



Adding Image Cleanup to Windows Workflows

An overview for IT Managers

Welcome to Workflow World

The inclusion of the Microsoft® Windows® Workflow Foundation (WF) in the .NET 3.0 Framework offers developers an easier, highly-integrated way to build applications that incorporate complex workflows. A “workflow” is a pre-defined set of operations for automatically processing documents and other objects. Windows Workflow Foundation includes a number of built-in “activities” for various kinds of program control, such as which workflow code is to be executed, and when. It also contains activities for routing of documents and many other purposes.

Workflow Foundation is highly integrated with Microsoft’s own products, such as Word®, Excel®, and Exchange Server, via Microsoft Office SharePoint® Services (MOSS). Importantly, WF capabilities are also available to business applications being developed by independent software vendors and by internal IT departments. These facilities are already built into Windows Vista® and Windows Server 2008, and can be added to any Windows XP or Windows Server 2003 machine simply by installing the latest .NET Framework.

Other Document Sources

It has become much easier to build an application that passes documents between Microsoft products, with a high degree of control over the processing steps. But not all documents begin their lives as Word documents, Excel spreadsheets, or Exchange Server E-mails. A tremendous volume of business information is created as scanned documents and FAXes. These documents often contain information that is vital to

business processes, such as invoices, forms, medical records, and many others. Frequently, they are already damaged when they arrive.

Any physical operation, such as scanning, subjects your documents to mechanical defects caused by misalignment (skew and irregular borders) and dust (visible as specks on the page). These artifacts can make documents unattractive and can greatly increase file sizes. But, far worse, they can greatly reduce the accuracy of later processes such as OCR and barcode detection. That can wreak havoc on business processes that rely on those capabilities.

ScanFix® to the Rescue

Pegasus Imaging has long been a leading provider of document image cleanup tools such as **ScanFix Xpress**. These are designed to be easily incorporated into applications as either ActiveX COM or .NET components, giving software application developers easy access to common image cleanup technologies such as deskew, despeckle, and border cropping, as well as more advanced algorithms that perform inverse text correction, removal of dot shading, line removal, comb removal, and character smoothing.

These functions, with many more to come, are now provided for use directly within the development environment Workflow Foundation. Each ScanFix Xpress cleanup operation is available as a fully-integrated WF activity for use within Visual Studio 2005 and Visual Studio 2008. Detailed parameters for each Activity can be set to fine-tune its operation, if necessary, to address any issues that are specific to the types of images you expect to process. Most activities

Adding Image Cleanup to Windows Workflows

return confidence values that can be used to conditionally apply cleanup activities to the image. The end result is significantly improved accuracy during OCR, ICR (recognition of hand printed characters), and barcode detection.

Profit-making Activities

The following ScanFix Xpress operations are now delivered as WF activities that can be used within the Workflow Designer or your own WF design implementation. Developers can also use Visual Studio to build these new activities into workflows hosted by MOSS.

- Binarize (*reduce to black-and-white*)
- Correct Inverse Text (*white on black*)
- Correct Negative Page (*i.e. microfiche*)
- Deskew (*correct turned page*)
- Despeckle (*remove dust specks*)
- Detect Blank Page
- Detect Blank Rectangular area
- Dilate (*build up black areas*)
- Erode (*shrink black areas*)
- Flip (*invert image*)
- Get Rectangle (*crop*)
- Mirror (*left-to-right*)
- Negate area (*switch black and white*)
- Register Image (*fix position on page*)
- Remove Blobs (*hole punches*)
- Remove Border (*black edges*)
- Remove Combs (*data entry underlines*)
- Remove Dot Shading
- Remove Lines (*and repair the text*)
- Rescale (*change image size*)
- Rotate (*any number of degrees*)
- Shear (*parallelogram shift*)
- Smooth Objects (*reduce jags*)
- Smooth and Zoom (*2X and fill in*)

Two ImagXpress® activities have also been added to the list of workflow activities, one to **open** document images (including single- and multi-page TIFFs) prior to processing, and one to **save** them back out after repair. This also enables conversion of images from one format to another. For more details on the function of each of these activities, review the ScanFix Xpress product at www.PegasusImaging.com or, better yet, download the entire toolkit with documentation and try them out. We also have a detailed tutorial to get you started.

Real-world Examples

To illustrate how one might apply this technology, suppose you're building an application that receives FAXed documents and then routes them for distribution or further processing. Users could place the TIFF images into a SharePoint folder where a workflow would automatically apply despeckle and deskew cleanup activities, convert them into PDF files, and then route them to a new folder.

Or, if you needed to make a massive microfiche collection accessible from anywhere, you might build a WF application that takes groups of images, negates them when needed, cleans them up, removes unsightly borders, positions the new image on a page, and saves them out as multi-page TIFF files that could be viewed in a browser with Pegasus Imaging's Prizm® Viewer.

Paul B. Firth, Product Manager
Pegasus Imaging Corporation

Pegasus™, ImagXpress™, ScanFix® and Prizm® are trademarks of Pegasus Imaging Corporation in the United States and/or other countries. Microsoft®, Windows®, Office®, BizTalk®, Word®, Excel®, Vista® and SharePoint® are registered trademarks of Microsoft Corporation.