

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
DEPARTAMENTO ACADÊMICO DE INFORMÁTICA  
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO**

**SYLVIO ALEXANDRE BIASUZ BLOCK**

**Estudo de Assinaturas Digitais para Identificação de Vídeos**

**TRABALHO DE CONCLUSÃO DE CURSO**

**CURITIBA  
2015**

SYLVIO ALEXANDRE BIASUZ BLOCK

**Estudo de Assinaturas Digitais para Identificação de Vídeos**

Trabalho de conclusão de curso apresentado como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação do Departamento Acadêmico de Informática da Universidade Tecnológica Federal do Paraná.

Orientador: Prof. Dr. Rodrigo Minetto

Coorientador: Prof. Dr. Ricardo Dutra da Silva

CURITIBA  
2015

## **Agradecimentos**

Gostaria de agradecer a Deus e aos meus pais pelo apoio e amor incondicional. E também aos meus orientadores, Prof. Dr. Rodrigo Minetto e Prof. Dr. Ricardo Dutra da Silva pela orientação e confiança, sempre mostrando-se dispostos e colaborativos. O autor também gostaria de agradecer a equipe do projeto QoS-TREAM, de número 295220, FP7-MC-IRSES e também Universal-CNPq-Brazil, de número 444789/2014-6.

## Resumo

Block, Sylvio Alexandre Biasuz. Estudo de Assinaturas Digitais para Identificação de Vídeos. 33f. Trabalho de Conclusão de Curso – Departamento Acadêmico de Informática, Universidade Tecnológica Federal do Paraná. Curitiba, 2015.

Uma assinatura de vídeo é um descritor único extraído do conteúdo do vídeo para identificação e recuperação do mesmo. Neste trabalho é apresentado um estudo comparativo de três métodos, Hua *et. al.* [6], Lee and Yoo [11] e Cook [3] para geração de assinaturas de vídeos. Esses métodos utilizam características espaciais, como luminância e gradiente da imagem, e características temporais, como movimentação da cena e de objetos. Neste trabalho foi utilizada a base de vídeos LIVE Video Quality, juntamente com vídeos distorcidos pelo autor, para avaliar a capacidade dos métodos estudados na identificação de vídeos que possuem alterações produzidas por compressão, transmissão, transformações geométricas e outras distorções, intencionais ou não intencionais.

## Lista de Figuras

1	Exemplo de cópia de vídeo. O vídeo copiado (a) possui alteração de características quando comparado ao vídeo original (b). Fonte: Youtube (copyright Fox Broadcast Company). . . . .	7
2	Propriedades de vídeo. . . . .	9
3	Divisão de vídeo. . . . .	10
4	Computação do descritor de Hua <i>et. al.</i> [6]: (a) divisão do quadro em blocos; (b) nível de cinza médio em cada bloco; (c) o descritor $\mathbf{d} = \langle 7, 9, 8, 3, 6, 5, 1, 4, 2 \rangle$ é formado pela permutação que ordena os níveis de cinza dos blocos de (b). . . . .	15
5	Assinatura por distribuição de gradiente. Fonte: Lee, Sunil e Yoo, Chang D. (2008. P 984). Traduzido pelo Autor. . . . .	17
6	Quadros iniciais dos dez vídeos de referência. Fonte LIVE Video Quality (LIVE-VQD). . . . .	19
7	Quadros com distorções, da esquerda para a direita e de cima para baixo: blur (ofuscamento), adição de borda vermelha, inversão de cores, recorte central do quadro, espelhamento, compressão JPEG do quadro, rotação para a direita, adição de legenda e, por fim, adição de marca d'água. . . . .	20
8	Performance dos métodos: curvas de precisão-revocação (esquerda) e ROC (direita) para consultas com <b>25</b> quadros considerando a base de dados LIVE-VQD original. . . . .	22
9	Performance dos métodos: curvas de precisão-revocação (esquerda) e ROC (direita) para consultas com <b>25</b> quadros considerando a base de dados LIVE-VQD com quatorze distorções adicionais para cada vídeo de referência. . . .	22
10	Performance dos métodos: curvas de precisão-revocação (esquerda) e ROC (direita) para consultas com <b>100</b> quadros considerando a base de dados LIVE-VQD com quatorze distorções adicionais para cada vídeo de referência. . . .	23

11	Performance dos métodos: curvas de precisão-revocação (esquerda) e ROC (direita) para consultas com <b>200</b> quadros considerando a base de dados LIVE-VQD com quatorze distorções adicionais para cada vídeo de referência. . . . .	23
12	Distância $L_1$ entre vídeos de mesma referência com mínimo evidente. Onde as linhas horizontais representam as técnicas e a linha vertical o ponto de início da consulta. . . . .	24
13	Distância $L_1$ entre vídeos de diferentes referências sem mínimo evidente. Onde as linhas horizontais representam as técnicas e a linha vertical o ponto de início da consulta. . . . .	25
14	Distância $L_1$ entre vídeos de mesma referência sem mínimo evidente. Onde as linhas horizontais representam as técnicas e a linha vertical o ponto de início da consulta. . . . .	25

## **Lista de Tabelas**

1 Parâmetros dos algoritmos e informações das assinaturas . . . . . 20

# **Sumário**

<b>1</b>	<b>Introdução</b>	<b>6</b>
1.1	Objetivo Geral . . . . .	8
1.2	Objetivos Específicos . . . . .	8
1.3	Estrutura do Documento . . . . .	8
<b>2</b>	<b>Estado da Arte</b>	<b>9</b>
2.1	Definição de Quadro . . . . .	9
2.2	Definição de Vídeo . . . . .	9
2.3	Definição de Assinatura de Vídeo . . . . .	10
2.3.1	Características da Assinaturas de Vídeo . . . . .	10
2.4	Recuperação de Vídeos Baseada em Conteúdo . . . . .	11
2.5	Descritores . . . . .	12
2.5.1	Descritores Globais . . . . .	12
2.5.2	Descritores Locais . . . . .	12
2.6	Trabalhos Prévios . . . . .	13
<b>3</b>	<b>Algoritmos</b>	<b>15</b>
3.1	Assinatura de Vídeo por Distribuição de Intensidade . . . . .	15
3.2	Assinatura de Vídeo por Distribuição de Gradientes . . . . .	16
3.3	Assinatura de Vídeo por Diferença entre Quadros . . . . .	17
<b>4</b>	<b>Metodologia</b>	<b>18</b>
4.1	Base de Dados . . . . .	18
4.2	Experimentos . . . . .	18
4.3	Recursos de Hardware e Software . . . . .	26
<b>5</b>	<b>Conclusão</b>	<b>27</b>

## 1 Introdução

A disseminação de dispositivos móveis de gravação de vídeo e também a adição desta funcionalidade aos celulares contribuiu para um rápido crescimento da produção de vídeos [16]. O crescimento pode ser verificado em redes sociais e sites especializados, como é o caso do YouTube, que conta hoje com mais de 100 milhões de vídeos on-line. Estes vídeos podem ser visualizados cada vez mais e por mais pessoas pois, segundo a International Telecommunication Union [21], são aproximadamente três bilhões de usuários de internet no mundo, ou seja, cerca de 40% da população mundial. Com tamanho volume de vídeos gerados e acessados, torna-se evidente que o correto tratamento do armazenamento, da descrição e da identificação desses vídeos são áreas de interesse para a computação.

Muitas vezes a descrição e identificação dos vídeos são feitas manualmente, utilizando-se palavras-chave (rótulos). Por exemplo, uma pessoa pode rotular um vídeo do litoral com as seguintes palavras: “praia”, “sol”, “litoral”. Outra pessoa, porém, pode utilizar as palavras: “mar”, “oceano”, “férias”. Portanto, o processo de identificar um vídeo é lento, visto que é preciso comparar diversas palavras e sinônimos, e o processo de descrever um vídeo pode ser redundante e impreciso [1]. Outras dificuldades advêm da quantidade de informações contidas em um único exemplar [23], bem como da facilidade de edição de seu conteúdo. Em vídeos de alta resolução e de duração considerável, a análise pixel a pixel torna-se demasiadamente custosa. Essas dificuldades geraram a demanda por métodos automáticos, compactos e precisos de identificação única de vídeos.

O interesse por identificar automaticamente um vídeo também provém da possibilidade de localizá-lo em bases de dados [5]. Esse ponto é importante, pois a facilidade de cópia e disseminação na internet facilita burlar questões de direitos autorais. Além da propriedade intelectual, existem situações onde vídeos privados são copiados e distribuídos sem a autorização de imagem das pessoas em cena.

Questões como as descritas anteriormente mostram a aplicação prática e a necessidade de uma assinatura de vídeo (descrição de vídeo) eficiente que possibilite a recuperação de vídeos e identificação de vídeos privados e protegidos por direitos autorais.

A dificuldade encontrada ao propor uma assinatura digital é a possibilidade de um vídeo manter seu conteúdo principal mas ter características não essenciais ao conteúdo alteradas. Exemplos de modificações que podem ocorrer são: mudança na compressão do arquivo digital, filmagens em telas de cinema ou de televisão, alteração de cores, adição de legendas, subtração do fundo do vídeo, inserção de tarjas (nas partes superior, inferior e laterais do vídeo), alteração do tamanho (altura e largura), entre outras.

Na Figura 1(a) pode-se observar um quadro extraído de um vídeo, publicado no site YouTube, que é a filmagem de um vídeo sendo mostrado em uma televisão. Quando comparado com o original, Figura 1(b), percebe-se diferenças de enquadramento dos personagens, de resolução, no tamanho e também nas cores. No entanto, o conteúdo principal do vídeo é mantido.



(a) Frame do vídeo copiado.



(b) Frame do vídeo original.

Figura 1: Exemplo de cópia de vídeo. O vídeo copiado (a) possui alteração de características quando comparado ao vídeo original (b). Fonte: Youtube (copyright Fox Broadcast Company).

Este trabalho apresenta um estudo a respeito da identificação única para um vídeo a partir da criação de uma assinatura digital, técnica que identifica um vídeo através de características como cor, forma, duração, textura e movimentação de objetos ou câmera. Diversos métodos abordando as mais variadas características foram propostos para executar tal tarefa [5], buscando sempre uma assinatura compacta, robusta e que minimize o custo computacional da identificação. No decorrer do trabalho apresentaremos três métodos utilizados para gerar as assinaturas e a comparação entre eles na identificação de vídeos alterados.

## 1.1 Objetivo Geral

Este trabalho tem como principal objetivo um estudo comparativo de métodos para assinatura digital de vídeos.

## 1.2 Objetivos Específicos

- Estudar e comparar assinaturas digitais para vídeos.
- Verificar a possibilidade de utilização das assinaturas digitais para identificação e recuperação de cópias de vídeos.
- Analisar a eficiência das assinaturas digitais na identificação de vídeos com efeitos de edição e com adição de ruídos.

## 1.3 Estrutura do Documento

Este trabalho está estruturado como segue. No Capítulo 2 são descritos os fundamentos básicos e revisados trabalhos correlatos. No Capítulo 3 apresentados os algoritmos utilizados no projeto. No capítulo 4 são apresentadas a metodologia, a base de dados utilizada e são mostrados os resultados obtidos. No Capítulo 5 são feitas as considerações finais.

## 2 Estado da Arte

Neste capítulo serão abordados alguns dos principais conceitos e definições que constam no texto, sendo necessários para entendimento de assinaturas digitais para vídeos e suas implicações.

### 2.1 Definição de Quadro

Uma quadro (*frame*), segundo Simões [20], é uma imagem em uma unidade de tempo dentro do domínio de tempo de um vídeo, ou seja, um quadro, de altura  $H$  e largura  $W$ , é uma função  $f_t(x, y)$  que representa a intensidade de um pixel em um tempo  $t$  com coordenadas espaciais  $(x, y)$ .

### 2.2 Definição de Vídeo

Um vídeo de tamanho  $n$  é uma sequência de quadros  $V = (f_0, f_1, \dots, f_{n-1})$ , relacionados temporalmente, em que  $f_t$  representa o  $t$ -ésimo quadro. Em um vídeo é possível observar dois tipos principais de propriedades. A amostra espacial, referente às dimensões como altura e largura, e a amostra temporal, referente à relação de tempo entre os quadros de um vídeo

Figura 2.

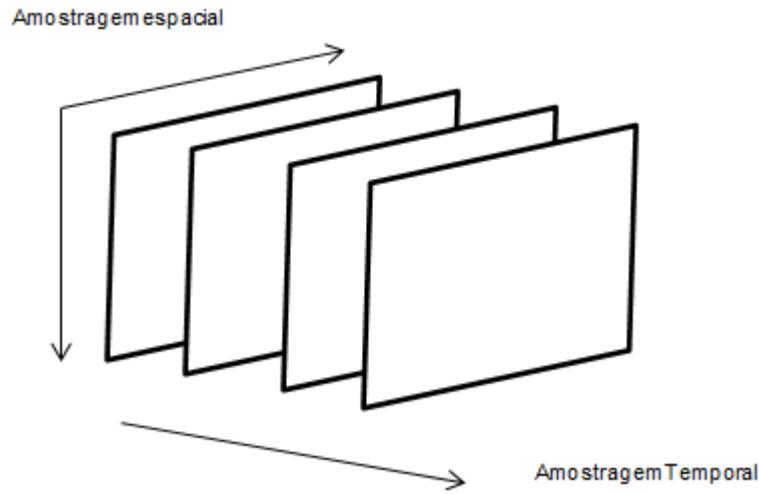


Figura 2: Propriedades de vídeo.

Tendo em vista que um vídeo conecta temporalmente os frames que o constituem, podemos dividir intervalos intermediários entre um frame e o vídeo completo. A Figura 3 ilustra a divisão do vídeo em cena, tomada e quadro, definida anteriormente, onde, uma cena é composta por tomadas e cada tomada por frames [18].

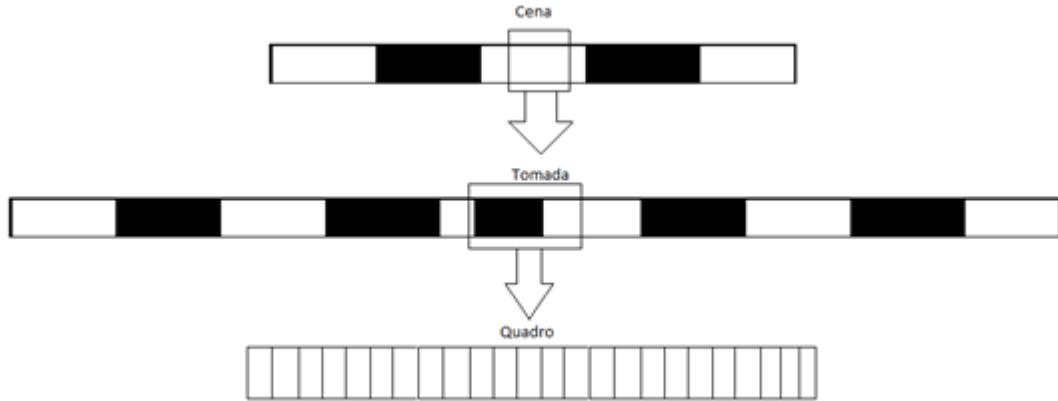


Figura 3: Divisão de vídeo.

## 2.3 Definição de Assinatura de Vídeo

Uma assinatura ou descritor de vídeo pode ser definida como um vetor de características que identifica unicamente um vídeo juntamente com uma medida de similaridade [16]. No escopo deste projeto, distintas assinaturas de vídeos serão avaliadas. Para tanto, é necessário caracterizar o que se espera de uma assinatura de vídeo.

### 2.3.1 Características da Assinaturas de Vídeo

A busca de uma assinatura digital capaz de identificar um vídeo com alta precisão e com baixo custo computacional produziu uma grande variedade de técnicas [5, 9, 12]. A maioria dessas técnicas utilizam propriedades temporais e espaciais para a extração de um descritor confiável, o que significa que a assinatura deve atender a alguns requisitos. A assinatura de vídeo deve possuir ao menos três características para atingir um certo grau de confiabilidade [11, 19, 22]:

- robustez: a assinatura de um determinado vídeo deve ser altamente similar à assinatura

do mesmo vídeo sujeito a alterações (distorções) de conteúdo;

- unicidade: as assinaturas de dois vídeos diferentes devem ser consideravelmente diferentes;
- eficiência de busca: a assinatura digital deve ser compacta e de baixo custo computacional para que seja eficiente na busca de vídeos em banco de dados.

Se a assinatura digital gerada por determinada técnica atender esses critérios, podemos inferir que ela é viável para ser utilizada na identificação de vídeos para cópia completa ou parcial de seu conteúdo.

## 2.4 Recuperação de Vídeos Baseada em Conteúdo

A área de estudo denominada *Content-Based Video Retrieval* (CBVR) que em tradução livre significa recuperação de vídeos baseada em conteúdo, trata do processo de geração de assinatura de vídeo e de identificação (recuperação) de vídeos [9]. Esse ramo da computação abrange desde o desenvolvimento de algoritmos para geração de assinaturas, o estudo de equivalência (*matching*) entre as assinaturas até a busca e recuperação de vídeos.

Outro campo de estudo correlacionado é o *Content-Based Copy Detection* (CBCD) que, em tradução livre, significa detecção de cópias baseada em conteúdo. Essa área de estudo trata da identificação de cópias de vídeos usando assinaturas digitais.

Como descrito no Capítulo 1, existem várias características de um vídeo que podem ser alteradas, dificultando a construção de soluções computacionais adequadas [4]. As diferentes técnicas utilizadas para gerar assinaturas em vídeos divergem tanto nas características usadas para extrair seus parâmetros quanto nas técnicas empregadas para avaliar o grau de similaridade entre as diferentes assinaturas. Nos tópicos a seguir serão apresentados algumas abordagens relevantes para o projeto.

## 2.5 Descritores

Do ponto de vista das características que os descritores exploram, é possível dividir as técnicas em dois grandes grupos. O primeiro grupo refere-se às técnicas que utilizam **descritores globais** para extrair as propriedades dos vídeos, o segundo é o grupo de técnicas que utilizam **descritores locais** para o mesmo propósito. Nas seções seguintes as duas abordagens serão definidas e discutidas.

### 2.5.1 Descritores Globais

Como visto na definição de assinatura digital, os descritores buscam sintetizar informações contidas nos vídeos a partir de características únicas dos vídeos. Dessa forma, os descritores globais tratam cada frame do vídeo como um objeto único, sendo que as informações usadas representam o frame como um todo. Geralmente os descritores globais são mais simples e possuem um bom desempenho computacional, porém mostram-se mais suscetíveis a certas distorções.

Os descritores globais utilizam-se de características como distribuição de cores [6] [3], bordas [11] e também histogramas de cor [14] que representam um frame. **Essas técnicas mostram-se robustas contra distorções geométricas como rotação, translação ou até mesmo recortes, porém são mais propensas a serem suscetíveis a distorções que atingem o espaço de cores.** Os algoritmos abordados neste trabalho (Capítulo 3) são todos descritores globais.

### 2.5.2 Descritores Locais

As técnicas que utilizam descritores locais geram seus dados a partir de **elementos específicos na imagem**, tais como: **pontos de interesse, bordas ou objetos específicos**. Busca-se caracterizar comportamentos das regiões ao longo do tempo e descritores mais robustos a distorções de geometria, de cores e de texturas. Porém, o custo computacional é mais alto em relação aos descritores globais. Dessa forma, essas técnicas procuram encontrar pontos de interesse que possam ser rastreados ao longo do vídeo produzindo descritores relevantes

e imunes a distorções. O descritor local deve possuir a capacidade de repetir-se ao longo do tempo. Dessa forma ao sofrer transformações espaciais o descritor mantém-se robusto. Pode-se citar a detecção de bordas em regiões específicas e também a técnica de *Space-Time Interest Points* [1] [8] que procura regiões no frame onde ocorrem movimentações mais abruptas.

## 2.6 Trabalhos Prévios

Neste trabalho foram estudados três métodos com o objetivo de avaliar a robustez e unicidade de seus descritores. Os métodos abrangem as técnicas mais comumente utilizadas para a geração de assinaturas digitais através de descritores globais.

A medida ordinal, utilizada por Hua et. al [6], é uma medida espacial de nível de cinza comumente utilizada para identificar cópias similares [2, 15]. A identificação de bordas, como descrito por Lee and Yoo [11], é outra informação a respeito das características espaciais de um vídeo utilizada para resumir o conteúdo de um quadro. Por fim, propriedade temporal, como descrito por Cook [3], reflete a mudança global dos frames e não apenas de um frame específico.

Dos métodos globais desenvolvidos para geração de um assinatura digital, alguns métodos utilizam abordagens espaciais para produzir seus descritores, calculando medidas baseadas em níveis de cinza e informações sobre bordas. Outros métodos utilizam medidas temporais, explorando informações entre quadros separados temporalmente. Na sequência será apresentada uma revisão dos métodos presentes na literatura.

Hua et. al. [6] calcularam uma medida ordinal que reflete a distribuição de níveis de cinza para cada quadro. A assinatura foi desenvolvida para ser robusta a diferentes formatos de compressão, taxa de frames, assim como diferenças no tamanho e composição espacial dos frames. Lee and Yoo [10, 11] propuseram um método baseado na orientação do centroide do gradiente para ser utilizada contra transformações geométricas, tais como rotação e translação, e ruído Gaussiano. Uma abordagem similar é apresentada por Massoudi et. al. [13]. A assinatura proposta por eles é baseada na orientação do gradiente e foi testada

contra transformações como compressão, recortes e borramento. Su *et. al.* [22] propôs um método baseado em regiões de atenção, este método foi avaliado contra vídeos contendo distorções de resolução, de quantidade de quadros por segundo e de adição de logotipos. Radhakrishnan and Bauer [17] apresentam um método que utiliza a Decomposição de Valor Singular (SVD) que busca robustez contra transformações geométricas, compressões e alterações nas taxas de quadros.

Cook [3] propôs uma assinatura temporal que utiliza a diferença de luminância entre frames para medir como a informação contida no vídeo é alterada através do tempo. A simplicidade deste método é utilizada para atingir eficiência computacional. O autor afirma que o método é robusto tanto contra transformações geométricas e não geométricas. No trabalho de Kim and Vasudev [7], a intensidade média dos frames é computada e comparada com a mesma medida em um frame subsequente, implicando em uma assinatura espaço-temporal. Chen e Stentiford [2] compararam a média do nível de cinza para dois frames consecutivos. Hampur *et. al.* [4] propõem um método baseado na captura de movimento entre dois frames subsequentes, utilizando como métrica o mínimo da diferença da soma absoluta dos pixels entre os frames.

### 3 Algoritmos

Nesta sessão serão apresentados e discutidos como os métodos apresentados neste trabalho geram as assinaturas digitais.

#### 3.1 Assinatura de Vídeo por Distribuição de Intensidade

Em 2004, Hua *et. al.* [6] propuseram um método baseado na intensidade de níveis de cinza de um quadro para gerar uma assinatura. Com este propósito, os autores dividem os quadros de um vídeo em  $M \times N$  blocos de tamanho igual, como mostrado na Figura 4(a).

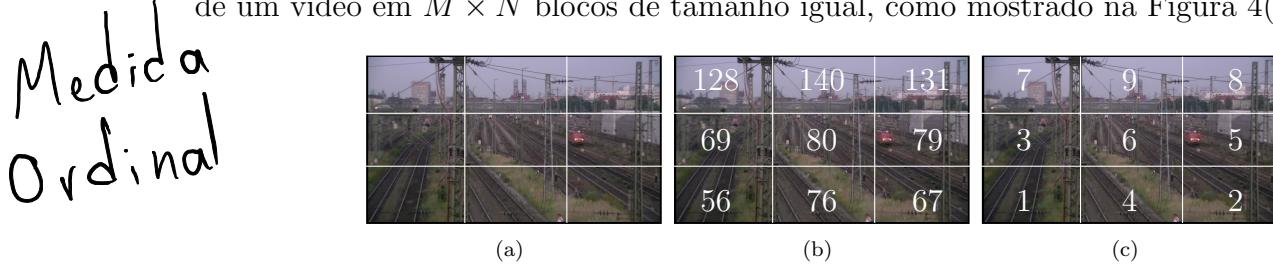


Figura 4: Computação do descriptor de Hua *et. al.* [6]: (a) divisão do quadro em blocos; (b) nível de cinza médio em cada bloco; (c) o descriptor  $\mathbf{d} = \langle 7, 9, 8, 3, 6, 5, 1, 4, 2 \rangle$  é formado pela permutação que ordena os níveis de cinza dos blocos de (b).

A média da intensidade do nível de cinza em cada bloco  $b[i]$ , de um quadro  $\mathbb{I}$ , para  $i = \{1, \dots, M \times N\}$ , é calculada como:

$$g[i] = \frac{\sum_{x,y \in b[i]} \mathbb{I}(x,y)}{|b[i]|}, \quad (1)$$

onde  $|b|$  é o número de pixels que compõem o bloco  $b$  (Figura 4(b)).

Então, o ranking dos blocos é obtido através da ordenação, em ordem crescente, do vetor da média da intensidade do nível de cinza  $\mathbf{g}$  e nomeando-os com inteiros consecutivos  $1, 2, \dots, M \times N$ . Este ranking também é conhecido como *medida ordinal*. O valor do descriptor  $d[i]$  é valor do inteiro que atribui-se ao block  $b[i]$  durante a ordenação. Como visto na Figura 4(c).

A assinatura de um vídeo é composta pela sequência de descritores de cada quadro. Os autores afirmam que a assinatura  $\mathbf{d}$  é robusta a certas alterações no formato de compressão, na taxa de quadros e no tamanho dos quadros.

### 3.2 Assinatura de Vídeo por Distribuição de Gradientes

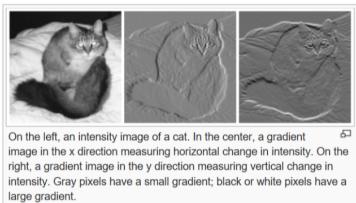
Em 2008, Lee and Yoo [11] propuseram uma assinatura baseada na distribuição dos gradientes dos quadros. Com este propósito, os autores primeiramente calculam o gradiente de um quadro  $\mathbb{I}$  para cada ponto  $(x, y)$  sendo:

$$\nabla \mathbb{I}(x, y) = \begin{bmatrix} \mathbb{G}_x \\ \mathbb{G}_y \end{bmatrix} = \begin{bmatrix} \partial \mathbb{I} / \partial x \\ \partial \mathbb{I} / \partial y \end{bmatrix} = \begin{bmatrix} \mathbb{I}(x+1, y) - \mathbb{I}(x-1, y) \\ \mathbb{I}(x, y+1) - \mathbb{I}(x, y-1) \end{bmatrix}. \quad (2)$$

Dado um gradiente, sua magnitude e sua orientação são obtidas por:

$$\omega(x, y) = \sqrt{\mathbb{G}_x^2 + \mathbb{G}_y^2} \quad \theta(x, y) = \tan^{-1} \left( \frac{\mathbb{G}_y}{\mathbb{G}_x} \right). \quad (3)$$

Como em Hua *et. al.* [6], os autores dividem os quadros do vídeo em  $M \times N$  blocos de tamanhos iguais e calculam a orientação do centroide dos gradientes para cada bloco. Especificamente, o descritor do centroide  $d[i]$  relativo ao bloco  $b[i]$ ,  $i = \{1, \dots, M \times N\}$ , é dado por:



$$d[i] = \frac{\sum_{x,y \in b[i]} \omega(x, y) \theta(x, y)}{\sum_{x,y \in b[i]} \omega(x, y)}. \quad (4)$$

Na Figura 5 pode-se observar o esquema de normalização do vídeo utilizado por Lee and Yoo [11], assim como o processo de geração da assinatura de gradiente.

Os autores afirmam que a assinatura está intimamente relacionada com a distribuição de bordas nos quadros, fornecendo informações visuais relevantes sobre o conteúdo dos quadros, como os limites de objetos. A assinatura é robusta contra modificações globais na intensidade dos pixels, tais como, alteração no brilho, cor e contraste.

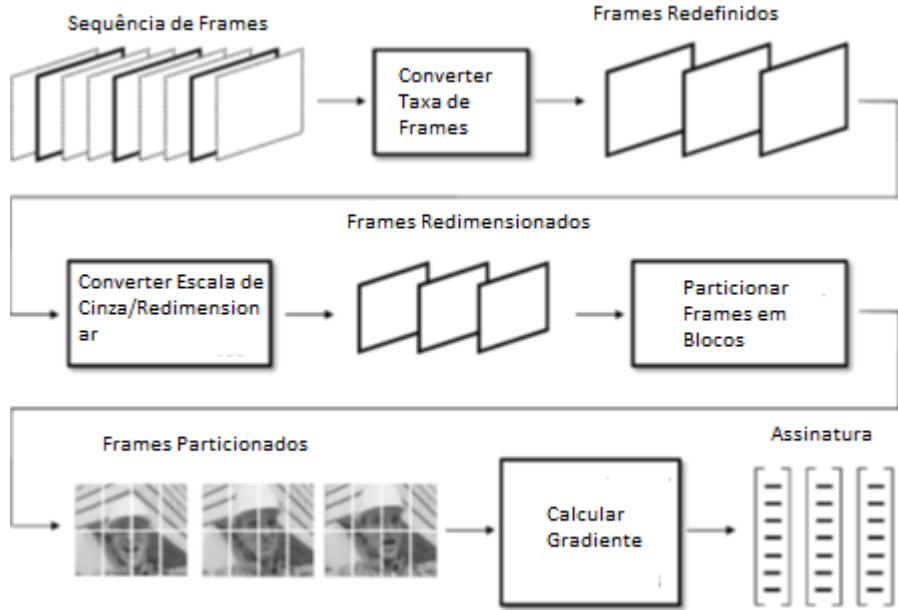


Figura 5: Assinatura por distribuição de gradiente. Fonte: Lee, Sunil e Yoo, Chang D. (2008. P 984). Traduzido pelo Autor.

### 3.3 Assinatura de Vídeo por Diferença entre Quadros

Cook [3] propôs, em 2011, uma assinatura baseada na informação temporal contida entre dois quadros deslocados no tempo. Sejam  $\mathbb{J}$  e  $\mathbb{I}$  dois quadros de um vídeo, não necessariamente subsequentes, a diferença entre os quadros é dada por

$$dY = \sum_{x,y \in \mathbb{I}, \mathbb{J}} |\mathbb{I}(x,y) - \mathbb{J}(x,y)|. \quad (5)$$

Os autores também calculam a soma dos níveis de cinza do quadro  $\mathbb{I}$

$$Y = \sum_{x,y \in \mathbb{I}} \mathbb{I}(x,y). \quad (6)$$

O descriptor para cada quadro é então composto pelas duas medidas,  $\mathbf{d} = \langle dY, Y \rangle$ .

Os autores afirmam que o valor  $dY$  reflete a mudança global dos quadros. Alterações que persistem entre quadros, tais como, translação, rotação, espelhamento, legendas, sobreposições estáticas, brilho, contraste e alterações no espectro de cores, são canceladas quadro a quadro, deixando um registro relativamente consistente das mudanças no vídeo.

## 4 Metodologia

Nesta seção serão apresentados a base de dados e as métricas de avaliação, assim como os resultados obtidos, recursos de hardware e software, cronograma e viabilidade.

### 4.1 Base de Dados

Neste trabalho foi utilizada a base de vídeos LIVE Video Quality (LIVE-VQD), para avaliar a robustez e a unicidade (características descritas na Seção 2.3.1) dos métodos estudados contra quatro tipos de distorções. A base LIVE-VQD contém dez vídeos de referência com alta qualidade e sem compressão. Os vídeos da base contém movimentação de câmera e de objetos. A quantidade de quadros varia entre 250 e 500. Além dos 10 vídeos originais de referência, existem outros 150 vídeos, criados a partir dos vídeos de referência (15 por vídeo de referência). Cada um desses vídeos possui um entre quatro tipos diferentes de alterações: compressão MPEG-2, compressão H.264, simulação de transmissão de dados comprimidos com H.264 através de uma rede IP e perda de dados na transmissão wireless. Na Figura 6 estão os quadros iniciais dos dez vídeos de referência.

Neste trabalho ainda foram incluídas quatorze distorções adicionais por vídeo de referência (Figura 7): rotação, recorte, **espelhamento**, legendas, marcas d'água, alteração de cores, redimensionamento, adição de bordas e alteração na taxa de quadros. Geralmente tais alterações são deliberadamente realizadas sobre vídeos proprietários com o intenção de despistar métodos de detecção de cópias.

### 4.2 Experimentos

Os experimentos foram realizados através da consulta da assinatura de um vídeo  $V^q$  contra a assinatura de cada vídeo  $V^t$  na base de dados. O vídeo de consulta é uma sequência de quadros  $V^q = \langle \mathbb{I}^1, \mathbb{I}^2, \dots, \mathbb{I}^n \rangle$  e o vídeo alvo é uma sequência de quadros  $V^t = \langle \mathbb{J}^1, \mathbb{J}^2, \dots, \mathbb{J}^m \rangle$ ,  $n \ll m$ . As assinaturas para esses vídeos são as sequências  $\mathbf{f}^q = \langle \mathbf{d}_1^q, \mathbf{d}_2^q, \dots, \mathbf{d}_n^q \rangle$  and  $\mathbf{f}^t = \langle \mathbf{d}_1^t, \mathbf{d}_2^t, \dots, \mathbf{d}_m^t \rangle$ , respectivamente.

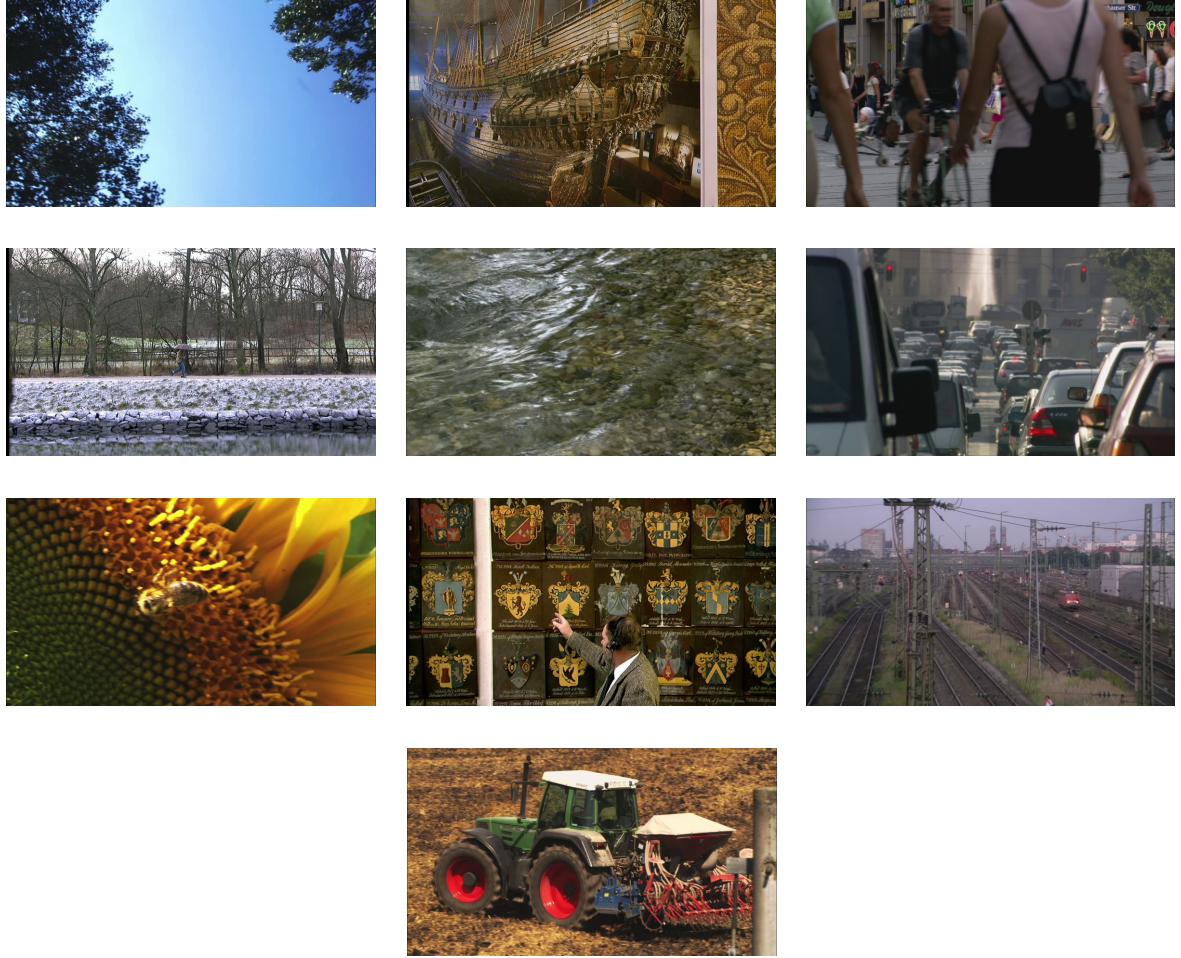


Figura 6: Quadros iniciais dos dez vídeos de referência. Fonte LIVE Video Quality (LIVE-VQD).

Seja  $\mathbf{d}^q \in \langle \mathbf{d}_1^q, \mathbf{d}_2^q, \dots, \mathbf{d}_n^q \rangle$  and  $\mathbf{d}^t \in \langle \mathbf{d}_1^t, \mathbf{d}_2^t, \dots, \mathbf{d}_n^t \rangle$ , a distância  $L_1$  entre os descritores  $\mathbf{d}^q = \langle d^q[1], d^q[2], \dots, d^q[k] \rangle$  e  $\mathbf{d}^t = \langle d^t[1], d^t[2], \dots, d^t[k] \rangle$  é dada por

$$L_1(\mathbf{d}^q, \mathbf{d}^t) = \|\mathbf{d}^q - \mathbf{d}^t\| = \sum_{i=1}^k |d^q[i] - d^t[i]|. \quad (7)$$

A distância entre as assinaturas do vídeo de consulta e de uma subsequência de quadros do vídeo alvo são calculadas por

$$D(\mathbf{f}^q, \mathbf{f}_j^t) = \frac{1}{nk} \sum_{i=1}^n L_1(\mathbf{d}_i^q, \mathbf{d}_{j+i}^t) \quad (8)$$



Figura 7: Quadros com distorções, da esquerda para a direita e de cima para baixo: blur (ofuscamento), adição de borda vermelha, inversão de cores, recorte central do quadro, espelhamento, compressão JPEG do quadro, rotação para a direita, adição de legenda e, por fim, adição de marca d'água.

onde o fator  $nk$  é utilizado para normalização e  $j = \{0, \dots, m - n\}$  são as possíveis posições onde a consulta pode ser comparada com o vídeo alvo.

O intervalo de valores possíveis dos descritores para cada método (ver Tabela 1) foram normalizados para o intervalo  $[0, 1]$ . Os ajustes dos parâmetros de cada método, como descritos originalmente pelos autores, também podem ser vistos na Tabela 1.

Tabela 1: Parâmetros dos algoritmos e informações das assinaturas

Algoritmo	Intervalo do Descritor	Número de blocos	Número de propriedades por bloco	Tamanho do descritor
Hua <i>et. al.</i> [6] (Ordinal)	$[1 : 9]$	9	1	9
Lee and Yoo [11] (Gradient)	$[-\pi/2 : +\pi/2]$	8	1	8
Cook [3] (Temporal)	$[0 :  \mathbb{I}  \times 255]$	1	2	2

Com o objetivo de avaliar as assinaturas testadas, foram computadas as curvas de Característica de Operação do Receptor (no inglês ROC) e precisão-revocação. Dado um limiar

(*threshold*)  $\tau$ , no intervalo  $[0, 1]$ , se  $D(\mathbf{f}^q, \mathbf{f}_j^t) \leq \tau$  então assume-se que o vídeo  $V^q$  corresponde à subsequência do vídeo  $V^t$  que se inicia no quadro  $j$ . Ao comparar-se a consulta com todos os vídeos da base de dados foram computadas as medidas de *precisão* ( $P$ ), *revocação* ( $R$ ) (também denominada *taxa de verdadeiros positivos* ( $TPR$ )), e *taxa de falsos positivos* ( $FPR$ )

$$P = \frac{TP}{TP + FP} \quad R(TPR) = \frac{TP}{TP + FN} \quad FPR = \frac{FP}{FP + TN} \quad (9)$$

onde  $TP$  é o número de verdadeiros positivos (acertos),  $FP$  é o número de falsos positivos,  $FN$  é o número de falsos negativos e  $TN$  é o número de verdadeiros negativos. Para um limiar específico, um verdadeiro positivo ocorre quando a consulta buscada de um vídeo corresponde a outra assinatura de um vídeo distorcido sendo ambos do mesmo vídeo de referência. Se eles não forem correspondentes, tem-se um falso negativo. Um falso positivo ocorre quando há a correspondência de assinaturas oriundas de vídeos de diferentes vídeos de referência, de outro modo ocorre um verdadeiro negativo.

Foram considerados três cenários para se avaliar a robustez e a unicidade das assinaturas:

1. Vídeos com compressão e erros de transmissão como aqueles contidos na base de dados LIVE-VQD;
2. Vídeos com compressão e erros de transmissão comuns, juntamente com vídeos que sofreram ataques de distorções para cópias;
3. Variação do tamanho da consulta sobre os vídeos do cenário 2.

As curvas de precisão-revocação e ROC para 50 consultas aleatoriamente selecionadas são apresentadas nas Figuras 8-10. A Figura 8 apresenta os resultados no cenário (1) considerando consultas que têm tamanho igual a 25 quadros. Todos os métodos saíram-se muito bem para estes tipos de distorções. O método de Hua *et. al.* [6] alcançou 100% de revocação e de precisão.

O comportamento não é o mesmo para o cenário(2), onde incluem-se vídeos deliberadamente atacados, ver Figura 9. Percebe-se que a performance cai rapidamente para estes

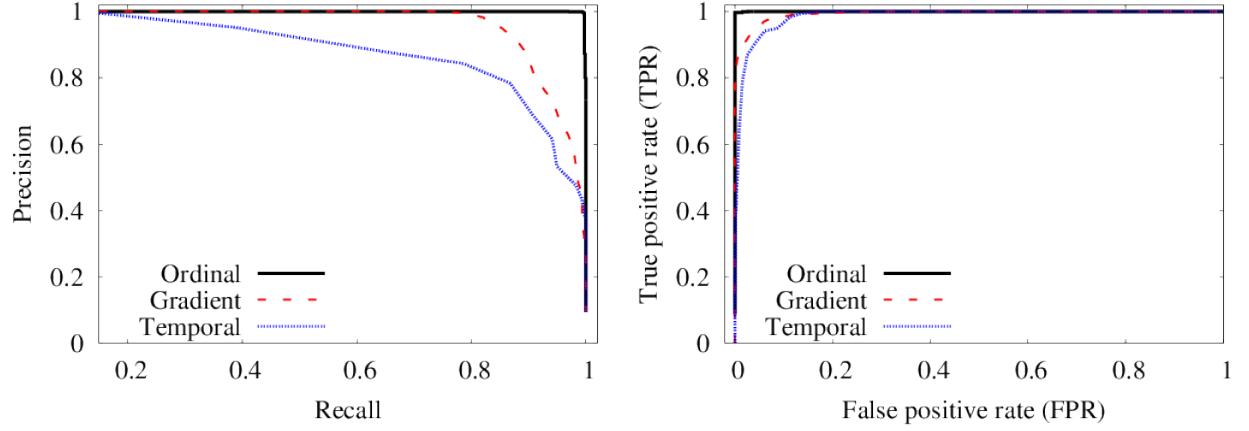


Figura 8: Performance dos métodos: curvas de precisão-revocação (esquerda) e ROC (direita) para consultas com **25** quadros considerando a base de dados LIVE-VQD original.

tipos de distorções. Os métodos de Hua *et. al.* [6] e Lee e Yoo [11] alcançaram resultados semelhantes, sendo ambos mais robustos que o método de Cook [3].

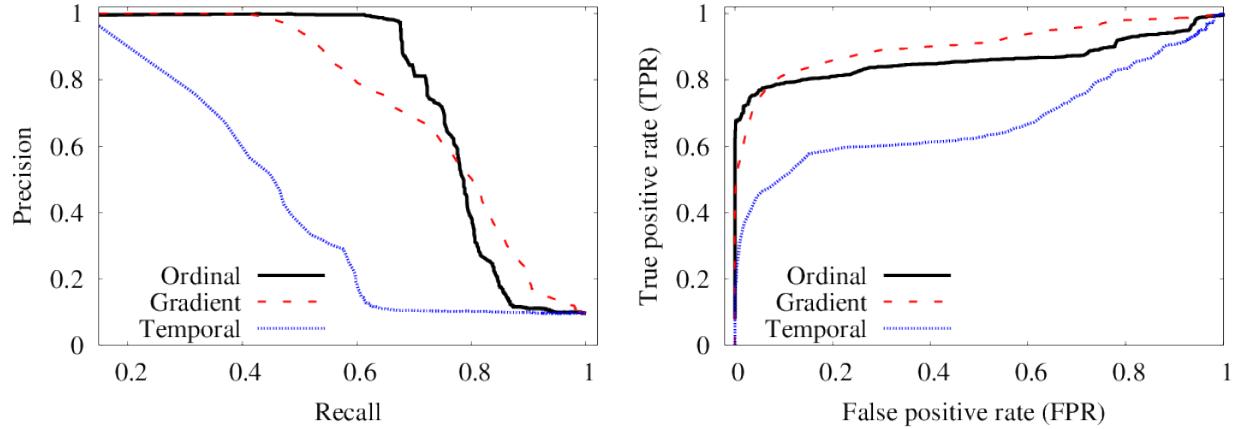


Figura 9: Performance dos métodos: curvas de precisão-revocação (esquerda) e ROC (direita) para consultas com **25** quadros considerando a base de dados LIVE-VQD com quatorze distorções adicionais para cada vídeo de referência.

Com o objetivo de verificar a influência do tamanho da consulta na recuperação de vídeos, foram utilizados os vídeos do cenário (2) e consultas com 100 e 200 quadros. Os resultados dos testes podem ser vistos nas Figures 10 e 11. Percebe-se que a performance de todos os algoritmos melhoraram nitidamente com consultas de 100 e 200 quadros. No entanto, o tamanho da consulta impacta diretamente no custo de computação dos métodos, visto que é preciso maior quantidade de cálculos para comparar as assinaturas. Esse ponto de troca

mostra-se muito importante na eficiência da busca em bases de dados.

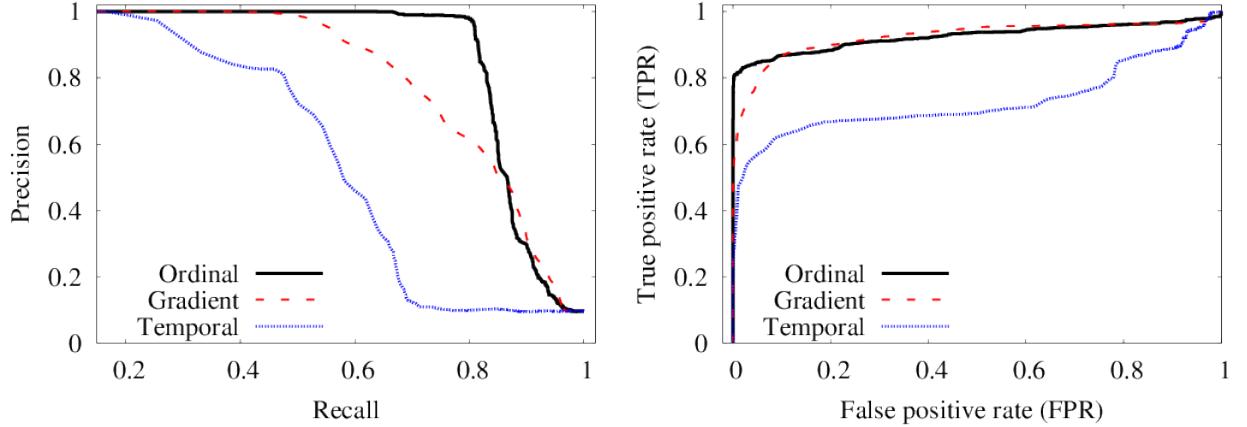


Figura 10: Performance dos métodos: curvas de precisão-revocação (esquerda) e ROC (direita) para consultas com **100** quadros considerando a base de dados LIVE-VQD com quatorze distorções adicionais para cada vídeo de referência.

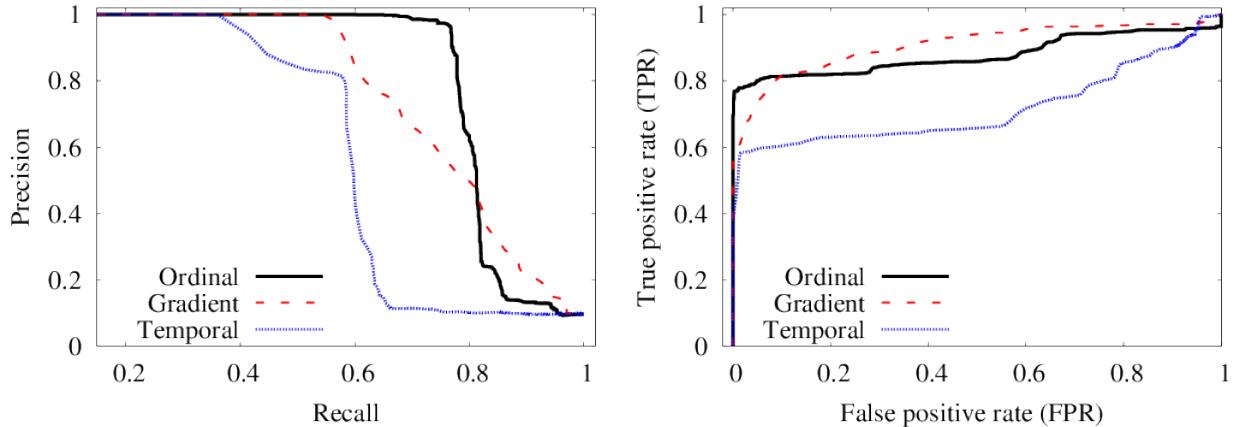


Figura 11: Performance dos métodos: curvas de precisão-revocação (esquerda) e ROC (direita) para consultas com **200** quadros considerando a base de dados LIVE-VQD com quatorze distorções adicionais para cada vídeo de referência.

Durante os testes também foram comparadas as curvas de distâncias  $L_1$  entre duas assinaturas. A análise destas curvas permite observar o comportamento das assinaturas em diferentes cenários. O comportamento esperado é que a distância  $L_1$  de dois descritores oriundos de um mesmo vídeo de referência, possuam um mínimo evidente, abaixo do limiar estabelecido, configurando assim uma correspondência. Por outro lado, para descritores provenientes de vídeos de referência diferentes espera-se que a distância  $L_1$  não apresente mínimo

abaixo do limiar. Para elaborar esse teste foram realizadas três comparações.

Primeiramente foram comparadas as assinaturas geradas entre o vídeo bs1 e o vídeo bs1.blur que são o mesmo vídeo porém o segundo conta com a adição de ruído do tipo blur. Pode-se observar a diferença entre os três métodos na Figura 12, onde verifica-se que os três métodos conseguiram identificar precisamente o ponto onde ocorre a correspondência entre a consulta e o vídeo alvo, caracterizando um verdadeiro positivo. Este ponto é representado pela linha vertical "Consulta".

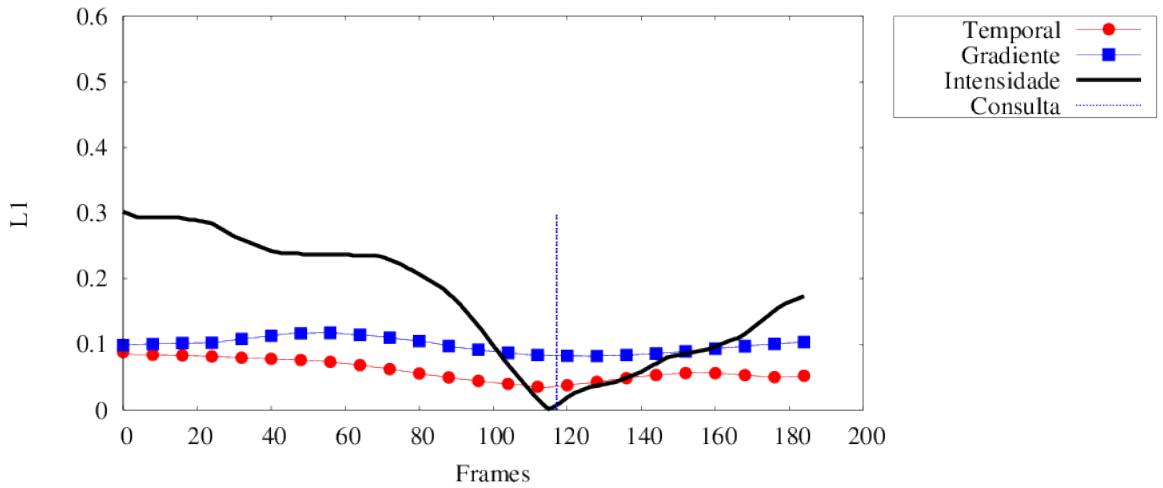


Figura 12: Distância  $L_1$  entre vídeos de mesma referência com mínimo evidente. Onde as linhas horizontais representam as técnicas e a linha vertical o ponto de início da consulta.

O segundo teste ocorreu comparando a distância  $L_1$  entre os vídeos bs1 e tr1, ou seja, dois vídeos completamente diferentes e sem a adição de distorções. Na Figura 13 observa-se que nas três curvas não se percebe um mínimo aparente, evidenciando que tratam-se de vídeos distintos, caracterizando um verdadeiro negativo.

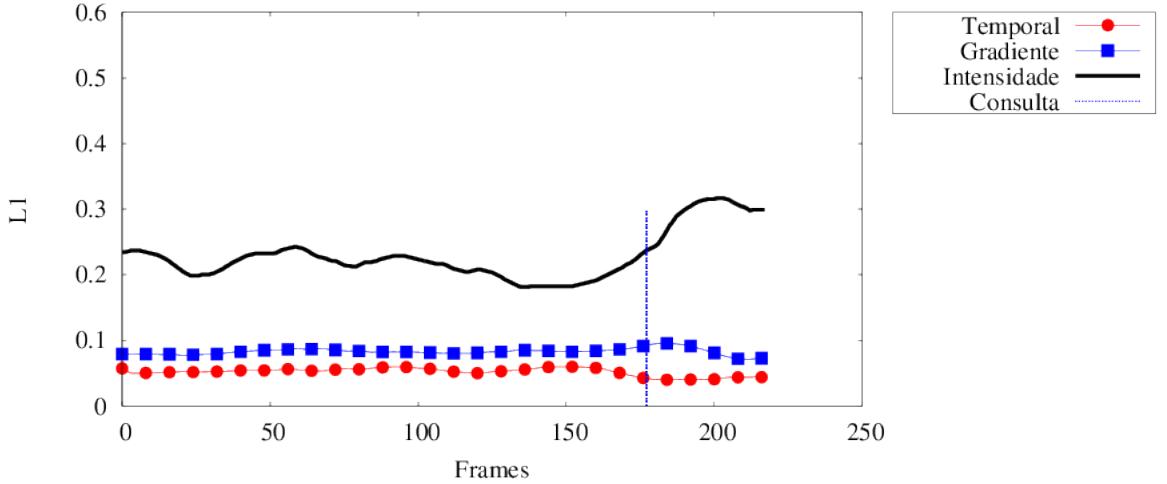


Figura 13: Distância  $L_1$  entre vídeos de diferentes referências sem mínimo evidente. Onde as linhas horizontais representam as técnicas e a linha vertical o ponto de início da consulta.

Por último foram comparados os vídeos bs1 e bs1\_Crop, ou seja, o vídeo de referência e sua cópia com recorte central. O comportamento esperado era que fosse encontrado um mínimo aparente, porém este mínimo não ocorreu evidenciando um falso negativo. Ver Figura 14.

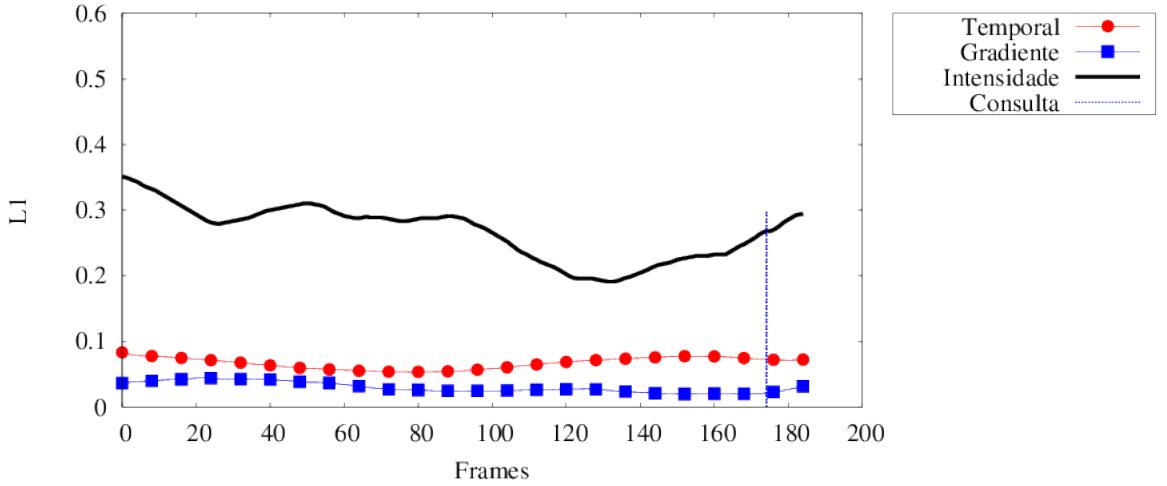


Figura 14: Distância  $L_1$  entre vídeos de mesma referência sem mínimo evidente. Onde as linhas horizontais representam as técnicas e a linha vertical o ponto de início da consulta.

Ao observar os gráficos de distância  $L_1$  percebe-se que os valores de  $L_1$  para a técnica de intensidade estão sempre maiores, porém quando encontra uma correspondência este valor cai drasticamente. Dessa forma corrobora os gráficos de precisão/revocação onde esta técnica obteve o melhor desempenho. Isto ocorre pois com um mínimo acentuado torna-se fácil para

a assinatura identificar uma correspondência. Nas outras técnicas os valores de  $L_1$  não obtém grande variação, dificultando a identificação de correspondência. Percebe-se que nas técnicas de gradiente e temporal os valores de  $L_1$  estão sempre muito próximos de zero, mesmo em situações onde não existe correspondência. Este fato pode indicar a razão da precisão inferior nas curvas de precisão/revocação e ROC.

### **4.3 Recursos de Hardware e Software**

Neste tópico serão apresentados os recursos de hardware e software, utilizados no desenvolvimento do projeto, tanto para o desenvolvimento dos algoritmos quanto para a realização dos testes.

#### **Recursos de Hardware**

Foi disponibilizado pelos professores orientadores do DAINF (Departamento de Informática) da UTFPR (Universidade Tecnológica Federal do Paraná) um computador para acesso remoto, utilizado para armazenar a base de dados (vídeos) utilizados nos testes, assim como para executar os testes. A configuração do hardware consiste em um processador Intel Core I7 com 1 terabyte de disco rígido e 32 gigabytes de memória RAM.

#### **Recursos de Software**

Para o desenvolvimento do projeto foi utilizado o sistema operacional Linux com distribuição Mint versão 17. Para codificar os algoritmos foi utilizada a linguagem C e bibliotecas para processamento de imagem e vídeos. O programa FFmpeg foi utilizado para compressão dos vídeos, extração de frames, adição de ruído aos vídeos e distorção dos vídeos.

## 5 Conclusão

Neste trabalho relatou-se uma avaliação experimental de três métodos, baseados em diferentes atributos, para a computação de assinaturas de vídeos digitais. Os métodos foram testados em uma base de dados de vídeos de qualidade reconhecida, a fim de investigar o efeito de distorções de vídeo e erros de transmissão, com o objetivo de recuperação de vídeo. Deste modo foram inclusas quatorze distorções de preservação de conteúdo, a fim de simular cópias de vídeo modificados. *Os experimentos mostraram que os métodos são robustos à tipos comuns de compressão e distorção, tais como os de vídeo da base de dados LIVE Video Quality, mas são sensíveis a cópias de vídeo modificados propositalmente.* Nos experimentos desenvolvidos neste trabalho, os melhores resultados foram obtidos utilizando-se o método da medida ordinal. Trabalhos futuros podem abordar a utilização conjunta de diferentes descritores para avaliar se existe melhora de desempenho, sendo possível também avaliar o desempenho de recuperação de vídeos em grande bases de dados.

A contribuição principal do projeto ocorreu durante o processo de desenvolvimento do artefato textual, pois foi necessária extensa pesquisa na área, assim como no processo de testes.

## Referências

- [1] Felipe dos Santos Pinto de Andrade. Combinação de descritores locais e globais para recuperação de imagens e vídeos por conteúdo. *Universidade Estadual de Campinas (UNICAMP). Instituto de Computação*, 2012.
- [2] Li Chen and F.W.M. Stentiford. Video sequence matching based on temporal ordinal measurement. *Pattern Recognition Letters*, 19:1824–1831, 2008.
- [3] R. Cook. An efficient, robust video fingerprinting system. In *IEEE International Conference on Multimedia and Expo*, pages 1–6, July 2011.
- [4] Arun Hampapur, Kiho Hyun, and Ruud M. Bolle. Comparison of sequence matching techniques for video copy detection. *Storage and Retrieval for Media Databases*, 4676:194–201, 2001.
- [5] Weiming Hu, Nianhua Xie, Li Li, Xianglin Zeng, and S. Maybank. A survey on visual content-based video indexing and retrieval. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 41(6):797–819, Nov 2011.
- [6] Xian-Sheng Hua, Xian Chen, and Hong-Jiang Zhang. Robust video signature based on ordinal measure. In *International Conference on Image Processing*, volume 1, pages 685–688, Oct 2004.
- [7] Changick Kim and B. Vasudev. Spatiotemporal sequence matching for efficient video copy detection. *IEEE Trans. on Circ. and Systems for Video Tech.*, 15(1):127–132, 2005.
- [8] Ivan Laptev. On space-time interest points. *International Journal of Computer Vision*, 64(2-3):107–123, 2005.
- [9] Julien Law-To, Li Chen, Alexis Joly, Ivan Laptev, Olivier Buisson, Valerie Gouet-Brunet, Nozha Boujemaa, and Fred Stentiford. Video copy detection: A comparative study. In *ACM International Conference on Image and Video Retrieval*, CIVR, pages 371–378, 2007.

- [10] Sunil Lee and C.D. Yoo. Robust video fingerprinting based on affine covariant regions. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pages 1237–1240, 2008.
- [11] Sunil Lee and C.D. Yoo. Robust video fingerprinting for content-based video identification. *IEEE Trans. on Circuits and Systems for Video Technology*, 18(7):983–988, 2008.
- [12] Jian Lu. Video fingerprinting for copy identification: from research to industry applications. *Media Forensics and Security*, 7254:725402–725402–15, 2009.
- [13] A. Massoudi, F. Lefebvre, C. Demarty, L. Oisel, and B. Chupeau. A video fingerprint based on visual digest and local fingerprints. In *IEEE Int. Conf. on Image Processing*, pages 2297–2300, 2006.
- [14] Babu M Mehtre, Mohan S Kankanhalli, A Desai Narasimhalu, and Guo Chang Man. Color matching for image retrieval. *Pattern Recognition Letters*, 16(3):325–331, 1995.
- [15] Sakrappee Paisitkriangkrai, Tao Mei, Jian Zhang, and Xian-Sheng Hua. Scalable clip-based near-duplicate video detection with ordinal measure. In *ACM International Conference on Image and Video Retrieval*, CIVR, pages 121–128, 2010.
- [16] Otavio Augusto Bizetto Penatti. Estudo comparativo de descritores para recuperação de imagens por conteúdo na web. *Universidade Estadual de Campinas (UNICAMP). Instituto de Computação*, 2009.
- [17] R. Radhakrishnan and C. Bauer. Robust video fingerprints based on subspace embedding. In *IEEE ICASSP*, pages 2245–2248, 2008.
- [18] Thiago Teixeira Santos. Segmentação automática de tomadas em vídeo. *Universidade Estadual de São Paulo (USP). Instituto de Computação*, 2004.
- [19] Isabelle Simand, Denis Pellerin, Stéphane Bres, and Jean-Michel Jolion. Spatio-Temporal Signatures for Video Copy Detection. In *Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS)*, pages 421–427, September 2004.

- [20] Nielsen Cassiano Simoes. Detecçao de algumas transições abruptas em sequencias de imagens. *Mestrado, Instituto de Computação, UNICAMP, Campinas*, 5, 2004.
- [21] ITU Statistics. Key ict indicators for developed and developing countries and the world (totals and penetration rates). URL: [http://www.itu.int/ITUD/ict/statistics/at\\_glance/KeyTelecom.html](http://www.itu.int/ITUD/ict/statistics/at_glance/KeyTelecom.html), 29:2012, 2014.
- [22] Xing Su, Tiejun Huang, and Wen Gao. Robust video fingerprinting based on visual attention regions. In *IEEE ICASSP*, pages 1525–1528, 2009.
- [23] Yueling Zhuang, Yong Rui, Thomas S Huang, and Sharad Mehrotra. Adaptive key frame extraction using unsupervised clustering. In *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, volume 1, pages 866–870. IEEE, 1998.