

Uma Medida Universal de Qualidade de Imagem

Alana Laryssa Seabra e Brunno Rafael
Centro de Informática
Universidade Federal de Pernambuco
Recife, Pernambuco, Brasil
alsas,brbc@cin.ufpe.br

Abstract—Ainda que o último teste de qualidade de uma imagem seja arbitrado pelos olhos humanos, a comunidade acadêmica tem realizado esforços para criar boas métricas objetivas de qualidade de imagem. As métricas são calculadas através de uma comparação entre duas imagens: uma chamada de ideal ou imagem perfeita e uma que foi degradada de alguma forma. Existem, basicamente, duas classes de medidas que resolvem esse tipo de problema: as que são matematicamente definidas e as que consideram informações do sistema visual humano (HVS). Neste projeto, replicamos e analisamos os resultados da nova medida proposta em A Universal Image Quality Index - trabalho de Wang e Bovik, publicado em 2002 -, que pertence a primeira classe de soluções e propõe a modelagem de qualquer distorção de uma imagem como uma combinação de três fatores: perda de correlação, distorções de luminosidade e distorções de contraste. Apesar de não considerar informação alguma sobre o sistema visual humano, a análise dos resultados usando imagens com diferentes tipos de distorções mostra que o desempenho deste método é significativamente melhor do que o popular método do erro quadrático médio.

Index Terms—Análise de imagem, medida de qualidade de imagem, erro quadrático médio.

I. INTRODUÇÃO

Imagens digitais são sujeitas a uma gama enorme de distorções durante seus processos de aquisição, processamento, compressão, armazenamento, transmissão e reprodução; e qualquer uma dessas etapas pode contribuir com uma degradação na qualidade visual da imagem final. Em aplicações onde as imagens são vistas, em última instância, por seres humanos, o único método “correto” de quantificar a qualidade visual da imagem é através de uma avaliação subjetiva. Entretanto, na prática, avaliação subjetiva é bastante inconveniente, demorada e cara. Somando este fator ao massivo número de aplicações de processamento de imagem que surgiram nas últimas duas décadas, podemos entender a crescente popularidade da área de pesquisa de avaliação de qualidade de imagens.

Qualidade da imagem é a característica de uma imagem digital que mede a degradação da imagem, geralmente comparada a uma imagem ideal ou perfeita. Portanto o intuito dos algoritmos que avaliam objetivamente a qualidade de uma imagem é desenvolver um indicador quantitativo que seja automaticamente gerado e que seja altamente consistente com o indicador da subjetividade humana. Uma medida objetiva de qualidade de imagem pode ser aplicada de várias maneiras em sistemas que contam com processamento de imagens.

Primeiro, ela pode ser utilizada para monitorar dinamicamente e ajustar a qualidade de imagem. Por exemplo, um sistema de aquisição ou streaming de imagem e vídeo pode usar a métrica de qualidade para controlar e, quando preciso, se auto-ajustar para obter a melhor qualidade de imagem e vídeo. Segundo, uma métrica como essa pode ser crucial em um sistema de benchmark. Suponha que, para uma tarefa específica, nós precisemos selecionar apenas uma de múltiplas imagens de um banco de dados. Uma medida de qualidade pode avaliar qual a melhor resposta a ser dada nesse sistema. Terceiro, ela pode ser incorporada a um sistema de processamento de imagens para otimizar o algoritmo e as configurações dos parâmetros. Tomando como exemplo um sistema de comunicação visual, uma medida de qualidade pode ajudar uma configuração ótima dos algoritmos de filtragem e atribuição de bits no codificador e dos algoritmos de pós processamento no decodificador.

Métricas objetivas de qualidade de imagem podem ser classificadas de acordo com a abordagem utilizada para o problema. Existem duas grandes classes que classificam as métricas. A primeira é a classe das medidas que são matematicamente definidas. Apesar de serem mais simples, são bastante utilizadas, como erro quadrático médio (MSE), relação sinal-ruído de pico (PSNR), raiz do erro quadrático médio (RMSE), média absoluta do erro (MAE), e relação sinal-ruído (SNR). A outra classe engloba métricas que consideram as características do sistema visual humano (HVS) em uma tentativa de incorporar medidas de qualidade que são mais perceptíveis [2][3]. Claramente esta última classe possui métodos mais refinados, entretanto, nenhum desses métodos objetivos mais rebuscados na literatura mostrou vantagens significativas em relação a medidas matemáticas como RMSE e PSNR em condições rigorosas de teste e imagens com diferentes cenários de distorção [4][5][6].

As métricas definidas matematicamente ainda são atraentes e consideradas vantajosas por duas razões principais. Primeiro, elas são fáceis de calcular e geralmente exigem uma baixa complexidade computacional, além de serem mais facilmente interpretadas fisicamente. Segundo, elas são independentes das condições de visualização e de observadores individuais. Acredita-se que as condições de visualização são importantes na percepção humana de qualidade da imagem, entretanto, na maioria dos casos, informações mais específicas não estão disponíveis para os sistemas de análise de imagem. Se existem N condições de visualização diferentes, um método que seja

dependente deste fator irá gerar N resultados de medidas diferentes que são inconvenientes para uso. Além de tudo, torna-se responsabilidade do usuário medir, calcular e informar as condições de visualização ao sistema. Por outro lado, um sistema independente de condições de visualização fornece um único valor de qualidade que dá uma ideia geral de quão boa é a imagem.

O trabalho [1] propõe uma medida matematicamente definida universal de qualidade de imagem. Por "universal", os autores querem dizer que a métrica é independente das imagens sendo testadas, das condições de visualização e de observadores individuais. Mais importante, a medida deve ser aplicável a várias aplicações de processamento de imagens e deve prover comparações significativas entre diferentes tipos de distorções. Atualmente, PSNR e MSE ainda são utilizados de forma universal, independentemente de suas performances questionáveis. O trabalho faz uma tentativa de desenvolvimento de um novo indicador para substituir os citados acima.

II. PROPOSTA

A. Ideia Geral

A ideia central desta métrica é representar a qualidade de uma imagem através de uma combinação de três fatores: perda de correlação, distorções de luminosidade e distorções de contraste. Seja $\mathbf{x} = \{x_i | i = 1, 2, 3, \dots, N\}$ e $\mathbf{y} = \{y_i | i = 1, 2, 3, \dots, N\}$ os sinais das imagens original e teste, respectivamente. A medida de qualidade Q é representada como um produto de três componentes:

$$Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\mu_x \mu_y}{\mu_x^2 + \mu_y^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$$

onde

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\mu_y = \frac{1}{N} \sum_{i=1}^N y_i$$

$$\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2$$

$$\sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \mu_y)^2$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$$

O primeiro componente é o coeficiente de correlação entre as duas imagens, o qual mede o grau de correlação linear entre \mathbf{x} e \mathbf{y} , e o seu intervalo dinâmico é $[-1, 1]$. O melhor valor 1 é obtido quando $y_i = ax + b$ para todo $i = 1, 2, 3, \dots, N$, onde

a e b são constantes e $a > 0$. Ou seja, o melhor valor para esta parcela é atingido quando as imagens são linearmente relacionadas. Entretanto, ainda assim pode-se haver distorções relativas entre as imagens, as quais serão contabilizadas nas segunda e terceira parcelas. O segundo componente, que possui intervalo $[0, 1]$, mede quão próxima é a luminosidade entre as imagens. A parcela atinge seu valor máximo se e somente se o valor médio dos sinais das imagens são iguais. σ_x e σ_y podem ser interpretados como estimativas do contraste de \mathbf{x} e \mathbf{y} , portanto o último componente da equação mede quão parecidos são os contrastes das imagens. Seu intervalo também é $[0, 1]$, e o melhor valor 1 é obtido se e somente se os contrastes forem iguais.

O intervalo dinâmico de Q é $[-1, 1]$. O seu melhor valor 1 é obtido se e somente se $y_i = x_i$ para todo $i = 1, 2, 3, \dots, N$. O menor valor -1 ocorre quando $y_i = 2\mu_x - x_i$ para todo $i = 1, 2, 3, \dots, N$. Reescrevendo a equação de forma mais sucinta, definimos a medida Q de qualidade de imagem como

$$Q = \frac{4\sigma_{xy}\mu_x\mu_y}{(\sigma_x^2 + \sigma_y^2)(\mu_x^2 + \mu_y^2)}$$

B. Detalhes da Implementação

Sinais de imagem, em geral, são não estacionários, por isso a qualidade de uma imagem, em teoria, também varia em relação ao espaço. Entretanto, na prática, o desejado é que a qualidade de uma imagem inteira seja representado com um único valor. Portanto, os autores do trabalho acharam mais apropriado que o cálculo das características estatísticas fossem realizados localmente e depois combinados. Os autores optaram pelo uso de uma janela deslizante para aplicar a medida de qualidade de maneira local. A abordagem da janela deslizante consiste em definir uma janela de tamanho $B \times B$ que se inicia no canto superior esquerdo da imagem. A janela então se move de pixel em pixel horizontalmente e verticalmente por toda imagem até alcançar o canto inferior direito. No j -ésimo passo, o índice local de qualidade Q_j é computado apenas dentro da janela. Considerando um total de M passos até o fim da imagem, o índice geral de qualidade é dado por

$$Q = \frac{1}{M} \sum_{j=1}^M Q_j$$

III. RESULTADOS

Para apresentar resultados concretos da técnica proposta, o artigo base se utilizou de diferentes tipos de distorções aplicados a uma imagem para realizar a comparação com o MSE e com avaliações subjetivas. Para tal, as imagens de teste foram distorcidas com os seguintes tipos de corrupção: ruído salt and pepper impulsivo, ruído Gaussiano aditivo, ruído Speckle multiplicativo, mean shift, realce de contraste, filtro de borramento e compressão JPEG. As imagens são mostradas na Fig.

1. As distorções foram ajustadas de modo a provocar um valor de MSE semelhante entre as imagens, exceto a compressão JPEG, na qual possui um MSE ligeiramente menor. O índice de qualidade universal (UQI) foi calculado usando uma janela deslizante de tamanho 8. Para os experimentos subjetivos, os autores mostraram as imagens para 22 indivíduos e pediram que eles classificassem as imagens em ordem de qualidade visual. Os valores médios do ranking, o MSE das imagens e o UQI são exibidos na Tabela I.

O artigo base propõe estas configurações de experimento para mostrar as notáveis falhas das medidas MSE e PSNR, quando comparadas a métricas um pouco mais sofisticadas. Os resultados do MSE são pobres no sentido de que imagens com valores quase iguais de MSE são drasticamente diferentes quanto à qualidade visual percebida.

Como é possível notar na Tabela I, os valores de UQI mostram uma correlação forte em relação às medidas subjetivas. Na verdade, a ordem dos valores de qualidade do novo índice é a mesma da que foi obtida com o experimento subjetivo. No artigo, os autores mencionam mais detalhes sobre o teste subjetivo e o que os indivíduos levaram em consideração na hora de propor o ranking.

Durante a implementação do método para este projeto, nós criamos nossas próprias imagens de teste aplicando diversos tipos de ruído. A nossa intenção era conseguir replicar este caso de teste e apresentar nossos próprios resultados neste trabalho. Primeiramente, utilizamos as imagens Lena e Peppers, bastante conhecidas no campo de processamento de imagens. Aplicamos diversos tipos de distorções nas imagens: o ruído multiplicativo Speckle, salt and pepper, ruído Gaussiano aditivo e filtros de borrramento (ou suavização) com núcleos gaussiano, de mediana e de média. Assim como os autores propõem, nós ajustamos os parâmetros das distorções de modo a gerar valores bem similares de MSE e PSNR na comparação das imagens distorcidas e a imagem original. Nós também utilizamos uma janela deslizante de tamanho 8 para computar o UQI.

Para o teste subjetivo, nós criamos um formulário com as 7 imagens (a original e 6 tipos de distorções) e pedimos para que os indivíduos ordenassem as imagens de acordo com a qualidade visual delas. 15 pessoas, englobando experientes ou não na área de processamento de imagem, responderam nosso formulário. As imagens apresentadas se encontram na Fig. 2 e Fig. 3 e os resultados obtidos podem ser visualizados nas Tabelas II e na Tabela III.

Nas nossas tabelas os índices de qualidade calculados não estão correlacionados com as métricas subjetivas. Acreditamos que essa falha pode ser atribuída ao tipo de teste subjetivo que fizemos. O nosso valor de MSE similar era bem inferior ao utilizado no artigo, o que significa que o erro entre as imagens que nós geramos era mais sutil do que o dos autores. O nosso formulário era digital e, das pessoas que o responderam, 53% responderam utilizando o celular, enquanto as outras utilizaram um computador ou laptop. O tamanho e a qualidade reduzidos da tela do celular podem ter influenciado significativamente as decisões dos participantes. Além disso,

acreditamos que um teste mais rígido deveria ser realizado se todos os participantes pudessem observar as imagens através de um dispositivo igual para todos. Portanto admitimos ser nossa a responsabilidade da falha na replicação dos testes propostos, de modo que não é possível a invalidação da métrica apresentada no trabalho.

IV. CONCLUSÃO E DISCUSSÃO

Um novo índice universal de qualidade de imagem foi proposto. Os resultados experimentais mostram que o método proposto tem uma performance significativamente melhor do que o MSE e PSNR em diferentes imagens com diferentes tipos de distorção. O que surpreende é que o método, além de matematicamente definido, é bem simples e fácil de entender, e ainda assim tem uma performance muito boa sem precisar depender de nenhuma informação do sistema visual humano.

Uma falha do MSE é que ele é sensível à energia dos erros, ao invés de ser sensível a distorções estruturais, como é este método. Talvez o sucesso da performance desta medida seja exatamente esse, a habilidade de modelar bem as distorções da imagem de forma estrutural.

Não restam dúvidas de que um bom método que tenha informações do sistema visual humano tenha sempre vantagem sobre os matematicamente definidos. Entretanto, sem um conceito matemático muito bem elaborado, os esforços na modelagem do sistema visual humano não irão obter bons resultados.

Os autores mencionam, no final deste trabalho, que acreditam que a ideia proposta é apenas um ponto de partida para o desenvolvimento de outros métodos de medição de qualidade de imagem e de vídeo. Os autores juntaram-se a Hamid Sheikh e Eero Simoncelli para continuar aprimorando as técnicas de medição de qualidade. E, dois anos após a publicação deste trabalho, elaboraram um método conhecido como índice de similaridade estrutural (SSIM), e o definiram no artigo [7]. Até o momento da escrita deste relatório, o artigo do SSIM já havia sido citado 13324 vezes, de acordo com o Google Scholar, sendo assim um dos artigos mais citados de todos os tempos nos campos de processamento de imagem e de vídeo. O artigo também ganhou o prêmio de melhor artigo da revista IEEE Signal Processing Society [8] no ano de 2009 e o método, hoje em dia, conta com inúmeras variantes.

Os autores perceberam que havia um gargalo no índice de qualidade universal (UQI). Como já citado acima, tal índice propõe a modelagem de qualquer distorção de uma imagem calculando (localmente) as distorções entre as luminâncias, contrastes e correlações entre a imagem de referência e a imagem distorcida. Um dos inconvenientes dessa medida é o fato de que ela se torna instável quando as médias das imagens se aproximam de zero ou quando as variâncias das imagens se aproximam de zero. O índice SSIM resolve este problema introduzindo constantes estabilizadoras. Ao fazer isso, ele continua possuindo os pontos positivos do UQI - método simples de ser entendido, independente de HSV e computacionalmente atrativo - enquanto evita o problema de instabilidade.



Fig. 1. Imagens utilizadas pelos autores para os resultados do artigo base. Todas as imagens tem 512x512 pixels, 8bits/pixel.

TABLE I
MEDIDAS DE QUALIDADE GERADA PELOS AUTORES

| Imagem | Tipo de distorção | Ranking subjetivo médio | MSE | Q |
|-----------|------------------------|-------------------------|-----|--------|
| Fig. 1(b) | Mean Shift | 1.59 | 225 | 0.9894 |
| Fig. 1(c) | Açuamento de contraste | 1.64 | 225 | 0.9372 |
| Fig. 1(d) | Salt and pepper | 3.32 | 225 | 0.6494 |
| Fig. 1(e) | Ruído Speckle | 4.18 | 225 | 0.4408 |
| Fig. 1(f) | Ruído Gaussiano | 4.27 | 225 | 0.3891 |
| Fig. 1(g) | Borramento | 6.32 | 225 | 0.3461 |
| Fig. 1(h) | Compressão JPEG | 6.68 | 215 | 0.2876 |



(a) Imagem original



(b) Ruído Gaussiano



(c) Ruído Speckle



(d) Filtro de média



(e) Filtro de mediana



(f) Salt and pepper



(g) Borramento

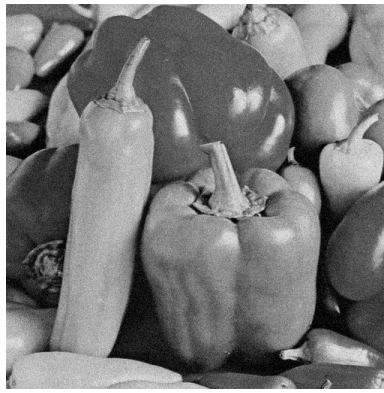
Fig. 2. Imagens "Lena" utilizadas por ns para a tentativa de replicação dos resultados. Todas as imagens tem 512x512 pixels, 8bits/pixel.

TABLE II
MEDIDAS DE QUALIDADE DA IMAGEM "LENA"

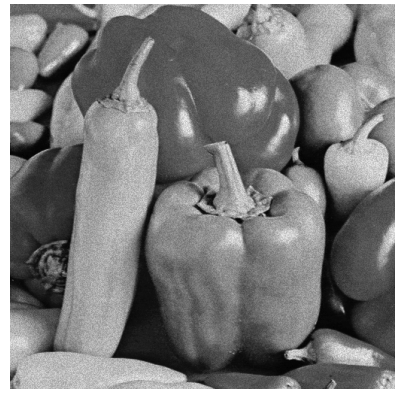
| Imagem | Tipo de distorção | Ranking subjetivo médio | MSE | PSNR | Q |
|-----------|-------------------|-------------------------|-------|-------|-------|
| Fig. 2(c) | Ruído Speckle | 1.66 | 125.6 | 27.13 | 0.513 |
| Fig. 2(b) | Ruído Gaussiano | 2.2 | 125.5 | 27.14 | 0.469 |
| Fig. 2(f) | Salt and pepper | 2.66 | 125.6 | 27.13 | 0.781 |
| Fig. 2(d) | Filtro de média | 3.8 | 126.1 | 27.12 | 0.582 |
| Fig. 2(e) | Filtro de mediana | 5.2 | 124.9 | 27.16 | 0.515 |
| Fig. 2(g) | Borramento | 5.46 | 125.4 | 27.14 | 0.510 |



(a) Imagem original



(b) Ruído Gaussiano



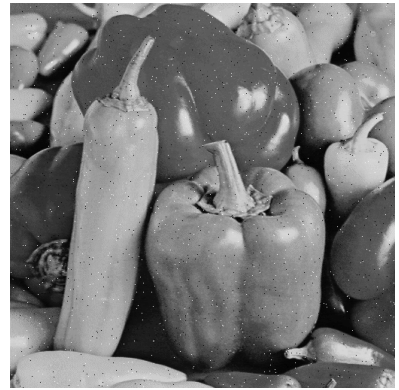
(c) Ruído Speckle



(d) Filtro de média



(e) Filtro de mediana



(f) Salt and pepper



(g) Borramento

Fig. 3. Imagens "Peppers" utilizadas por ns para a tentativa de replicação dos resultados. Todas as imagens tem 512x512 pixels, 8bits/pixel.

TABLE III
MEDIDAS DE QUALIDADE DA IMAGEM "PEPPERS"

| Imagem | Tipo de distorção | Ranking subjetivo médio | MSE | PSNR | Q |
|-----------|-------------------|-------------------------|-------|-------|-------|
| Fig. 3(c) | Ruído Speckle | 1.73 | 116.6 | 27.49 | 0.576 |
| Fig. 3(b) | Ruído Gaussiano | 2.2 | 116.8 | 27.46 | 0.513 |
| Fig. 3(f) | Salt and pepper | 2.46 | 117.2 | 27.21 | 0.801 |
| Fig. 3(d) | Filtro de média | 4.06 | 119.3 | 27.32 | 0.550 |
| Fig. 3(e) | Filtro de mediana | 4.86 | 112.2 | 27.62 | 0.473 |
| Fig. 3(g) | Borramento | 5.66 | 116.6 | 27.46 | 0.482 |

REFERENCES

- [1] Z. Wang and A. C. Bovik, *A Universal Image Quality Index*, in IEEE Signal Processing Letters, March 2002.
- [2] T. N. Pappas and R. J. Safranek, *Perceptual criteria for image quality evaluation*, in Handbook of Image and Video Processing (A. C. Bovik, ed.), Academic Press, May 2000.
- [3] *Special issue on image and video quality metrics*, Signal Processing, vol. 70, Nov. 1998.
- [4] J.-B. Martens and L. Meesters, *Image dissimilarity*, Signal Processing, vol. 70, pp. 155-176, Nov. 1998.
- [5] VQEG, *Final report from the video quality experts group on the validation of objective models of video quality assessment*, <http://www.vqeg.org/>, Mar. 2000.
- [6] A. M. Eskicioglu and P. S. Fisher, *Image quality measures and their performance*, IEEE Trans. Communications, vol. 43, pp. 2959-2965, Dec. 1995.
- [7] Z. Wang et al, *Image Quality Assessment: From Error Visibility to Structural Similarity*, in IEEE Transactions on Image Processing, vol. 13, April 2004.
- [8] IEEE Signal Processing Society, *Best Paper Award*, http://signalprocessingsociety.org/uploads/awards/Best_Paper.pdf