Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Detecção de Adulterações Espaciais e Temporais em Vídeos Digitais Utilizando o Algoritmo de Marca D'Água por Modulação do Índice de Quantização

Ronaldo Rigoni

Documentação apresentada como requisito parcial
para conclusão do Mestrado em Informática

Orientadora
Prof.ª Dr.ª Mylène Christine Queiroz de Farias

Brasília
2013

Universidade de Brasília — UnB
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Mestrado em Informática

Coordenador: Prof. Dr. Dr. Ricardo Pezzuol Jacobi

Banca examinadora composta por:

Prof.ª Dr.ª Mylène Christine Queiroz de Farias (Orientadora) — CIC/UnB
Prof. Dr. Bruno Luiggi Macchiavello Espinoza — CIC/UnB
Prof. Dr. José Gabriel Rodriguez Carneiro Gomes — UFRJ

CIP — Catalogação Internacional na Publicação

Rigoni, Ronaldo.

Detecção de Adulterações Espaciais e Temporais em Vídeos Digitais Utilizando o Algoritmo de Marca D'Água por Modulação do Índice de Quantização / Ronaldo Rigoni. Brasília : UnB, 2013.
94 p. : il. ; 29,5 cm.

Dissertação (Mestrado) — Universidade de Brasília, Brasília, 2013.

1. Detecções de adulterações em imagens, 2. detecções de adulterações em vídeos, 3. QIM, 4. marca d'água digital, 5. autenticação de imagens, 6. autenticação de vídeo, 7. ataques temporais

CDU 11/0054733

Endereço: Universidade de Brasília
Campus Universitário Darcy Ribeiro — Asa Norte
CEP 70910-900
Brasília–DF — Brasil

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Detecção de Adulterações Espaciais e Temporais em Vídeos Digitais Utilizando o Algoritmo de Marca D'Água por Modulação do Índice de Quantização

Ronaldo Rigoni

Documentação apresentada como requisito parcial
para conclusão do Mestrado em Informática

Prof.ª Dr.ª Mylène Christine Queiroz de Farias (Orientadora)
CIC/UnB

Prof. Dr. Bruno Luiggi Macchiavello Espinoza
CIC/UnB

Prof. Dr. José Gabriel Rodriguez Carneiro Gomes
UFRJ

Prof. Dr. Dr. Ricardo Pezzuol Jacobi
Coordenador do Mestrado em Informática

Brasília, 26 de junho de 2013

Dedicatória

A Deus, à minha mãe e à memória de meu pai, Neivor Rigoni.

Agradecimentos

Antes de tudo preciso dizer que meus agradecimentos não são formais. Eu não me reconheceria neles se assim fora. Quero agradecer a todas as pessoas que se fizeram presentes, que se preocuparam, que foram solidárias, que torceram por mim. De qualquer forma, todos os que realizam um trabalho de pesquisa sabem que não o fazem sozinhos, embora seja solitário o ato da leitura (em nossos tempos) e o de escrever. O resultado de nossos estudos foi possível apenas pela cooperação e pelo esforço de outros antes de nós. Albert Einstein certa vez falou: "Não descobri a teoria da relatividade apenas com um pensamento racional". Isso leva a questionar-me sobre quanto deste trabalho é meu e quanto é dos outros com quem convivi e com quem convivo, então chego a conclusão de que este trabalho não é só meu.

Gostaria de agradecer primeiramente a Deus! Sem Ele, jamais teria conseguido chegar até aqui. Há muito o que percorrer ainda, mas tenho certeza que a sua companhia me fortalece, me dá paz, saúde e sabedoria. Muito obrigado, Senhor!

Gostaria de agradecer à minha orientadora, a profa. Dra. Mylène Christine Queiroz de Farias, por todo o apoio e também por ter me ajudado durante todo o processo.

A minha mãe que, mesmo distante, sempre me deu forças. Obrigado pelo amor, pelo carinho e compreensão. A meu pai Neivor (in memoriam), por ter registrado em mim valores sem os quais este trabalho não teria sido possível. A meus irmãos, Ben-hur e Graziela, pelas brigas, pelo apoio e pelo incentivo.

Gostaria de agradecer imensamente aos meus amigos, pela compreensão, pelo apoio recebido. Em especial, ao meu amigo e colega, Pedro Garcia, sem o qual este trabalho não teria sido possível.

A Tupã, por abençoar seu filhos, semeando nas terras do sul a Illex paraguaiensis, pelo sorver de sua seiva, companheira nas madrugadas.

A Little Richard, Jerry Lee Levis, Chuck Berry pela trilha sonora durante esse tempo de estudos.

Por fim, gostaria de agradecer ao meu lado teimoso, perseverante-batalhador por mais esta conquista.

"Damnant quod non intelligunt".
Provérbio Latim.

"The greatest good we can do to others is not to provide them with our wealth but to reveal theirs.".
Louis Lavelle.

Resumo

A integridade e confiabilidade de conteúdos multimídia tornaram-se um desafio perante o fácil acesso às informações e à grande gama de softwares disponíveis para edição e editoração. A literatura classifica os métodos de adulterações de mídias digitais em duas classes: ativos e passivos. Os métodos passivos consistem na análise de catacterísticas típicas dos conteúdos adulterados, enquanto que os métodos ativos utilizam uma marca d'água ou uma assinatura digital para detecção de adulterações. Em particular, o uso de técnicas de marca d'água digital é uma das abordagems mais promissoras para detecção de adulterações em mídias digitais. Diversas técnicas de detecção de adulterações foram propostas na literatura. No entanto, todas as técnicas concentram-se na deteção de uma pequena gama de adulterações. O objetivo deste trabalho é o desenvolvimento de uma técnica de proteção contra adulterações em imagens e vídeos utilizando uma técnica de inserção de marca d'água digital com o algoritmo por modulação do índice de quantização (QIM). O sistema proposto é capaz de detectar adulterações locais, globais e temporais apresentando bom desempenho.

Palavras-chave: Detecções de adulterações em imagens, detecções de adulterações em vídeos, QIM, marca d'água digital, autenticação de imagens, autenticação de vídeo, ataques temporais

Abstract

The integrity of multimedia content, such as images, videos and audios, have become a challenge because of the easy access to information and the wide range of software available for edition. The literature classifies tampering of digital media methods in two types: active and passive. Passive methods analyse the intrinsic caracteristics of the media to detect tampered areas. The active methods use a watermarking algorithm or a digital signature for tampering detection. Using a watermarking algorithm is one of the most promising ways for the detection of tampering or any modification on the media. Several techniques are proposed in the literature for different types of digital tampering detection. However, most techniques concentrate on a small range of tampering detection. The goal of this work is to develop a technique for tampering protection in images and videos using a watermarking technique combined with a Quantization Index Modulation algorithm (QIM). The proposed system is capable of detecting local, global and temporal tampering and presents good performance.

Keywords: Video tampering detection,image tampering detection, QIM, watermark, image authentication, video authentication, temporal attacks

Sumário

Lista de Figuras

Lista de Tabelas

Glossário

BACM Blocking Artifacts Characteristics Matrix.

CSF Contrast Sensitive Function.

DC-QIM Distortion Compensated QIM.

DCT Discrete Cosine Transform.

DFT Discrete Fourier Transform.

DWT Discrete Wavelet Transform.

FFT Fast Fourier Transform.

IDWT Inverse Discrete Wavelet Transform.

JND Just Noticeable Difference.

LSB Least Significant Bit.

MPEG-2 Motion Picture Experts Group 2.

OTP One Time Pad.

QIM Quantization Index Modulation.

RBF Radial Basis Function.

ST-QIM Spread Transform Quantization Index Modulation.

SVD Single Value Decomposition.

SVH Sistema Visual Humano.

SVM Support Vector Machine.

Capítulo 1

Introdução

As tecnologias digitais permitem a criação de imagens de alta qualidade, animações, jogos, efeitos especiais e editoração de vídeos com um realismo incrível. As mídias digitais (imagens e vídeos) podem ser aprimoradas, comprimidas, transmitidas, re-codificadas em formatos diversos e exibida em uma variedade de dispositivos.

O formato de armazenamento e representação digital de mídias agrega vários benefícios. Dentre eles, a facilidade e flexibilidade de manipulação, edição, criação e armazenamento das informações de vídeos utilizando softwares e a distribuição em larga escala em formatos otimizados de acordo com o canal de transmissão.

A principal demanda por vídeos, imagens e áudios digitais se concentra no entretenimento, cinema, televisão digital, ensino a distância, marketing e Internet. O aumento na demanda é devido, também, ao crescimento da infraestrutura de rede disponível, satélites de transmissão direta de TV digital, popularização da Internet, aumento dos dispositivos móveis conectados, etc.

A Internet torna fácil a distribuição e compartilhamento de conteúdos digitais. No entanto, toda distribuição pode colocar em risco a segurança da mídia e a confiabilidade de seu conteúdo. Com isso, cópias ilegais, redistribuições, apropriações ou outras violações de direitos autorais podem ser efetuadas por um usuário não autorizado, causando grandes prejuízos, que podem alcançar a escala de bilhões de dólares [1, 2].

Cresce na Internet o uso de softwares para compartilhamento de arquivos via redes peer-to-peer (P2P), tais como *Kazaa*, *BitTorrent*, *eMule*, *SoulSeek* dentre outros. Alguns softwares utilizam criptografia em sua arquitetura de compartilhamento. Esta característica objetiva o anonimato e atrai um grande número de usuários, o que aumenta a segurança no compartilhamento e dificulta o rastreamento dos usuários infratores [3, 4]. A utilização desses softwares torna o controle de direitos autorais e da manipulação ilegal um grande desafio [5].

Com a facilidade de acesso a vídeos, a grande gama de softwares disponíveis para edição e o anonimato provido pelas ferramentas de compartilhamento, as adulterações se tornam um perigo real e eminente. Nos últimos anos, várias técnicas foram propostas na literatura com o objetivo de identificar adulterações e direitos autorais em vídeos.

Lin propôs um método que utiliza várias técnicas para detecção de uma pequena gama de adulterações temporais e globais, mas não detecta adulterações locais no quadro do vídeo [6]. Wang *et al.* propuseram um algoritmo para detecção de adulterações em vídeos com formato MPEG-2 [5]. Cross *et al.* também propuseram um método de detecção de adulterações em vídeos MPEG-2 capaz de detectar adulterações e perda de pacotes [7]. Infelizmente, este método não á capaz de localizar onde as adulterações ocorreram e pode ser utilizado apenas em vídeos formato MPEG-2. Do mesmo modo, Chen *et al.* propuseram uma técnica para autenticação e detecção de adulterações em vídeos MPEG-2 utilizando um algoritmo de *Compressive Sensing* [8]. Hou *et al.* propuseram um método capaz de detectar adulterações em blocos de 4×4 na área dos quadros do vídeo [9]. Este método não detecta adulterações temporais, tais como remoção de quadros e a alteração dos vetores de movimento.

Conforme exposto, as técnicas presentes na literatura para proteção de adulterações em vídeos são concentradas em apenas um tipo de adulteração ou formato do vídeo. Apenas a técnica proposta por Lin [6] aborda algumas adulterações globais e temporais. Deste modo, é de grande importância o desenvolvimento de ferramentas capazes de detectar adulterações locais, globais e temporais, independentemente do formato de codificação

do vídeo. Ou seja, capazes de detectar um número maior de adulterações de preferência utilizando uma única técnica.

O proposta deste trabalho é o desenvolvimento de uma algoritmo capaz de detectar adulterações de granularidade de pixel (locais), globais e temporais para vídeos, independente do formato de codificação do vídeo e capaz de estimar o tipo de adulterações temporais.

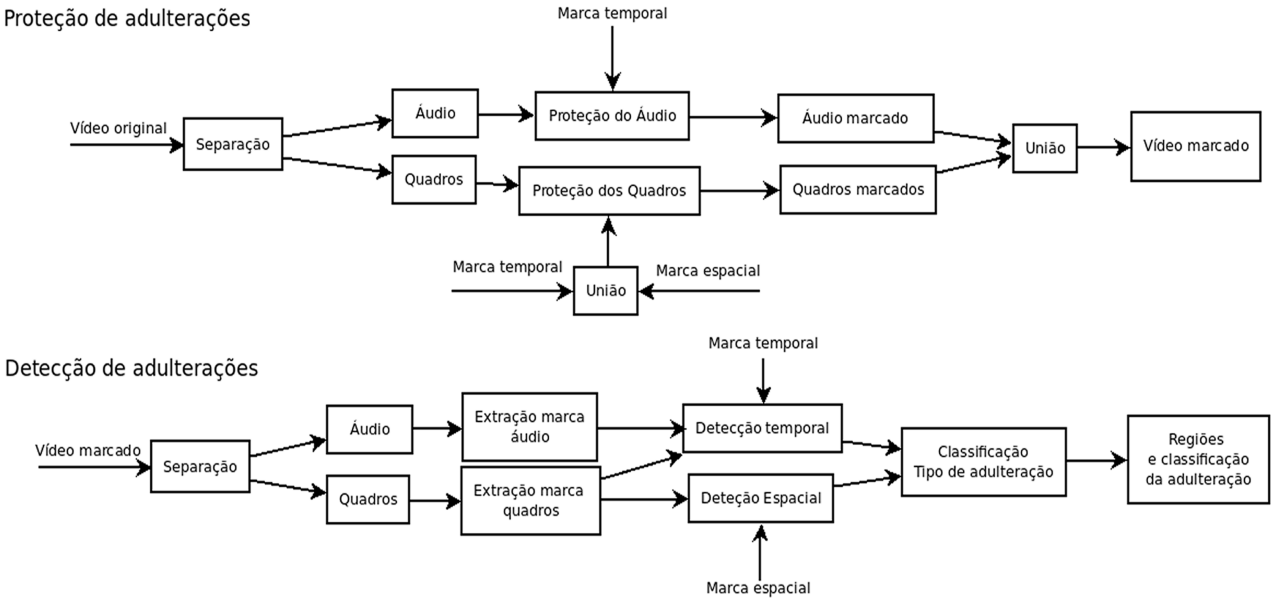


Figura 1.1: Diagrama de blocos de visão macro da técnica proposta.

A técnica proposta é dividida em duas partes: proteção de adulterações e detecção de adulterações, conforme ilustrado na Figura 1.1. Na proteção de conteúdo é utilizada uma técnica de marca d'água para cada quadro do vídeo, sendo que uma marca é inserida no domínio espacial e outra é inserida no domínio temporal. A marca temporal é inserida nos quadros e replicada para o canal de áudio. Esta replicação permite distinguir entre adulterações espaciais e temporais no processo de detecção de adulterações.

No processo de detecção de adulterações, as marcas são extraídas e comparadas com as marcas inseridas para identificar áreas adulteradas. Após a identificação das áreas adulteradas, a técnica classifica o tipo de adulteração temporal ocorrido.

Os resultados mostram que o algoritmo possui alta eficiência e acurácia na detecção de adulterações. Mais especificamente, o algoritmo apresenta baixa taxa de falsos-negativos, e consegue identificar áreas adulteradas espacialmente na granularidade de pixel, adulterações globais, temporais e classifica o tipo de ataque sofrido através da análise temporal de cada quadro.

Organização da Dissertação

Esta dissertação está dividida conforme segue. No Capítulo 2, apresentamos os tipos de adulterações mais comuns, suas classificações e métodos para suas detecções. No Capítulo 3, são apresentadas as principais classificações de marca d'água e suas principais características são apresentadas. No Capítulo 4, são apresentados a técnica proposta, a implementação, a proteção de conteúdo e a detecção de adulterações. No Capítulo 5, os resultados são discutidos. E finalmente, no Capítulo 6, concluímos o trabalho e discutimos os trabalhos futuros.

Capítulo 2

Adulterações de Mídias Digitais

Neste capítulo, damos uma introdução às principais adulterações em mídias digitais, fundamentos, suas características e tipos. Além disso, apresentamos o estado da arte dos tipos de adulterações e algoritmos.

2.1 Fundamentos

2.1.1 Proteção de Conteúdo

Um dos principais objetivos dos produtores ou proprietários de vídeos é assegurar seus direitos de propriedade intelectual. As regras de proteção de conteúdo são definidas pela lei vigente de cada país. O mecanismo de proteção fica sob responsabilidade dos proprietários dos vídeos.

A proteção de conteúdo é classificada em três tipos:

- *Controle de acesso* [10]: Protege o conteúdo da mídia para que ela esteja acessível apenas a usuários autorizados, mas não oferece proteção contra cópias não autorizadas ou distribuição ilegal.
- *Controle de cópias* [11]: Protege uma mídia contra criação de cópias indevidas. Na maioria das vezes, a cópia é feita por meio de softwares que conseguem "quebrar" o protocolo de criptografia. Em outros casos, as cópias são realizadas através de

gravação de uma exibição em um monitor ou cinema, conhecidas como "buraco analógico"[11].

- *O rastreamento de mídias* [6]: Garante que uma determinada mídia seja destinada a apenas um usuário ou a um grupo. Caso esta mídia seja encontrada em posse de alguém não autorizado, a identificação do comprador permitindo identificar o usuário infrator.

A vulnerabilidade aplicada pelo canal de distribuição de vídeos afetam os direitos autorais e a integridade; em consequência, a confiabilidade do seu conteúdo. A seguir, descrevemos cada um destes termos:

- *Direitos Autorais*: Grande parte dos mecanismos de proteção visam assegurar os interesses do proprietário. Em alguns casos, os direitos autorais são vendidos ou repassados a terceiros, que passam a gerenciá-los. O controle sobre direitos autorais engloba: o controle de cópias, o controle de distribuição, a integridade do conteúdo e a confiabilidade do conteúdo e da fonte [1].

- *Integridade*: Entende-se como integridade a garantia de que um vídeo quando acessado por um determinado usuário está representando exatamente a mesma informação que foi enviada pelo proprietário no momento da distribuição do vídeo na rede. A integridade garante ao usuário final que o vídeo não sofreu nenhuma adulteração, tanto no formato quanto no conteúdo. A criptografia é utilizada como ferramenta para assegurar integridade [1].

- *Confiabilidade*: A confiabilidade garante que apenas pessoas interessadas e devidamente autorizadas tenham acesso a determinado vídeo. Os acessos são concedidos a critérios definidos pelo proprietário tais como, licença de uso, compra de direitos sobre o vídeo ou livre acesso a determinado grupo. Algoritmos criptográficos são utilizados para obter a identificação do usuário, autenticação e autorização [1].

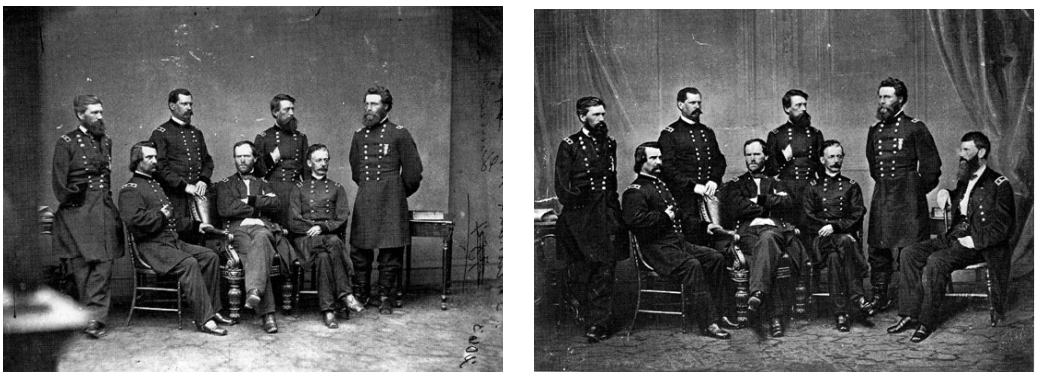
A ferramenta mais adequada para se garantir confiabilidade, integridade do conteúdo de um vídeo, controle de acesso, controle de cópias e rastreamento são providas pela

criptografia [12]. Segundo Sale [13], criptografia é o estudo dos princípios e técnicas pelas quais a informação pode ser transformada da sua forma original para outra ilegível, de forma que o conteúdo cifrado possa apenas ser revertido por seu destinatário (detentor de uma chave secreta).

A criptografia provê confiabilidade pelo uso de uma chave secreta que apenas o usuário autorizado possui para a recuperação do conteúdo original. Uma das limitações significativas da criptografia na proteção contra adulterações em vídeos é não oferecer um mecanismo de proteção após o vídeo ser decifrado [1]. Isso implica que o uso somente da criptografia como mecanismo de proteção não é suficiente para proteger e detectar adulterações em vídeos.

Na próximas seções abordaremos as principais adulterações em imagens e vídeos digitais e técnicas para assegurar sua proteção.

2.2 Adulterações em Imagens



(a) (b)

Figura 2.1: (a) Foto original sem a presença do General Blair. (b) Foto Adulterada onde Blair (mais à direita) foi adicionado por interesses militares. Fonte: *http://www.fourandsix.com.*

A fotografia, desde meados de 1800, foi alvo de manipulações fraudulentas, principalmente fotos de figuras públicas onde a adulteração teve o objetivo de ocultar ou distorcer informação nela contida. São exibidos nas Figuras 2.1, 2.2 e 2.3 alguns exemplos de manipulações em fotografias ocorridas ao longo da história.

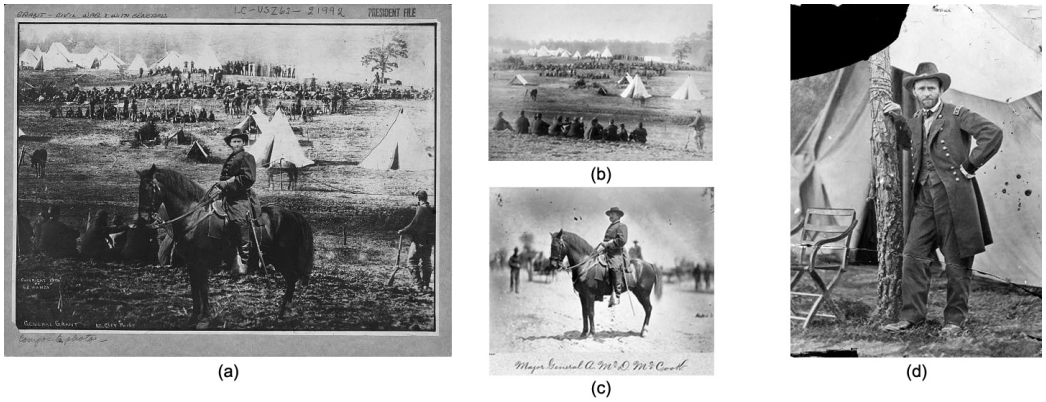


Figura 2.2: (a) Composição entre: (b) plano de fundo de um cenário da Guerra Civil Americana, (c) o corpo do Major Alexander M. McCook e (d) cabeça do General Ulysses S. Grant. Fonte: *www.fourandsix.com.*

Figura 2.3: Composição de uma foto de antílopes e de um trem, capturadas separadamente por Liu Weiqiang. Esta foto recebeu o prêmio de "Uma das fotos mais impressionantes do ano de 2006". Fonte: *www.fourandsix.com.*

As imagens na Figura 2.1(a) mostra uma foto do famoso fotógrafo Mathew Brandy onde o General Sherman esta posando junto com outros generais. Na Figura 2.1(b) o General Francis P. Blair (mais à direita) foi adicionado à fotografia original (Figura.2.1(a)).

A Figura 2.2(a) mostra uma composição da foto do General Ulysses S. Grant composta a partir de outras três fotos: a cabeça do General Grant (Figura.2.2(a)), o cavalo e o corpo do Major Alexander M. McCook (Figura.2.2(c)) e o plano de fundo de uma foto da Guerra Civil Americana (Figura. 2.2(b)).

Outro exemplo de adulteração maliciosa pode ser visto na Figura 2.3, que corresponde a uma fotografia de Liu Weiqiang, que recebeu o prêmio do Jornal *Daqing Evening News*

como uma das fotos mais impressionantes do ano de 2006. Entretanto, a foto é uma composição de uma foto de antílopes e uma foto de um trem. O próprio fotógrafo admitiu ter adulterado a imagem posteriormente a ter recebido o prêmio.

Os tipos mais comuns de adulterações presentes na literatura são adulteração por Composição, Copiar e Colar e Adulterações Locais. Nas próximas seções, descrevemos estes tipos de adulterações.

2.2.1 Composição

Adulteração por Composição é a união de uma ou mais imagens para a geração de outra imagem (composta). Para a criação de uma composição sem deixar rastros é necessário redimensionar, rotacionar ou apagar partes das imagens que estão sendo manipuladas para não criar bordas perceptíveis na região composta. Este processo gera regiões não homogêneas entre as bordas das áreas compostas, tornando possível a detecção. Encontram-se disponíveis na literatura vários algoritmos específicos para detecção de adulterações por composição em imagens [14], [15].

Cao *et al.* [15] propuseram um método para detecção de Composições utilizando a diferença da uniformidade de cor dos objetos presentes na composição. Neste método, primeiramente os objetos e o plano de fundo da imagem são segmentados. Em seguida, são extraídas a estatística dos histogramas locais e uma média do balanço de branco de cada objeto. A estatística de histogramas para cada objeto é calculada. Então, são gerados dois histogramas em três dimensões, sendo um referente ao plano de fundo e outro ao objeto. Para detectar inconsistências de luminância entre objetos é utilizada a distância entre os histogramas do objeto e do plano de fundo.

A media do balanço de branco dos objetos é calculada utilizando a média das diferenças de cores de todos os pixels do objeto segmentado. No entanto, se todos os pixels forem utilizados para calcular a média de diferença de cor de um objeto, podem ocorrer erros quando o plano de fundo for monocromático, pois ele alterará drasticamente a me-

dida calculada quando comparada àquela percebida pelo olho humano. Para evitar este problema, apenas alguns pixels são selecionados com um limiar de luminância mínima.

Ambas as características extraídas das regiões segmentadas são inseridas em um vetor estatístico de aprendizagem denominado Máquina de Vetores Suporte (SVM, do inglês Support Vector Machine). Uma Função de Base Radial (RBF, do inglês Radial Basis Function) é aplicada sobre o vetor SVM para classificar as áreas que possuem a maior discrepância das informações com relação ao restante do vetor. As áreas selecionadas pela função RBF que são maiores do que um limiar pré-definido correspondem às áreas adulteradas.

Os resultados apresentados por Cao *et al.* alcançaram resultados de cerca de 70.4% de acertos. Os experimentos envolveram imagens com composições realistas e composições propositalmente criadas.

Wang e Ping propuseram um método baseado em Decomposição de Valores Singulares (SVD, do inglês *Single Value Decomposition*) [14]. A técnica SVD decompõe uma imagem A no produto de duas outras matrizes ortonormais $U_{m\times m}$ e $V_{n\times n}$ e uma matriz diagonal $S_{m\times n}$. A diagonal principal da matriz $S_{m\times n}$ corresponde a um vetor I de valores singulares não negativos. Primeiramente, a matriz A é normalizada para média zero. Então, para obter um vetor Z_v para cada bloco, a SVD é aplicada em 50% dos blocos sobrepostos. Ou seja, $B = w \times w(w = 3, 4, ..., 16)$, no qual w corresponde ao tamanho do bloco. Assume-se que B é o número de blocos de tamanho $w \times w$.

Em seguida, são coletadas estatísticas do vetor I e de cada bloco resultante da decomposição SVD. Em particular, são coletadas a média dos números que não possuam valor singular em W e a média de números que possuam valores singulares em I iguais a zero. Os blocos que não mantiverem uma correlação ajustada de acordo com um limiar pré-definido com os blocos vizinhos correspondem a áreas possivelmente adulteradas.

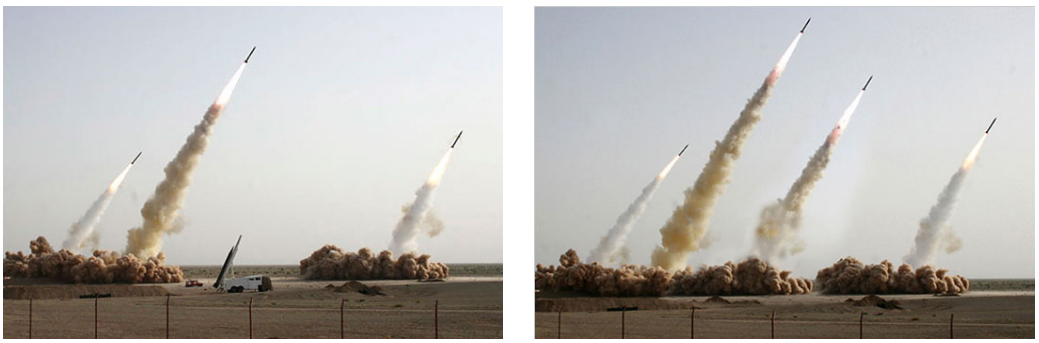
Os resultados apresentados por Wang *et al.* mostram que o algoritmo é eficiente para detectar adulterações por composição e redimensionamento, pois analisa os rastros deixados pelas dependências lineares das adulterações.

2.2.2 Copiar e Colar

"Copiar e Colar" é um dos tipos de adulterações mais comuns, consistindo em copiar e colar partes da imagem em outras áreas dela mesma, de forma a esconder áreas da imagem ou corrigir erros [16], [17]. Quando este tipo de adulteração é empregada, a sua detecção é difícil pois as regiões podem ser de qualquer tamanho e podem estar replicadas em vários locais da imagem, conforme exibido nas Figuras 2.4 e 2.5. A grande parte dos métodos presentes na literatura para detecção de "Copiar e Colar" são baseados na análise e comparação de blocos no domínio da frequência. Em outras palavras, a detecção é realizada através da busca por blocos com grande similaridade no domínio transformado.

(a) (b)

Figura 2.4: (a) Imagem original. (b) Adulteração da imagem utilizando o método Copiar e Colar no qual a casa e caminhão velho são ocultados por folhas copiadas e coladas da imagem original. Fonte *www.fourandsix.com*.



(a) (b)

Figura 2.5: (a) Imagem original capturada no momento do lançamento de um míssil das Forças Revolucionárias do Iran. (b) Imagem adulterada onde o segundo míssil da direita foi copiado e colado propositalmente para esconder um outro míssil em solo que não disparou. Fonte: *www.fourandsix.com*.

Wang *et al.* propuseram um método passivo para detectar ataques do tipo "Copiar e Colar" que é baseado em análises estatísticas dos blocos da imagem utilizando a Transformada Discreta Wavelet (DWT, do inglês *Discrete Wavelet Transform*) e a Transformada Discreta de Cosseno (DCT, do inglês *Discrete Cosine Transform*) [16]. Nesta técnica, a imagem é subdividida em blocos de 8×8 e são aplicadas as transformadas DCT e DWT em paralelo em cada bloco, obtendo-se duas matrizes de coeficientes b_m e c_m, respectivamente. São selecionados e multiplicados alguns blocos das matrizes, resultando em uma matriz $r_m = (b_m) \times (c_m)$ para cada bloco. Cada matriz r_m é comparada com todas as outras matrizes resultantes de cada bloco da imagem utilizando, um limiar pré-definido. As matrizes dos blocos que apresentam valor maior ao limiar são caracterizadas como regiões adulteradas. Os resultados apresentados por Wang e Ping demonstraram uma baixa taxa de falsos positivos, mas alto custo computacional.

Barni *et al.* propuseram um método para detecção de "Copiar e Colar" utilizando segmentação de blocos da transformada DCT. Esta técnica é capaz de identificar áreas adulteradas mesmo se varias áreas foram comprimidas utilizando taxas de compressão diferentes [17]. A técnica consiste em segmentar a imagem (em formato JPEG) em uma matriz de blocos de 8×8. E, então, cada bloco é comparado de forma independente aos outros blocos da imagem utilizando uma Matriz de Características de Blocos (BACM, do inglês *Blocking Artifacts Characteristics Matrix*). Os resultados exibidos por Barni *et al.* mostram que o algoritmo é robusto, porém com alto custo computacional. Além disso, o método é restrito a imagens em formato JPEG.

2.2.3 Adulterações Localizadas

Adulterações localizadas são pequenas operações locais nas imagens ou vídeos, como por exemplo borramento, recortes, distorção e suavização. Na maioria dos casos este tipo de adulteração é utilizada no entretenimento e publicidade, como por exemplo na criação de caricaturas e imagens com partes distorcidas.

Dois exemplos de Adulterações Locais são mostrados nas Figuras 2.6 e 2.7. A Figura 2.6 (b) é uma adulteração maliciosa onde os níveis de fumaça da foto original (Figura 2.6(a)) foram intensificados antes da publicação em um jornal. A Figura 2.7 mostra uma adulteração localizada, onde a boca do garoto propaganda foi adulterada para fins de entretenimento.

(a) (b)

Figura 2.6: (a) Foto capturada por Adnan Hajj logo após um ataque aéreo Israelense na capital da Líbia. (b) Versão manipulada da imagem com os níveis de fumaça intensificados publicada no Jornal The Reuters. Fonte: *www.fourandsix.com.*

Figura 2.7: Exemplo Adulteração Localizada, onde a boca do garoto propaganda foi distorcida para fins de entretenimento. Fonte: *www.fourandsix.com.*

Dentre os métodos disponíveis na literatura para detecção de Adulterações Locais, destaca-se citar a técnica proposta por Roy e Sun [18], que é baseada em uma função de *Hash One Way*. Funções de Hash One Way são funções que extraem uma quantidade fixa de bits de um arquivo de qualquer tamanho, sendo bastante utilizadas como assinatura ou identificador único [18]. No trabalho de Roy e Sun a imagem é dividida em n blocos de

8×8 e um Hash é gerado para cada um dos n blocos. Em seguida, o hash gerado de cada bloco é cifrado com uma chave privada e inserido em seu respectivo bloco. Posteriormente, o hash é removido da imagem adulterada e comparado com o hash gerado aplicando-se a função de hash em todos os blocos da imagem. O hash extraído é comparado com o hash gerado para o respectivo bloco.

Liu *et al.* propuseram um método de detecção de adulterações locais em imagens utilizando marca d'água frágil e permutação aleatória sob coeficientes da DWT [19]. Este método efetua uma permutação aleatória dos coeficientes de altas frequências da DWT gerando um vetor V. Posteriormente, uma marca d'água binária m é inserida nos índices do vetor V, escolhidos aleatoriamente de acordo com uma chave secreta k. Uma permutação inversa é aplicada no vetor V resultando numa marcação dos coeficientes de altas frequências. Em seguida, tira-se a inversa da DWT (IDWT, do inglês *Inverse Discrete Wavelet Transform*) para gerar a imagem marcada. No processo de detecção, a DWT é aplicada na imagem e as altas frequências são selecionadas. Então, uma permutação aleatória controlada pela chave k é aplicada nos coeficientes da DWT. Finalmente, a marca é extraída e comparada com a marca inserida. Os resultados apresentados por Liu *et al.* demonstram que o algoritmo possui alta capacidade de identificar áreas adulteradas por ataques locais devido a sensibilidade a ataques que afetam pequenas áreas.

2.3 Adulterações em Vídeos Digitais

A indústria produtora de vídeos é a maior interessada em proteçcão contra adulterações. O crescimento de fraudes envolvendo vídeos está criando um impacto na sociedade. Embora poucos vídeos foram adulterados e expostos até os dias atuais, eles já são suficientes para diminuir a confiabilidade dos vídeos disponíveis.

Nas Figuras 2.8, 2.9, 2.10, 2.11 e 2.12 são exibidos alguns quadros de vídeos adulterados de maior repercussão nos últimos tempos.

Na Figura 2.8, uma marca d'água da Televisão CBS foi inserida em uma transmissão ao vivo para esconder o símbolo da "NBC" que estava em exposição ao fundo. A técnica é utilizada também em eventos esportivos para exibir anúncios em painéis nas laterais de estádios. A Figura 2.9 mostra um quadro de um vídeo produzido pelo Comitê Nacional Republicano (RNC, do inglês *Republican National Committee*) onde é exibido um soldado originalmente assistindo um filme de desenho animado. A adulteração é semelhante à Figura 2.9, onde a imagem exibida pela televisão é substituída por uma propaganda política do Partido Democratas.

Na Figura 2.10 Mikhail Delyagin foi removido do vídeo após atacar o primeiro-ministro Russo Vladimir Putin no programa de televisão "O Povo Quer Saber". Apenas uma parte de Delyagin foi removida, ainda restam suas pernas e suas mãos visíveis à direita do homem que segura o microfone. Este é um exemplo de adulteração mal sucedida onde vestígios permanecem no vídeo, permitindo sua detecção.

A Figura 2.11 refere-se a uma adulteração de um vídeo onde uma águia supostamente captura um bebê e o carrega por alguns metros e, em seguida, o solta. Este vídeo foi facilmente classificado como fraude, pois a sombra da águia e do bebê são projetadas por outra fonte de iluminação. A superfície mais alta do objeto e da sombra devem convergir para um único ponto (linhas azuis) mas as sombras da águia e do bebê convergem para pontos diferentes. Logo, conclui-se que águia e bebê foram adicionados propositalmente no vídeo.

A Figura 2.12 mostra outro exemplo de fraude em um vídeo. Este vídeo causou polêmica na Internet. Trata-se de uma manipulação manual onde o homem supostamente atropelado foi adicionado ao vídeo. A detecção foi possível pois o ângulo da sombra projetada pelo homem é diferente da sombra projetada pelo resto dos objetos em cena. Só foi possível detectar a fraude após correções de ângulo ocasionadas pelas lentes e uma análise quadro a quadro do vídeo.

As adulterações apresentadas nas Figuras 2.11 e 2.12 demonstram a sofisticação dos ataques e a dificuldade de detecção.

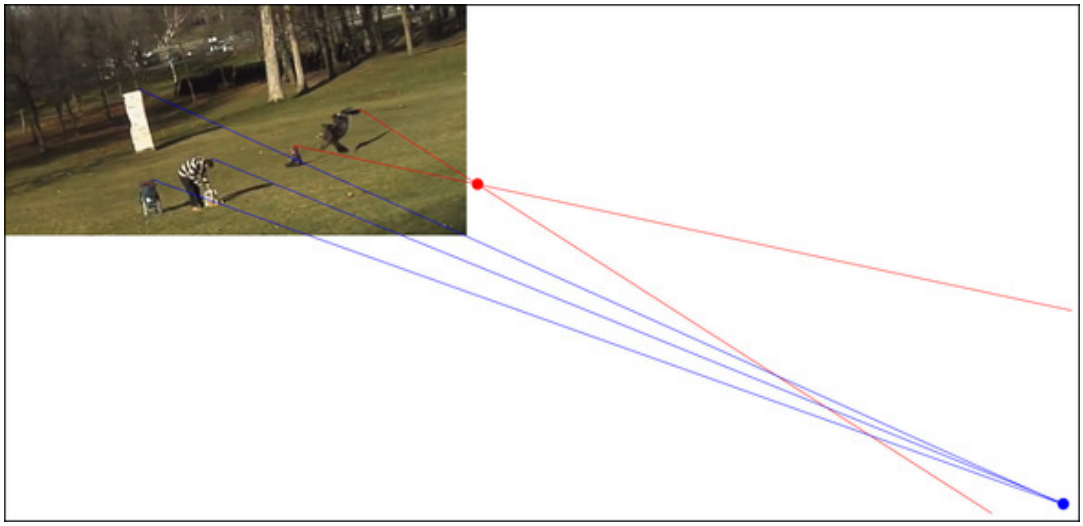
Figura 2.8: Marca d'água "CBS" inserido para esconder a marca "NBC" ao fundo. Fonte: *www.fourandsix.com*.

Figura 2.9: Vídeo político produzido pelo Comitê Nacional Republicano (RNC) onde originalmente o soldado estava assistindo um filme e a imagem foi adulterada para a legenda do Partido Democratas. Fonte: *www.fourandsix.com*.

Figura 2.10: Mikhail Delyagin foi removido do vídeo após atacar Vladimir Putin em um programa de televisão. Fonte: *www.fourandsix.com*.

Adulterações em vídeos são classificadas na literatura [20, 21, 22] em três grupos: adulterações locais, globais e temporais.

(a)



(b)

Figura 2.11: Dois quadros de um vídeo onde uma águia supostamente captura um bebê e o solta após alguns metros. Adulteração por composição onde vários objetos foram adicionados no vídeo. Fonte: *www.fourandsix.com*.

- Adulterações locais: Semelhante a imagens, conforme abordado no ínicio deste capítulo.
- Adulterações globais: São modificações que afetam a área do quadro como um todo. Como exemplo podemos citar o ajuste do brilho de um quadro, a conversão de formato do vídeo, redução das dimensões do vídeo, dentre outras.
- Adulterações temporais (Sincronização): Um dos tipos de adulterações que mais

(a)

(b) (c)

Figura 2.12: Vídeo de um acidente em uma rodovia da Rússia onde o homem supostamente atropelado foi manualmente adicionado ao vídeo. A detecção foi possível pois a sombra projetada pelo homem possuí ângulo diferente da iluminação do quadro. Fonte: *www.fourandsix.com.*

receberam atenção na comunidade científica foram adulterações temporais ou de Sincronização [6]. Um ataque temporal constitui na modificação da disposição dos quadros do vídeo ao longo da linha do tempo, isso inclui trocar quadros de posição, apagar quadros, ou movê-los para outra posição. Este tipo de ataque é bastante difícil de ser detectado pois sua intenção é confundir o algoritmo de detecção no processo de extração da marca d'água. Se o detector não conseguir detectar a

presença da marca d'água, nenhuma detecção temporal poderá ser efetuada. Esta é a maior vulnerabilidade no processo de detecção de adulterações utilizando marcas d'água digital [6].

Adulterações globais e locais possuem caracteristicas semelhantes às adulterações de imagens. A seguir definiremos mais detalhes de adulterações temporais.

Todas as adulterações apresentadas na Seção 2.2 (Composição, "Copiar e Colar" e Localizadas) são também aplicáveis a vídeos.

Encontram-se presentes na literatura também métodos de adulterações que modificam a ordem natural dos quadros ao longo do tempo (exibição). Os tipos mais comuns são:

- Duplicação de quadros: Ocorre quando quadros do vídeo são copiados e colados na posição subsequente do original. Isso aumenta o taxa de quadros por segundo, aumentando também o tamanho do mesmo. Este tipo de ataque visa dessincronizar a marca d'água inserida, dificultando sua remoção.

- Embaralhamento de quadros: Ocorre quando quadros do vídeo são trocados de posição. Este tipo de ataque é bastante difícil de ser detectado pois as substituições são feitas em um curto intervalo entre quadros para não ocasionar uma mudança visual que seja perceptível ao olho humano.

- Remoção de quadros: Quando quadros do vídeo são removidos de maneira aleatória. Este tipo de adulteração é de difícil detecção pois a informação da marca d'água inserida nestes quadros foi perdida.

- Inserção de quadros: Quando "novos" quadros são inseridos no vídeo. Os quadros podem ser do próprio vídeo ou de outras fontes (geralmente composições). Este tipo de adulteração é considerada a mais fácil de ser detectada pois os quadros inseridos na maioria das vezes não estão marcados e, geralmente são composições ou adulterações mais sofisticadas.

Adulterações temporais em vídeos são difíceis de detectar pois dessincronizam a marca d'água inserida. Isso exige que a marca d'água tenha redundância suficiente em outros quadros e que a perda não comprometa a detecção do resto do vídeo.

Diversas propostas estão disponíveis na literatura para detecção de adulterações no conteúdo de vídeos digitais. Na Tabela 2.1 é apresentado um resumo das principais referências abordadas neste trabalho. A seguir, detalhamos os principais métodos de detecção de adulterações em vídeos.

Lin propôs um método de detecção de ataques temporais que utiliza uma combinação de várias técnicas de marca d'água e redundância temporal [6]. Esta técnica utiliza redundância temporal e espacial, de modo a funcionar com vídeos ou uma sequência de imagens. A marca é inserida em uma série de blocos de alta correlação, característica que o torna robusto a operações geométricas tais como rotação, ampliação de imagens ou vídeos. Segundo os resultados apresentados por Lin, a acurácia é de 75% para imagens com fator de compressão JPEG de 90%. Além disso o método é capaz de detectar adulterações temporais de transposição de quadros, remoção e redução do número de quadros por segundo, mas não é capaz de detectar adulterações locais.

Wang *et al.* propuseram um algoritmo capaz de detectar vestígios de adulterações em vídeos sem utilização de técnicas de marca d'água digital [5]. Esta técnica assume que o vídeo contém certas características como: o ângulo de iluminação, aberrações cromáticas derivadas da captura do vídeo, etc. A premissa básica é que toda e qualquer modificação sobre o vídeo deixará vestígios que podem ser detectados. Entrelaçamento, de-entrelaçamento, dupla compressão, duplicação de quadros e re-projeção de vídeos (projetar o vídeo em um display e capturá-lo com outro dispositivo) não comprometem a eficiência do algoritmo.

Hou *et al.* propuseram um método para autenticação de vídeos utilizando marcas d'água digitais [9]. Este método consiste em armazenar um bit de verificação em blocos de coeficientes da DCT de tamanho 4×4. Conforme resultados apresentados por Hou *et al.*, o algoritmo é robusto e produz baixa degradação no vídeo marcado. Em contrapartida,

Autor	Temporais	Globais	Locais	Principal Técnica
Lin [6]	Sim	Sim	Não	Redundância da marca
Wang [5]	Sim	Não	Não	Método Passivo
Hou [9]	Não	Não	Sim	DCT
Cross [7]	Sim	Não	Não	MPEG-2 e Quadros "I"
Chen [8]	Sim	Não	Não	*Compressive Sensing*

Tabela 2.1: Tabela comparativa das principais técnicas de derecção de adulterações em vídeos presentes na literatura.

o método consegue apenas detectar adulterações com resolução mínima de tamanho 4×4. Além disso, o algoritmo não consegue detectar adulterações externas (temporais) a um quadro, tais como a remoção de quadros e a alteração dos vetores de movimento.

Cross *et al.* propuseram um método de detecção de adulterações em vídeos digitais em formato MPEG-2 [7]. Este método insere uma marca d'água em posições aleatórias dos quadros dos intra-quadros (quadros do tipo "I" usados no formado do MPEG-2 do vídeo). De acordo com os resultados apresentados por Cross *et al.*, o algoritmo se mostrou robusto e resistente à compressão MPEG-2, permitindo detectar perdas de pacotes e remoção de quadros. Em contrapartida, este algoritmo não é capaz identificar quais as áreas onde as adulterações ocorreram e funciona apenas para vídeos em formato MPEG-2.

Chen *et al.* propuseram uma técnica para autenticação de detecção de adulterações em vídeos no formato MPEG-2 [8]. Este método combina técnicas de marca d'água com *Compressive Sensing*. A marca d'água utilizada na autenticação e detecção de áreas adulteradas com base no conteúdo dos intra-quadros do vídeo, usando um algoritmo de *Compressive Sensing*. A marca d'água é inserida nos coeficientes de baixa frequências do domínio transformado dos intra-quadros através da DCT. Nos experimentos apresentados por Chen o algoritmo apresentou alta capacidade e acurácia de detecção.

Segundo Thangavel *et al.* uma das abordagens mais promissoras para detecção de adulterações em vídeo e imagens é o uso de técnicas de marca d'água digital [23]. No próximo capítulo apresentaremos as principais técnicas e algoritmos de marca d'água.

Capítulo 3

Marca d'água

Marca d'água digital é uma técnica de inserção de uma mensagem (marca) em um sinal tolerante a ruído, tais como áudios, vídeos e imagens. Normalmente é utilizado para identificar a propriedade dos direitos autorais do sinal marcado. Segundo Cox [24], é o processo de esconder uma informação digital em um sinal hospedeiro. A informação escondida pode, mas não necessita, conter uma relação com o sinal hospedeiro. Técnicas de inserção de marca d'água digital podem ser utilizadas para verificar a autenticidade e integridade de mídia ou para resgatar a identidade dos seus proprietários [25, 26].

O processo de inserção de uma marca introduz distorções na qualidade do vídeo original. Para reduzir esta degradação foram propostos modelos de inserção adaptativos, de modo a inserir a marca em regiões onde o Sistema Visual Humano (SVH) é menos sensível a alterações [27].

Um dos maiores desafios dos algoritmos de inserção de marca d'água digital é a resistência contra adulterações, ou seja a robustez da marca. Adulterações podem ocorrer de forma intencional ou não. As adulterações intencionais são realizadas com o objetivo de danificar ou remover a proteção do vídeo, enquanto que as adulterações não-intencionais correspondem a processamentos ou alterações sem intenção de danificar o conteúdo ou remover informações de proteção. Os tipos mais comuns de adulterações não-intencionais são: melhoria na qualidade da mídia, compressão, transcodificação para outro formato

para otimizar a largura de banda da rede em que será distribuído, etc.

A seguir descrevemos as principais aplicações das técnicas de marcas d'água.

3.1 Aplicações de Marca d'água

Os principais usos de marca d'água digital presentes na literatura [1, 28, 26, 29] são:

- Direitos autorais: Proteção contra violações de direitos autorais. O proprietário da mídia pode inserir uma informação que o identifique legalmente. Esta informação pode ser uma chave privada ou uma assinatura digital. A marca representará os direitos do proprietário sobre aquela mídia quando alguém violar seus direitos.

- Impressão digital: Quando o proprietário deseja rastrear a origem de cópias ilegais, ele pode usar uma técnica de impressão digital. Nesta técnica, o proprietário pode inserir diferentes marcas d'água no conteúdo de mídias destinadas a diferentes consumidores. A marca inserida contém uma informação que identifica o consumidor cuja mídia foi destinada. Desta forma, pode se identificar o comprador que violou os termos da licença de uso.

- Proteção contra cópias: A marca d'água inserida em uma mídia pode diretamente controlar dispositivos de gravação contra cópias ilegais [30]. Neste caso, a marca d'água armazena uma informação que, quando interpretada pelo dispositivo, irá determinar se a mídia pode ou não ser copiada.

- Monitoramento de difusão: No cenário comercial é comum o uso de marcas d'água para monitoramento de difusão. Emissoras de TV por assinatura ou satélite, emissoras de rádio, dentre outras, podem monitorar a difusão do sinal através do uso de marca d'água [29].

- Autenticação de mídias: Na autenticação de mídias, marcas d'água podem controlar não somente adulterações de uma mídia, mas também medir o quão profunda foi a adulteração [31].

- Indexação: Mecanismos de busca podem utilizar marcas d'água de forma imperceptível para indexar mídias.
- Segurança médica: Imagens médicas são comumente protegidas por marcas d'água. A proteção inclui dados do paciente, data do exame, dentre outros [29].
- Ocultação de dados (Esteganografia): Marcas d'água são também utilizadas para transferência de mensagens secretas. Nesta técnica a marca d'água utiliza a mídia como um canal hospedeiro, fazendo com que a mensagem seja transferida sem ser percebida [24].

3.2 Geração, Codificação e Classificação de Marca d'Água

3.2.1 Classificação

Marcas d'água são classificadas na literatura segundo sua visibilidade e robustez [28, 26]. A seguir detalhamos as classificações.

Uma marca pode ser classificada de acordo com sua percepção como sendo visível ou invisível. A marca d'água visível corresponde a uma informação inserida e perceptível a olho nu. Na maioria dos casos, corresponde a uma logomarca ou uma identificação dos proprietários. De outro modo, marcas invisíveis são inseridas de forma imperceptível ao olho humano.

Marcas d'água são classificadas quanto sua robustez em três tipos: frágil, semi-frágil e robusta.

- Frágil: Marcas d'água frágeis são projetadas para serem facilmente removíveis e corrompidas por qualquer processamento que a mídia sofrer. Este tipo de marca d'água é geralmente utilizado para checar a integridade e autenticidade de uma mídia. Em outras palavras, marcas d'água frágeis fornecem uma garantia de que uma mídia marcada não foi editada ou adulterada.

- Semi-frágil: Marcas d'água semi-frágeis também são utilizadas para autenticação de mídias. Estas possibilitam distinguir alterações que modificam uma imagem substancialmente das alterações que não modificam o conteúdo visualmente.
- Robustas: As marcas d'água robustas são projetadas para resistirem à maioria dos procedimentos de manipulação da mídia marcada. Por este motivo, elas são geralmente utilizadas para proteção de propriedade intelectual.

3.2.2 Codificação

Figura 3.1: Processo de codificação de Marca D'água.

A marca d'água pode ser um logotipo, uma imagem binária, uma assinatura digital, um padrão, uma sequência de bits, etc. Em alguns casos, a marca gerada leva em consideração o conteúdo da mídia, como uma mascara binária da imagem, um padrão de cores, etc.

O processo de codificação de uma marca d'água em uma mídia digital, conforme visto na Figura 3.1, possui três entradas: o sinal da mídia no formato original e uma marca d'água e, em alguns casos, uma chave secreta. É, então, efetuada a inserção da marca por um algoritmo específico, resultando na mídia marcada.

3.2.3 Decodificação

O processo de decodificação é separado em duas etapas. Primeiro, a marca d'água é extraída da mídia marcada no processo de extração. Em seguida, a autenticidade da marca d'água é verificada no processo de comparação.

Extração

A decodificação possui três entradas, conforme apresentado na Figura 3.2: a mídia marcada, a chave secreta e, em alguns casos, a mídia original. A saída é a marca d'água recuperada. É importante salientar que existem dois tipos de extrações. O primeiro tipo não utiliza a mídia original, também chamado de detecção sem referência. Enquanto que o segundo tipo faz uso da mídia original para extração da mara d'água, também conhecido como detecção com referência.

Figura 3.2: Processo de extração de Marca D'água.

Comparação

Figura 3.3: Processo de comparação de Marca D'água.

O processo de detecção , conforme demonstrado na Figura 3.3, consiste na comparação da marca d'água inserida com a marca recuperada. Esta comparação pode ser realizada de diversas formas, como por exemplo por comparação bit-a-bit ou usando uma função de correlação. Dependendo do tipo da função de comparação utilizado é possível determinar o quão severa foi a adulteração ou em quais as regiões da mídia ela ocorreu.

3.3 Algoritmos de Inserção de Marca D'Água

Vários algoritmos foram propostos para inserção de marcas d'água em imagens e vídeos. Nesta seção, descrevemos um conjunto representativo dos algoritmos de marca d'água disponíveis na literatura.

3.3.1 Espalhamento Espectral

Segundo Pickholtz *et al.*, espalhamento espectral é uma técnica de codificação para a transmissão digital de sinais [32]. Esta técnica foi desenvolvida originalmente pelos militares durante a Segunda Guerra Mundial com o objetivo de transformar as informações a serem transmitidas em um sinal parecido com um ruído radioelétrico. Dessa forma, conseguia-se camuflar uma mensagem no sinal, evitando que as forças inimigas que monitorassem o canal pudessem decodificar as mensagens transmitidas [32].

O espalhamento espectral é uma técnica de modulação em que a largura de banda usada para transmissão é muito maior que a banda mínima necessária para transmitir a informação. Dessa forma, a energia do sinal transmitido passa a ocupar uma banda muito maior do que a da informação.

A ideia básica de utilização do método de espalhamento espectral para inserção de marca d'água consiste em adicionar uma marca d'água pseudo-aleatória (de banda larga) ao sinal hospedeiro. Essa adição pode se dar no domínio do tempo ou no domínio de uma transformada; neste último caso, são comuns as transformadas DCT, FFT (a FFT é a versão rápida da DFT) e a DWT [24]. Em outras palavras, considere que x é o sinal original, que é transformado para o domínio da frequência utilizando a DFT, gerando um vetor y. Um vetor u, com valores aleatórios entre -1 e 1, é multiplicado por um vetor contendo a mensagem b. A mensagem b é binária, ou seja, os seus valores são iguais a 0 ou 1. O sinal marcado s é, então, obtido utilizando a seguinte equação:

$$s = y + b \cdot u. \tag{3.1}$$

A marca d'água é detectada pela comparação da marca extraída b' com a marca original b. A similaridade entre b' e b é medida pela função de correlação

$$sim(b, b') = \frac{b'.b}{\sqrt{b'.b}} > T, \tag{3.2}$$

no qual o valor da medida é comparado com um limiar T para determinar se a marca está presente no sinal. O limiar T possui valor arbitrário.

O algoritmo de marca d'água utilizando espalhamento espectral é bastante utilizado devido à sua robustez à compressão. Isto se deve ao fato da força da marca estar espalhada por todas as frequências da imagem. Dessa forma, quando as altas frequências são descartadas no processo de compressão, a marca não é perdida.

3.3.2 Modelos Perceptivos

Modelos perceptivos de inserção de marca d'água exploram as características do Sistema Visual Humano (SVH), visando maximizar a capacidade de inserção sem causar degradações visíveis à imagem ou ao vídeo. Com este objetivo, muitos algoritmos utilizam a métrica de Limiar Perceptível de Distorção (JND, do inglês, *Just Noticeable Distortion*) que mede os níveis de distorção que podem ser visíveis em 50% dos casos [33]. JND é geralmente utilizado como uma unidade para medida da distância (erro) entre dois sinais. Segundo Watson *et al.*, uma quantidade maior de informação pode ser inserida em áreas de menor importância visual. Enquanto que, em áreas de maior importância visual, uma quantidade menor de informação pode ser inserida, o que de forma geral maximiza a capacidade de armazenamento [33]. Este procedimento garante que seja inserida uma quantidade máxima de informação com um mínimo de distorção.

Ellinas *et al.* [34] propuseram um algoritmo de marca d'água digital que explora as características do SVH inserindo a marca d'água nas componentes de frequência da imagem. Isto é feito utilizando a função de sensibilidade de contraste (CSF, do inglês

Contrast Sensitivity Function) definida pela seguinte equação:

$$CSF(f) = 2.6(0.293 + 0.224f)e^{-(0.114f)^{1.1}}, \tag{3.3}$$

no qual f corresponde à frequência do sinal. Logo, se calcularmos os valores de CSF para todas as frequências visíveis teremos a curva de sensibilidade. Resultados experimentais mostram que a sensibilidade é maior para frequências muito baixas ou muito altas.

No modelo de Ellinas *et al.*, a imagem é decomposta utilizando a transformada DWT. Em seguida, os coeficientes LL de baixa frequência são selecionados e um filtro de *Sobel* é aplicado em cada sub-banda de alta frequência (HL, HH e LH). Nos coeficientes das sub-bandas LL a informação marca é inserida utilizando a seguinte equação:

$$s = x + \alpha_l w_l x m, \tag{3.4}$$

em que s corresponde ao coeficiente de borda modificado pela inserção, α_l corresponde a uma constante que determina a força do sinal da marca inserida, w_l corresponde ao peso visual do coeficiente l. O peso w_l é definido pela função CSF.

O processo de detecção da marca é realizado combinando o sinal original x com o sinal marcado $\hat{x}$ contendo a marca d'água possivelmente deteriorada. A DWT é aplicada nas imagens e o filtro de *Sobel* é aplicado em todos os coeficientes wavelets do sinal. A CSF é, então, utilizada para selecionar os coeficientes de $\hat{x}$ que possuem a marca m. Em seguida, os coeficientes de x são comparados aos coeficientes do sinal $\hat{x}$, resultando em um fator de correlação p, através de seguinte equação:

$$p = \frac{x}{\sqrt{\hat{x}}}. \tag{3.5}$$

A presença da marca d'água é verificada comparando se p é maior que um limiar T, onde $T > 0$. Conforme os experimentos apresentados no trabalho de Ellinas *et al.*, o algoritmo apresentou alta robustez em ataques de compressão, filtragem passa-baixa e

recortes.

Algoritmos perceptivos são pouco utilizados para detecção de adulterações em mídias devido à sua deficiência na localização da marca inserida. Por se tratar de um modelo adaptativo, grande quantidade de informações são inseridas em locais de menor sensibilidade, o que pode comprometer a detecção de adulterações em locais onde uma quantidade menor de informação foi inserida.

3.3.3 Modulação por Índice Quantizado

Coria *et al.* definem quantização como um processo de mapeamento de um conjunto grande de valores a um conjunto menor [35]. O algoritmo de modulação por índice quantizado (QIM, do inglês *Quantization Index Modulation*) foi proposto por Chen *et al.* [36]. A ideia básica do QIM é quantizar a amostra do sinal, x, de acordo com a informação a ser inserida, m. Este processo é ralizado escolhendo-se um quantizador $Q(.)$, a partir de um conjunto de quantizadores.

Considere uma mensagem $m \in \left\{1, 2, \ldots, 2^{NR_m}\right\}$ para ser inserida em um sinal $x \in \Re^N$, no qual R_m corresponde a taxa de inserção em *bits* por amostragem do sinal. Especificamente, $s \in \Re^N$ é gerado através de m e x utilizando uma função de modulação $Q(x, m)$, a qual determina a distorção entre x e s [36]. Uma vez marcado, o sinal s está sujeito a ataques intencionais e não intencionais. O sinal que chega no decodificador é $x' \in \Re^N$. O decodificador gera uma estimativa $\hat{m}$ da marca d'água m inserida, utilizando x' e sem referenciar o sinal original x.

A quantização é definida pela equação seguinte:

$$Q(x_i, \delta) = \left\lfloor \frac{x_i}{\delta} \right\rfloor \delta, \tag{3.6}$$

no qual $\lfloor . \rfloor$ corresponde à operação matemática chão, $Q(x_i, \delta)$ corresponde à função de quantização, que tem como entradas o valor do pixel original e a constante do quantizador escalar (δ), que é ajustado de acordo com a capacidade de inserção desejada. O sinal

quantizado é obtido utilizando a seguinte equação:

$$s(x_i, m) = Q(x_i, \delta) + d(m), \tag{3.7}$$

onde $d(m)$ corresponde a função de modulação de um bit.

A extração da marca d'água é realizada através da seguinte equação:

$$\hat{m} = \hat{x} \mod \delta, \tag{3.8}$$

em que $\hat{m}$ corresponde à marca d'agua extraída e $\hat{x}$ corresponde ao sinal possivelmente adulterado.

Além do algoritmo original, Chen *et al.* propuseram dois outros algoritmos baseados no QIM. O DITHER-QIM e o DC-QIM (do inglês, *Distortion Compensated QIM*) [36].

O *Dither-QIM* consiste na inserção da marca m em um vetor de ruído aleatório *dither* v utilizando a Equação 3.6 resultando em um vetor marcado v'. Este ruído pseudo-aleatório tem por objetivo distorcer a informação da marca quantizada. Em seguida, o vetor v' é inserido na imagem utilizando a Equação 3.7. Este modelo é mais robusto do que a implementação padrão do QIM, pois distribui a potência da marca d'água sob um vetor de ruído v e, em seguida, este vetor é inserido na imagem hospedeira. Com isso, torna-se mais difícil remover a marca d'água uma vez que o vetor v é desconhecido.

O DC-QIM consiste no uso de grandes conjuntos de números aleatórios utilizados para distorcer a marca inserida. A inserção da marca d'água é realizada pela seguinte equação:

$$s = Q(\alpha x_i + c, m) - c, \tag{3.9}$$

em que c é o próximo número inteiro no dicionário aleatório. Note que, se $\alpha = 1$ a informação será inserida sem distorção, ou seja, quantizada apenas como quantizador escalar. Neste modelo, a utilização de um quantizador escalar aumenta a robustez por fator de $1/\alpha^2$ e aumenta a distorção por um fator de $1/\alpha^2$. Em suma, este modelo adiciona

uma distorção em x (adicionando valores de um dicionário de números inteiros aleatórios). Esta característica dificulta a remoção da marca d'água, pois o usuário mal intencionado não tem acesso ao dicionário de números aleatórios.

Dentre os algoritmos de marca d'água presentes na literatura, o QIM é o que possui a maior capacidade de inserção, menor distorção e maior capacidade de localização da marca inserida. Em especial, o QIM nos permite detectar adulterações com granularidade de 1 pixel mantendo a alta taxa de inserção. Por este motivo, neste trabalho escolhemos o QIM.

Capítulo 4

Algoritmo Proposto

Conforme apresentado no Capítulo 2, adulterações são classificadas como: global, local ou temporal. Dentre os métodos de detecção de adulterações disponíveis na literatura, quase a totalidade dos métodos disponíveis é capaz de detectar apenas um tipo de adulteração. Este fato pode limitar o uso destes algoritmos em aplicações reais. Em outras palavras, quando o número de adulterações cobertas pelo algoritmo é limitado, a adoção destes algoritmos como ferramentas de detecção é dificultada, uma vez que teriam que ser adotadas várias soluções para proteger os vídeos contra um número limitado de ataques. Um outro problema comum encontrado nos algoritmos disponíveis na literatura é o fato de algumas técnicas abordarem apenas alguns ataques espaciais e raramente abordam ataques temporais.

Neste capítulo, propomos um algoritmo para detecção de adulterações em vídeos digitais. Este algoritmo utiliza o algoritmo de marca d'água QIM, que possui muitas características interessantes para detecção de adulterações, tais como capacidade de localização de adulteração, baixa degradação do conteúdo, baixa complexidade, sensibilidade a pequenas adulterações, dentre outras.

O algoritmo proposto é dividido em duas unidades principais, proteção de adulterações e detecção de adulterações. A unidade de proteção de adulterações consiste em inserir duas marcas d'água concatenadas e cifradas no canal de vídeo e áudio. Já a unidade de

detecção de adulterações consiste em extrair a informação inserida previamente e buscar possíveis modificações na marca d'água de forma a identificar áreas adulteradas.

4.1 Proteção de Adulterações em Vídeos

Nesta seção, detalhamos o processo de proteção de adulterações do algoritmo proposto. Este processo é dividido em três partes: geração da marca d'água, cifragem e inserção. Na Figura 4.1 é apresentado o diagrama de bloco do processo de proteção de adulterações.

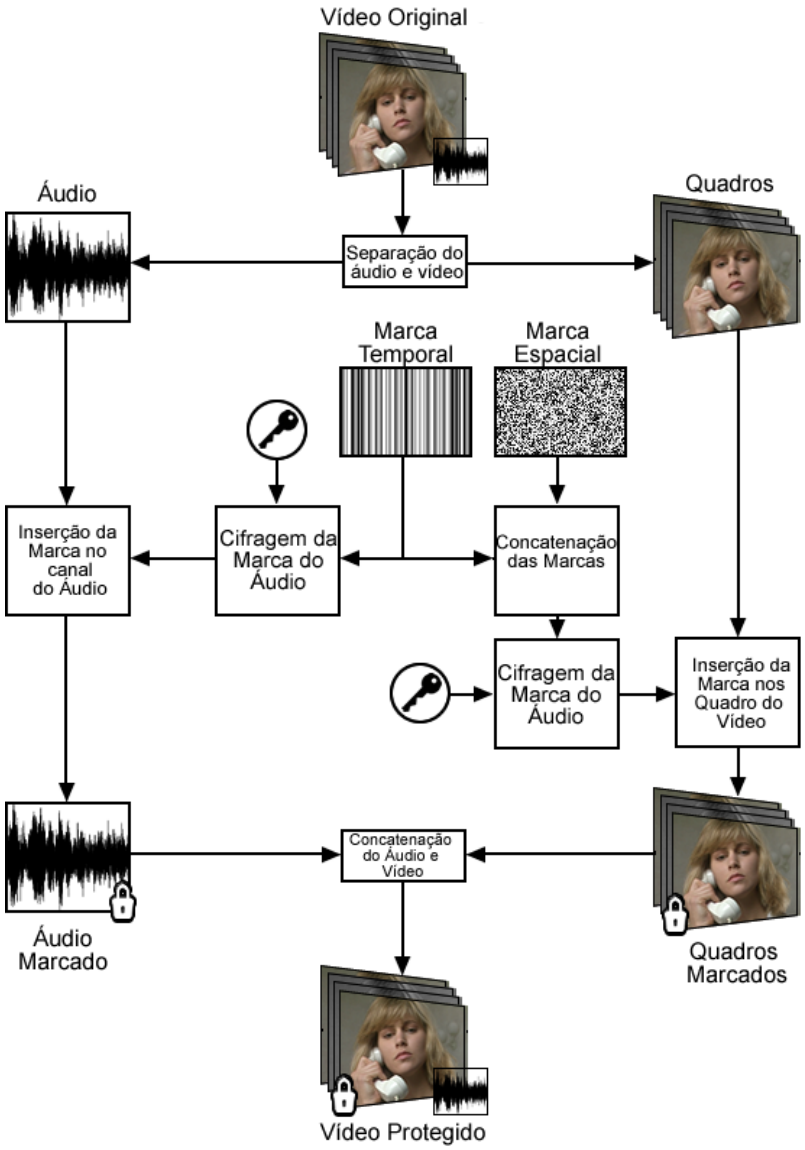


Figura 4.1: Diagrama de Blocos do processo de Proteção de Adulterações.

4.1.1 Geração da Marca d'Água

O primeiro passo do algoritmo proposto é a geração da marca a ser inserida no vídeo. Objetivando a proteção do canal de áudio e vídeo, primeiramente nós decodificamos o vídeo em duas partes: o canal de áudio (A) e o canal de vídeo (F). Usando um gerador de números pseudo-aleatórios e uma chave privada k, geramos uma marca para proteger o conteúdo espacial (M_s) dos quadros do vídeo e uma marca para proteger as informações temporais (M_t).

Para cada quadro do vídeo, a marca M_t corresponde a uma matriz binária com profundidade de um bit e as mesmas dimensões espaciais do vídeo. A marca M_t contém uma sequência binária aleatória de 16 bits, onde cada bit é replicado de modo a preencher as dimensões da matriz. A marca espacial M_s corresponde a uma matriz com profundidade de 5 bits sendo utilizada para detectar adulterações globais e locais no quadros do vídeo. Marcas diferentes são geradas para cada quadro do vídeo, de forma a evitar persistência visual.

A utilização de duas marcas aumenta a robustez e a sensibilidade do algoritmo na detecção de adulterações. Além disso, as duas marcas nos permitem diferenciar entre ataques espaciais e temporais.

4.1.2 Cifragem da Marca d'água

Após a geração de ambas as marcas para cada quadro do vídeo, elas são cifradas antes de serem inseridas no sinal. A marca espacial M_s é cifrada utilizando o algoritmo criptográfico *One Time Pad (OTP)* criado por *Gilbert Vernam* [37]. O algoritmo OTP é baseado na aplicação da operação matemática ou-exclusivo (XOR) entre a marca e uma chave binária aleatória do tamanho da mensagem. Segundo Zhihua *et al.*, o algoritmo OTP é considerado teoricamente seguro, pois não fornece nenhuma informação da mensagem original à criptoanálise [37]. Em suma, a marca d'água é cifrada de acordo com a equação

abaixo:

$$\Psi_s[x, y] = M_s[x, y] \oplus k[x, y], \tag{4.1}$$

em que $\oplus$ corresponde a operação matemática XOR, k a chave criptográfica, e Ψ_s a marca M_s cifrada. Do mesmo modo, a marca temporal M_t é cifrada usando a equação:

$$\Psi_t[x, y] = M_t[x, y] \oplus k[x, y], \tag{4.2}$$

no qual Ψ_t corresponde a marca M_t cifrada.

4.1.3 Inserção da Marca d'Água

Em posse das duas marcas (Ψ_t e Ψ_s) cifradas, cada elemento $\Psi_t[x, y]$ é concatenado em posições aleatórias de cada elemento da marca $\Psi_s[x, y]$. Então, a matriz Ψ_t é inserida em posições aleatórias de Ψ_s, em cada coordenada, $\Psi_t[x, y]$ é inserida em posições aleatórias (na profundidade da matriz) da matriz $\Psi_s[x, y]$ conforme mostrado na Figura 4.2, no qual $i \in [0, 5]$ é um numero aleatório gerado através de um gerador de números pseudo-aleatórios usando a chave k e Φ é a matriz (de profundidade 6 bits) que contém as duas marcas concatenadas.

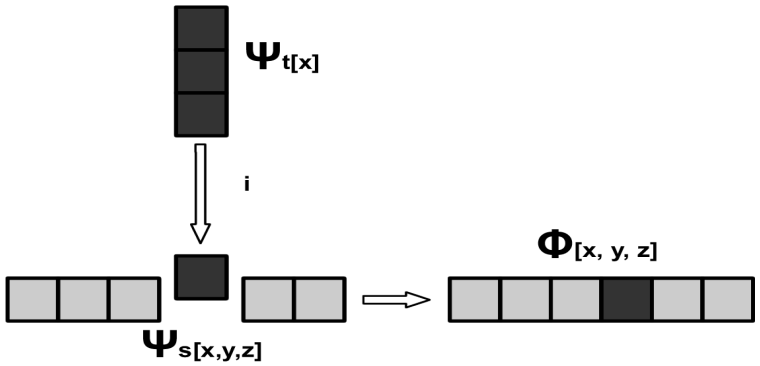


Figura 4.2: Demonstração do processo de concatenação das marcas d'água.

Note que o processo de concatenação das marcas as torna interdependentes, pois quando concatenamos dois números binários os transformamos em outra representação

binária completamente diferente. Além disso, a alteração de um único bit no valor binário concatenado irá comprometer ambas as marcas.

Em seguida, a marca Φ (resultado da concatenação) é inserida em F (quadros do vídeo). Cada posição $\Phi[x, y]$ é inserida na posição x, y de um quadro do vídeo. Cada posição $\Phi[x, y]$ é dividida em três partes iguais com profundidade de 2 bits. Então, são convertidas para base decimal, resultando em três marcas decimais (Φ_1, Φ_2 e Φ_3). Em seguida, são posteriormente inseridas nos três canais de cores (RGB) do quadro nas coordenadas $[x, y]$.

Neste trabalho, nós usamos uma versão modificada do algoritmo QIM que aumenta sua capacidade de inserção. A modificação consiste em usar a função de modulação como função identidade, ou seja $d(m) = m$. Com isso podemos inserir um número inteiro, que representará uma quantidade maior de bits. As três marcas Φ_1, Φ_2 e Φ_3 são inseridas no pixel correspondente no quadro do vídeo, respectivamente, nos três canais de cores R, G e B usando a seguinte equação:

$$F_m[x, y, c] = Q(F_[x, y, c], \delta) + \Phi_c[x, y], \tag{4.3}$$

no qual F_m é o canal de cor marcado, δ é a constante do quantizador escalar, c é o canal de cor (R,G ou B) correspondente e x e y correspondem às posições da matriz do quadro do vídeo. Para este trabalho usamos $\delta = 4$, valor que permite inserir valores inteiros e sua representação decimal entre $(0 - 3)$.

O próximo passo consiste em inserir uma cópia de Ψ_t no canal de áudio. Para isso, Ψ_t é redimensionada em um vetor unidimensional Ψ'_t. Então, Ψ'_t é inserida em A usando a equação:

$$A_m[x] = Q(A_o[i], \delta) + \Psi'_t[i], \tag{4.4}$$

em que A_m é o canal de áudio marcado, A_o é o conteúdo original de A e i é o índice dos vetores.

Note que para cada quadro do vídeo, inserimos também uma cópia da marca temporal

no canal do áudio. Para vincularmos uma sub parte do áudio com um quadro do vídeo, subdividimos o áudio relativo a 1 segundo pelo FPS (do inglês, *Frames Per Second*). Em outras palavras, cada quadro do vídeo tem sua marca replicada em uma subparte do áudio.

Após a inserção das marcas, os canais de áudio e vídeo são re-multiplexados e codificados, resultando no vídeo marcado. Assim, o vídeo está pronto para ser distribuído pelo proprietário.

4.2 Detecção de Adulterações em Vídeos

O processo de detecção de adulterações pode ser divido em quatro estágios: extração da marca, detecção de adulteração espacial, detecção de adulteração temporal e classificação de adulteração temporal. Esse processo é executado no receptor da mídia (decodificador). Na Figura 4.3 é apresentado o diagrama de blocos para o processo de detecção de adulterações.

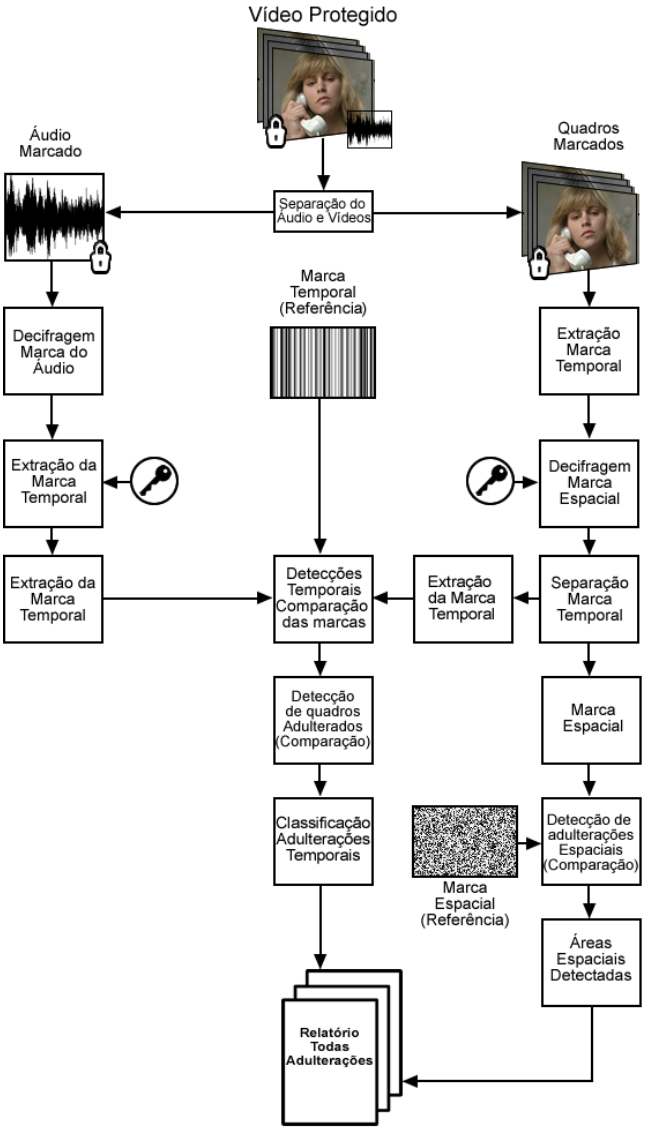


Figura 4.3: Diagrama de blocos do processo de detecção de adulterações.

4.2.1 Extração da Marca d'água

A extração da marca d'água é realizada sobre o vídeo previamente marcado. O primeiro passo consiste em separar os canais de áudio e vídeo e extrair as marcas inseridas em cada um dos canais. Primeiramente, para o canal de áudio, a marca é extraída utilizando a seguinte equação:

$$\hat{\Psi}'_{A,t}[x] = A_m[x] \bmod \delta, \tag{4.5}$$

no qual $\hat{\Psi}'_{A,t}$ é a marca temporal extraída e $A_m[x]$ é o canal do áudio marcado.

Após a extração, a marca unidimensional $\hat{\Psi}'_{A,t}$ é transformada novamente em suas dimensões originais resultando na matriz $\hat{\Psi}_{A,t}$. Então, a marca $\hat{\Psi}_{A,t}$ é decifrada usando a equação:

$$\hat{M}_{A,t}[x,y] = \hat{\Psi}_{A,t}[x,y] \oplus k, \tag{4.6}$$

no qual $\hat{M}_{A,t}$ é a marca temporal decifrada extraída de A.

A marca inserida no canal F (quadros do vídeo) é uma combinação de duas marcas: a marca espacial e a marca temporal. Então, nós precisamos extrair ambas as marcas desta combinação, decifrá-las e separá-las. Para cada quadro do vídeo, a extração da marca é realizada utilizando a equação:

$$\sigma[x,y,c] = F_m[x,y,c] \bmod \delta, \tag{4.7}$$

na qual $\sigma[x,y,c]$ é a marca extraída correspondendo à posição espacial (x,y,c) e o canal de cor c. É importante notar que $\sigma[x,y,c] \in [0,3]$. Em outras palavras, $\sigma[x,y,c]$ possui valores no intervalo $(0-3)$, ou seja, números de 2 bits em sua representação na base decimal.

Em seguida, $\sigma[x,y,c]$ é transformada para uma representação binária, resultando em três variáveis γ_r, γ_g e γ_b que correspondem aos três canais de cores (R,G e B). Então, as três variáveis são concatenadas usando a seguinte equação:

$$\hat{\Phi}[x,y] = \gamma_r \mid\mid \gamma_g \mid\mid \gamma_b, \tag{4.8}$$

na qual $\hat{\Phi}[x,y]$ corresponde às três marcas concatenadas e $||$ denota a operação de concatenação. Note que $\hat{\Phi}[x,y,z]$ tem profundidade de $3 \times \delta$ e z corresponde à coordenada de profundidade de $\hat{\Phi}$. Para $\delta = 2$, $\hat{\Phi}$ possui 6 níveis de profundidade. Cada bit é endereçado separadamente através da coordenada z.

Uma vez que $\hat{\Phi}[x,y,z]$ esteja construída, o próximo passo consiste na separararação da

marca espacial e temporal. Para isso, é gerado uma sequência aleatória $i \in [0, 5]$ utilizando a mesma chave secreta k do processo de inserção. Então, as marcas são separadas, conforme ilustrado na Figura 4.4.

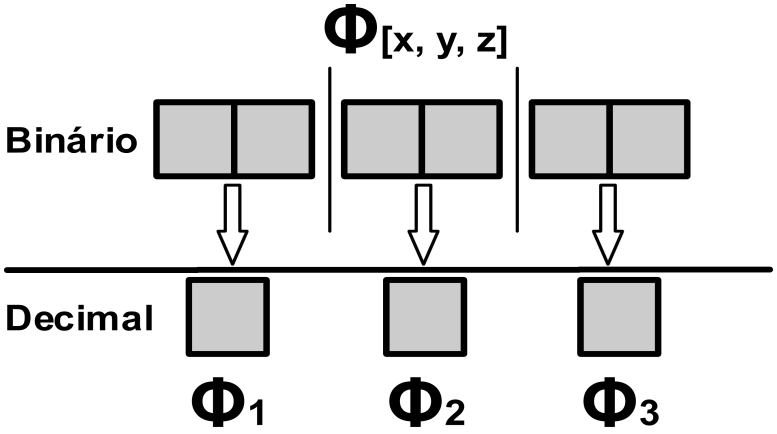


Figura 4.4: Processo de separação das marcas temporal e espacial.

A coordenada $\hat{\Psi}_{F,t}[x, y]$ recebe o bit correspondente à marca temporal. Esse bit está na i-*ésima* posição, conforme demonstrado no processo de inserção. Para todos os outros casos, $\hat{\Psi}_{F,s}$ recebe os outros 5 bits. Note que $\hat{\Psi}_{F,s}$ possui profundidade de 5 bits, enquanto $\hat{\Psi}_{F,t}[x, y]$ um nível apenas. Neste ponto, ambas marcas estão cifradas e possuem o mesmo tamanho (matrizes) do quadro do vídeo.

O próximo passo consiste em decifrar as marcas extraídas. A marca $\hat{\Psi}_{F,s}$ é decifrada pela seguinte equação:

$$\hat{M}_{F,s}[x, y] = \hat{\Psi}_{F,s}[x, y] \oplus k, \tag{4.9}$$

no qual $\hat{M}_{F,s}$ é uma matriz contendo a marca espacial decifrada com profundidade de 5 bits. Do mesmo modo, a marca $\hat{\Psi}_{F,t}$ é decifrada usando a equação:

$$\hat{M}_{F,t}[x, y] = \hat{\Psi}_{F,t}[x, y] \oplus k, \tag{4.10}$$

no qual $\hat{M}_{F,t}$ é a marca temporal extraída do quadro do vídeo.

Depois da extração da marca temporal e espacial e de sua decifragem, elas são comparadas com as marcas originalmente inseridas. Para isso, as marcas originais são geradas

usando o mesmo processo da inserção. O processo de detecção é dividido em detecção temporal e espacial.

4.2.2 Detecção de Adulterações Espacial

Na detecção espacial, o objetivo é detectar adulterações locais e globais. Para cada quadro do vídeo, a detecção local é realizada comparando-se a marca extraída ($\hat{M}_{F,s}$) com a marca originalmente inserida (M_s) usando a seguinte equação:

$$\rho_{F,s}[x,y] = \begin{cases} verdadeiro, & \text{se } \hat{M}_{F,s}[x,y,z] \neq M_s[x,y] \\ falso, & \text{caso contrário} \end{cases} \tag{4.11}$$

no qual $\rho_{F,s}$ é uma *Matriz de Detecção Espacial* (MDE) booleana. Nessa matriz, cada posição que possuir o valor 'verdadeiro' corresponde a um pixel adulterado. Note que, com esta abordagem de detecção é possível identificar a exata localização de uma adulteração.

Quando uma ou mais posições da matriz MDE forem encontradas, a porcentagem total de áreas adulteradas é calculada em relação ao vídeo. A ideia é estimar a classificação e o tipo da adulteração. Se a porcentagem for maior que um limiar $\tau_{F,s}$, o quadro é classificado como globalmente adulterado. Caso contrário o quadro é classificado como localmente adulterado. Nas simulações, utilizamos como limiar o valor $\tau_{F,s} = 85\%$.

Note que qualquer tipo de ataque espacial (local ou global) adultera a marca temporal inserida no quadro do vídeo. Isso é um problema pois quando a marca temporal for perdida, sua detecção poderá falhar. Para solucionar este problema, quando uma adulteração espacial danificar a marca temporal em mais de 85% do quadro, a marca temporal utilizada será a marca replicada no canal A (áudio). Isso permite diferenciar entre ataques temporais, locais ou globais, conforme descrito a seguir.

4.2.3 Detecção de Quadros Temporal

O processo de detecção temporal utiliza ambas as marcas temporais extraídas dos dois canais F e A. Primeiramente, o processo é iniciado usando a seguinte equação:

$$\rho_{F,t}[x,y] = \begin{cases} verdadeiro, & \text{se } \hat{M}_{F,t}[x,y] \neq M_t[x,y] \\ falso, & \text{caso contrário} \end{cases} \tag{4.12}$$

no qual M_t é a marca original (referência) e $\rho_{F,t}$ é a *Matriz de Detecções Temporais* booleana (MDT). Quando a matriz MDT tiver valores 'verdadeiro', uma perda da marca temporal ocorreu e o quadro do vídeo correspondente sofreu uma provável adulteração temporal.

Para detectar se o quadro está *temporalmente adulterado*, o percentual de valores 'verdadeiro' presentes na matriz MDT é comparado com um limiar $\tau_{F,t}$. Para este trabalho, o limiar foi definido como $\tau_{F,t} = 15\%$. Caso a porcentagem encontrada seja menor que $\tau_{F,s}$, o quadro é classificado como não adulterado temporalmente. Isso estabelece que a marca d'água temporal inserida no quadro do vídeo foi perdida em menos de 15%. Caso a marca ultrapasse o limiar $\tau_{F,s}$, a porcentagem se torna uma medida imprecisa pois uma adulteração local ou global pode degradar a marca temporal. Neste caso, a marca temporal inserida no canal de áudio é utilizada para verificar adulterações temporais. Então, a marca $\hat{M}_{A,t}$ é comparada com a marca temporal original usando a seguinte equação:

$$\rho_{A,t}[x,y] = \begin{cases} verdadeiro, & \text{se } \hat{M}_{A,t}[x,y] \neq M_t[x,y] \\ falso, & \text{caso contrário} \end{cases} \tag{4.13}$$

em que $\rho_{A,t}$ é a MDT para o canal de áudio A. Usando $\rho_{A,t}$, a porcentagem de valores 'verdadeiro' é calculada e comparada com outro limiar $\tau_{A,t}$. Quando a porcentagem de perda da marca d'água for menor que $\tau_{A,t}$, o quadro é classificado como não sendo temporalmente adulterado.

Quando a validação da presença da marca d'água temporal falhar para os dois canais (áudio e vídeo), o quadro é classificado como temporalmente adulterado. É utilizada uma dupla checagem, o que torna possível diferenciar entre adulterações espaciais e temporais.

Em seguida, um vetor Θ é criado para estimar o tipo de adulteração temporal ocorrida. Este vetor armazena o resultado da detecção temporal. Caso o quadro seja classificado como temporalmente adulterado, Θ armazena a cópia da marca temporal extraída. Caso contrário, Θ armazena $\varnothing$ (vazio) na posição correspondente no quadro.

4.2.4 Classificação de Adulterações Temporais

Uma vez que a localização das adulterações temporais são conhecidas, pode-se estimar o tipo de adulteração ocorrida. Para isso, primeiramente é gerado um vetor Ω contendo as marcas temporais originalmente inseridas. Então, o vetor Θ é analisado usando Ω como referência. A seguir, a lista de análises realizadas para classificação dos ataques.

1. A primeira análise consiste em encontrar todas as posições de Ω que não possuem em Θ. Caso forem encontradas posições que satisfação este critério, calculamos a distância entre as posições originais e adulteradas. Em seguida, verificamos se todas as ocorrências possuem a mesma distância temporal e, caso constatado, a adulteração é classificada como Redução de Quadros. Caso contrário, a adulteração é classificada como Remoção de Quadros.

2. A segunda análise consiste em verificar se todas as ocorrências de Ω estão presentes em Θ. Em seguida, é verificado se Θ contém elementos que não estão presentes em Ω. Se as duas condições forem satisfeitas, Θ contém mais elementos que o vetor de marcas originais Ω. Isso indica que novos quadros foram adicionados ao vídeo marcado. Essa adulteração é classificada como Inserção de Quadros.

3. A terceira análise consiste em buscar todos os elementos de Ω que ocorrem mais de uma vez em Θ. O resultado desta análise indica que um ou mais quadros foram

copiados e colados em outras posições do vídeo. Essa adulteração é classificada como Duplicação de Quadros.

4. Finalmente, na quarta análise, para cada elemento marcado como adulterado na posição i em Θ, é buscada sua posição correta j em Ω. Então, é verificada se a posição do elemento $\Theta[j]$ é semelhante a $\Omega[i]$. Se as posições forem semelhantes, a adulteração é classificada como Embaralhamento de Quadros.

A análise das características dos ataques sobre a linha do tempo do vídeo nos permite identificar a relação entre as adulterações e, na maioria dos casos, estimar sua classificação e tipo. Isso é possível pois a marca temporal corresponde a uma sequência aleatória onde as suas posições na linha do tempo são conhecidas. Esta característica permite identificar quadros removidos, quadros inseridos, quadros com posições trocadas entre si, cópias de quadros, dentre outros.

A utilização de uma sequência aleatória replicada por toda a área espacial do vídeo e no canal de áudio introduz um segundo nível de detecção e amplia sua redundância. No próximo capítulo são apresentados os resultados detecção de adulterações temporais e espaciais utilizado o algoritmo proposto.

Capítulo 5

Simulações e resultados

5.1 Simulações

Neste trabalho, nós usamos uma biblioteca de quinze vídeos, de licença livre, baixados do banco de dados ReefVid mantido pela Universidade de Queensland, Austrália [38]. Um quadro vídeos utilizados em nossos são mostrados na Figura 5.1. É importante ressaltar que esses vídeos têm diferentes características espaciais e temporais. Alguns possuem alta atividade espacial (textura), enquanto que outros têm alta atividade temporal (movimentos). Esta diversidade é um requerimento importante na escolha de vídeos para testes de adulterações. Todos os vídeos testados possuem formato AVI sem compressão e espaço de cores RGB.

Em nossas simulações foram testados dez tipos de ataques, temporais e espaciais. Três ataques temporais foram considerados: Redução de Quadros ou Reduçao de Framerate, Embaralhamento de Quadros e Duplicação de Quadros. No ataque conhecido como Duplicação de Quadros, os quadros do vídeo são copiados e inseridos nas posições subsequentes. As posições dos quadros adulterados são aleatórias e a sua quantidade é estabelecida como um parâmetro do algoritmo. Neste trabalho, consideramos perdas de quadros entre 5% e 25% do número total de quadros do vídeo. O ataque conhecido como Redução de Quadros consiste em remover quadros do vídeo. Neste trabalho, removemos 1 quadro em uma

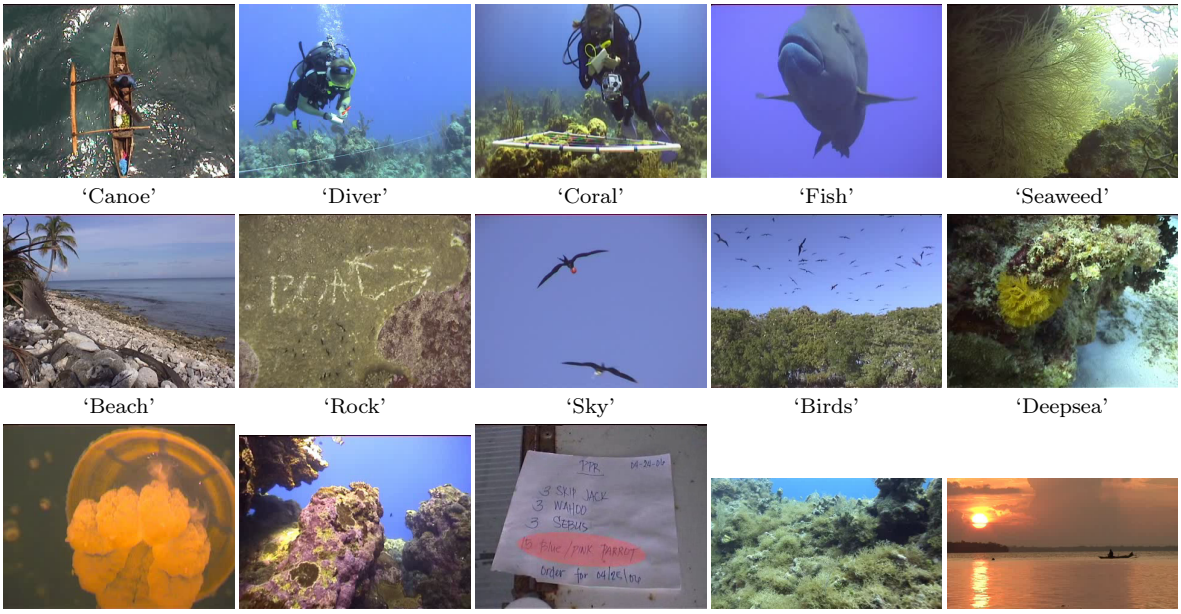


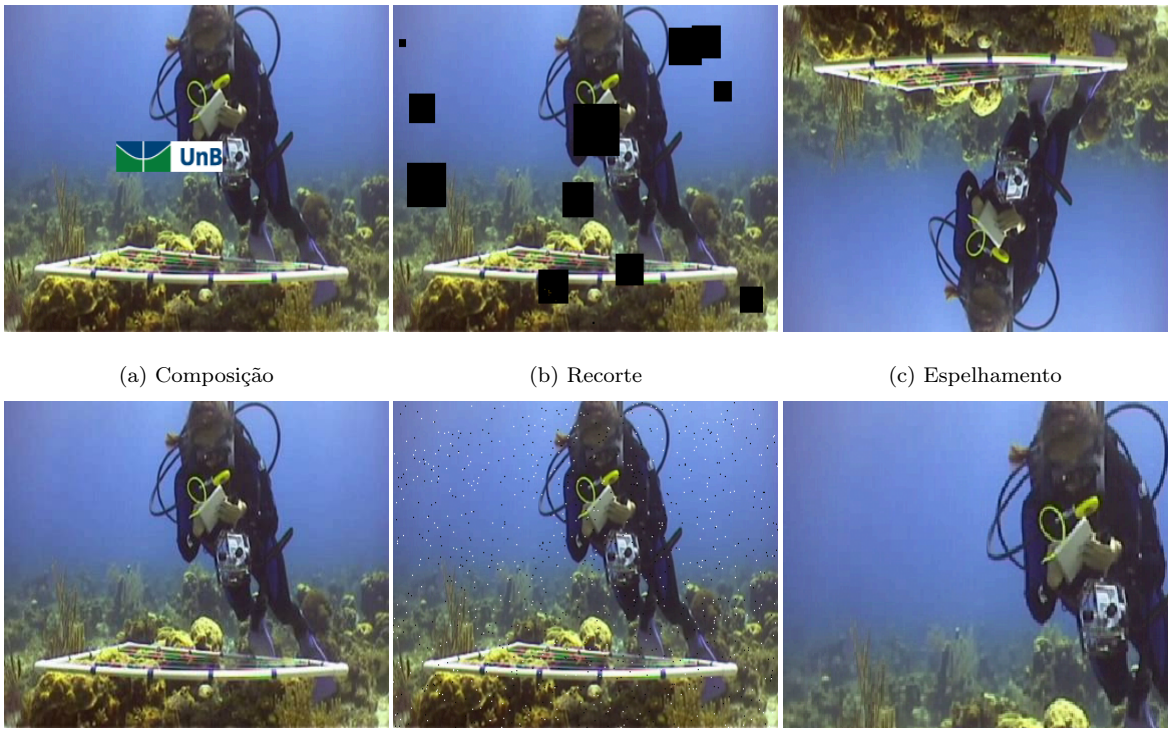
Figura 5.1: Quadros dos vídeos dos vídeos utilizados nas simulações.

janela de 5 quadros. O último tipo de adulteração temporal testado é o Embaralhamento de Quadros. Nesta adulteração são escolhidos pares de quadros aleatoriamente e, para cada par, são trocadas suas posições na linha do tempo. Neste trabalho, a quantidade de pares foi configurada entre 5% e 25% sobre o total de quadros do vídeo.

As adulterações espaciais testadas em nossas simulações são: Composição, Recorte, Espelhamento, Compressão JPEG, Rotação, Adição de Ruído Sal e Pimenta e Redimensionamento. Trataremos a adulteração de Adição de Ruído Sal e Pimenta apenas como Sal e Pimenta.

Na adulteração por Composição foi utilizado o logotipo da UnB de tamanho 106×27 pixels. O logo é colado (sobreposto) em posições aleatórias dos quadros do vídeo. Os quadros adulterados são escolhidos aleatoriamente. Um exemplo deste ataque é exibido na Figura 5.2(a). Na adulteração por Recorte, são escolhidos quadros aleatoriamente na linha do tempo do vídeo. Em seguida, uma quantidade aleatória de blocos (retângulos) de tamanhos também aleatórios são aplicados no quadro em forma de recorte. Em outras palavras, o recorte é aplicado substituindo-se toda a área de todos os retângulos por zero. Um exemplo deste ataque é exibido na Figura 5.2(b). Na adulteração por Espelhamento são escolhidos quadros aleatórios do vídeo. Então, cada quadro é rotacionado horizon-

talmente sobre seu próprio eixo. Um exemplo deste ataque é exibido na Figura 5.2(c). Na adulteração por Compressão JPEG são escolhidos quadros aleatórios do vídeo. Cada quadro é comprimido com fator de compressão JPEG entre 75% e 95% escolhidos aleatóriamente. Um exemplo deste ataque é exibido na Figura 5.2(d). Na adulteração por Sal e Pimenta são escolhidos quadros aleatoriamente do vídeo. Para cada quadro é aplicado o ruído Sal e Pimenta em 2% da área do quadro. Um exemplo deste ataque é exibido na Figura 5.2(e). Na adulteração por Rotação, cada quadro é rotacionado aleatoriamente entre 90º, 180ºou 270º. E, finalmente, para o ataque de Redimensionamento também são escolhidos quadros aleatórios do vídeo. Cada quadro é ampliado entre 50% e 100% do seu tamanho utilizando-se interpolação bi-cúbica. Em seguida, o quadro é recortado, partindo-se do centro, para seu tamanho original. Um exemplo deste ataque é exibido na Figura 5.2(f).



(a) Composição (b) Recorte (c) Espelhamento

(d) JPEG (e) Sal-e-Pimenta (f) Redimensionamento

Figura 5.2: Ataques espaciais utilizando o décimo quinto quadro do vídeo 'Diver'.

5.2 Resultados

5.2.1 Análise da Qualidade Objetiva das Imagens Marcadas

O primeiro teste realizado consiste em analisar a qualidade dos vídeos *marcados*. Embora o nosso objetivo seja proteger o conteúdo do vídeo de qualquer adulteração, é importante que a qualidade geral do vídeo não seja afetada pelo algoritmo de marca d'água. Esta análise foi realizada comparando-se os quadros do vídeo marcado e os quadros originais usando duas das mais populares métricas: PSNR (do inglês, Peak signal-to-noise ratio) e SSIM (do inglês, Structural SIMilarity) [39]. Os valores obtidos com estas métricas são exibidos na Tabela 5.1. Os valores de PSNR obtidos todos acima de 45dB, enquanto que os valores do SSIM são todos acima de 0.999. Estes valores indicam que os vídeos marcados possuem alta qualidade com defeitos imperceptíveis.

Video	Quadros	PSNR	SSIM
Canoe	413	45.53459	0.99966
Diver	337	45.40895	0.99968
Coral	314	45.51629	0.99967
Fish	612	45.56291	0.99964
Seaweed	191	45.44542	0.99937
Beach	803	45.59478	0.99955
Rock	19	45.37220	0.99967
Sky	273	45.45621	0.99966
Birds	102	45.38711	0.99967
Deepsea	58	45.41717	0.99958
Aquamarine	1123	45.51627	0.99940
Rocks	751	45.57977	0.99973
Bill	900	45.54493	0.99936
Alga	576	45.55015	0.99966
Sunset	1938	45.64242	0.99970

Tabela 5.1: Valores de PSNR e SSIM calculados entre os quadros originais e os quadros dos vídeos marcados utilizados nas simulações.

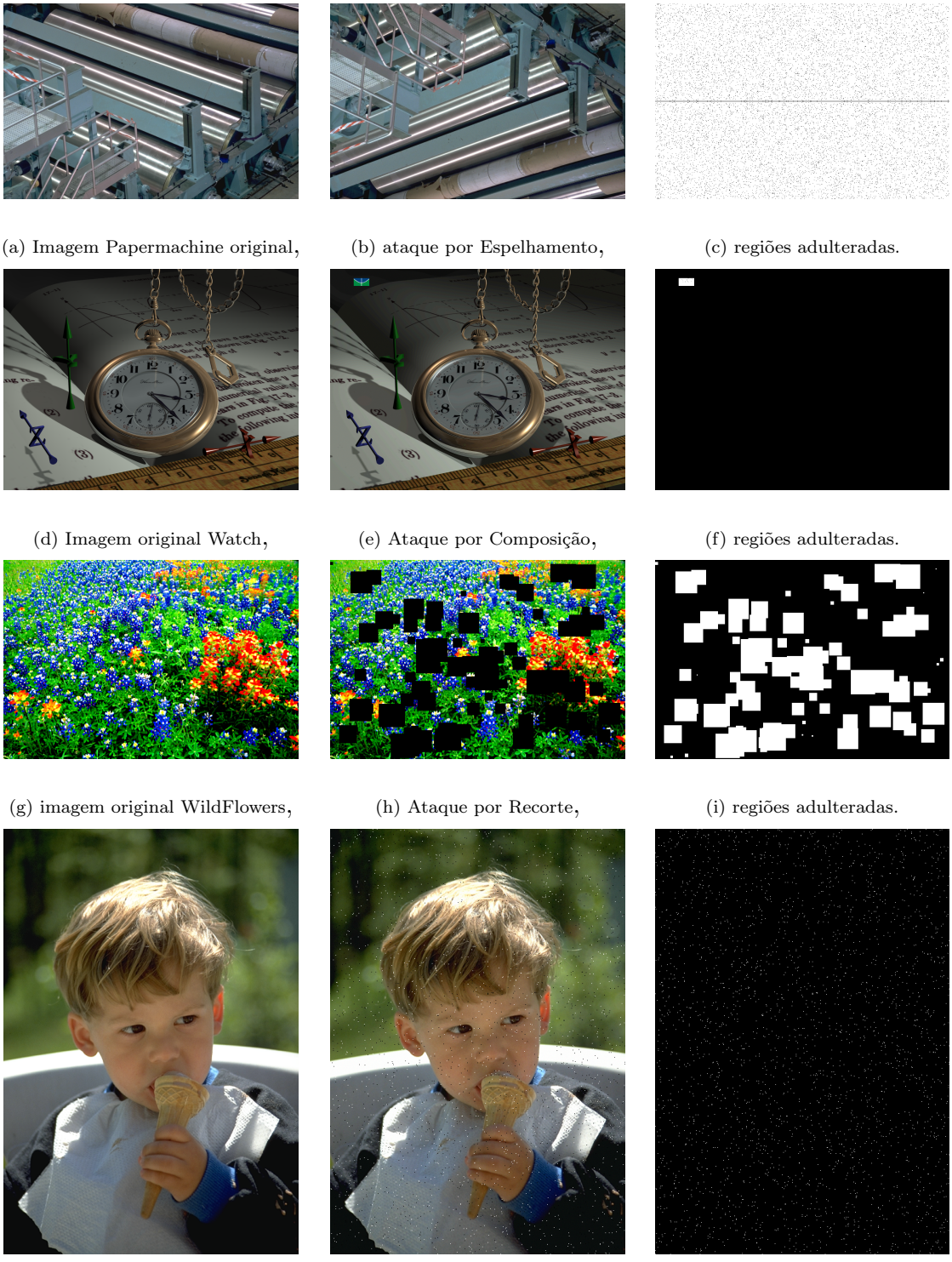
5.2.2 Resultados com Imagens e Vídeos sem Áudio

Os primeiros testes do algoritmo foram efetuados com imagens estáticas e vídeos sem áudio. Neste momento testamos a capacidade do algoritmo para detectar adulterações locais e globais (espaciais). Para os testes com imagens utilizamos a marca d'água com uma matriz binária aleatória e o algoritmo QIM configurado com $\delta = 4$. Deste modo,

adquirimos uma menor degradação da imagem marcada e aumentamos a robustez para ataques de baixa granularidade.

Na Figura 5.3 são apresentados alguns exemplos de ataques e sua detecções. Na Figura 5.3(a) é apresentado a imagem "Papermachine" original e na Figura 5.3(b) a adulteração do tipo "Espelhamento" desta imagem, que caracteriza-se pela rotação da imagem em torno do seu próprio eixo. Na Figura 5.3(c) é apresentada a detecção obtida utilizando o algoritmo proposto, na qual pode-se ver claramente uma linha horizontal ao centro da imagem indicando o eixo ao redor do qual a imagem foi rotacionada. Na Figura 5.3(d) é apresentada a imagem "Watch" original, enquanto que na Figura 5.3(e) é apresentada a sua versão adulterada que consiste na adição do logo da UnB no canto superior esquerdo da imagem. Na Figura 5.3(f) é apresentada a área detectada como adulterada. Na Figura 5.3(g), a imagem "WildFlowers" original é apresentada, enquanto que na Figura 5.3(h) é apresentada a sua versão adulterada com recortes aleatórios de vários tamanhos. Na Figura 5.3(i) são apresentadas as áreas detectadas como adulteradas. Finalmente, na Figura 5.3(j) é apresentada a imagem "Kid" original, enquanto que na Figura 5.3(l) é apresentada a sua versão adulterada por adição de ruído Sal e Pimenta e na Figura 5.3(m) as áreas adulteradas detectadas pelo algoritmo.

Os testes executados demonstram a boa capacidade do algoritmo na localização espacial e sensibilidade a ataques de baixa granularidade. Na Figura 5.3(c) podemos notar claramente uma linha horizontal no centro da imagem. Isso ocorre pois a imagem foi rotacionada sobre seu próprio eixo e a informação da marca d'água inserida no eixo central permaneceu no mesmo local mesmo após a adulteração. Na Figura 5.3(m) notamos a robustez do algoritmo para adulterações de baixa granularidade por adição de ruído Sal e Pimenta.



(a) Imagem Papermachine original, (b) ataque por Espelhamento, (c) regiões adulteradas.

(d) Imagem original Watch, (e) Ataque por Composição, (f) regiões adulteradas.

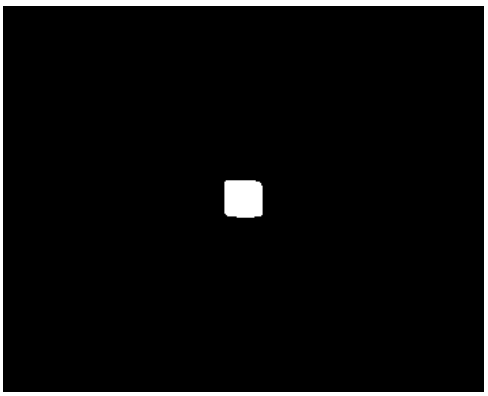
(g) imagem original WildFlowers, (h) Ataque por Recorte, (i) regiões adulteradas.

(j) imagem original Kid, (l) Ataque de Ruído Sal e Pimenta, (m) regiões adulteradas.

Figura 5.3: Resultado da aplicação do algoritmo proposto para imagens estáticas. A primeira coluna exibe as imagens originais, áreas adulteradas na segunda coluna e áreas detectadas pelo algoritmo na terceira coluna.

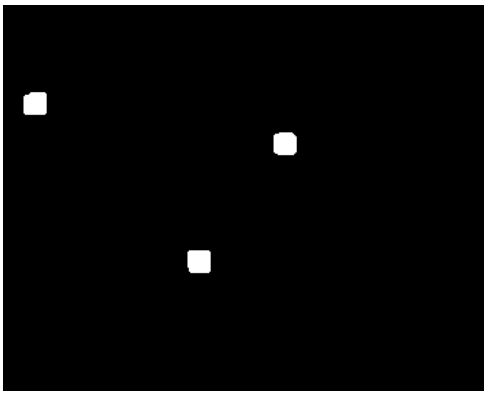
A próxima análise é feita utilizando vídeos sem áudio. A detecção é semelhante a uma sequência de imagens estáticas. A Figura 5.4 exibe um exemplo para o vídeo "Container" e três tipos de ataques locais. Na primeira coluna são exibidas as imagens adulteradas e, na segunda coluna, suas detecções.

(a) Borramento aleatório de uma parte do vídeo.



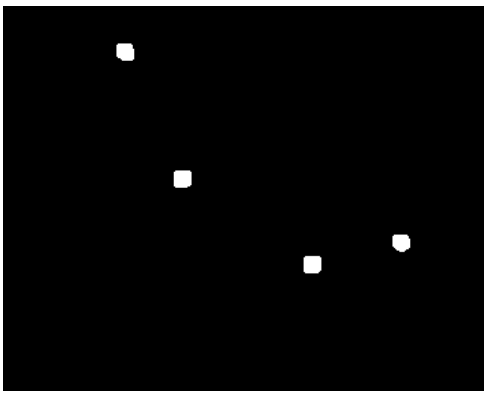
(b) Detecção da adulteração.

(c) "Copiar e Colar" de áreas aleatórios.



(d) Resultado da detecção.

(e) 'Borramento' de áreas aleatórias.



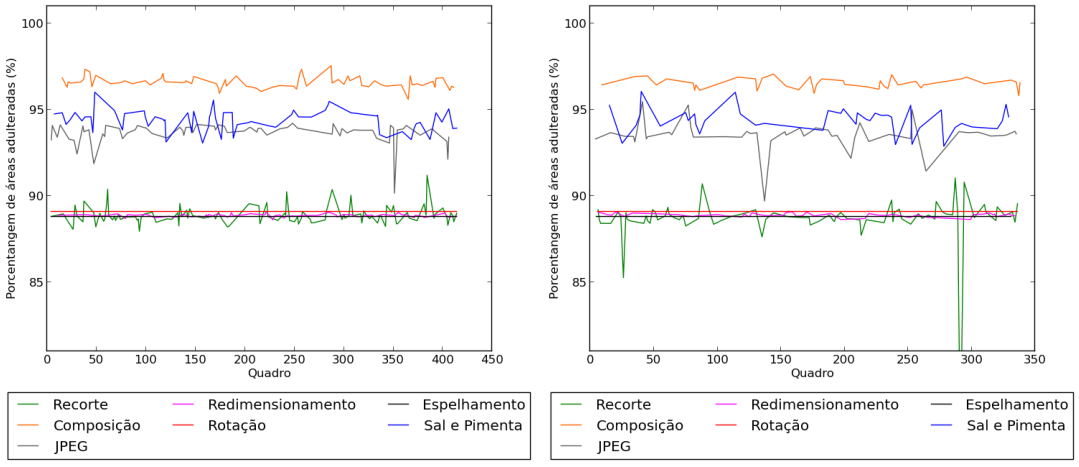
(f) Resultado da detecção.

Figura 5.4: Exemplo de detecção de adulterações no vídeo "Container" (vídeo sem áudio). Fonte: *http://www.cdvl.org.*

5.2.3 Resultados de Vídeos com Áudio

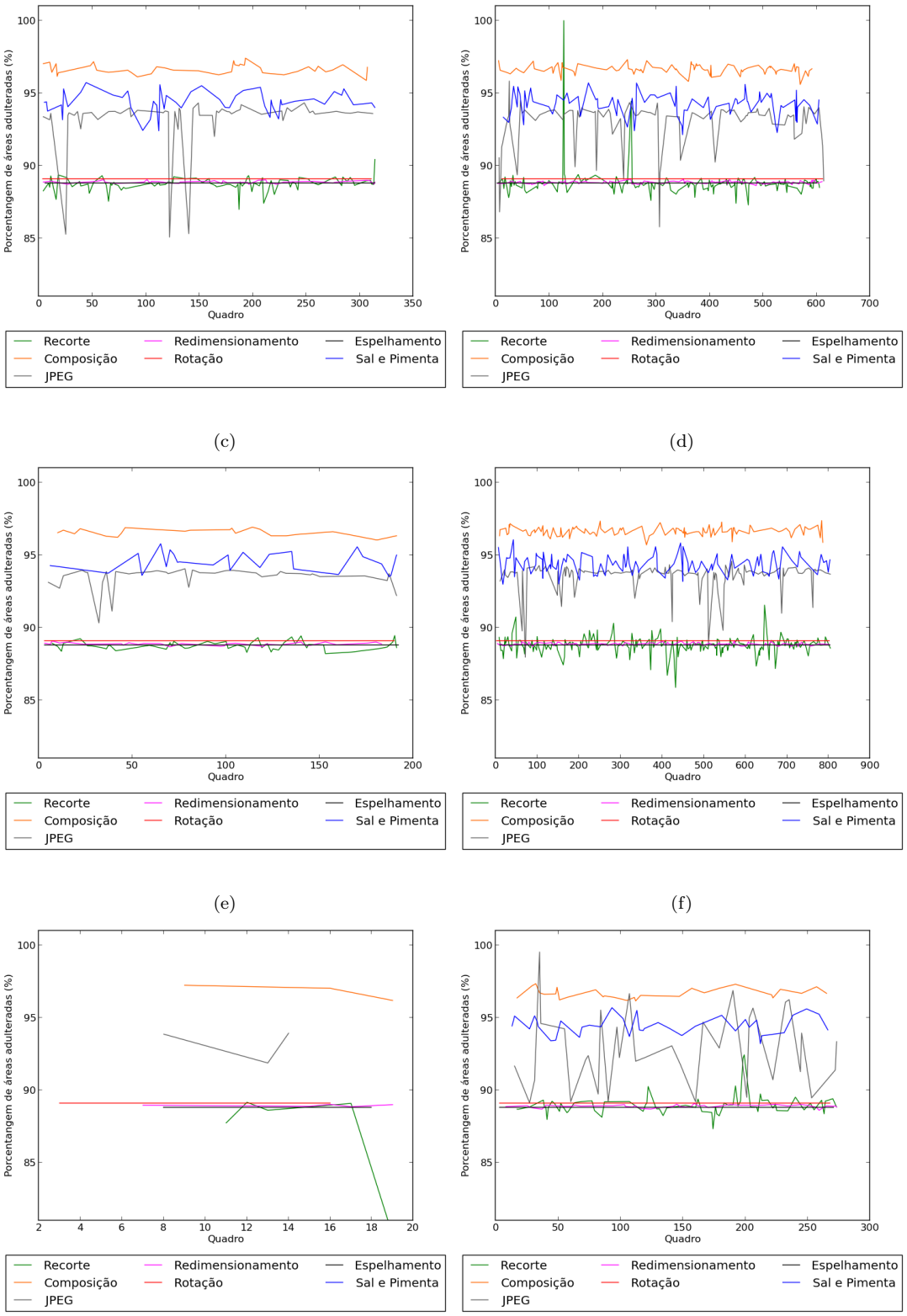
Em seguida, nós testamos os 10 tipos de ataques (conforme explicado no início deste capítulo). Para os ataques espaciais, foram efetuados os mesmos testes para todos os vídeos exibidos na Figura 5.1. Nossa segunda análise consiste na comparação geral entre adulterações e detecções (espaciais e temporais). Para estas simulações utilizamos a técnica proposta no Capítulo 4.

Nas Figuras 5.5, 5.6, 5.7 e 5.8 são exibidos a porcentagem de pixels detectados como adulterados por vídeo. Em outras palavras, representam a taxa de áreas detectadas sobre adulteradas. O algoritmo proposto é capaz de detectar mais de 85% das adulterações espaciais e globais testadas. Como pode ser visto nas Figuras 5.5, 5.6, 5.7 e 5.8. O pior caso foi obtido com as adulterações por Recorte e Compressão JPEG. Nestes casos, foram encontrados menos de 10% de falso-negativos. No caso do Recorte, muitas áreas recortadas podem sobrescrever outras áreas já recortadas, isso se deve à aleatoriedade na configuração da simulação. Neste caso, o algoritmo consegue detectar apenas uma área adulterada. No caso de Compressão JPEG, algumas frequências altas não são eliminadas. Com isso, as informações da marca d'água nestes pixels não são perdidas. Em consequência, estas regiões não são classificadas como adulteradas.



(a) (b)

Figura 5.5: Porcentagem da detecção das sete adulterações espaciais por quadro dos vídeos 'Canoe' (a) e 'Diver' (b).



(c) (d)

(e) (f)

(g) (h)

Figura 5.6: Porcentagem da detecção das sete adulterações espaciais por quadro dos vídeos 'Coral'(c), 'Fish' (d), 'Seaweed' (e), 'Beach' (f), 'Rock' (g) e 'Sky' (h).

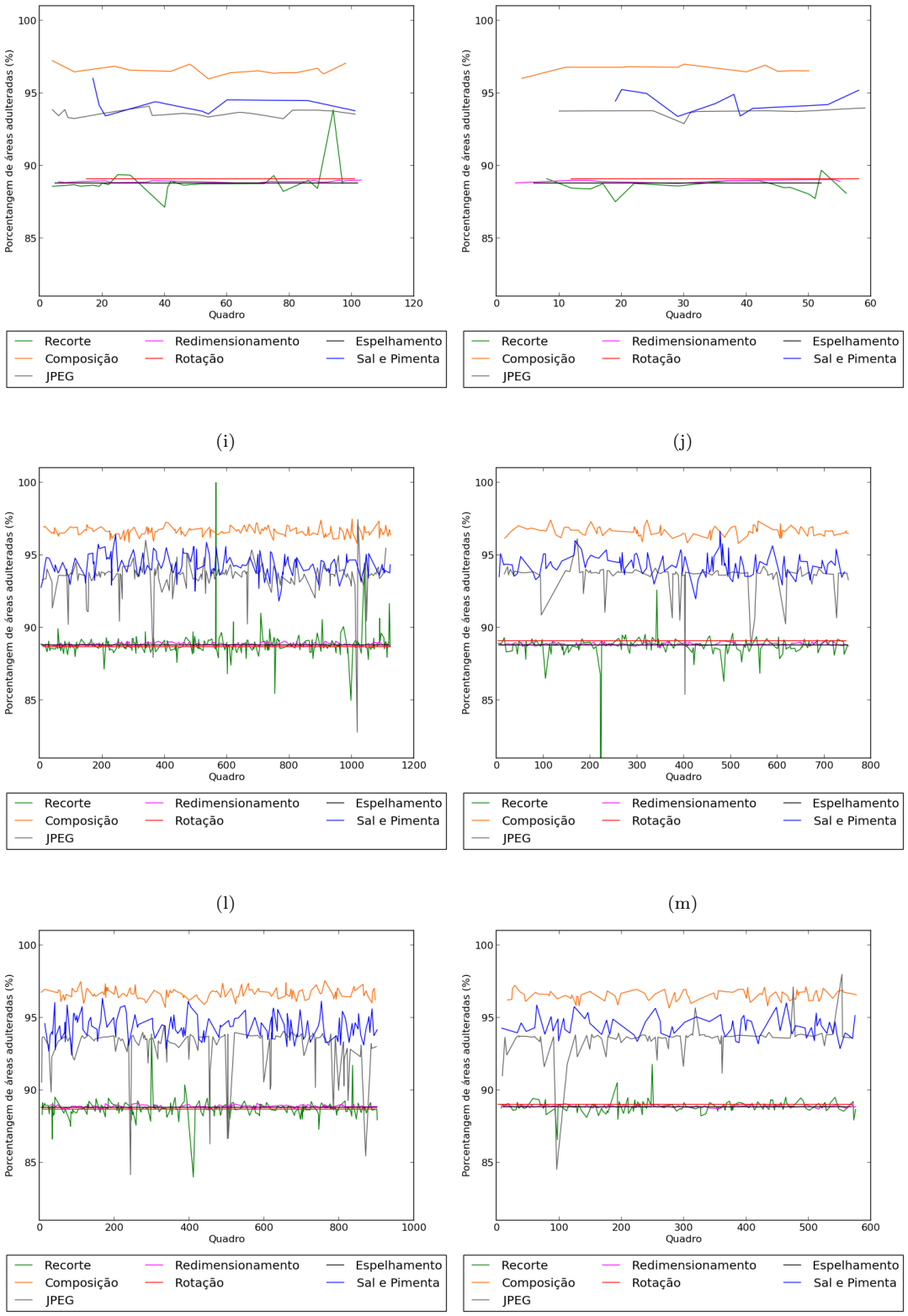
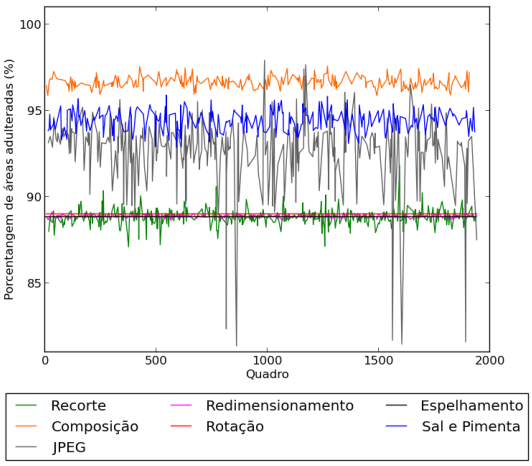


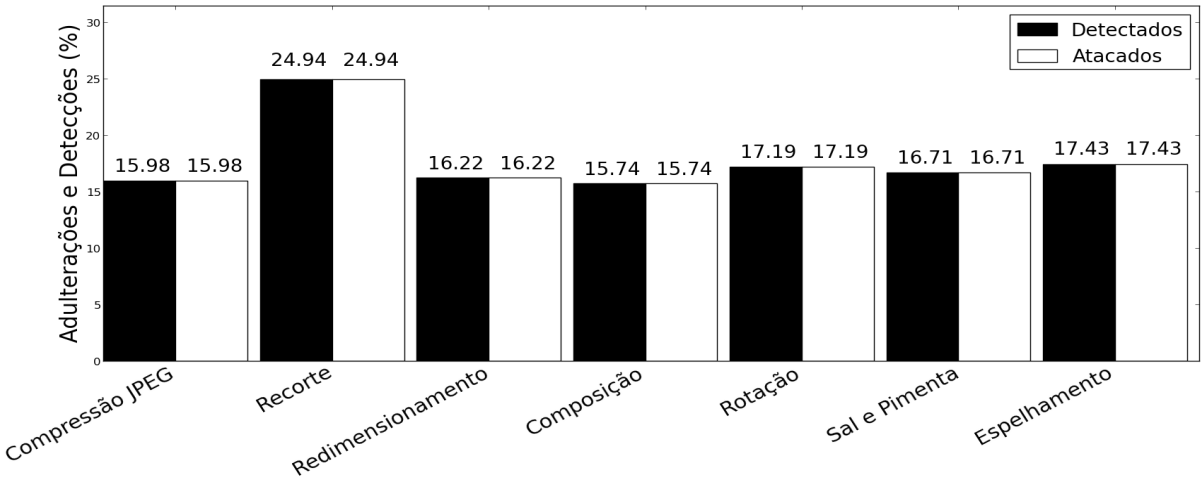
Figura 5.7: Porcentagem da detecção das sete adulterações espaciais por quadro dos vídeos ‘Birds’ (i), ‘Deepsea’ (j), ‘Aquamarine’ (l), ‘Rocks’ (m), ‘Bill’ (n) e ‘Alga’ (o).



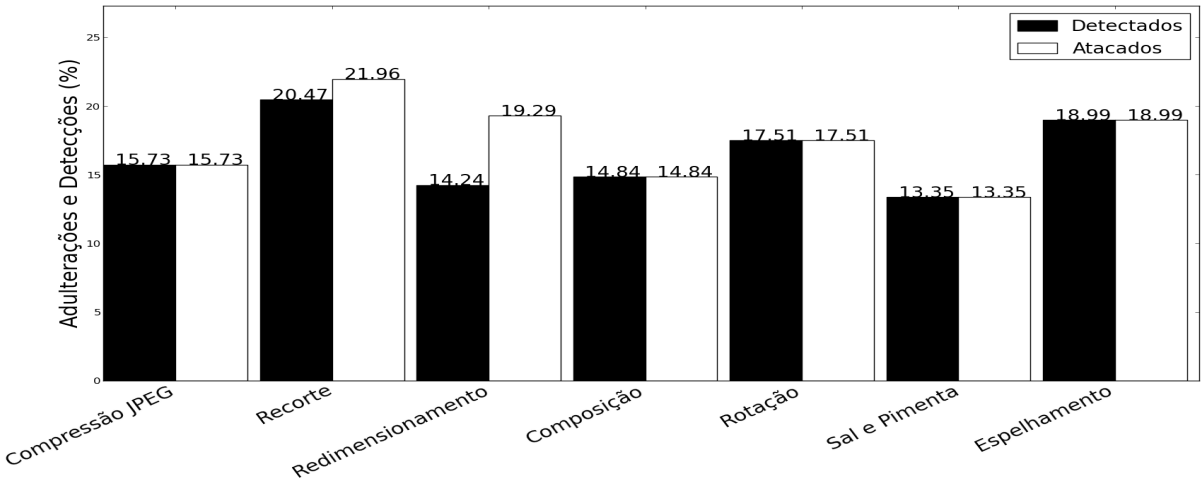
(p)

Figura 5.8: Porcentagem da detecção das sete adulterações espaciais por quadro do vídeo 'Sunset' (p).

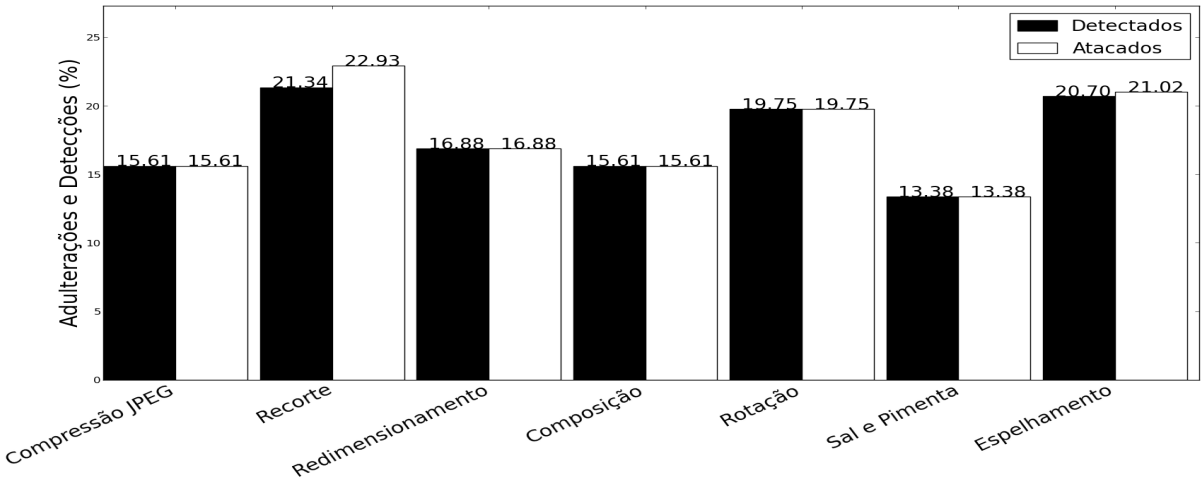
Enquanto que as Figuras 5.6, 5.7 e 5.8 mostram a taxa de detecção das adulterações espaciais por frame, as Figuras 5.9, 5.10, 5.11, 5.12 e 5.13 mostram a porcentagem de quadros atacados espacialmente comparados com a porcentagem de quadros detectados para todos os ataques espaciais e todos os vídeos. Nestes gráficos, as barras brancas correspondem à porcentagem de quadros atacados, enquanto que as barras pretas correspondem a porcentagem de quadros detectados como adulterados espacialmente. Quando as barras pretas e brancas possuem o mesmo tamanho, o algoritmo proposto é capaz de detectar todas as adulterações aplicadas ao vídeo. Quando as barras pretas possuírem tamanho maior que as brancas o algoritmo apresenta falsos positivos. Caso contrário, quando as barras brancas apresentarem tamanho maior que as barras pretas, o algoritmo apresenta falsos negativos.



(a)

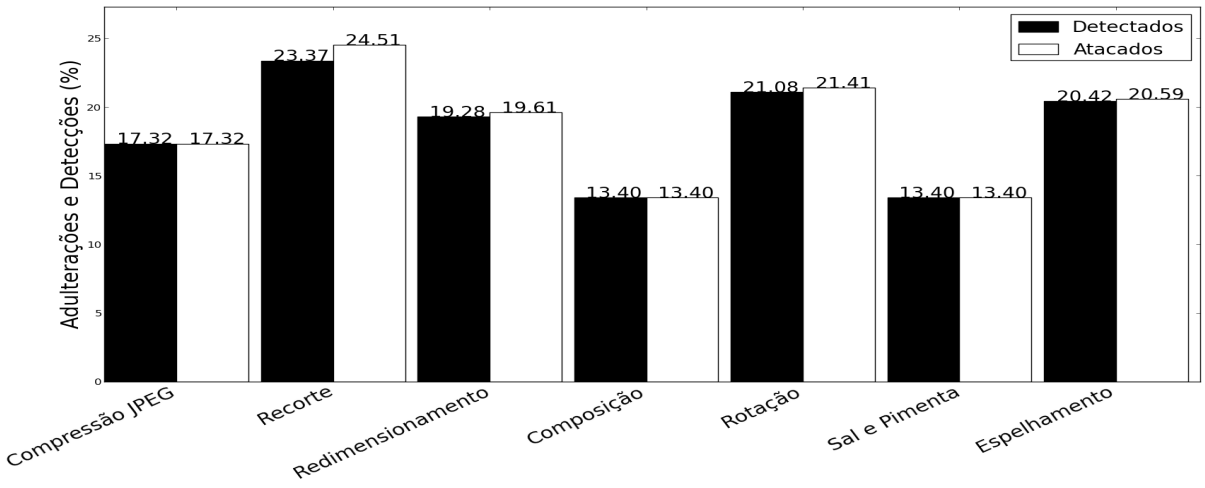


(b)

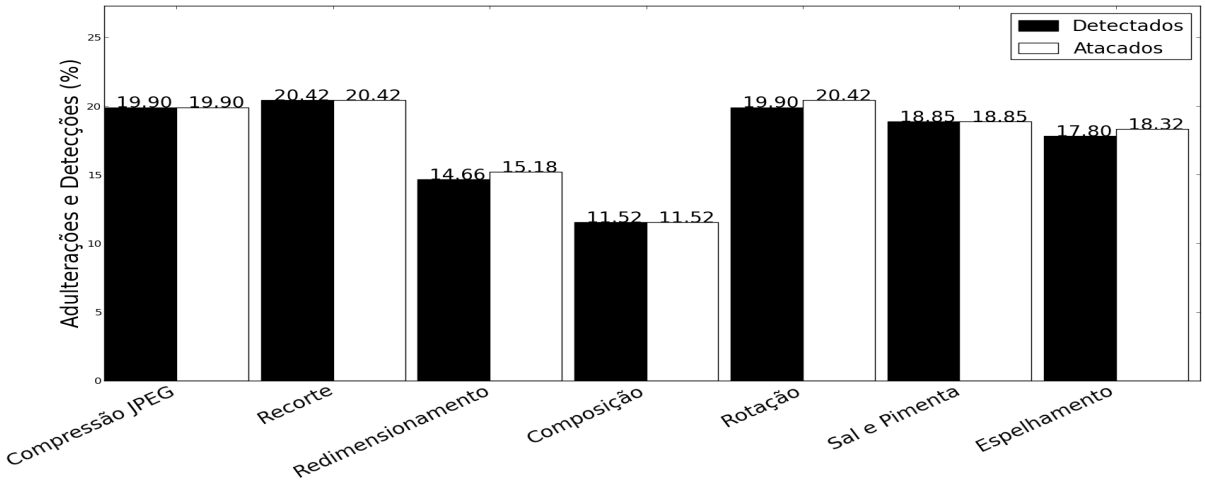


(c)

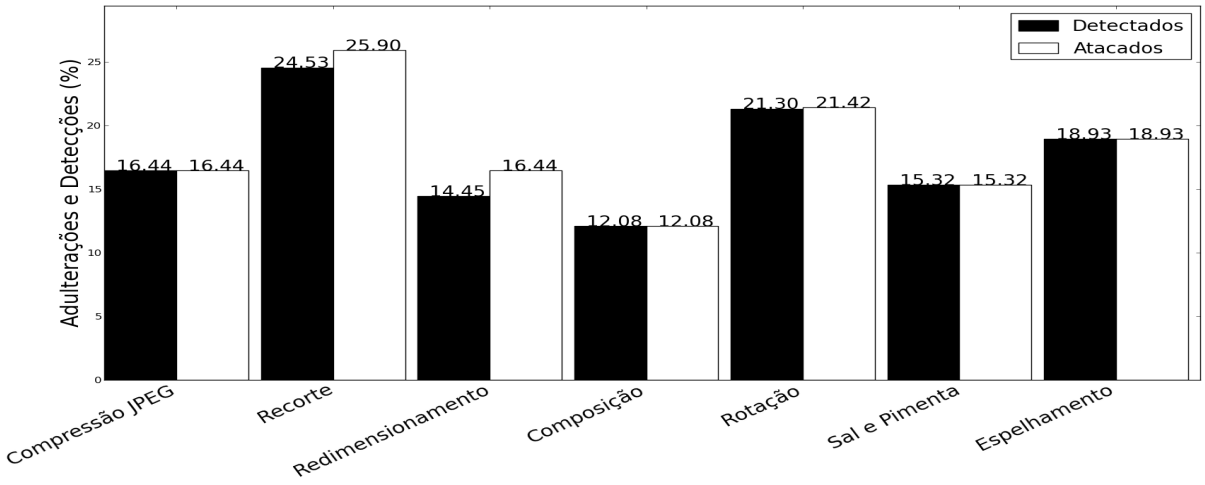
Figura 5.9: Porcentagem das detecções sobre as adulterações espaciais dos vídeos ‘Canoe’ (a), ‘Diver’ (b) e ‘Coral’ (c).



(d)

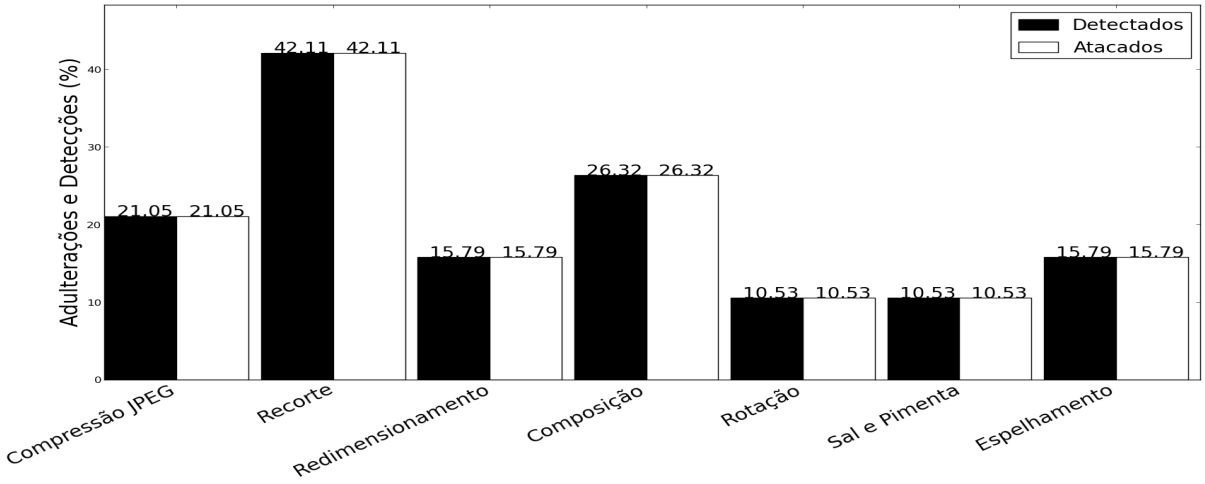


(e)

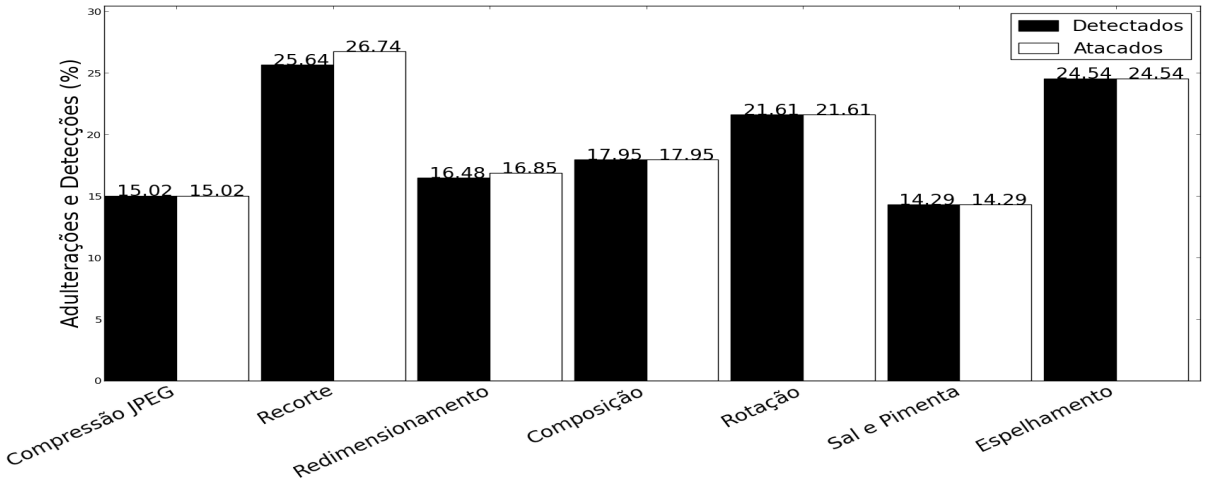


(f)

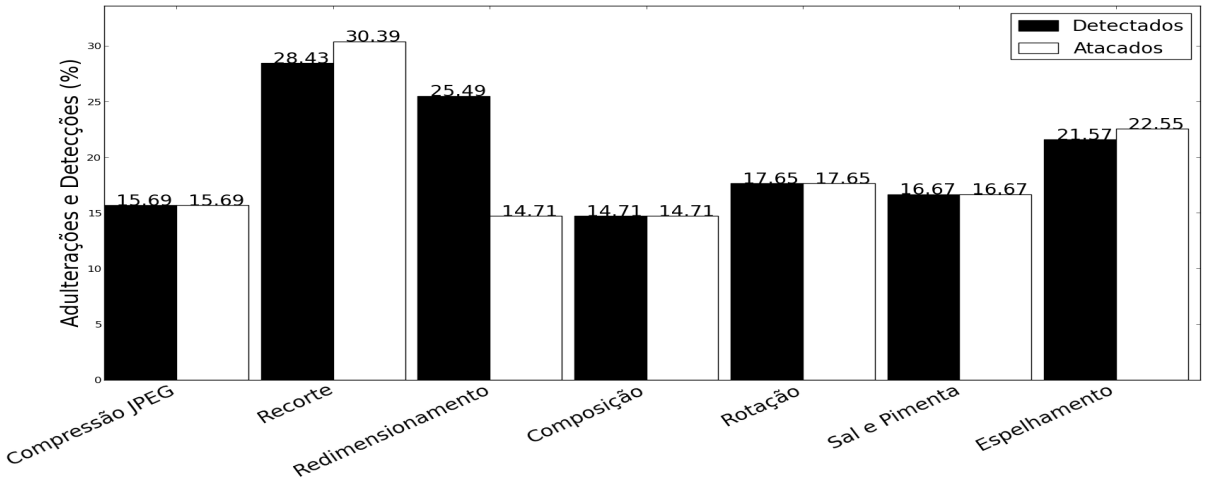
Figura 5.10: Porcentagem das detecções sobre as adulterações espaciais dos vídeos ‘Fish’ (d), ‘Seaweed’ (e) e ‘Beach’ (f).



(g)

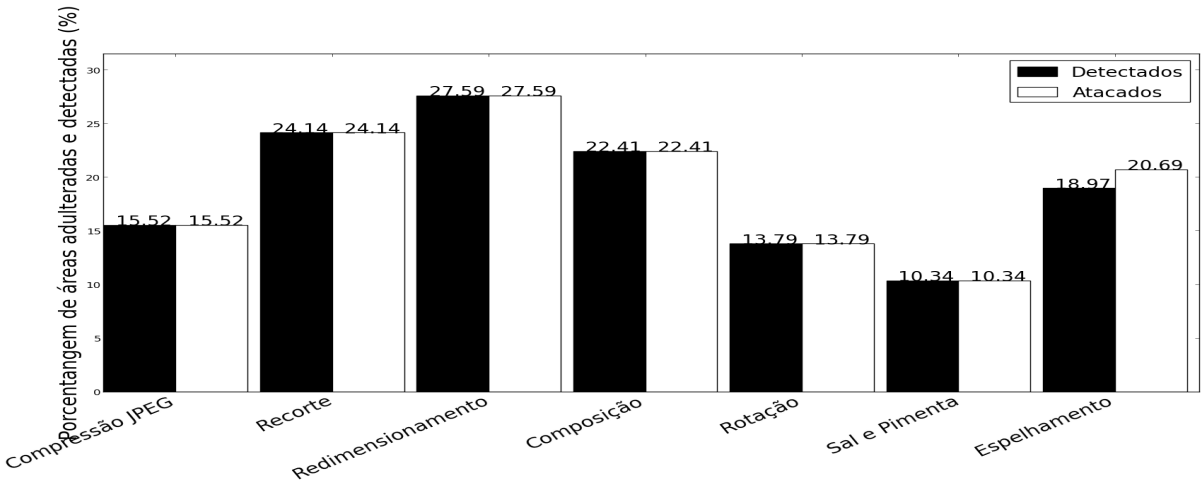


(h)

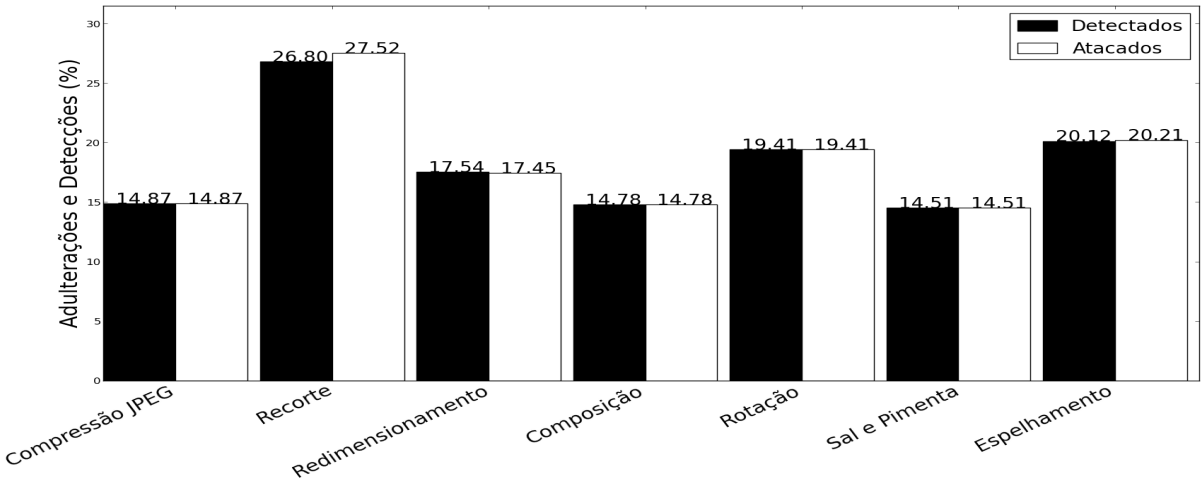


(i)

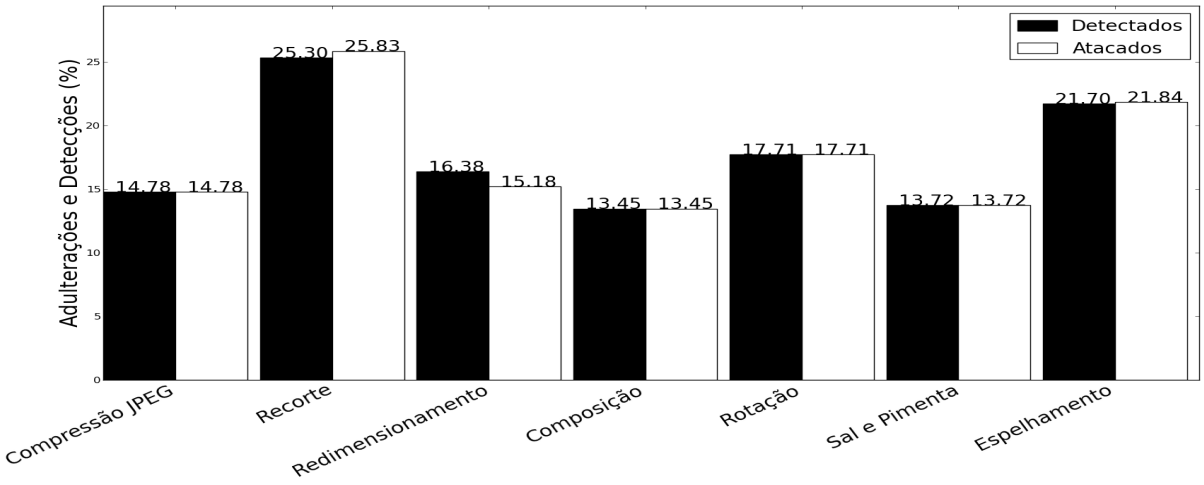
Figura 5.11: Porcentagem das detecções sobre as adulterações espaciais dos vídeos 'Rock' (g), 'Sky' (h) e 'Birds' (i).



(j)

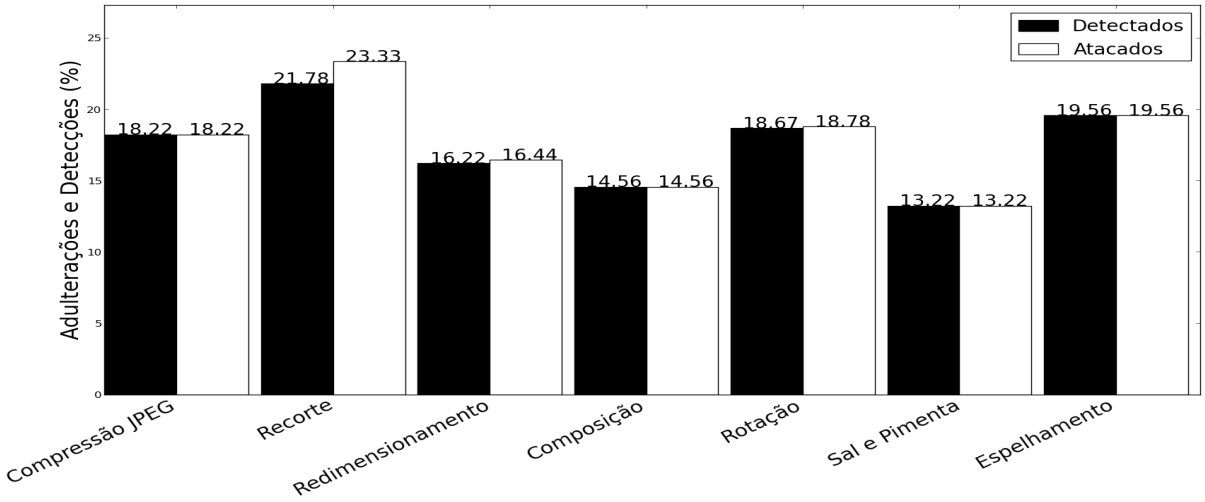


(l)

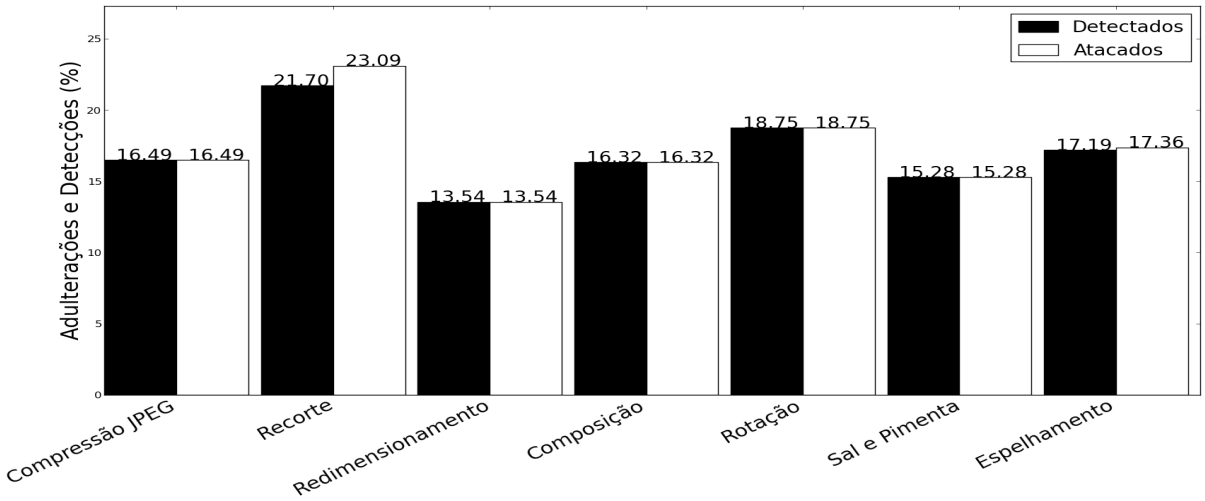


(m)

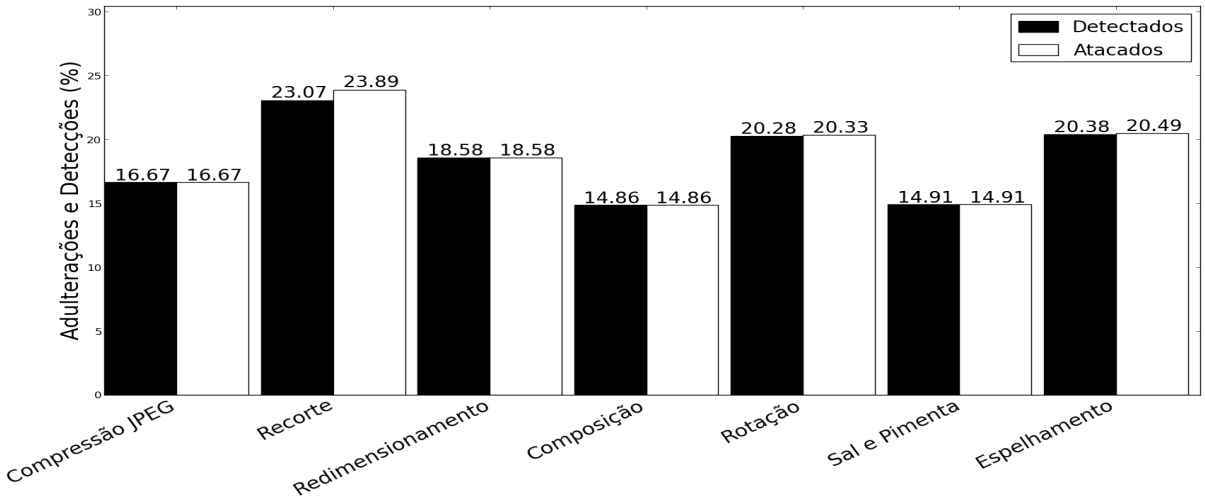
Figura 5.12: Porcentagem das detecções sobre as adulterações espaciais dos vídeos 'Deepsea' (j), 'Aquamarine' (l) e 'Rocks' (m).



(n)



(o)



(p)

Figura 5.13: Porcentagem das detecções sobre as adulterações espaciais dos vídeos 'Bill' (n), 'Alga' (o) e 'Sunset' (p).

As figuras .5.14, 5.15 e 5.16 mostram a quantidade de quadros adulterados e detectados temporalmente para todos os vídeos. Do mesmo modo, as barras brancas correspondem à quantidade de quadros atacados para cada vídeo e as barras pretas à quantidade de quadros detectados como adulterados pelo algoritmo. O algoritmo proposto foi capaz de detectar 100% das adulterações de Redução e Duplicação para a maioria dos vídeos. Note que para a adulteração por Embaralhamento existe uma pequena taxa de falso-negativos. Essa taxa é devido a alguns quadros retornarem à sua posição original durante o ataque. Com isso, a detecção falha quando comparada ao número total de pares de quadros reposicionados durante a adulteração.

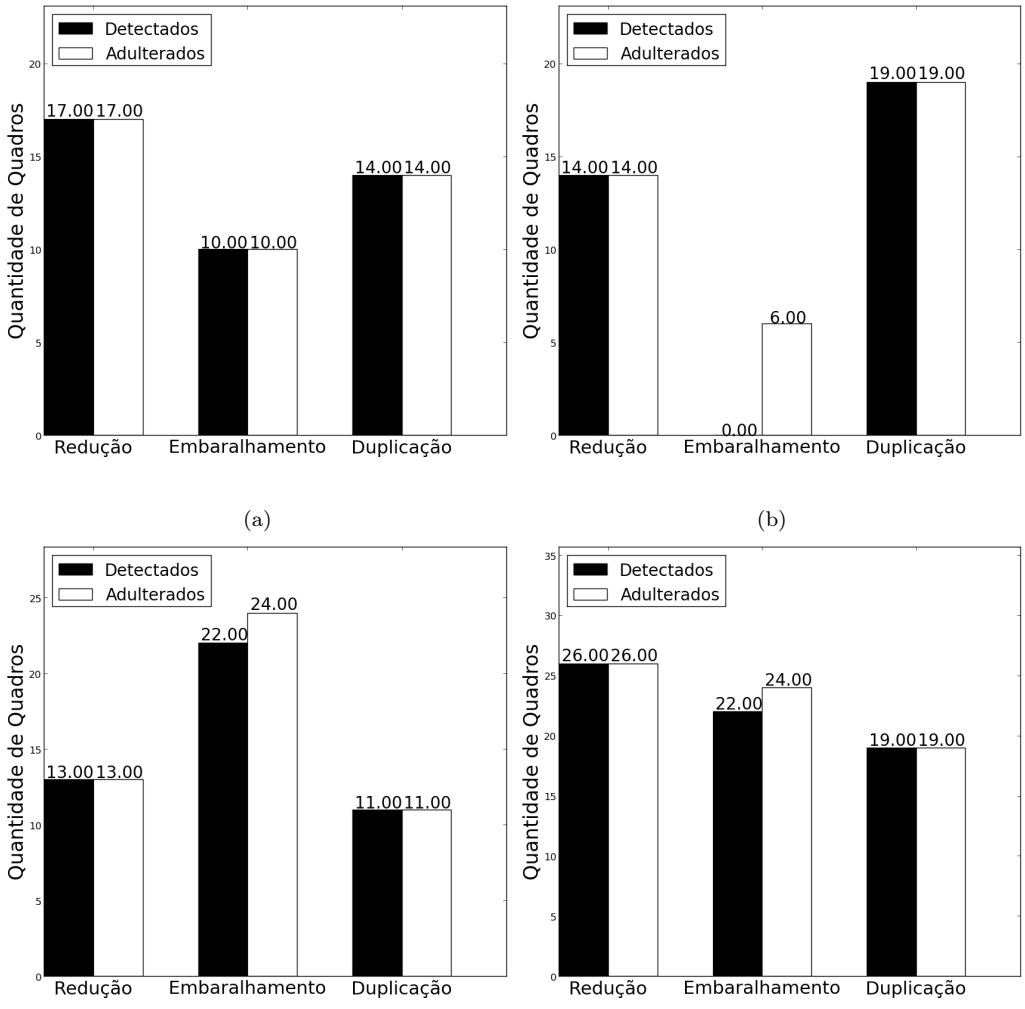


Figura 5.14: Comparação entre a quantidade de quadros adulteracos e detectados temporalmente para os vídeos: 'Canoe' (a), 'Diver' (b), 'Coral' (c), 'Fish' (d).

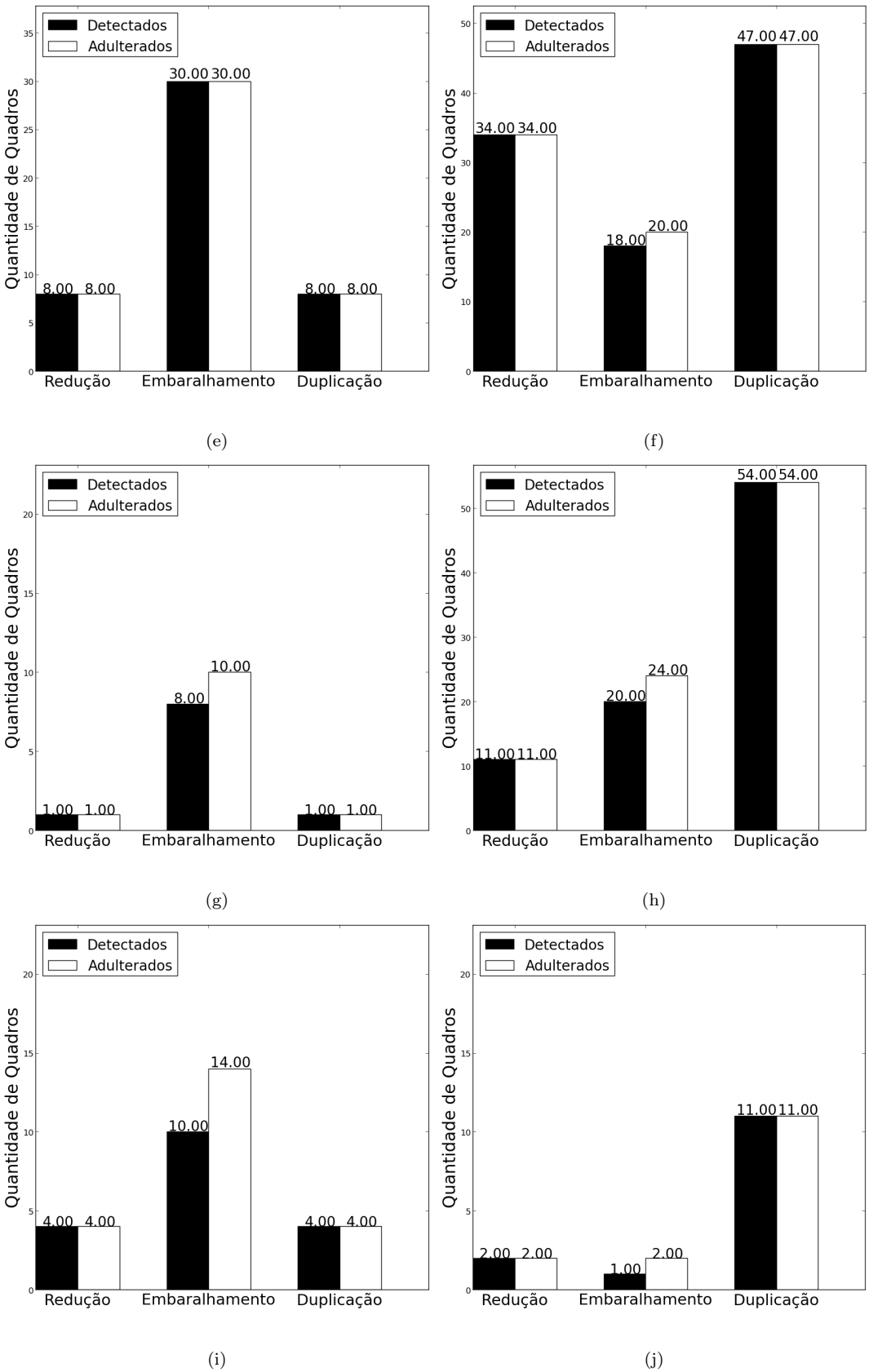
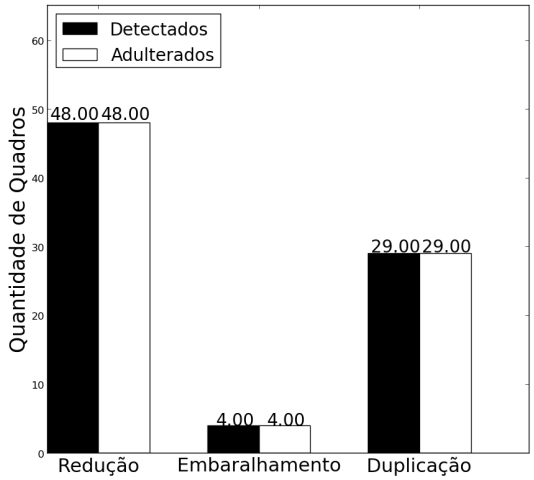
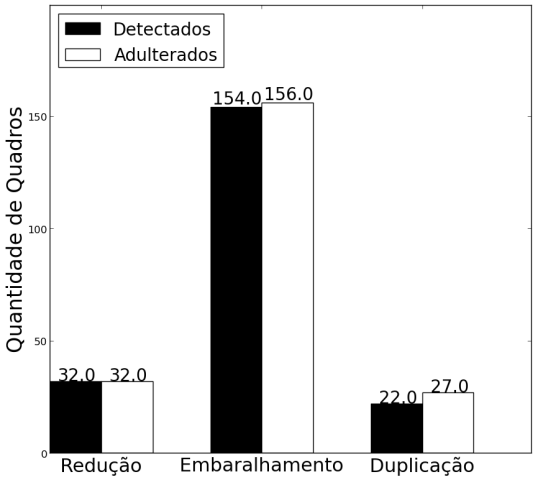


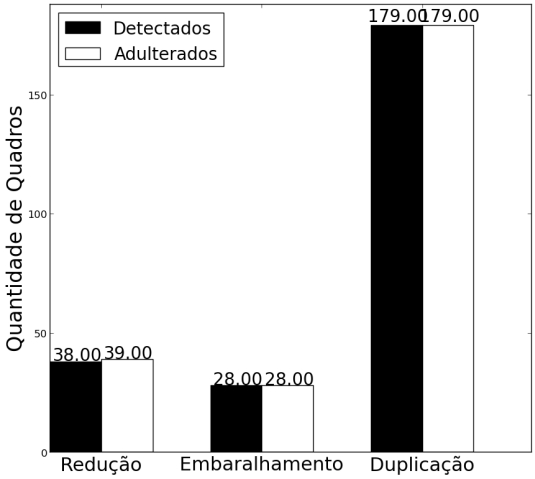
Figura 5.15: Comparação entre a quantidade de quadros adulteracos e detectados temporalmente para os vídeos: 'Seaweed' (e), 'Beach' (f), 'Rock' (g), 'Sky' (h), 'Coral' (i) e 'Fish' (j).



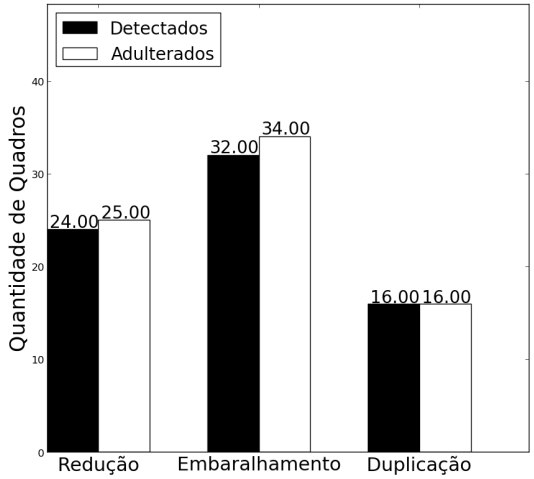
(l)



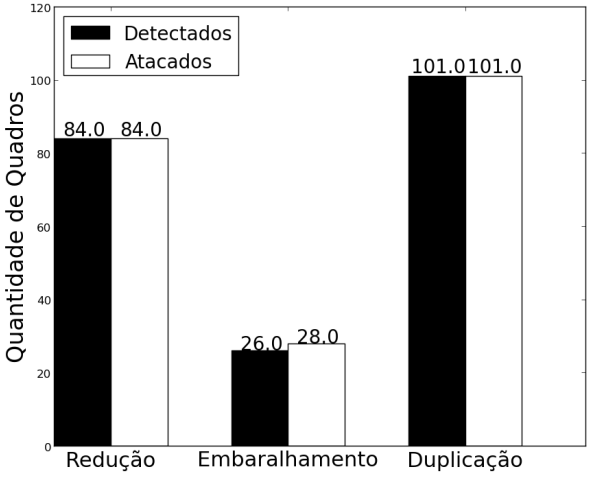
(m)



(n)



(o)



(p)

Figura 5.16: Comparação entre a quantidade de quadros adulteracos e detectados temporalmente para os vídeos: 'Seaweed' (l), 'Beach' (m), 'Bill' (n), 'Alga' (o) e 'Sunset' (p).

Nossa terceira análise é realizada sobre a medida da eficiência das detecções sobre as adulterações. A medida da eficiência é a porcentagem da razão entre a quantidade de adulterações e a quantidade de detecções. A Tabela 5.2 mostra a eficiência de sete ataques espaciais para todos os vídeos testados, enquanto que a Tabela 5.3 exibe a média de falsos positivos para todas as adulterações espaciais testadas. Em suma, essas tabelas exibem a porcentagem de quadros que foram adulterados e o algoritmo não foi capaz de detectar.

Nos resultados da Tabela 5.2 notamos que o algoritmo proposto possui aproximadamente 96% de sucesso nas detecções para a adulteração por Composição. A média de 4% de falsos positivos ocorre porque a imagem inserida na adulteração é sempre do mesmo tamanho. As adulterações por Espelhamento, Rotação e Redimensionamento possuem uma média de 12% de falso-negativos, uma vez que são ataques globais e a marca d'água é perdida de maneira semelhante para os três ataques. Em outras palavras, ataques globais danificam a marca d'água em sua totalidade e, devido a alta localização da marca d'água no pixel, é possível extrair uma informação igual à inserida. Isso contribui para o percentual de 12% de falsos negativos. Para Compressão JPEG, percebe-se o fator de compressão utilizado varia entre 75% e 95% e, conforme já relatado, algumas altas frequências permanecem intactas, mantendo a marca d'água íntegra nestes pixels. Assim, a média de detecção é 93%. Como a adulteração de Sal e Pimenta consiste na aplicação deste ruído aleatoriamente no quadro algumas posições podem se sobrepor, o que contribui para a média de 4% de falsos-negativos. Uma situação similar ocorreu para a adulteração de Recorte, na qual se obteve uma média de 12% de falsos-negativos.

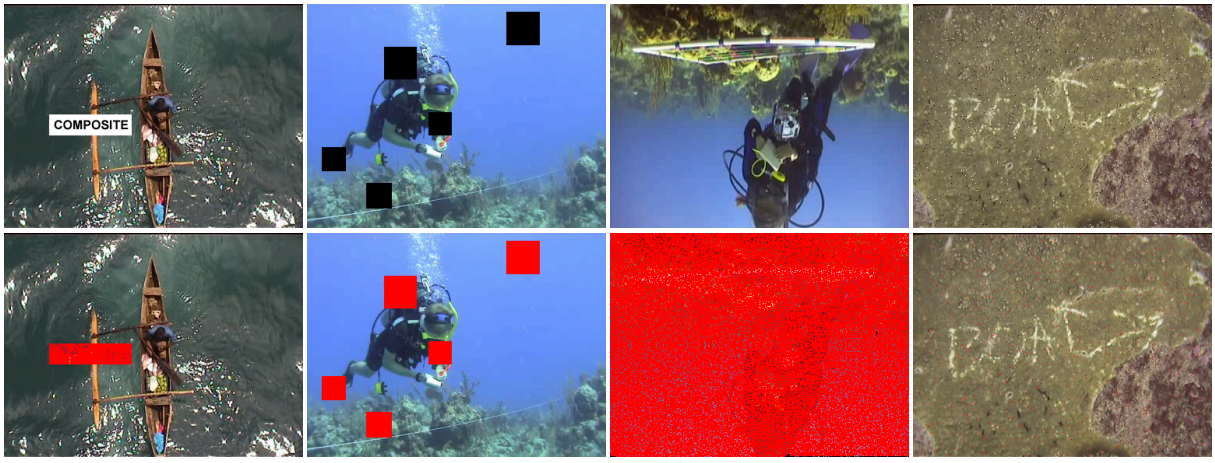
Vídeo	Composição	Espelhamento	JPEG	Rotação	Sal e Pimenta	Redimensionamento	Recorte
Canoe	96.57532	88.89811	93.65289	88.86159	94.37949	88.84412	88.88089
Diver	96.53922	88.79136	93.52155	88.85033	94.41239	88.85801	88.65209
Coral	96.66685	88.95969	93.20273	88.86600	94.38747	88.85448	88.74452
Fish	96.62966	88.90802	92.99444	88.86235	94.22194	88.83048	88.81851
Seaweed	96.53025	88.86941	93.49474	88.86328	94.52187	88.83860	88.77541
Beach	96.65479	88.88090	93.54141	88.86420	94.40788	88.84218	88.79031
Rock	96.82041	88.92413	93.22163	88.88834	94.61698	88.83916	86.91857
Sky	96.66859	88.88664	92.80452	88.86852	94.46368	88.86577	88.98801
Birds	96.59843	88.90845	93.60656	88.86725	94.24075	88.85656	88.94274
Deepsea	96.65206	88.87395	93.66885	88.86175	94.40545	88.80427	88.55688
Aquamarine	96.62299	88.81509	93.52120	88.85723	94.29038	88.83239	88.83305
Rocks	96.62892	88.81875	93.48457	88.86136	94.34785	88.79764	88.68016
Bill	96.68402	88.81955	92.89044	88.85585	94.37704	88.86138	88.71293
Alga	96.54274	88.88281	93.49796	88.87459	94.37877	88.86033	88.87534
Sunset	96.65064	88.84509	92.69742	88.86846	94.42169	88.86054	88.83518
Total	96.63099	88.87213	93.32006	88.86474	94.39158	88.84306	88.66697

Tabela 5.2: Porcentagem da eficiência média de detecções espaciais para todos os vídeos.

Vídeo	Composição	Espelhamento	JPEG	Rotação	Sal e Pimenta	Redimensionamento	Recorte
Canoe	3.42468	11.10189	6.34711	11.13841	5.62051	11.15588	11.11911
Diver	3.46078	11.20864	6.47845	11.14967	5.58761	11.14199	11.34791
Coral	3.33315	11.04031	6.79727	11.134	5.61253	11.14552	11.25548
Fish	3.37034	11.09198	7.00556	11.13765	5.77806	11.16952	11.18149
Seaweed	3.46975	11.13059	6.50526	11.13672	5.47813	11.1614	11.22459
Beach	3.34521	11.1191	6.45859	11.1358	5.59212	11.15782	11.20969
Rock	3.17959	11.07587	6.77837	11.11166	5.38302	11.16084	13.08143
Sky	3.33141	11.11336	7.19548	11.13148	5.53632	11.13423	11.01199
Birds	3.40157	11.09155	6.39344	11.13275	5.75925	11.14344	11.05726
Deepsea	3.34794	11.12605	6.33115	11.13825	5.59455	11.19573	11.44312
Aquamarine	3.37701	11.18491	6.4788	11.14277	5.70962	11.16761	11.16695
Rocks	3.37108	11.18125	6.51543	11.13864	5.65215	11.20236	11.31984
Bill	3.31598	11.18045	7.10956	11.14415	5.62296	11.13862	11.28707
Alga	3.45726	11.11719	6.50204	11.12541	5.62123	11.13967	11.12466
Sunset	3.34936	11.15491	7.30258	11.13154	5.57831	11.13946	11.16482
Total	3.36901	11.12787	6.67994	11.13526	5.60842	11.15694	11.33303

Tabela 5.3: Porcentagem da média de falsos negativos da detecções de adulterações espaciais para todos os vídeos.

A Figura 5.17 mostra quatro adulterações espaciais e suas detecções. Nesta figura, as marcas em vermelho correspondem a regiões detectadas como adulteradas. A primeira linha exibe as adulterações por Composição, Recorte, Espelhamento e Sal e Pimenta. A segunda linha exibe as suas respectivas detecções. Assim, a Figura 5.17-(a) mostra o ataque por Composição e sua detecção para um quadro do vídeo "Canoe". A Figura 5.17-(b) exibe a adulteração por Recorte e sua detecção para um quadro do vídeo "Diver". A Figura 5.17-(c) mostra um exemplo de Espelhamento e sua detecção para o mesmo vídeo. Para este ataque, a maior parte do quadro foi identificada como adulterada, uma vez que a maior parte da marca d'água foi alterada. Finalmente, Figura 5.17-(d), a exibe o ataque de ruído Sal e Pimenta e sua detecção.



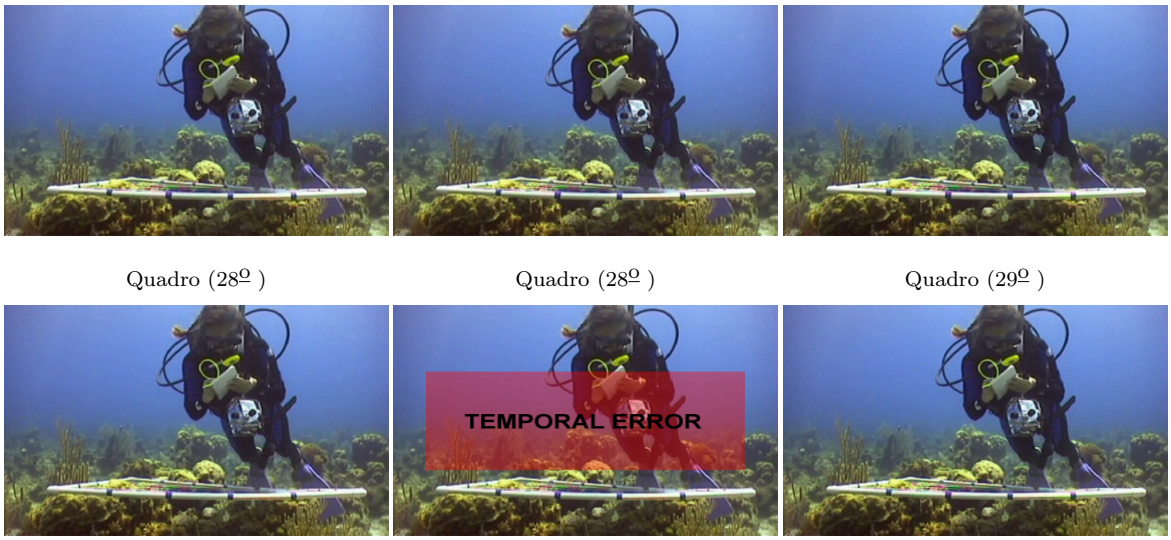
Composição (a) Recorte (b) Espelhamento (c) Sal e Pimenta (d)

Figura 5.17: Ataques espaciais (primeira linha) e as respectivas detecções (segunda linha) de 4 vídeos testados.

No próximo passo, nós analisamos os resultados dos três tipos de adulterações temporais. Para comparar todas as adulterações na linha do tempo do vídeo, o algoritmo requer um *buffer*. Então, primeiro nós efetuamos a leitura de todos os quadros do vídeo para a memória e, em seguida, efetuamos a análise conforme descrito no Seção 4.2.4 do Capítulo 4. Os resultados dos testes são exibidos nas Figuras 5.18, 5.19 e 5.20. A Figura 5.18 ilustra a adulteração temporal de Duplicação de Quadros do vídeo 'Coral'. Na primeira linha da figura, o quadro 28º foi copiado e colado entre as posições 28 e 29. O resultado da detecção é exibido na segunda linha.

A Figura 5.19 mostra o resultado da detecção da adulteração por Embaralhamento de Quadros para o vídeo 'Sunset'. Na primeira linha, os 163º e 529º quadros foram trocados de posições. Na segunda linha da figura, é apresentada sua detecção. O algoritmo foi capaz de detectar a transposição do par de quadros e identificou suas posições originais.

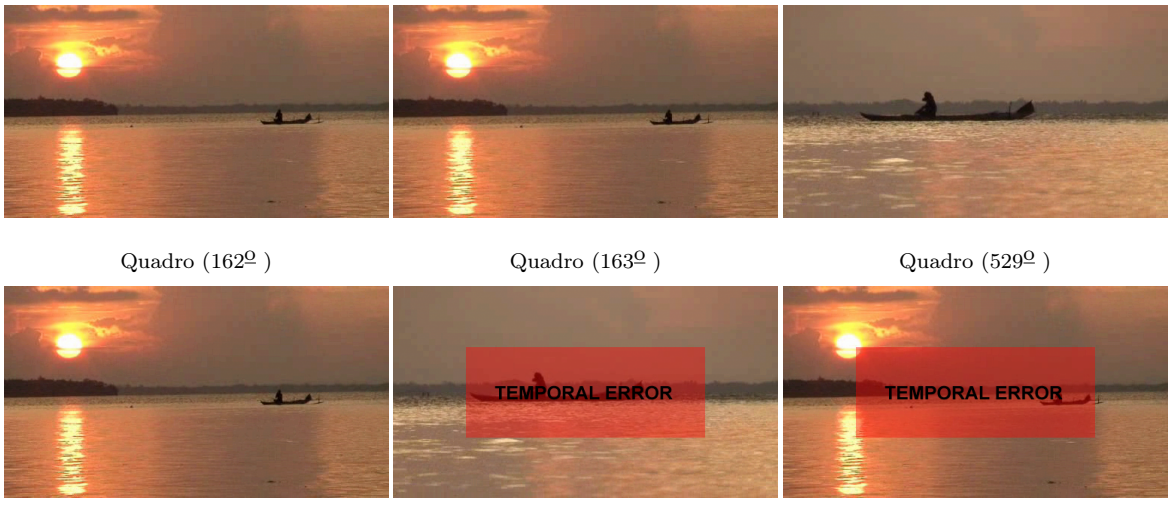
A última adulteração temporal testada é por Redução de Quadros, conforme exibido na Figura 5.20. Nesta ataque o algoritmo foi capaz de detectar a remoção do 24º quadro.



Quadro (28º) Quadro (28º) Quadro (29º)

Quadro (28º) Detecção da Duplicação (28º) Quadro (29º)

Figura 5.18: Adulteração por Duplicação de Quadros e sua detecção para o vídeo 'Coral'.



Quadro (162º) Quadro (163º) Quadro (529º)

Quadro (162º) Embaralhamento Quadro (529º) Embaralhamento Quadro (163º)

Figura 5.19: Adulteração temporal por Embaralhamento de Quadros para o vídeo 'Sunset' e sua detecção.

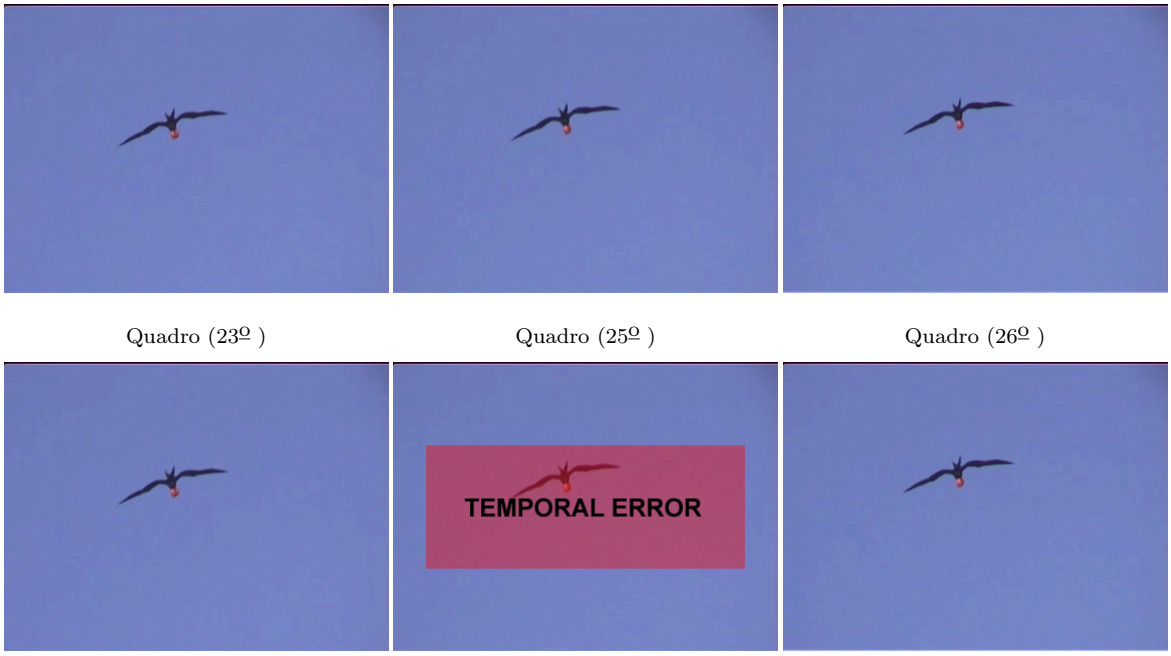


Figura 5.20: Adulteração por Redução de Quadros e sua detecção para o vídeo 'Sky'.

A média da eficiência e acurácia da detecção espacial é exibida na Tabela 5.2 para todos os vídeos. Para a adulteração por Composição, a média total para todos os vídeos é de 96.63% de sucesso. Esta média é devido ao tamanho da imagem (logo da UnB) utilizada em posições aleatórias dos quadros. Para as adulterações por Espelhamento, Redimensionamento e Recorte manteve-se média de 88% de sucesso devido a todos as três adulterações serem globais. A perda da marca d'água é semelhante para todas as três adulterações (conforme já relatado). Para a adulteração por Compressão JPEG obteve-se a média de 93%. Para a adulteração por adição de ruído Sal e Pimenta obteve-se 94% de sucesso. E finalmente, para a adulteração por Recorte obteve-se média de 88%.

Enquanto que a Tabela 5.4 mostra a média da eficiência para os três tipos de adulterações temporais, a Tabela 5.5 apresenta a média de falsos negativos para estas adulterações. Para o Embaralhamento de Quadros obteve-se média de 90.5% devido ao problema de sobreposição de quadros. No entanto, para as adulterações de Redução e Duplicação de Quadros obteve-se taxas de 99.5% de sucesso para todos os vídeos testados. No último caso, isso ocorre devido ao fato da inserção ou remoção de quadros não introduzir erros.

Vídeo	Embaralhamento	Redução	Duplicação
Canoe	100.0	100.0	100.0
Diver	83.3	100.0	100.0
Coral	91.6	100.0	100.0
Fish	91.6	100.0	94.7
Seaweed	100.0	100.0	100.0
Beach	90.0	100.0	100.0
Rock	80.0	100.0	100.0
Sky	63.1	100.0	100.0
Birds	71.4	100.0	100.0
Deepsea	100.0	100.0	100.0
Aquamarine	100.0	100.0	100.0
Rocks	100.0	100.0	100.0
Bill	100.0	97.4	100.0
Alga	94.1	96.0	100.0
Sunset	92.8	100.0	100.0
Total	90.5	99.5	99.6

Tabela 5.4: Porcentagem da média da eficiência da detecção de adulterações temporais para todos os vídeos.

Vídeo	Embaralhamento	Redução	Duplicação
Canoe	0.0	0.0	0.0
Diver	16.7	0.0	0.0
Coral	8.4	0.0	0.0
Fish	8.4	0.0	5.3
Seaweed	0.0	0.0	0.0
Beach	10.0	0.0	0.0
Rock	20.0	0.0	100.0
Sky	36.9	0.0	0.0
Birds	28.6	0.0	0.0
Deepsea	0.0	0.0	0.0
Aquamarine	0.0	0.0	0.0
Rocks	0.0	0.0	0.0
Bill	0.0	2.6	0.0
Alga	3.9	4.0	0.0
Sunset	7.2	0.0	0.0
Total	9.5	0.5	0.4

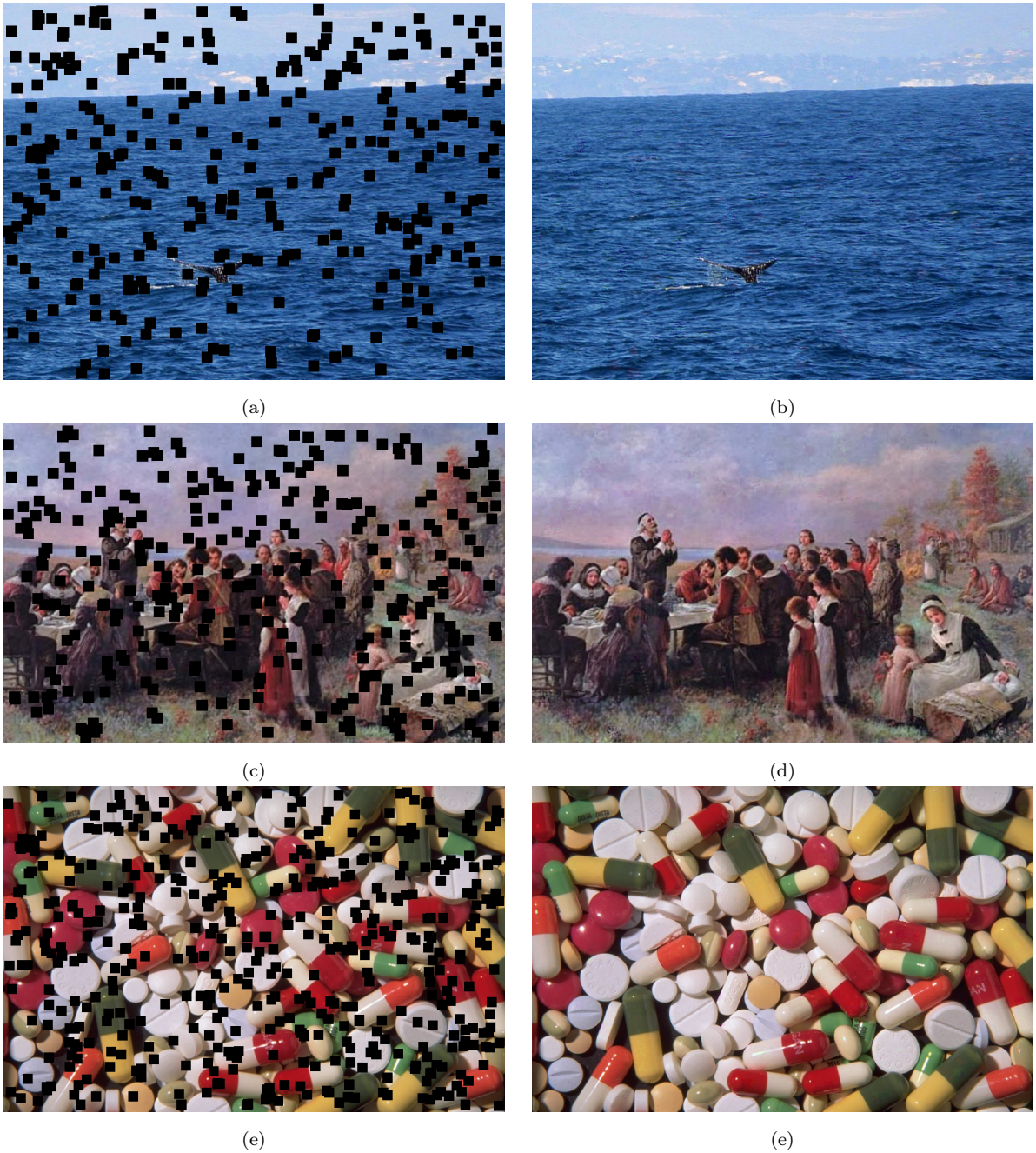
Tabela 5.5: Porcentagem da média de falsos negativos de detecções temporais.

5.2.4 Outras Aplicações: Mitigação de Erros

Considerando a alta capacidade de inserção do algoritmo proposto e a baixa degradação causada, foi proposto em conjunto com *Garcia* [40] uma técnica de mitigação de erros em vídeos e imagens. Esta proposta utiliza o algoritmo proposto neste trabalho e uma técnica Inversa de Meio-tons (do inglês, *Halftoning Technique*). A técnica Inversa de Meio-tons gera uma versão binária da imagem original. A versão binária é utilizada como marca d'água que é inserida na imagem original. Posteriormente, a informação é extraída e utilizada para restaurar as áreas adulteradas ou perdidas. A técnica é capaz de detectar e restaurar o conteúdo perdido ou adulterado com uma qualidade próxima ao original.

Os primeiros testes foram realizados com imagens estáticas. A Figura 5.21 exibe três

exemplos de restauração do conteúdo adulterado. Na primeira coluna são apresentadas três imagens 5.21 ((a), (c) e (e)) marcadas e adulteradas. Na segunda coluna, são apresentadas as imagens restauradas utilizando a técnica proposta por Garcia *et al.* e o algoritmo proposto neste trabalho. Em seguida efetuamos testes para vídeos sem áudio. A idéia básica é detectar adulterações ou perda de pacotes e restaurá-las.

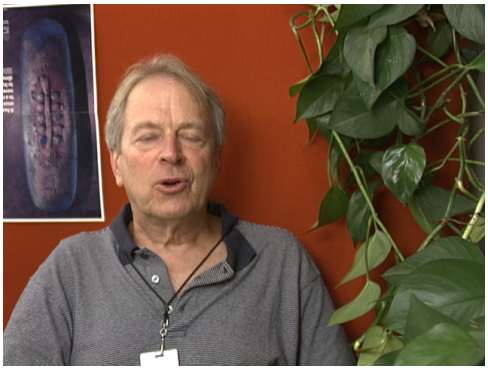


(a) (b)

(c) (d)

(e) (e)

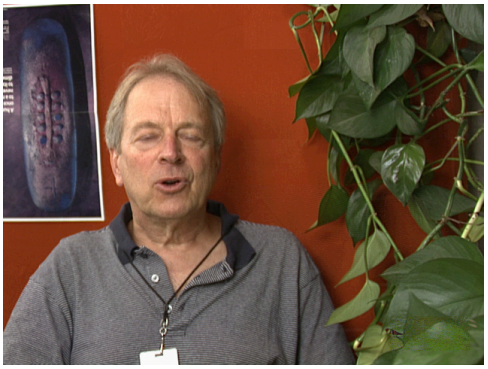
Figura 5.21: Adulteração e restauração para as imagens 'Whalelost', 'Thanks Giving' e 'Pills'.



(a) Quadro original do vídeo “NTIA cat joke”.

(b) Composição com adição do logotipo da UNB.

(c) Áreas detectadas como adulteradas do quadro.



(d) Versão restaurada do quadro do vídeo.

Figura 5.22: Exemplo de restauração de erros/adulterações no vídeo “NTIA cat joke”. Fonte: *http://www.cdvl.org.*

Nas Figuras 5.22,5.23 e 5.24 são exibidos três exemplos de detecção e restauração dos vídeos “NTIA cat joke”, “Container” e “Akiyo”. Na Fig 5.22(a) é apresentado o quadro original do vídeo “NTIA cat joke”, enquando que na Figura 5.22(b) sua versão adulterada por composição, onde o logotimo da Universidade de Brasília(UnB) é adicionado ao quadro do vídeo. Na Figura 5.22(c) são apresentadas as áreas detectadas como adulteradas pelo algoritmo. Na Figura 5.22(d) é apresentado o quadro do vídeo com as áreas adulteradas restauradas. Na Figura 5.23 é apresentado outro exemplo de funcionamento do algoritmo proposto utilizando o vídeo “Container”. Na Figura 5.23(a) é apresentado o quadro original do vídeo, enquanto que na Figura 5.23(b) é apresentado sua versão adulterada com o borramento de um quadro ao centro da imagem. Na Figura 5.23(c) são apresentadas as áreas detectadas como adulteradas. Figura 5.23(d) é apresentado o quadro do vídeo com as áreas restauradas. Finalmente, na Figura 5.24 é apresentado o último exemplo para o vídeo “Akiyo”. Na Figura 5.24(a) é apresentado o quadro original. Na Figura 5.24(b) a

(a) Quadro original do vídeo “Container”.

(b) Borramento de uma região específica.

(c) Áreas detectadas como adulteradas do quadro.

(d) Versão restaurada do quadro do vídeo.

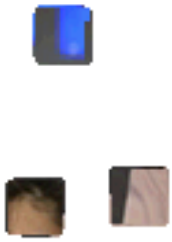
Figura 5.23: Exemplo de restauração de erros/adulterações no vídeo “Container”. Fonte: *http://trace.eas.asu.edu/yuv*.

versão adulterada onde três quadros de tamanhos aleatórios foram copiados e trocados de local. Na Figura 5.24(c) são exibidas as áreas restauradas e na Figura 5.24(d) o quadro restaurado.

Note que a restauração dos retângulos adulterados possui qualidade próxima ao conteúdo original e superior às técnicas disponíveis na literatura. Isso se deve ao fato de que a modificação do algoritmo QIM proposta permite a inserção de grande quantidade de informação. Esta capacidade aliada à restauração do conteúdo da técnica Inversa de Meio-tons permitem restaurar detectar e restaurar adulterações em vídeos e imagens mesmo se a perda do conteúdo atingir níveis próximos a 25% [40].

(a) Quadro original do vídeo "Container".

(b) Borramento de uma região específica.



(c) Áreas detectadas como adulteradas do quadro.

(d) Versão restaurada do quadro do vídeo.

Figura 5.24: Exemplo de restauração de erros/adulterações no vídeo "Akiyo". Fonte: *http://trace.eas.asu.edu/yuv*.

Capítulo 6

Conclusões e Trabalhos Futuros

6.1 Conclusões e Trabalhos Futuros

Nesta dissertação foi proposto um algoritmo para detecção de adulterações em vídeos digitais utilizando técnicas de marca d'água. O trabalho é baseado em um algoritmo de marca d'água simples e com baixa degradação do vídeo marcado. O algoritmo permite identificar adulterações locais, globais e temporais na granularidade de pixel.

A combinação das marcas temporal e espacial permite aumentar a sensibilidade e robustez do algoritmo, quando comparado com outras técnicas presentes na literatura. A replicação da marca d'água temporal no canal de áudio e vídeo permite identificar adulterações temporais mesmo se o conteúdo total do quadro do vídeo for adulterado ou perdido. O algoritmo proposto também é capaz de estimar o tipo da adulteração temporal analisando as características de um vetor de adulterações do vídeo.

O método proposto utiliza apenas uma técnica de marca d'água simples e robusta. Isso o diferencia da maioria das técnicas presentes na literatura, as quais, ao que tange nosso conhecimento, são capazes de detectar baixa gama de adulteração.

O algoritmo proposto não degrada visualmente o vídeo marcado, conforme pode ser comprovado pelos altos valores de PSNR e SSIM. A média de falsos positivos para adulterações espaciais apresentada pelo sistema é de 8.61% enquanto que para adulterações

temporais é de 3.46%.

Por fim, o algoritmo tem um bom desempenho, apresentando uma alta precisão, eficiência e uma baixa taxa de falsos positivos.

6.2 Contribuições

As contribuições deste trabalho são:

- Algoritmo capaz de detectar adulterações espaciais e temporais em granularidade de pixel, com baixo custo computacional, robusto e expansível;
- Detecção das três maiores classes de adulterações utilizando apenas uma técnica;
- Modificação do algoritmo QIM para inserir mais informações sem degradar visivelmente o vídeo marcado;
- Estimação do tipo de ataque sofrido pela combinação de duas marcas d'água (espacial e temporal).

6.3 Trabalhos Futuros

Estudos futuros podem concentrar-se na ampliação da gama de ataques suportados pelo algoritmo, ampliação da proteção no canal de áudio e otimização do modelo de geração de marca d'água. Vale a pena enfatizar que algoritmo proposto é extensível a novas adulterações ou ataques.

Outra possível linha de pesquisa pode incluir a proteção do canal do áudio. A ideia consiste em utilizar uma marca d'água no canal de áudio de modo protegê-lo. A marca deverá ser redundante o suficiente para não comprometer detecção temporal do vídeo.

Uma outra possível linha de pesquisa pode concentrar-se na otimização do modelo de geração de marca d'água. Neste caso, o objetivo é que a informação marca d'água contenha informações semânticas além de dados binários. Desta forma, seria possível obter uma melhor classificação dos ataques temporais.

6.4 Publicações

A partir deste trabalho, foram elaborados 2 artigos, um aceito e outro submetido nas seguintes conferências:

- *Error Concealment Using a Halftone Watermarking Technique: Full Paper* aceito para a Conference on Graphics, Patterns and Images (SIBGRAPI) realizada em Ouro Preto, Brasil em 2012 (SIBGRAPI/2012).
- *A Novel Approach for Digital Video Tampering Detection: Full Paper* submetido para a *Conference on Graphics, Patterns and Images* (SIBGRAPI) que será realizada em Arequipa, Peru em 2013 (SIBGRAPI/2013).

Referências

[1] EI Lin, A.M. Eskicioglu, R.L. Lagendijk, and E.J. Delp. Advances in digital video content protection. *Proceedings of the IEEE*, 93(1):171–183, 2005. 1, 6, 7, 23

[2] F. Hartung. *Digital watermaking and fingerprinting of uncompressed and compressed video*. PhD thesis, Universitat, Erlangen, Nurnberg, 2000. 1

[3] R. Parloff. Morpheus falling? *Spectrum, IEEE*, 40(12):18–19, 2003. 2

[4] P. Biddle, P. England, M. Peinado, and B. Willman. The darknet and the future of content distribution. In *ACM Workshop on Digital Rights Management*, volume 6, page 54, 2002. 2

[5] Weihong Wang. *Digital video forensics*. PhD thesis, Citeseer, 2009. 2, 20, 21

[6] E.T. Lin et al. Video and image watermark synchronization. 2005. 2, 6, 18, 19, 20, 21

[7] D. Cross and B.G. Mobasseri. Watermarking for self-authentication of compressed video. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 2, pages II–913 – II–916 vol.2, 2002. 2, 21

[8] Chen Xiaoling and Zhao Huimin. A novel video content authentication algorithm combined semi-fragile watermarking with compressive sensing. In *Intelligent System Design and Engineering Application (ISDEA), 2012 Second International Conference on*, pages 134 –137, jan. 2012. 2, 21

[9] Hou Zhi-yu and Tang Xiang-hong. Integrity authentication scheme of color video based on the fragile watermarking. In *Electronics, Communications and Control (ICECC), 2011 International Conference on*, pages 4354 –4358, sept. 2011. 2, 20, 21

[10] J. Zhang and A.T.S. Ho. An efficient authentication method for h.264/avc. In *Integration of Knowledge, Semantics and Digital Media Technology, 2005. EWIMT 2005. The 2nd European Workshop on the (Ref. No. 2005/11099)*, pages 157 –164, 30 2005-dec. 1 2005. 5

[11] J.A. Bloom, I.J. Cox, T. Kalker, J.P.M.G. Linnartz, M.L. Miller, and C.B.S. Traw. Copy protection for dvd video. *Proceedings of the IEEE*, 87(7):1267–1276, 1999. 5, 6

[12] M. Johnson, P. Ishwar, V. Prabhakaran, D. Schonberg, and K. Ramchandran. On compressing encrypted data. *Signal Processing, IEEE Transactions on*, 52(10):2992–3006, 2004. 7

[13] A.E. Sale. Lorenz and colossus [military cryptography]. In *Computer Security Foundations Workshop, 2000. CSFW-13. Proceedings. 13th IEEE*, pages 216 –222, 2000. 7

[14] Ran Wang and Xijian Ping. Detection of resampling based on singular value decomposition. In *Image and Graphics, 2009. ICIG '09. Fifth International Conference on*, pages 879 –884, sept. 2009. 9, 10

[15] Gang Cao, Yao Zhao, and Rongrong Ni. Image composition detection using object-based color consistency. In *Signal Processing, 2008. ICSP 2008. 9th International Conference on*, pages 1186 –1189, oct. 2008. 9

[16] Xiaofeng Wang, Xiaoni Zhang, Zhen Li, and Shangping Wang. A dwt-dct based passive forensics method for copy-move attacks. In *Multimedia Information Networking and Security (MINES), 2011 Third International Conference on*, pages 304 –308, nov. 2011. 11, 12

[17] M. Barni, A. Costanzo, and L. Sabatini. Identification of cut amp; paste tampering by means of double-jpeg detection and image segmentation. In *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, pages 1687 –1690, 30 2010-june 2 2010. 11, 12

[18] Sujoy Roy and Qibin Sun. Robust hash for detecting and localizing image tampering. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 6, pages VI –117 –VI –120, 16 2007-oct. 19 2007. 13

[19] Huajian Liu and M. Steinebach. Semi-fragile watermarking for image authentication with high tampering localization capability. In *Automated Production of Cross Media Content for Multi-Channel Distribution, 2006. AXMEDIS '06. Second International Conference on*, pages 143 –152, dec. 2006. 14

[20] I.J. Cox and J.-P.M.G. Linnartz. Some general methods for tampering with watermarks. *Selected Areas in Communications, IEEE Journal on*, 16(4):587 –593, may 1998. 16

[21] B.G. Mobasseri, M.J. Sieffert, and R.J. Simard. Content authentication and tamper detection in digital video. In *Image Processing, 2000. Proceedings. 2000 International Conference on*, volume 1, pages 458 –461 vol.1, 2000. 16

[22] Peng Yin and Hong H. Yu. Classification of video tampering methods and countermeasures using digital watermarking. pages 239–246, 2001. 16

[23] P. Thangavel and T. Kumaran. Fragile watermark for tamper detection using structural distortion measure. In *Industrial Electronics, 2007. ISIE 2007. IEEE International Symposium on*, pages 1755 –1760, june 2007. 21

[24] I.J. Cox, J. Kilian, F.T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *Image Processing, IEEE Transactions on*, 6(12):1673 –1687, dec 1997. 22, 24, 27

[25] M. Barni and F. Bartolini. Data hiding for fighting piracy. *Signal Processing Magazine, IEEE*, 21(2):28–39, 2004. 22

[26] M.D. Swanson, M. Kobayashi, and A.H. Tewfik. Multimedia data-embedding and watermarking technologies. *Proceedings of the IEEE*, 86(6):1064–1087, 1998. 22, 23, 24

[27] R.B. Wolfgang, C.I. Podilchuk, and E.J. Delp. Perceptual watermarks for digital images and video. *Proceedings of the IEEE*, 87(7):1108–1126, 1999. 22

[28] I. Cox, M. Miller, J. Bloom, and C. Honsinger. Digital watermarking. *Journal of Electronic Imaging*, 11(3):414–414, 2002. 23, 24

[29] R.J. Anderson and F.A.P. Petitcolas. On the limits of steganography. *Selected Areas in Communications, IEEE Journal on*, 16(4):474–481, 1998. 23, 24

[30] GC Langelaar. Conditional access to television service. *Wireless Commun. Interactive Multimedia CD-ROM*, 1999. 23

[31] R.B. Wolfgang and E.J. Delp III. Fragile watermarking using the vw2d watermark. In *Electronic Imaging'99*, pages 204–213. International Society for Optics and Photonics, 1999. 23

[32] R. Pickholtz, D. Schilling, and L. Milstein. Theory of spread-spectrum communications–a tutorial. *Communications, IEEE Transactions on*, 30(5):855 – 884, may 1982. 27

[33] A.B. Watson. Visually optimal dct quantization matrices for individual images. In *Data Compression Conference, 1993. DCC '93.*, pages 178 –187, 1993. 28

[34] J.N. Ellinas and D.E. Manolakis. A robust watermarking scheme based on edge detection and contrast sensitivity function. In *VISAPP Proc. Int. Conf. Computer Vision Theory and Applications*, 2007. 28

[35] Lino Coria, Panos Nasiopoulos, and Rabab Ward. A region-specific qim-based watermarking scheme for digital images. In *Broadband Multimedia Systems and Broadcasting, 2009. BMSB '09. IEEE International Symposium on*, pages 1 –6, may 2009. 30

[36] B. Chen and G.W. Wornell. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *Information Theory, IEEE Transactions on*, 47(4):1423–1443, 2001. 30, 31

[37] Zhihua Chen and Jin Xu. One-time-pads encryption in the tile assembly model. In *Bio-Inspired Computing: Theories and Applications, 2008. BICTA 2008. 3rd International Conference on*, pages 23 –30, 28 2008-oct. 1 2008. 35

[38] Peter Mumby. Reefvid: A resource of free coral reef video clips for educational use, April 2013. 46

[39] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 49

[40] Pedro Garcia Freitas, Ronaldo Rigoni, Mylene CQ Farias, and Aletela PF de Araujo. Error concealment using a halftone watermarking technique. In *Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on*, pages 308–315. IEEE, 2012. 70, 73