# Transcription Pearl

## *User Manual*

**1.0.0 Beta**

***Written by Mark Humphries with Assistance from Claude Sonnet-3.5***

## Table of Contents

## 1.0 Requirements

This program is for Windows only and should ideally run on Windows 10 or later. It requires an internet connection and at least one API key from OpenAI, Anthropic, or Google (see section 3.0 below).

## 2.0 Installation

You can download the executable file for Transcription Pearl from the GitHub repository or directly from this link. Once downloaded, create a directory for the program and move the TranscriptionPearl.exe file to that directory.

As this program is open-source software, it may be read as a virus or potentially malicious program by Windows Defender or other antivirus software. If this happens, you can choose to add an exception and run the program anyway.

## 3.0 API Key Setup

### *3.1 Introducing API Keys*

API Keys are like usernames and passwords combined, allowing an external server to identify you and your account when you make requests. When you enter you API key into Transcription Pearl, it is stored locally on your machine.

### *3.2 Costs and Tokenization*

Your API keys will be used to send requests from Transcription Pearl to the servers providing access to OpenAI, Anthropic, and/or Google Large Language Models (LLMs). You will be billed by those companies for the use of their servers on a per-token basis, both for input tokens and output tokens.

A token is about ¾ of a word and rates vary from one company to another from $1.25 to $3.00 per million input tokens. Image pricing also varies: Gemini is about $0.0003 per image, OpenAI about $0.002 per image, and Claude about $0.005 per image. Users should check current pricing with OpenAI, Anthropic, and Google Gemini before using the Transcription Pearl software.

For each page that you choose to transcribe, Transcription Pearl sends a high-resolution image of the page and about 160 tokens of text (that is the size of the default prompts, see section 5.0 below) to the relevant API. For each page you wish to correct, the program sends a high-resolution image of the page and about 150 tokens of text (for the default prompts) plus the number of tokens in the initial transcription. For reference, most handwritten pages will not have more than about 500 tokens of text at maximum. If you want to determine the approximate size of a request, simply copy and paste the initial transcription from Transcription Pearl into the OpenAI Tokenizer, a program that counts tokens, add 150 tokens to the total, and the cost of the image. Note that Gemini and Claude count tokens slightly differently, but the OpenAI Tokenizer will give you an approximate length.

### 3.3 Obtaining API Keys

Obtaining an API key is about as easy as signing up for any web-based service. There are lots of web-based tutorials on how to do this. Below is some basic information. While the process will vary somewhat, after you create a key you will be prompted to copy and paste it to store it for future use. You can then input it directly into Transcription Pearl (see Section 3.4 below) but you should also save it to a local document for future reference as you won't be able to access it again. If you lose it, you will need to create a new key. To use the API Key, you will need to enter billing information and/or buy credits. You can also set usage limits to limit costs.

#### *3.3.1 OpenAI API Keys*

To register for an OpenAI API key, first create an account on the OpenAI platform (note that this will be different than a ChatGPT account). Click your profile icon in the top right-hand corner and select "Your Profile". In the menu on the left, click API-keys and then click the green "+Create New Secret Key" at the top right. Give your Key a name and create the key.

#### *3.3.2 Anthropic API Keys*

To register for an Anthropic API key, first create an account on the Anthropic consol (this will be different than your Claude chat account). After you create an account, cock your profile icon at the top right and select "API Keys" from the dropdown and then select the orange "+Create Key" button at the top right.

#### *3.3.3 Gemini API Keys*

The process with Google can be somewhat more confusing as it provides a number of different ways to access their models via API, specifically via the Google AI Studio and Vertex platforms. You need to use the Google AI Studio platform. First, create a Google AI Studio Account then, once you are logged in, select the blue "Get API Key" button on the left. In the next pane, click the blue "Create API Key" button.

## 4.0 Interface

The Transcription Pearl interface consists of a main window divided into two primary sections: a text editor panel on the left and an image viewer panel on the right. At the top of the window, there is a navigation bar displaying the current text type (None, Original Text, or Initial Draft) and page navigation controls. The navigation controls include buttons for first page (<<), previous page (<), next page (>), and last page (>>), along with a page counter showing the current page number and total number of pages.

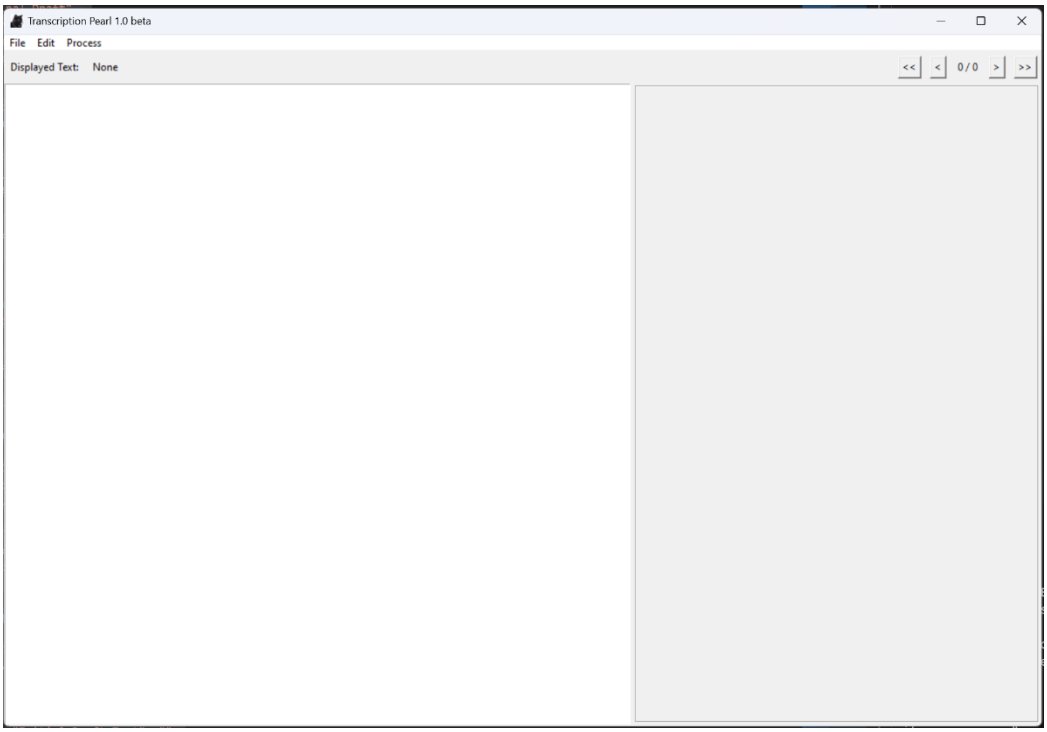

*Figure 1.0: Basic Transcription Pearl Interface*

The text editor panel (left) features a standard text editing interface with a white background and black text. The text is displayed in a clear, readable font size and supports standard text editing functions including find and replace; copy, cut and paste; and undo and redo. The image viewer panel (right) displays the current document page and supports zooming (ctrl+mouse wheel), panning (click and hold the left mouse button and move the mouse), and scrolling (mouse wheel) functions to examine details of the handwritten text. The relative size of the two panes can be adjusted as needed by hovering over the vertical boundary and dragging the image pane left or right.

The program uses a menu-based interface to access functions (with keyboard shortcuts) with three main menus: File, Edit, and Process. These menus contain all the core functions needed to manage projects, edit text and images, and process documents using AI transcription and correction tools. The File menu handles project management and file operations, the Edit menu contains text and image editing functions, and the Process menu provides access to the AI-powered transcription and correction tools.

The interface is designed to be intuitive for historians working with historical documents, with clear labeling and straightforward navigation between pages. The split-screen design allows users to easily compare the transcribed text with the original document image, making it simple to verify accuracy and make corrections as needed. All major functions are accessible through both the menu system and keyboard shortcuts, providing flexibility in how users interact with the program.

## 5.0 Setup and Settings

Transcription Pearl's settings interface provides comprehensive control over the program's AI transcription and correction functions through an easy-to-navigate menu system. To access the

settings, click on "Settings" in the File menu, which opens a new window with options organized into categories on the left side: APIs and Login Settings, HTR Settings, and Correct Text Settings.

The APIs and Login Settings section is where users enter their API keys for OpenAI, Anthropic, and Google's AI services. These keys are required to use the program's AI functions and are securely stored between sessions. Be sure there are no leading or trailing spaces at the start or end of your API keys and that the keys have been entered exactly as provided.

The program saves these settings locally on your computer (so long as you save the settings), so you only need to enter them once unless you want to change to a different API key. in the following folder (by default): *C:/Windows/Users/[your_name]/AppData/Roaming/TranscriptionPearl/settings.json.* Do not share this file with anyone!

The HTR (Handwritten Text Recognition) Settings section allows users to configure how the program processes handwritten documents. Users can select which AI model to use for transcription from a dropdown menu of available options. The General Instructions field contains the system-level prompts that guide the AI in how to approach transcription tasks (technically called the System Message), while the Detailed Instructions field allows for more specific guidance (technically called the User Message). The Validation Text field is of critical importance because this is used by Transcription Pearl to extract text (see 7.2 below).

The Correct Text Settings section follows a similar structure to the HTR Settings but is specifically focused on configuring how the program corrects existing transcriptions. Users can select a different AI model for correction tasks, and modify both general and detailed instructions to achieve optimal results. These settings determine how the program approaches correcting transcription errors while preserving historical accuracy.

At the bottom of the settings window, users find options to Load Settings (restore previously saved configurations), Save Settings (store current configurations), and Restore Defaults (reset all settings to their original values). The Done button closes the settings window and returns to the main interface. Any changes made to settings are automatically saved when closing the window, ensuring your preferences are preserved for future sessions.

## 6.0 Importing Files

Transcription Pearl provides several methods for importing historical documents into the program. The primary import options are accessed through the File menu and include “Import Images Only”, “Import Text and Images”, and “Import PDF.” These options are designed to accommodate different source materials and project requirements.

The program also supports drag-and-drop functionality, allowing users to directly drag image files (JPG, JPEG, PNG) or PDF files into the main window. When importing through any method, the program automatically processes the images to optimize them for AI analysis while preserving image quality. This includes resizing very large images to ensure consistent performance, automatically rotating images taken with a digital camera to the correct original orientation, while maintaining the readability of the handwritten text.

When importing files, Transcription Pearl creates a structured working environment by organizing the imported materials into a standardized format. Each imported page is assigned a sequential identifier and paired with a corresponding text file to store transcriptions. The program maintains this organizational structure throughout the project, ensuring that image files and their associated transcriptions remain properly linked.

The program's import functions are designed to handle both individual documents and large collections of historical materials. When importing multiple pages, Transcription Pearl automatically sorts them in numerical order based on file names. PDFs that contain HRT or OCR text will be imported so that the text appears on the left and the image on the right.

### *6.1 Importing Transcriptions from Transkribus or Other Sources*

Users can import Transkribus transcriptions—or transcriptions from any source—for automatic correction. The easiest way to do this is to export the file from the original transcription program as a PDF with recognized text. Sometimes, though, image resolution is lost when the PDF is created. If you don't obtain good results, you can also import the transcriptions manually. To do so, you need to have an image file and text file with the same names (but with jpg and txt extensions) in the same folder. Transkribus can do this automatically: choose export, then select both images and text files and "one file per page" then click the blue "Start Export" button. When your file is finished downloading, you will have a zipped folder that contains a folder with images labeled something like 0001_0001_filename.jpg. Inside that folder is another folder called "txt" containing matching files with "txt" extensions. To import these files into Transcription Pearl, create a new folder on your computer, and copy and paste all the jpgs and the matching txt files into the new directory. Then in Transcription Pearl, go to File and Import Text and Images, then navigate to the new directory you created.

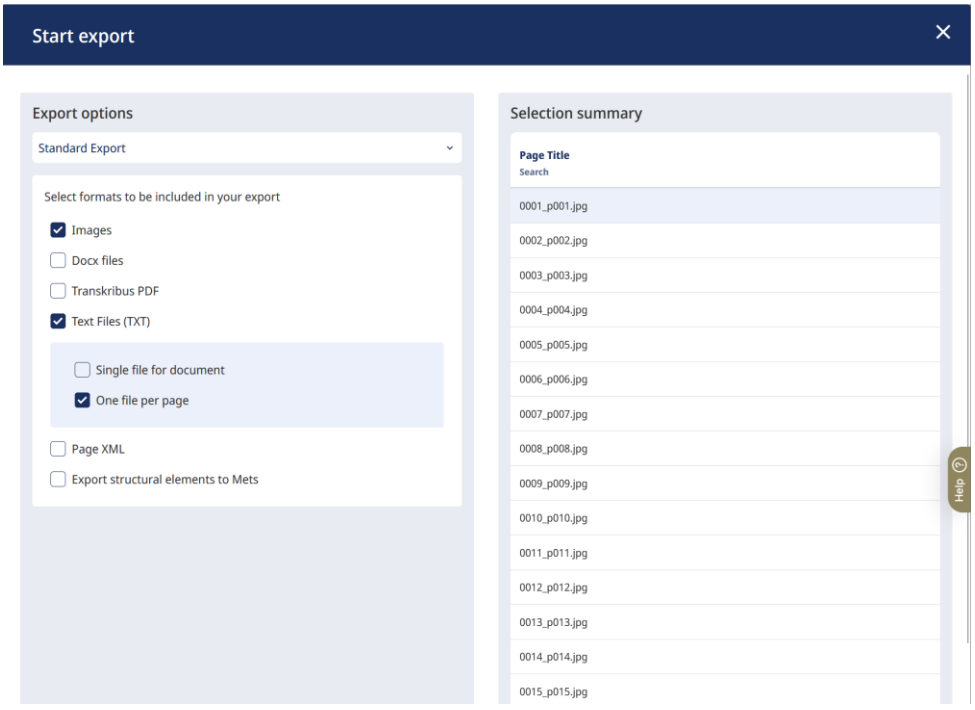


*Figure 2.0: Transkribus export functions*

## 7.0 HTR and Correct Text Functions

Transcription Pearl offers two primary AI-powered functions for working with historical documents: Handwritten Text Recognition (HTR) and Text Correction. These functions can be accessed through the Process menu or via keyboard shortcuts, and each can be applied to either the current page or all pages in a project.

The HTR function is designed to create initial transcriptions of handwritten documents. When activated, it sends the current page image (or all page images) to the selected AI model along with the instructions specified in the HTR Settings. The program processes multiple pages simultaneously to maximize efficiency, with a progress bar displaying the current status. For each page, the AI model analyzes the handwritten text and returns a transcription that appears in the text editor panel. These transcriptions are automatically stored as "Original Text" in the project.

The Correct Text function builds upon the initial transcriptions by sending both the image and the existing transcription to the AI model for refinement. This function is particularly useful for correcting errors in existing transcriptions, whether they were created using the HTR function or imported from other sources. By default, the AI model compares the original document image with the transcription, making corrections while preserving historical accuracy in spelling, punctuation, and formatting. The corrected text is saved as "Initial Draft" in the project.

Both functions utilize a batch processing system that can handle up to 50 pages simultaneously, though users should be mindful of their API rate limits and costs. The program automatically manages the API requests to prevent timeouts or overload, and if any errors occur during processing, they are logged and reported to the user. A progress window shows both the number of pages processed and the percentage complete, allowing users to monitor large jobs which may take several minutes to complete. Note that sometimes API errors can occur on the server side either delaying completion or causing an operation to fail. Users can access the current status of the OpenAI and Anthropic APIs through these links to their websites. For Google, users should consult Down Detector or a similar website.

The effectiveness of both functions depends significantly on the quality of the input images and the settings configured in the HTR Settings and Correct Text Settings panels. Even small changes to a prompt (or the order of instructions) can have a significant effect on the quality of the output. Users can fine-tune these settings based on their specific needs and the characteristics of their historical documents (see 7.1 below). The program also maintains separate text versions (Original Text and Initial Draft) for each page, allowing users to track changes and revert to previous versions if needed.

For optimal results, users should ensure their images are clear and properly oriented before processing. The program includes an image preprocessing tool to help prepare documents for AI analysis, including rotation, splitting, and cropping functions, which is accessible through the Edit menu (see 11.0 below). Additionally, users can adjust the AI model selection and instruction parameters in settings to better handle specific types of historical documents or handwriting styles.

### 7.1 Prompting

All transcription and correction tasks are accomplished via prompting, although this process remains largely hidden from the user. In Transcription Pearl, prompting involves two distinct components: General Instructions and Detailed Instructions. The General Instructions establish the general framework and rules for how the AI should approach the task, while the Detailed Instructions provide specific guidance and largely serve to pass the contents to the LLM (ie. the image and/or text to correct).

For HTR tasks, the General Instructions should outline the fundamental approach to transcription, emphasizing (as relevant) accuracy, preservation of historical spelling and punctuation, capitalization, and the handling of special elements like marginalia or insertions. The Detailed Instructions then pass the specific contents. The default HTR prompt is:

> General Instructions: Your task is to accurately transcribe handwritten historical documents, minimizing the CER and WER. Work character by character, word by word, line by line, transcribing the text exactly as it appears on the page. To maintain the authenticity of the historical text, retain spelling errors, grammar, syntax, and punctuation as well as line breaks. Transcribe all the text on the page including headers, footers, marginalia, insertions, page numbers, etc. If these are present, insert them where indicated by the author (as applicable). In your response, write: "Transcription:" followed only by your accurate transcription
>
> Detailed Instructions: Carefully transcribe this page from an 18th/19th century document. In your response, write: "Transcription:" followed only by your accurate transcription.
>
> (the page image is included automatically with the detailed instructions).

For Correction tasks, the default General Instructions focus on comparing and correcting existing transcriptions while maintaining historical accuracy. The Detailed Instructions must include the {text_to_process} placeholder, which tells the program where to insert the existing transcription that needs correction. The default Text Correction prompt is:

> General Instructions: Your task is to compare handwritten pages of text with corresponding draft transcriptions, correcting the transcription to produce an accurate, publishable transcript. Be sure that the spelling, syntax, punctuation, and line breaks in the transcription match those on the handwritten page to preserve the historical integrity of the document. Numbers also easily misread, so pay close attention to digits. You must also ensure that the transcription begins and ends in the same place as the handwritten document. Include any catchwords at the bottom of the page. In your response write "Corrected Transcript:" followed by your corrected transcription.
>
> Detailed Instructions: Your task is to use the handwritten page image to correct the following transcription, retaining the spelling, syntax, punctuation, line breaks, catchwords, etc of the original.
>
> {text_to_process}

Both types of prompts must include instructions for the AI to begin its response with the appropriate Validation Text (e.g., "Transcription:" or "Corrected Transcript:", see section 7.2 below). This text must match exactly what you enter in the Validation Text field in Settings (see section 5.0).

Remember that even small changes to prompt wording or order can significantly impact results. Users are encouraged to experiment with different prompting approaches while maintaining the required Validation Text and {text_to_process} placeholder (for correction tasks) to find what works best for their specific documents.

### 7.2 Validation Text

Depending on what they are asked to do, LLMs can produce outputs in a variety of formats that might sometimes include preambles like "Sure, I would be happy to help!" as well as "chain of thought" outputs in which the model is explicitly told to think through a given task step by step before giving a final answer. To enable Transcription Pearl to identify the text to use for the transcription or corrected text, users need to tell the LLM to explicitly identify this portion of the response. While this can also be accomplished via JSON or other structured API outputs, this approach allows the most flexibility for non-expert users.

By default, the LLM will be instructed to begin its transcriptions with "Transcription:" and corrected text with "Corrected Transcription:". These "Validation Texts" are then used by Transcription Pearl to verify that the LLM completed the assigned task and then to find and extract the relevant text. You need to ensure that your General Instructions end with a statement like: "In your response, write: "Transcription:" ". You would then put "Transcription:" in the Validation Text box (for more on prompting see 7.1).

## 8.0 Saving and Opening Projects

Transcription Pearl allows users to save their work and return to it later. Projects in Transcription Pearl contain all the images, transcriptions, and associated metadata organized in a structured format. When saving a project for the first time, users should select "Save Project As" from the File menu. This will prompt them to choose a location and name for their project. The program creates a new directory with this name containing all project files, including a project file (with .pbf extension) that stores the relationships between images and transcriptions.

For ongoing work, users can simply click "Save Project" or use Ctrl+S to update the existing project files with any changes. This saves all current transcriptions, image modifications, and project settings. The program also maintains separate versions of transcriptions (Original Text, Initial Draft, and Final Draft) within the project, allowing users to track the evolution of their work.

To open an existing project, users select "Open Project" from the File menu or use Ctrl+O. This allows them to navigate to and select a previously saved project directory. Upon opening, Transcription Pearl loads all associated images and transcriptions, restoring the project exactly as it was last saved. The program also verifies the integrity of project files during loading, ensuring all components are present and properly linked.

Users can create new projects at any time by selecting "New Project" from the File menu or using Ctrl+N. This clears the current workspace and prepares the program for a fresh start. The program will prompt users to save any unsaved changes before creating a new project, helping prevent accidental data loss.

## 9.0 Exporting Files

Transcription Pearl offers straightforward options for exporting transcribed documents through the File menu's Export function. When users select "Export" or use Ctrl+E, they're prompted to choose a location and name for their export file. The program combines all transcriptions into a single text file, maintaining the sequential order of the documents.

The export function automatically formats the text to create a clean, readable document. The program uses the currently displayed version of each transcription (Original Text or Initial Draft) when creating the export file, allowing users to control which version of their work is included in the final output.

For quality control, the program automatically removes any excessive line breaks or formatting artifacts that might have been introduced during the transcription process. This ensures the exported document maintains a professional appearance while preserving the historical accuracy of the transcriptions. The resulting text file can be easily imported into other programs for further editing or publication preparation.

## 10. Transcription Pearl Keyboard Shortcuts

*Project Management*

- Ctrl+N: Create new project
- Ctrl+O: Open existing project
- Ctrl+S: Save current project
- Ctrl+E: Export project

*Navigation*

- Ctrl+Home: Go to first page
- Ctrl+Left: Go to previous page
- Ctrl+Right: Go to next page
- Ctrl+End: Go to last page

*Image Controls*

- Ctrl+MouseWheel: Zoom in/out
- MouseWheel: Scroll image up/down
- Left Mouse Button (hold): Pan image
- Ctrl+]: Rotate image clockwise
- Ctrl+[: Rotate image counter-clockwise

*Image Editing*

- Ctrl+I: Edit current image
- Ctrl+Shift+I: Edit all images
- Ctrl+D: Delete current image

*Text Editing*

- Ctrl+Z: Undo
- Ctrl+Y: Redo
- Ctrl+X: Cut
- Ctrl+C: Copy
- Ctrl+V: Paste
- Ctrl+F or Ctrl+H: Find and replace
- Ctrl+R: Revert current page
- Ctrl+Shift+R: Revert all pages

*AI Processing*

- Ctrl+1: HTR on current page
- Ctrl+Shift+1: HTR on all pages
- Ctrl+2: Correct text on current page
- Ctrl+Shift+2: Correct text on all pages

## 11.0 Image Preprocessing Tool

The Image Preprocessing Tool in Transcription Pearl is a specialized interface designed to help historians prepare document images for optimal AI transcription. While LLMs require less preprocessing than other HRT solutions, we've found that it is often beneficial to split pages and ledger columns into discreet images. Sometimes straightening pages and cropping are also useful if additional text is visible or the page is relatively small inside the overall image. This tool provides essential functions for handling common issues with archival document scans and photographs, including splitting two-page spreads, cropping excess margins, straightening text, and correcting page orientation.

The tool can be accessed through the Transcription Pearl Edit menu by selecting "Edit Current Image" or "Edit All Images" or by pressing Ctrl+I. When launched, it opens in a new window that displays the current image in a large viewing area. The tool provides both manual and automatic options for image manipulation. Each image (or all images) can be reverted to their original state and changes will only be applied in the main Transcription Pearl window if the user decides to save the images or saves when they quit the program. To discard all changes, simply close the editing window and choose not to save images.

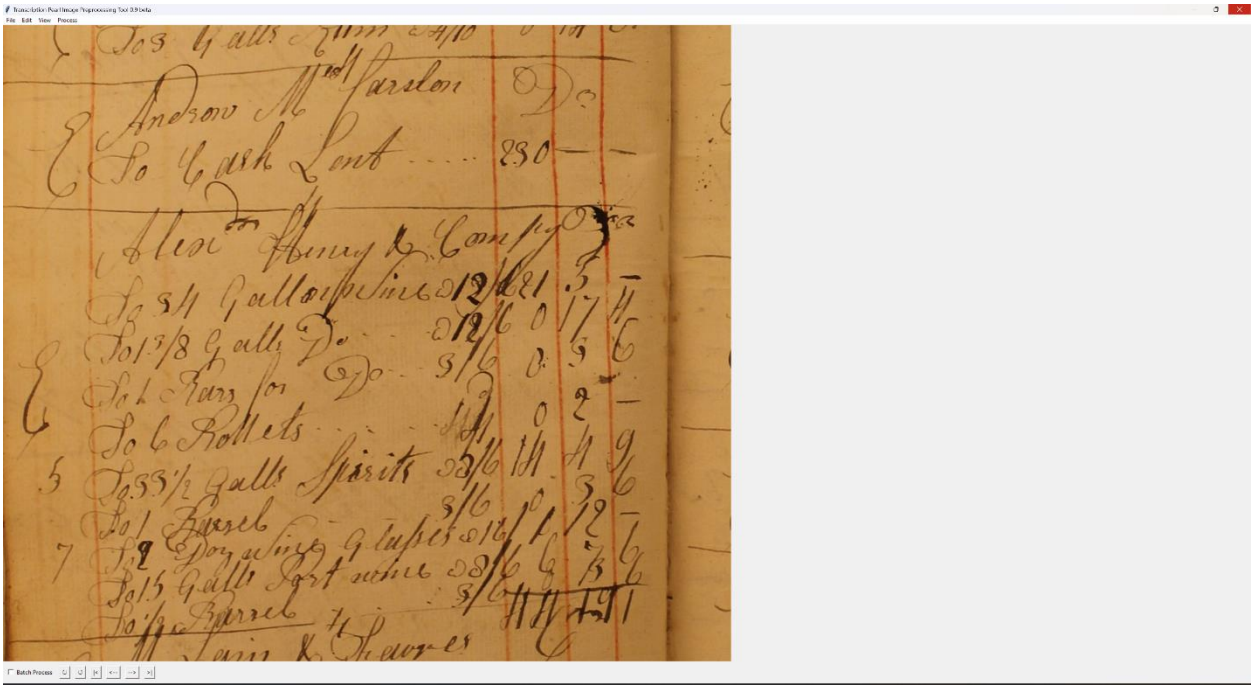

*Figure 3.0: Image Preprocessing Tool Interface*

### *11.1 Interface*

The Image Preprocessing Tool interface features a large central canvas displaying the current document image, with a navigation bar at the bottom containing rotation buttons (↻ and ↺), page navigation controls (|<, <--, -->, >|), and a "Batch Process" checkbox. The navigation controls allow users to move through multiple images in a project, while the Batch Process option enables automatic progression to the next image after completing an operation to speed up the image editing process.

At the top of the window, a menu bar provides access to all editing functions through File, Edit, View, and Process menus. The File menu handles saving and reverting changes, while Edit contains options for undoing operations. The View menu provides navigation controls, and the Process menu contains all image manipulation tools including splitting, cropping, straightening, and auto-cropping functions.

The interface is designed to be responsive to both mouse and keyboard input, with the canvas supporting zoom functionality using Ctrl+mouse wheel and panning via click-and-drag. When tools are activated, the cursor changes to indicate the current mode (e.g., crosshair for cropping, vertical/horizontal line for splitting), and visual guides appear on the canvas to assist with precise adjustments. A red line indicator helps users visualize where splits or crops will occur, updating in real-time as the mouse moves across the image. A blue line is used to align the page edge for straightening.

### *11.2 Splitting Images*

The Image Preprocessing Tool allows for splitting document images, particularly useful for ledger pages or two-page spreads. Users can initiate splitting by selecting "Split Image" from the Process menu or pressing Ctrl+V for vertical splits and/or Ctrl+H for horizontal splits. This activates a red guide line on the canvas that follows mouse movement. The splitting line can also be rotated clockwise and counterclockwise using the [ and ] keys, allowing users to split pages along non-square lines. When “Batch Processing” is enabled (see 11.6 below), users will automatically advance to the next un-split image when they click the mouse, enabling them to quickly split a number of images with as few clicks as possible.

When splitting is active, users can click anywhere on the image to execute the split at that position. For vertical splits, the program creates separate left and right images, while horizontal splits create top and bottom images. The left image and/or top image will be placed first in sequence. The tool maintains the original image resolution.

The splitting function also includes an auto-split feature, accessible through "Split All Images" in the Process menu, which can apply the same split position to multiple images in batch. This is particularly useful for consistent document formats like scanned books or microfilm scans. The split operation can be cancelled at any time by pressing Escape.

### *11.3 Cropping Images*

The cropping tool allows users to remove excess margins or focus on specific sections of a document. It can be activated through "Crop Image" in the Process menu or by pressing Ctrl+Shift+C. Once active, users click and drag to create a rectangular selection area, which appears as a red outline on the image.  The crop can be adjusted by clicking and dragging any edge or corner of the selection rectangle. When satisfied with the selection, users can apply the crop by either releasing the mouse button (if Batch Process is checked) or pressing Enter. The cropping tool maintains aspect ratio and image quality while removing unwanted portions of the image.

For batch operations, users can activate the Batch Process checkbox (see section 11.6 below) to automatically move to the next image after each crop operation. This streamlines the process of cropping multiple images. The crop operation can be cancelled at any time by pressing Escape.

### *11.4 Straightening Images*

The straightening tool helps correct skewed document images. The manual straightening function is accessed through "Straighten Image by Line" in the Process menu or by pressing Ctrl+L. This tool allows users to draw a blue line along a top or side page edge and the program automatically rotates the image to align this line with the horizontal or vertical plane.

Users can also perform precise rotations through "Rotate Image by Angle" in the Process menu, which allows input of specific degree values. For quick adjustments, the rotate buttons (↻ and ↺) in the navigation bar or Ctrl+[ and Ctrl+] provide 90-degree rotations. All rotation operations maintain image quality through high-quality resampling.

### *11.5 Auto-Cropping*

The auto-cropping feature automatically detects document edges and removes excess margins. It can be accessed through "Auto Crop Active Image" or "Auto Crop All Images" in the Process menu. This function uses edge detection algorithms to identify the main content area of documents and so works best when the main page is clearly defined against a contrasting background (as in the case of microfilm). It automatically adjusts for both color and grayscale images, maintaining consistent results across different document types. For batch processing, the auto-crop function can be applied to all images in a project, significantly speeding up the preprocessing workflow.

### *11.6 Batch Processing*

When checked, the batch processing checkbox enables users to quick apply a manual operation to sequential operations with one a single click per image. With splitting, this automatically advances the image to the next unsplit image, each time an image is split (with a mouse click) while maintaining the cursor orientation. With cropping, it advances automatically to the next image when the user releases the mouse button after drawing a cropping box.

### *11.7 Image Preprocessing Tool Keyboard Shortcuts*

*Splitting*

- Ctrl+V: Activate vertical split tool
- Ctrl+H: Activate horizontal split tool
- [ and ]: Rotate cursor line (when splitting)
- Escape: Cancel current operation

*Cropping*

- Ctrl+Shift+C: Activate crop tool
- Enter: Apply current operation
- Escape: Cancel current operation

*Rotation and Straightening*

- Ctrl+L: Activate straighten tool
- Ctrl+[ and Ctrl+]: Rotate image 90 degrees
- Escape: Cancel current operation

*Navigation*

- Arrow keys: Navigate between images
- Ctrl+Mouse wheel: Zoom in/out
- Mouse drag: Pan image when zoomed