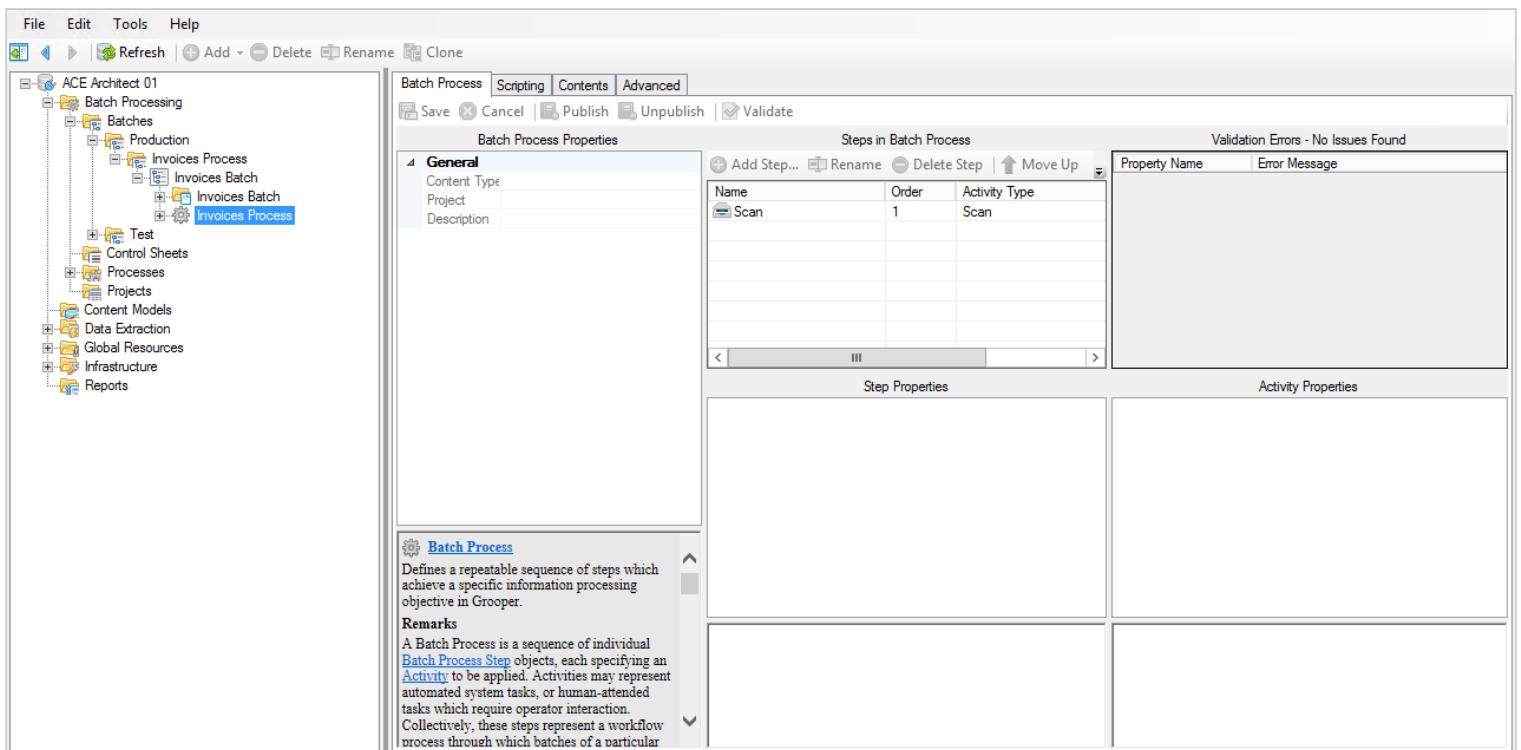
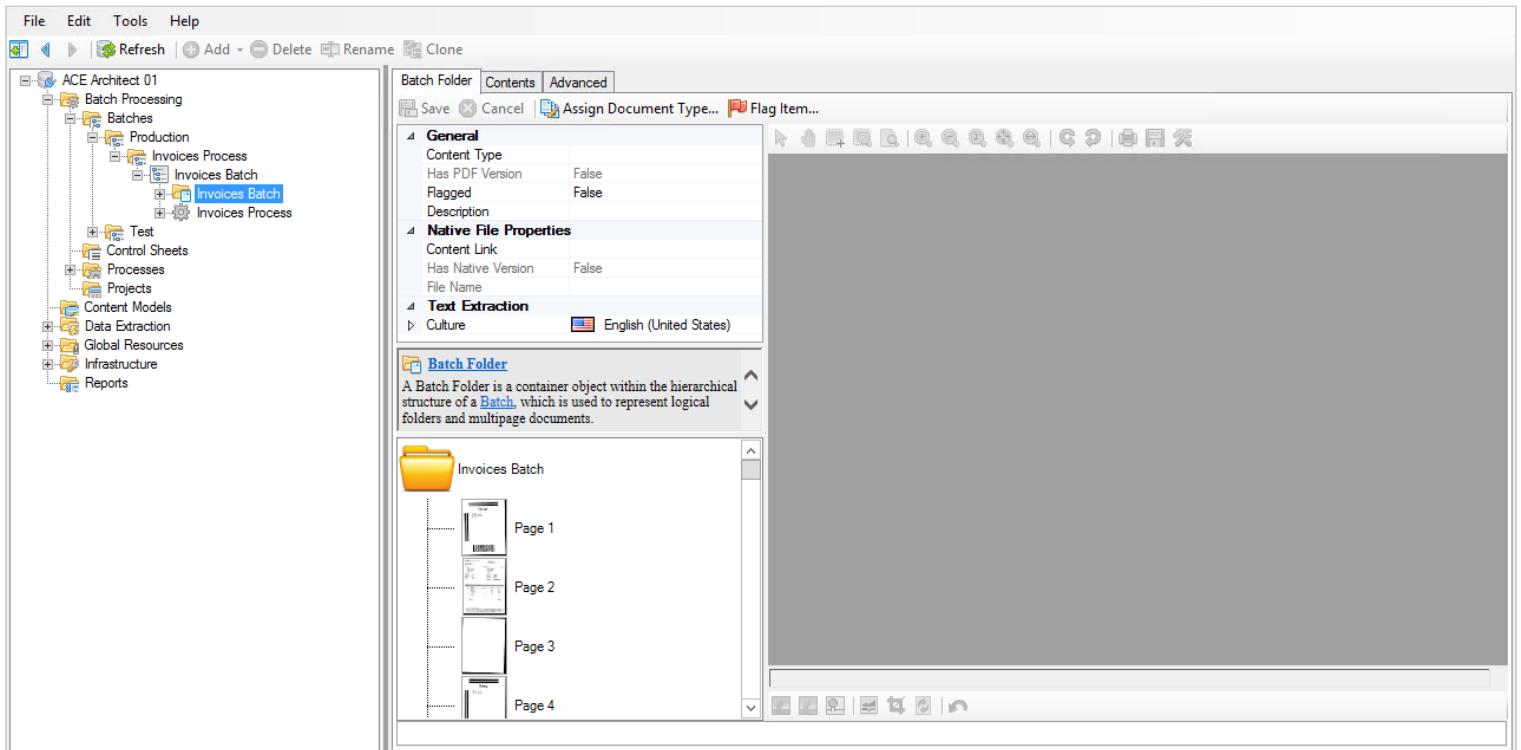
Click the **+** button next to the batch in the node tree to expand it.



Notice that there are two child objects here:

└ Invoices Batch
 └ Invoices Batch
 └ Invoices Process

1. a folder with the same name as the batch, and
2. a gear icon with the name of the Batch Process we used to create the batch.



When you create a production batch and assign it a particular Batch Process, it attaches a copy of that Process to that batch. This is to ensure that, should there be any changes to the process after the batch is created, it won't affect how this batch is processing.

An important note is that because this Process is only a copy of the original, it doesn't maintain a link to the original.

If we want to make changes to the Process and subsequently test those changes against an batch, we need to tell that batch to update.

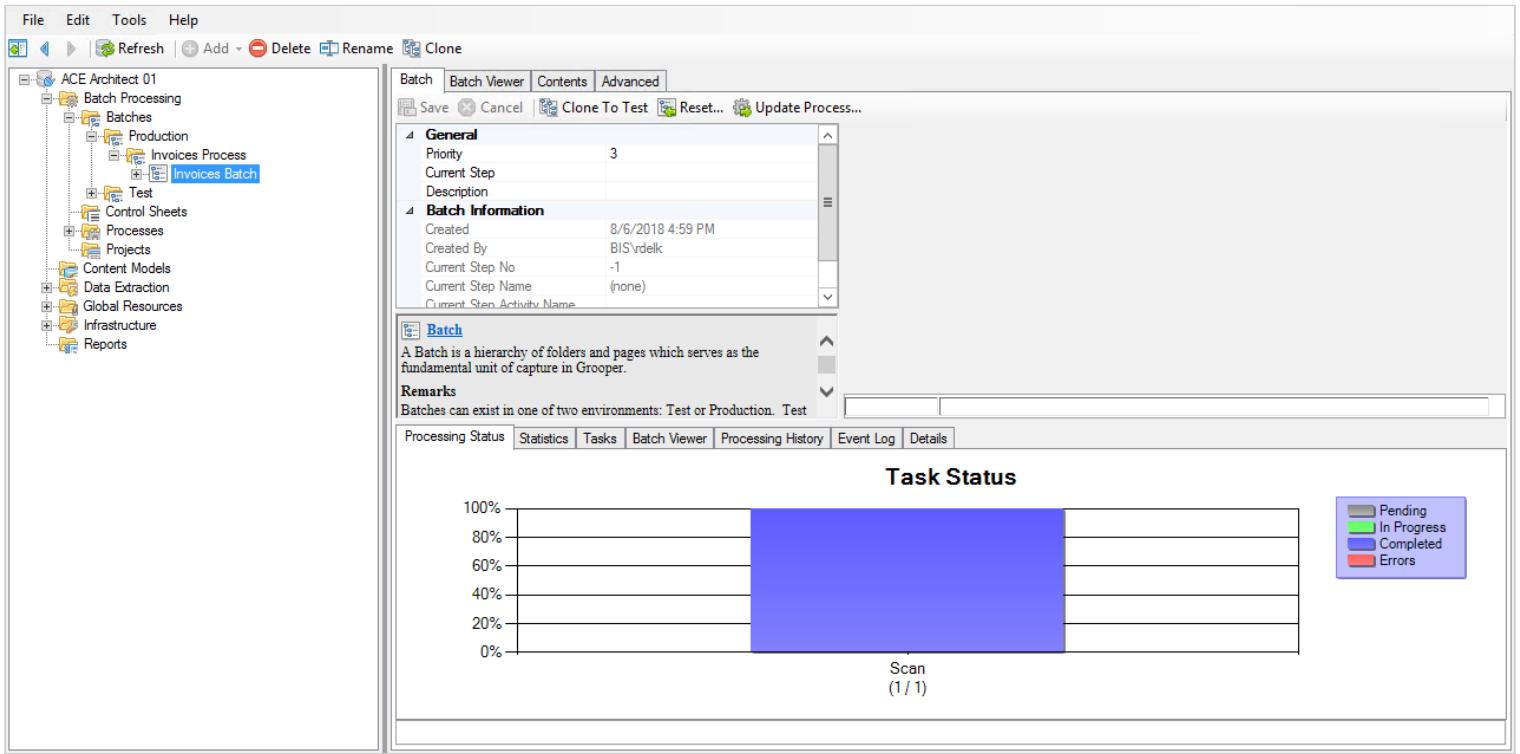
This is like telling Grooper, "Hey, I've changed the steps I want you to take when you process."

We'll be doing this quite a few times, so it won't hurt to become familiar with the technique!

Cloning to test

Step 1

Navigate back up the node tree to `(root) > Batch Processing > Batches > Production > Invoices Process > Invoices Batch`.



We need to clean up our batch images, so we're going to create an Image Processing Profile to do that. However, profiles can be tested only against test batches, so we can't use our production batch.

We can, however, clone our production batch to our test batches. This way we don't have to create a brand new test batch to work with.

Step 2

Click the **Clone To Test** button, and then **Execute** on the window that appears.

S | Refresh | Add | Delete | Rename | Clone

File Edit Tools Help

ACE Architect 01

- Batch Processing
 - Batches
 - Production
 - Invoices Process
 - Invoices Batch
 - Test
 - Control Sheets
 - Processes
 - Projects
 - Content Models
 - Data Extraction
 - Global Resources
 - Infrastructure
 - Reports

Batch Batch Viewer Contents Advanced

Save Cancel Clone To Test Reset... Update Process...

General

Priority 3 Clones this production batch to the 'Production Clones' folder of the test branch.

Current Step

Description

Batch Information

Created 8/6/2018 4:59 PM
 Created By BISvdekk
 Current Step No -1
 Current Step Name (none)
 Current Step Activity Name

Batch

A Batch is a hierarchy of folders and pages which serves as the fundamental unit of capture in Grooper.

Remarks

Batches can exist in one of two environments: Test or Production. Test

Processing Status Statistics Tasks Batch Viewer Processing History Event Log Details

Task Status

Scan (1 / 1)

Pending In Progress Completed Errors

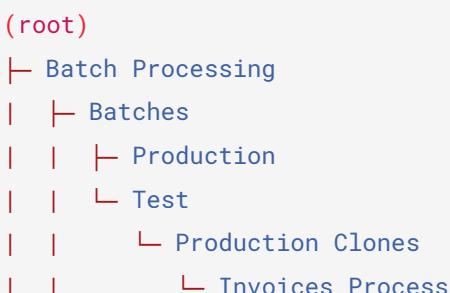
E0 02 Warning

Make sure you click the **Clone To Test** button and NOT the **Clone** button in the upper toolbar.

- **Clone To Test** makes an exact replica of the batch in the test batches.
- **Clone** is a way to create an exact copy of an object in-place in the node tree. If we used this option instead, we would have two identical production batches (and we'd have to rename the second one).

Once the clone is complete, you will see a confirmation window; click **OK**.

Now you can view the cloned batch by navigating to **(root) > Batch Processing > Batches > Test > Production Clones > Invoices Process > Invoices Batch**





The screenshot displays the ACE Architect 01 application window. The left sidebar shows a tree view of the project structure, including 'ACE Architect 01', 'Batch Processing', 'Batches', 'Production', 'Test', 'Control Sheets', 'Processes', 'Projects', 'Content Models', 'Data Extraction', 'Global Resources', 'Infrastructure', and 'Reports'. The main panel is titled 'Batch' and contains tabs for 'Batch Viewer', 'Contents', and 'Advanced'. The 'General' section shows Priority: 3, Current Step: 3, and Description: 'Invoices Batch'. The 'Batch Information' section shows Created: 8/6/2018 4:59 PM, Created By: BIS\vdelk, Current Step No: -1, Current Step Name: '(none)', and Current Step Activity Name: ' '. Below these sections is a 'Batch' description: 'A Batch is a hierarchy of folders and pages which serves as the fundamental unit of capture in Grooper.' A 'Remarks' section states: 'Batches can exist in one of two environments: Test or Production. Test'. At the bottom, there are tabs for 'Processing Status', 'Statistics', 'Tasks', 'Batch Viewer', 'Processing History', 'Event Log', and 'Details'. A 'Task Status' chart is present, showing a vertical axis from 0% to 100% and a horizontal axis for tasks. A legend indicates four status colors: Pending (grey), In Progress (green), Completed (blue), and Errors (red). The chart area shows 'Scan (0 / 0)'.

Image Processing

Our current objective is to clean up the images, which we learned is done via an Image Processing Profile (or "IP Profile").

We can start creating one now that we've got a batch against which we can test it.

Creating an IP Profile

Step 1

Navigate to `(root) > Global Resources > IP Profiles`.

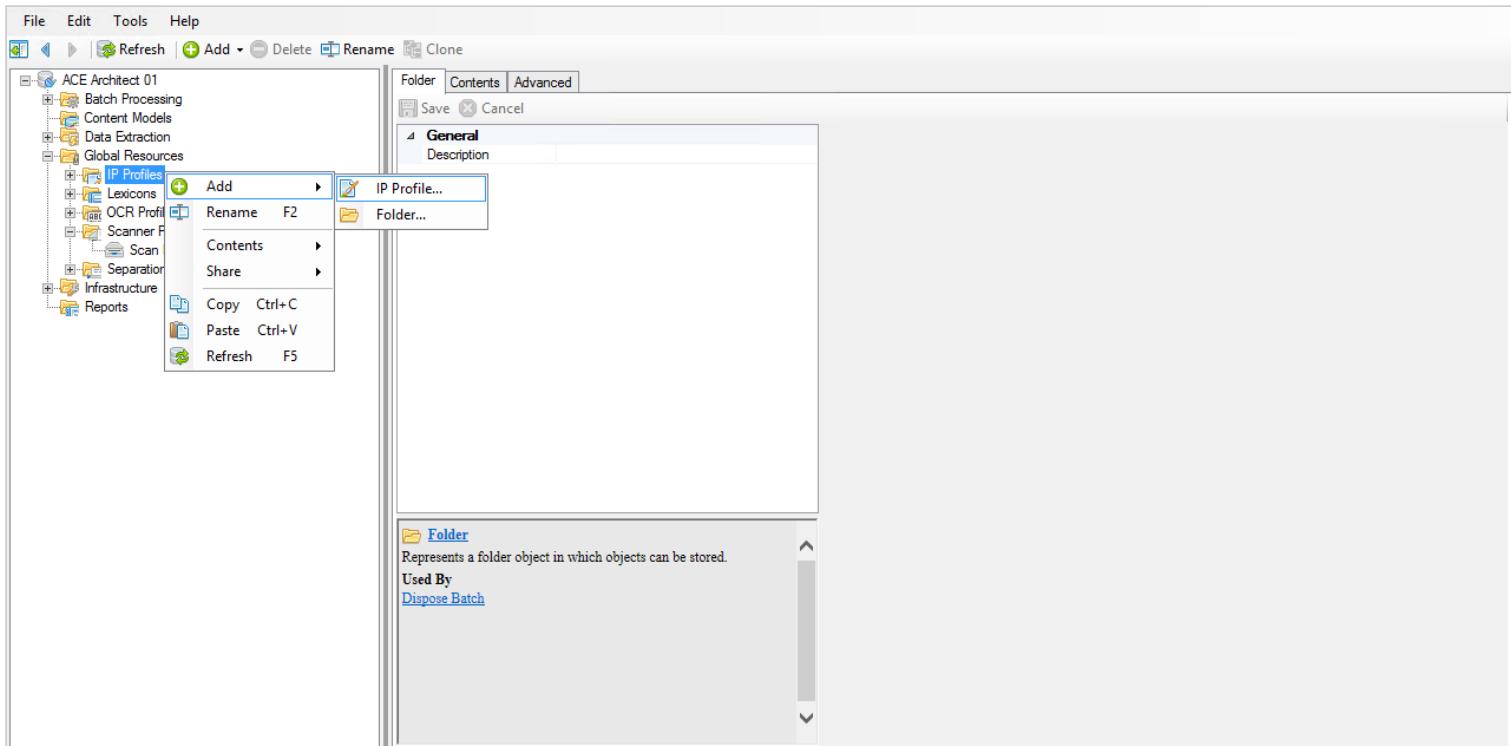
`(root)`

└ Control Sheets
└ Processes
└ Projects
└ Content Models
└ Data Extraction
└ ...

- Data Extraction
- Global Resources
- IP Profiles
- Infrastructure
- Reports

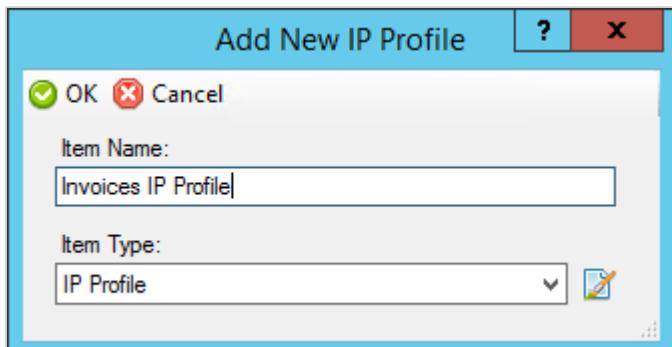
Step 2

Right-click on this node and click Add > IP Profile... .



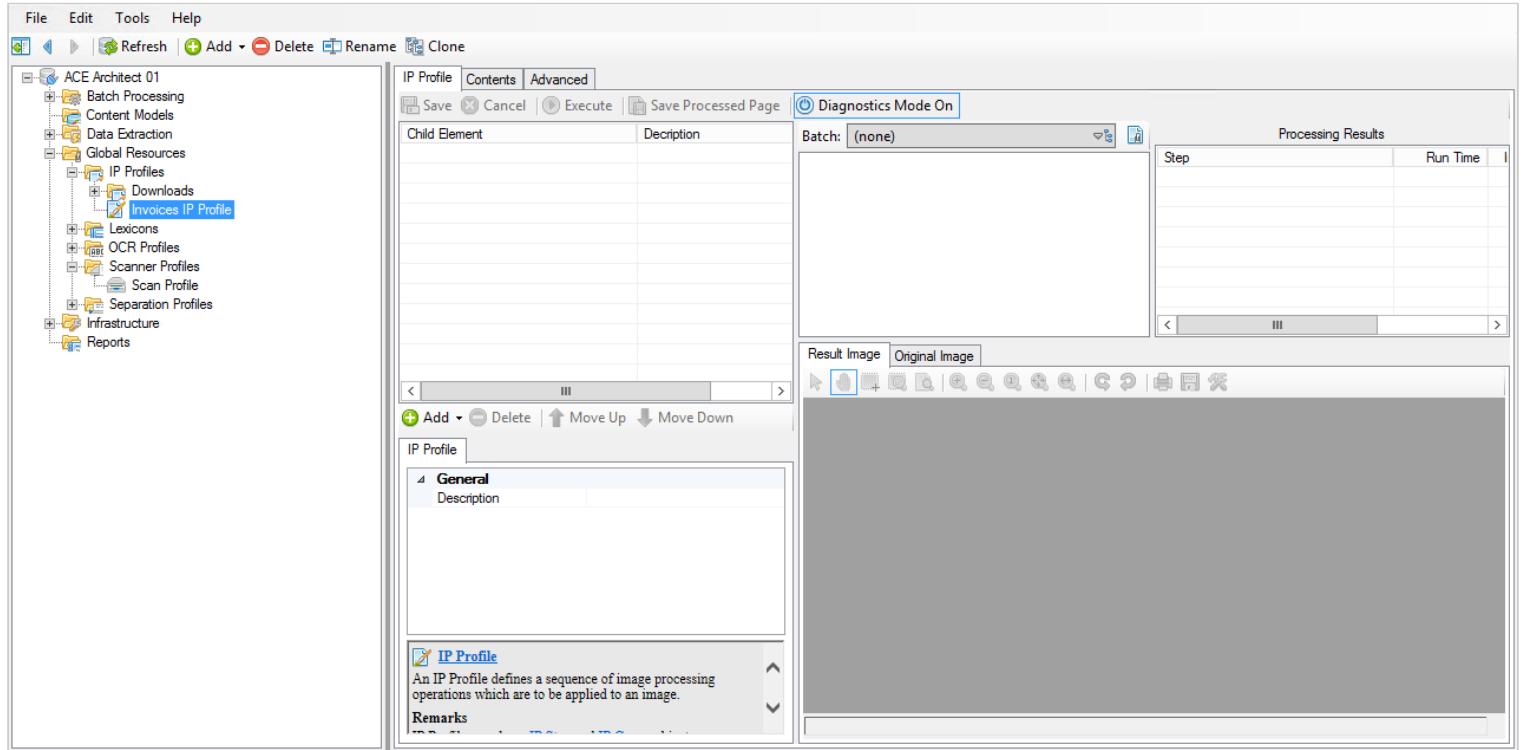
Step 3

Give the profile a name, such as Invoices IP Profile , and click OK .



Configuring the IP Profile

Once you have your IP Profile created, you should see the configuration screen.



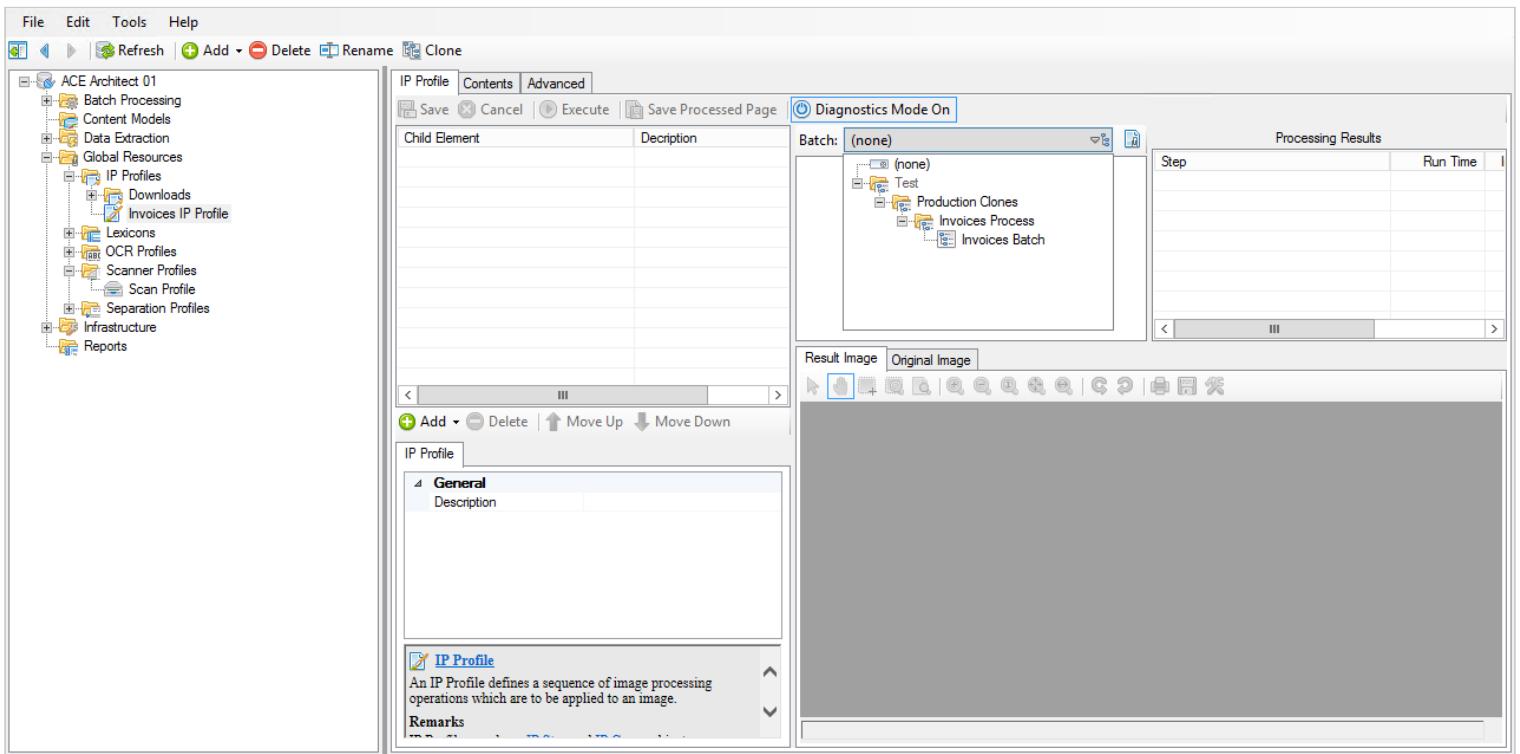
Setting up an IP Profile is a bit like setting up a Batch Process. We'll create a list of things to do (called "commands") when the profile runs against a page. Then we'll actually get to test it out against our test batch before we put it into production.

Selecting a batch

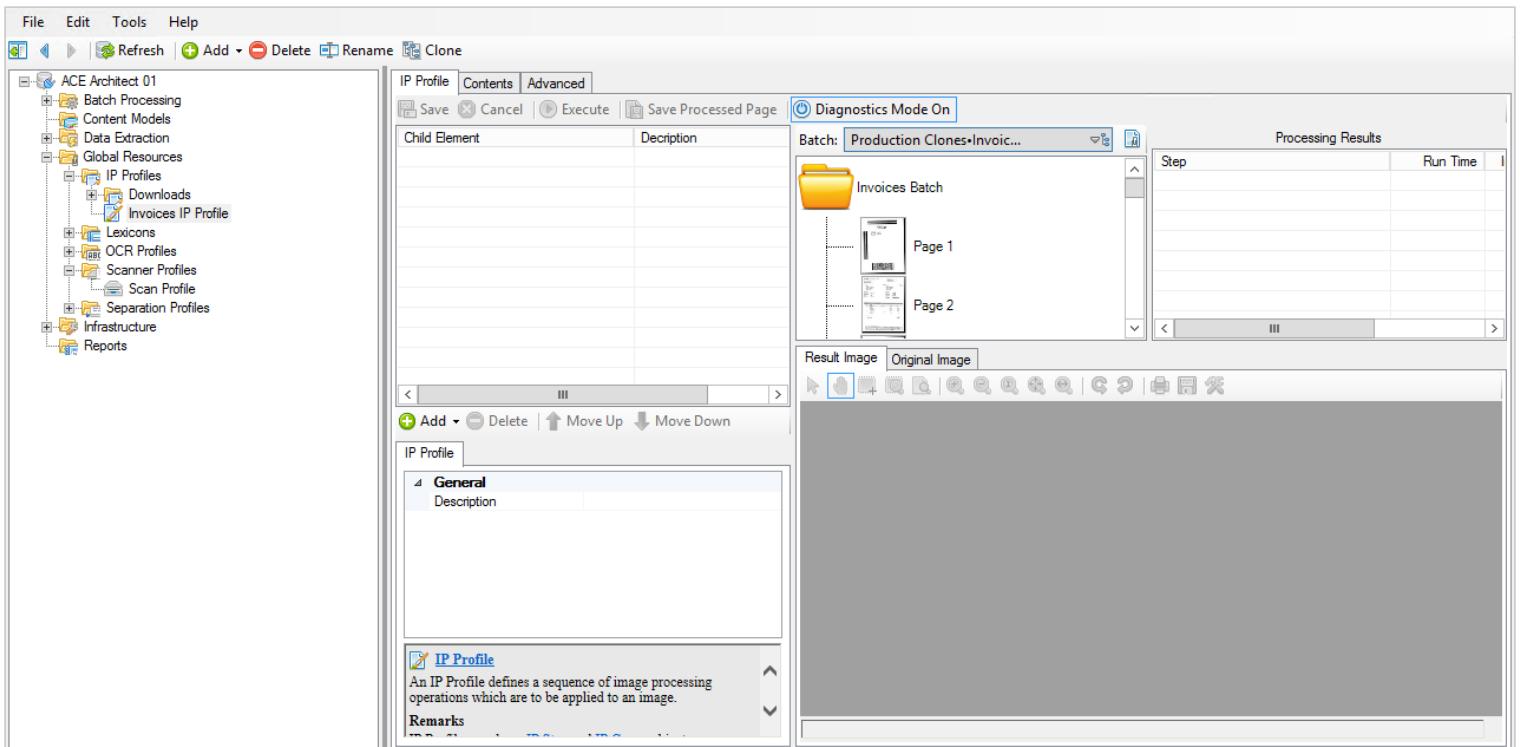
Step 1

From the **Batch** dropdown, select our cloned **Invoices Batch**.

We're going to build the profile, but we first need to select a batch for testing, otherwise we won't know if the commands we're adding will work.



After you select the batch, it will appear in the batch viewer below.



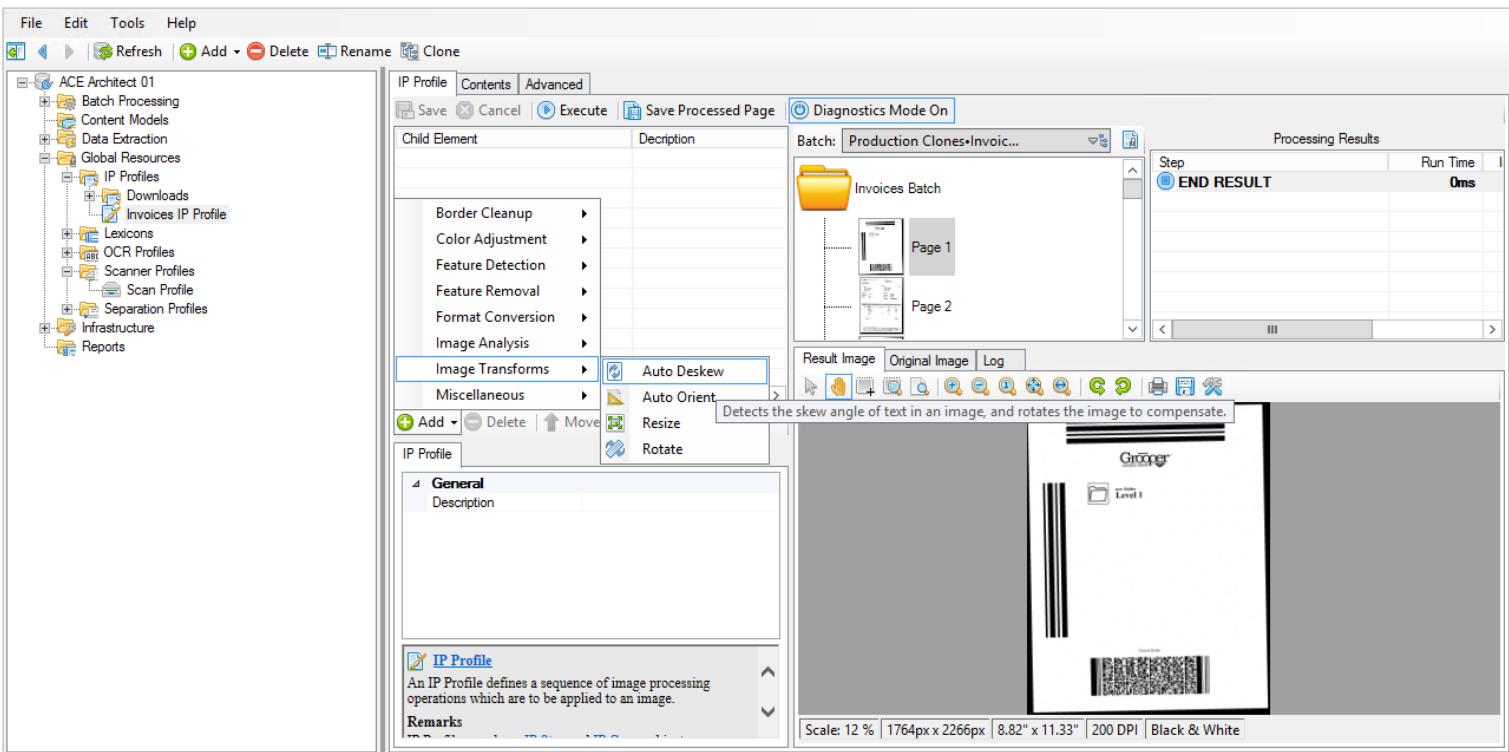
Let's add a command!

Adding a Deskew command

The first thing we want to do is deskew the images.

E3 Step 2

Click the **Add** button, and then click on **Image Transforms > Auto Deskew**.



You should now see the "Auto Deskew" command in the list in the "Child Element" column.

E8 Tip

If you're not seeing an image in the lower right panel, click on one of the pages in the batch.

E9 Tip

When adding commands, if you're unsure what their purposes are, hover over them with your cursor. A tooltip will pop up and give you a brief description.

In the panel below, there are different properties that you can modify for each command, but we're going to leave these ones at their default configurations.

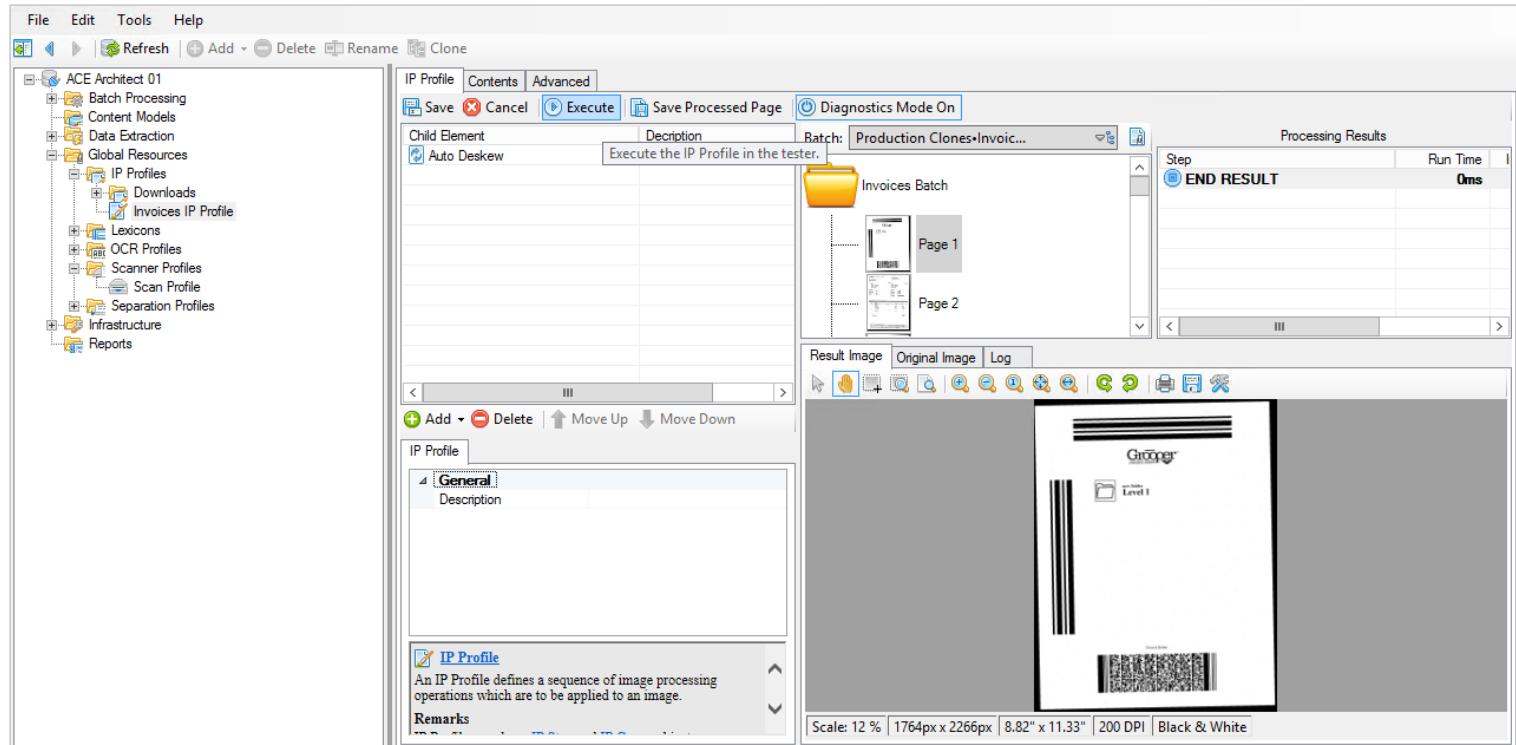
Let's test it out so far to make sure it's working.

Testing the command

Step 3

Click the **Execute** button in the toolbar. Keep your eye on the image in the lower right panel.

This will run all commands in the list against the selected page from top to bottom.

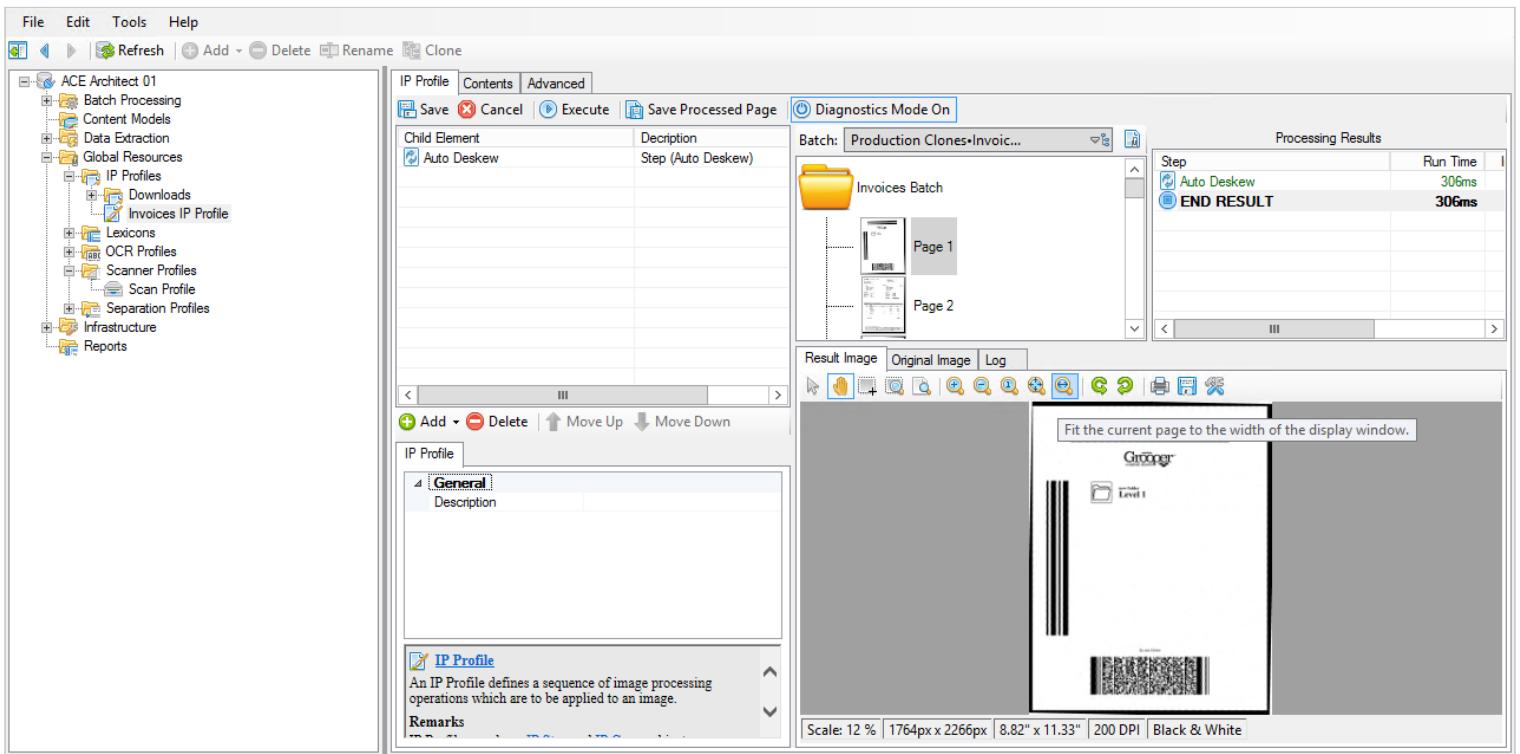


If you were watching, you probably noticed that the image did rotate, but now we have some extra white space around the edges that we didn't have before. That's okay, because we have plenty of other commands at our disposal to take care of that. Right now our primary concern is that the *text* is oriented correctly (we'll talk about why here in a bit).

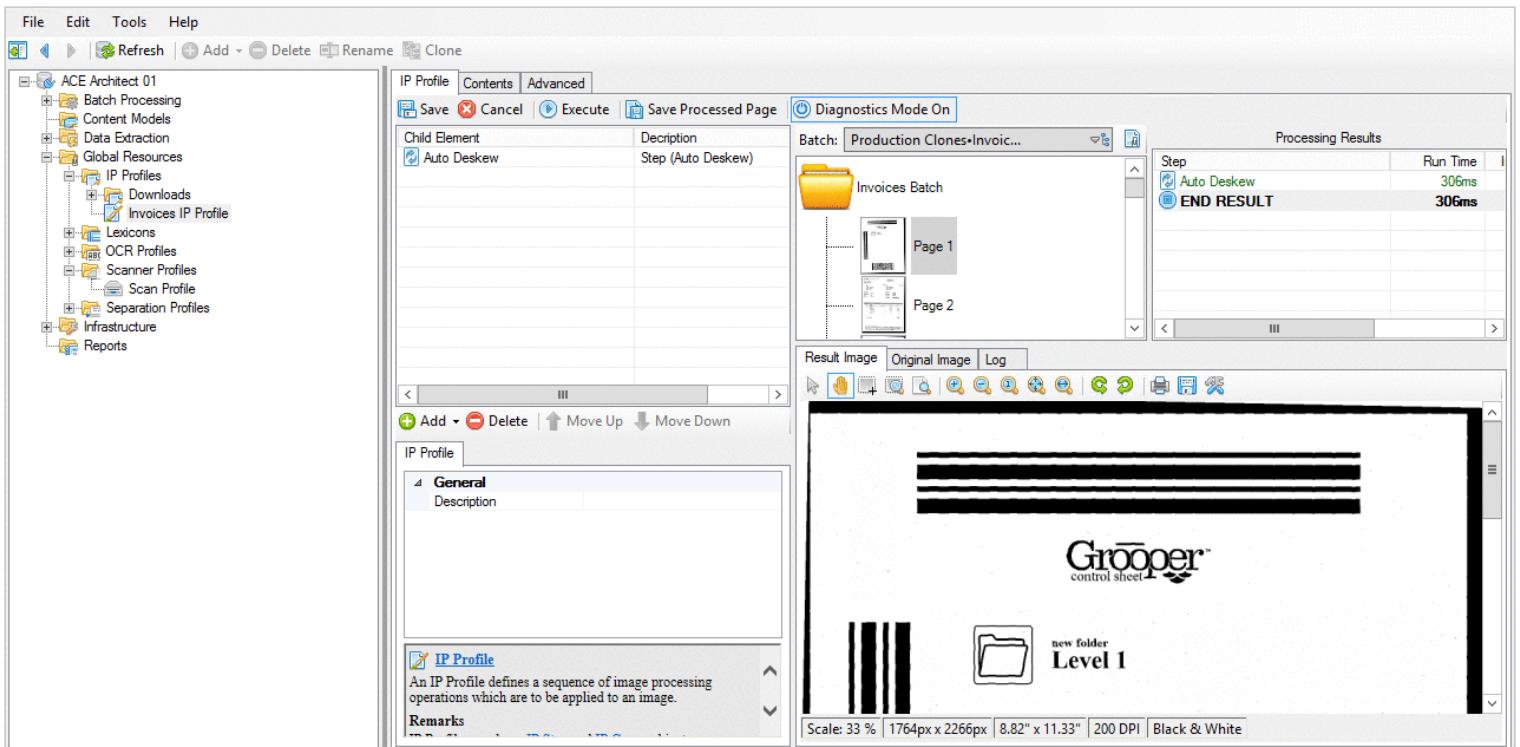
Step 4

In the lower right panel, click the rightmost magnifying glass to zoom the image to fit the panel.

This will make it easier to see.



You can see the effect your commands have on your image by switching back and forth between Result Image and Original Image .



Check it out on page two.

E15 Step 5

The screenshot shows the ACE Architect software interface. On the left, there's a tree view of project resources under 'ACE Architect 01'. In the center, a dialog box titled 'IP Profile' is open, showing a child element 'Auto Deskew' with a description 'Step (Auto Deskew)'. Below this, there's an 'IP Profile' section with tabs for 'General', 'Description', 'IP Profile' (which contains a note about defining sequences of image processing operations), and 'Remarks'. To the right, a 'Diagnostics Mode On' window shows a 'Batch: Production Clones-Invo...' with an 'Invoices Batch' containing 'Page 1' and 'Page 2'. A 'Processing Results' table shows a single step 'Auto Deskew' with a run time of 385ms. At the bottom, a preview window displays an invoice from 'ACME | INTERNATIONAL' to 'Grooper Industries'.

Click Save .

This screenshot is identical to the one above, showing the same software interface and steps. The difference is that the 'IP Profile' dialog now has a message at the top: 'Save the changes to the IP Profile. p (Auto Deskew)'. This indicates that the changes made in Step 5 have been saved.

E8 OE Tip

**E3
13**

Save often!

A good rule of thumb is to save every time you make a change
you know you're going to keep.

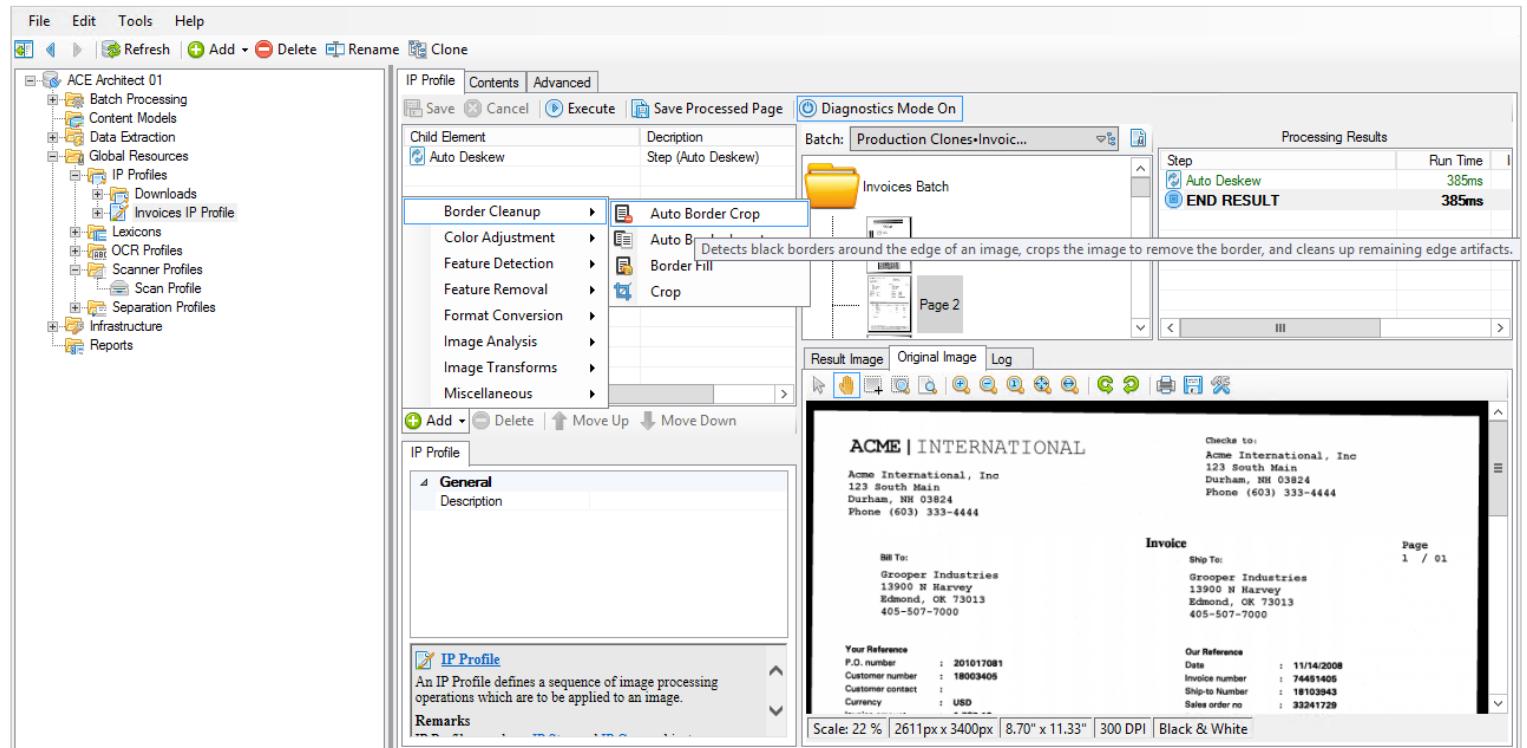
We're done with the Deskew command. Remember that the goal is to make sure the text is displayed in horizontal lines, so try not to focus on the border skew.

Speaking of borders, let's take care of them.

Border cleanup commands

Step 6

Click **Add**, and then **Border Cleanup > Auto Border Crop**.



We aren't going to change any of the default properties, so let's see what happens.

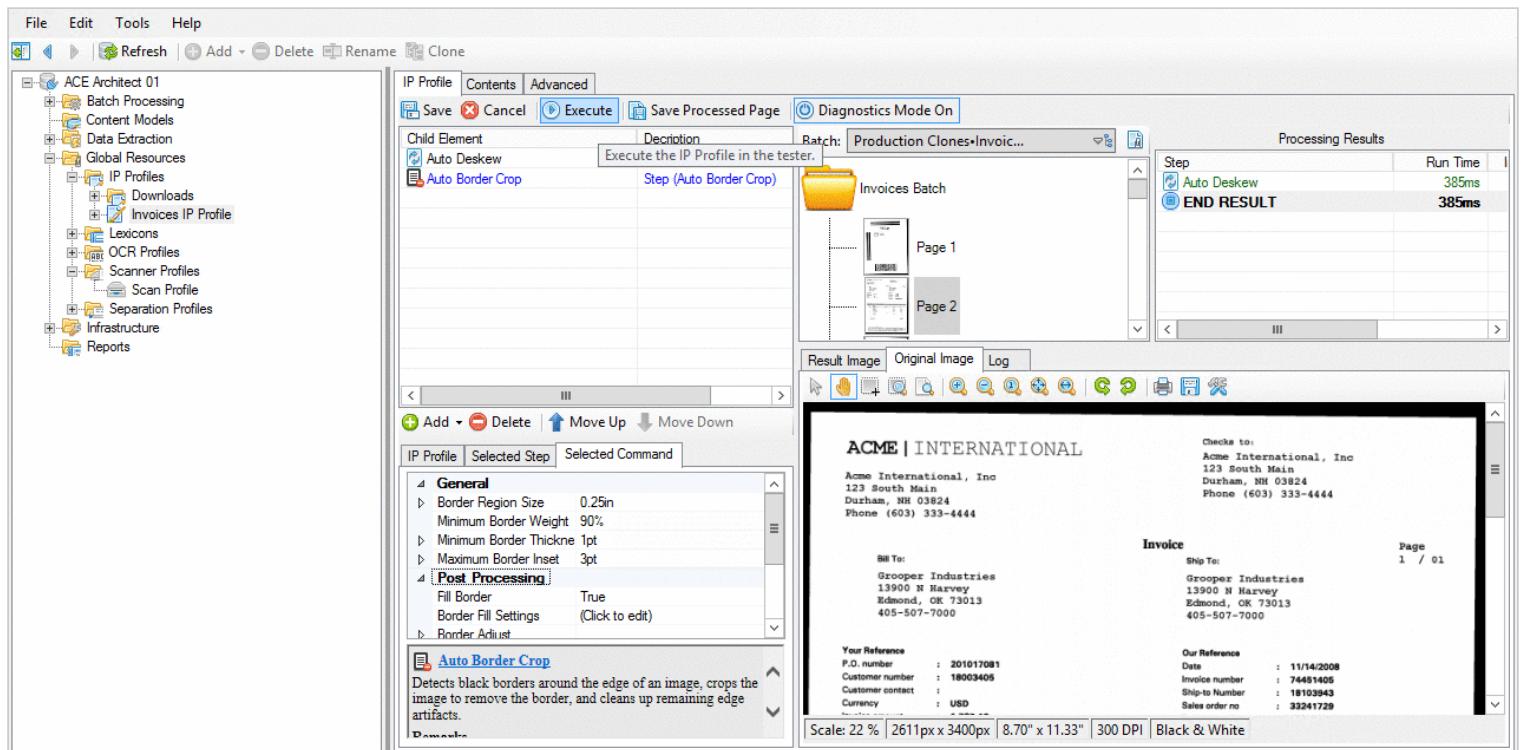
Step 7

Click **Execute**, and click back and forth between the images to see the changes.

Tip

E3
1 2 3

As you're building an IP Profile and adding and testing commands, make sure you are on the [Result Image](#) to see if your configuration is working. It's not unusual to be adding commands and not see any changes in your image, only to realize you never switched over from the [Original Image](#) view!



Note

E3
1 2 3

The [Border Crop](#) command changes the size of your image. You can see this in the information panel below the image itself.

Our Border Crop works pretty well, but there are some images that still have a bit leftover. We don't necessarily want to increase the size of our crop because we risk cutting off any text that might be close to the edge of the page (for example, [Page 7](#) in our batch). We'll take care of that another way.

Step 8

Save the profile.

E3 Step 9

The screenshot shows the ACE Architect 01 software interface. On the left, the project tree shows 'ACE Architect 01' with various modules like Batch Processing, Content Models, Data Extraction, Global Resources, IP Profiles, Lexicons, OCR Profiles, Scanner Profiles, Separation Profiles, Infrastructure, and Reports. The 'IP Profiles' node is expanded, showing 'Downloads' and 'Invoices IP Profile'. The main workspace is divided into several panes:

- IP Profile Editor:** Shows the 'IP Profile' tab selected. A table lists child elements: 'Auto Deskew' (Description: 1) and 'Auto Border Crop' (Description: 2). Below is a detailed configuration pane for 'Auto Border Crop' under 'Post Processing'.
- Diagnostics Mode On:** Shows the 'Batch: Production Clones-Invo...' status.
- Invoices Batch:** Displays two pages: 'Page 1' and 'Page 2'.
- Processing Results:** A table showing steps: 'Auto Deskew' (Run Time: 415ms), 'Auto Border Crop' (Run Time: 437ms), and 'END RESULT' (Run Time: 852ms).
- Result Image:** Shows the processed invoice document with header information for 'ACME | INTERNATIONAL' and 'Grooper Industries'.
- Original Image:** Shows the original invoice document.
- Log:** Shows processing logs.

Select Page 1 in our batch. Click on Add , and then select Border Cleanup > Border Fill.

The screenshot shows the ACE Architect 01 software interface with the same project structure and IP Profile editor as the previous screenshot. The 'Auto Border Crop' step has been modified:

- Border Cleanup:** 'Border Fill' has been added to the list of sub-steps under 'Border Cleanup'.
- Processing Results:** The 'Step' table now includes 'Border Fill' (Run Time: 92ms) along with 'Auto Deskew' and 'Auto Border Crop'.
- Result Image:** The processed invoice document shows a clean version with no borders or artifacts.

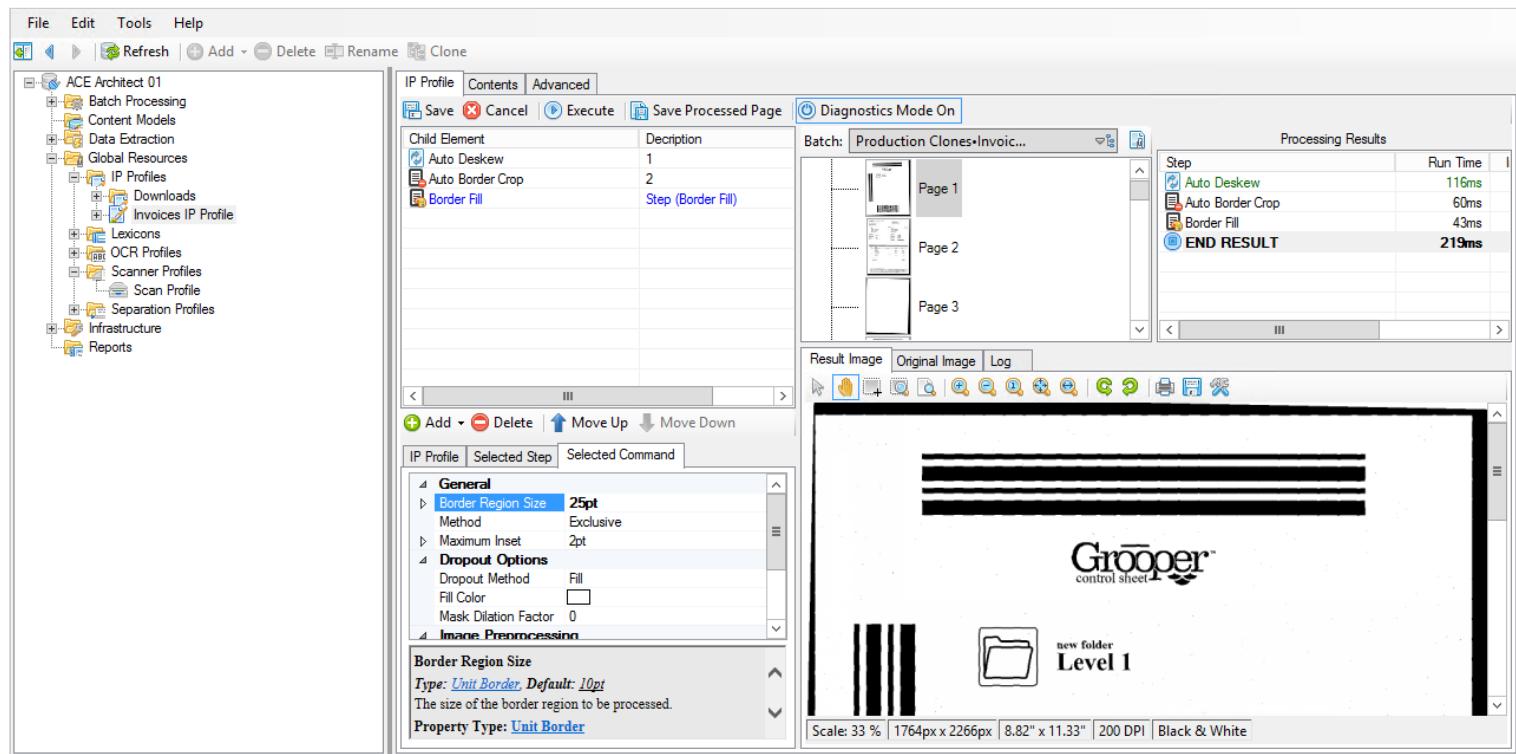
E3 C9 Note E3 C13

When you click on a page in the batch viewer, it automatically runs all commands against that page. The only time you need to use the `Execute` button is when you're adding commands and you want to run them against your current page without navigating away from it.

If you click on `Execute`, nothing happens. That tells us that the default properties for this command probably need to be tweaked for us to see results.

Step 10

1. Select the `Border Fill` command.
2. Change the `Border Region Size` property to `25pt`.

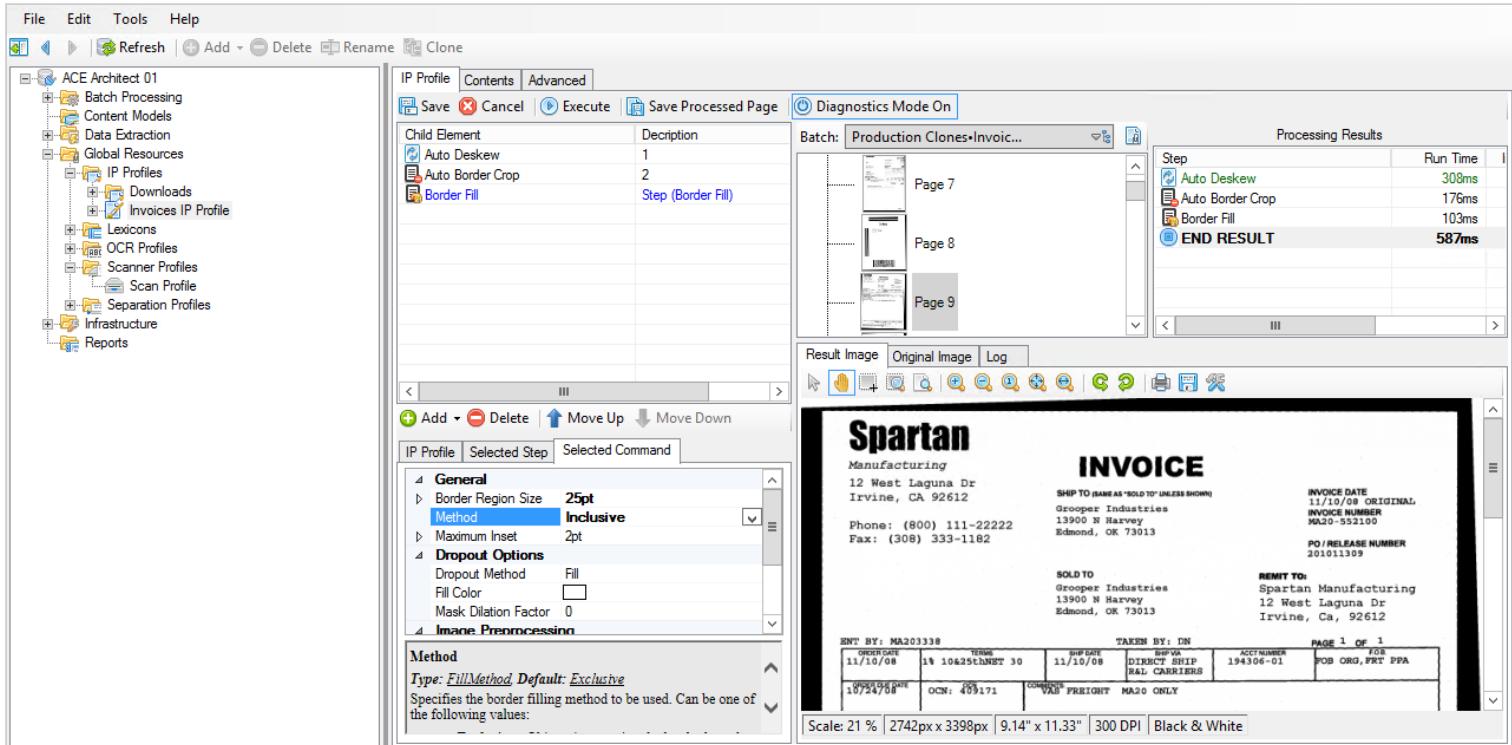


This is increasing how far inward from the border Grooper will look when running this command.

Yet, once again, `Execute` yields no results.

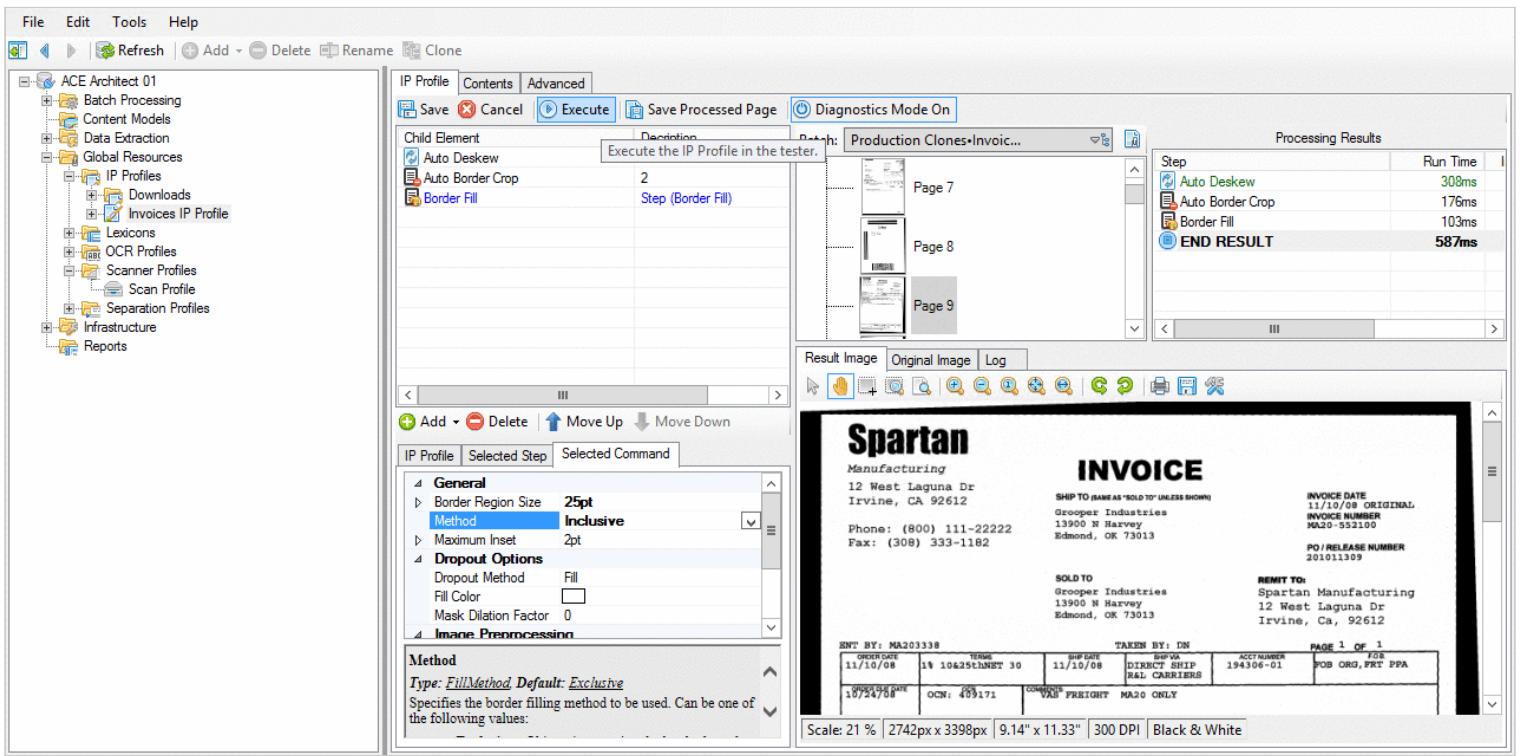
Step 11

1. Select Page 9 in our batch.
2. Change the Method property to Inclusive .



Step 12

Click Execute and check out the results.



Note

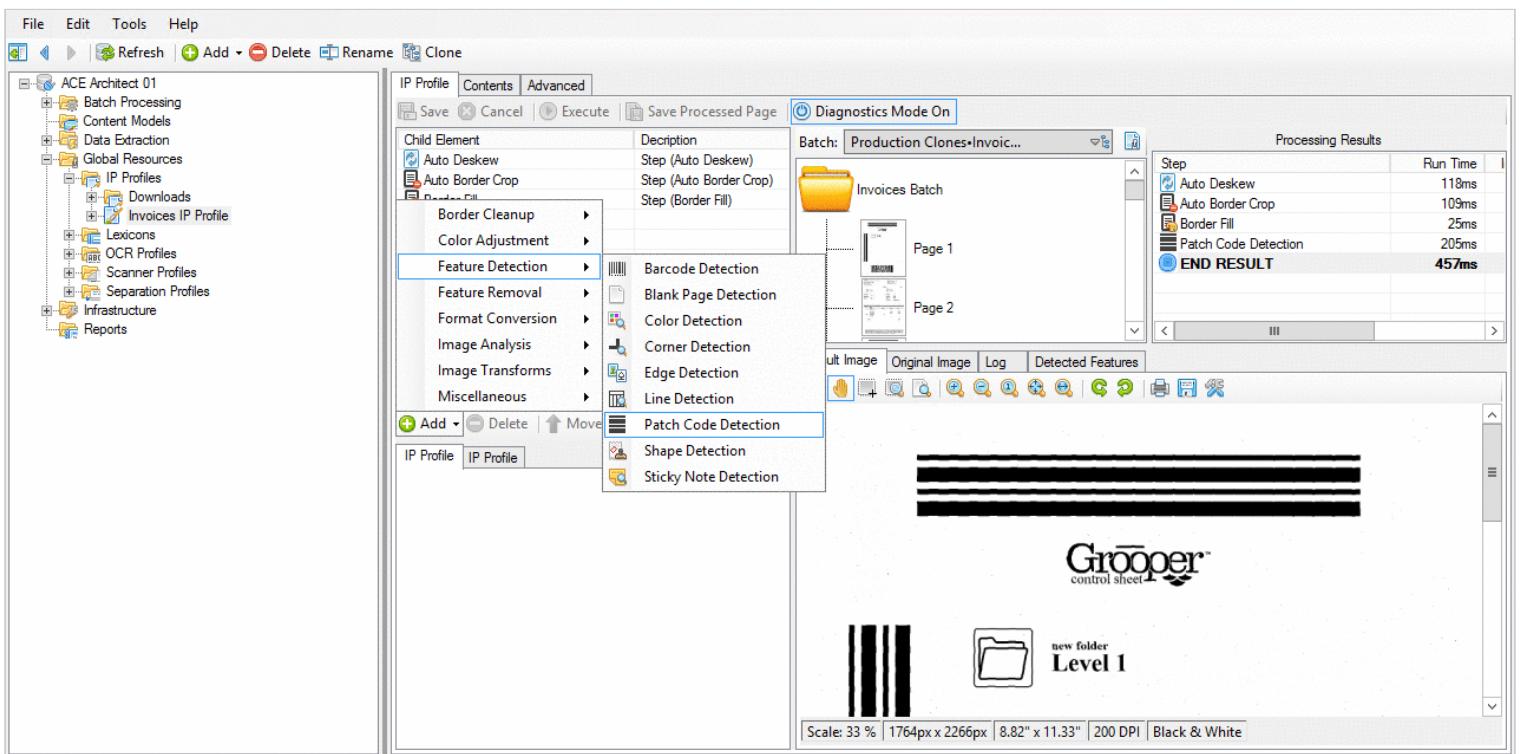
Unlike the `Border Crop` command, `Border Fill` does not change the size of your image. This is because this command serves to fill in the edge of the page with whatever color is set in the `Fill` property.

We're not changing the size of the page, only the stuff on it.

Step 13

Click `Add` and select `Feature Detection > Patch Code Detection`.

Save the profile.



Adding to the Batch Process

Step 1

Navigate to (root) > Batch Processing > Processes > Working > Invoices Process .

Name	Order	Activity Type
Scan	1	Scan

Step 2

Click **Add Step...**.

The screenshot shows the ACE Architect 01 software interface. On the left is a navigation tree with categories like Batch Processing, Processes, and Reports. The main area is titled "Batch Process Properties". It has tabs for General, Scripting, Contents, and Advanced. Under General, there's a section for "Content Type" with "Project" selected. Below that is "Publishing Info" with a publish date of 8/6/2018 4:16:37 PM and a published version of c1f25984f516-4440-a5d3-d73ef01. A "Steps in Batch Process" table is shown, containing one row with "Scan" as the name and "Add a step to the Batch Process." as the activity type. A tooltip for "Batch Process" defines it as a sequence of steps for processing information. A "Remarks" section notes that a Batch Process is a sequence of individual steps, each specifying an activity.

Step 3

In the **Properties** for our new empty step, change the **Activity Type** to **Image Processing**.

Set the **IP Profile** on the right to the IP Profile we just got making.

Step 4

Save and Publish our process.

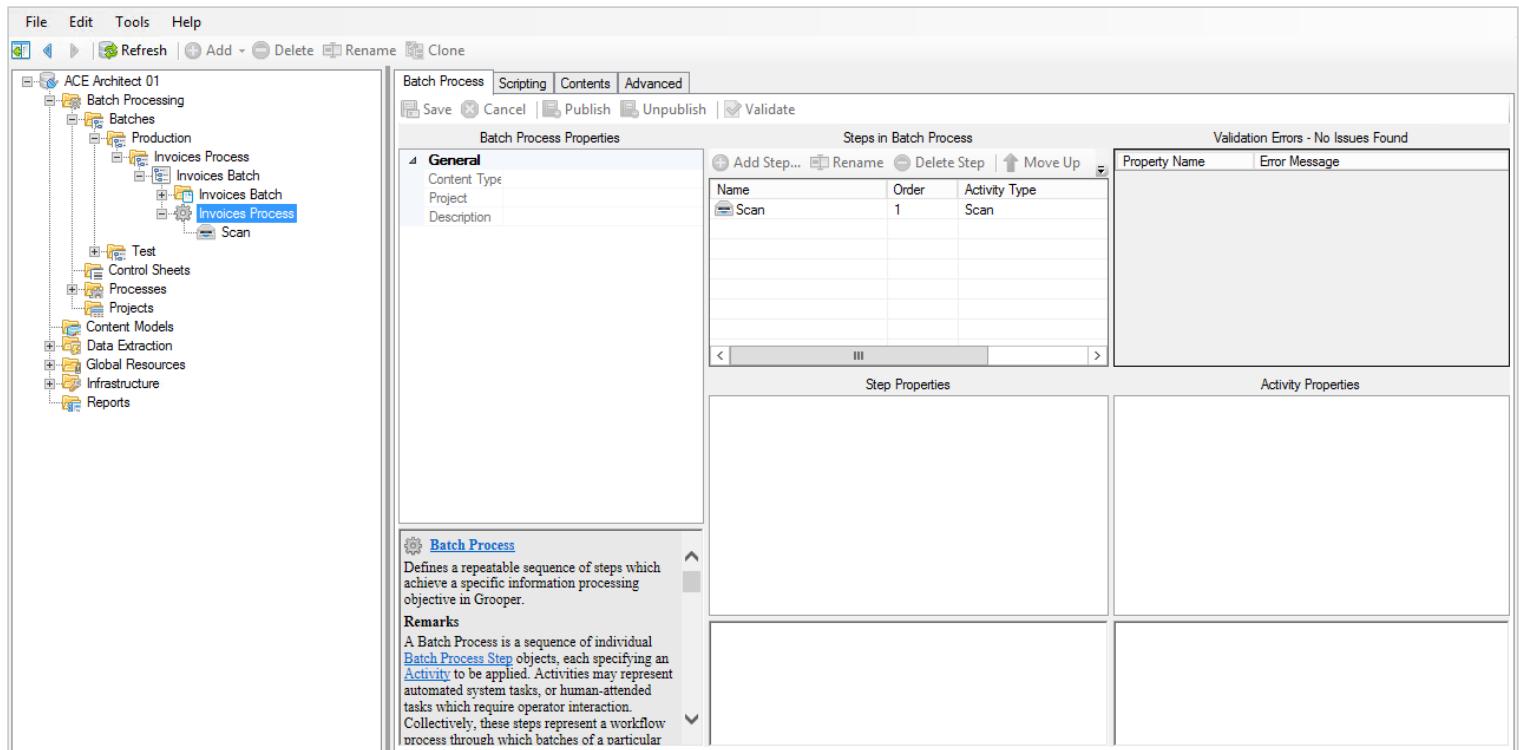
Step 4

Save and Publish our process.

Updating the production batch

Step 1

Navigate to `(root) > Batch Processing > Batches > Production > Invoices Process > Invoices Batch > Invoices Process`.



Notice how the Batch Process that is attached to this batch didn't receive the new step we just added. Remember that when this batch was created, the process had only the "Scan" step. We want to tell this batch to check out the changes we made to the process so that it can run through the new Image Processing step.

Step 2

Navigate back up to `(root) > Batch Processing > Batches > Production > Invoices Process`.

Select the batch from the list, and then go to `Batch > Update Process...`.

Screenshot of the ACE Architect 01 software interface showing the Batch Processing module. The left sidebar contains navigation links like File, Edit, Tools, Help, Refresh, Add, Delete, Rename, and Clone. The main area displays a tree view of batches under 'Batch Processing' and a grid view of batches. A context menu is open over a row named 'Invoices Batch'. The menu items include Clone To Test, Reset..., Update Process..., Share, New..., New From, Delete, and Properties... . A tooltip for 'Update Process...' states: 'Updates or changes the Batch Process associated with this batch.' Below the grid is a 'Task Status' chart showing a single task named 'Scan' at 100% completion.

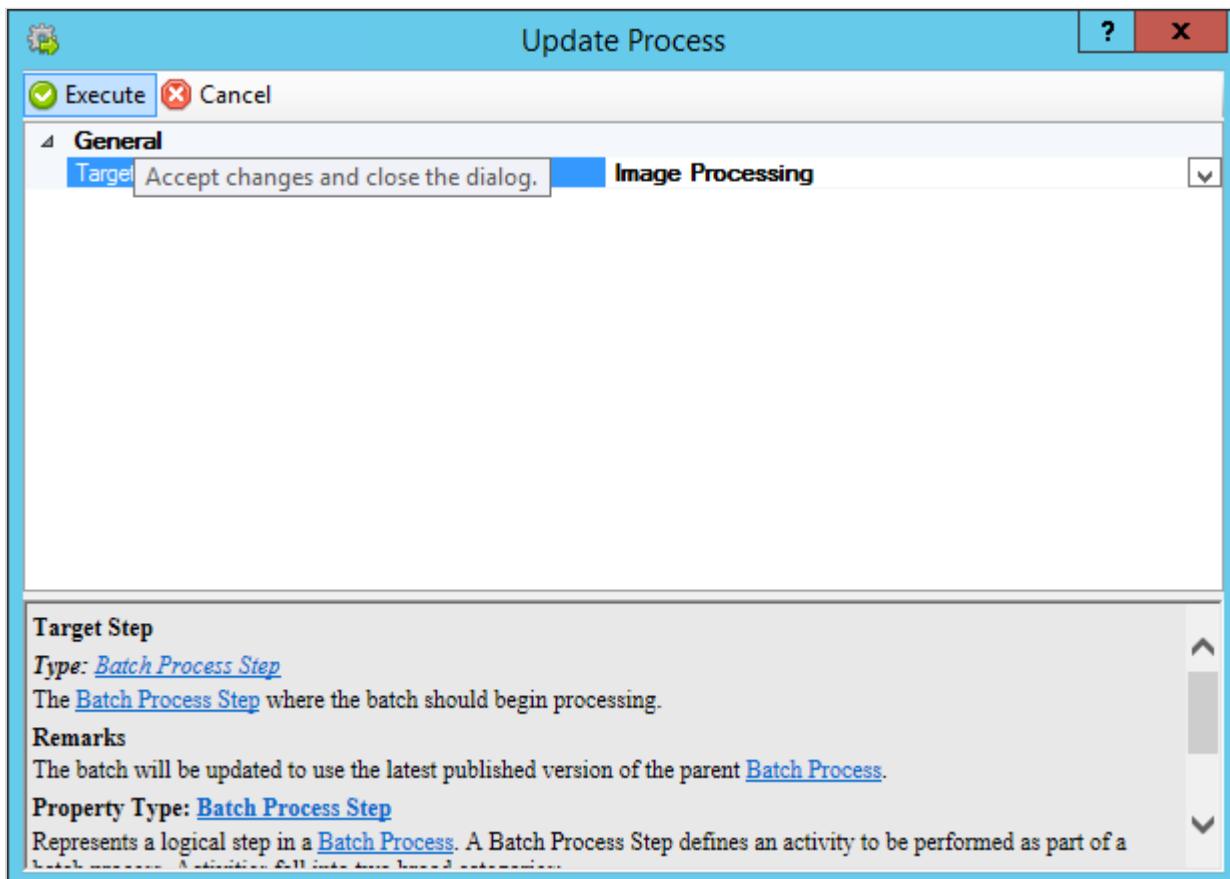
The **Update Process** window will appear.

Screenshot of the 'Update Process' dialog box. The title bar says 'Update Process' with a help (?) icon and a close (x) icon. Below the title bar are 'Execute' and 'Cancel' buttons. The main area has a 'General' tab selected, showing a dropdown labeled 'Target Step' with 'Scan' currently selected. At the bottom of the dialog, there is a section titled 'Batch - Update Process' with the note: 'Updates or changes the [Batch Process](#) associated with this batch.' It also includes 'Remarks' and 'See Also' sections.

Step 3

Select **Target Step**. From the dropdown on the right, select the **Image Processing** step, and then click **Execute**.

We're telling Grooper, "Update the process on this batch. We're using this process, and I want you to start processing at this step."



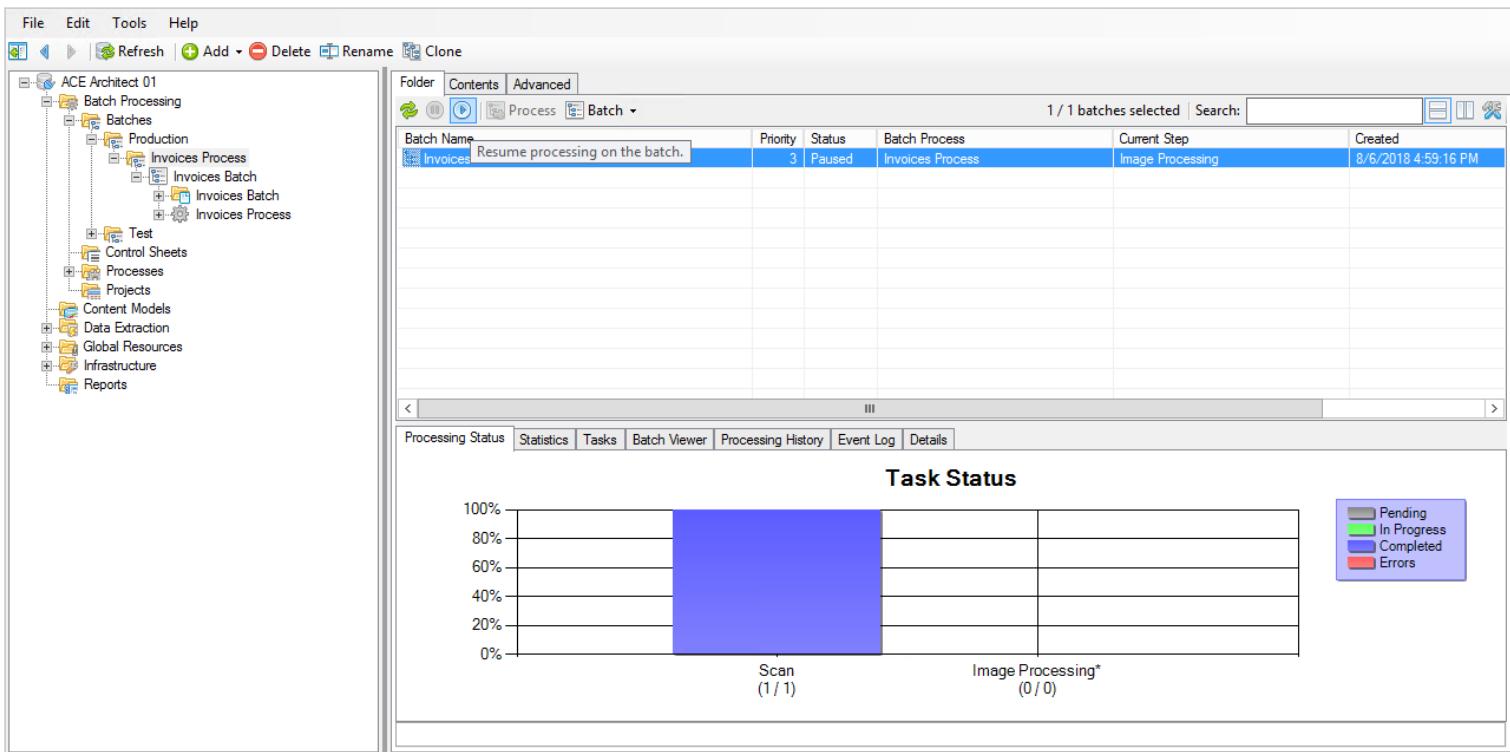
We will return to Grooper Design Studio, only now our batch's [Task Status](#) panel has another step in it.

The screenshot shows the Grooper Design Studio interface. On the left is a navigation tree with 'ACE Architect 01' expanded, showing 'Batches', 'Production', 'Test', 'Control Sheets', 'Processes', 'Projects', 'Content Models', 'Data Extraction', 'Global Resources', 'Infrastructure', and 'Reports'. In the center, there's a 'Batch Processing' view with a 'Batches' table. The table has columns: Batch Name, Priority, Status, Batch Process, Current Step, and Created. One row is selected: 'Invoices Batch' with Priority 3, Status Paused, Batch Process 'Invoices Process', Current Step 'Image Processing', and Created '8/6/2018 4:59:16 PM'. Below this is a 'Task Status' panel. It has tabs: Processing Status, Statistics, Tasks, Batch Viewer, Processing History, Event Log, and Details. Under 'Task Status', there's a chart titled 'Scan' (1 / 1) and 'Image Processing*' (0 / 0). A legend indicates task status colors: Pending (grey), In Progress (green), Completed (blue), and Errors (red).

The task doesn't have a progress bar because the batch is still paused.

Step 4

Click on the **Resume Batch** button, and then **Execute** on the confirmation window that appears.



The screenshot shows the ACE Architect 01 software interface. The left sidebar displays a hierarchical tree structure under 'ACE Architect 01', including 'Batch Processing', 'Batches', 'Production', 'Invoices Process', 'Invoices Batch', and 'Invoices Process'. The main workspace shows a table titled 'Batch' with one row selected. The table columns are 'Batch Name', 'Priority', 'Status', 'Batch Process', 'Current Step', and 'Created'. The 'Batch Name' column contains 'Invoices' with the note 'Resume processing on the batch.' The 'Status' column shows 'Paused'. The 'Batch Process' column is 'Invoices Process'. The 'Current Step' column is 'Image Processing'. The 'Created' column is '8/5/2018 4:59:16 PM'. Below the table is a 'Task Status' section with a chart. The chart has two categories: 'Scan' (1 / 1) and 'Image Processing*' (0 / 0). A legend indicates task status: Pending (gray), In Progress (green), Completed (blue), and Errors (red). The 'Completed' status is represented by a blue bar for the 'Scan' task.

Now the batch has a gray progress bar for the **Image Processing** step. We know this means that it's ready and waiting for us to tell it to start this activity.

Before we do that, let's look at what we just did.

Step 5

Navigate to **(root) > Batch Processing > Batches > Production > Invoices Process > Invoices Batch > Invoices Process** and expand it.

The screenshot shows the ACE Architect 01 application window. On the left is a navigation tree with categories like Batch Processing, Production, Invoices Process, and Test. The main area displays the 'Batch Process Properties' dialog. At the top are tabs: Batch Process, Scripting, Contents, and Advanced. Below them are buttons: Save, Cancel, Publish, Unpublish, and Validate. A table titled 'Steps in Batch Process' lists steps: Scan (Order 1, Activity Type Scan) and Image Processing (Order 2, Activity Type Image Processing). To the right is a table titled 'Validation Errors - No Issues Found'. A detailed description of 'Batch Process' is shown in a sidebar.

Now we see that there's an `Image Processing` step that wasn't here the last time we looked. This is the result of the `Update Process...` that we did to the batch.

The batch received the latest updates from our `Invoices Process`, which had a new step, so now it's ready to run it!

Step 6

Navigate back up to `(root) > Batch Processing > Batches > Production > Invoices Process`, select the batch, and click the `Process` button.

The screenshot shows the ACE Architect 01 software interface. On the left is a file tree with categories like Batch Processing, Production, Test, Processes, Projects, Content Models, Data Extraction, Global Resources, Infrastructure, and Reports. The 'Batch Processing' node is expanded, showing 'Batches' (Production, Invoices Process, Test), 'Control Sheets', and 'Processes'. The 'Invoices Process' node under 'Production' is also expanded. On the right, there's a main processing window with a toolbar at the top. Below the toolbar is a table titled 'Batch Name' with one row: 'Invoices Batch', Priority '3', Status 'Ready', Batch Process 'Invoices Process', Current Step 'Image Processing', and Created '8/6/2018 4:59:16 PM'. Below the table is a navigation bar with tabs: Processing Status, Statistics, Tasks, Batch Viewer, Processing History, Event Log, and Details. Under 'Processing Status', there's a 'Task Status' section with a chart. The chart has two segments: 'Scan (1 / 1)' in blue and 'Image Processing* (0 / 153)' in grey. A legend on the right indicates: Pending (grey), In Progress (green), Completed (blue), and Errors (red).

You'll see the **Grooper Unattended Client** window appear. This is the part of the Grooper suite that processes **unattended activities**.

When this shows up, you don't need to press anything - it will start working automatically.

The screenshot shows the Grooper Unattended Client window. On the left, there's a configuration pane with sections for General (Thread Pool: Default, Thread Count: 1, Thread Priority: Normal, Batch: Invoices Batch, Step: Image Processing) and Processing (Threads Running: 1, Start Time: 10:18 AM, Time Elapsed: 00:00:09, Tasks Completed: 20, Errors: 0, Tasks/Minute: 124.27, Est. Time Remaining: 00:01:03). On the right, there's a main processing window with a toolbar at the top. Below the toolbar is a table titled 'Batch Name' with one row: 'Invoices Batch', Priority '3', Status 'Ready', Batch Process 'Invoices Process', Current Step 'Image Processing', and Created '8/6/2018 4:59:16 PM'. Below the table is a navigation bar with tabs: Processing Status, Statistics, Tasks, Batch Viewer, Processing History, Event Log, and Details. Under 'Processing Status', there's a 'Task Status' section with a chart. The chart has two segments: 'Scan (1 / 1)' in blue and 'Image Processing* (18 / 153)' in blue. A legend on the right indicates: Pending (grey), In Progress (green), Completed (blue), and Errors (red).

If you take a look at the **Image Processing** step, you can see it working on the pages.

When it's done, it will have a blue progress bar similar to the Scan step.

The screenshot shows the ACE Architect software interface. On the left is a navigation tree with categories like Batch Processing, Production, Invoices Process, Test, Processes, Projects, Content Models, Data Extraction, Global Resources, Infrastructure, and Reports. The main area has tabs for Folder, Contents, and Advanced. Under Advanced, there are tabs for Process, Batch, and Batch. A table titled 'Batch' lists one entry: 'Invoices Batch' with Priority 3, Status Working, Batch Process 'Invoices Process', Current Step 'Image Processing', and Created date '8/6/2018 4:59:16 PM'. Below the table is a 'Task Status' section with a bar chart. The y-axis ranges from 0% to 100%. The x-axis shows two tasks: 'Scan (1 / 1)' and 'Image Processing* (153 / 153)'. The 'Image Processing*' task is shown in blue, indicating it is in progress. A legend on the right identifies colors: grey for Pending, green for In Progress, blue for Completed, and red for Errors.

But how do we know that it actually worked, and that our images have been correctly cleaned up?

Image Review

In a production environment, you might want to have users review the images before they move on to any other steps in your Batch Process.

In Grooper, we can do this with an `Image Review` step. This is an attended activity that will display the batch to the users and give them an opportunity to approve or rescan any images that need it.

Adding an Image Review step

Step 1

1. Navigate to `(root) > Batch Processing > Batches > Production > Invoices Process`.
2. Click `Add Step...`.
3. Under `Properties of Image Review Step`, set the `Activity Type` to `Image Review`.
4. Save and Publish the process.

Batch Process Properties

Steps in Batch Process

Name	Order	Activity Type
Scan	1	Scan
Image Processing	2	Image Processing
Image Review	3	Image Review

Properties of Image Review Step

General

- Activity Type: Image Review
- Activity: Image Review
- Scope: Batch
- Description:

Expressions

- Should Submit Expression
- Next Step Expression

Properties of Image Review Activity

General

- Pre-Defined Flag Messages: Image Review Messages
- Allowed IP Profiles: (0 IP Profile objects)
- Require Review of All Pages: True
- Allow Completion with Flags: True

UI Configuration

- Command Options: (0 Command Options objects)
- User Activity Timeout: 0

Batch Process
Defines a repeatable sequence of steps which achieve a specific information processing objective in Grooper.

Remarks
A Batch Process is a sequence of individual Batch Process Step objects, each specifying an Activity to be applied. Activities may represent automated system tasks, or human-attended tasks which require operator interaction. Collectively, these steps represent a workflow process through which batches of a particular

Batch Process Step
Represents a logical step in a Batch Process.

Remarks
A Batch Process Step defines an activity to be performed as part of a batch process. Activities fall into two broad categories:

- Attended Activities - Activities which require a human

Image Review
Image Review is an attended activity which provides a user interface optimized for quickly reviewing and/or correcting Batch Page images.

Remarks
Launches Attended Client with the Batch Viewer and Thumbnail

Now we need to go update our batch.

Step 2

1. Navigate to (root) > Batch Processing > Batches > Production > Invoices Process .
2. Pause the batch.
3. Select the Batch dropdown and select Update Process... .
4. In the Update Process window, select the Image Review step from the Target Step dropdown.
5. Click Execute .

File Edit Tools Help

Refresh | Add | Delete | Rename | Clone

ACE Architect 01

Batch Processing

- Batches
 - Production
 - Invoices Process
- Test
- Control Sheets
- Processes
- Projects

Content Models

- Data Extraction
- Global Resources
- Infrastructure
- Reports

Folder Contents Advanced

Process Batch

1 / 1 batches selected | Search: []

Batch Name	Priority	Status	Batch Process	Current Step	Created
Invoices Batch	3	Completed	Invoices Process	(none)	8/6/2018 4:59:16 PM

Processing Status Statistics Tasks Batch Viewer Processing History Event Log Details

Task Status

Pending In Progress Completed Errors

Update Process

Execute Cancel

General

Target Step Scan

Target Step

Type: [Batch Process Step](#)

The [Batch Process Step](#) where the batch should begin processing.

Remarks

The batch will be updated to use the latest published version of the parent [Batch Process](#).

Property Type: [Batch Process Step](#)

Represents a logical step in a [Batch Process](#). A Batch Process Step defines an activity to be performed as part of a

We should see another step in the Task Status panel, just like when we added Image Processing.

File Edit Tools Help

Refresh | Add | Delete | Rename | Clone

ACE Architect 01

- Batch Processing
 - Batches
 - Production
 - Invoices Process
 - Test
 - Control Sheets
 - Processes
 - Projects
- Content Models
- Data Extraction
- Global Resources
- Infrastructure
- Reports

Folder Contents Advanced

Process Batch

Batch Name Invoices Batch **Priority** 3 **Status** Paused **Batch Process** Invoices Process **Current Step** Image Review **Created** 8/6/2018 4:59:16 PM

Processing Status Statistics Tasks Batch Viewer Processing History Event Log Details

Task Status

Scan (1 / 1)	Image Processing (153 / 153)	Image Review* (0 / 0)
-----------------	---------------------------------	--------------------------

Pending In Progress Completed Errors

Step 3

Resume the batch.

File Edit Tools Help

Refresh | Add | Delete | Rename | Clone

ACE Architect 01

- Batch Processing
 - Batches
 - Production
 - Invoices Process
 - Test
 - Control Sheets
 - Processes
 - Projects
- Content Models
- Data Extraction
- Global Resources
- Infrastructure
- Reports

Folder Contents Advanced

Process Batch

Batch Name Invoices | Resume processing on the batch. **Priority** 3 **Status** Paused **Batch Process** Invoices Process **Current Step** Image Review **Created** 8/6/2018 4:59:16 PM

Processing Status Statistics Tasks Batch Viewer Processing History Event Log Details

Task Status

Scan (1 / 1)	Image Processing (153 / 153)	Image Review* (0 / 0)
-----------------	---------------------------------	--------------------------

Pending In Progress Completed Errors

Step 4

Click **Process** to start this activity.

The screenshot shows the ACE Architect 01 software interface. On the left, there is a navigation tree with categories like ACE Architect 01, Batch Processing, Control Sheets, Processes, Projects, Content Models, Data Extraction, Global Resources, Infrastructure, and Reports. Under Batch Processing, there are sub-folders for Batches, Production, Invoices Process, and Test. The main workspace is titled "Batch Processing" and shows a table with one row selected. The table columns are Batch Name, Priority, Status, Batch Process, Current Step, and Created. The selected row is "Invoices Batch" with the status "Process the selected batch through the current step." The "Batch Process" column shows "Invoices Process". The "Current Step" column shows "Image Review". The "Created" column shows "8/6/2018 4:59:16 PM". Below the table, there is a "Task Status" section with a bar chart. The chart has three segments: "Scan" (1/1), "Image Processing" (153 / 153), and "Image Review" (0 / 1). A legend indicates that grey represents Pending, green represents In Progress, blue represents Completed, and red represents Errors. The "In Progress" status is shown in blue for the first two segments.

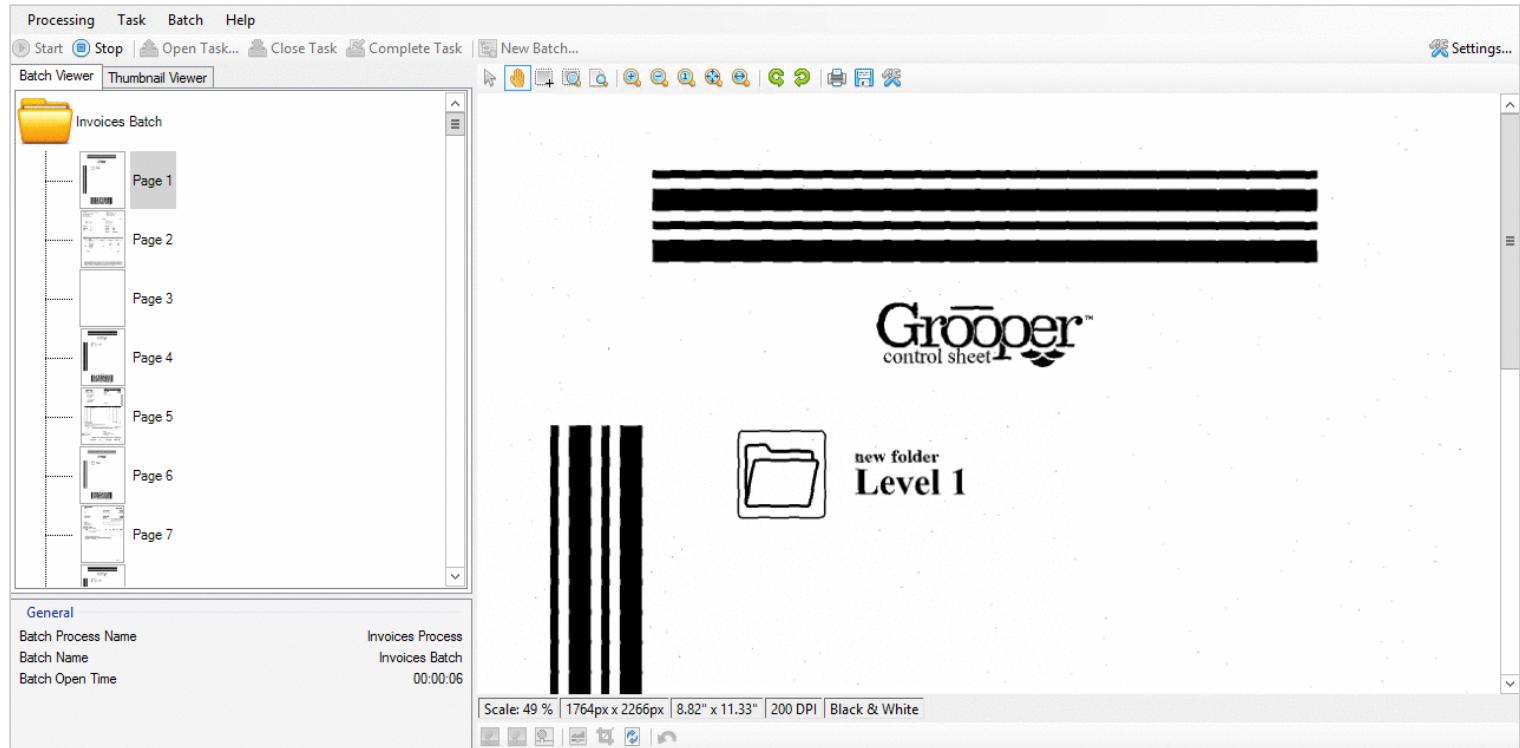
Image Review is an attended activity, so clicking **Process** will open up the **Grooper Attended Client** module (like it did when we scanned).

The screenshot shows the Grooper Attended Client software interface. The top menu includes Processing, Task, Batch, and Help. The toolbar includes Start, Stop, Open Task..., Close Task, Complete Task, New Batch..., and Settings... buttons. The main area is titled "Batch Viewer" and shows a "Thumbnail Viewer" window. Inside the viewer, there is a folder icon labeled "Invoices Batch" containing thumbnails for "Page 1" through "Page 7". To the right of the viewer, there is a large preview area showing a document page with horizontal black bars and the "Grooper control sheet" logo. Below the viewer, there is a "new folder" button and the text "Level 1". At the bottom, there is a "General" section with fields for "Batch Process Name" (Invoices Process), "Batch Name" (Invoices Batch), and "Batch Open Time" (00:00:06). The bottom status bar displays "Scale: 49 %", "1764px x 2266px", "8.82" x 11.33", "200 DPI", and "Black & White".

Opening Image Review

Step 5

Click on the **Thumbnail Viewer** tab.

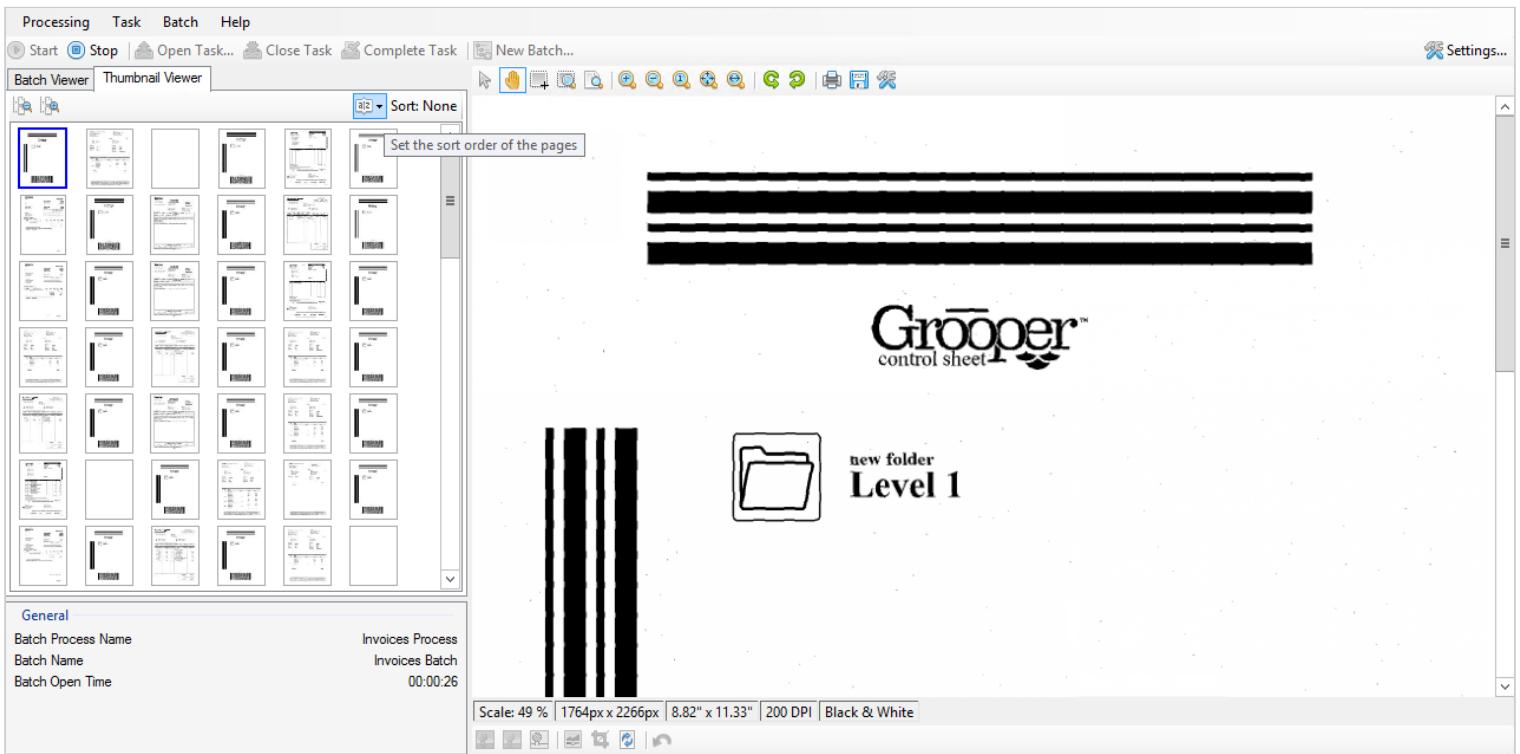


This gives us an easy way to see all of the thumbnails of the batch at a glance.

If you haven't noticed yet, there are a few blank pages in this batch. We don't really care about processing those, so let's get rid of them.

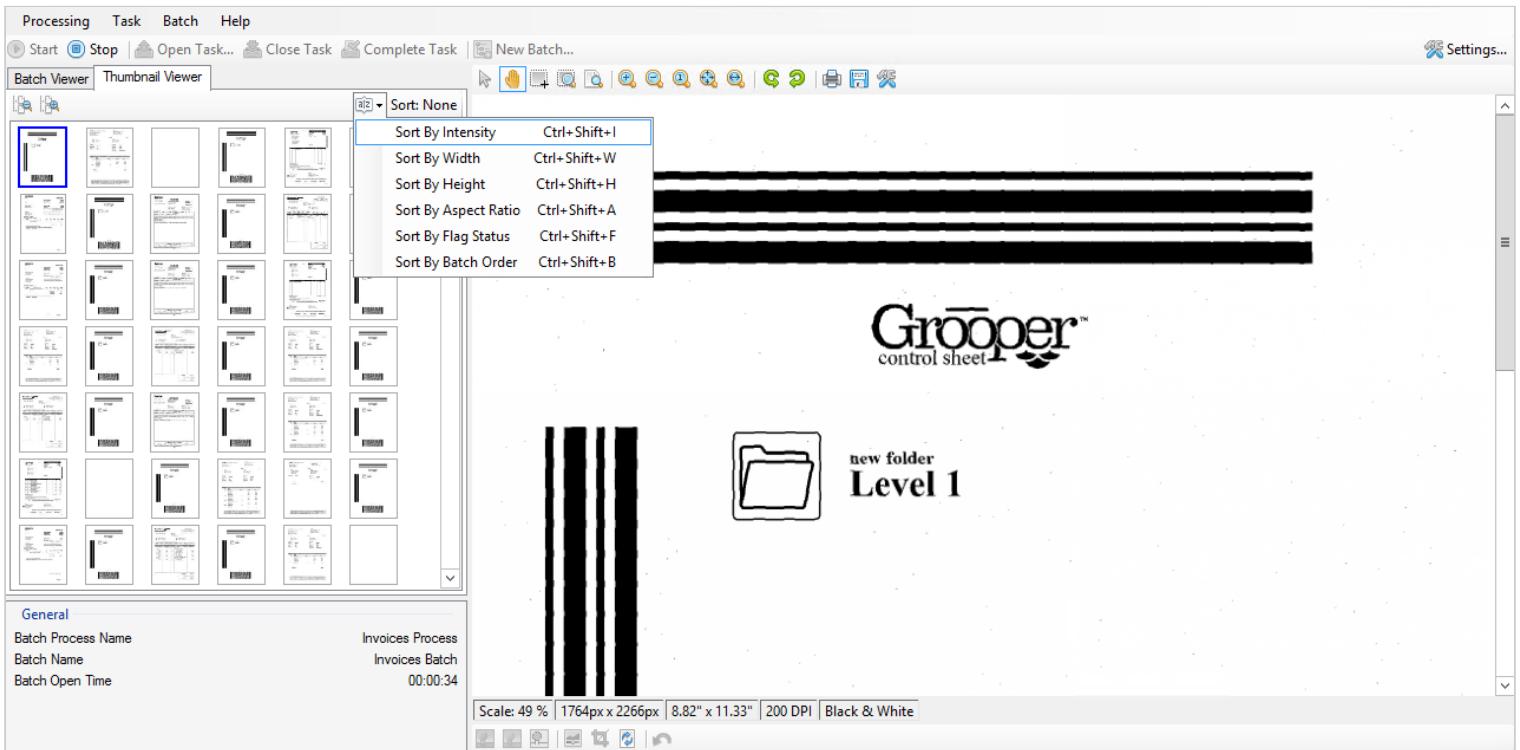
Step 6

In the **Thumbnail Viewer**, click the **Sort** button in the upper right corner.



E15 Step 7

Select the Sort By Intensity option.

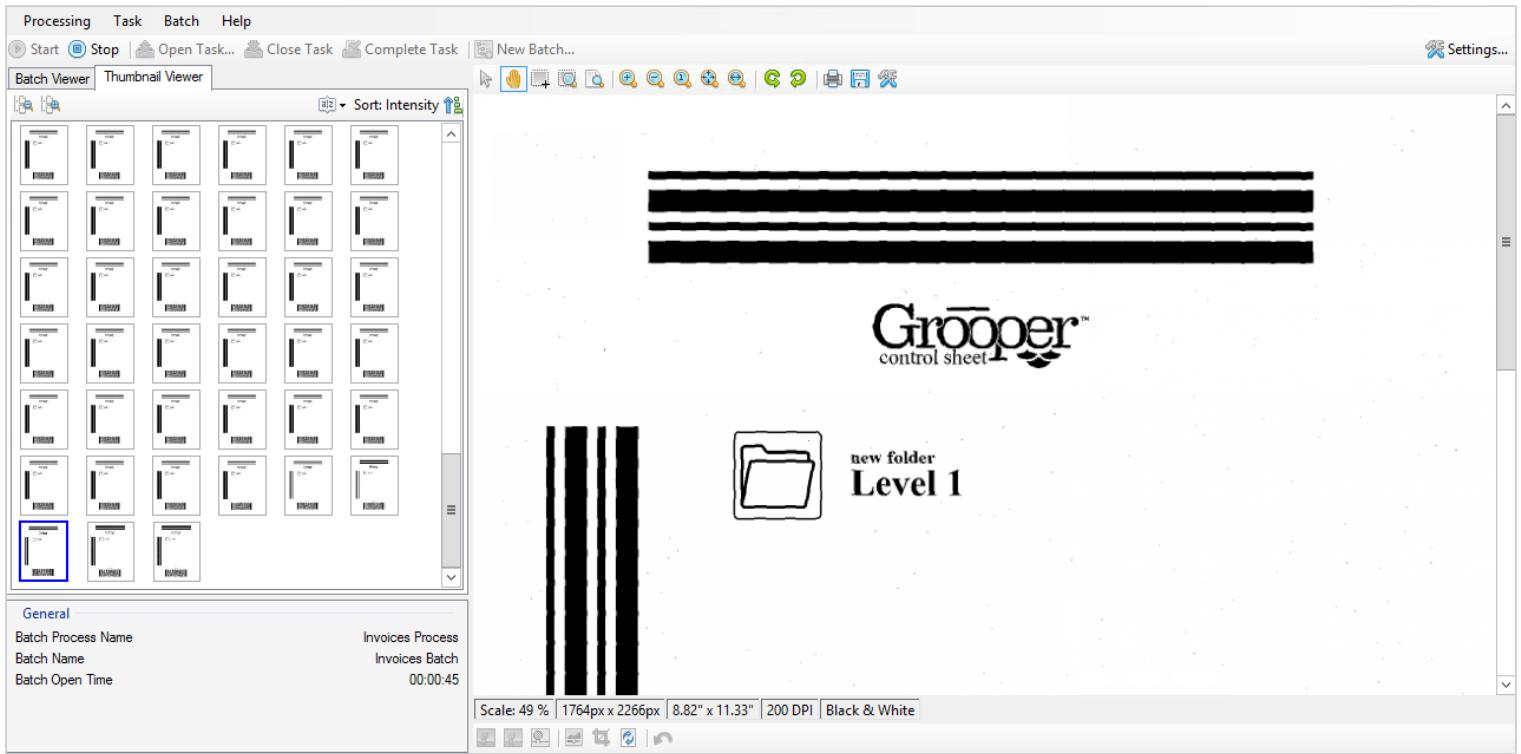


This will arrange the images by how intense they are, based on the black pixel count. Images will be arranged from top to bottom by least amount of black pixels to most.

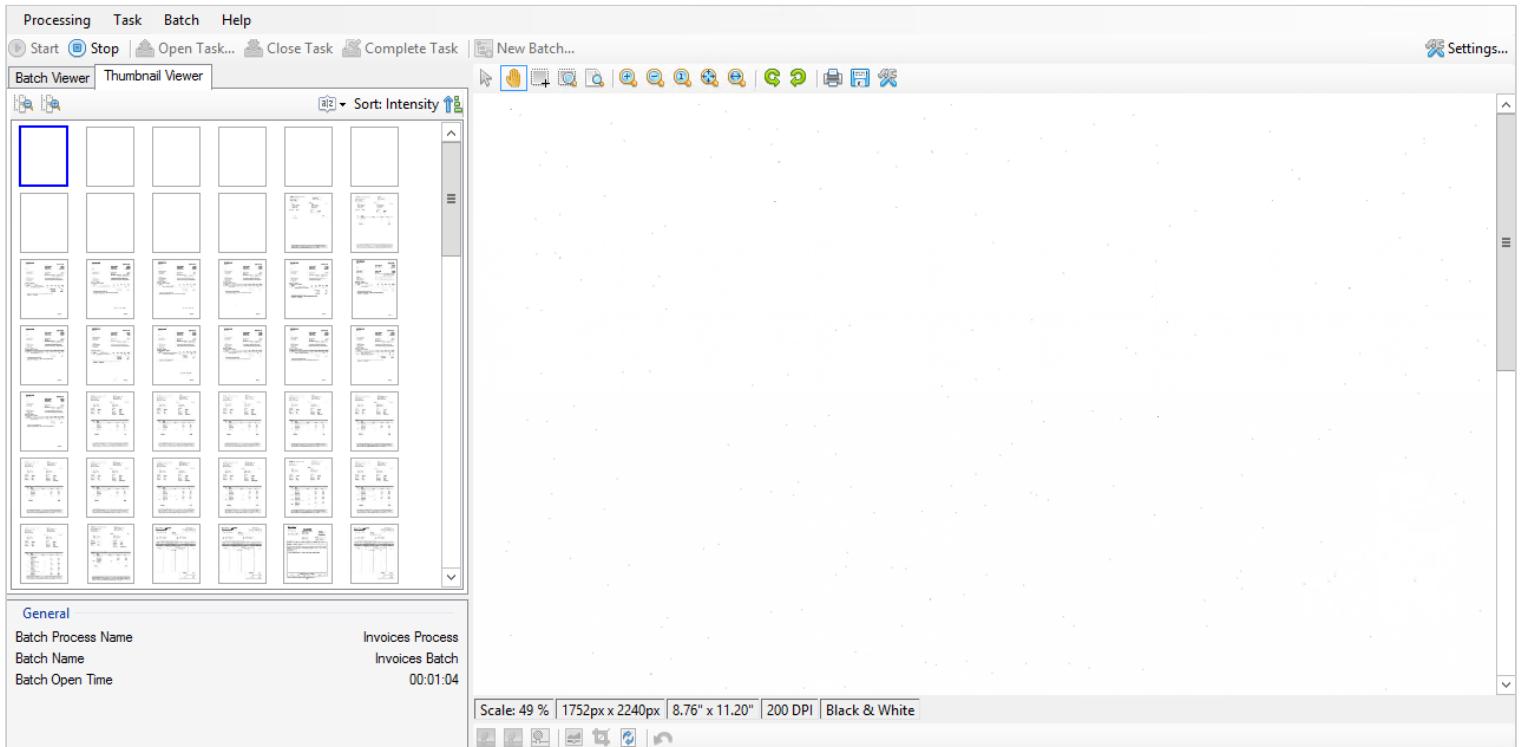
Note

E 3
1 3

This does not permanently arrange the images in this order. This is just a technique to see how many blank pages we have.



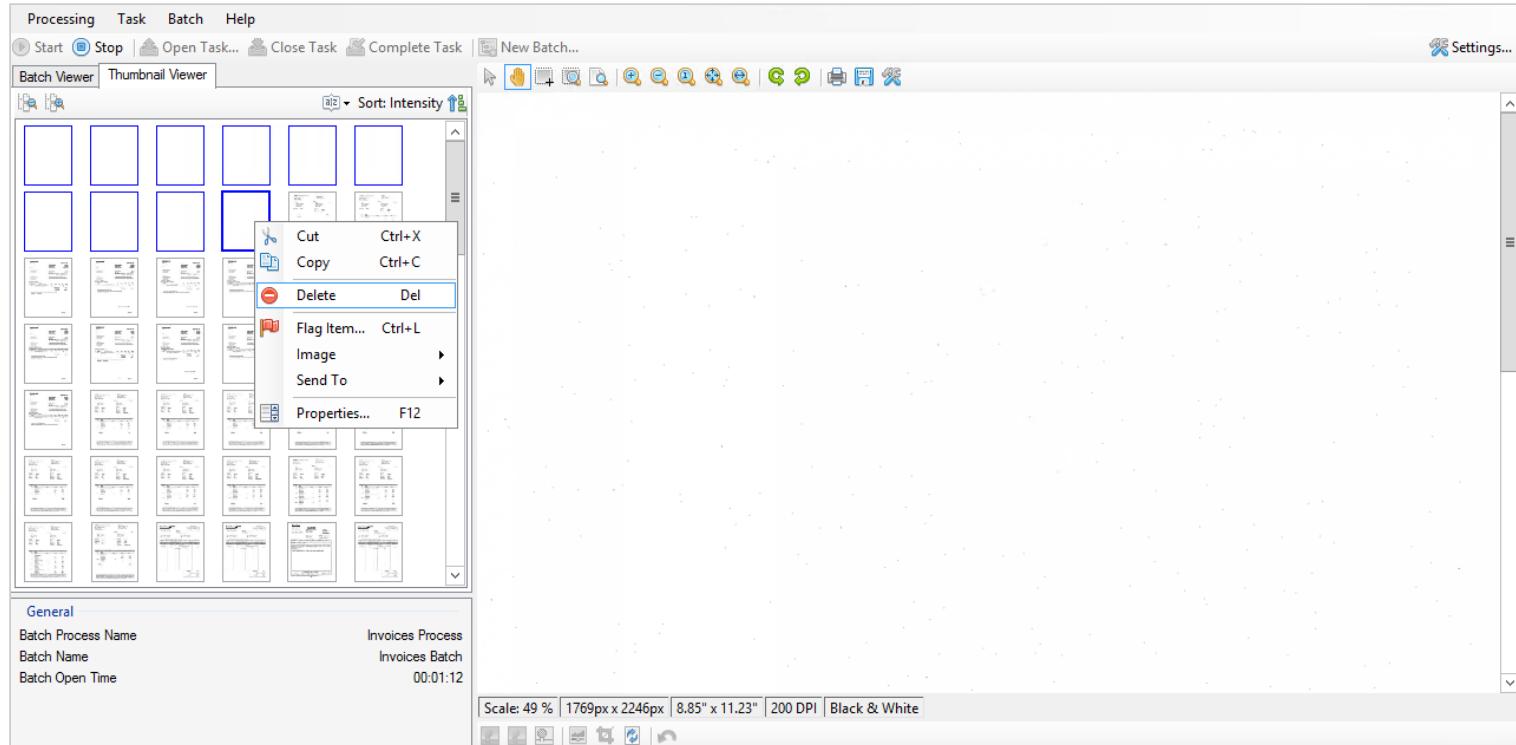
If you scroll to the top, you can see all of the pages with the least amount of black pixels - the blank pages we want to remove.



Deleting blank pages

E3 Step 8

1. Select the blank pages. You can do this in the same way you select files on your computer.
 - Using your `Ctrl` key and clicking on the page you want to select
 - Clicking the first blank page, holding `Shift` on your keyboard, and then clicking on the last blank page
2. Delete the pages. Either:
 - press the `Delete` key on your keyboard, or
 - right click on the selected pages and select `Delete`.
3. Confirm the deletion when the confirmation window comes up.



Ta-da! They're all gone.

Processing Task Batch Help

Start Stop Open Task... Close Task Complete Task New Batch...

Batch Viewer Thumbnail Viewer Sort: Intensity

ACME | INTERNATIONAL

Acme International, Inc
123 South Main
Durham, NH 03824
Phone (603) 333-4444

Checks to:
Acme International, Inc
123 South Main
Durham, NH 03824
Phone (603) 333-4444

Invoice

Bill To:
Grooper Industries
13900 N Harvey
Edmond, OK 73013
405-507-7000

Ship To:
Grooper Industries
13900 N Harvey
Edmond, OK 73013
405-507-7000

Page
2 / 02

Your Reference
Customer number : 18003405

Our Reference
Date : 12/05/2008
Invoice number : 74454835
Ship-to Number : 18103443

Item total 2,432.98

Invoice amount 2,432.98

General

Batch Process Name
Batch Name
Batch Open Time
00:02:04

Invoices Process
Invoices Batch
00:02:04

Image Review
Pages Deleted 10

Scale: 32% 2735px x 3426px 9.12" x 11.42" 300 DPI Black & White

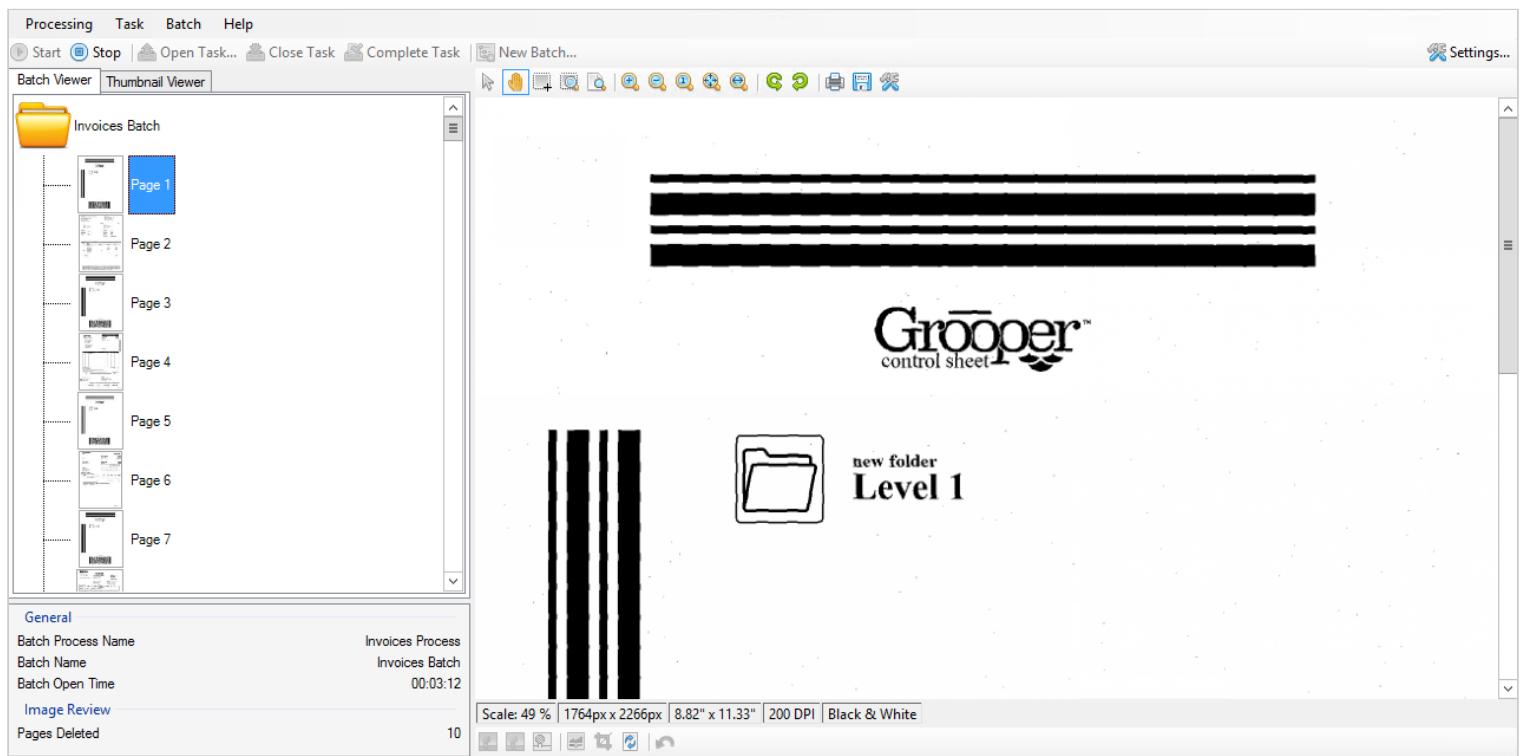
Note

This is only one of many ways that blank pages can be removed in Grooper.

Step 9

Click back over to the Batch Viewer tab.

Notice that the order of the pages didn't actually change, and now our batch is free of any blank pages.



Verifying the images

In order to complete this activity, the images have to be verified. This means that they need to be checked to make sure the previous step (Image Processing) worked correctly on all of the pages.

To verify that an image has been reviewed, you can: 1. right click on the image and click `Mark As Reviewed`, or 2. press the `Enter` key on your keyboard.

You know an image has been reviewed when there is a green on the image.

Each image has to be reviewed before the `Complete Task` button in the toolbar will light up, so it's probably easiest to press the `Enter` key for each of these.

Step 10

Make sure every image in the batch has been marked as reviewed.

Tip

For this exercise, Image Processing works on all of the pages.

To complete this step quickly, select the first page in the batch and hold your **Enter** key down. It will quickly verify all of the pages in the batch.

When you have reviewed all of the pages in the batch, the **Complete Task** button in the toolbar will light up.

The screenshot shows the Batch Viewer application interface. On the left, there is a tree view titled "Invoices Batch" with nodes for "Page 1" through "Page 7". "Page 6" is highlighted with a blue selection box. Below this is a "General" section with fields: "Batch Process Name" (Invoices Process), "Batch Name" (Invoices Batch), "Batch Open Time" (00:08:49), "Image Review" (checkbox checked), "Mark As Reviewed" (checkbox checked), and "Print Deleted" (checkbox checked). A scroll bar indicates more content below. On the right, a detailed view of an invoice is shown. The header reads "EXPRESS Dallas, TX INVOICE". The invoice details are as follows:

Invoice Number:	16862865
Invoice Date:	12/2/2008
Purchase Order:	710015038
Invoice Total:	\$216.80
Date Due:	1/1/2009
Mail Payment To:	Express Manufacturing 12333 N Dallas Tollway Dallas, TX 75032

Below the header, it says "Bill To:" followed by the recipient information: Grooper Industries, 13900 N Harvey, Edmond, OK 73013, 405-507-7000. Under "Shipped To:", it lists the same address. The vendor number is M00000032 and the customer account number is 138463500. At the bottom, there is a table with columns: Line, McMaster Part Number, Description/Customer ID #, You Ordered, We Shipped, Balance Due, Unit Price, and Extended Amount. A note at the bottom states: "You may deduct \$4.34 if paid on or before 12/12/08. Terms: 2% 10 Days, Net 30 Days, on Merchandise Only."

Step 11

Click the **Complete Task** button to finish this activity and return to Grooper Design Studio.

Processing Task Batch Help

Start Stop Open Task... Close Task Complete Task New Batch... Settings...

Batch Viewer Thumbnail Viewer Complete the current task.

Page 136
Page 137
Page 138
Page 139
Page 140
Page 141
Page 142
Page 143

Spartan
Manufacturing
12 West Laguna Dr
Irvine, CA 92612
Phone: (800) 111-22222
Fax: (308) 333-1182

INVOICE

SHIP TO (SAME AS "SOLD TO" UNLESS SHOWN)
Grooper Industries
13900 N Harvey
Edmond, OK 73013

INVOICE DATE
11/04/08 ORIGINAL
INVOICE NUMBER
IN62-757792

PO / RELEASE NUMBER
0020112357

SOLD TO
Grooper Industries
13900 N Harvey
Edmond, OK 73013

REMIT TO:
Spartan Manufacturing
12 West Laguna Dr
Irvine, Ca, 92612

ENT BY: IN629817 TAKEN BY: DF PAGE 1 OF 1

ORDER DATE 10/23/08	TERMS 1% 10&25thNET 30	SHIP DATE 10/27/08	SHIP VIA DIRECT SHIP UPS GROUND SERVICE	ACCT NUMBER 776665-02	F.O.B. FOB ORG, FRT PPA
ORDER DUE DATE 11/24/08	OCN: 146839	COMMENTS: DIRECT FROM MFG			

LINE VEN MNUO DESCRIPTION CUSTOMER INFORMATION CUST PO QUANTITIES UNIT PRICE UNIT NET AMOUNT
 1 00999 Z 65000 727-15 BLOW OFF YELLOW NOZZLE
 CPNO: 165926

Scale: 31 % | 2753px x 3434px | 9.18" x 11.45" | 300 DPI | Black & White

And that's it! We've successfully cleaned up our images. But now what do we do?

OCR

You may be wondering why it's so important for us to have the images as clean as we can get them? Why can't we just store them however they came in? They weren't that bad, right?

Well, believe it or not, there is a reason behind the madness.

About OCR

Remember that our overarching goal with Grooper is to automate the process of processing documents, collecting information from them, and sending them on their way.

We're well on our way to getting that information from the documents, but we have to do a few things to prepare Grooper to get the information for us.

What is OCR?

We're going to be performing what's called OCR, which stands for "optical character recognition." This means we're going to tell Grooper to look at the images in our batch, identify what on these images is text, and store that text for us to use later.

Image someone who doesn't know how to read. Letters on a page would simply look like a bunch of symbols without meaning, right?

Now image that that person is a computer. For a scanned document, a computer doesn't even know the symbol is a letter, but instead an arbitrary combination of pixels. The OCR process is how the computer takes an image and, **line by line**, finds combinations of pixels that it ultimately determines are letters, numbers, spaces, special characters, and so on.

For this to happen successfully, the quality of the images has to be as high as possible. If we tried to OCR low-quality images, the text identified by the OCR engine won't be very good, which essentially breaks down everything we do from this point forward.

When do I need to OCR?

To harness all the power of Grooper, pages must be OCRed. There are perhaps very simple batch processes that could be created that utilized human interaction for every step, and in that case OCR would not be required, but that's like buying a Ferrari to take trips to the grocery store.

All the power of Grooper, from separation, to classification, to extraction, leverages the computer's ability to read the document, and to that end require a page be OCRed. Therefore it is *crucial* to have the highest quality documents we can get in order for the automation process to run as smoothly as possible.

How does OCR work?

Check out [this video](#) on Wikipedia. It's a nice demonstration of a very manual approach to OCR. You can see the person live scanning the document, line by line, and the computer "reads" the letters as the scanning happens.

With Grooper, page scanning happens (usually) all at once. During the OCR process the image is broken into vertical and horizontal lines of pixels to identify individual letter characters and spacing.

TO-DO

THESE STEPS NEED EXPLANATION.

Adding an OCR step

Step 1

1. Navigate to `(root) > Batch Processing > Processes > Working > Invoices Process`.
2. Click `Add Step...`.
3. Under `Properties of OCR Step`, set the `Activity Type` to `OCR`.
4. Under `Properties of OCR Activity`, set the `OCR Profile` to `Full Text - Accurate`.
5. `Save` and `Publish` the process.

The screenshot shows the ACE Architect 01 software interface. On the left is a navigation tree with nodes like 'ACE Architect 01', 'Batch Processing', 'Processes', 'Invoices Process', and 'Reports'. The main area is titled 'Batch Process Properties' with tabs for 'Scripting', 'Contents', and 'Advanced'. It shows a list of 'Steps in Batch Process' with columns for 'Name', 'Order', and 'Activity Type'. There are four steps: 'Scan' (Order 1, Scan), 'Image Processing' (Order 2, Image Processing), 'Image Review' (Order 3, Image Review), and 'OCR' (Order 4, OCR). Below this is a 'Properties of OCR Step' panel with sections for 'General' (Activity Type set to 'OCR') and 'Expressions'. To the right is a 'Properties of OCR Activity' panel with sections for 'General' (OCR Profile set to 'Full Text - Accurate') and 'Processing Options'. A 'Remarks' section at the bottom provides a detailed explanation of what a Batch Process is.



Step 2

1. Navigate to (root) > Batch Processing > Batches > Production > Invoices Process .
2. Pause the batch.
3. Select the Batch dropdown and select Update Process... .
4. In the Update Process window, select the OCR step from the Target Step dropdown.
5. Click Execute .

The screenshot shows the ACE Architect 01 software interface. On the left, there is a navigation tree with the following structure:

- ACE Architect 01
 - Batch Processing
 - Batches
 - Production
 - Invoices Process
 - Test
 - Control Sheets
 - Processes
 - Projects
 - Content Models
 - Data Extraction
 - Global Resources
 - Infrastructure
 - Reports

The main workspace displays the "Batch Processing" module. At the top, there are tabs: Folder, Contents, Advanced, Refresh, Add, Delete, Rename, and Clone. Below these are buttons for Process, Batch, and a search bar. A message indicates "1 / 1 batches selected | Search: []".
A table lists the batches:

Batch Name	Priority	Status	Batch Process	Current Step	Created
Invoices Batch	3	Completed	Invoices Process	(none)	8/6/2018 4:59:16 PM

Below the table is a "Task Status" chart. The y-axis ranges from 0% to 100%. The x-axis shows three tasks: Scan (1 / 1), Image Processing (153 / 153), and Image Review (1 / 1). The chart uses a color legend: Pending (grey), In Progress (green), Completed (blue), and Errors (red). The "Image Processing" task is shown as a large blue bar at 100%, while the other tasks are at 0%.

Update Process



?

X

Execute Cancel

General

Target Step

Scan



Target Step

Type: [Batch Process Step](#)

The [Batch Process Step](#) where the batch should begin processing.

Remarks

The batch will be updated to use the latest published version of the parent [Batch Process](#).

Property Type: [Batch Process Step](#)

Represents a logical step in a [Batch Process](#). A Batch Process Step defines an activity to be performed as part of a

Step 3

1. the batch.
2. Click to kick off the OCR activity.

File Edit Tools Help

Refresh | Add | Delete | Rename | Clone

ACE Architect 01

- Batch Processing
 - Batches
 - Production
 - Invoices Process
 - Test
 - Control Sheets
 - Processes
 - Projects
- Content Models
- Data Extraction
- Global Resources
- Infrastructure
- Reports

Folder Contents Advanced

Process Batch

1 / 1 batches selected | Search: []

Batch Name	Priority	Status	Batch Process	Current Step	Created
Invoices	3	Paused	Invoices Process	OCR	8/6/2018 4:59:16 PM

Processing Status Statistics Tasks Batch Viewer Processing History Event Log Details

Task Status

Task	Sub-Task	Count
Scan	(1 / 1)	1
Image Processing	(153 / 153)	153
Image Review	(1 / 1)	1
OCR*	(0 / 0)	0

Pending In Progress Completed Errors

The Grooper Unattended Client should kick off again and start processing against the pages in the batch.

File Edit Tools Help

Refresh | Add | Delete | Rename | Clone

ACE Architect 01

- Batch Processing
 - Batches
 - Production
 - Invoices Process
 - Control Sheets
 - Processes
 - Projects
- Content Models
- Data Extraction
- Global Resources
- Infrastructure
- Reports

Folder Contents Advanced

Grooper Unattended Client

Start Stop

General

Thread Pool	Default
Thread Count	1
Thread Priority	Normal
Batch	Invoices Batch
Step	
Activity Type	OCR

Unattended Client Configuration

Settings class representing configurable properties for the Grooper Unattended Client application.

See Also

Thread Pool, Batch

1 / 1 batches selected | Search: []

Current Step	Created
OCR	8/6/2018 4:59:16 PM

Processing Status

Task Status

Task	Sub-Task	Count
Scan	(1 / 1)	1
Image Processing	(153 / 153)	153
Image Review	(1 / 1)	1
OCR*	(7 / 143)	7

Pending In Progress Completed Errors

TO-DO

Insert information about OCR and unattended client (images 36-39)

Recap

That's it for the `Condition` phase! We've successfully prepped our images and extracted text from them.

Here's a list of everything we learned: - how to clone a batch from production to test, - how to create an Image Processing Profile (or "IP Profile"), - how to configure our IP Profile to clean up our images, - how to update an existing batch when you change its Batch Process, - how to perform Image Review on the images, and - how to extract text from the images using Optical Character Recognition (or "OCR")

And thus your arsenal of Grooper knowledge grows!

Up next

We currently have a batch of loose pages with some patch code sheets. This batch consists of multiple invoices, but Grooper isn't aware of that yet. You and I know where one invoice ends and another begins, so it's up to us to tell Grooper how to recognize that as well.