

*Associação Nacional de História – ANPUH*  
**XXIV SIMPÓSIO NACIONAL DE HISTÓRIA - 2007**

**A Contribuição da Matemática no Realce de Imagens Digitais das Fontes Paroquiais**

Dario Scott \*

**Resumo:** Os problemas enfrentados pelos historiadores que se utilizam de fontes documentais manuscritas, como os registros paroquiais, colocam questões que nos estimularam a encontrar soluções viáveis para o tratamento de fontes documentais que podem ter sofrido a ação do tempo e/ou as más condições de conservação e armazenamento. Além das vantagens óbvias da melhoria da definição e qualidade das imagens das Fontes Paroquiais, deve ser ressaltada ainda, a possibilidade da preservação de um patrimônio histórico documental, que pode estar correndo risco de se perder.

Uma das formas mais eficientes de conseguir alcançar o realce dessas imagens digitais é através da morfologia matemática, ramo do processamento e análise não linear de imagens, que permite processar imagens com o objetivo de realçar, segmentar e detectar bordas, entre várias outras aplicações. O objetivo desta comunicação é mostrar as possibilidades de melhoria na qualidade das imagens digitais manuscritas, o que facilitaria o trabalho do pesquisador na leitura e interpretação dos documentos originais.

**Palavras chave:** Imagem digital, realce, fontes históricas.

**Abstract:** This paper presents the advantages of the use of digital image enhancement, applied in historical sources, like parish registers. Besides the advantages of the use of this technique, it is important, also, to underline the possibilities of preservation from manuscript original sources, used by historians.

**Keywords:** digital image, enhancement, historical sources.

Devido ao avanço tecnológico, cada vez mais pessoas têm acesso a imagens digitais, seja por meio de câmeras, celulares, scanners entre outros aparelhos, que posteriormente podem ser modificadas através de programas de tratamento de imagens. Isto permitiu que profissionais das mais variadas áreas se beneficiassem desses avanços, que passaram a ser integrados no seu dia-a-dia. Mais comumente, esses recursos são utilizados nos campos de conhecimento onde a tecnologia está mais presente, como nas ciências da saúde e nas ciências exatas. Por outro lado, a disseminação das câmeras digitais através da redução de seu custo possibilitou uma maior abrangência dos usuários dessa tecnologia, menos familiarizados com as potencialidades, os recursos e os benefícios do tratamento de imagem : 84).

---

\* UNISINOS, graduando do curso de Licenciatura em Matemática.

De acordo com : 1), o interesse em métodos de processamento de imagens digitais está vinculado a duas aplicações principais: 1) melhoria de informação visual para interpretação humana; 2) técnicas de processamento para percepção através de máquinas. Independentemente do tipo de aplicação, para interpretação humana ou para percepção através de máquinas, a Matemática tem um papel fundamental nas técnicas de tratamento de imagens digitais, uma vez que toda imagem digitalizada é armazenada em forma de matriz.

Em tempos de multidisciplinaridade e interdisciplinaridade, uma temática que merece uma reflexão mais aprofundada diz respeito à contribuição que o realce de imagens digitais pode trazer às mais diferentes áreas do conhecimento. Em áreas como a medicina, por exemplo, já existem diversas aplicações desenvolvidas comercialmente para o tratamento de imagens digitais. Por outro lado, outras áreas praticamente não se beneficiam dessas inovações tecnológicas, e por isso é um desafio mostrar as possibilidades que o tratamento de imagens digitais de registros paroquiais oferecem as ciências humanas, e mais especificamente para a História.

Os problemas específicos enfrentados pelos historiadores colocam questões que nos estimularam a encontrar soluções viáveis para o tratamento de fontes documentais que podem ter sofrido a ação do tempo e/ou as más condições de conservação e armazenamento.

Uma das matérias primas fundamentais da História é o documento escrito. Esse pode ser dividido em duas categorias básicas: 1) os documentos manuscritos; 2) os documentos impressos. Por se tratar de documentos em suporte de papel, estão sujeitos às vicissitudes do tempo e das condições de conservação e de armazenamento. Quando se trata de um documento manuscrito, não é viável o emprego de um programa de OCR (*Optical Character Recognizer*). A alternativa, neste caso, é tentar melhorar a qualidade da imagem para facilitar a leitura e interpretação pelo pesquisador.

Para além das vantagens óbvias que representam a melhoria da definição e qualidade das imagens dos documentos históricos, sobretudo tratando-se de manuscritos, uma outra ordem de considerações deve ser ressaltada. Ela está ligada à própria preservação de um patrimônio histórico documental, que em certos casos pode estar correndo risco de se perder. Ter os documentos reproduzidos digitalmente permite que mais pesquisadores tenham acesso ao mesmo, sem que isso implique na manipulação dos originais.

A idéia e a prática de digitalização de documentos históricos não é recente, são diversos os projetos que reúnem importantes instituições que se empenham para a preservação desse patrimônio documental. Temos por exemplo, o “Projeto Resgate Barão do Rio Branco” desenvolvido por iniciativa do Ministério da Cultura do governo brasileiro que se propunha,

em 1994, a disponibilizar para consulta aos pesquisadores brasileiros os documentos históricos relativos ao período colonial existentes em arquivos de outros países, através de microfilmes e CD-ROM : 7-11).

Um passo importante, sem dúvida, é a digitalização que implica em salvamento, preservação e disponibilização das fontes originais, como é o caso do projeto referido. Entretanto, é possível ir além, propondo-se o desenvolvimento de programas que permitam a melhoria da imagem original do documento, facilitando desta maneira o trabalho do especialista.

Nos últimos anos, tenho participado de inúmeros projetos que se utilizam desses documentos como fonte de dados, sobretudo através da exploração de registros paroquiais, e acompanhei, de perto, as dificuldades inerentes ao trabalho da pesquisa histórica.

O principal objetivo desta pesquisa é fornecer uma ferramenta que facilite o trabalho dos historiadores, através do processo de realce de imagens de documentos históricos, visando a melhoria da qualidade visual das imagens digitais, ao mesmo tempo que se permitirá a preservação dos originais e uma maior disponibilidade e divulgação dos mesmos. Os pesquisadores vinculados ao campo da História, têm ocupado uma boa parte do tempo de suas pesquisas em visitas a arquivos levantando fontes. Se pudermos disponibilizar as imagens digitalizadas desses documentos, além de mantermos a integridade dos documentos originais, poderemos atender muitos mais pesquisadores com os mesmos documentos podendo estes, abordarem problemáticas distintas com as mesmas fontes.

Em Setembro de 2006, participamos do XV Encontro Anual da ABEP (Associação Brasileira de Estudos Populacionais), onde apresentamos um estudo comparativo entre dois softwares que trabalham com fontes nominativas históricas, e apesar de nenhum deles fazer qualquer tratamento com imagens digitais. o SRP (Sistema de Reconstituição de Paróquias) já permite se fazer um link com a imagem digital original , isso nos mostra que, com a disseminação das máquinas fotográficas digitais, disponíveis com mais recursos a um custo aceitável, os pesquisadores estão se utilizando desse recurso para agilizar seu trabalho em arquivo.

Nossa meta principal é procurar obter um realce de imagens digitais coletadas em arquivos, possibilitando uma maior clareza na leitura e, na fase posterior, de interpretação desses documentos por parte dos pesquisadores. Sendo assim, este trabalho pretende utilizar ferramentas matemáticas para melhorar a qualidade visual das imagens digitais, facilitando assim o trabalho dos pesquisadores.

Recentemente, foi publicado um estudo que apresenta uma discussão sobre a utilização de câmeras para a análise de textos e documentos. Entre as questões tratadas pelos autores, destacam-se as possibilidades de realce de imagens a partir de diversas técnicas, como por exemplo a utilização dos recursos de brilho e contraste que representam as ferramentas básicas para o início do processo : 95).

Sabemos que as imagens digitais são definidas em *pixel*<sup>1</sup>, neste sentido, podemos trabalhar com as mesmas através de operações com matrizes como é bem definido por Gonzáles e Woods.

“O termo imagem refere-se a uma função de intensidade luminosa bidimensional, denotada por  $f(x,y)$ , em que o valor ou amplitude de  $f$  nas coordenadas espaciais  $(x,y)$  dá a intensidade (brilho) de imagem naquele ponto. Como a luz é uma forma de energia,  $f(x,y)$  deve ser positiva e finita, isto é,” : 21-22.

$$0 < f(x,y) < \infty \quad (1)$$

Uma imagem contínua  $f(x,y)$  é normalmente aproximada por amostras igualmente espaçadas arranjadas na forma de uma matriz  $N \times M$  como mostrado na equação 2, em que cada elemento é uma quantidade discreta (tal discretização, chamada de quantização, é necessária para armazenamento em forma digital), e cada elemento da matriz determina um elemento de imagem, *pixel*.

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,M-1) \\ f(1,0) & f(1,1) & \dots & f(1,M-1) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ (N-1,0) & f(N-1,1) & \dots & f(N-1, M-1) \end{bmatrix}, \quad (2)$$

<sup>1</sup> O Pixel (abreviatura de “*picture element*”, elementos de figura) é a menor unidade de uma Imagem, e quanto maior for o número de pixels, melhor a resolução que a imagem terá., em imagens em tons de cinza, a quantidade desses tons de cinza é determinada em BITS, sempre demonstrados em potência de 2, ou seja, uma imagem com oito bits, por exemplo, refere-se a 2 na potência 8, ou seja, 256 tons de cinza, e estes tons de cinza variam entre o branco e o preto.

Para que possamos efetuar este trabalho, primeiramente devemos definir qual será nosso ponto de partida, ou seja, qual o tipo de imagem que vamos contemplar para aplicar as técnicas mais adequadas para o realce de imagens.

Após a análise de algumas imagens geradas a partir de máquinas fotográficas digitais, optamos por utilizar imagens em tons de cinza e não uma imagem binarizada como por exemplo as imagens geradas dos documentos da Torre do Tombo conforme . Imagens binárias só apresentam dois tons-de-cinza: preto e branco, e são normalmente obtidas a partir da imagem original através de um processo de limiarização. Na limiarização, existe um ponto de corte (chamado limiar), de modo que intensidades maiores do que tal limiar são saturadas para o branco, enquanto que valores menores do que o limiar são saturados para preto. Tal processo gera perda de informação que não pode ser recuperada (por exemplo, o ruído da imagem do verso da folha fotografada que devido a fragilidade do papel e ação do tempo pode ser vista na frente do papel). Por outro lado, as imagens digitais armazenadas em tons de cinza possibilitam um ajuste mais fino do que é ruído e que deve ser retirado e o que é imagem que deve ser mantido.

Trabalhamos com a utilização de alguns filtros (funções matemáticas que podem ser aplicadas em matrizes) para propor o mais adequado ao tratamento das imagens digitais de documentos históricos. Para isto utilizamos o software MATLAB<sup>®2</sup> que permite não só efetuar diversas operações com matrizes, como também a possibilidade de gerar um aplicativo para que os pesquisadores utilizem em suas imagens digitais.

Uma forma eficiente de se conseguir alcançar o realce das imagens digitais de documentos históricos é através da morfologia matemática.

*“como em biologia onde morfologia refere-se ao estudo da estrutura dos animais e das plantas, a morfologia matemática, elaborada inicialmente por Georges Matheron e Jean Serra, concentra seu esforço no estudo da estrutura geométrica das entidades presentes numa imagem” : 109).*

A morfologia matemática, ramo do processamento e análise não linear de imagens permite processar imagens com o objetivo de realçar, segmentar e detectar bordas, entre várias outras aplicações.

Na morfologia matemática, a erosão e a dilatação são a base para a maioria das operações que podemos utilizar no realce de imagem. No caso de aplicarmos a erosão : 292), definimos uma imagem pela função  $f$  e o elemento estruturante por  $s$ , a erosão em escala de cinza  $f \ominus s$  no pixel  $(x,y)$  será:

<sup>2</sup>

© 1994-2006 The MathWorks, Inc.

$$(f \ominus s)(x,y) = \min_{jk}[f(x+j, y+k) - s(j,k)] , \quad (3)$$

onde  $j$  e  $k$  é o endereço do pixel na matriz.

Já a dilatação de uma imagem definida pela função  $f$  e elemento estruturante  $s$ , será  $f \oplus s$  no *pixel*  $(x,y)$  será:

$$(f \oplus s)(x,y) = \max_{jk}[f(x-j, y-k) + s(j,k)] \quad (4)$$

Além das operações morfológicas, podemos também aplicar um filtro delimitando os níveis de cinza através de histograma para definir qual a faixa mais relevante da imagem que devemos manter e qual poderia ser desconsiderada. Tal procedimento seria interessante para binarizar a imagem, com o objetivo de reduzir o tamanho de armazenagem sem comprometer a qualidade da informação

Com este trabalho, espera-se disponibilizar um programa que permita aos pesquisadores que se utilizam de imagens digitais de fontes manuscritas, uma melhoria na definição das mesmas, quando possível.

Entre as principais repercussões e impactos deste trabalho destacam-se ao menos duas. Em primeiro lugar a possibilidade de, através da digitalização de imagens de fontes documentais, contribuir para a preservação do patrimônio histórico e da memória nacional, regional e/ou local, evitando o manuseio contínuo de documentos originais e que poderia levar a uma degradação irreversível do material em questão. Em segundo lugar, o desenvolvimento de um programa que tenha como objetivo o realce de imagem, que disponibilizará uma ferramenta de trabalho que facilitará, e muito, a atividade não só de historiadores e pesquisadores das ciências humanas em geral, mas também será de grande valia para os profissionais da área de arquivística, que poderão se valer de seus recursos para salvaguardar um patrimônio documental que é do interesse da sociedade como um todo.

A importância deste tema pode ser avaliada nas recomendações do NARA (*National Archives and Records Administration*) para o trabalho de digitalização de arquivos.

*“Num dos maiores projetos de escaneamento de manuscritos até o presente momento, o Archivo General de Indias, em Sevilha, Espanha, está escaneando cerca de oito milhões de páginas que documentam a colonização espanhola das Américas. Bons resultados foram obtidos com a utilização de uma resolução de 100 dpi (“Dots per inch” pontos por polegada) com 4 bits de cinza (inicialmente capturados a 256 níveis de cinza, mas somente os níveis contíguos mais significativos são retidos)<sup>3</sup>. Contudo, o Projeto Sevilha não tem como objetivo o uso da reprodução digital de imagem para propósitos de substituição. Não ficou claro, a partir dos relatórios publicados, que este nível de resolução proporcione a captura completa de informação. Peter Robinson, em uma contribuição importante à literatura, *The digitization of primary textual sources*, recomenda acentuatadamente que materiais manuscritos “de qualquer tipo” sejam capturados como imagens em cores, com um mínimo de 300 dpi e cor de 24 bits<sup>4</sup>”: 23).*

Como podemos observar, a problemática de digitalização de documentos históricos já é estudada há algum tempo, porém, com este trabalho, queremos ir além da simples digitalização, trabalhando efetivamente no realce e melhoria dessas imagens.

## **Referências Bibliográficas**

---

<sup>3</sup> Hans Rütimann; M. Stuart Lynn. Computerization project of the Archivo General de Indias. Sevilhe, Spain: a report to the Commission on Preservation and Access, p.7. Apud :23

<sup>4</sup> Peter Robinson. *The digitization of primary textual sources*. P. 29. Apud : 24.