

Detecção e Reconhecimento de Faces Distorcidas por Artefatos de Compressão

*Fernando Rodrigues Trindade Ferreira, Carla Liberal Pagliari

Instituto Militar de Engenharia (IME)

Seção de Engenharia Elétrica (SE/3)

Praça General Tibúrcio, 80, 22290-270, Praia Vermelha, Rio de Janeiro, RJ, Brasil.

*fernando_rodrigues1990@hotmail.com

RESUMO: Os sistemas de reconhecimento de faces permitem diversas aplicações, tais como controle de acesso, identificação biométrica e vigilância. É apresentado um estudo sobre a resiliência dos detectores e reconhedores faciais perante faces distorcidas por artefatos de compressão. São utilizados os seguintes padrões de compressão JPEG, JPEG2000, H.264/AVC e HEVC. Para reconhecimento de faces, são estudados dois algoritmos: o LBP o PCA. Este trabalho propõe uma combinação das versões do detector Viola-Jones, Face e UpperBody, visando mitigar a ocorrência de falsos-positivos. Uma outra contribuição deste trabalho é o uso de sequências de imagens/vídeos, originais e codificadas, além do uso dos tradicionais bancos de faces para a detecção e reconhecimento facial. Os resultados mostram que a taxa de desempenho do LBP em relação ao PCA foi bem maior, uma vez que o LBP é mais robusto às alterações impostas pelos artefatos de codificação, uma vez que efetua a análise de texturas das imagens.

PALAVRAS-CHAVE: Codificação, reconhecimento de faces, detecção de faces, artefatos de codificação.

ABSTRACT: Face recognition systems allow several applications, such as access control, biometric identification and surveillance. A study is presented on the resilience of facial detectors and facial recognition methods to faces distorted by compression artifacts. The JPEG, JPEG2000, H.264 / AVC and HEVC compression standards are used. For face recognition, two algorithms are studied: LBP and PCA. This work proposes a combination of two versions of the Viola-Jones detector to mitigate the occurrence of false positives. Another contribution of this work is the use of original and coded sequences of images / videos, as well as the use of traditional face datasets for face detection and recognition. The results show that the LBP method outperforms the PCA method, as the LBP is more robust to the changes imposed by the coding artifacts since it analyzes the textures of the images.

KEYWORDS: Codification, face recognition, face detection, coding artifacts.

1. INTRODUÇÃO

A área de análise de faces pode ser dividida em diversas subáreas, como detecção/localização de face, reconhecimento de face, reconhecimento de expressões faciais e análise de poses [1]. A detecção de faces é frequentemente utilizada como uma etapa inicial para o reconhecimento facial. O reconhecimento de faces consiste em identificar um indivíduo por intermédio da análise de sua face, comparando-a com outras faces armazenadas em um banco de dados.

São inúmeras as aplicações para o reconhecimento facial, como identificar criminosos, terroristas em aeroportos e grandes centros urbanos ou um visitante não autorizado a entrar em um determinado local para alertar as autoridades competentes. Devido a isso, vem crescendo a relevância dada à pesquisa de sistemas inteligentes de identificação e reconhecimento de faces capazes de identificar indivíduos nas mais diversas situações, de forma que auxiliem com eficiência e agilidade o trabalho realizado pelas equipes e atuando principalmente em sistemas de segurança.

O armazenamento de vários ângulos de rostos em um banco de dados para realizar o reconhecimento facial demanda um elevado gasto de área de armazenamento. Para reduzir o volume de dados, além de reduzir a quantidade de dados a serem transmitidos e/ou armazenados [2], muitas bases de dados armazenam as imagens das faces em formato comprimido. A compressão, que pode ser com ou sem perdas, pode imputar artefatos, no caso de compressão com perdas. Tais artefatos podem modificar algumas características das faces, dificultando o desempenho das técnicas de detecção e reconhecimento facial.

O principal objetivo e contribuição deste trabalho é estudar a resiliência dos detectores e reconhedores faciais, perante imagens/vídeos comprimidos com os padrões de codificação JPEG [3], JPEG 2000 [4], H.264/MPEG-4 AVC [5] e HEVC [4]. O

algoritmo Viola-Jones [7] foi a técnica de detecção facial selecionada, enquanto as técnicas *Local Binay Pattern* (LBP) [8] e *Principal Components Analysis* (PCA) (referencia). O diagrama em blocos da Figura 1 ilustra todo o processo.

Uma das finalidades deste artigo é propor parâmetros de codificação, para codificadores com perdas, que não prejudiquem o desempenho de técnicas de detecção e reconhecimento facial. Uma outra contribuição é a proposta da combinação de duas versões do método de detecção facial selecionado para mitigar a ocorrência de falsos-positivos.

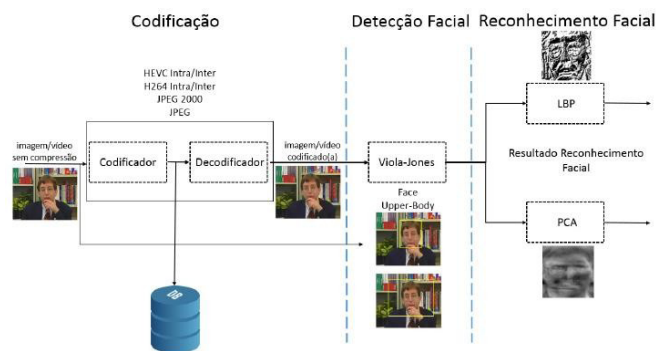


Fig 1: Diagrama em blocos do sistema de reconhecimento facial.

Dessa forma, são objetivos específicos da pesquisa:

Investigar o desempenho de técnicas de detecção facial com imagens/vídeos não comprimidos;

Investigar o desempenho e a robustez de técnicas de detecção facial com imagens/vídeos comprimidos com diferentes padrões de codificação e diferentes taxas de compressão;

Investigar o desempenho de técnicas de reconhecimento facial com imagens/vídeos não comprimidos;

Investigar o desempenho e a robustez de técnicas de reconhecimento facial com imagens/vídeos comprimidos com diferentes padrões de codificação e diferentes taxas de compressão;

Investigar o desempenho da combinação de dois métodos de detecção facial para mitigar a ocorrência de falsos-positivos.

Este trabalho está dividido em cinco seções além desta introdução, da seguinte maneira: a seção 2 apresenta os padrões de codificação de imagem/vídeo utilizados nesta dissertação, analisando o desempenho de cada padrão. Os fundamentos da detecção facial e técnicas selecionadas são apresentadas na seção 3. A seção 4 introduz o desafio do reconhecimento facial e as soluções empregadas. Na seção 5 são apresentadas as metodologias experimentais e os resultados obtidos. Já na seção 6 são apresentadas as conclusões.

2. CODIFICAÇÃO DE IMAGENS/VÍDEO

As técnicas de compressão podem ser classificadas como: sem perdas e com perdas. No primeiro caso, apesar de proporcionar uma taxa de compressão bem mais modesta do que as taxas das técnicas de compressão com perdas, a imagem reconstruída é idêntica à imagem original. No segundo caso, a imagem reconstruída é similar à imagem original, mas não idêntica, como no caso anterior.

Quando o objetivo da compressão é visualizar as imagens/vídeo, como é o caso de entretenimento, eventuais artefatos podem não comprometer a finalidade da aplicação. No entanto, os artefatos de codificação podem alterar o desempenho dos algoritmos de detecção e reconhecimento de faces. Os ganhos em taxa e distorção é que vão determinar o desempenho de um determinado sistema de compressão. Taxas de compressão muito elevadas tendem a gerar imagens com baixa qualidade (alta distorção).

A exploração das redundâncias espaciais com finalidade de redução do número de *bits* para representá-la é denominada de compressão intra-quadro. Desta forma, a informação que antes estava espalhada por todos os elementos da imagem fica concentrada em poucos elementos. O segundo passo é quantizar os novos elementos de forma a reduzir o número de bits necessário para representar a imagem agora codificada. O estágio seguinte, denominada de codificação por entropia, codifica os elementos quantizados para explorar ainda mais as redundâncias introduzidas pela quantização.

O uso de blocos para o processo de compressão tem a vantagem de reduzir o ônus computacional. Este tipo de técnica é chamada de codificação baseada em blocos, que vai explorar a similaridade dentro de um bloco. Ou seja, a imagem é dividida em blocos, em que os blocos é que vão ser transformados, quantizados e codificados separadamente.

Muitos padrões de codificação são somente para imagens estáticas, tais como o JPEG e o JPEG-2000. Entretanto, apenas as redundâncias espaciais foram exploradas para os vídeos codificados com estes padrões. Ou seja, os quadros foram codificados independentemente, sem explorar as redundâncias temporais exibidos na taxa de quadros pré-definida.

Um vídeo é uma sequência de imagens estáticas denominadas quadros, que representam cenas em movimento. Além das redundâncias dentro de cada quadro (intra-quadro), existem as similaridades entre quadros vizinhos (inter-quadros). Desta forma, os codificadores de vídeo empregam técnicas de predição de movimento para localizar em um quadro padrões similares existentes no quadro de referência. A exploração das similaridades temporais não é realizada *pixel a pixel* e sim bloco a bloco [9].

Os padrões de compressão utilizados neste artigo são o JPEG, o JPEG-2000, o H.264/MPEG-4 AVC e o HEVC. Este último é o atual estado-da-arte.

As imagens e seqüências de imagens (vídeos) utilizadas apresentam diferentes faces, em diferentes poses (frontal, perfil), com diferentes expressões e capturadas em diferentes condições de iluminação. Tal variedade visa desafiar os métodos de detecção e reconhecimento facial.

3. DETECÇÃO FACIAL

O detector facial tem a finalidade de informar se uma imagem de tamanho arbitrário contém faces humanas, e em caso afirmativo, onde estas estão localizadas no interior da imagem. As técnicas de detecção de faces podem apresentar diferentes abordagens, conforme mencionado por [1]:

Poses: As faces podem variar conforme o perfil da pessoa diante da câmera;

Expressão Facial: Apresenta influência diretamente nos modos de como o algoritmo pode determinar a classificação;

Presença de Elementos externos: Uma pessoa pode apresentar barba, bigode, cabelos grandes, cabelos curtos, óculos, chapéu, boné, variações na tonalidade da pele, tamanho diferente de face em imagens distintas, variações de luminosidade do ambiente, etc.

3.1 Algoritmo de detecção de faces Viola-Jones

O trabalho de Viola-Jones [7] estabeleceu o uso de filtros de Haar para representar a imagem em um espaço de características de Haar (*Haar features*), usando o conceito de “imagem integral”. Esta representação da imagem integral permite que as características usadas pelo detector sejam processadas através de uma única passagem pela imagem. Uma segunda contribuição foi o uso de um classificador simples e eficiente baseado em *Boosting*, o *AdaBoost* [7], que seleciona um pequeno conjunto de características visuais relevantes a partir de um conjunto muito maior de características. A velocidade de processamento é garantida pela terceira contribuição do trabalho de Viola-Jones, que combina classificadores em cascata com foco nas regiões de interesse

O algoritmo é dividido em três partes:

- A criação de uma representação integral da imagem;
- Montagem de um classificador de aprendizado *Boosting* chamado de *AdaBoost*, capaz de selecionar as características relevantes;
- Criação de uma estrutura em árvore, chamada cascata de classificadores.

O método funciona através da quantificação de detalhes da imagem por *features*, que são exibidos na Figura 2, varrem a imagem em busca de detalhes. A varredura se sobrepõe às *features* em posições arbitrárias da imagem e o valor da *feature* é calculado em múltiplas posições, conforme mostra a Figura 3.

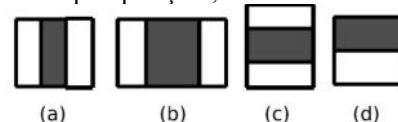


Fig 2: Features retangulares, extraído de [7].



Fig 3: Varredura da imagem através das *features* retangulares, extraído de [7].

Conforme a Figura 3, o método de varredura, em que as *features* são dadas pelo somatório dos *pixels* que se situam dentro dos retângulos brancos, e são subtraídos da soma dos *pixels* em retângulos pretos [7]. Então, esse resultado irá representar o valor encontrado pelas *features* para determinada região. Outros métodos, baseados do trabalho de Viola-Jonas, adicionaram outros conjuntos de *features* com a finalidade de produzir métodos mais robustos à variação de poses.

O método upper-body de detecção facial proposto por [10] consiste basicamente em acrescentar novas características utilizadas no detector, no qual as diferenças estão ponderadas de acordo com as representações das imagens integrais sobre as subregiões retangulares. A Figura 4 apresenta um conjunto de vários tipos de *features* disponíveis, que são os *features* originais propostos por Viola-Jones rotacionados para atender a região do dorso (*UpperBody*).

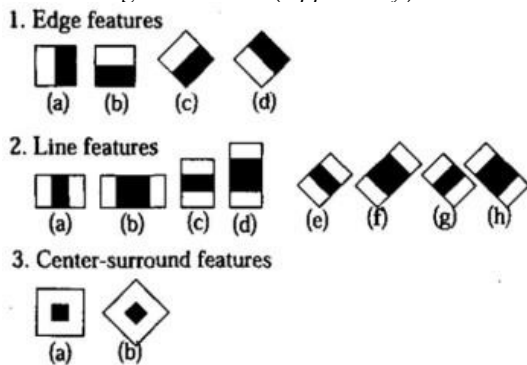


Fig 4: Features usados pelo UpperBody, extraído de [10].

A Figura 5 exhibe um exemplo de como o método Upper-Body elimina (Figura 5 (b)) um falso positivo (Figura 5(a)) obtido pelo método Viola-Jones, que detectou como face a região da boca até parte da gravata do apresentador.



Fig 5: Exemplos de classificação do detecto facial em relação ao contexto local (a) e centrada na face (b), respectivamente no quadro 18 do vídeo News.

4. RECONHECIMENTO DE FACES

O reconhecimento facial é uma técnica que visa identificar padrões em características faciais, tais como, distância entre os olhos e o formato dos olhos, do nariz e da boca, por exemplo. Nesta etapa é necessário estabelecer um conjunto de imagens de treinamento e um outro conjunto de imagens para teste. O primeiro conjunto, como o nome já implica, serve para treinar o algoritmo de acordo com o que estamos procurando relacionar. Enquanto que o segundo conjunto serve para verificar se os relacionamentos são válidos.

Os trabalhos [11], [12] e [13] apresentam uma revisão dos diferentes métodos de reconhecimento facial, onde o leitor pode encontrar comparações entre os diferentes métodos. O método LBP foi escolhido por apresentar altas taxas de reconhecimento,

bem como ter baixo ônus computacional.

4.1 Análise de componentes principais - PCA

Consiste em uma técnica utilizando ferramentas matemáticas para a redução da dimensionalidade das características faciais, cujo objetivo é extrair componentes principais do espaço de busca. A ideia é determinar se uma imagem teste pertence a uma determinada classe. Isto pode ser feito calculando-se a distância entre a imagem teste e os conjuntos de imagens de treino.

A Análise das Componentes Principais (Principal Component Analysis - PCA) mapeia os tons de cinza das imagens em um novo conjunto de variáveis aleatórias. Esta transformação linear visa desconectar as variáveis, reduzindo a sua dimensionalidade. A ideia de compressão ou de redução de dimensionalidade, utilizando o PCA tem como base a ideia de que projeções lineares são capazes de capturar o essencial da informação envolvida.

4.2 Padrão binário local - LBP

De acordo com o trabalho produzido pelo [14], este é um método não-paramétrico que trata as estruturas locais das imagens de forma eficiente, comparando cada *pixel* com seus *pixels* vizinhos. Uma das grandes virtudes deste método está na sua resistência a variações de intensidades de iluminação, escala, pose, ruídos e em sua simplicidade computacional [15].

O LBP do inglês (Local Binary Pattern, em português Padrão Binário Local [13] é um método não-paramétrico que trata as estruturas locais das imagens de forma eficiente, comparando cada *pixel* com seus *pixels* vizinhos. Uma das grandes virtudes deste método está na sua resistência à variações de iluminação, escala, pose, ruído e em sua simplicidade computacional [15]. Este método extrai um padrão local relativo ao *pixel* central e utiliza comparação utilizando histogramas dessas regiões locais. Ele foi proposto inicialmente para análise de textura em imagens em escala de cinza. No entanto, também tem sido muito empregado para o reconhecimento facial em imagens [16].

As faces podem ser vistas como uma composição de pequenos padrões, que se encaixam no contexto do LBP. A codificação da textura é feita em torno de um *pixel* é calculada por meio de um valor binário, “0” ou “1”, que é atribuído a um raio ao redor do *pixel* central, conforme a Figura 6. Este método extrai um padrão local relativo ao *pixel* central e utiliza comparação utilizando histogramas dessas regiões locais.

A codificação da textura é feita em torno de um *pixel* é calculada por meio de um valor binário, “0” ou “1”, que é atribuído a um raio, que é atribuído a um raio R ao redor do *pixel* central, conforme Figura 6.

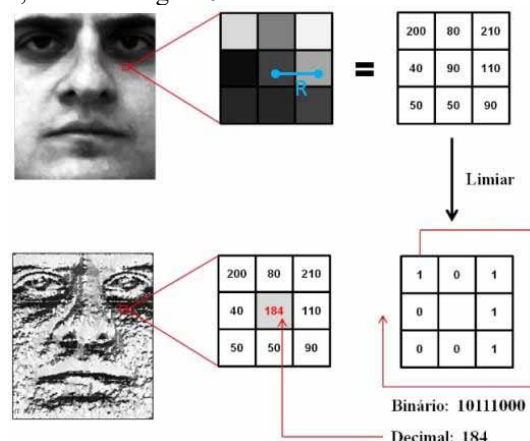


Fig 6: Operações referentes ao LBP, adaptado de [16].

Se o valor do pixel do vizinho for menor que o valor do pixel central, é atribuído o valor 1 àquela posição do pixel, caso contrário, será atribuído zero. A disposição dos zeros e uns garante nesse método a contagem de transições entre eles, fazendo dessa transição um padrão que é adotado pelo classificador. A disposição dos zeros e uns garante nesse método a contagem de transições entre eles, fazendo dessa transição um padrão que é adotado pelo classificador. Assim, detalhes praticamente imperceptíveis podem ser detectados, mas os detalhes que variam de acordo com a escala da imagem já seriam mais complicados de serem detectados, uma vez que a vizinhança apresenta raio fixo. No entanto, uma forma de contornar este problema seria implementando uma vizinhança de raio variável.

Nota-se que pela Figura 6 a matriz apresenta raio R igual a 1 e 8 pixels vizinhos. Com R=2, teríamos 16 pixels vizinhos, e assim por diante. Em [10] a tonalidade dos pixels vizinhos, em torno do pixel central, estabelecem uma relação entre si, no qual os valores nos tons de cinza a serem utilizados correspondem aos *pixels* delimitados por um círculo de raio R com uma quantidade de vizinhos P. Desta forma, a imagem gerada é dividida em pequenas regiões, conforme ilustrado no canto inferior esquerdo da Figura 6.

Uma importante modificação no cálculo do LBP para representação de face é a ideia de particionar a imagem de face em pequenos blocos (que podem ser sobrepostos ou não) e calcular o histograma LBP para cada bloco individualmente. Os blocos depois são concatenados em um único bloco, desta forma retendo a informação espacial. Portanto, a imagem de face é descrita em três diferentes níveis, conforme a Figura 7 nível de pixel: com o cálculo de cada operador LBP individualmente; nível regional: com o cálculo do histograma de cada bloco; e um nível global: com a concatenação de todos os histogramas de cada bloco; e uma nível global, com a concatenação de todos os histogramas de bloco. A Figura 7 mostra os três níveis de descrição que são abordados pelo LBP.

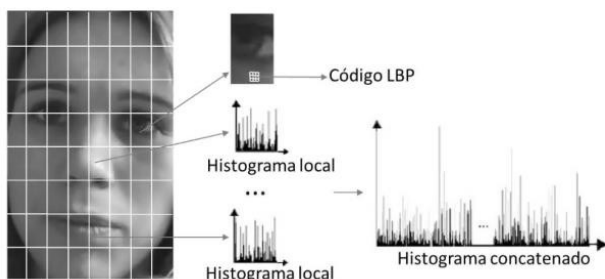


Fig 7: Descrição da face nos três diferentes níveis: nível de pixel (código LBP), nível regional (histograma local) e nível global (histograma concatenado), extraído de [17].

Conforme [18], para calcular a distância entre regiões correspondentes de duas imagens de faces, utilizam-se basicamente métricas que permitam explorar a relevância de algumas características faciais sobre outras. No processo de reconhecimento, é possível definir pesos específicos para cada região através dos histogramas que foram gerados a partir da Figura 7, como por exemplo, no trabalho feito por [16], foi adotado para fins de cálculos a distância Qui-Quadrado, conforme equação 1.

$$X_w^2(x,y) = \sum w_j \frac{(x_{i,j} - y_{i,j})^2}{x_{i,j} + y_{i,j}} \quad (1)$$

Onde x e y são as matrizes de histogramas a serem comparadas, i é um bin do histograma correspondente a j-ésima região (bloco da face, vide Figura 7) e w_j é o peso pré-defi-

nido para a região j. Contudo, este método não requer treinamento a cada adição de amostra na base de busca. Conforme [15], a utilização de múltiplas amostras por pessoa aumenta a probabilidade de reconhecimento. Este método LBP é conhecido como LBP de padrões uniformes, que está relacionado à quantidade de transições (de 0 para 1, ou de 1 para 0) que ocorre em cada padrão de textura. Neste trabalho, u define a quantidade de transições, tendo sido definido como uniformes, todos os padrões de textura com $u \leq 2$.

O LBP uniforme invariante a rotação é igualmente invariante contra transformações monotônicas da escala de cinza, sendo uma boa representação nos níveis de tons de cinza, portanto, da textura, na vizinhança de cada pixel [8].

A importância da diversidade na base de dados termos de etnia, idade e sexo, apresentando cenários do mundo real com diferentes ambientes, condições de iluminação e comportamentos humanos, garante uma análise do desempenho dos sistemas de detecção e reconhecimento de facial mais próxima do “mundo real”. Para a avaliação do método proposto, foram utilizados os vídeos extraídos da base de dados [19], tais como: Akiyo, Bowling, News, Crew, Deadline, Sign Irene, Silent e Foreman.

A Tabela 1 apresenta as médias das taxas de reconhecimento facial para os vídeos e imagens treinados com as suas respectivas versões originais. Onde, u2-R8 significa LBP uniforme (u2) com raio (R) igual a 1 e vizinhança (P) igual a 8. De maneira análoga, riu2-R8 significa LBP uniforme com R=1 e P=8 sendo invariante à rotação. As siglas u2-R16 e riu2-R16 apresentam R=2, P=16 na versão uniforme e na versão uniforme e invariante à rotação, respectivamente. Na Tabela 1, tanto o desempenho do detector, como o do reconhecedor facial não estão sendo testados usando as versões codificadas, que podem apresentar artefatos de compressão. Claramente, o desempenho das diferentes versões do LBP (u2R8, u2R16, riu2R8 e riu2R16) são superiores ao método PCA. Os métodos LBP u2R8 e u2R16 são dados pela representação uniforme do LBP com raio de 8 e 16 *pixels*, respectivamente. Já os métodos LBP riu2R8 e riu2R16 são as representações uniforme e invariante a rotação do LBP com raio de 8 e 16 *pixels*, respectivamente.

Tab 1: Taxa de reconhecimento facial.

	Treino Original				
	Reconhecimento Original (%)				
	LBP				PCA
	u2-R8	riu2-R8	u2-R16	riu2-R16	
Akiyo	85,411	86,196	86,584	91,629	56,931
Deadline	82,765	83,431	85,954	88,867	54,517
Silent	76,396	77,369	80,247	89,231	65,845
Bowling	29,356	33,168	35,513	41,946	23,650
Sign Irene	74,563	82,644	84,132	85,297	56,060
FACE96	88,587	86,541	90,233	92,000	64,000
FACE Pain	86,350	84,400	85,250	89,410	65,380

5. RESULTADOS OBTIDOS E ANÁLISE DAS MODIFICAÇÕES PROPOSTAS

Com o objetivo de mitigar a ocorrência de falsos-positivos, este trabalho propõe um detector facial híbrido que combina as versões Face e *UpperBody* do detector Viola-Jones. A Figura 8 apresenta o diagrama esquemático do método de

deteção de faces proposto. As transições 8(a), 8(c) e 8(d) mostram regiões da imagem/vídeo que foram classificadas como face pelo Viola-Jones. Pode-se perceber que a região gerada pelos classificadores do Viola-Jones Face 8(c) se refere à falso-positivo. Já as regiões 8(d) e 8(a) foram classificadas corretamente utilizando Viola-Jones Face e Viola-Jones *UpperBody*, respectivamente. A tarefa desenvolvida visa aplicar os métodos *UpperBody* e Face do Viola-Jones para a redução de falsos-positivos no processo de detecção facial. A transição 8(a) indica a classificação *UpperBody*, em seguida essas regiões são introduzidas à classificadores Face do método Viola-Jones dado pela transição 8(b). Essa proposta de redutor de falsos-positivos apesar de simples, reforça os classificadores, ao informar regiões que realmente, a probabilidade de ter encontrado face seja muito grande.

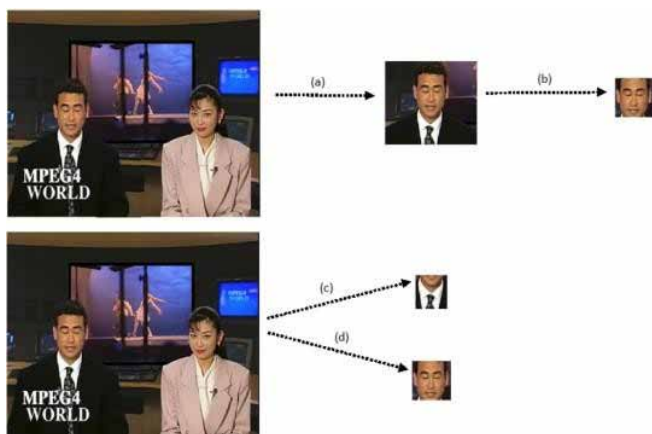


Fig 8: Diagrama esquemático da redução de regiões de busca pelo reconhecimento de padrões no processo de detecção de faces.

A Tabela 2 mostra a taxa de detecção de faces usando os métodos ViolaJones Face e Face+*UpperBody* unificados, apenas para vídeos originais, conforme o diagrama esquemático da Figura 8.

Tabela 2: Taxa de detecção de faces no método Viola-Jones.

	Taxa de Detecção (%)	
	Face + <i>UpperBody</i>	Face
	Original	Original
Akiyo	100,000	93,000
Deadline	98,840	92,667
Bowing	58,600	54,667
Silent	97,600	91,000
Sign Irene	98,880	88,888
Crew	79,471	77,157

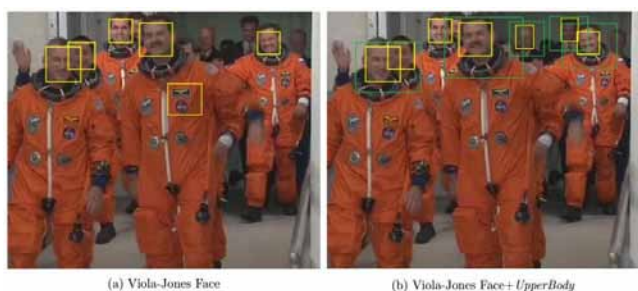


Fig 9: Comparativo no quadro 117 no vídeo Crew.

No vídeo Crew, mostrado na Figura 9, utilizando o Viola-Jones Face, foram detectadas 764 faces corretamente, e o número de falsos positivos foi de 136, já utilizando o Viola-Jones Face+*UpperBody* foram detectadas 841 faces corretamente e foram detectados 75 falsos positivos. Todos estes valores foram conferidos manualmente, visando garantir a veracidade das informações fornecidas. Analogamente, as ocorrências de falsos positivos foram eliminados nos quadros do vídeo Akiyo, Bowing e Sign Irene e Silent

5.1 Resultados do detector facial híbrido com imagens/vídeos codificados

Nesta seção são apresentados e discutidos os resultados obtidos pelo processo de detecção facial, empregando a proposta da combinação das versões Face e *UpperBody*, seguido do processo de reconhecimento facial aplicados a imagens e vídeos originais e codificados. É importante ressaltar que os artefatos de compressão imputados por cada codificador vão impactar no desempenho dos métodos de detecção e reconhecimento facial.

A tabela 3 apresenta as médias das taxas de detecção que foram obtidas nos conjuntos de teste de cada banco de vídeos e imagens usando a combinação Face e *UpperBody* do método Viola-Jones. Estes valores indicam as taxas média de acerto, levando em consideração os falsos positivos. A coluna original apresenta os vídeos/imagens originais não codificadas, enquanto que as colunas QP22, QP28, QP32 e QP40 apresentam os vídeos/imagens originais codificadas com os diferentes passos de quantização e diferentes codificadores (HEVC Intra, HEVC Inter, H.264 Intra, H.264 Inter, JPEG-2000 e JPEG).

Tab 3: Taxa de detecção de faces no método Viola-Jones.

		Taxa de Detecção (%)				
		Original	QP22	QP28	QP32	QP40
Crew	HEVC Intra	85,474	82,541	79,756	68,243	62,489
	HEVC Inter	85,474	81,265	78,741	67,648	61,368
	H264 Intra	85,474	78,133	75,692	65,962	59,330
	H264 inter	85,474	77,647	74,877	64,046	58,155
	JPEG 2000	85,474	73,114	71,004	61,423	55,699
	JPEG	85,474	70,333	67,385	60,222	53,988
Foreman	HEVC Intra	77,157	76,523	75,142	69,584	62,284
	HEVC Inter	77,157	75,844	74,978	67,262	60,837
	H264 Intra	77,157	75,158	72,521	65,113	57,661
	H264 inter	77,157	74,257	71,745	64,855	56,238
	JPEG 2000	77,157	72,698	69,866	61,423	54,347
	JPEG	77,157	70,283	67,789	58,327	52,411
Face96	HEVC Intra	92,864	88,667	84,511	78,165	71,667
	H264 Intra	92,864	87,667	82,933	75,773	69,881
	JPEG 2000	92,864	81,685	76,979	72,654	66,655
	JPEG	92,864	79,741	74,000	71,000	64,689
Face Pain	HEVC Intra	91,243	88,975	83,333	77,052	71,338
	H264 Intra	91,243	87,742	81,856	76,454	68,653
	JPEG 2000	91,243	80,123	76,442	71,333	64,644
	JPEG	91,243	77,327	72,966	68,977	59,254

Analisando os resultados obtidos pela simulação computacional do sistema de reconhecimento facial, na Tabela 4 foram utilizadas imagens/vídeos originais como treino e para

5.2 Resultados do reconhecedor facial híbrido com imagens/vídeos codificados

Tabela 4: Taxa de reconhecimento de faces no método Viola- Jones.

		Treino Original					Treino Original				
		Taxa de Reconhecimento QP22 (%)					Taxa de Reconhecimento QP32 (%)				
		LBP				PCA	LBP				PCA
		u2-R8	riu2-R8	u2-R16	riu2-R16		u2-R8	riu2-R8	u2-R16	riu2-R16	
Akiyo	HEVC Intra	81,328	74,977	83,897	87,344	57,690	52,165	43,403	37,403	55,424	24,203
	HEVC Inter	79,845	70,852	81,577	85,647	54,994	49,587	42,854	35,698	53,988	23,058
	H264 Intra	64,108	59,102	71,924	66,134	53,469	41,694	30,392	29,484	51,096	21,426
	H264 Inter	57,828	60,280	62,430	57,047	42,265	38,765	28,257	30,966	47,506	21,043
	JPEG	41,277	49,160	53,490	57,385	35,905	17,844	19,847	21,633	23,877	5,874
	JPEG 2000	50,148	53,285	54,470	58,592	37,073	29,738	21,536	30,088	29,936	7,586
		Treino Original					Treino Original				
		Taxa de Reconhecimento QP28 (%)					Taxa de Reconhecimento QP40 (%)				
		LBP				PCA	LBP				PCA
		u2-R8	riu2-R8	u2-R16	riu2-R16		u2-R8	riu2-R8	u2-R16	riu2-R16	
Akiyo	HEVC Intra	68,807	72,868	75,304	86,587	48,072	2,436	3,066	4,715	4,529	1,608
	HEVC Inter	65,756	70,922	71,569	82,541	46,776	1,587	2,841	4,682	3,560	1,169
	H264 Intra	54,239	53,618	59,360	59,825	42,556	0,000	1,729	1,768	2,761	0,000
	H264 Inter	53,982	56,069	55,190	56,451	37,357	0,000	1,004	1,217	1,883	0,000
	JPEG	25,845	28,111	30,453	28,865	20,411	0,000	0,000	0,000	0,000	0,000
	JPEG 2000	46,535	30,139	48,504	46,093	23,251	0,000	0,000	0,943	1,544	0,000

reconhecimento foram utilizadas imagens/vídeos codificados com diferentes passos de quantização (QP22, 28, 32 e 40). Entretanto, pode-se perceber que ao aumentar o passo de quantização ocorre uma degradação na imagem/vídeo devido os artefatos impostos pela codificação e como consequência, os métodos LBP e PCA vão perdendo sua eficiência. Para vídeos codificados com QP22, em média, a taxa de reconhecimento varia em torno de 73% (LBP) e 34,58% (PCA), enquanto que vídeos codificados com QP40 a taxa média cai para 4,3% (LBP) e 1,05% (PCA) aproximadamente. Estas análises foram feitas para a configuração riu2R16 que apresentou melhores resultados.

6 CONCLUSÃO

Este trabalho visou efetuar um estudo relacional entre os métodos de detecção e reconhecimento facial com os artefatos de codificação impostos por codificadores. Com isso, o paradigma de consumo de vídeo atual assenta sobre a internet e isso reforça a importância da utilização de codecs eficientes e tolerantes a erros e perda de pacotes. Também, a redução da taxa de bits oferece vantagens na transmissão, pois permite a proliferação de conteúdos em alta qualidade sem que seja necessário ocupar uma maior largura de banda, garantindo-se uma utilização mais eficiente do limitado espectro eletromagnético. Por isso foram utilizados codificadores HEVC, H.264, JPEG-2000 e JPEG com a finalidade de identificar artefatos de compressão que são impostos por esses codificadores e que afetam de maneira significativa os métodos de detecção e reconhecimento de faces.

As diferenças de textura entre os conjuntos de treino e teste refletem nos métodos de reconhecimento facial, pois o LBP começa a ficar perdido entre as suas variações (u2R0, riu2R8, u2R16 e riu2R16), uma vez que as transições entre o pixel central e seus vizinhos fica muito prejudicada pelos

artefatos de codificação. No entanto, os resultados mostrados nas tabelas, nos leva a concluir que caso queiramos reconhecer faces com QP22, temos que ter no conjunto de treino imagens/vídeos codificados com QP22 ou valores de QPs muito próximos deste valor.

40. Isto é explicado, em parte, pelo princípio de funcionamento do LBP que analisa a textura. Ou seja, quanto maior a similaridade entre os vídeos de treino e teste, maior é a probabilidade de sucesso no reconhecimento. Este princípio começa a falhar quando os artefatos de codificação degradam significativamente os vídeos de treino e/ou de teste.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] YANG, M.-H., KRIEGMAN, D. J. and AHUJA, N. Detecting faces in images: A survey. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 24(1):34–58, 2002.
- [2] ORTEGA, A. and RAMCHANDRAN, K. Rate-distortion methods for image and video compression. *Signal Processing Magazine*, IEEE, 15(6):23–50, 1998.
- [3] PART-1-ISO/IEC-10918-1. JPEG stands for Joint Photographic Experts Group. *Computer Vision and Image Understanding*, 1994.
- [4] ISO/IEC-15444-1. JPEG 2000 stands for Joint Photographic Experts Group. ITU, 2000.
- [5] ISO/IEC-14496-10-MPEG-4-PART10. MPEG-4 Part 10, Advanced Video Coding. ITU, 2000.
- [6] ISO/IEC-23008-2. MPEG-4 Part 10, Advanced Video Coding. ITU-T H.265, 2013.
- [7] VIOLA, P. and JONES, M. Rapid object detection using a boosted cascade of simple features. *Em Computer Vision and Pattern Recognition*, 2001. *CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, págs. 1–511. IEEE, 2001.
- [8] OJALA, T., PIETIKÄINEN, M. and MÄENPÄÄ, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 24(7):971–987, 2002a
- [9] RICHARDSON, I. E. H. 264 and MPEG-4 video compression: video coding for next-generation multimedia. John Wiley & Sons, 2004.
- [10] KRUPPA, H., CASTRILLON-SANTANA, M. and SCHIELE, B. Fast

- and robust face finding via local context. Em Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VSPETS), págs. 157–164, 2003.
- [11] LU, J., LIONG, V. E., ZHOU, X. and ZHOU, J. Learning compact binary face descriptor for face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(10):2041–2056, 2015.
- [12] SCHROFF, F., KALENICHENKO, D. and PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. Em *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, págs. 815–823, 2015
- [13] YANG, J., ZHANG, D., YONG, X. and YANG, J.-Y. Two-dimensional discriminant transform for face recognition. *Pattern recognition*, 38(7):1125–1129, 2005.
- [14] AHONEN, T., HADID, A. and PIETIKAINEN, M. Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12):2037–2041, 2006.
- [15] HUANG, D., SHAN, C., ARDABILIAN, M., WANG, Y. and CHEN, L. Local binary patterns and its application to facial image analysis: a survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *IEEE Transactions on*, 41(6):765–781, 2011.
- [16] DO AMARAL, V. and THOMAZ, C. E. Extração e Comparação de Características Locais e Globais para o Reconhecimento Automático de Imagens de Faces. Dissertação de Mestrado, Dissertação de Mestrado, Centro Universitário da FEI, SP, Brasil, 2011.
- [17] RODRIGUEZ, Y. . M. fce authentication using adaptive local binary pattern histograms. *Proceedings of European Conference on Computer Vision*, (1):321– 332, 2006.
- [18] BONFÁ, C. M. Um Sistema de Reconhecimento Facial em Vídeo Baseado em uma Implementação Multithread do Algoritmo TLD. Dissertação de Mestrado, PUCRio, 2013.
- [19] XIPH.ORG, VIDEO, T. M. collection of test sequences and clips for evaluating compression technology, 2003. URL <http://media.xiph.org/video/derf/>.