

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO



# **Deteção de Publicidade em Conteúdos de Difusão**

**Tiago José Rodrigues Pereira**

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Orientador: Pedro Miguel Machado Soares Carvalho

Co-orientador: Luís António Pereira de Meneses Corte-Real

25 de Julho de 2018



# Resumo

A detecção automática de conteúdos publicitários em transmissões televisivas é um problema abordado há várias décadas em diferentes estudos com diferentes interesses, sejam eles por parte das empresas de difusão, por parte das empresas publicitadas ou dos próprios consumidores. Um dos casos de maior interesse aborda a substituição de anúncios publicitários por outros mais direcionados ao consumidor final em situações de redistribuição de conteúdos, como transmissões *online*. No entanto, são várias as dificuldades associadas a este processo.

Nesta dissertação são propostos dois métodos baseados em *fingerprinting* de vídeo para realizar a detecção de conteúdos publicitários de modo automático num ambiente de *broadcasting*, tendo em conta as restrições temporais e computacionais existentes num contexto de transmissão linear. Para a geração e armazenamento de *fingerprints* é considerado o conhecimento prévio dos segmentos publicitários a serem transmitidos enquanto que a técnica de resumo de vídeo *Hash Diferencial - dHash* e a estratégia de base de dados SQLite são também abordadas.

Tendo em vista o teste e avaliação da metodologia proposta é utilizado um *dataset* de *streams* televisivas e é apresentado um método de avaliação com o objetivo de analisar os resultados obtidos em pontos cruciais para o alcance dos objetivos propostos. Os métodos desenvolvidos apresentaram resultados distintos, sendo que um deles cumpre os objetivos propostos, nomeadamente ao nível do acerto e velocidade na detecção de diferentes anúncios publicitários, o que leva a considerar a possibilidade da sua utilização num sistema real.

Após discussão dos resultados obtidos, são ainda destacadas algumas medidas tendo em vista o aprimoramento dos métodos propostos.



# Abstract

The automatic detection of advertising contents in television broadcasting is a problem approached since many decades ago in different studies with different interests for the broadcasting companies, for the advertised companies or the consumers themselves. One of the most interesting cases approaches the substitution of commercials with different ones more targeted to the final consumer in situations of contents redistribution like online broadcasting. However, there are several difficulties associated to this task.

In this dissertation are proposed two video fingerprint based methods to automatically perform the detection of advertising contents in a broadcasting environment, taking into account the temporal and computational constraints associated to a linear broadcasting context. For the generation and storing of fingerprints it is considered the previous knowledge of the advertising segments to be broadcasted while the Difference Hash - dHash technique for video summary generation and SQLite database strategy are also approached.

For the testing and evaluation of the proposed methodology a dataset of television streams is used and an evaluation method is introduced with the goal of analyzing the obtained results in crucial points to the achievement of the proposed objectives. The developed methods present distinct results, one of which fulfills the proposed objectives, namely at the level of accuracy and speed in the detection of different commercials, which leads to the consideration of its use in a real system.

After the discussion of the results, some measures are also highlighted in order to improve the proposed methods.



# Agradecimentos

O meu maior agradecimento é feito aos meus pais pelo esforço que realizaram para que eu pudesse realizar os meus sonhos, por toda a paciência que tiveram comigo mesmo nos tempos mais difíceis e por tudo o que são na minha vida.

Um agradecimento também a toda a minha família por todo o bom ambiente em que pude crescer e aprender para chegar onde me encontro.

Aos meus orientadores, Professor Pedro Carvalho e Professor Luís Corte-Real, com quem tive o prazer de trabalhar, por toda a confiança e motivação demonstrada ao longo destes últimos meses, que tiveram um peso fundamental para o sucesso desta dissertação.

Aos colaboradores do INESC TEC, pelo bom ambiente em que pude trabalhar e pela maneira como fui sempre bem recebido, em especial ao Hooshir Zolfagharnasab por todo o apoio e amizade demonstrados durante a realização desta dissertação.

Aos meus amigos de infância com quem tive oportunidade de crescer e conviver e a todas as excelentes pessoas que tive oportunidade de conhecer durante a faculdade e que tanto me ajudaram e contribuíram para que eu fosse uma pessoa mais feliz nestes anos que foram difíceis mas que sem dúvida valeram a pena.

A todos, um enorme obrigado!

Tiago Pereira





*“However difficult life may seem, there is always something you can do and succeed at.  
It matters that you don’t just give up.”*

Stephen Hawking



# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Contexto e Motivação . . . . .	1
1.2	Objetivos . . . . .	2
1.3	Estrutura da Dissertação . . . . .	3
<b>2</b>	<b>Revisão Bibliográfica</b>	<b>5</b>
2.1	Caraterísticas de Publicidades . . . . .	5
2.2	Deteção de Conteúdos Publicitários Baseada em Visão . . . . .	6
2.3	Deteção Multimodal de Conteúdos Publicitários . . . . .	7
2.4	Deteção de Conteúdos Publicitários Baseada em Reconhecimento . . . . .	8
2.4.1	<i>Fingerprinting</i> de Vídeo . . . . .	9
2.4.2	Comparação de <i>Fingerprints</i> . . . . .	11
<b>3</b>	<b>Ambiente de Desenvolvimento</b>	<b>13</b>
3.1	Caraterização do <i>Dataset</i> . . . . .	13
3.2	Anotação de Caraterísticas . . . . .	15
3.3	Método de <i>Hashing</i> . . . . .	15
3.4	Estratégia de Base de Dados . . . . .	17
3.4.1	Estudo dos Objetivos . . . . .	17
3.4.2	Avaliação das Alternativas . . . . .	17
<b>4</b>	<b>Métodos de Deteção de Publicidade</b>	<b>19</b>
4.1	Métodos Desenvolvidos . . . . .	20
4.1.1	Método dos Coeficientes da DCT . . . . .	20
4.1.2	Método <i>Visual Rhythm</i> . . . . .	21
<b>5</b>	<b>Teste e Avaliação</b>	<b>25</b>
5.1	Método de Avaliação . . . . .	25
5.2	Resultados Experimentais . . . . .	26
5.2.1	Testes com o Método dos Coeficientes da DCT . . . . .	27
5.2.2	Testes com o Método <i>Visual Rhythm</i> . . . . .	27
5.3	Discussão de Resultados . . . . .	28
<b>6</b>	<b>Conclusões</b>	<b>31</b>
6.1	Conclusões Gerais . . . . .	31
6.2	Trabalho Futuro . . . . .	32
	<b>Referências</b>	<b>33</b>



# Lista de Figuras

3.1	Imagens representativas das <i>streams</i> que compõem o <i>dataset</i> . . . . .	14
3.2	Método de <i>hashing dHash</i> . . . . .	16
3.3	Armazenamento de uma janela de <i>frames</i> consecutivos . . . . .	17
4.1	Visão geral dos algoritmos desenvolvidos . . . . .	19
4.2	Fase de geração de <i>fingerprints</i> do Método dos Coeficientes da DCT . . . . .	20
4.3	Fase de detecção do Método dos Coeficientes da DCT . . . . .	21
4.4	Exemplo da utilização do <i>Visual Rhythm</i> . . . . .	22
4.5	Conjugação da concatenação de três sequências de pixels . . . . .	22
4.6	Fase de geração de <i>fingerprints</i> do Método <i>Visual Rhythm</i> . . . . .	23
4.7	Fase de detecção do Método <i>Visual Rhythm</i> . . . . .	24



# Lista de Tabelas

3.1	Descrição geral do <i>dataset</i> . . . . .	14
3.2	Divisão dos conteúdos do <i>dataset</i> . . . . .	15
3.3	Tempos médios de pesquisa numa estrutura de dados . . . . .	18
3.4	Tempos médios de pesquisa numa base de dados SQLite . . . . .	18
5.1	Parâmetros utilizados no Método dos Coeficientes da DCT . . . . .	27
5.2	Resultados do Método dos Coeficientes da DCT . . . . .	27
5.3	Parâmetros utilizados no Método <i>Visual Rhythm</i> . . . . .	28
5.4	Resultados do Método <i>Visual Rhythm</i> . . . . .	28





# Abreviaturas e Símbolos

ALISP	Automatic Language Independent Speech Processing
API	Application Programming Interface
COG	Centroid of Gradient Orientations
DCT	Discrete Cosine Transform
DoG	Digital on-screen Graphics
ECR	Edge Change Ratio
HMM	Hidden Markov Models
LSH	Locality Sensitive Hashing
LUT	Lookup Table
RBT	Random Bases Transform
RGB	Red Green Blue
SVD	Singular-Value Decomposition
SVM	Support Vector Machine
TDT	Televisão Digital Terrestre



# Capítulo 1

## Introdução

### 1.1 Contexto e Motivação

A identificação e classificação de conteúdos televisivos é uma operação que tem ganho relevância ao longo dos anos, sendo importante em diversos aspectos para os intervenientes nas transmissões dos mesmos, no entanto, atualmente este processo é tipicamente realizado manualmente por um operador humano em ambiente profissional como parte do processo [1]. Os anúncios publicitários, sendo parte integrante dos conteúdos televisivos, são um caso de particular interesse e importância, nomeadamente do ponto de vista económico e da influência que estes têm sobre os telespetadores, afetando diretamente os seus estilos de consumo. Devido a estes fatores e tendo em conta a evolução das estratégias de difusão ao longo dos últimos anos torna-se pertinente o estudo e desenvolvimento de métodos automáticos que permitam a sua deteção.

Os conteúdos publicitários são uma parte crucial de qualquer transmissão televisiva, trazendo enormes benefícios à cadeia de televisão que os transmite, pois, tipicamente, trata-se da sua principal fonte de rendimento, o que leva a que as cadeias televisivas pretendam rentabilizar ao máximo os espaços dedicados aos anúncios publicitários de forma a aumentar as suas receitas e assim obter maiores lucros. A transmissão deste tipo de conteúdos está sujeita a diversos requisitos contratuais entre as cadeias de televisão e as empresas publicitadas, para as quais a transmissão dos anúncios é também de grande relevância por se tratar de uma importante ferramenta de *marketing* que permite atrair a atenção dos consumidores. Para além das variações dos custos de transmissão de um anúncio publicitário associadas à duração e ao horário de transmissão dos mesmos [2], geralmente, estas empresas têm ainda custos acrescidos relativamente à verificação de que estes anúncios são corretamente transmitidos, de acordo com o contratualizado [3]. Em [1] destaca-se a influência que os anúncios publicitários exercem nos consumidores, sendo evidente a fraca consideração por este tipo de conteúdo, salientando-se ainda o desenvolvimento de novos métodos de distribuição de conteúdos que têm em conta a substituição de anúncios publicitários por outros mais direcionados às preferências de cada telespetador. Do ponto de vista das audiências, torna-se também relevante evitar os conteúdos publicitários, por exemplo, em gravações televisivas pessoais [3], [4]. Ainda neste contexto, [5] apresenta um método que se baseia na deteção de intervalos publicitários para

realizar uma mudança de canal automática de modo a permitir aos telespectadores evitarem este tipo de conteúdo.

Tendo em conta as novas estratégias de difusão, as cadeias de televisão têm apostado cada vez mais em novas plataformas multimédia, como, por exemplo, as transmissões de conteúdos em plataformas *online*. Devido à contratualização relativa à transmissão de anúncios publicitários ser geralmente apenas referente à transmissão televisiva em si, as empresas de difusão pretendem reutilizar os espaços dedicados a publicidade nas transmissões *online*, ou em outras plataformas onde o material televisivo seja redistribuído, para transmitir anúncios publicitários diferentes, podendo desta maneira explorar melhor os interesses pessoais de cada consumidor, o que ao mesmo tempo leva ao aumento da relevância dos conteúdos publicitários transmitidos, abrindo assim portas à realização de novos contratos com o objetivo de rentabilizar ao máximo estes espaços, no entanto são diversos os problemas associados a este tipo de abordagem, entre os quais se destacam a pouca informação disponível relativamente aos conteúdos das transmissões televisivas originais, o que dificulta o processo da sua segmentação e deteção [6]. Um outro desafio que este tipo de abordagem enfrenta refere-se às restrições temporais existentes para realizar todo o processo de deteção e substituição de conteúdos, que deve ser suficientemente rápido e eficaz, garantindo a integridade dos segmentos programáticos.

## 1.2 Objetivos

O problema consiste no desenvolvimento de um conjunto de algoritmos que permita a deteção automática dos segmentos de publicidade, previamente conhecidos e passíveis de serem difundidos numa transmissão televisiva, para que estes possam ser posteriormente substituídos numa situação de redistribuição de conteúdos, nomeadamente uma situação de transmissão *online*. Pretende-se que este processo seja realizado numa situação em tempo real para que os conteúdos publicitários transmitidos na plataforma de redistribuição sejam automaticamente substituídos por segmentos diferentes e ao mesmo tempo garantindo a integridade dos conteúdos programáticos. As dificuldades deste processo focam-se na necessidade de desenvolver um método de deteção de conteúdos que seja suficientemente preciso e rápido de modo a satisfazer a necessidade de solucionar este problema numa situação em tempo real, reaproveitando de uma forma eficaz os espaços dedicados a publicidade nas diferentes plataformas de difusão.

Para a realização desta dissertação foram definidos como objetivos principais, o estudo e implementação de um conjunto de técnicas de *fingerprinting* para proceder à deteção de conteúdos publicitários num ambiente de *broadcasting*, tendo em conta as restrições temporais e computacionais existentes num contexto de transmissão linear. Para tal, é efetuada uma contribuição inicial na anotação de características relevantes para a preparação de um *dataset* representativo que serve como elemento de teste e avaliação das soluções desenvolvidas.

## **1.3 Estrutura da Dissertação**

O restante conteúdo desta dissertação divide-se nos seguintes capítulos: No capítulo 2 é realizada uma revisão bibliográfica de modo a analisar a evolução de trabalhos relacionados com este tema. No capítulo 3 é apresentado o contexto em que o trabalho desenvolvido se insere. No capítulo 4 é realizada uma apresentação teórica dos métodos estudados e aplicados. No capítulo 5 são apresentados os resultados do trabalho realizado e a discussão dos mesmos. Por fim, no capítulo 6 são expostas algumas conclusões acerca do trabalho realizado bem como a verificação da satisfação dos objetivos propostos.



## Capítulo 2

# Revisão Bibliográfica

### 2.1 Características de Publicidades

A estrutura de um bloco de publicidades apresenta variações dependendo do canal ou do país onde é realizada a transmissão televisiva, o que é tido em conta em diferentes trabalhos, como por exemplo [7], que propõe um método de deteção de publicidades em tempo real no caso específico de canais de televisão na Alemanha, no entanto, em [8] é referido que a generalidade das características dos blocos publicitários são comuns nas transmissões televisivas noutros países. Neste artigo é realizada uma divisão entre dois tipos de características presentes nos blocos publicitários, que abordam aspetos de baixo nível como as janelas temporais dos anúncios, a presença de faixas monocromáticas negras (*black frames*) ou o aumento do volume de áudio durante a transmissão de publicidades (caraterísticas diretamente mensuráveis) e as que exploram atributos como quantidade de movimento e ação, a existência de edições de vídeo ou a presença de caixas de texto (caraterísticas indiretamente mensuráveis).

Em [4] é abordado com maior detalhe um sistema de classificação de conteúdos televisivos baseado na deteção de logótipos que são removidos durante os anúncios publicitários (analisando o caso concreto da televisão Espanhola), associado ao facto dos planos de vídeo (*video shots*) terem uma duração mais curta, também durante as publicidades. É referido que os logótipos podem ser opacos, transparentes ou animados, sendo que os animados são mais difíceis de detetar devido às variações dos valores dos pixels. Também [9] aborda a deteção de logótipos de canais de televisão, explorando as características estacionárias das regiões onde os logótipos estão localizados e destaca a utilidade da sua localização em diferentes aplicações como a deteção de anúncios publicitários ou a monitorização de transmissões televisivas.

Na proposta para a deteção e segmentação de blocos publicitários apresentada em [10] é destacado que caraterísticas como faixas monocromáticas negras entre anúncios publicitários ou a presença do logótipo do canal durante conteúdos programáticos pode não apresentar bons resultados em certas regiões, com destaque para alguns países Asiáticos. Em alternativa, é explorada a presença de blocos de texto sobrepostos, que ao contrário dos conteúdos programáticos, ocorrem com grande frequência em anúncios publicitários. Também em [11] é salientada a pouca robustez

apresentada por métodos que se baseiam na existência de faixas monocromáticas antes de anúncios publicitários, sendo apresentado em alternativa um método baseado em reconhecimento de deteção de anúncios previamente conhecidos e armazenados numa base de dados e deteção de possíveis anúncios não armazenados.

Em análise a dados do projeto [12], relativamente a quatro canais portugueses (dois públicos e dois privados), verificou-se uma menor percentagem de tempo dedicado à transmissão de anúncios publicitários nos canais públicos. É também de destacar a constante existência do logótipo referente ao canal de televisão (que pode conter variações ao nível do pixel) num dos cantos da imagem durante conteúdos programáticos, assim como a frequente existência de faixas monocromáticas negras (*black frames*) associadas a períodos de silêncio nas transições entre anúncios publicitários, no entanto este tipo de característica também ocorre por vezes durante conteúdos programáticos.

## 2.2 Deteção de Conteúdos Publicitários Baseada em Visão

No domínio visual, podem ser encontradas na literatura abordagens a diferentes características e métodos na deteção e reconhecimento de publicidades em conteúdos de difusão. Na abordagem de [8], inicialmente são identificados no vídeo de interesse os segmentos onde potencialmente existem blocos publicitários. Para detetar as suas fronteiras foi considerado o facto de todos os anúncios individuais estarem separados entre si por curtas sequências de *frames* monocromáticos negros. São ainda abordados cortes de cena (*scene breaks*), como cortes rígidos (*hard cuts*) e desvanecimentos (*fades*), como características chave de modo a identificar segmentos publicitários. No caso dos cortes rígido, uma alta frequência de cortes é um indicador de que um segmento possa ser uma publicidade e podem ser monitorizados através da análise dos histogramas de cores entre *frames* consecutivas.

Em [13] é proposto um método baseado na deteção de *black frames* associadas a períodos de silêncio, o que requereria a descodificação de vídeo a partir do formato comprimido MPEG-1, de modo a aceder às amostras de pixeis e de áudio do vídeo. No entanto, devido a esta abordagem apresentar um peso computacional elevado é proposto em alternativa a análise dos coeficientes da Transformada Discreta do Cosseno (*DC-DCT*) e do peso dos fatores de escala das sub-bandas de frequência do sinal de áudio.

A técnica de deteção de publicidades baseada na presença de logótipos em conteúdos programáticos, abordada em [4], baseia-se na deteção de uma área com contornos estáveis, restringida aos quatro cantos da imagem. Esta técnica associa a presença do logótipo à duração do plano em que o mesmo está presente, apresentando uma vantagem relativamente a outras técnicas de deteção de logótipos, pois pode ser realizada sem qualquer intervenção humana. Numa abordagem específica sobre canais de televisão generalistas Portugueses, em [14] é explorada a presença do logótipo do canal de televisão, sendo para tal utilizado um mecanismo de deteção e distinção de gráficos digitais no ecrã (*Digital on-screen Graphics - DoG*). Nesta proposta são realçadas algumas dificuldades associadas à deteção do logótipo do canal de televisão como a variação de cor,



posição ou alterações temporárias durante, por exemplo, épocas festivas. Em relação a outros métodos propostos apresenta a vantagem de não necessitar de uma base de dados de *DoGs* previamente construída e com o tempo de processamento apresentado pode permitir a sua utilização numa situação em tempo real.

Em [15] é apresentada uma abordagem que se baseia na deteção das fronteiras dos segmentos publicitários através da análise de cortes por minuto de modo a identificar as variações de planos dentro e entre segmentos publicitários respetivamente. A isto é ainda associada uma análise às variações temporais de cor e luz em planos de vídeo consecutivos. O método apresentado em [16] também se centra na deteção de mudanças no plano de vídeo, recorrendo à ferramenta de abstração de vídeo *visual rhythm*, que é uma representação sub-amostrada do conteúdo de um vídeo, permitindo a deteção e classificação de diversos efeitos de vídeo, de forma a aumentar a robustez do algoritmo proposto e assim permitir uma maior percentagem de deteções corretas.

Dos principais modos de análise de conteúdos de multimédia que são explorados em [17] destacam-se a deteção de objetos (como texto ou faces), deteção de limites nos segmentos de vídeo e ainda a classificação destes em géneros através de *Hidden Markov Models - HMM* [18].

Relativamente aos métodos utilizados no cálculo de similaridades entre cenas (sequências de planos de vídeo com conteúdos do mesmo contexto) para realizar a segmentação temporal de vídeos, [19] salienta que em geral estes métodos dividem-se em três fases: extração de características ao nível do *frame*, medição de similaridades entre *frames* e deteção de limites entre *frames* dissimilares. De entre as características extraídas são destacadas o histograma de cor ([20] explorou mais aprofundadamente esta característica usando histogramas em vários espaços de cores), *Edge Change Ratio - ECR* e vetores de movimento. Quanto às abordagens para a deteção de limites entre *frames*, podem ser baseadas no valor de similaridade medido comparativamente a um valor de *threshold* ou baseadas numa aprendizagem estatística, nas quais os planos de vídeo são classificados como variações de planos, dependendo das características que contêm. Numa outra abordagem, [21] propôs o uso de matrizes de auto-semelhança para a deteção de fronteiras de cenas de vídeo, que representam graficamente sequências similares numa série de dados, o que apresenta um melhor desempenho quando comparado com métodos que usam medidas de diferença entre histogramas.

### 2.3 Detecção Multimodal de Conteúdos Publicitários

Os métodos de deteção de conteúdos publicitários baseados em visão podem tornar-se pouco eficientes, pois, em geral, são computacionalmente exigentes, o que torna importante a exploração de métodos que abordem outro género de características, como por exemplo as características da componente áudio, que em termos computacionais são menos complexas, no entanto, também estas apresentam alguns problemas. De modo a contornar estes problemas, diferentes autores propuseram abordagens multimodais que combinam as características visuais com as características de áudio extraídas.

Em [22] são utilizados os modos acústico e visual com o objetivo de verificar a classificação de diferentes segmentos de vídeo em diferentes gêneros como desenhos animados, desporto ou notícias e verificou-se um melhor funcionamento quando utilizadas características acústicas comparativamente com o domínio visual, sendo que a combinação de ambos os domínios apresenta ainda melhores resultados ao nível da precisão da classificação.

Abordando diferentes métodos de deteção de publicidade, em [23] são utilizados dois métodos de deteção de publicidades de modo a complementarem-se. O primeiro método aborda a repetição de publicidades ao longo de uma transmissão enquanto que o segundo método aborda as diferenças das características básicas ao nível da cor e do áudio entre conteúdos publicitários e programáticos. O primeiro método apresenta melhores resultados ao nível da precisão enquanto que o segundo método apresenta melhores valores de *recall*.

Na abordagem para a deteção de segmentos publicitários de [24] é proposto um método baseado em aprendizagem onde são extraídas diferentes características visuais e de áudio básicas, a partir das quais é derivado um conjunto de características baseadas em contexto, mais eficientes na deteção de publicidades, sendo que a classificação de segmentos é posteriormente realizada através de um classificador *Support Vector Machine - SVM*, apresentando uma elevada precisão de deteções, tanto de conteúdos programáticos, como publicitários.

Em [6] é apresentada uma proposta de um método multimodal tendo em vista a deteção e substituição de publicidades através da pesquisa de segmentos de vídeo repetidos. O método proposto aborda, em primeiro lugar, uma fase onde são detetadas repetições a partir das características de áudio que são posteriormente validadas recorrendo a descritores visuais. Por fim são procurados os limites de cada segmento de modo a determinar os seus pontos inicial e final. Este método foi desenvolvido tendo em vista a eficiência computacional de todo o processo e apresenta bons resultados perante situações de ruído e variações.

## 2.4 Deteção de Conteúdos Publicitários Baseada em Reconhecimento

Como analisado anteriormente, algumas propostas de deteção e segmentação de conteúdos publicitários recorrem a métodos baseados em reconhecimento. Este tipo de deteção é abordado em [3] onde é destacado o elevado peso computacional associado à pesquisa numa base de dados de vídeo e ainda que a maioria dos algoritmos de pesquisa utiliza uma característica de baixo nível para procurar, numa primeira fase, um grupo de vídeos candidatos, diminuindo assim o peso computacional associado à pesquisa na base de dados.

Para a monitorização da transmissão de anúncios publicitários, [25] destacou que várias companhias, que realizam essa tarefa, introduzem marcas de águas (*watermarking*) nos segmentos publicitários para proceder à sua monitorização automática, num processo de elevado custo e limitado apenas à deteção dos anúncios inseridos pela própria empresa. Em [26] é salientado ainda que no processo de *watermarking*, as marcas introduzidas nos segmentos de vídeo correm o risco de ser distorcidas perante as transformações no formato do vídeo ou mesmo na transmissão dos mesmos.

O método proposto por [6] procede à deteção de potenciais repetições de conteúdos publicitários na faixa de áudio, que após serem validadas e segmentadas são armazenadas numa base de dados. Este método de deteção é adaptado a partir de [27], onde foi proposto um método para identificar uma música através do reconhecimento de apenas alguns segundos de áudio sujeito a ruído, com o objetivo de tornar todo o processo computacionalmente eficiente. Também no domínio do áudio, [28] abordou a indexação automática com o objetivo de detetar anúncios publicitários conhecidos numa *stream* de rádio, propondo o uso da ferramenta *Automatic Language Independent Speech Processing - ALISP* associada a *Hidden Markov Models - HMM*.

Para a deteção de novos segmentos publicitários, que não sejam anteriormente conhecidos e armazenados, [29] apresenta uma proposta baseada na monitorização contínua de uma transmissão televisiva. O sistema proposto começa por realizar a segmentação temporal de vídeo, dividindo-o em planos individuais, procedendo à deteção de planos de vídeo previamente conhecidos, através de um método baseado em *hashing*, que são posteriormente classificados como conteúdo publicitário ou programático. Na fase de identificação de sequências de vídeo repetidas, esta proposta combina planos de vídeo adjacentes que são posteriormente identificados e classificados através de um processo de segmentação que confere a possibilidade de ser utilizado num sistema em tempo real.

### 2.4.1 *Fingerprinting* de Vídeo

Associado à deteção e segmentação de conteúdos publicitários, nas situações em que ocorre repetição de conteúdos, verifica-se em vários dos métodos analisados a aplicação de técnicas de *fingerprinting*, que, numa alusão ao método de identificação de pessoas através de uma impressão digital, no domínio do vídeo consiste na criação de um resumo do mesmo a partir do qual possa ser identificado. Na abordagem realizada por [8], é analisada a possibilidade de armazenar anúncios publicitários, que normalmente permanecem no dia-a-dia das transmissões televisivas por longos períodos de tempo, de modo a facilitar a deteção dos limites dos mesmos no processo de reconhecimento. Neste processo de reconhecimento é abordada a necessidade de criar uma base de dados para armazenar os segmentos publicitários assim como gerar as *fingerprints* dos mesmos de maneira a aumentar a eficiência do processo de comparação. Em [11] é abordada a deteção de segmentos publicitários desconhecidos, propondo um método que armazena, num *buffer*, novas *fingerprints* de potenciais segmentos publicitários desconhecidos, que servem de comparação para a deteção de uma possível repetição desse mesmo segmento publicitário, que, caso ocorra, permita a sua identificação como um novo anúncio a adicionar a uma base de dados de publicidades já conhecidas.

Numa abordagem a este tema e dando especial foco ao domínio de vídeo, [26] identifica as técnicas de *fingerprinting* como sendo uma poderosa ferramenta na deteção e gestão de cópias, que enfrenta uma série de desafios como distorções causadas pela mudança de formato dos vídeos que afetam métodos que se baseiam em informações de características básicas das imagens. É também salientada a elevada importância que as características a utilizar em *fingerprinting* têm, pois a sua escolha afeta diretamente todo o desempenho do sistema. São ainda destacadas várias

características, propostas na literatura, para gerar *fingerprints*, como histogramas de cor ou cor dominante, realçando que as mesmas são pouco robustas perante variações de cor ou intensidade de luz. Os diferentes tipos de características podem ser classificados em três dimensões, sendo elas cor, espaço e tempo, (em [30] é identificada ainda uma quarta dimensão, no domínio das transformadas) sendo que a utilização de um maior número de dimensões na geração de uma *fingerprint* oferece-lhe uma maior robustez. É também destacada uma classificação de características, que se refere à informação extraída das imagens, que pode ser global, extraindo informações de toda uma imagem, ou local, extraindo informações de apenas algumas partes de uma imagem, permitindo assim a exploração de pontos-chaves da mesma. Numa referência a [31], são destacadas três características fundamentais para a obtenção de uma *fingerprint* de qualidade, sendo elas a robustez perante distorções, a independência entre *fingerprints* e o modo como as mesmas afetam a eficiência de uma posterior pesquisa numa base de dados.

Na pesquisa realizada em [30], são caracterizadas as diferentes dimensões utilizadas no desenvolvimento de *fingerprints* de vídeo:

- ***Fingerprint* de Cor**

Nas *fingerprints* de cor é calculado o histograma para as diferentes componentes do espaço de cores YUV, servindo a interseção dos mesmos como base para a medição de similaridades.

- ***Fingerprints* Espaciais**

Neste tipo de *fingerprints*, é caracterizada a componente da luminância de uma imagem obtida através da conversão da mesma no espaço de cores YUV, a partir da qual é gerada uma grelha de blocos de tamanho fixo, originando uma *fingerprint* compacta. Em alternativa, é destacado o cálculo de *fingerprints* espaciais em torno de pontos de interesse em *frames* chave, que pode gerar *fingerprints* de diferentes tamanhos, tornando este método computacionalmente mais complexo comparativamente com o anterior.

Comparativamente com os outros tipos de *fingerprints*, as espaciais são mais compactas e apresentam melhores resultados em termos de robustez, discriminabilidade e complexidade computacional.

- ***Fingerprints* Temporais**

As *fingerprints* temporais consistem na duração dos diferentes planos de um vídeo, obtida através da segmentação do mesmo, sendo que a sequência da duração destes planos gera a *fingerprint* temporal, tornando assim importante a deteção de *frames* chave, como as das fronteiras em que ocorrem mudanças de plano.

- ***Fingerprints* no Domínio das Transformadas**

Diferentes métodos calculam *fingerprints* no domínio de transformadas, devido a propriedades invariáveis que estas oferecem. Como exemplo, é referido o uso da Transformada de Fourier, Transformada de Radon ou Decomposição em Valores Singulares (*Singular-Value*

*Decomposition - SVD*), para a geração de *hashes* robustas a variações geométricas, como rotações ou translações. Como alternativa, surge ainda o uso da Transformada Discreta do Cosseno (*Discrete Cosine Transform - DCT*) ou da Transformada de Bases Aleatórias (*Random Bases Transform - RBT*), ambas associadas a técnicas de *hashing*.

No método proposto por [32] para identificação de vídeo baseada em conteúdo, inicialmente no processo de extração de *fingerprints* o vídeo de entrada é reamostrado a uma taxa (*frame rate*) fixa e depois cada *frame* é convertido para escala de cinza (*grayscale*) e o seu tamanho ajustado de modo a normalizar a sua altura e largura. Por fim, cada *frame* é dividido numa grelha de blocos, a partir dos quais é calculado o Centróide de Orientação de Gradientes (*Centroid of Gradient Orientations - COG*) para cada um deles, gerando assim a sua *fingerprint*. Este método apresenta a vantagem de ser robusto perante variações da taxa de *frames*, variações de cor e brilho e variações de tamanho, mostrando assim melhores resultados comparando com outras características utilizadas noutros métodos de *fingerprinting* desenvolvidos até à data da sua publicação.

No processo de extração de *fingerprints* proposto em [33], os *frames* de vídeo são inicialmente processados de modo a tornar as suas características mais robustas e com a qualidade necessária para realizar uma extração de *fingerprints* de qualidade. Posteriormente, para cada *frame*, são calculados mapas a partir de características visuais, como cor ou intensidade, que após serem normalizados, são combinados para formar um único Mapa de Saliência que é subsequentemente dividido numa grelha de blocos, para cada qual é calculada a média do valor de saliência, sendo assim robusto perante vários tipos de distorções. De um modo semelhante aos objetivos propostos nesta dissertação, neste artigo é ainda proposto um método de comparação de *fingerprints* associado a uma estratégia de gestão de base de dados de modo a acelerar o processo de comparação (*matching*) e recuperação (*retrieval*).

#### 2.4.2 Comparação de *Fingerprints*

No âmbito da deteção de cópias, no qual as técnicas de *fingerprinting* se inserem e após o armazenamento das mesmas numa base de dados, é necessário proceder à comparação das mesmas através de algoritmos de *matching*, sendo importante que estes sejam rápidos e eficientes ao nível dos recursos computacionais que exigem. Na literatura, são diversos os métodos que abordam este tema, destacando-se os apresentados em seguida.

O processo de comparação é visto em [30] como um problema de pesquisa de similaridades, às quais encontra-se geralmente associada uma métrica como forma de quantificação. São exemplo de medidas de similaridade a distância da Manhattan e a distância Euclidiana ou ainda a distância de Hamming (assinaturas binárias). Como forma de complemento a estas medidas de similaridade são ainda destacadas as taxas relativas à precisão e *recall* (*precision rate* e *recall rate*) como medidas de discriminabilidade e robustez, respetivamente. De um modo simplificado, recorrendo ao método de pesquisa exaustiva, é descrito que o processo de *matching* consiste na pesquisa numa base de dados por *fingerprints* candidatas com uma distância de similaridade o mais próxima possível da *fingerprint* alvo, que são posteriormente sujeitas a um processo de decisão. No entanto,

devido às grandes proporções que a base de dados de *fingerprints* pode tomar e ao elevado peso computacional associado ao cálculo das medidas de distância ou similaridade entre a *fingerprint* a ser pesquisada e as que estão armazenadas na base de dados, este método de pesquisa exaustiva torna-se impraticável, sendo explorado em alternativa um método de pesquisa aproximada como o *Locality Sensitive Hashing - LSH*, anteriormente abordado em [34] e [35], que tem como objetivo a pesquisa da *fingerprint* de valor mais próximo através de uma ligeira e negligenciável redução na precisão de pesquisa que permite acelerar a velocidade de pesquisa e diminuir consideravelmente a complexidade computacional deste processo, podendo assim ser considerada a sua utilização num sistema em tempo real como o proposto.

A estratégia de base de dados discutida em [33] utiliza a distância de Hamming como medida de similaridade e consequente classificação das *fingerprints* que são compostas por um conjunto de dados binários de tamanho fixo. É ainda apresentada uma estratégia para aumentar a eficiência de pesquisa na base de dados, caso esta contenha um elevado número de *fingerprints* armazenadas, recorrendo a uma *Lookup Table - LUT* que apresenta baixos custos computacionais.

Numa alternativa mais recente, em [36] são utilizados métodos de *hashing*, para gerar *fingerprints* mais robustas ao nível da aplicabilidade, discriminação e escalabilidade, permitindo assim uma maior eficiência das abordagens utilizadas na pesquisa de *fingerprints* similares. *Locality Sensitive Hashing - LSH* é um dos métodos destacados, no entanto, sendo inconsciente relativamente à informação da estrutura de dados torna-se ineficiente, mas pode utilizar múltiplas tabelas de *hash* de modo a obter uma maior precisão. São também referidos vários métodos que preservam a informação da estrutura de dados recorrendo a ferramentas de aprendizagem de máquina (*machine learning*), tornando assim o código de *hash* e a respetiva pesquisa de similaridade mais eficientes e precisos.

## Capítulo 3

# Ambiente de Desenvolvimento

De modo a testar e avaliar os métodos de deteção de publicidade desenvolvidos, foi utilizado um conjunto de transmissões televisivas utilizado no âmbito do projeto de I&DT [12]. Neste capítulo é apresentada uma descrição do *dataset* assim como uma análise aos dados extraídos a partir da sua anotação manual. São ainda abordadas em detalhe algumas ferramentas utilizadas nos métodos desenvolvidos como o método de *hashing* utilizado e o modo como a base de dados foi preparada e estruturada.

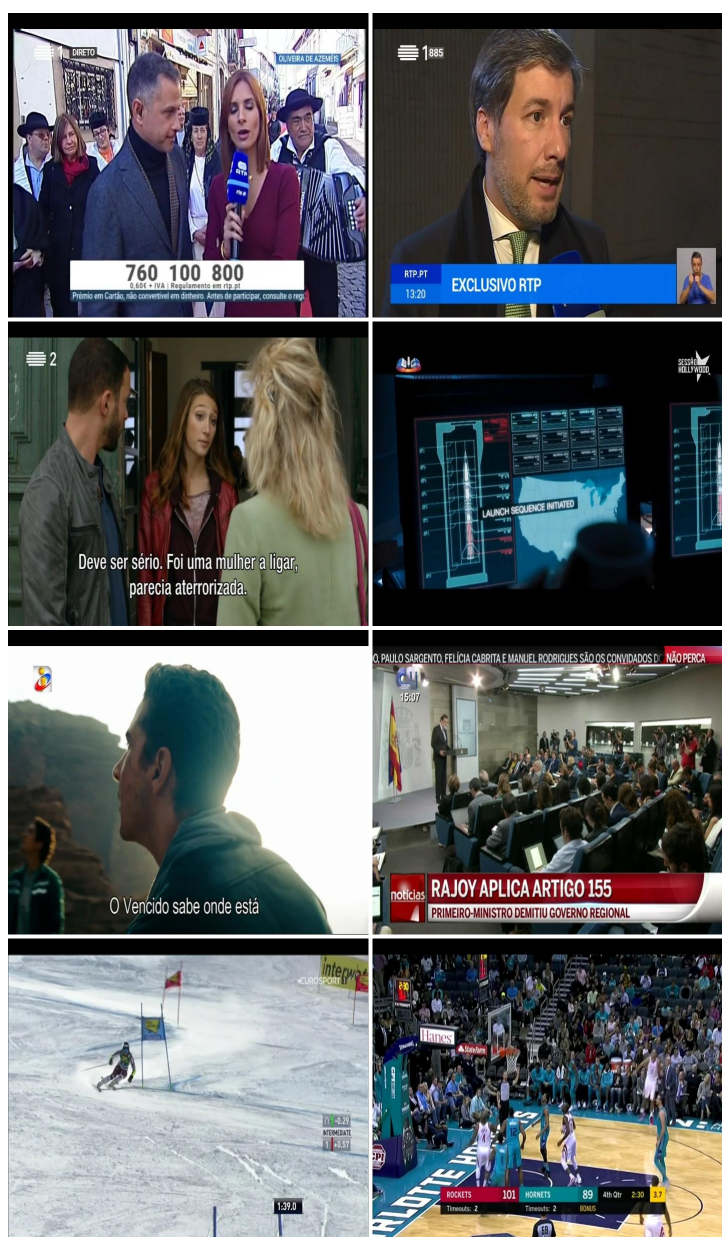
### 3.1 Caraterização do *Dataset*

O *dataset* de *streams* televisivas disponibilizado é composto por segmentos de transmissões televisivas de sete canais de televisão diferentes recolhidos a partir da Televisão Digital Terrestre (TDT) e a partir de um operador de cabo privado português:

- A partir da TDT, foram recolhidas transmissões de quatro canais portugueses: RTP1, RTP2, SIC e TVI. Estes canais transmitem conteúdos generalistas, contendo uma combinação razoável entre conteúdos programáticos e publicitários, pelo que o seu uso torna-se pertinente tendo em conta os objetivos propostos. De notar que foram recolhidos e utilizados dois segmentos diferentes relativos ao canal de televisão RTP1.
- A partir de um operador de cabo privado, foram recolhidas transmissões de três canais: CMTV HD, Eurosport HD e NBA TV. Ao contrário da CMTV HD que transmite conteúdos generalistas, a Eurosport HD e a NBA TV são canais temáticos dedicados à transmissão de conteúdos desportivos.

De modo resumido, a tabela 3.1 apresenta as caraterísticas principais do *dataset* recolhido, que foi posteriormente anotado para ser utilizado no teste e avaliação dos métodos de deteção de publicidade desenvolvidos. Na figura 3.1 são apresentadas imagens representativas de um *frame* de cada uma das *streams* que compõem o *dataset* (da esquerda para a direita, de cima para baixo: RTP1a, RTP1b, RTP2, SIC, TVI, CMTV, Eurosport, NBA TV).

Canal	País de Origem	Tipo de Conteúdo	Duração
<b>RTP1a</b>	Portugal	Generalista	6h52m
<b>RTP1b</b>	Portugal	Generalista	1h10m
<b>RTP2</b>	Portugal	Generalista	8h25m
<b>SIC</b>	Portugal	Generalista	8h37m
<b>TVI</b>	Portugal	Generalista	8h37m
<b>CMTV</b>	Portugal	Generalista	0h42m
<b>Eurosport</b>	Portugal	Desporto	0h42m
<b>NBA TV</b>	E.U.A.	Desporto	0h42m

Tabela 3.1: Descrição geral do *dataset*Figura 3.1: Imagens representativas das *streams* que compõem o *dataset*



## 3.2 Anotação de Caraterísticas

Após a recolha das *streams* que compõem o *dataset*, procedeu-se à anotação das mesmas, num processo manual onde foram registados os períodos e respetiva contagem de *frames* onde ocorrem as fronteiras entre blocos programáticos e publicitários e entre os diferentes anúncios dentro dos blocos publicitários, ou seja, o número dos *frames* inicial e final de cada segmento publicitário e programático, assim como algumas informações básicas e subjetivas acerca de cada segmento individual. A informação obtida sobre os *frames* iniciais e finais de cada segmento é fundamental para efetuar o processo de corte e segmentação dos diferentes conteúdos necessários para o desenvolvimento dos métodos baseados em reconhecimento propostos. Neste processo de segmentação de vídeo foi utilizado o *software* FFmpeg que oferece um vasto leque de ferramentas de multimédia e edição de vídeo. A tabela 3.2 revela uma análise ao modo como as *streams* do *dataset* se dividem e são constituídas.

Canal	Duração	Percentagem de Publicidade	Segmentos Publicitários	Percentagem de Programa	Segmentos Programáticos
RTP1a	06:51:36	7,02%	53	92,98%	8
RTP1b	01:10:23	15,07%	22	84,93%	3
RTP2	08:25:21	6,30%	67	93,70%	14
SIC	08:36:51	26,72%	413	73,28%	11
TVI	08:36:39	24,97%	347	75,03%	14
CMTV	00:42:00	14,23%	17	85,77%	3
Eurosport	00:42:00	19,25%	16	80,75%	4
NBA TV	00:42:26	29,34%	30	70,66%	5

Tabela 3.2: Divisão dos conteúdos do *dataset*

A partir desta análise, verifica-se que de um modo geral, nos diferentes canais de televisão que constituem o *dataset*, a maior percentagem da transmissão é dedicada a conteúdos programáticos e ainda que os blocos de conteúdos publicitários se encontram divididos numa grande quantidade de segmentos.

## 3.3 Método de *Hashing*

A estratégia utilizada para a deteção e identificação de segmentos de publicidade recorre a técnicas de resumo de vídeo tendo a vista a geração de *fingerprints* identificativas dos vários segmentos de vídeo, para tal foram estudadas diferentes estratégias de *hashing*. Com uma estratégia de *hashing* é possível originar um resumo de vídeo que pode ser usado para gerar a sua *fingerprint* identificativa de baixa complexidade. Esta abordagem é efetuada de modo a melhorar a eficiência de uma pesquisa numa base de dados e tendo em vista a necessidade de proceder a esta tarefa no caso de deteção de conteúdos durante uma transmissão televisiva em tempo real.

Resumidamente, as técnicas de *hashing* utilizam informações relativas ao conteúdo e características de uma imagem para gerar um código identificativo simplificado e fácil de armazenar e pesquisar. Neste contexto, recorreu-se ao método de *Hash* de Diferença - *dHash*. Este método procede inicialmente a um redimensionamento do tamanho da imagem de entrada para um tamanho  $9 \times 8$  e a uma conversão desde o modelo de cor RGB para a escala de cinza (*grayscale*) de modo a contornar variações de cor. Posteriormente é efetuado o cálculo das diferenças entre pixels adjacentes, o que após as duas reduções iniciais de tamanho e cor, dá origem a uma imagem de tamanho  $8 \times 8$  que após ser binarizada (cada bit é convertido para 0 ou 1 comparando o nível do brilho entre dois pixels adjacentes) fica com um tamanho total correspondente a 64 bits. Escolhendo o sentido da esquerda para a direita e de cima para baixo, estes bits são então convertidos numa *hash* final de 16 dígitos hexadecimais, tal como se pode verificar através da figura 3.2.

Comparado com outros métodos de *hashing* como *Hash* de Média - *aHash* ou *Hash* Percetiva - *pHash*, o método *dHash* apresenta uma melhor combinação de resultados ao nível da velocidade e precisão. Enquanto que a *aHash* apresenta vantagens a nível da velocidade (consiste numa redução do tamanho da imagem e posterior binarização através da simples comparação de cada pixel com o valor médio de intensidade de todos os pixels da imagem reduzida), por outro lado tem um baixo nível de precisão. A *pHash* apesar de ser mais precisa, apresenta baixa velocidade (de um modo semelhante a *aHash* recorre ao cálculo da média, mas apresenta uma maior robustez devido ao uso da Transformada Discreta do Cosseno, de modo a reduzir a redundância da imagem, obtendo-se uma matriz de coeficientes, a partir dos quais é realizada a binarização da imagem) [37].

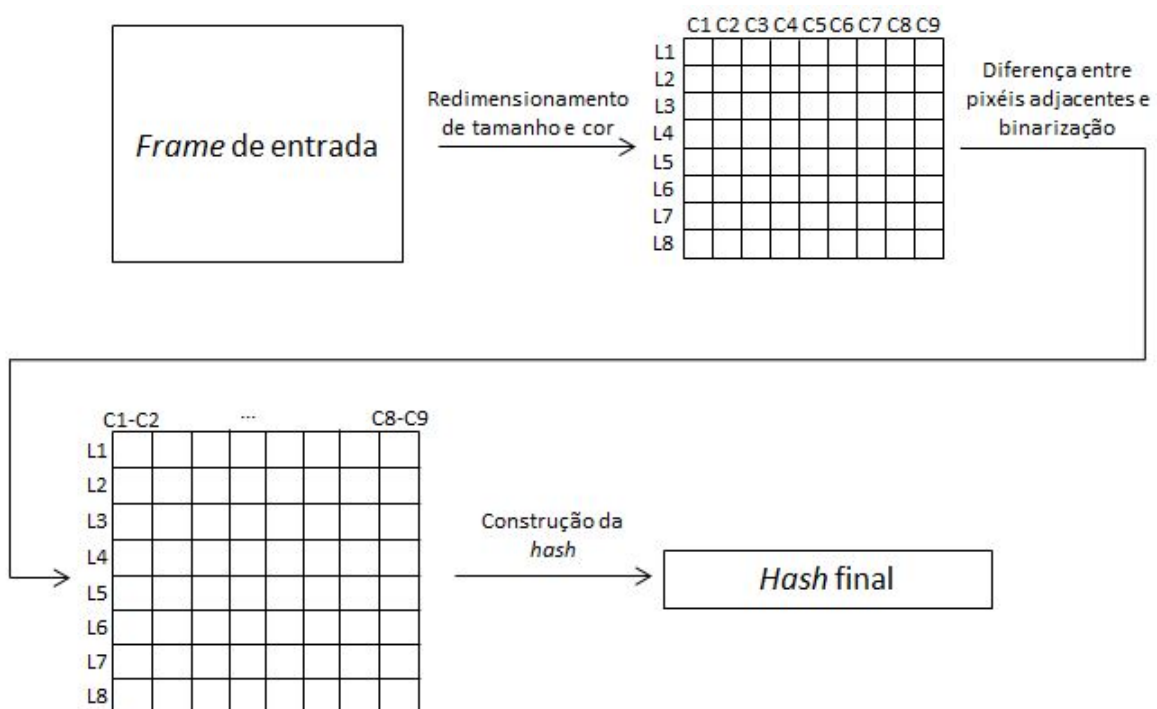


Figura 3.2: Método de *hashing* *dHash*

## 3.4 Estratégia de Base de Dados

### 3.4.1 Estudo dos Objetivos

Após a geração dos resumos de vídeo relativos aos segmentos a serem identificados durante a fase de deteção de cada um dos métodos desenvolvidos, é necessário proceder ao armazenamento dos dados gerados. No entanto, neste problema em específico, devido à necessidade de atingir o objetivo de deteção numa transmissão em tempo real, este aspeto é de elevada importância pois para além da utilização de um método de *hashing* rápido é também necessário construir uma interface de armazenamento de dados que permita operações de alta velocidade, nomeadamente ao nível da pesquisa.

Tendo em conta a elevada quantidade de *frames* presente numa *stream* televisiva, gerar e armazenar uma *fingerprint* para cada *frame* tornaria todo o processo de deteção lento e pouco eficiente. Torna-se então benéfico gerar *fingerprints* para conjuntos de *frames* consecutivos, tal como ilustrado na figura 3.3, o que permite diminuir consideravelmente a quantidade de dados armazenados assim como o peso computacional associado a este processo.

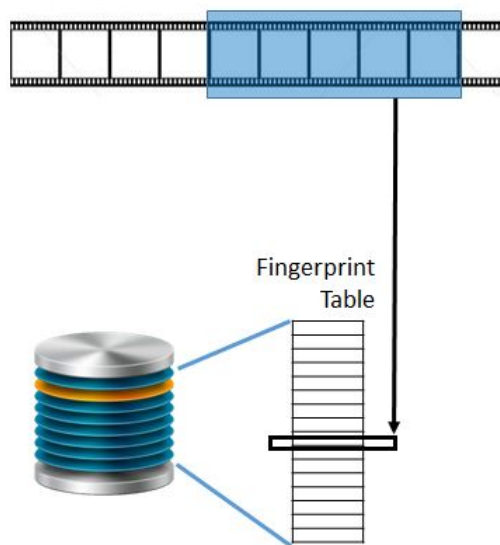


Figura 3.3: Armazenamento de uma janela de *frames* consecutivos

### 3.4.2 Avaliação das Alternativas

Uma alternativa possível para o problema de armazenamento e pesquisa de *fingerprints* propõe que estas sejam guardadas numa estrutura de dados, no entanto, perante a necessidade de armazenar um elevado número de entradas, esta opção pode não ser eficiente. A tabela 3.3 revela os resultados dos tempos médios de pesquisa por uma *fingerprint* numa estrutura de dados em duas situações de armazenamento de um elevado número de entradas.

<b>Número de <i>Fingerprints</i> Armazenadas</b>	4000	45000
<b>Tempo Médio de Pesquisa (<i>ms</i>)</b>	30	100

Tabela 3.3: Tempos médios de pesquisa numa estrutura de dados

Como alternativa, surge a possibilidade de usar um sistema de base de dados. Comparativamente com outras tecnologias de base de dados existentes, a tecnologia *open-source* SQLite apresenta diversas vantagens: não necessita de ser instalado ou configurado, pode ler ficheiros diretamente a partir do disco rígido sem necessidade de servidores, possibilitando o armazenamento de uma base de dados de grandes dimensões num único ficheiro e o armazenamento de dados em diferentes formatos. Para além de ser rápida, os dados são armazenados através de uma implementação *B-tree* [38], que tem uma complexidade de tempo logarítmica [39], é também eficiente e escalável, podendo ser usado em diferentes sistemas operativos e suportando diversas linguagens de programação com as quais se pode interagir com os ficheiros de base de dados através de uma API de fácil utilização [40], [41].

Através da implementação de uma base de dados SQLite, a tabela 3.4 demonstra os resultados dos tempos médios de pesquisa por uma *fingerprint* numa base de dados SQLite em duas situações de armazenamento semelhantes à realizada na alternativa anterior.

<b>Número de <i>Fingerprints</i> Armazenadas</b>	4000	45000
<b>Tempo Médio de Pesquisa (<i>ms</i>)</b>	5	10

Tabela 3.4: Tempos médios de pesquisa numa base de dados SQLite

Com esta análise, é possível verificar que o uso de um sistema de base de dados como SQLite é mais apropriado para a realização dos objetivos propostos, pois apresenta tempos médios de pesquisa suficientemente baixos perante um elevado número de dados armazenados que influenciam diretamente a necessidade de implementação de um método de deteção de conteúdos publicitários em tempo real.

## Capítulo 4

# Métodos de Detecção de Publicidade

Neste capítulo são apresentadas em detalhe as diferentes abordagens desenvolvidas para proceder à deteção automática de anúncios publicitários, tendo em conta os objetivos propostos. São propostos dois métodos baseados em *fingerprinting* de vídeo que apresentam características comuns mas que diferem essencialmente no modo de geração e armazenamento das *fingerprints* associadas aos diferentes segmentos publicitários.

Apresentando uma visão geral do trabalho desenvolvido, demonstrada através da figura 4.1, numa fase inicial foram anotadas características relevantes num conjunto de transmissões televisivas para a preparação do *dataset*. Com os dados recolhidos, o conjunto de vídeos que compõe o *dataset* foi segmentado de modo a permitir a utilização individual de cada segmento publicitário e programático. Através disto, dois métodos de deteção baseados em *fingerprinting* foram desenvolvidos para proceder à deteção dos anúncios publicitários numa situação de *broadcasting*.

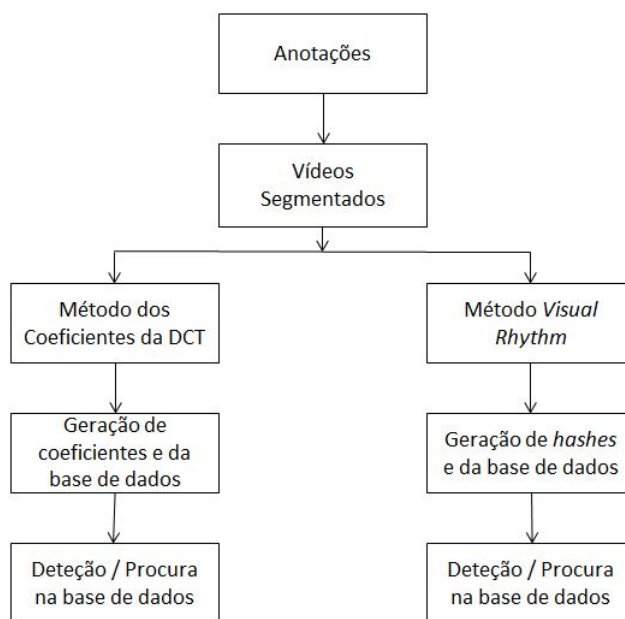


Figura 4.1: Visão geral dos algoritmos desenvolvidos

## 4.1 Métodos Desenvolvidos

De modo a compreender melhor os métodos desenvolvidos, considera-se que estes se dividem em duas fases: uma primeira fase onde são geradas e armazenadas as *fingerprints* dos vários segmentos publicitários que foram anteriormente segmentados, de modo a serem processados individualmente e uma fase final onde se procede à deteção dos diferentes segmentos numa transmissão televisiva.

### 4.1.1 Método dos Coeficientes da DCT

- Fase de geração de *fingerprints*

A implementação deste método aborda numa primeira fase um processo de resumo de vídeo onde se procede à geração de *fingerprints* dos diferentes anúncios publicitários e ao seu armazenamento na base de dados. Para tal, inicialmente é definido um tamanho fixo de uma janela de *frames* no qual o vídeo de entrada é dividido. Dentro de cada janela, cada *frame* é sujeito ao método de *hashing dHash*, gerando assim uma *hash* de caracteres hexadecimais para cada *frame*. Cada uma das *hashes* é convertida carácter a carácter para valores na escala de 0 a 15 (de modo a converter todas as *hashes* para valores numéricos inteiros), valores estes que após serem concatenados é aplicada a Transformada Discreta do Cosseno (DCT) o que gera uma série de coeficientes para cada uma das janelas de *frames*, tal como representado pela figura 4.2. Posteriormente um número fixo de coeficientes é selecionado e armazenado na base de dados como *fingerprint* de cada janela de *frames*, permitindo assim a identificação de diferentes partes de um anúncio publicitário durante a fase de deteção, o que consequentemente torna essa fase mais robusta.

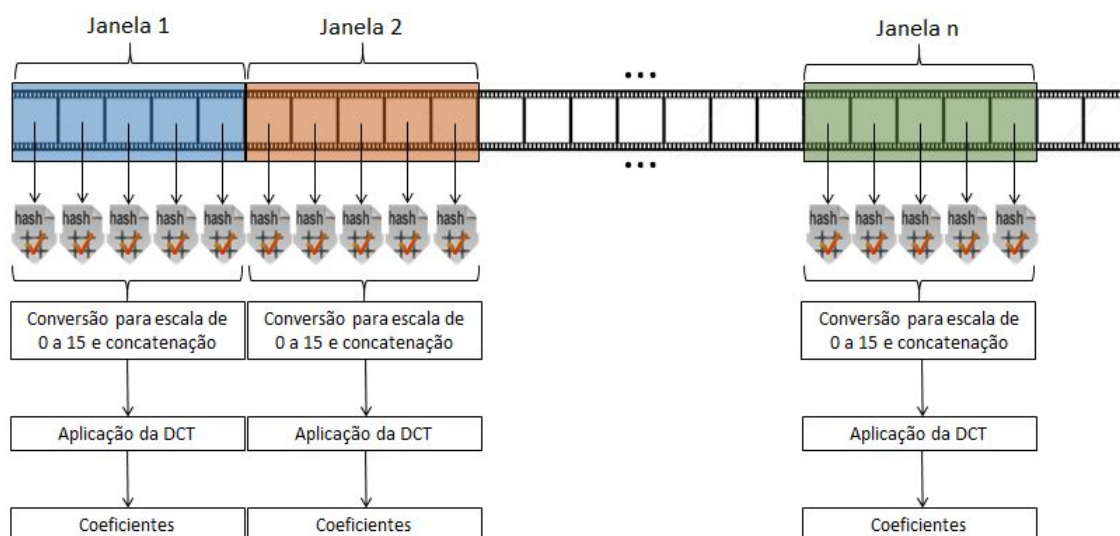


Figura 4.2: Fase de geração de *fingerprints* do Método dos Coeficientes da DCT

- **Fase de detecção**

Depois do armazenamento dos coeficientes a serem usados como *fingerprint* dos diferentes anúncios publicitários, na fase de detecção procede-se a uma abordagem similar à realizada na fase de geração de *fingerprints*. Inicialmente são geradas as *hashes* para cada *frame* da primeira janela de tamanho fixo e igual ao tamanho utilizado na primeira fase que após a conversão para valores na escala de 0 a 15 e respectiva concatenação é aplicada a Transformada Discreta do Cosseno (DCT) sendo assim gerados os coeficientes da respectiva janela. Posteriormente ocorre a fase de *matching* onde cada coeficiente é comparado com o respectivo em cada *fingerprint* armazenada na base de dados. Este processo de comparação está sujeito a um valor de *threshold* previamente definido de modo a contornar pequenas distorções que possam ter ocorrido na fase de segmentação de vídeo.

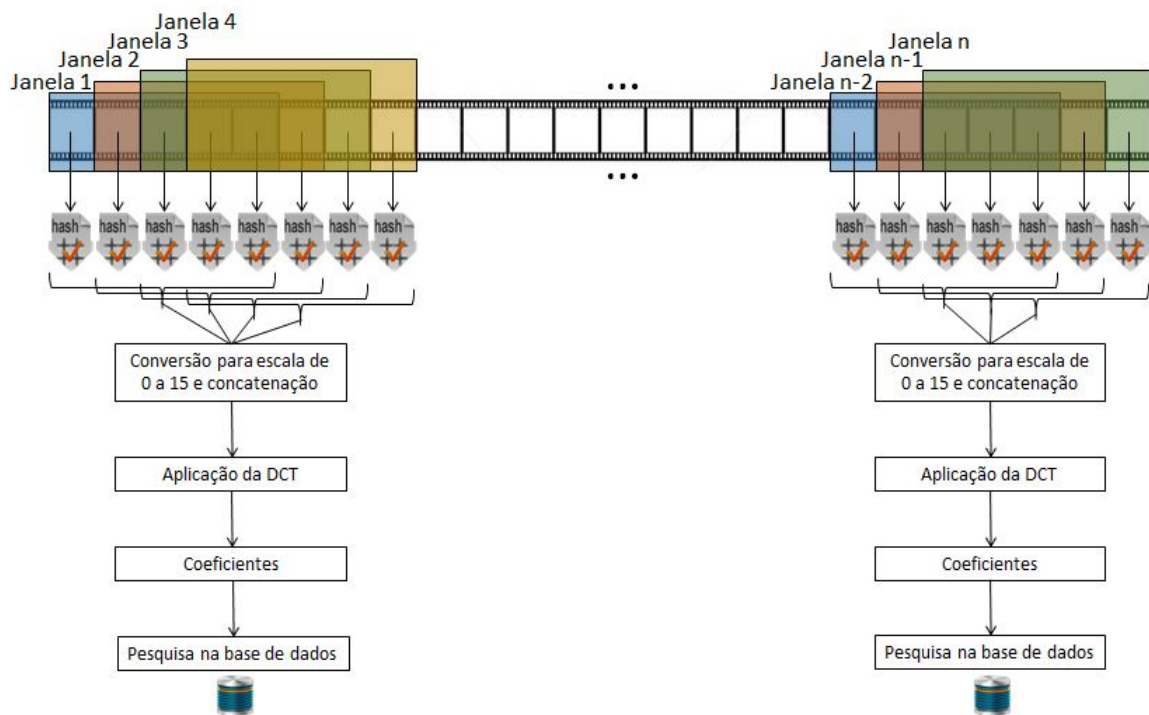


Figura 4.3: Fase de detecção do Método dos Coeficientes da DCT

#### 4.1.2 Método *Visual Rhythm*

- **Fase de geração de *fingerprints***

De um modo semelhante ao primeiro método, os segmentos publicitários individuais passam por um processo de resumo onde as suas *fingerprints* são geradas e armazenadas. Neste método recorreu-se ao *Visual Rhythm*, que aborda o uso de padrões e características visuais de modo a representar todo o conteúdo de um vídeo [16]. O *Visual Rhythm* é também

abordado em [42], que o destaca como sendo uma simplificação de vídeo utilizando fatias espaço-temporais de cada *frame*. Em termos mais simples, o *Visual Rhythm* utiliza sequências de pixels de cada *frame* de um vídeo, que são concatenadas de modo a originar uma representação visual dessa sequência de *frames*. Na figura 4.4 é exemplificada a representação em escala de cinza de um segmento publicitário a partir da concatenação da sequência de pixels correspondente à coluna central de cada *frame* do segmento.



Figura 4.4: Exemplo da utilização do *Visual Rhythm*

Inicialmente é definido um tamanho fixo de uma janela de *frames* no qual o vídeo de entrada é dividido. Dentro de cada janela, em cada *frame*, após a sua conversão para escala de cinza, são recolhidas três sequências de pixels que podem corresponder a colunas, a linhas e/ou a uma das duas diagonais do *frame* em questão. Ao longo de uma janela, as sequências de pixels recolhidas de cada *frame* são concatenadas formando três novas imagens (uma para cada posição das sequências de pixels recolhidas). Após o preenchimento de uma janela de *frames*, estas três imagens, representando uma abstração de uma sequência de *frames*, são inseridas nas três camadas RGB de uma imagem final (tal como exemplificado na figura 4.5) que é sujeita ao método de *hashing dHash*, gerando assim uma *hash* de caracteres hexadecimais para cada janela de *frames*, correspondente à sua *fingerprint* e armazenada numa base de dados.

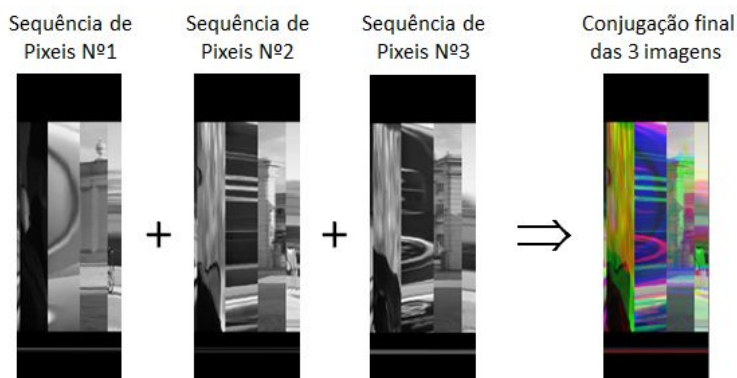


Figura 4.5: Conjugação da concatenação de três sequências de pixels



Esta fase de geração de *fingerprints* é exemplificada através da figura 4.6 onde, a título de exemplo, são recolhidas três colunas de cada *frame*. Nos casos em que o tamanho entre as sequências de pixels sejam diferentes (por exemplo, o tamanho das sequências diagonais é diferente do tamanho das sequências horizontais), as sequências de tamanho inferior são complementadas com valores de pixels iguais a zero, num processo designado por *zero padding*.

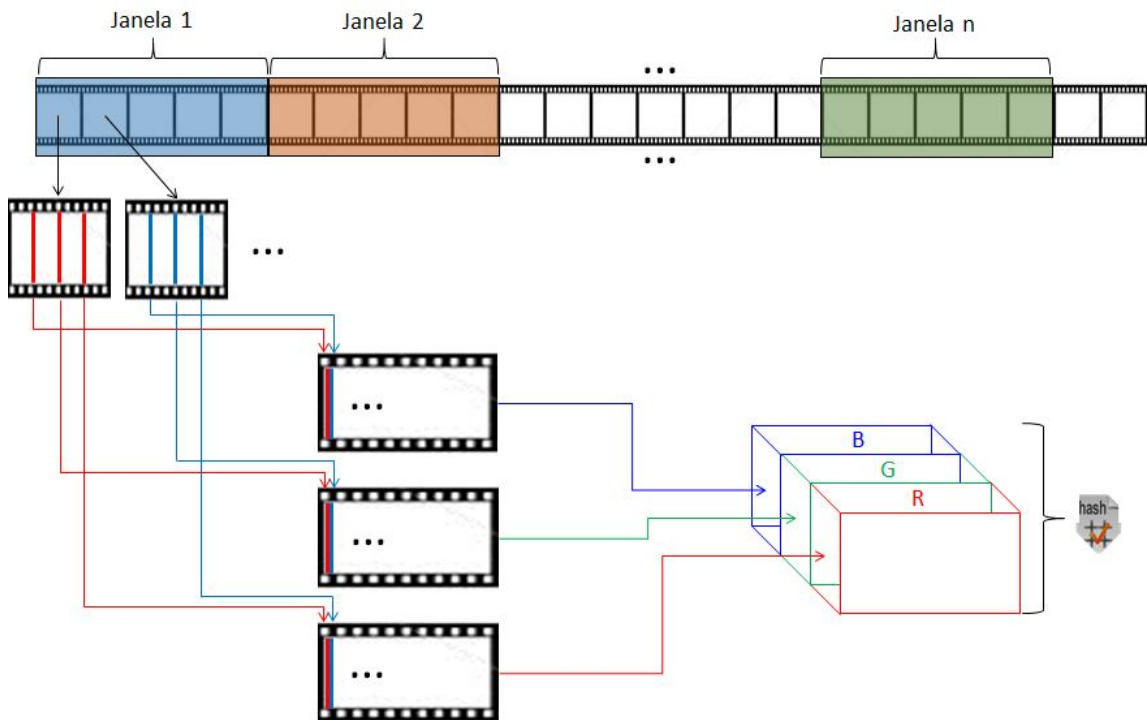


Figura 4.6: Fase de geração de *fingerprints* do Método *Visual Rhythm*

#### • Fase de detecção

De um modo similar ao primeiro método apresentado, após o armazenamento das *fingerprints* numa base de dados procede-se à fase de detecção. Inicialmente são recolhidas as sequências de pixels nas mesmas posições da fase de geração, para cada *frame* da primeira janela, do mesmo modo, estas sequências de pixels são concatenadas, formando três imagens que são inseridas nas três camadas RGB de uma imagem final para a qual é gerada uma *fingerprint* através do mesmo método *dHash*. Após a recolha da primeira janela de *frames*, são pesquisadas na base de dados *fingerprints* próximas da gerada, para isso utilizando uma métrica de similaridade, a distância de Hamming, que compara dois códigos do mesmo tamanho e calcula a sua distância contando o número de símbolos que diferem em cada posição [43]. Na implementação desta métrica, as *fingerprints* compostas por dígitos hexadecimais são convertidas para o sistema numérico binário a partir do qual se calcula o número de bits diferentes, definindo assim a sua distância. Depois da conclusão da pesquisa

na base de dados a janela temporal avança um *frame*, repetindo-se o mesmo processo até ao final da transmissão televisiva. Esta fase é exemplificada através da figura 4.7 onde, a título de exemplo, são recolhidas três colunas de cada *frame*.

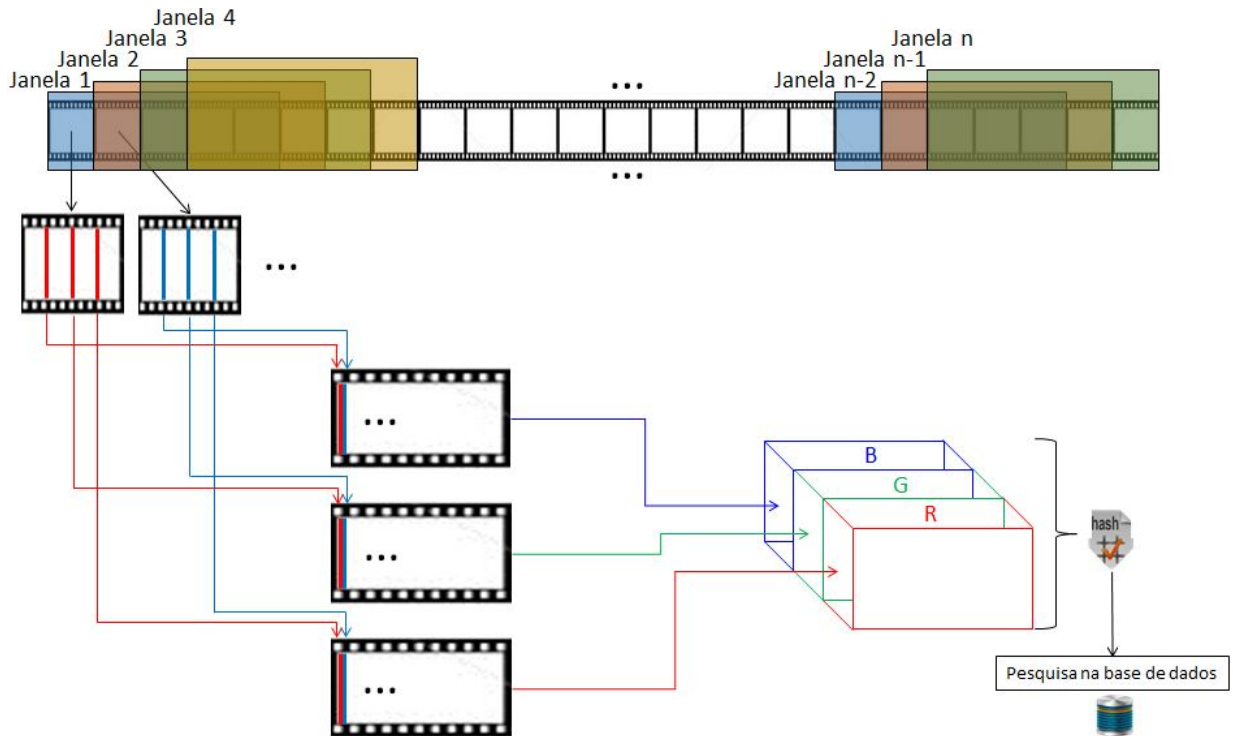


Figura 4.7: Fase de detecção do Método *Visual Rhythm*

## Capítulo 5

# Teste e Avaliação

Neste capítulo é descrita a metodologia utilizada para a avaliação dos testes realizados assim como os resultados obtidos para cada um dos métodos desenvolvidos e a sua respetiva análise e posterior discussão.

### 5.1 Método de Avaliação

Para efetuar a avaliação do desempenho dos métodos desenvolvidos, utilizando o *dataset* de *streams* disponível, havia a necessidade de utilizar métricas que reproduzissem informações sobre a eficiência do processo de deteção de anúncios publicitários, para tal, foram empregues as métricas de precisão (*precision*), revocação (*recall*) e pontuação *FI* (*FI score*). A métrica de *precision* reflete a fração de deteções corretas de entre o total de deteções realizadas, a métrica de *recall* reflete a fração de deteções corretas efetuadas de entre o conjunto de segmentos passíveis de serem corretamente detetados enquanto que a métrica *FI* reflete a média harmónica das métricas de *precision* e *recall*.

Para o cálculo das métricas propostas são utilizadas quatro formas de classificação das decisões efetuadas pelos algoritmos de deteção:

- **Positivo Verdadeiro:** ocorre quando é detetada uma publicidade no momento em que está a ser transmitido um anúncio publicitário;
- **Negativo Verdadeiro:** ocorre quando é detetado um programa no momento em que está a ser transmitido um segmento programático;
- **Positivo Falso:** ocorre quando é detetada uma publicidade no momento em que está a ser transmitido um segmento programático;
- **Negativo Falso:** ocorre quando é detetado um programa no momento em que está a ser transmitido um anúncio publicitário.

As métricas de avaliação são então calculadas através das seguintes fórmulas:

$$Precision = \frac{PositivoVerdadeiro}{PositivoVerdadeiro + PositivoFalso} \quad (5.1)$$

$$Recall = \frac{PositivoVerdadeiro}{PositivoVerdadeiro + NegativoFalso} \quad (5.2)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5.3)$$

De modo a aumentar o alcance dos resultados e extrair mais informação sobre o desempenho dos métodos de detecção implementados, foram definidos três tipos de erros que permitem analisar o comportamento de cada algoritmo em três fases distintas de um segmento, no início, meio e fim. Estes três tipos de erros definem-se da seguinte maneira:

- **Tipo I:** ocorrência de Positivo Falso ou Negativo Falso no início de um segmento, permitindo verificar o desempenho de um algoritmo de detecção após uma transição de segmentos;
- **Tipo II:** ocorrência de Positivo Falso ou Negativo Falso no fim de um segmento, permitindo verificar o desempenho de um algoritmo de detecção antes de uma transição de segmentos;
- **Tipo III:** ocorrência de Positivo Falso ou Negativo Falso no meio de um segmento, permitindo verificar o desempenho de um algoritmo de detecção no núcleo de um segmento.

Esta notação é utilizada de modo a verificar o desempenho dos métodos propostos em diferentes partes de cada segmento e assim identificar onde estes funcionam melhor e pior com o objetivo de identificar as causas do pior funcionamento para que cada método seja afinado de modo a obter melhores resultados.

Os erros do Tipo III têm uma importância de maior relevo, tendo em conta os objetivos propostos, pois a sua ocorrência significa uma falha de detecção a meio de um segmento, o que coloca em risco a sua integridade.

## 5.2 Resultados Experimentais

Tendo em conta o método de avaliação proposto, os algoritmos desenvolvidos foram testados de maneira a entender qual o seu melhor funcionamento através da afinação de alguns parâmetros variáveis podendo assim gerar os seus resultados finais. Para uma melhor compreensão, os resultados referentes a conteúdos publicitários e programáticos são apresentados de modo separado. Esta divisão deve-se à elevada diferença de tamanho entre os segmentos publicitários e programáticos, o que faz com que o impacto de uma falha de detecção durante um segmento publicitário, que têm durações consideravelmente mais curtas que os programas, tenha um impacto muito maior. Ao separar os resultados dos dois tipos de conteúdos, o desempenho de cada algoritmo desenvolvido é apresentado de forma mais clara e correta.

### 5.2.1 Testes com o Método dos Coeficientes da DCT

Devido à longa duração das *streams* do *dataset* utilizado e de modo a gerir o tempo dedicado a testes, numa fase inicial procedeu-se à realização de testes preliminares gerando uma base de dados com um pequeno conjunto de segmentos publicitários presentes num excerto de uma das *streams* do *dataset*. Esta fase de testes preliminares serviu para explorar os diferentes parâmetros variáveis do algoritmo tendo em vista encontrar os valores mais apropriados para a obtenção dos melhores resultados possíveis. Neste caso específico foram definidos os parâmetros apresentados na tabela 5.1 a serem executados utilizando o *dataset* completo.

Tamanho da janela de <i>frames</i>	Número de coeficientes armazenados	Valor de <i>threshold</i>
100	5	10

Tabela 5.1: Parâmetros utilizados no Método dos Coeficientes da DCT

Com esta afinação de parâmetros foi possível observar os efeitos da variação de cada um deles, o que permitiu concluir que um maior tamanho da janela de *frames* utilizada aumenta o peso computacional da geração de cada *fingerprint* ao mesmo tempo que permite utilizar uma base de dados de menor tamanho, um maior número de coeficientes utilizado permite melhorar a percentagem de deteções corretas, mas aumenta o tamanho da base de dados, diminuindo a velocidade de pesquisa. Relativamente ao valor de *threshold* utilizado, quanto maior este for, maior a probabilidade de realizar falsas deteções.

Com os parâmetros definidos, a tabela 5.2 reflete os resultados obtidos testando o algoritmo com o *dataset* completo.

Métrica	Publicidade			Programa		
	<i>Precision</i>	<i>Recall</i>	<i>F1</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>
Tipo I	46,51%	26,24%	31,03%	56,39%	72,09%	59,75%
Tipo II	58,18%	34,76%	41,58%	48,92%	62,16%	53,47%
Tipo III	59,92%	55,13%	56,90%	60,89%	79,42%	66,59%
Base	60,30%	54,40%	56,60%	60,80%	78,90%	66,00%

Tabela 5.2: Resultados do Método dos Coeficientes da DCT

### 5.2.2 Testes com o Método *Visual Rhythm*

De um modo similar ao método anterior, inicialmente foram realizados alguns testes preliminares de modo a gerir recursos. Para tal foi também gerada uma base de dados com um pequeno conjunto de segmentos publicitários que foi testada perante um excerto de uma *stream* presente no *dataset*. Esta fase de testes preliminares serviu também para explorar os diferentes parâmetros variáveis deste algoritmo tendo em vista encontrar os valores mais apropriados para a obtenção dos melhores resultados possíveis. No caso específico deste algoritmo, foram definidos os parâmetros apresentados na tabela 5.3 a serem executados utilizando o *dataset* completo.

Tamanho da janela de <i>frames</i>	Sequências de pixels	Distância de Hamming
100		4

Tabela 5.3: Parâmetros utilizados no Método *Visual Rhythm*

Com esta afinação de parâmetros, de um modo semelhante, foi possível observar os efeitos da variação de cada um deles, o que permitiu concluir que um maior tamanho da janela de *frames* utilizada aumenta o peso computacional da geração de cada *fingerprint* ao mesmo tempo que permite utilizar uma base de dados de menor tamanho, a utilização das diagonais dos *frames* é mais vantajosa pois permite a recolha se uma sequência de *frames* maior, aumentando a probabilidade de efetuar deteções corretas. Relativamente ao valor de *threshold* utilizado na medição da distância de Hamming, quanto maior este for, maior a possibilidade de realizar falsas deteções.

Com os parâmetros definidos, a tabela 5.4 reflete os resultados obtidos testando o algoritmo com o *dataset* completo.

Métrica	Publicidade			Programa		
	Precision	Recall	F1	Precision	Recall	F1
Tipo I	75,61%	87,99%	77,81%	85,44%	61,30%	68,59%
Tipo II	67,90%	80,42%	71,81%	97,13%	70,06%	79,20%
Tipo III	79,42%	98,26%	85,00%	98,71%	93,35%	95,81%
Base	78,90%	97,70%	84,50%	98,60%	93,10%	95,60%

Tabela 5.4: Resultados do Método *Visual Rhythm*

### 5.3 Discussão de Resultados

Analisando os resultados obtidos nos dois métodos desenvolvidos, verifica-se facilmente que o método *Visual Rhythm* apresenta relativamente bons resultados que são consideravelmente melhores comparativamente com o método dos Coeficientes da DCT, em qualquer uma das métricas consideradas, tanto na sinalização de segmentos publicitários como programáticos.

De um modo geral, os resultados obtidos no primeiro método implementado revelam-se insuficientes para ser tida em conta a sua utilização num sistema real, apresentando percentagens de deteção baixas, nomeadamente na fase inicial dos segmentos publicitários. Num olhar especial aos erros do Tipo III, devido à sua importância perante os objetivos definidos, também neste ponto as taxas de *precision* e *recall* apresentam resultados pouco satisfatórios, o que coloca em risco a integridade dos conteúdos numa situação de substituição de segmentos num cenário de redistribuição. O principal problema associado a este método recai sobre a quantidade de coeficientes

utilizados como *fingerprint* de cada janela de *frames*, pois a DCT calcula um número de coeficientes igual ao número de dígitos das *hashes* concatenados na sua entrada, o que faz com que o uso de apenas uma pequena fração desses coeficientes não ofereça uma *fingerprint* suficientemente robusta. O armazenamento de uma maior quantidade de coeficientes poderia ser considerado, mas iria aumentar o peso computacional do processo de pesquisa de *fingerprints*.

Relativamente ao segundo método implementado, este revela resultados promissores, destacando-se a taxa de *recall* base na detecção de anúncios publicitários situada nos 97,70%, apesar da taxa de *precision* ligeiramente mais baixa de um modo geral. Analisando os erros do Tipo III, os resultados apresentados são razoáveis, sendo quase ideais ao nível da taxa de *recall*, durante a detecção de publicidades. De notar os resultados obtidos neste método na detecção de conteúdos programáticos, que foi consideravelmente positiva, exibindo excelentes resultados nas três métricas consideradas, nomeadamente nas zonas intermédias dos segmentos, o que leva a considerar a possibilidade de utilizar este método num sistema real.

Em termos de tempo de processamento despendido, o método de *Visual Rhythm* mostrou ser suficientemente rápido para ser considerado o seu uso numa situação de *broadcasting* em tempo real ao apresentar um tempo de processamento total, no processo de geração e pesquisa de *fingerprints* durante a fase de detecção perante uma *stream* televisiva, inferior a 40ms, valor de referência correspondente ao rácio de *frames* exibidos por unidade de tempo. Por outro lado, o método dos coeficientes da DCT, devido á sua diferente estrutura de armazenamento e pesquisa de dados, revelou-se consideravelmente mais lento.





## Capítulo 6

# Conclusões

### 6.1 Conclusões Gerais

Nesta dissertação, dois métodos de detecção de conteúdos publicitários em transmissões televisivas foram propostos e implementados. Os métodos propostos basearam-se num conjunto de técnicas de *fingerprinting* tendo em vista a sua utilização num ambiente de *broadcasting*, tendo em conta as restrições temporais e computacionais associadas a este contexto e basearam-se no pressuposto de que os segmentos a detetar são previamente conhecidos.

Após uma revisão bibliográfica das técnicas de interesse anteriormente propostas verificou-se o estudo de diferentes características para proceder à detecção de segmentos publicitários. Características do domínio visual e do domínio de áudio foram exploradas, assim como a combinação de ambos os domínios em abordagens multimodais. Diversas propostas baseiam-se no reconhecimento de conteúdos, explorando técnicas de resumo de vídeo.

O ambiente de desenvolvimento foi estudado para melhor entender a constituição e caracterização do *dataset* a usar na fase de testes, que foi alvo de um processo de anotações manuais de modo a extrair informações relevantes para as fases de desenvolvimento e avaliação dos algoritmos. Também foram alvo de estudo algumas técnicas de *hashing* para geração de *fingerprints*, assim como estratégias de armazenamento de dados de modo a identificar as que melhor se adaptavam aos objetivos propostos.

Uma descrição teórica dos métodos desenvolvidos foi realizada, de forma a detalhar o seu funcionamento, abordando as diferentes técnicas utilizadas e salientando os parâmetros que podem ser ajustados para a obtenção de melhores resultados. Nesta descrição, os métodos apresentados dividem-se em duas fases, uma primeira onde são geradas as *fingerprints* dos anúncios publicitários a serem detetados e uma segunda fase onde se realiza a detecção propriamente dita. No primeiro método, foi apresentada uma proposta baseada na comparação de um conjunto de coeficientes gerados a partir da aplicação da Transformada Discreta do Cosseno (DCT) que na sua entrada recebe um conjunto de *hashes* geradas a partir de uma janela de *frames* que são previamente processadas e concatenadas. O segundo método apresentado propõe a comparação de *hashes* geradas a partir

da concatenação de sequências de pixels de cada *frame* de uma determinada janela temporal, num processo baseado em *Visual Rhythm*.

Numa fase de testes e avaliação, foram evidenciadas as métricas de avaliação utilizadas e apresentados os resultados finais, obtidos após uma fase de testes preliminares. A avaliação é feita de modo a analisar a precisão dos algoritmos em diferentes etapas, de modo a perceber onde estes podem ser melhorados. O segundo método proposto apresentou bons resultados, sendo consideravelmente melhores comparativamente com os obtidos no primeiro método, assim como um tempo de processamento total suficientemente baixo para se considerar o seu uso numa situação de *broadcasting* em tempo real.

## 6.2 Trabalho Futuro

Apesar dos bons resultados obtidos, essencialmente no segundo método desenvolvido, e tendo em conta as restrições temporais existentes, há ainda espaço para a execução de algumas melhorias, nomeadamente ao nível da robustez da metodologia implementada e do modo como a fase de avaliação é realizada. Alguns dos aspetos a serem melhorados residem na realização de uma maior quantidade de testes preliminares para encontrar, com maior precisão, os parâmetros que melhor se adaptam aos métodos desenvolvidos. A utilização de um *dataset* maior e mais variado (contendo *streams* de canais de diferentes géneros e de um maior número de países) permitiria testar a robustez dos métodos propostos mais intensivamente e perante diferentes cenários, tendo em vista a sua utilização a um nível global. A implementação de uma abordagem multimodal, através do uso de características de áudio combinadas com características visuais para gerar *fingerprints* dos diferentes segmentos de vídeo, poderá permitir aumentar a robustez da metodologia desenvolvida, nomeadamente ao nível da precisão de deteção, tendo no entanto como contrapartida, o aumento do tamanho da base de dados de *fingerprints* e do peso computacional associado ao processo de *matching*.

# Referências

- [1] David Conejero e Xavier Anguera. Tv advertisements detection and clustering based on acoustic information. Em *Computational Intelligence for Modelling Control & Automation, 2008 International Conference on*, páginas 452–457. IEEE, 2008.
- [2] Najeed Ahmed Khan, Umair Amin, e Muhammad Umer. Unsupervised commercials identification in videos. *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS*, 8(2):127–133, 2017.
- [3] Brandon Satterwhite e Oge Marques. Automatic detection of tv commercials. *IEEE Potentials*, 23(2):9–12, 2004.
- [4] Alberto Albiol, MJ Ch, FA Albiol, e Luis Torres. Detection of tv commercials. Em *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 3, páginas iii–541. IEEE, 2004.
- [5] Pujadevi Mourya1 Sakshi Jadhav, Reshma Pawar, e YB Gurav. Automatic channel switch system to escape commercial breaks. 2018.
- [6] Michele Covell, Shumeet Baluja, e Michael Fink. Advertisement detection and replacement using acoustic and visual repetition. Em *Multimedia Signal Processing, 2006 IEEE 8th Workshop on*, páginas 461–466. IEEE, 2006.
- [7] Ronald Glasberg, Cengiz Tas, e Thomas Sikora. Recognizing commercials in real-time using three visual descriptors and a decision-tree. Em *Multimedia and Expo, 2006 IEEE International Conference on*, páginas 1481–1484. IEEE, 2006.
- [8] Rainer Lienhart, Christoph Kuhmunch, e Wolfgang Effelsberg. On the detection and recognition of television commercials. Em *Multimedia Computing and Systems' 97. Proceedings., IEEE International Conference on*, páginas 509–516. IEEE, 1997.
- [9] Nedret Ozay e Bulent Sankur. Automatic tv logo detection and classification in broadcast videos. Em *Signal Processing Conference, 2009 17th European*, páginas 839–843. IEEE, 2009.
- [10] Nan Liu, Yao Zhao, Zhenfeng Zhu, e Hanqing Lu. Exploiting visual-audio-textual characteristics for automatic tv commercial block detection and segmentation. *IEEE Transactions on Multimedia*, 13(5):961–973, 2011.
- [11] Yijun Li, Danqing Zhang, Xiangmin Zhou, e Jesse S Jin. A confidence based recognition system for tv commercial extraction. Em *Proceedings of the nineteenth conference on Australasian database-Volume 75*, páginas 57–64. Australian Computer Society, Inc., 2008.

- [12] MOG CLOUD SETUP – Plataforma de Preparação de Conteúdos Audiovisuais para Ingest na Cloud. Projeto de I&DT Empresas em Copromoção nº 17561, 2018.
- [13] David A Sadlier, Sean Marlow, Noel O’Connor, e Noel Murphy. Automatic tv advertisement detection from mpeg bitstream. *Pattern Recognition*, 35(12):2719–2726, 2002.
- [14] Alexandre Gomes, Maria Paula Queluz, e Fernando Pereira. Automatic detection of tv commercial blocks: A new approach based on digital on-screen graphics classification. Em *Signal Processing and Communication Systems (ICSPCS), 2017 11th International Conference on*, páginas 1–6. IEEE, 2017.
- [15] Jen-Hao Yeh, Jun-Cheng Chen, Jin-Hau Kuo, e Ja-Ling Wu. Tv commercial detection in news program videos. Em *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, páginas 4594–4597. IEEE, 2005.
- [16] Hyeokman Kim, Jinho Lee, Jae-Heon Yang, Sanghoon Sull, Woonkyung M Kim, e S Moon-Ho Song. Visual rhythm and shot verification. *Multimedia Tools and Applications*, 15(3):227–245, 2001.
- [17] Nevenka Dimitrova. Multimedia content analysis: the next wave. *Image and video retrieval*, páginas 415–420, 2003.
- [18] Nevenka Dimitrova, Lalitha Agnihotri, e Gang Wei. Video classification based on hmm using text and faces. Em *Signal Processing Conference, 2000 10th European*, páginas 1–4. IEEE, 2000.
- [19] Weiming Hu, Nianhua Xie, Li Li, Xianglin Zeng, e Stephen 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, 2011.
- [20] Ullas Gargi, Rangachar Kasturi, e Susan H Strayer. Performance characterization of video-shot-change detection methods. *IEEE transactions on circuits and systems for video technology*, 10(1):1–13, 2000.
- [21] Matthew Cooper e Jonathan Foote. Scene boundary detection via video self-similarity analysis. Em *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 3, páginas 378–381. IEEE, 2001.
- [22] Matthew Roach, John Mason, e Li-Qun Xu. Video genre verification using both acoustic and visual modes. Em *Multimedia Signal Processing, 2002 IEEE Workshop on*, páginas 157–160. IEEE, 2002.
- [23] Pinar Duygulu, Ming-yu Chen, e Alexander Hauptmann. Comparison and combination of two novel commercial detection methods. Em *Multimedia and Expo, 2004. ICME’04. 2004 IEEE International Conference on*, volume 2, páginas 1267–1270. IEEE, 2004.
- [24] Xian-Sheng Hua, Lie Lu, e Hong-Jiang Zhang. Robust learning-based tv commercial detection. Em *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, páginas 4–pp. IEEE, 2005.
- [25] Patrick Cardinal, Vishwa Gupta, e Gilles Boulianne. Content-based advertisement detection. Em *Eleventh Annual Conference of the International Speech Communication Association*, 2010.

- [26] Wei-Lun Chao. Introduction to video fingerprinting. *Tutorial Report, DISP lab, NTU*, 4, 2009.
- [27] Yan Ke, Derek Hoiem, e Rahul Sukthankar. Computer vision for music identification. Em *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, páginas 597–604. IEEE, 2005.
- [28] Houssemeddine Khemiri, Gérard Chollet, e Dijana Petrovska-Delacrétaz. Automatic detection of known advertisements in radio broadcast with data-driven alisp transcriptions. *Multimedia Tools and Applications*, 62(1):35–49, 2013.
- [29] John M Gauch e Abhishek Shivadas. Identification of new commercials using repeated video sequence detection. Em *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 3, páginas II–1252. IEEE, 2005.
- [30] Jian Lu. Video fingerprinting for copy identification: from research to industry applications. Em *Media Forensics and Security*, volume 7254, 2009.
- [31] Sunil Lee e Chang D Yoo. Video fingerprinting based on centroids of gradient orientations. Em *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 2, páginas II–II. IEEE, 2006.
- [32] Sunil Lee e Chang D Yoo. Robust video fingerprinting for content-based video identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(7):983–988, 2008.
- [33] Xing Su, Tiejun Huang, e Wen Gao. Robust video fingerprinting based on visual attention regions. Em *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, páginas 1525–1528. IEEE, 2009.
- [34] Piotr Indyk e Rajeev Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. Em *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, páginas 604–613. ACM, 1998.
- [35] Aristides Gionis, Piotr Indyk, Rajeev Motwani, et al. Similarity search in high dimensions via hashing. Em *Vldb*, volume 99, páginas 518–529, 1999.
- [36] Fuhao Zou, Yunpeng Chen, Jingkuan Song, Ke Zhou, Yang Yang, e Nicu Sebe. Compact image fingerprint via multiple kernel hashing. *IEEE Transactions on Multimedia*, 17(7):1006–1018, 2015.
- [37] Mengjuan Fei, Jing Li, e Honghai Liu. Visual tracking based on improved foreground detection and perceptual hashing. *Neurocomputing*, 152:413–428, 2015.
- [38] SQLite. Architecture of SQLite. Disponível em <https://www.sqlite.org/arch.html>, Acedido pela última vez em 21 de junho de 2018.
- [39] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, e Clifford Stein. *Introduction to algorithms*. MIT press, 2009.
- [40] Chunyue Bi. Research and application of sqlite embedded database technology. *WSEAS Transactions on Computers*, 1(8):83–92, 2009.
- [41] Lv Junyan, Xu Shiguo, e Li Yijie. Application research of embedded database sqlite. Em *Information Technology and Applications, 2009. IFITA'09. International Forum on*, volume 2, páginas 539–543. IEEE, 2009.

- [42] Silvio Jamil Ferzoli Guimar, Michel Couprie, Neucimar Jeronimo Leite, et al. A method for cut detection based on visual rhythm. Em *Computer Graphics and Image Processing, 2001 Proceedings of XIV Brazilian Symposium on*, páginas 297–304. IEEE, 2001.
- [43] Hong Yang e Yiding Wang. A lbp-based face recognition method with hamming distance constraint. Em *Image and Graphics, 2007. ICIG 2007. Fourth International Conference on*, páginas 645–649. IEEE, 2007.