Avaliação Objetiva e Subjetiva da Heurística *Early Skip/DIS*

Qualidade em Sistemas de Vídeo Digital

Érick Moreira, Heitor Almeida, Marcos Bueno, Ruhan Conceição, Thiago Bubolz

Universidade Federal de Pelotas

Ciência, Engenharia e Pós-Graduação em Computação

Pelotas, Brazil

*Resumo*—Este trabalho apresenta uma avaliação objetiva e subjetiva para a heurística *Early Skip*/DIS. A avaliação objetiva considerou três métricas: PSNR, SSIM e MAD. Além disso, a avaliação subjetiva utilizou vídeos em três dimensões. Resultados das avaliações demonstraram que a solução não apresentou impacto significativo a qualidade de imagem, quando comparado com o vídeo reconstruído a partir do 3D-HEVC sem modificações.

Palavras-chave—Codificação de Vídeos; Análise Subjetiva; Early Skip/DIS

# Introdução

O *High Efficiency Video Coding* (HEVC) [1] é o mais novo padrão de codificação de vídeos, sendo lançado em janeiro de 2013. Entre as principais metas do HEVC quando fora proposto, destacava-se atingir o dobro da taxa de compressão – quando comparado ao seu antecessor, o padrão H.264/AVC [2] – mantendo a mesma qualidade de imagem e complexidade computacional. Além disso, esperava-se que o HEVC apresentasse suporte à codificação eficiente de vídeos de resolução até UHD 8K (7680x4320 pixels), 16 vezes maior que a resolução HD 1080p para qual seu predecessor, H.264/AVC (REF-H.264), fora desenvolvido. Devido ao número de ferramentas e estruturas de codificação utilizadas para obter os excelentes resultados de compressão de vídeo, não foi possível ao HEVC atingir sua meta de manter a mesma complexidade computacional do H.264/AVC. Assim, o esforço computacional para o HEVC codificar um vídeo chega a ser 502% maior que o H.264/AVC [3].

A fim de expandir a eficiência de codificação a diversos outros tipos de aplicações, extensões para os padrões atuais de codificação de vídeos foram desenvolvidos visando dar suporte a vídeos capturados por mais de uma câmera simultaneamente. O *Multiview Video Coding* (MVC) [4] foi o padrão de codificação de vídeos multivista desenvolvido com base no padrão H.264/AVC. Com base no HEVC, experts desenvolveram o *3D-High Efficiency Video Coding* (3D-HEVC) [4]. O 3D-HEVC foi desenvolvido visando codificar eficientemente vídeos 3D, levando em conta as características específicas deste tipo de aplicação. A complexidade computacional apontada no HEVC cresce significativamente quando considerado o processo de codificação de vídeos 3D, visto que o número de imagens processadas para um mesmo instante de tempo (vistas) é maior. Além das redundâncias exploradas pelo codificador convencional, o 3D-HEVC também explora redundâncias entre imagens pertencentes ao mesmo instante de tempo, elevando ainda mais a complexidade computacional do 3D-HEVC.

Somando-se a isto, o 3D-HEVC introduziu o conceito de mapas de profundidade (*depth maps*) [4] na codificação de vídeos, que são imagens monocromáticas onde informações de distância dos objetos da cena em relação à câmera são armazenadas. Mapas de profundidade são essenciais no processo de síntese de vistas, o qual, a partir de dois quadros de textura (imagem convencional) capturados em posições diferentes (duas vistas), juntamente com seus respectivos mapas de profundidade, possibilitam a geração de uma vista intermediária entre as duas câmeras –denominada vista sintética [4]. Desta forma, além da complexidade adicional em relação ao HEVC, devido a exploração das redundâncias entre as imagens capturadas durante o mesmo instante de tempo, o codificador 3D-HEVC também apresenta um acréscimo na complexidade ao processar mapas de profundidade.

Tendo isto em vista, justifica-se a necessidade de estratégias inteligentes que busquem reduzir a complexidade computacional do processo de codificação de vídeos do 3D-HEVC. Estas estratégias não devem gerar um *bitstream* incompatível com o padrão, possibilitando a integração da solução em dispositivos que capturem e codifiquem o vídeo de forma compatível com o 3D-HEVC. Além disso, estas soluções de redução de complexidade não podem apresentar impacto significativo na eficiência da codificação, visando não comprometer a qualidade de imagem (objetiva e subjetiva) e a taxa de compressão obtidas pelo codificador.

Destaca-se que estratégias propostas neste cenário geralmente são inseridas no módulo de decisão de modo (*Mode Decision* – MD) do codificador, onde as diversas possibilidades de particionamento de blocos em unidades de codificação, predição e transformada são avaliados; além dos diversos modos e ferramentas disponíveis.

Dentre diversas soluções propostas na literatura para reduzir a complexidade do processo de codificação para o 3D-HEVC, este trabalho apresenta uma avaliação do impacto proporcionado pela heurística *Early Skip/DIS* [5-] [6] na qualidade subjetiva de imagens sintetizadas. Além disso, uma discussão e comparação entre três métricas objetivas de qualidade de imagem também são apresentadas neste trabalho.

A heurística *Early Skip*/DIS visa testar, primeiramente, os modos *Skip* e DIS do codificador de mapas de profundidade do 3D-HEVC e, dado o custo de codificação que relaciona a taxa de bits e qualidade de imagem (*Rate Distortion* – RD) destes modos, é tomada a decisão de avaliar ou não as demais ferramentas de codificação. Dado que esta solução visa selecionar antecipadamente a utilização dos modos *Skip* e/ou DIS, ela é nomeada *Early Skip*/DIS. Os valores limiares (*thresholds*) que determinarão se os demais modos serão ou não avaliados são definidos durante a codificação dos primeiros quadros do vídeo, ou em intervalos pré-definidos, visando obter uma pré-determinada taxa de acerto. Assim, a definição destes valores é realizada de forma dinâmica e adaptativa para cada vídeo e configurações de codificação utilizadas, aprimorando a eficiência de redução de complexidade da solução *Early Skip*/DIS.

A Seção II descreve a forma e a metodologia empregada para realizar as avaliações subjetivas, além de descrever as métricas empregadas na avaliação objetiva dos vídeos. Resultados objetivos e subjetivos são apresentados e discutidos na Seção III. Por fim, a Seção IV conclui este trabalho.

# Configuração das Análises Realizadas

## Sequências de Vídeo Utilizadas

Os vídeos utilizados na análise subjetiva foram os recomendados nas condições comuns de teste (CCT) [7] di 3D-HEVC, sendo eles: *Balloons, Kendo, Newspaper –* resolução 1024×768; e *Undo\_Dancer, GhostTownFly (GTFly), Poznan\_Hall2, Poznan\_Street* – resolução 1920×1088.

Destaca-se que este trabalho avaliou dois *Target Hit Rates* (THR) da solução *Early Skip*/DIS [6]: 95 e 80. Assim, um conjunto de experimentos foi realizado para avaliar o THR95 e, posteriormente, outro conjunto de experimentos fora realizado para avaliar o THR 80. Desta forma, é possível perceber os espectadores que avaliaram o THR 95 não são exatamente os mesmos que avaliaram o THR 80, embora havendo uma intersecção entre ambos os grupos.

Por fim, é importante destacar que o par de imagens exibidas aos espectadores são sintetizadas (vistas virtuais) de acordo com o documento das CCT [7].

## Métricas de Avaliação Objetiva

Neste trabalho adotou-se três diferentes métricas objetivas para quantificar a qualidade dos vídeos processados: *Peak signal-to-noise ratio* (PSNR), *Structural similarity* (SSIM) e *Most Apparent Distortion* (MAD) [8]. Tendo em vista a familiaridade da comunidade de processamento de imagens com as duas primeiras métricas, este trabalho se reserva a brevemente descrever apenas a métrica MAD.

### Most Apparent Distortion

O algoritmo que calcula a métrica MAD considera que o sistema visual humano emprega diferentes estratégias para julgar a qualidade das imagens. Quando a imagem contém uma baixa distorção (quase-limiar), a imagem original é mais aparente, então o sistema visual humano tenta olhar além da imagem buscando as distorções (estratégia baseada em distorção). Quando a distorção é claramente perceptível, o sistema visual humano tenta olhar além das distorções, avaliando o objetivo da imagem (estratégia baseada em aparência).

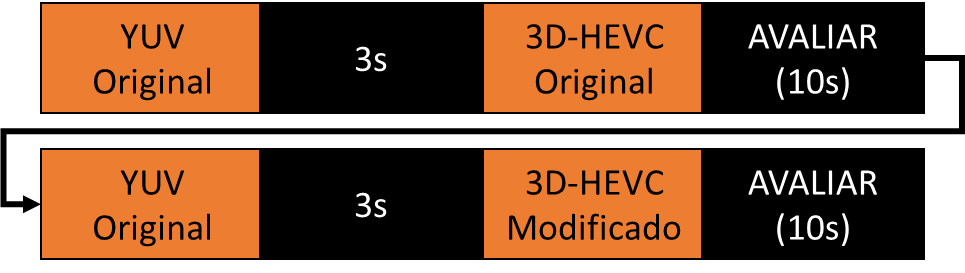
A métrica MAD explora estas duas estratégias de forma separada. Um “mascaramento local” de luminância e contraste é utilizado para estimar a distorção baseada em detecção para imagens de alta qualidade. E, em imagens de baixa qualidade, o MAD emprega um modelo estatístico local nas componentes de frequência. Destaca-se que ao comparar duas imagens idênticas, o resultado do MAD é igual a zero.

## Métrica de Avaliação Subjetiva

Na avaliação subjetiva das imagens, empregou-se a medida *Mean Opinion Score* (MOS), onde os usuários poderiam atribuir cinco notas após assistir duas sequências de vídeo, respondendo se a distorção percebida entre as imagens era:

1. Muito incômoda
2. Incômoda
3. Pouco incômoda
4. Perceptível, mas não incômoda
5. Imperceptível

A Figura 1 demonstra o fluxo utilizado no processo de avaliações dos vídeos. Primeiramente, exibiu-se aos avaliadores o par vídeo original (*YUV Original*) e vídeo codificado com o 3D-HEVC sem modificação alguma (*3D-HEVC Original*), então os usuários foram convidados a avaliar as distorções percebidas no vídeo. Após, um novo par de vídeos utilizando a mesma sequência foi exibida aos avaliadores. Primeiramente o vídeo original (*YUV Original*) e então o vídeo codificado com o 3D-HEVC com o *Early Skip/*DIS empregado (*3D-HEVC Modificado*). Após a exibição destes dois vídeos, os avaliadores foram novamente convidados a avaliar a distorção percebida entre estes.



1. Fluxo de exibição e avaliação das sequências de vídeo

## Ambiente das Avaliações Subjetivas

O ambiente onde fora realizada as avaliações subjetivas seguiu as recomendações descritas no documento [9]. O televisor 3D utilizado é da marca Samsung de 46 polegadas. A intensidade luminosa da sala entre o espectador e o televisor alternava entre 120 a 350 lux. Além disso, a intensidade luminosa atrás do televisor manteve-se sempre abaixo dos 20 luxes, conforme recomendado no documento [9].

# Resultados e Discussões

Esta seção apresenta e discute os resultados obtidos na avaliação objetiva e subjetiva dos vídeos utilizados pela heurística *Early Skip*/DIS.

É importante destacar que, em ambas avaliações (objetiva e subjetiva) utilizou-se o QP 35 nos experimentos. Além disso, embora não fixou-se o *bitrate* nas simulações, observou-se que em nenhum caso a diferença de taxa entre o vídeo codificado com o 3D-HEVC original e o modificado atingiu mais que 1%.

## Avaliação Objetiva

A Tabela I sumariza os resultados de distorção obtidos através das métricas PSNR, SSIM e MAD. É importante destacar que resultados da avaliação subjetiva também estão presentes na Tabela I (*ΔMOS*), entretanto estes serão comentados posteriormente. Além disso, visto que duas vistas são apresentadas ao telespectador (vídeo 3D), cada resultado apresentado na Tabela I é média obtida entre o PSNR, SSIM ou MAD obtido em para cada uma das vistas. Como é possível observar, o emprego da heurística *Early Skip*/DIS não acarreta em perdas significativas na qualidade objetiva de imagem (redução em PSNR e SSIM e acréscimo em MAD).

## Avaliação Subjetiva



Fig. 2 Gráfico *Boxplot* sumarizando resultados da avaliação subjetiva da heurística *Early Skip*/DIS utilizando (a) THR 95 e (b) THR 80.

Além dos resultados de qualidade objetiva de imagem, a Tabela I também apresenta resultados da avaliação subjetiva. Visto que cada usuário avaliou a degradação a qualidade proporcionada pelo 3D-HEVC original e também pelo 3D-HEVC implementando o *Early Skip*/DIS, a Tabela I apresenta a diferença entre a pontuação dada ao 3D-HEVC original e o modificado, conforme apresentado na equação (1).

|  |  |
| --- | --- |
| *ΔMOS = MOSmod – MOSorg* | (1) |

Nota-se que em alguns casos, principalmente considerando THR 95, há um *ΔMOS* positivo, indicando que a degradação na qualidade de imagem percebida pelos usuários é menor quando o vídeo é codificado através do 3D-HEVC com a heurística implementada.

Visto que as avaliações subjetivas foram realizadas com 30 avaliadores cada, a média dos *ΔMOS* para cada vídeo não traz muita informação acerca da dispersão dos dados. Tendo isto em vista, a Figura 2 apresenta um gráfico do tipo *boxplot*, visando sumarizar as avaliações de forma mais representativa.

Considerando o THR 95, é possível observar que exceto no vídeo *Kendo* os dados não são dispersos, ocorrendo apenas alguns *outliers*. Com isto, infere-se que apenas no vídeo *Kendo* os avaliadores tenderam a perceber uma maior degradação na qualidade de imagem do vídeo proporcionado pela heurística *Early Skip*/DIS quando comparado ao 3D-HEVC original.

Por outro lado, considerando o THR 80, é possível observar na maioria dos casos (exceto nos vídeos *GTFly* e *Poznan\_Street*) uma tendência dos avaliadores a perceberem uma maior degradação na qualidade da imagem. É importante destacar que o comportamento dos dados no vídeo *Kendo* é bem semelhante para ambos os casos, demonstrando que a variação do THR não causou significativa variação na qualidade de imagem desta sequência.

1. Resultados Obtidos

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Resolução** | **Vídeo** | **HEVC Original** | | | ***Early Skip*/DIS THR 95** | | | | ***Early Skip*/DIS THR 80** | | | |
| **PSNR** | **SSIM** | **MAD** | **PSNR** | **SSIM** | **MAD** | ***ΔMOS*** | **PSNR** | **SSIM** | **MAD** | ***ΔMOS*** |
| *1024*  *×*  *768* | *Balloons* | 40,8419 | 0,5537 | 24,6028 | 40,8415 | 0,5537 | 24,6445 | 0,0333 | 40,8121 | 0,5536 | 25,2057 | -0,0667 |
| *Kendo* | 41,2651 | 0,4860 | 26,6814 | 41,2617 | 0,4861 | 26,6073 | -0,2667 | 41,2169 | 0,4858 | 27,2538 | -0,2000 |
| *Newspaper* | 38,6174 | 0,6902 | 27,0885 | 38,6210 | 0,6901 | 27,0705 | -0,1000 | 38,5890 | 0,6900 | 27,5768 | -0,0333 |
| *1920*  *×*  *1088* | *Undo\_Dancer* | 36,0371 | 0,5417 | 40,7482 | 36,0232 | 0,5416 | 40,7640 | 0,1333 | 35,8282 | 0,5410 | 46,2936 | -0,1333 |
| *GT\_Fly* | 37,8582 | 0,3530 | 47,9941 | 37,8377 | 0,3529 | 48,4748 | -0,0333 | 37,7941 | 0,3526 | 49,7379 | 0,0333 |
| *Poznan\_Hall2* | 42,0491 | 0,2573 | 39,1993 | 41,9792 | 0,2571 | 41,3101 | 0,0333 | 41,9029 | 0,2570 | 43,7910 | -0,0667 |
| *Poznan\_Street* | 38,3882 | 0,5723 | 27,9870 | 38,3852 | 0,5722 | 28,1233 | 0,1000 | 38,3566 | 0,5720 | 28,6531 | 0,1667 |
| ***Média*** | | **39,2938** | **0,4934** | **33,4716** | **39,2784** | **0,4933** | **33,8563** | **-0,0143** | **39,2142** | **0,4931** | **35,5017** | **-0,0429** |

Na Tabela I é possível identificar que a sequência *Undo\_Dancer* apresentou a maior diferença absoluta entre as médias de *ΔMOS* em ambas as avaliações (0,2666). Tendo isto em vista, a Figura 3 apresenta uma comparação entre os histogramas de ambos os cenários para esta sequência de vídeo.

É possível perceber que em ambos THR’s, 18 avaliadores não perceberam uma degradação na qualidade de imagem proporcionado pela heurística *Early Skip*/DIS, quando comparada a degradação proporcionada pelo próprio 3D-HEVC. Entretanto, considerando THR 80 é possível perceber a tendência dos avaliadores em perceber maior impacto à qualidade de imagem do vídeo – comparado ao THR 90 – de acordo com o gráfico apresentado na Figura 2.

### Dados sobre os avaliadores

Além de avaliar as sequências de vídeo, os espectadores também responderam um breve questionário, informando sexo, idade, expertise em codificação de vídeos e escolaridade. A Figura 4 apresenta dois gráficos do tipo *boxplot* mostrando as informações dos avaliadores no que diz respeito às suas idades e conhecimento no tópico de codificação de vídeos. É importante ressaltar que os THR’s da heurística *Early Skip*/DIS não foram avaliados no mesmo momento, assim o conjunto de avaliadores não foi exatamente o mesmo em ambas as análises.

Como é possível perceber, os avaliadores da heurística utilizando THR 80 eram “mais velhos”, entretanto possuíam a mesma experiência no tópico de codificação de vídeos. Outro fator importante a ser mencionado é que, dos 30 avaliadores, 29 eram estudantes de graduação na análise do THR 95, havendo apenas um avaliador cursando pós-graduação. Por outro lado, na análise do THR 80, 25 avaliadores ainda não concluíram a graduação, 3 são pós-graduandos e 2 já possuem título de pós-graduação. Acredita-se que este fato justifica a tendência da população que avaliou o THR 80 possuir mais idade.

Embora os grupos de avaliadores que analisaram a heurística utilizando THR 95 e 80 não foram exatamente os mesmos, oito deles estiveram presentes em ambos os experimentos. Sendo assim, estes tiveram que avaliar duas vezes a degradação na qualidade de imagem proporcionada pelo 3D-HEVC sem nenhuma modificação. Tendo isto em vista, a Figura 5 apresenta um gráfico que demonstra a discrepância entre as notas dadas em suas análises em ambos os experimentos para o mesmo par de vídeos. Além disso, o gráfico também apresenta a discrepância na resposta à pergunta “experiência em codificação de vídeos” de cada avaliador.



Fig. 5 Discrepância nos votos



Fig. 3 Comparação entre os histogramas de ΔMOS das avaliações subjetivas utilizando THR 95 e THR 80.

Como é possível perceber, todos os avaliadores (AV) não foram 100% consistentes em suas análises – embora AV3 e AV7 tenham atribuído apenas um valor diferente entre ambas as análises. Além disso, é possível perceber que três avaliadores atribuíram notas diferentes quando questionados sobre sua expertise no tema codificação de vídeos. Estes fatos demonstram que a avaliação subjetiva, embora seja extremamente importante, sofre com algumas “aleatoriedades” por parte dos avaliadores.



Fig. 4 Comparação entre os histogramas de ΔMOS das avaliações subjetivas utilizando THR 95 e THR 80.

# Conclusões

Através dos resultados obtidos, foi possível perceber que a proposta *Early Skip*/DIS não apresentou uma variação significativa na qualidade objetiva da imagem. Além disso, considerando o THR 95, percebeu-se que não houve uma forte tendência dos avaliadores penalizarem a qualidade de imagem proporcionada pela heurística. Por outro lado, considerando THR 80, ficou evidenciado uma tendência nos avaliadores em classificarem darem notas inferiores para a qualidade de vídeo proporcionado pelo *Early Skip*/DIS. Por fim, verificou-se que os avaliadores tendem a ser incoerentes em suas respostas em avaliações subjetivas.

##### Referências

1. JCT-VC Editors, Recommendation ITU-T H.265 – High Efficiency Video Coding (ITU-T Rec.H.265), April 2013.
2. JVT Editors (T. Wiegand, G. Sullivan, A. Luthra), Draft ITU-T Recommendation and final draft international standard of joint video specification (ITU-T Rec.H.264|ISO/IEC 14496-10 AVC), 2003.
3. G. Correa, P. Assuncao, L. Cruz, and L. Agostini, “Computational complexity control for HEVC based on coding tree spatio-temporal correlation” in IEEE 20th International Conference on Electronics, Circuits and Systems (ICECS), Abu Dhabi, United Arab Emirates, 2013.
4. G. Tech; Y. Chen; K. Muller J. Ohm; A Vetro; Y Wang; “Overview of the Multiview and 3D Extensions of High Efficiency Video Coding” IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 35-49, Jan. 2016
5. R. Conceição; et al., “Complexity reduction for 3D-HEVC depth map coding based on *Early Skip* and *Early DIS* scheme” in 2016 IEEE Conference on Image Processing (ICIP), Phoenix, USA, 2016.
6. R. Conceição, “Early Skip/DIS: Uma Heurística para Redução de Complexidade no Codificador de Mapas de Profundidade do 3D-HEVC” Dissertação de Mestrado, Universidade Federal de Pelotas, Pelotas, Brasil, 2016.
7. K. Muller; A. Vetro; “Common Test Conditions of 3DV Core Experiments.” Document: JCT3c-G1100. San José, 2014.
8. P. Mohammadi; et al.; “Subjective and Objective Quality Assessment of Image: A Survey”. Disponível em: https://arxiv.org/ftp/arxiv/papers/1406/1406.7799.pdf
9. ITU-T; “Methodology for the subjective assessment of the quality of television pictures”. Document BT.500-13, 2012.
10. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

##### apêndice

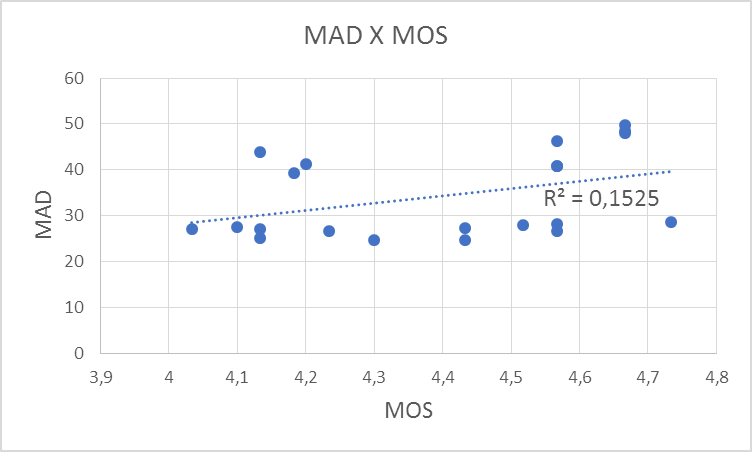
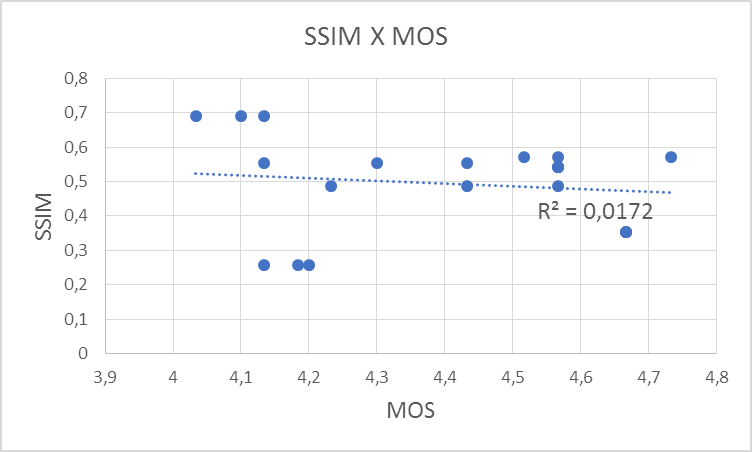
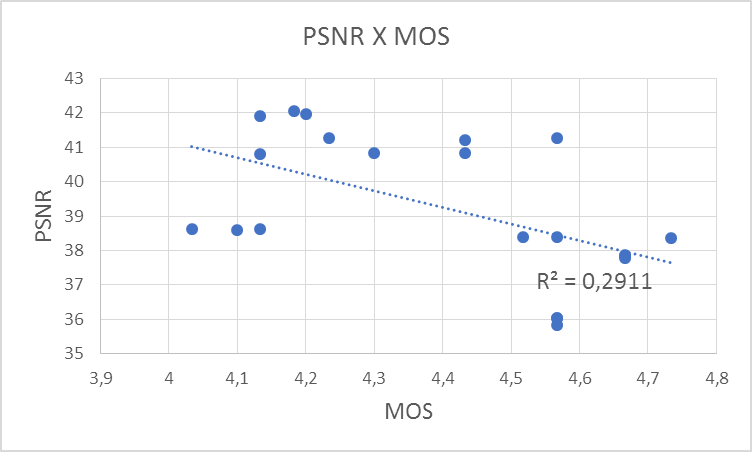


Fig. A1 Gráficos de dispersão relacionando métricas objetivas com o MOS.

Não há uma relação forte entre nenhuma das métricas objetivas com o MOS. Além disso, a linha de tendência sempre teve um sentido inverso ao esperado.



Fig. A2 Gráficos de dispersão relacionando idade e experiência em codificação de vídeo dos avaliadores com a média dos ΔMOS atribuídos.

Como é possível perceber, nem a a idade dos avaliadores, nem suas experiências em vídeo coding estão relacionados à nota atribuída na avaliação dos vídeos

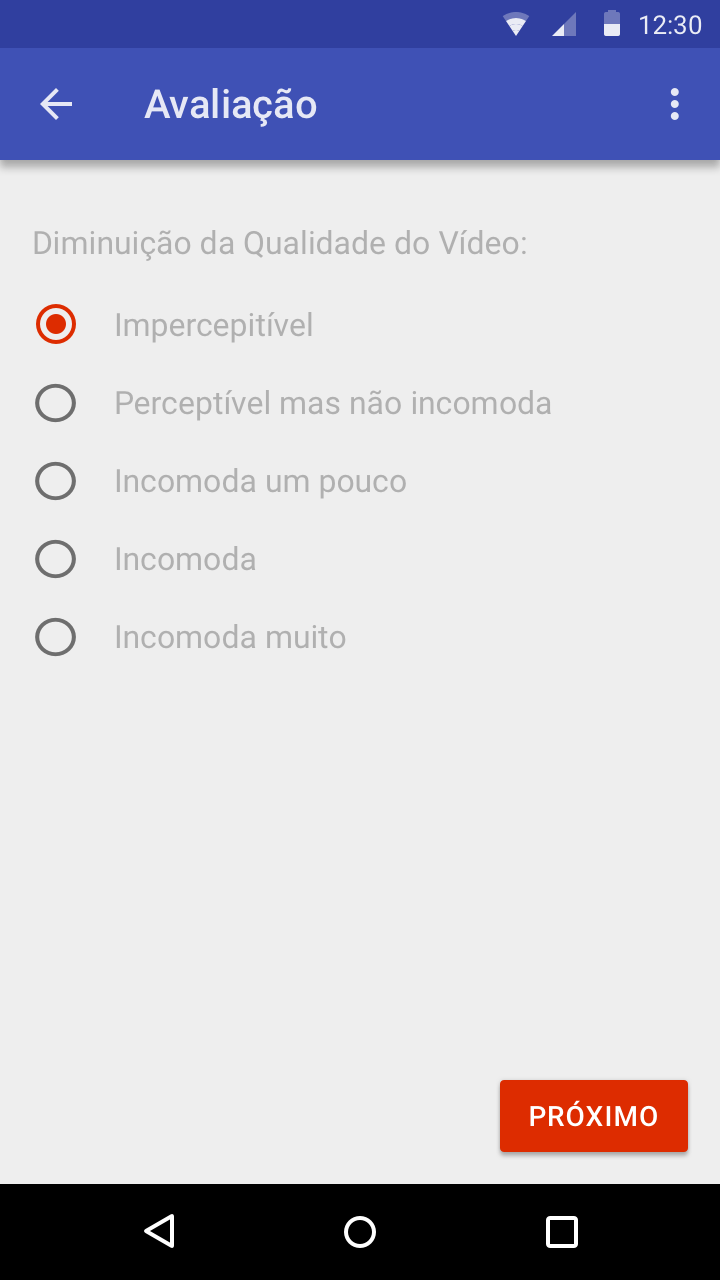
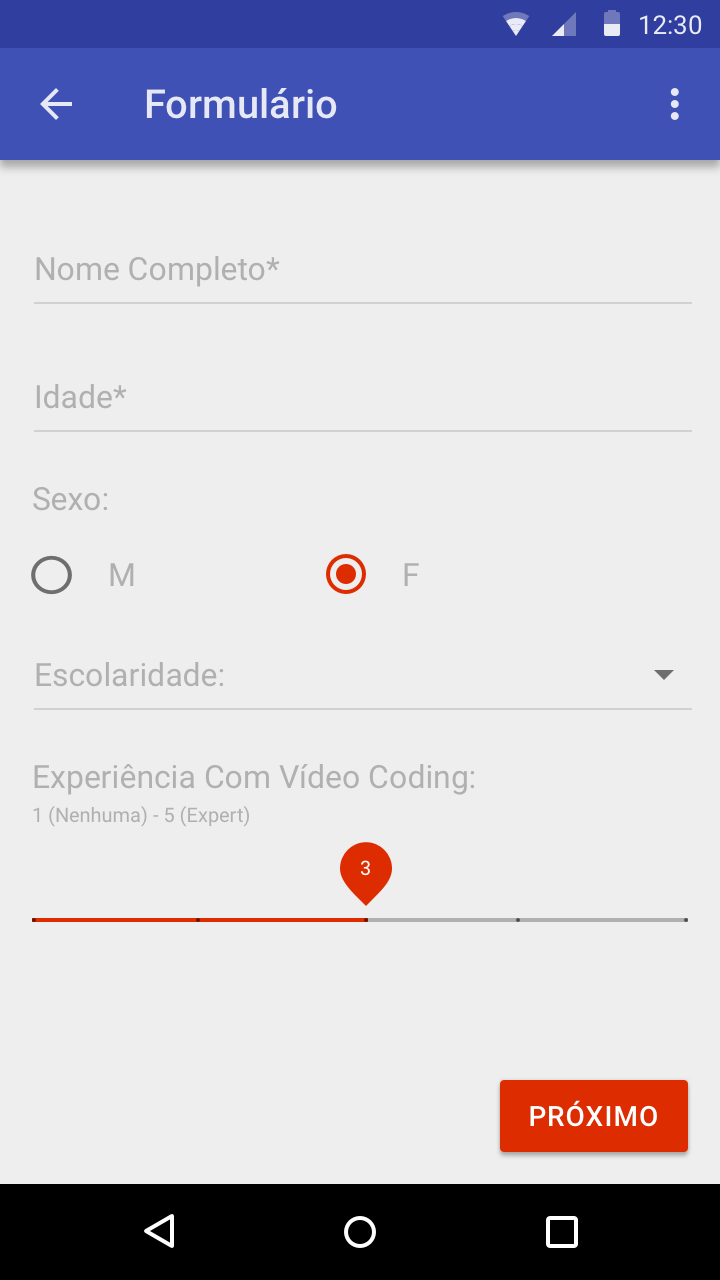


Fig. A4 *Mockups* do aplicativo

Cada usuário inseriu suas informações e avaliações em um aplicativo desenvolvido pelo grupo.



Fig. A3 Avaliador T.I.B realizando a avaliação subjetiva.