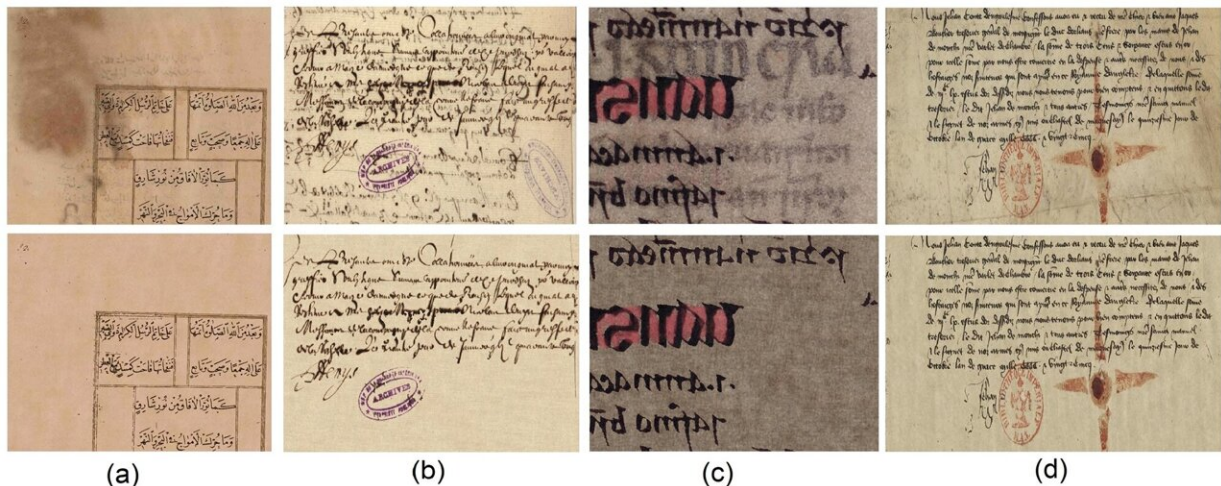# Digital restoration of historical documents

March 24 2023, by Justin Jackson

Visual examples of color document images: Input degraded image (top row), restored image using the proposed method (bottom row). Credit: Hanif et al., 2023, PLOS ONE, CC-BY 4.0 (creativecommons.org/licenses/by/4.0/)

One of the best ways to learn about any historical period is by conversing with the people who lived through it. Speaking with people from the distant past is very one-sided, as they are typically dead and have stopped listening long ago. However, they speak volumes if you have the patience to listen, or rather, read what they say in letters, diaries and primitive post-it notes with no sticky back sides.

An international group of computer scientists from Italy, the U.K. and Pakistan have teamed up to resurrect the dead from writings that have

been degraded by time by developing a computer-assisted method to virtually return documents to a more legible and decipherable condition. In their research paper, "Restoration and content analysis of ancient manuscripts via color space based segmentation," published in the journal *PLOS ONE*, the team details their digital restoration technique's method and experimental results.

We get a sense of ancient civilizations from their writings, both trivial and profound. The Sumerian cuneiform writing on [clay tablets](#) reveals 4,000-year-old merchant transactions, geometric calculations, and poetry detailing the fall of a great city. Had they been written on paper and not in clay we would likely not have them today.

We do have historians writing about ancient events, as they themselves were living in what is now ancient Greece. We have letters from soldiers in every major conflict over the past four hundred years, a written history of most major or even minor events since the industrial revolution, and all of it on paper that is susceptible to aging and degradation.

There is a minor void in the record around the late 1980s as people used floppy disks to store important documents only to throw them all away a few years later—thankfully, nothing important happened in those years.

Reading old or ancient documents, even when in good condition, can require a reader to delve into the role of a cryptologist deciphering encoded messages. Not so much because the writer sought to be intentionally secretive with their texts but because combinations of a writer's literacy, handwriting legibility, lettering styles, obsolete spelling and grammar conventions or ad-hoc abbreviations to save space can make even a native language seem unfamiliar.

Take that same document and reduce its legibility even more by fading

the ink, widening the lettering from moisture exposure, bleeding through or transferring pigment from one page to another and puzzling the information back together could become a truly daunting task.

One aspect of the current effort that differs greatly from past methods is the use of color. Typical digital reconstruction converts a document into black text on a white background in an attempt to reveal the text through noise filtering and contrast enhancement. In the current color method, the image is much more of a restoration effort, keeping much of the look and virtual feel of the original document.

Researchers applied a complex integration of several digital color interpretation techniques and Gaussian mixture models to identify and separate features of text, paper, and artifacts. This allows for the selective extraction of different layers of information based on sometimes subtle spectral differences. So a user could decide if a scribbled notation, page decoration, stamp, coffee mug ring stain or other feature was critical to the understanding by adding or removing layers.

In some cases, the technique could preserve features that would otherwise be lost, such as the faint watermarks of a paper manufacturer, which could offer vital insights to historians. For example, Leonardo da Vinci's famous works, which have survived in remarkable condition and were not part of the current study, are written on pages with watermarks from various paper suppliers that allowed document historians to group them into likely clusters of writing periods.

According to the paper, the processing can be accomplished on a standard desktop computer (core i5, 4 GB RAM, Windows 10). The algorithm takes just a few minutes to virtually restore a document to greater legibility. This level of user-friendliness would make it a great tool for both historical document researchers and people attempting to

decipher old family recipes alike.

More information: Muhammad Hanif et al, Restoration and content analysis of ancient manuscripts via color space based segmentation, *PLOS ONE* (2023). DOI: 10.1371/journal.pone.0282142