Tips for Drawing Zones

Tips for Drawing Zones

When defining zones on your **sample page**, use the following guidelines:

- If all of the index data is separated from other information on your sample page by at least .25", draw your **index zones** to include a moderate amount of white space (about .25") all around the index data. The white space around the data allows for horizontal or vertical offsets that might be introduced by a scanner. Zones may overlap.

- If at least some of the index data is close together (that is, not separated from other information on your sample page by at least .25"), draw your index **zones** with only a small amount of white space. This will prevent unwanted data from appearing in a zone which could reduce the accuracy of character **recognition**. But, because a small amount of white space does not allow for horizontal or vertical offsets that might be introduced by a scanner, an **OCR registration zone** may be required.

For OMR Zones

- Each **OMR** shape must be zoned separately. You cannot draw one **OMR zone** around multiple OMR shapes. (Note that you can draw another type of zone around multiple OMR zones. Then, you can use scripts or SmartGroups to process the results from the zones.)

- All your OMR zones should be consistent in size. If the OMR shapes are of different sizes, create different groups of OMR zones. Make the zones for each group consistent within that group. Use a different **recognition profile** for each group of OMR zones.

- To allow for potential registration offsets, the zones should include the OMR shape and as much white space around the shape as prudent.

- Do not create zones whose boundaries touch or get near any other text or unrelated OMR shapes on the page.

- For square or rectangular shapes, you can use an aggressive **line removal** scheme to remove the shapes without destroying the inner marks. Once the shapes are removed, you can use the **despeckle** feature to cleanup extraneous noise. Then, set your pixel threshold to a low value (for example, 0 or 1). During processing, "empty" zones will return 0s (the shape was not filled in) or 1s (the shape was filled in).

- The **confidence** returned from an OMR zone is really the percentage of black pixels found in the zone. Therefore, you can use the confidence returned from a zone to determine the most appropriate threshold.

- The accuracy of the OMR engine depends, in part, on your threshold setting. The threshold is the percentage of pixels that must be colored (non-white) for the OMR engine to consider the OMR shape (usually a square, rectangle, or circle) to be marked. In addition, the actual size of your OMR zone affects the results based on the threshold setting, and therefore the accuracy of the results.

  When you draw a zone around an area for OMR processing, you should include only the OMR shape with little white space outside of the shape within the zone. This allows for a wider range of mark size to yield correct results.

  For example, if you draw a large zone box around an OMR item, the percentage of filled pixels will be relatively small. Consequently, small check marks within the OMR shape may not significantly change the percentage of filled pixels for the entire zone. If this small change does not put the percentage of filled pixels over the threshold, the OMR engine will not detect the zone as marked. Therefore, the smaller the zone box around the OMR item, the easier it is to detect a mark.

Tips for Drawing Zones

Figure 1 shows how the size of the zone box can affect the accuracy of OMR results.

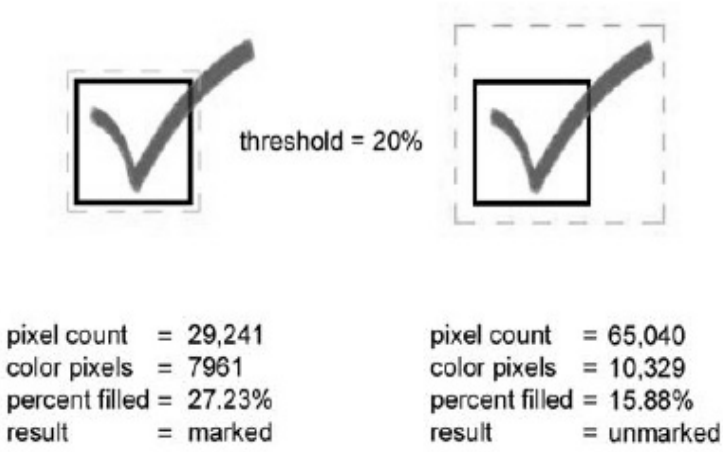


Figure 1. Effect of Zone Box Size on Accuracy

The zone around the check box on the left contains 29,241 pixels, of which 7961 (about 27 percent) are filled in. The larger zone around the check box on the right contains 65,040 pixels of which 10,329 (about 16 percent) are filled in. Because the zone on the right does not equal or exceed the threshold (20 percent), it will be regarded as unmarked by the OMR engine. This is clearly not the desired result.

For OCR and IMR Zones:

- In general, characters and numbers must reside completely within the zone boundaries to be read.
- Characters or numbers that do not reside completely within the boundaries will not be reflected in the captured data.

For Bar Code Recognition:

- Include at least a .5" boundary on the leading and trailing edges of the **bar code** when drawing your zone.
- Make sure you include the entire bar code within the zone.

For OCR Registration Zones:

- If possible, include at least .25" of white space around the **registration** data to allow for horizontal or vertical offsets that might be introduced by a scanner. The registration process will measure and correct any scan offset by shifting the index zones into place.
- Make sure the border of the zone does not touch other text or elements on the image.
- When manually resetting a registration point, select it carefully. Position the crosshair cursor at the bottom left corner of an imaginary box around the first character of the text in the OCR registration zone. If the first character has a descender (for example, a lower case "y"), use the baseline of the character, not the bottom of the character, when selecting the registration point.
- Test whether a registration point is properly defined. To do so: Create a test batch by **importing** the sample image file used to define your OCR index zones. If the registration point is properly defined, your OCR index zones will contain the expected data. If the registration point is not properly defined, your OCR index zones will be

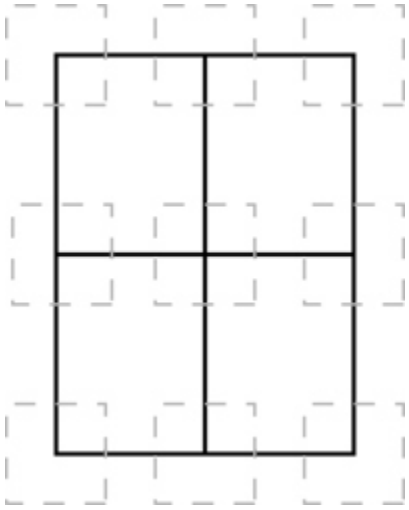
Tips for Drawing Zones

shifted and will contain unexpected data.

- For best results, keep the search text as short as possible.
- The OCR engine will read only the first four characters in your zone text string for Form ID and OCR Registration zones. For Separator zones, the OCR engine reads exactly the string defined by the zone definition.

For Shape Registration Zones:

- Include only that portion of the shape that you are trying to match. If your zone includes excess components, the recognition engine will probably return code 100 (general shape). For example, if you are searching for a cross shape, center the zone only on the cross portion of the graphic.



- For best results, keep the shape limitations in mind.