

Análise de Métricas de Similaridade Estrutural

Jean N. R. Araujo¹ Claudio de Casto Monteiro²

¹Instituto Federal de Educação, Ciência e Tecnologia do Tocantins - IFTO
Palmas – TO – Brasil

jean.to@gmail.com, ccm@ifto.edu.br

Abstract. *The approach which evaluates the image quality by checking structural similarity is based on the assumption that the HVS is highly adapted for extracting structural information of a scene and hence structural similarity metrics can provide a good approximation of the image quality perceived. The purpose of this article is to demonstrate the use of three metrics (SSIM, MS-SSIM and CW-SSIM) using this approach.*

Resumo. *A abordagem que avalia a qualidade da imagem por meio de verificação de similaridade estrutural é baseada na suposição de que o sistema visual humano (HVS) é altamente adaptado para extrair informação estrutural de uma cena e, portanto, métricas de similaridade estrutural podem fornecer uma boa aproximação da qualidade da imagem percebida. O objetivo deste artigo é demonstrar o uso de três métricas (SSIM, MS-SSIM e CW-SSIM) que utilizam desta abordagem.*

1. Introdução

As métricas objetivas podem ser classificadas de acordo com a disponibilidade da imagem original. A maioria das abordagens existentes são conhecidas como *full-reference* (com referência completa), o que significa que a imagem original é conhecida e utilizada na medição. Em muitos casos a imagem original não está disponível, neste caso a abordagem utilizada é conhecida como *null-reference* (sem referência). No terceiro tipo de método, as informações da imagem de referência são apenas parciais, conhecido como *reduced-reference* (com referência reduzida). Os algoritmos de similaridade estrutural se concentram entre os métodos objetivos de referência completa (*full-reference*).

A mais simples e mais amplamente utilizada métrica de qualidade de referência completa é o MSE (*mean squared error*), juntamente com a relação entre sinal e ruído (PSNR). Estes métodos são atraentes devido à simplicidade do cálculo e são matematicamente convenientes em um contexto de otimização, mas não são bem adaptados a qualidade percebida pelo espectador, e. g. [Girod 1993], [Teo e Heeger 1994], [Wang, Bovik e Lu 2002]. Nas últimas três décadas, um grande esforço foi canalizado para o desenvolvimento de métodos de avaliação da qualidade que se aproveitam de características conhecidas do sistema visual humano (HVS).

A abordagem de similaridade estrutural fornece uma alternativa complementar para resolver o problema da avaliação da qualidade da imagem, e. g. [Wang, Sheik e Bovik 2003]. Este método considera que o HVS é altamente adaptado para extrair informações

estruturais de determinada cena, e, portanto, uma medida de similaridade estrutural pode ser uma boa aproximação da qualidade da imagem percebida.



Em e. g. Wang (2004) é demonstrado que uma simples aplicação desta metodologia, a saber, o índice de semelhança estrutural (*SSIM – Single-Scale Structural Similarity*), supera métricas amplamente utilizadas como o PSNR. No entanto, o algoritmo SSIM apresentado é uma abordagem de única dimensão. Em e. g. Wang (2003) observa-se nos experimentos que o melhor desempenho é alcançado quando aplicado em uma escala adequada. O autor propõe então um método *multi-escala* de similaridade estrutural (MS-SSIM) e apresenta uma abordagem para calibrar os parâmetros referentes a importância relativa de diferentes escalas.

Outras métricas de similaridade estrutural vêm sendo propostas, em razão de algumas particularidades do algoritmo SSIM. Apesar do desempenho superior do SSIM sobre o MSE, ambos são muito sensíveis a distorções de imagem geométricas, tal como rotação e translação. A fim de eliminar esta dificuldade, a métrica CW-SSIM (*complex wavelet SSIM*) foi proposta em e. g. Wang (2005), e tem demonstrado ser uma métrica útil em uma série de aplicações, incluindo avaliação de qualidade de imagem, comparação de segmentação, reconhecimento de face baseado em intervalos e reconhecimento de impressão digital. Estaremos demonstrando a utilização das três métricas de similaridade estrutural citadas até aqui.

2. Metodologia e Resultados

As transmissões foram realizadas em um ambiente que dispunha de um servidor multimídia disponibilizando conteúdo de vídeo sob demanda através do *vlcserver*. No host do usuário conectado a rede foi implantado um *framework* de otimização e automatização das requisições de mídia e da coleta dos dados de transmissão referentes à rede e ao vídeo. As características dos vídeos utilizados são descritas na Tabela 01.

Tabela 1. Descrição das sequências de vídeo utilizadas nas transmissões

Vídeo	Info	Imagem
<i>akiyo</i>	300 frames 9 segundos 4:3	
<i>container</i>	300 frames 8 segundos 4:3	

Foi definida a quantidade de 120 transmissões para cada vídeo a fim de aplicar um valor de amostra populacional suficiente para demonstrar a diferença entre as três métricas de similaridade em um ambiente de rede ideal.

Após as transmissões dos conteúdos multimídia definidos pelas duas sequencias de vídeo, foi aplicado nas imagens extraídas dos vídeos transmitidos as três métricas de similaridade estrutural (SSIM, MS-SSIM e CW-SSIM). Com base nestes procedimentos, foram analisados os dados obtidos nas transmissões. As informações coletadas foram sintetizadas e dispostas em gráficos e tabelas. Os resultados são descritos a seguir:

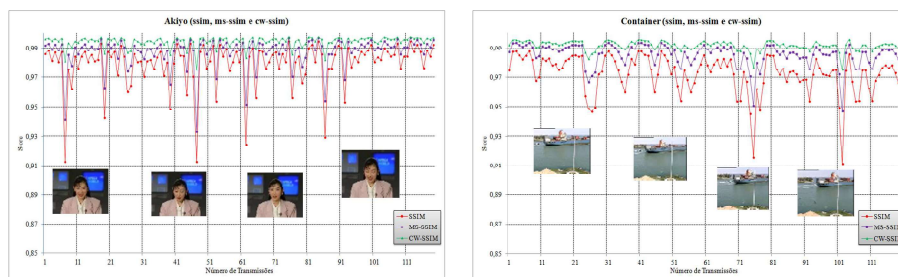


Figura 1. SSIM, MS-SSIM e CW-SSIM Score

Tabela 2. Média, máximo, mínimo e intervalo de confiança medidos para atraso, jitter, perda, ssim, ms-ssim e cw-ssim para o vídeo akiyo

	Atraso	Jitter	SSIM	MS-SSIM	CW-SSIM
Avg	0.25	0.067	0.9796	0.9874	0.9944
Max	1.231	1.156	0.995	0.997	0.998
Min	0.186	0.011	0.913	0.933	0.975
DP	0.0950	0.1042	0.0151	0.0151	0.0151
Erro	0.0171	0.0187	0.0027	0.0027	0.0027
IC	0.233 – 0.267	0.049 – 0.086	0.977 – 0.982	0.985 – 0.990	0.992 – 0.997

Tabela 3. Média, máximo, mínimo e intervalo de confiança medidos para atraso, jitter, perda, ssim, ms-ssim e cw-ssim para o vídeo container

	Atraso	Jitter	SSIM	MS-SSIM	CW-SSIM
Avg	0.236	0.051	0.9725	0.9855	0.9919
Max	0.311	0.124	0.987	0.994	0.996
Min	0.193	0.009	0.911	0.947	0.975
DP	0.021	0.0228	0.0131	0.0151	0.0151
Erro	0.0038	0.0041	0.0024	0.0027	0.0027
IC	0.232 – 0.240	0.047 – 0.055	0.970 – 0.975	0.983 – 0.988	0.989 – 0.995

Na figura 1, pode-se observar os resultados das 120 transmissões realizadas para as sequencias de vídeo, em um ambiente de rede ideal. Entende-se que as médias de jitter e atraso nas duas sequências demonstram que as condições da rede se encontram adequadas para transmissão de vídeo.

Os valores encontrados para as métricas SSIM, MS-SSIM e CW-SSIM demonstram que destas métricas a CW-SSIM apresenta valor um pouco mais condizente em relação a qualidade da imagem comparando com as outras duas métricas. No entanto, esta métrica CW-SSIM provoca maior custo de processamento e consequentemente menor grau de adaptação para utilizar em um ambiente de otimização. Na realidade, as métricas de similaridade estrutural, mesmo sendo mais eficientes que métricas como PSNR e MSE, não podem ser utilizadas em situações onde as imagens de referência não são conhecidas, dificultando o seu uso em ambientes mais realísticos de disponibilização de *streams* em uma rede.

3. Conclusão

A abordagem tradicional de avaliação de qualidade de imagem com base na sensibilidade ao erro, representada principalmente pelo PSNR, possui suas limitações já amplamente

estudadas e expostas. O uso de algoritmos de similaridade estrutural se mostra um caminho alternativo na definição de medidas de qualidade de imagem.

Inicialmente, existem duas abordagens de métricas de similaridade. A abordagem *single scale* representado pelo índice SSIM, e a abordagem *multi-scale* representado pelo índice MS-SSIM. A abordagem *multi-scale* de similaridade estrutural tem se mostrado mais flexível, pois incorpora as variações de resolução da imagem e as condições de visualização.

Outro método de similaridade estrutural é o índice CW-SSIM, que é, simultaneamente, insensível às mudanças de luminosidade, alteração do contraste e pequenas transições, dimensionamento e rotação de imagens. Estas características tornam esta uma abordagem flexível que tem boas potencialidades que podem ser aplicadas a uma grande variedade de problemas de classificação de imagens.

Referências

- Girod, B. (1993) “What’s wrong with mean-squared error,” in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, pp. 207–220.
- Teo, P. C. and Heeger, D. J. (1994) “Perceptual image distortion,” in *Proc. SPIE*, vol. 2179, pp. 127–141.
- Wang, Z. (2001) “Rate scalable Foveated image and video communications,” Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Texas at Austin, Austin, TX.
- Wang, Z. and Bovik, A. C. (2002) “A universal image quality index,” *IEEE Signal Processing Letters*, vol. 9, pp. 81–84.
- Wang, Z., Bovik, A. C. and Lu, L. (2002) “Why is image quality assessment so difficult,” in *Proc. IEEE Conf. Acoustics, Speech, and Signal Processing*, vol. 4, pp. 3313–3316.
- Wang, Z., Sheikh, H. R. and Bovik, A. C. (2003) “Objective video quality assessment,” in *The Handbook of Video Databases: Design and Applications* (B. Furht and O. Marques, eds.), pp. 1041–1078, CRC Press, Sept. 2003.
- Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P. (2004) “Image quality assessment: From error measurement to structural similarity,” *IEEE Trans. Image Processing*, vol. 13.
- Wang, Z. and Bovik, A. C. (2004) “Video quality assessment based on structural distortion measurement,” *Signal Processing: Image Communication*, special issue on objective video quality metrics, vol. 19.
- Wang, Z., Bovik, A. C. and Simoncelli, E. P. (2003) “Multi-scale Structural Similarity for Image Quality Assessment,” *37th IEEE Asilomar Conference on Signals, Systems and Computers*.
- Wang, Z. and Simoncelli, E. P. (2005) “Translation insensitive image similarity in complex wavelet domain,” in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, Philadelphia, PA.