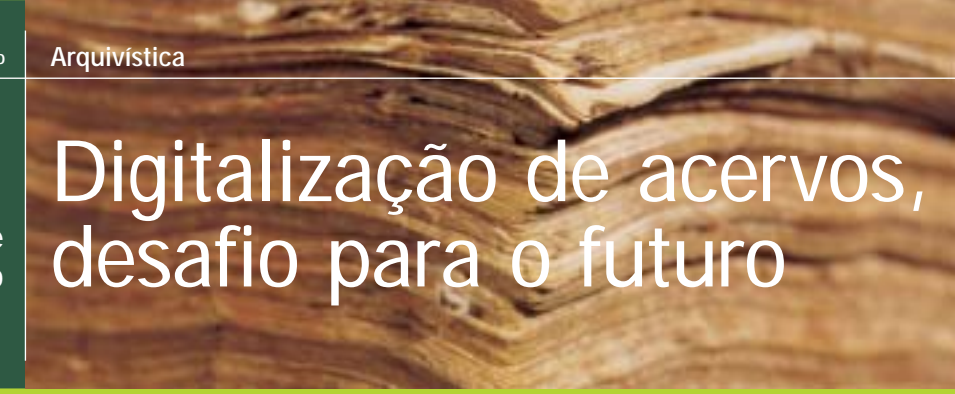



Revista do Arquivo Público Mineiro

Arquivística

Eduardo Valle
Arnaldo Araújo



Digitalização de acervos, desafio para o futuro



As inúmeras possibilidades que a digitalização oferece à preservação de acervos supõem a necessidade de estratégias a longo prazo para sua utilização, sob pena de se colocarem artefatos de valor permanente à mercê da fragilidade da tecnologia digital.

> Se o computador é uma invenção recente, sua vulgarização se deu com tal ímpeto que se vem assistindo ao crescimento exponencial de praticamente todos os indicadores a ele relacionados. A maior parte dos trabalhos de impressão e escrita de textos é feita em computadores, a gravação profissional de som é quase toda digital, câmeras fotográficas digitais estão substituindo as baseadas em filmes e mesmo a gravação de vídeo caminha inexoravelmente no mesmo sentido. O fato de a tecnologia digital ser ao mesmo tempo tão recente, tão ubíqua e de evolução tão veloz provoca severas consequências para a preservação dos documentos criados sob sua égide. [TFADI, 96].

Os dados digitais são afetados pela fragilidade de seus suportes – fitas e discos ópticos e magnéticos –, questão que não abordaremos aqui, para limitar nosso escopo. De qualquer forma, muito mais severos são os problemas provocados pela obsolescência da tecnologia. A política da indústria de informática (incluindo a nefasta *obsolescência programada*, ou seja, o abandono calculado das tecnologias por razões mercadológicas), bem como certos comportamentos dos usuários (como compressão e criptografia) também não facilitam a tarefa de preservação.

Besser [BESSER 00] estabelece a natureza da informação digital quanto à questão da longevidade, identificando desafios técnicos, informacionais e políticos:

visualização - a informação digital requer manutenção de uma infra-estrutura e de uma base de conhecimento para ser visualizada. Não basta preservar os dados digitais sem o *software* necessário para visualizá-lo, e o *hardware* necessário para executar o *software*. No mínimo, é preciso reter o conhecimento acerca da codificação dos arquivos para que se possa interpretar seu conteúdo;

embaralhamento - práticas comuns para resolver problemas de curto prazo acabam resultando em problemas de preservação. Dois exemplos são a compressão de dados e

a criptografia. A compressão de dados adiciona uma camada de complexidade à interpretação, impedindo que os dados sejam interpretados, a não ser que o método de descompressão seja conhecido ou o *software* de descompressão esteja disponível. A criptografia apresenta ainda mais problemas, pois mesmo que os métodos sejam conhecidos pode ser impossível resgatar as mensagens sem as chaves, e é fácil que estas sejam acidentalmente perdidas no decorrer da vida dos dados;

inter-relação - no mundo digital, as informações estão cada vez mais inter-relacionadas através de recursos como a incorporação e hiperligação. Páginas da *web*, por exemplo, mesmo que sejam exibidas de forma unificada, são compostas por diversos arquivos que, às vezes, residem em computadores muito distantes entre si. Está se tornando progressivamente mais difícil determinar os limites dos itens de informação e identificar seus contextos;

tradução - quando o conteúdo é traduzido para novos formatos, freqüentemente a mudança de forma provoca uma certa alteração no conteúdo. Sucessivas migrações de formato em arquivos digitais provocam constantes oportunidades para erros e imprecisões de tradução. Esforços de emulação nem sempre conseguem capturar todos os aspectos do ambiente simulado. Além disso, para que a operação de salvar um arquivo seja fiel, é preciso não apenas armazenar seu conteúdo, mas permitir que o comportamento do usuário ou consultante e sua forma de interação com o documento sejam preservados;

custódia - embora as organizações tradicionalmente tenham desenvolvido a preocupação de preservar e manter vários tipos de material analógico, esse cuidado não se estendeu ainda aos dados digitais. Por isso, a maior parte do material produzido digitalmente não é atribuída a responsáveis por sua custódia e provavelmente não estará disponível para as futuras gerações.

Diversos incidentes já atingiram o *status* de contos admonitórios, como os que envolveram os dados do censo norte-americano de 1960 ou o mapeamento de uso das terras do Estado no Arquivo do Estado de Nova York, gerado no final da década de 1960. O alerta para a ameaça à memória contemporânea estimulou a formação de consórcios como a força-tarefa da *Comission on Preservation and Access e o Research Libraries Group*, que identificou que a preservação dos dados digitais está sujeita aos seguintes fatores [TFADI, 96]:

conteúdo - a preservação do conteúdo digital é complexa e aparece em diferentes níveis de abstração. É necessário, mas não suficiente, preservar a seqüência de números zero e um que compõe a informação em seu nível elementar. É preciso também preservar os mecanismos que tornam essa seqüência inteligível aos seres humanos;

imutabilidade - é preciso preservar com segurança a autenticidade dos documentos, ou seja, garantir que não haja possibilidade de adulteração ou supressão dos dados. Em caso contrário, o valor comprobatório do registro digital fica nulo e mesmo seu valor cultural é grandemente diminuído;

referência - é preciso criar mecanismos consistentes de referência aos dados, de forma que eles possam ser associados a identificadores únicos, como números de chamada, códigos de notação etc;

proveniência - toda a arquivística moderna é centrada no princípio da proveniência. Esse princípio estabelece que parte da integridade de um documento reside em conhecer sua origem. Para preservar a integridade da informação digital, deve ser possível registrar a sua origem e a sua cadeia de custódia;

contexto - para que a informação permaneça íntegra é preciso preservar o seu contexto. No caso da informação digital, há um contexto técnico — *hardware, software*, a

tecnologia de que eles dependem; mas também um contexto de interdependência informacional, pois no mundo digital, freqüentemente, os itens documentais fazem referências uns aos outros. Há ainda que se considerar o contexto comportamental, isto é, a forma como o usuário se relaciona com os dados, que é dependente da tecnologia disponível no momento em que eles foram criados.

Os documentos digitais não sobrevivem sem uma estratégia constante de proteção aos seus mecanismos de armazenamento e visualização, uma vez que estes estão sujeitos a se tornarem indisponíveis devido à rápida obsolescência. Vários métodos têm sido estudados para permitir essa proteção: o "refrescamento" dos suportes *envelhecidos*; a migração dos dados para novos formatos e/ou plataformas, a "emulação" de sistemas ou plataformas obsoletos; a padronização dos formatos dos arquivos e até mesmo o "encapsulamento" dos dados junto aos programas de visualização. O mérito relativo dessas técnicas é alvo de muita controvérsia, embora, provavelmente, um plano de preservação eficaz deva combinar diversas delas. [VALLE, 03] [AMORIM, 05]

Digitalização como técnica de reformatação

Dada a dificuldade de se obter longevidade digital, parece contraproducente utilizar a técnica de digitalização para reformar dados com objetivo de preservação [WEBER, 97]. Entretanto, certos benefícios tornam os sistemas digitais tão atraentes que, às vezes, compensa sua utilização como meios ou, ao menos, como coadjuvantes na preservação de acervos.

O primeiro desses benefícios, e talvez o mais importante, é a perfeita replicabilidade dos dados digitais: uma cópia de um artefato digital é um clone indistinguível do seu original, ao contrário do que acontece com os dados analógicos em que as cópias são sempre imperfeitas, sujeitas a ruídos e atenuações.

O segundo fator é a facilidade de manipulação e análise dos dados digitais por causa de sua natureza numérica. Imagens digitais permitem uma vasta gama de operações de realce, restauração, filtragens, ampliações e correções que só estão disponíveis para fotografias analógicas através de morosas e delicadas intervenções de laboratório.

Não se pode negar que, no mínimo, a tecnologia digital pode propiciar um excelente acesso ao conteúdo intelectual dos artefatos digitalizados, resguardando os originais da manipulação desnecessária, e retardando sua deterioração. Para impedir a degradação provocada pelo acesso direto e constante, itens valiosos são guardados em arquivos seguros, disponíveis para poucos pesquisadores. Isso é frustrante, pois, quando os documentos estão fora do alcance do público, a tarefa de manter a memória viva não está sendo cumprida adequadamente. A tecnologia digital surge como uma possibilidade de romper o compromisso entre preservação e acesso: de opostas, essas dimensões se tornam cooperantes.

Em muitos casos, porém, a tecnologia digital passa de coadjuvante a protagonista da preservação de acervos. Os resultados obtidos na análise científica dos dados digitais podem ser usados no processo de conservação e restauração dos artefatos originais [LAHANIER, 02]. Em alguns casos, a natureza dos artefatos a serem preservados é tão complexa que a tecnologia digital pode oferecer uma forma mais conveniente de representar seu conteúdo intelectual — como na preservação de sítios arqueológicos.

A perfeita replicabilidade também é um fator de interesse nos casos em que a reprodução analógica introduz perdas inaceitáveis de uma geração a outra: é o caso da fotografia colorida e dos filmes. Esforços de reformatação para o meio digital, além de todas as preocupações relativas à longevidade, acrescentam-se à necessidade de capturar o mais fielmente possível os artefatos.

Natureza da imagem digital

A forma mais convencional de representar e armazenar imagens em sistemas digitais é através das chamadas "imagens de varredura" (*raster images*). Essas imagens são formadas por uma grade bidimensional de amostras de valores de tonalidade, para imagens coloridas, ou luminosidade, para imagens monocromáticas. Cada uma dessas amostras é chamada de *pixel* ou *pel*, uma abreviatura para "elemento da figura" (*PICTure Element*).

A correspondência entre as dimensões de cada amostra da imagem digital e as dimensões da imagem representada no mundo real fornece a "resolução" da imagem. A resolução pode ser expressa como a dimensão das amostras (e.g., *pixels* de 250 μ m), ou mais comumente pelo número de amostras por unidade linear, como 300 *pixels* por polegada ou 300dpi (*Dots Per Inch*). No universo das imagens analógicas, especialmente na fotografia à micrografia, a "resolução" é freqüentemente expressa em pares de "linha por milímetro". Como cada par de linhas equivale a dois *pixels*, e uma polegada tem 25,4mm, a conversão entre as duas unidades é simples (equação 1).

Para imagens monocromáticas, cada amostra representa um valor numérico de luminosidade, em uma escala que vai desde o preto até o branco, passando pelos cinzas intermediários. A capacidade de representação tonal da imagem é determinada pelo número de *bits* das amostras, às vezes chamado de "profundidade de" *bits* (*bit depth*). No caso limítrofe, em que cada amostra é composta por apenas um *bit*, só é possível representar duas luminosidades: a presença e a ausência da luz, formando as "imagens bitonais". À medida que se aumenta o número de *bits* por amostra, cresce a capacidade de representar luminosidades intermediárias. Valores típicos são 4, 8, 10, 12 e 16 *bits* por amostra, que permitem representar, respectivamente 16, 256, 1024, 4096 e 65.536 níveis diferentes.

Para imagens coloridas, cada amostra deve conter não apenas a representação da luminosidade, mas também da tonalidade. A informação de cor está ligada ao espectro da luz, isto é, à distribuição dos diferentes comprimentos de onda presentes no raio luminoso. Se fosse necessário que cada amostra descrevesse exatamente o espectro da luz coletado por ela, a representação de cor se tornaria inexecutável nos sistemas digitais, mas, felizmente, o sistema visual humano simplifica bastante a percepção da cor, dividindo-a em três faixas de comprimento de onda: mais curtas, correspondendo ao azul; mais longas, correspondendo ao vermelho, e intermediárias, correspondendo ao verde.

Assim, para efeitos de percepção humana, a complexa informação espectral só precisa ser descrita através de três valores referentes às cores primárias: verde, vermelho e azul. Isso faz com que uma imagem colorida seja equivalente, grosso modo, a três imagens monocromáticas superpostas (Figura 2). De fato, para uma dada profundidade de *bit*, uma imagem colorida será três vezes maior que a equivalente monocromática.

Uso das imagens

A primeira consideração a fazer na determinação dos parâmetros de qualidade das imagens digitais é o seu propósito. Essa questão afeta todas as decisões subsequentes, desde a resolução até o formato de armazenamento. Em ordem crescente de necessidade de fidelidade ao original, as imagens digitais podem ter o propósito de:

índice - imagens de pequeno tamanho e baixa qualidade, utilizadas como miniaturas em resultados de pesquisa ou em listagens impressas. Seu propósito é apenas a rápida identificação do conteúdo geral do documento para busca e identificação;

acesso - imagens de média qualidade que permitam uma visualização razoavelmente detalhada do conteúdo do

documento. Seu propósito é reduzir o acesso ao documento original, mas não substituí-lo. As imagens de acesso podem ser pré-tratadas para permitir melhor visualização do documento, pelo aumento de contraste, remoção de ruído etc;

reprodução - imagens de alta qualidade que permitam a duplicação do documento em impressora ou outro equipamento de imagem. Seu propósito é capturar a aparência do documento original em grau suficiente para permitir uma reprodução satisfatória, tão próxima do original quanto possível;

preservação - imagens da mais alta qualidade possível que procuram capturar todos os detalhes de interesse do documento, não apenas para as aplicações atuais, mas para possíveis aplicações futuras. As imagens de preservação não devem sofrer tratamentos de melhoria de contraste, retoques, alisamentos, aplicações de marca d'água, acréscimos de texto, nem outros tipos de alterações comuns nas imagens de acesso. Ela deve ser o testemunho mais fiel possível do documento em seu estado original.

Uma segunda dimensão é o tipo de análise que se espera fazer com a imagem. Se a imagem se destina apenas a olhos humanos, os padrões de qualidade podem ser mais relaxados: menor resolução, compressão mais agressiva. Entretanto, para análises quantitativas, restauração assistida por computador, reconhecimento de caracteres e visão computacional, os requisitos de qualidade são mais estritos.

É recomendável ser o mais conservador possível no sentido da alta qualidade, considerando não apenas os usos presentes, mas os possíveis usos futuros das imagens. O custo mais expressivo do processo de reformatação consiste na tarefa de separar e preparar os materiais para serem reformatados e posteriormente devolvê-los aos seus depósitos permanentes. De fato, acessar cada página ou

item a ser convertido representa o maior percentual do dispêndio. Uma certa extrapolação da qualidade de escaneamento evita incorrer no ônus inaceitável de passar por sucessivos processos de digitalização, à medida que surgem novas aplicações e necessidades.

Quando for necessário, por razões de acesso, disponibilizar arquivos de menor tamanho, é possível gerá-los automaticamente a partir dos registros de alta qualidade, por meio de programas que reduzem a resolução e comprimem as imagens. O contrário, entretanto, não é possível — nenhum procedimento é capaz de aumentar a resolução real de uma imagem e a compressão é frequentemente irreversível.

Dessa forma, um esquema comumente adotado é o escaneamento na mais alta qualidade possível, visando à geração de um máster ou matriz digital de preservação. Desses arquivos, grandes demais para serem acessados diretamente pelo usuário ou mesmo para residirem no sistema da informação *on-line*, são derivadas, automaticamente, as cópias para as finalidades de índice, acesso e reprodução. As matrizes são, então, transferidas para mídias terciárias (fitas magnéticas ou discos ópticos), onde ficam armazenadas até que seja necessário utilizá-las novamente — por exemplo, para criar uma nova geração, melhorada, de imagens de acesso.

Decidindo a resolução

Freqüentemente, a primeira decisão a tomar é a resolução de captura do material. Quanto maior a resolução, maior a qualidade da imagem digital, que equivale ao maior tamanho dos arquivos. De fato, devido à natureza bidimensional das imagens, quando a resolução dobra, o tamanho do arquivo resultante quadruplica. Assim, procura-se utilizar a menor resolução que atenda com segurança aos propósitos da imagem digital.

Para documentos iconográficos (fotografias, mapas etc.), em que o valor visual do documento é essencial, torna-se desejável que as matrizes digitais capturem o máximo de informação; se possível, todos os detalhes do original. Para isso, é preciso utilizar uma resolução de escaneamento igual ou superior à do documento original.

A resolução de um negativo ou *slide* fotográfico é determinada por fatores complexos, que incluem o tipo do filme, as qualidades ópticas e mecânicas da câmera e os cuidados e decisões no processamento do filme. Os melhores filmes fotográficos coloridos atualmente disponíveis atingem uma resolução de 100 lp/mm. Filmes em preto-e-branco podem atingir enormes valores de resolução (até 600 lp/mm em condições especiais). Entretanto, considerando o conjunto do sistema de captura fotográfica, com seus múltiplos componentes ópticos e mecânicos, esses níveis de qualidade podem ser considerados utopias teóricas. Valores mais típicos situam-se em torno de 20 a 80lp/mm (1000 a 4000dpi) [DIGITAL, 02] [PUTS, 01]. Na prática, isso implica em tamanhos de arquivos bastante elevados, embora os *scanners* de filmes atualmente disponíveis já ofereçam resoluções dessa ordem.

Quando a digitalização é feita através da cópia fotográfica em papel, deve-se considerar que resoluções menores são necessárias. De fato, a resolução de uma cópia fotográfica será sempre inversamente proporcional à sua ampliação e, às vezes, menor ainda devido às limitação da óptica do ampliador. Considere uma fotografia no formato mais popular atualmente, 10-15cm, gerada a partir de um negativo de 35mm (24-36mm), cuja resolução seja de 50lp/mm. A resolução da cópia, considerando um ampliador perfeito, será dada pela equação 2.

Tomemos uma fotografia de 10-15cm, ampliada de um negativo colorido de 35mm, com sensibilidade ISO400. A resolução adequada para digitalizar esse filme, que tem resolução entre 20 e 40lp/mm, extraindo toda a informação útil, estaria entre 1000 e 2000dpi.

Com a digitalização feita a partir da ampliação, a resolução adequada ficaria em torno de 240 e 480dpi. No experimento que fizemos, o escaneamento a 600dpi já revelou o grão do filme e em 1.200dpi a granularidade foi bastante acentuada.

Imagens produzidas em matrizes tipográficas ou *offset*, como pôsters, gravuras de livro ou de revistas, utilizam "meios-tons" na reprodução da gama tonal. A imagem não é formada por tons contínuos, como numa fotografia, mas decomposta em minúsculos pontos de cor pura, cujas dimensões ou densidade irão promover a sensação de cores intermediárias, quando o documento é visto à distância de leitura. A resolução da imagem equivale à da retícula utilizada para gerar o padrão de meios-tons, que é medida em linhas por polegada ou *lpi* (*Lines Per Inch*). Jornais utilizam retículas de cerca de 85lpi, um valor tão baixo, que os pequenos pontos da imagem são visíveis até mesmo a olho nu. Imagens utilizadas em publicações de alta qualidade podem atingir até 200lpi.

A regra simples para fazer o escaneamento de imagens impressas em meios-tons é usar o dobro da resolução da retícula. Assim, por exemplo, para um original gerado em retícula de 80lpi, deve-se usar um escaneamento de 160dpi.

Tomemos uma impressão em *offset*, colorida, gerada em tela de fotocomposição de 150lpi. A resolução adequada para digitalizar esse original seria de 300dpi. No experimento que realizamos, verificou-se uma melhoria de nitidez da imagem no aumento de 150dpi para 300dpi, mas aumentos de resolução subsequentes simplesmente ampliam a granulação da imagem.

Na digitalização de imagens de *offset*, um problema que, às vezes, surge é a formação de padrões de interferência entre o padrão da retícula de meios-tons e a grade de amostragem de digitalização. Esse efeito é chamado *moiré* e produz uma texturização bastante desagradável

na imagem resultante. O problema pode ser amenizado no pós-processamento da imagem, mas não removido completamente. Às vezes, uma ligeira rotação do original ou alteração na resolução de escaneamento resolve o problema [WIKIPEDIA, 05], mas a única forma garantida de eliminá-lo é fazer o escaneamento em uma resolução pelo menos duas vezes superior à da retícula.

As imagens de acesso e reprodução, quer sejam diretamente digitalizadas, quer sejam produzidas pela compressão das matrizes digitais, têm requisitos de resolução menos estritos do que essas últimas. Ao se determinar a resolução utilizada nessas imagens, considera-se mais a capacidade do sistema visual humano do que a natureza dos artefatos. A resolução da visão humana é medida em unidades angulares e não em unidades lineares, revelando o fato óbvio de que a capacidade de distinguir um par de linhas depende da distância desse par de linhas. A resolução de visão não excede um minuto de grau, o que equivale, para um objeto a 25cm (a distância normal de leitura de um material impresso), a cerca de 8lp/mm ou 400dpi.

Contudo, se o usuário deseja ter a possibilidade de ampliar a imagem sem perdas de resolução (como se ele estivesse utilizando uma lupa para observar o documento original), a digitalização deve levar isso em consideração. A fórmula da resolução de aquisição é dada pela equação 3. Portanto, supondo um original de 10-15cm, que se deseja permitir ao usuário imprimir nas dimensões de 15-22cm em uma impressora de 300dpi, deve-se escaneá-lo a 450dpi.

Ao determinar a resolução de imagens de acesso e reprodução, devem-se considerar os seguintes fatores:

□ se o objetivo for a visualização da imagem em um monitor de vídeo, a resolução desses dispositivos não ultrapassa 150dpi (valores mais típicos estando em torno de 90dpi). Assim, uma imagem de 450dpi ultrapassa em três vezes a resolução necessária (desconsiderando a ampliação):

❑ se não se estiver salvando as matrizes de digitalização, deve-se extrapolar, dentro das possibilidades do projeto, a qualidade das imagens de acesso e reprodução. Isso irá garantir que elas não se tornem o fator limitante de qualidade, à medida que a tecnologia de monitores e impressoras avança;

❑ não faz sentido utilizar uma resolução de escaneamento superior à de preservação, mesmo que se deseje possibilitar grandes ampliações. O resultado final sempre será limitado pela resolução do original. As chamadas *resoluções* "interpoladas", ou seja, um aumento "simulado" da resolução do original podem ser utilizadas (embora a resolução real nunca aumente). Contudo, é melhor armazenar o arquivo na resolução do original e interpolar no momento do uso para evitar desperdício de espaço de estocagem.

Para documentos de natureza textual, a perfeita reprodução da aparência visual não é tão importante quanto a legibilidade. Kenney e Chapmam adaptam um método para medição da qualidade em microfilmes — o Índice de Qualidade (IQ) — para a digitalização de textos. Segundo eles:

As normas técnicas mais exigentes foram desenvolvidas para a indústria de micrográficos e se baseiam no método do Índice de Qualidade. De fato, os procedimentos de controle de qualidade para inspeção de microfilme e o método do IQ para descrever a legibilidade de texto são bem adequados — com certas modificações — para prever e avaliar o desempenho de sistemas digitais de reprodução de imagens. [...] Contanto que uma câmera ou um *scanner* estejam operando em seus níveis ótimos, o IQ pode ser utilizado para prever os níveis de qualidade marginal (3,6), médio (5,0) ou alto (8,0) — que serão consistentemente alcançados na cópia de uso." [KENNEY, 97]

O método do Índice de Qualidade é muito simples e relaciona a altura do menor caractere de interesse no texto (normalmente a altura da letra "e" minúscula) e a resolução do documento. Ele pressupõe que, se o menor caractere do texto for representado por número razoável de pontos, o texto como um todo será legível. No IQ tradicional, para sistemas micrográficos, a altura é medida em milímetros e a resolução, em pares de linha por milímetro. Para se adaptar o sistema à tecnologia digital é preciso converter a fórmula para a resolução em pontos por polegada. A Equação 4 mostra as fórmulas.

Sumarizamos, na Tabela 1, algumas resoluções necessárias para atingir o IQ desejado em documentos com tipos de diferentes tamanhos.

A resolução utilizada no cálculo do IQ não é simplesmente a resolução nominal de escaneamento das imagens, e sim a resolução efetivamente obtida, ou a "resolução de saída". Esta resolução normalmente é menor que a resolução nominal por causa dos chamados "erros de registro", que fazem com que os pontos da imagem nem sempre correspondam exatamente às amostras coletadas pelo elemento do *scanner*. Problemas de foco, trepidação mecânica e outros também podem resultar numa resolução de saída inferior à nominal. A resolução de saída pode ser medida através do uso de um cartão de resolução, o que garante a fidelidade dos resultados.

Na falta desse cartão, Kenney e Chapmam recomendam, em escaneamentos multitonais, considerar a resolução de saída igual à resolução nominal e, em escaneamentos bitonais, considerar a resolução de saída como dois terços da resolução nominal [KENNEY, 97]. Nossos experimentos indicaram, entretanto, que esses números nada têm de absolutos, sendo sempre melhor usar o cartão de resolução. O comportamento da resolução de saída depende muito do equipamento utilizado para a aquisição, uma razão adicional para se usar o cartão em vez dos fatores de ajuste. [VALLE, 03]

Tonalidades

A importância relativa da cor e da reprodução tonal varia de acordo com a natureza dos documentos. Na maioria dos acervos textuais, cor e tonalidade são consideradas irrelevantes. Isso é particularmente verdadeiro na digitalização de microformas (microfilmes, microfichas), em que a informação de cor foi perdida e mesmo a tonalidade foi drasticamente reduzida. Em acervos iconográficos, entretanto, a boa preservação de cores e cinzas é crítica.

Uma imagem bitonal é oito vezes menor que uma em escala de 256 cinzas (o suficiente para capturar toda a gama perceptível pelo olho humano). Por causa disso e do alto contraste dos documentos textuais – tinta escura sob fundo claro – muito frequentemente se advoga o uso da imagem bitonal na digitalização de textos. Entretanto, a imagem multitonal preserva melhor a informação do documento e, em muitos casos, torna-o mais legível.

Textos com variações de contraste, borrados ou esmaecidos são mais bem capturados em imagens com tons de cinza. Mesmo documentos relativamente bem preservados, se impressos sobre marcas d'água ou com fundos coloridos, repetidamente têm sua digitalização mais bem sucedida em modo multitonal. Uma vantagem adicional para o usuário é a possibilidade de ajustar, interativamente e a seu gosto, brilho e contraste. Um fator adicional, o alisamento dos caracteres, também aumenta o conforto de leitura do usuário, especialmente no caso de tipografias delicadas.

Nosso parecer recomenda o uso de imagens bitonais apenas para documentos especialmente regulares: tipografados ou impressos, com boa legibilidade, sem grande degradação e em que o contraste entre o texto e o fundo seja elevado e constante. Nos outros casos, especialmente em textos datilografados, manuscritos, com diferentes níveis de contraste ou que apresentem deterioração, recomendamos fortemente a digitalização multitonal. Um

estudo detalhado sobre a digitalização bitonal de textos *versus* multitonal encontra-se em [VALLE, 03].

Quando, por razões de tamanho de arquivo, for considerado essencial utilizar imagens bitonais, um procedimento que resulta em melhores resultados é fazer o escaneamento em tons de cinza e depois converter a imagem utilizando um algoritmo selecionado, ao invés de confiar na bitonalização feita pelo próprio *scanner* que é mais grosseira. A imagem bitonal gerada por esse processo pode ser colocada em uso normalmente, e a imagem em tons de cinza usada como intermediária pode ser descartada ou armazenada como matriz digital.

O método mais simples utilizado para bitonalização é a "aplicação de limiar", e consiste em escolher uma tonalidade de cinza como limite entre os tons mais claros, que serão convertidos para branco, e os tons mais escuros, que serão convertidos em preto. A maioria dos *scanners*, ao fazer a digitalização em modo bitonal, o faz pela aplicação de limiar.

Para algumas imagens, entretanto, pode ser vantajoso utilizar os chamados métodos de meios-tons ou "pontilhamento" (*dithering*). Esses métodos distribuem uma combinação de pontos pretos e brancos na imagem, procurando simular o efeito de escala de cinza original. As imagens em meios-tons podem, inclusive, ser visualizadas em tons de cinza quando o usuário as estiver utilizando abaixo de sua resolução original, proporcionando excelente legibilidade. Elas não devem, porém, ser usadas em aplicações de reconhecimento de texto.

Se a decisão for pela digitalização multitonal, resta definir a profundidade de *bit*. Para aplicações convencionais, 8 *bits* costumam ser suficientes — o olho humano não distingue intervalos em uma variação contínua de 256 cinzas. Valores mais elevados são utilizados em aplicações especiais de visão computacional e tratamento de imagens. Para essas aplicações, 12 *bits* apresentam uma

margem segura. Em alguns casos, como na digitalização de microfílm de alto contraste, a gama tonal é muito limitada e 4 *bits* (16 tons de cinza) são suficientes.

Em busca da fidelidade

A preservação da informação de cor, quando é necessário alto grau de fidelidade, apresenta um desafio adicional para os atuais sistemas de imagem. Embora toda imagem colorida seja formada por uma composição de cores primárias, existe uma grande variação na escolha exata das primárias, bem como uma enorme disparidade na capacidade de os dispositivos representarem as milhões de cores perceptíveis ao olho humano.

A forma como os diferentes scanners fazem a leitura dos valores de verde, vermelho e azul não corresponde exatamente ao modo como os diferentes monitores definem e exibem essas cores. Pior do que isso, a maioria dos sistemas de impressão não se baseia nessas primárias aditivas, mas nas primárias subtrativas amarelo, magenta e ciano (frequentemente acrescidas de uma quarta "primária", o preto, para melhoria de contraste). Existe uma consideração adicional que é o fato de que quase todos sistemas de imagem — monitores, impressoras, *scanners* e até mesmo o olho humano — não se comportam de maneira linear quanto à percepção ou produção das cores, mas de forma exponencial. Na prática, isso provoca as seguintes dificuldades:

- ❑ os números que representam as cores precisam ser traduzidos entre sistemas que usam cores primárias diferentes;
- ❑ o comportamento não-linear dos diversos sistemas varia em expoente, necessitando de ajustes para fazer a correspondência entre os valores de cor;
- ❑ as diferenças de capacidade entre os dispositivos fazem com que a representação de algumas cores seja impossí-

vel — gerando a necessidade de encontrar uma cor aproximada. Os critérios para essa aproximação podem variar bastante (preservar as relações tonais da imagem, modificar o menor número de cores possível, maximizar a saturação, um compromisso entre esses fatores etc).

Para endereçar essas questões, vários sistemas de fidelidade de cor foram propostos para fazer corresponder as cores em diferentes dispositivos e solucionar a questão das cores não-representáveis. O sistema proposto pelo *International Color Consortium* requer que os dispositivos indiquem seus perfis de comportamento quanto à cor em um formato padronizado, e especifica técnicas para traduzir as cores entre equipamentos diferentes [ICC, 05]. Outro sistema, endereçado especificamente para a Internet e para a visualização de imagens em monitores, envolveu a criação de um espaço de cores padronizado, mas bastante restrito, chamado *sRGB* ou *RGB* padrão (*standard RGB*). [STOKES, 96].

Para trabalhos em que a representação da cor é realmente crítica, existem equipamentos e diagramas de calibração que permitem estabelecer relações de conversão entre os espaços de cor dos diferentes equipamentos utilizados no procedimento de digitalização. Usando adequadamente esses equipamentos, é possível, por exemplo, inspecionar no monitor o resultado do que foi escaneado e, em seguida, imprimir o documento com excelente previsão do resultado final. Infelizmente, os fenômenos de cor são complexos e a calibração feita para um certo iluminante (luz do dia) não funcionará necessariamente para outro (luz de lâmpada fluorescente). Além disso, é preciso recalibrar periodicamente os equipamentos, pois monitores, *scanners* e câmeras digitais apresentam desvios em seus espaços de cor ao longo de sua vida útil.

Formatos de arquivo: compressão

Uma vez adquirida a imagem digital, é preciso escolher o formato de arquivo usado para armazená-la. Para fins de

preservação, essa decisão é muito importante. Deve-se utilizar um formato amplamente compatível, bem documentado, de preferência definido por um consórcio aberto (em contraposição aos formatos proprietários de empresas), que preserve fielmente as informações coletadas na aquisição e ainda permita o acréscimo de alguns metadados à imagem digital — no mínimo os parâmetros utilizados na sua digitalização, e uma identificação do documento original. Na prática, entretanto, nenhum formato atualmente disponível atende plenamente a esses objetivos.

O TIFF (*Tagged Image File Format*), formato de arquivo etiquetado de imagem, aproxima-se dessa proposta ao permitir armazenar a imagem digital preservando toda a integridade dos dados. A documentação do formato é propriedade da *Adobe Systems*, mas o formato em si é amplamente documentado e a *Adobe* divulga livremente a especificação técnica [ADOBE, 92]. O formato conta também com uma vasta comunidade de usuários, que mantêm e trocam informações a seu respeito [RITTER, 97]. O TIFF possui o conceito de *etiquetas* - metadados que podem ser acrescentados ao cabeçalho do arquivo, com informações de identificação da imagem e detalhadas descrições de seu formato digital.

Contudo, por ser tão flexível, o TIFF tornou-se um formato complexo. Isso faz com que ele apresente problemas de compatibilidade, uma vez que nem todos os aplicativos dão suporte a todos os detalhes de sua especificação. É comum dois aplicativos não conseguirem compartilhar seus arquivos, mesmo quando ambos usam o TIFF. Tentativas de normalização, como o TIFF/IT, tiveram o efeito contrário de ampliar as variações encontráveis para o formato. O TIFF versão 6 permanece, porém, como um dos principais formatos de escolha para preservação digital. Sugere-se utilizar com parcimônia os recursos opcionais mais avançados do formato, de forma a se evitar problemas de compatibilidade.

Uma alternativa pode ser o formato PNG (*Portable Network Graphics*) — gráficos portáteis de rede. O PNG foi criado como um sucessor do GIF, um formato limitado e que, por usar um método de compressão patenteado, começou a ter seu uso desencorajado na *Internet* [ROELOFS 05]. As especificações do PNG foram publicadas pelo W3C e admitem menos variações de implementação que o TIFF, tornando o formato menos sujeito a incompatibilidades [LILLEY, 03]. O PNG oferece a possibilidade de anotação de metadados textuais, bem como características interessantes como correção de gama embutida e suporte aos principais espaços de cor dependentes e independentes de dispositivo.

O PNG é um formato de compressão sem perdas. Isso significa que um algoritmo reversível é utilizado para reduzir o tamanho da imagem armazenada, preservando-a sem nenhuma alteração visível ou invisível. Entretanto, mesmo a compactação sem perdas pode afetar a longevidade digital, dificultando eventuais tarefas de resgate de arquivos obsoletos. Por outro lado, um arquivo PNG é tipicamente menor que um TIFF — sem nenhuma perda de qualidade.

O formato JFIF – formato de arquivo para imagem JPEG (*Jpeg File Image Format*) – utiliza compressão com perdas para reduzir dramaticamente o tamanho das informações. Aproveitando-se de características do sistema visual humano, a compressão JPEG descarta parte da informação da imagem de forma a tornar o arquivo propício à compactação e em seguida aplica métodos que lhe permitem reduzir o tamanho do arquivo de 10 a 50 vezes e, em algumas ocasiões, ainda mais [JPEG, 04].

Embora o JFIF seja uma escolha pobre para as matrizes digitais, ele deve ser considerado para as imagens de acesso ou reprodução por causa da sua grande capacidade de compressão, mantendo a aparência da imagem. O grau de compressão determina a qualidade resultante do arquivo. Para reduções de até 10 vezes, as variações

costumam ser imperceptíveis. O fato, entretanto, de o JPEG ser um formato com perdas faz com que sua compressão não seja reversível: uma vez que um arquivo foi gravado em formato JPEG, a imagem original jamais pode ser reconstruída, exatamente, a partir dele. Além disso, a cada vez que o arquivo é salvo, novas perdas acontecem, fazendo com que o JPEG seja um formato inadequado para arquivos que precisem sofrer vários ciclos de retoque ou alteração. Um fator adicional a pesar na decisão é o fato de os defeitos de compressão serem invisíveis em condições normais de visualização, mas poderem se tornar muito evidentes se o usuário resolver ampliar ou mudar o contraste da imagem.

Estratégias de acesso

No universo das técnicas convencionais de conservação, a preservação e o acesso são dimensões que não só se distinguem, como freqüentemente se opõem. Muitas vezes, a única forma de garantir a preservação de um item é reduzir sua circulação.

Com a aplicação da tecnologia digital, esse cenário é radicalmente transformado, pois essas dimensões se tornam relacionadas e cooperantes. O desenvolvimento de uma estratégia de recuperação eficaz da informação é necessário para que a digitalização possa ser considerada uma atividade de preservação. "[...] gerar uma cópia de preservação de um livro deteriorado [...] sem tornar possível sua localização [...] é um desperdício de dinheiro. A preservação no universo digital descarta toda e qualquer noção dúbia que entenda preservação e acesso como atividades distintas." [CONWAY, 97] Isso se coaduna com a idéia de que a memória tem uma importância social que ultrapassa o simples arquivamento de artefatos.

É preciso impedir que os grandes volumes documentais se degenerem numa massa amorfa de informações mal classificadas e mal indexadas. Sem uma boa estrutura

informativa, dificulta-se o acesso ao acervo, pois cada consulta resulta em uma grande quantidade de respostas espúrias mescladas à informação útil. Assim, a coleção deve ser protegida contra sua própria desmesura, o que é feito através de mecanismos de arranjo e indexação e da criação de instrumentos de pesquisa que garantam a possibilidade de recuperação da informação.

Em acervos textuais, a busca por texto livre, associada aos dicionários de sinônimos e ao tratamento semântico, está revolucionando a forma como os consulentes — e até mesmo arquivistas — percebem as coleções. Se antes uma indexação cuidadosa com vocabulários controlados e tesouros era uma condição *sine qua non*, hoje é possível um tratamento sumário acompanhado da busca direta pelo conteúdo do texto. Entretanto, para possibilitar a busca textual não basta obter a imagem do texto, é preciso fazer o reconhecimento de caracteres (cujas vantagens vão além da simples busca, pois o texto pode ser copiado, colado e manipulado com mais conforto).

Existem duas tecnologias de conversão de texto em imagens para texto codificado em caracteres. A primeira, chamada de OCR (*Optical Character Recognition*) – reconhecimento óptico de caracteres – é normalmente utilizada para texto tipografado ou impresso em alta qualidade, em que o tipo dos caracteres é bastante legível e regular. O OCR utiliza métodos mais expressos e convencionais de reconhecimento da forma das letras. Uma segunda técnica, chamada de ICR (*Intelligent Character Recognition*) – reconhecimento inteligente de caracteres – é baseada em inteligência artificial e métodos estatísticos e aplica-se a textos mais problemáticos, como impressos matriciais, tipografias antigas, dactilografia e até mesmo manuscritos.

Quando essas tecnologias são utilizadas para fins de apresentação do texto convertido, taxas de reconhecimento muito elevadas são requeridas, pois os erros serão visíveis e, se freqüentes, torna-se mais econômico redigir o texto

do que reconhecê-lo e corrigi-lo. Felizmente, para simples operações de busca, é possível usar a técnica do OCR *sujo*: a busca por texto livre é feita a partir dos caracteres convertidos, com erros e tudo, usando um método de busca robusto contra imprecisões. Para a apresentação, porém, a imagem da página é utilizada.

Os acervos iconográficos também já se beneficiam da busca baseada no conteúdo. De fato, uma das aplicações mais interessantes da tecnologia digital decorre da possibilidade de se utilizarem métodos que interpretam diretamente as imagens, ao invés de confiarem apenas no texto a elas associados. Embora esses métodos ainda sejam incipientes, eles produzem bons resultados em algumas aplicações, e sistemas comerciais já estão disponíveis. Num futuro breve, o usuário poderá utilizar essas técnicas para consultar todas as imagens em que aparece determinada pessoa ou objeto ou, ainda, identificar as fotografias de exteriores ou de interiores, de paisagem urbana ou de paisagem natural, dentre outras possibilidades.

Conclusões

Quase sempre, a digitalização no universo de arquivos e bibliotecas é motivada pela facilidade de acesso ao acervo que propicia. Contudo, muitas vezes a preservação é um dos objetivos visados, especialmente quando se trata de documentos cuja complexidade dificulta o uso das técnicas convencionais. Objetos tridimensionais, discos e fitas de áudio, filmes e muitos outros artefatos são passíveis de representação no mundo dos computadores, mas neste artigo nos limitamos à aquisição de imagens bidimensionais, aplicável a uma vasta gama de acervos: mapas, plantas, pinturas, desenhos, livros, manuscritos, fotografias, microfilmes etc. Fizemos uma discussão dos desafios propostos pela presença do acervo digital e procuramos responder à questão de como ajustar os parâmetros técnicos da digitalização de forma a otimizar os resultados do processo.

A digitalização traz inúmeras possibilidades para o universo da preservação de acervos. Entretanto, sua aplicação em artefatos de valor permanente deve ser conduzida com cuidado, acompanhada de uma estratégia a longo prazo, sob pena de colocar o acervo à mercê da fragilidade da tecnologia digital. Em particular, a questão da longevidade digital e do acesso deve ser abordada em qualquer plano de reformatação para o meio digital.

A facilidade no uso corrente da informação digital provoca, freqüentemente, a ilusão de que o usuário detém sua custódia completa, e de que os recursos disponíveis para sua manipulação são, em si, flexíveis e de alta acessibilidade. Uma análise mais esclarecida demonstra, porém, que os dados digitais são fortemente dependentes de todo um contexto tecnológico/social para que se obtenha o mínimo grau de intelegibilidade.

Ainda assim, a preservação dessa informação se impõe como uma necessidade cada vez mais urgente, devido não apenas ao crescente patrimônio produzido através de computadores, mas também ao uso da digitalização como técnica de preservação em casos nos quais a reformatação convencional não apresenta bons resultados.

A representação digital sintetiza radicalmente o artefato original, pois consiste em uma coleção de "amostras numéricas" de mensurações feitas neste original. Por isso, é preciso decidir *a priori* quais os aspectos do documento se deseja preservar, em que grau de fidelidade, e garantir que o processo de conversão atenda a essas metas de qualidade. Assim, o uso da digitalização como forma principal ou colateral de preservação de acervos requer um cuidadoso ajuste de parâmetros de qualidade, e decisões que variam desde a compra dos equipamentos de aquisição até o formato de armazenamento dos arquivos.

Na digitalização de imagens, os principais parâmetros de qualidade são a resolução; a decisão de capturar uma imagem: bitonal, multitonal ou em cores; a inclusão ou

não de técnicas de calibração de cores; o formato de armazenamento dos arquivos e o uso ou não da compressão de imagens.

Ao decidir com que qualidade digitalizar, deve-se evitar cair em duas situações extremas. A primeira, ser muito modesto nos parâmetros de qualidade, ignorando toda extrapolação das necessidades presentes. Como o principal custo do processo consiste no "virar das páginas", é preciso esquivar-se do ônus de redigitalizar as imagens cada vez que surge uma nova aplicação.

A outra é dimensionar os parâmetros acima da qualidade dos próprios originais ou incluir tratamentos dispensáveis. Gerar arquivos desnecessariamente grandes irá onerar a etapa de estocagem a longo termo, o que, em última análise, é danoso para a preservação do acervo.

Esperamos que as orientações contidas neste texto possam auxiliar o profissional a trilhar o "caminho do meio", determinando, através da análise dos documentos originais, os parâmetros de digitalização ao mesmo tempo econômicos e seguros para a preservação de seus acervos a longo prazo.

Tabela e Equações

Tabela 1: Relação entre resolução e índice de qualidade em documentos textuais

Altura do menor tipo h (em mm)	Resolução (em dpi) necessária para		
	Marginal (3,6)	Médio (5,0)	Alto (8,0)
0,5	366	508	813
1,0	183	254	406
1,5	122	169	271
2,0	91	127	203
2,5	73	102	163

Equação 1:

Conversão das medidas de resolução em lp/mm para dpi

$$\frac{\text{Resolução Original}}{\text{Ampliação}} = \frac{\text{Resolução Original}}{\text{Dimensão da Cópia/Dimensão Original}}$$

$$\frac{50 \text{ lp/mm}}{10\text{cm} \cdot 24\text{cm}} = \frac{50 \text{ lp/mm}}{100\text{mm} \cdot 24\text{mm}} = \frac{50 \text{ lp/mm}}{4,2} = 12 \text{ lp/mm} = 610 \text{ dpi}$$

Equação 2:

Resolução necessária para a digitalização de uma cópia fotográfica

$$300 \text{ dpi} \times \frac{15\text{cm}}{10\text{cm}} = 450 \text{ dpi}$$

Equação 3:

Determinando a resolução das imagens de acesso

$$a) IQ = h \times p \quad b) IQ = h \times (r \downarrow 50,8)$$

Equação 4:

A fórmula do índice de qualidade para resolução em lp/mm (a); e em pontos por polegada (b). h é a altura em milímetros da letra "e" minúscula de menor tipo presente no texto; p é a resolução em lp/mm; r é a resolução em dpi

Referências bibliográficas

Um grande número de estudos tem sido conduzido a respeito da digitalização de imagens, seus formatos de armazenamento e sua conversão para texto, com fins de preservação de acervos. O Conselho em Bibliotecas e Recursos de Informação editou um conjunto de guias para a qualidade em projetos de imagem digital. Esses guias provêm diretrizes para todo o processo de captura do acervo desde o planejamento da digitalização, passando pela seleção do *scanner*, até a mensuração da qualidade e o formato de armazenamento para as matrizes digitais [CLIR 00]. Para os usuários e instituições interessadas em trabalhar com gerência de fidelidade de cor, Stone explica detalhadamente as questões envolvidas no processamento de imagens coloridas, bem como os diferentes sistemas [STONE 01]. McCarthy descreve um fluxo de trabalho para digitalização de materiais coloridos [MCCARTHY 02].

[ADOBE 92] Adobe Developers Association. *TIFF Specification Revision 6.0*. Adobe Systems Inc., Junho / 1992. Disponível na WWW: <<http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf>> 05/oct/2005.

[AMORIM 05] AMORIM, Eliane; LOPES, Carlos; CRUZ, Emília e VALLE, Eduardo. *Introdução à Preservação de Acervos Digitais*. Belo Horizonte: Secretaria de Cultura de Minas Gerais, Arquivo Público Mineiro, 2005. (In Prelo).

- [BESSER 00] BESSER, Howard. Digital Longevity. In: SITTS, Maxine (ed.). *Handbook for Digital Projects: a management tool for preservation and access*. Andover MA: Northeast Document Conservation Center, 2000. p.155-166. Disponível na WWW: <<http://www.gseis.ucla.edu/~howard/Papers/sfs-longevity.html>> 20/set/2005
- [CLIR 00] Council on Library and Information Resources. *Guides to Quality in Visual Resource Imaging*. Research Libraries Group, 2000. Disponível na World Wide Web <<http://www.rlg.org/visguides/index.html>> 05/oct/2005.
- [CONWAY 97] CONWAY, Paul. *Preservação no universo digital*. Coord. Ingrid Beck, Trad. Olga Marder. Rio de Janeiro, Arquivo Nacional, 1997. 24p. Grifo do autor. (Tradução de Preservation in the digital world). Disponível na WWW: <<http://www.clir.org/pubs/reports/conway2/index.html>> 20/set/2005.
- [DIGITAL 02] *Digital Photography vs Traditional Photography (How to compare)*. TheImage.com. Junho/2002. Disponível na WWW: <<http://www.theimage.com/photography/index.htm>> 05/oct/2005.
- [ICC 05] International Color Consortium. *Homepage*. 2005. Disponível na WWW: <<http://www.color.org>> 05/oct/2005.
- [JPEG 04] Joint Picture Expert Group. *Homepage*. 2004. Disponível na WWW: <<http://www.jpeg.org/>> 05/oct/2005.
- [KENNEY 97] KENNEY, Anne and CHAPMAN, Stephen. *Requisitos de resolução digital para textos: métodos para o estabelecimento de critérios de qualidade de imagem*. Coord. Ingrid Beck, Trad. José L. Pedersoli. Rio de Janeiro: Arquivo Nacional, 1997. 26 p. (Tradução de *Tutorial — Digital Resolution Requirements for Replacing Text-based Materials: Methods for Benchmarking Image Quality*). Disponível na WWW: <http://siarq02.siarq.unicamp.br/cpba/cadtec/cadtec_51.htm> 05/oct/2005.
- [LAHANIER 02] LAHANIER, Christian et al. CRISATEL: high definition spectral digital imaging of paintings with simulation of varnish removal. In: ICOM-CC 13th Triennial Meeting, 2002, Rio de Janeiro. *Annals of...* p. 295-300.
- [LILLEY 03] LILLEY, Chris. PNG (*Portable Network Format*). World Wide Web Consortium, 2003. Disponível na WWW: <<http://www.w3.org/Graphics/PNG/>> 05/oct/2005.
- [MCCARTHY 02] MCCARTHY, Ann. *Color Imaging Workflow Primitives*. International Color Consortium, 2002. Disponível na WWW: <<http://www.color.org/primitivelong.pdf>> 05/oct/2005.
- [PUTS 01] PUTS, Erwin. *Exploring the limits of 35mm BW Photography*. imX Photosite. 21/Maio/2001. Disponível na WWW: <<http://www.imx.nl/photosite/technical/highres.html>> 05/oct/2005.
- [RITTER 97] RITTER, Niles. *The Unofficial TIFF Home Page*. 1997. Disponível na WWW: <<http://home.earthlink.net/~ritter/tiff/>> 05/oct/2005.
- [ROELOFS 05] Roelofs, Greg. *Portable Network Format Home Site*. 2005. Disponível na WWW: <<http://www.libpng.org/pub/png/>> 05/oct/2005.
- [STOKES 96] STOKES, Michael et al. *A Standard Default Color Space for the Internet*. Version 1.1. International Color Consortium, 1996. Disponível na WWW: <<http://www.color.org/sRGB.html>> 05/oct/2005.
- [STONE 01] STONE, David. *Color Matching. Extreme Tech*. Ziff Davis Media: June / 2001. Disponível na WWW: <<http://www.extremetech.com/article2/0,3973,15467,00.asp>> 05/oct/2005.
- [TFADI 96] Task Force on Archiving of Digital Information. *Preserving Digital Information*. The Commission of Preservation and Access and The Research Libraries Group, 1996. 64p. Disponível na WWW: <http://www.rlg.org/en/page.php?Page_ID=114> 20/set/2005;
- [VAKULENKO 00] VAKULENKO, Alex. *Removing Moiré Patterns*. 2000. Disponível na WWW: <<http://www.oberonplace.com/dtp/moire/index.htm>> 05/oct/2005.
- [VALLE 03] VALLE, Eduardo. *Sistemas de informações multimídia na preservação de acervos permanentes*. Belo Horizonte: Universidade Federal de Minas Gerais, Departamento de Ciência da Computação, 2003. Dissertação de mestrado. Disponível na WWW: <<http://www.eduardovalle.com/permalink.htm?doc=doc00001>> 20/set/2005
- [WEBER 97] WEBER, Hartmut, DÖRR, Marianne. *Digitization as a Means of Preservation?*. European Commission on Preservation and Access, 1997. Disponível na WWW: <<http://www.clir.org/pubs/reports/digpres/digpres.html>> 05/oct/2005.
- [WIKIPEDIA 05] Wikipedia contributors. *Nyquist-Shannon sampling theorem*. Wikipedia, the free encyclopedia; 4/October/2005 11:42 UTC. Disponível na WWW: <http://en.wikipedia.org/wiki/Nyquist%E2%80%93Shannon_sampling_theorem> 05/oct/2005.

* Este artigo foi adaptado da dissertação de mestrado de Eduardo Valle, desenvolvida sob orientação de Arnaldo Araújo. Disponível na Internet no endereço <http://www.eduardovalle.com/permalink.htm?doc=doc00001> > 20/set/2005

Eduardo Valle é mestre em Ciência da Computação pela UFMG e doutorando no Équipes Traitement des Images et du Signal, laboratório partilhado entre a Université de Cergy-Pontoise e a École Nationale Supérieure de l'Électronique et ses Applications, França.

Arnaldo Araújo é professor adjunto do Departamento de Ciência da Computação da UFMG e coordenador do Núcleo de Processamento Digital de Imagens.