



INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA (ISEL)

DEPARTAMENTO DE ENGENHARIA ELETRÓNICA E DE
TELECOMUNICAÇÕES E COMPUTADORES (DEETC)

LEIM

LICENCIATURA EM ENGENHARIA INFORMÁTICA E MULTIMÉDIA
UNIDADE CURRICULAR DE PROJETO

Anotação de Eventos Sonoros em Vídeo

(eventual) imagem ilustrativa do trabalho – *dimensão*: até 13cm x 4cm

Carina Fernandes (45118)

Orientador(es)

Professor [Doutor] Joel Paulo

Professor [Doutor] Paulo Trigo

Professor [Doutor] Paulo Vieira

Setembro, 2022

Resumo

O desempenho de um atleta, numa determinada modalidade desportiva, melhora quando este é acompanhado de uma perspetiva externa no decurso da sua atividade desportiva. Neste sentido, o atleta pode ser monitorizado por um treinador para atingir melhores resultados. Como complemento, é possível registar em vídeo e analisar posteriormente o desempenho do atleta.

Neste sentido, é vantajoso o desenvolvimento de uma ferramenta capaz de realizar essa análise externa. A ferramenta desenvolvida neste projeto, permite extrair e reconhecer eventos relevantes (i.e. períodos em que ocorra uma maior troca de bolas) com base no vídeo da atividade desportiva do atleta (em treinos ou competições de padel /ténis). O processamento é realizado com base no áudio extraído do vídeo, e tem como base a extração de características ou padrões identificativos dos eventos, com o auxílio de técnicas de aprendizagem automática, baseadas em séries temporais. No final, são realizadas estatísticas, que permitem obter um resumo detalhado do que foi registado no vídeo, dando uma perspetiva mais abrangente e objetiva do desempenho do atleta. A ferramenta identifica corretamente cerca de 85% dos eventos em análise, sendo necessários alguns ajustes para melhorar o processo de reconhecimento.

Os ensaios são realizados no Laboratório de Áudio e Acústica do ISEL, LAA.

Palavras-chave: algoritmos de software de deteção do som, algoritmos de software de inteligência artificial.

Abstract

In a given sport, the athlete's performance improves when he is supervised from an external perspective. In this scenario, the athlete can be supervised by a coach in order to achieve better results, or as a complement, it is possible to video record and analyze his performance afterwards.

Thus, it is advantageous to develop a tool capable of performing this external analysis. The tool developed in this project, allows extracting and recognizing relevant events (i.e., periods when a greater exchange of balls occurs) based on the video of the athlete's sport activity (in the padel/tennis sport activity). The processing is done based on the audio extracted from the video, and is based on the extraction of patterns identifying the events, with the help of machine learning techniques, based on time series. At the end, statistics are performed, allowing a detailed summary of what was recorded in the video, giving a more comprehensive and objective perspective of the athlete's performance. The tool correctly identifies about 85% of the events under analysis, with some adjustments needed to improve the recognition process.

The tests are carried out at ISEL's Audio and Acoustics Laboratory, LAA.

Keywords: sound detection software algorithms, artificial intelligence software algorithms.

Agradecimentos

Escrever aqui eventuais agradecimentos ...

Eventual texto de dedicatória . . .

. . . mais texto,

. . . e o fim do texto.

Índice

Resumo	i
Abstract	iii
Agradecimentos	v
Índice	ix
Lista de Tabelas	xi
Lista de Figuras	xiii
1 Introdução	1
1.1 Motivação	2
1.2 Processo de Desenvolvimento	2
1.3 Validação e Testes	3
1.4 Apreciação Global dos Objetivos Alcançados	3
1.5 Estrutura do Relatório	4
2 Trabalho Relacionado	5
3 Modelo Proposto	7
3.1 Requisitos	8
3.1.1 Requisitos Funcionais	8
3.1.2 Requisitos Não Funcionais	8
3.1.3 Casos de Utilização	8
3.2 Fundamentos	8
3.3 Abordagem	9

4	Implementação do Modelo	11
5	Validação e Testes	13
6	Conclusões e Trabalho Futuro	15
6.0.1	Conclusões	15
6.0.2	Trabalho Futuro	15
A	Um Detalhe Adicional	17
B	Outro Detalhe Adicional	19
	Bibliografia	21

Lista de Tabelas

5.1	Uma tabela	13
-----	----------------------	----

Lista de Figuras

3.1	Processo de obtenção do dataset.	7
5.1	Uma figura	13

Capítulo 1

Introdução

Para que a atividade desportiva de um atleta, numa determinada modalidade, melhore é necessário que este seja monitorizado de uma perspetiva externa. Para esse efeito, é necessário que o atleta seja supervisionado durante a sua atividade física, ou que a última seja registada e posteriormente analisada (registo em papel ou em vídeo).

A anotação de eventos sonoros em vídeo consiste no processo de análise do áudio extraído de um vídeo previamente gravado, no sentido de recolher informações relativas a possíveis eventos relevantes.

A deteção e anotação de eventos sonoros potencialmente relevantes, constitui uma tarefa complexa, já que envolve o processamento de um grande volume de dados (referenciar na bibliografia). Um grande volume de dados pressupõe um processamento em modo *offline*, devido ao custo computacional associado. Embora esse processamento possa ter aplicações nas mais variadas componentes, este trabalho destina-se apenas à análise da componente desportiva de padel, no sentido de fornecer a atletas e/ou treinadores um complemento no processo de treino.

O padel é um desporto praticado com o auxílio de raquetes, jogado a pares, que se realiza num campo semelhante ao do ténis(referenciar na bibliografia). O campo é dividido por uma rede e encontra-se rodeado por paredes de vidro e redes metálicas. As suas dimensões são de 20 metros de comprimento por 10 de largura(referenciar na bibliografia).

Neste sentido, pretende-se desenvolver uma ferramenta que auxilie na atividade desportiva do atleta ou na monitorização de quem o treina. A aplicação processa o vídeo introduzido, realiza um processamento interno e

devolve uma análise intensiva dos períodos onde existem eventos que possam interessar aos utilizadores, como por exemplo, períodos onde existem mais batidas de bola.

1.1 Motivação

A aplicação em desenvolvimento implica que se implemente um sistema onde a máquina é capaz de reconhecer batidas de bola. No entanto, a forma como uma máquina reconhece um som difere de como esse mesmo som é percebido pelo ouvido humano. O ouvido humano consegue fazer a distinção entre sons, considerando as características dos sons emitidos: duração, intensidade, timbre e tom (referenciar na bibliografia).

Pelo que, é necessário proporcionar à máquina a capacidade de reconhecer batidas de bola de forma semelhante à humana. Neste sentido, o que se pretende, é utilizar a máquina para simular o ouvido humano. Para esse efeito, far-se-á uso de aprendizagem automática.

A aprendizagem automática permitirá que o sistema em desenvolvimento receba como *input* dados previamente tratados, com o objetivo de treinar um modelo capaz de identificar batidas de bola.

A identificação de padrões nos dados constitui uma forma de a máquina aprender a reconhecer os eventos relevantes ou batidas de bola. A implementação de uma sistema com estas particularidades constitui uma grande vantagem, já que permitirá obter uma análise mais detalhada de treinos ou jogos de padel aos utilizadores da ferramenta em questão.

1.2 Processo de Desenvolvimento

O processo de reconhecimento batidas de bola, requer um trabalho adicional correspondente à etiquetagem prévia de eventos presentes no áudio. Ou seja, para que a máquina possa aprender o que é uma batida de bola, é necessário que se “diga” à mesma em que instantes de tempo ocorrem batidas de bola. Assim, esta adquire a capacidade de distinguir eventos correspondentes a batidas de bola, de eventos que não correspondem a batidas de bola. Os dados serão colocados numa estrutura de dados denominada *dataset*. O processo de construção do *dataset* envolve os seguintes passos:

- Extração do áudio a partir do vídeo – corresponde à conversão do vídeo no áudio correspondente;
- Anotação dos eventos relevantes presentes no áudio – refere-se à etiquetagem de batidas de bola no áudio obtido no ponto anterior;
- Extração das características do áudio – refere-se à obtenção de padrões no áudio que possam identificar potenciais batidas de bola;
- Junção dos dados resultantes dos dois pontos anteriores.

Após a obtenção do *dataset*, estes dados são submetidos a um algoritmo, neste caso, uma rede neuronal. A rede neuronal recebe o *dataset* e identifica no mesmo, padrões com o intuito de reconhecer os eventos associados a esses padrões (referenciar bibliografia). Os eventos em contexto, são batidas de bola (na raquete, no chão, nas paredes de vidro, entre outras).

A aplicação do algoritmo sobre os dados denomina-se modelo ou classificador. O treino do modelo terá como principal objetivo, reconhecer as batidas de bola. Depois do processo de treino, o classificador deve ser capaz de “dizer” que tipo de evento ocorre num dado instante de tempo, tendo em conta os padrões encontrados nos dados.

Tendo construído o modelo, o passo seguinte corresponde a introduzir novos dados (nunca antes processados pelo modelo), classificá-los e, posteriormente o sistema efetuará estatísticas acerca dos resultados retornados pelo modelo. As estatísticas referem-se por exemplo, a períodos de tempo onde o modelo identifica mais batidas de bola e poderão ter mais relevância para os utilizadores.

1.3 Validação e Testes

(o que deve se colocado aqui na realidade?)

1.4 Apreciação Global dos Objetivos Alcançados

Fazer uma apreciação global dos resultados alcançados (como está no Resumo e Abstract).

Falar um pouco sobre os resultados obtidos em termos de percentagens e a forma como o sistema se comporta face a novos dados de input.

1.5 Estrutura do Relatório

O relatório está organizado da seguinte forma:

- O Capítulo 1 introduz o conceito de anotação de eventos sonoros em vídeo na modalidade desportiva de padel. Na secção de Motivação, é abordada a intenção de utilizar a aprendizagem automática para reconhecer padrões identificativos de batidas de bola. O processo de desenvolvimento descreve, de forma geral, as etapas para desenvolvimento da aplicação, dando algum destaque à construção do *dataset* como trabalho adicional. De seguida, é realizada uma apreciação global dos resultados obtidos.
- O trabalho relacionado analisado durante o desenvolvimento do trabalho é descrito no Capítulo 2, que se centra no uso da inteligência artificial para reconhecer eventos em desporto.
- O Capítulo 3 divide-se em três componentes: requisitos; fundamentos e abordagem. A primeira componente refere-se aos requisitos funcionais e não funcionais (atributos) do sistema e aos casos de utilização. Na secção de Fundamentos 3.2 são abordados os conceitos relacionados com áudio, bem como a inteligência artificial. São também abordadas as tecnologias utilizadas. Por último, na secção Abordagem 3.3 são abordados os métodos através dos quais se obteve o *dataset*, bem como os algoritmos utilizados para treinar o modelo (conceitos de aprendizagem automática e redes neuronais).
- (Falar do capítulo 4)
- (Falar do capítulo 5)
- (Falar do capítulo 6)

Capítulo 2

Trabalho Relacionado

No fundo, pegar nos artigos científicos e relacionar/ comparar com o que se pretende no trabalho: tecnologias utilizadas, métodos de desenvolvimento, área a que se aplica (desporto).

Trabalho relacionado aqui . . .

Aqui terá certamente necessidade de citar (fazer referência) a vários trabalhos anteriormente publicados e que foi analisando ao longo de todo o seu projeto. Esses trabalhos devem ser apresentados com os seguintes objetivos essenciais:

- delimitar o contexto onde o seu projeto se insere,
- definir claramente os aspetos diferenciadores (inovadores) do seu projeto,
- identificar e caracterizar os pressupostos (teóricos ou tecnológicos) em que o projeto se baseia.

Cada trabalho a que fizer referência precisa de ser corretamente identificado. Essa identificação depende do tipo de publicação do trabalho. Um trabalho terá sido publicado em revista científica, e.g., [Elzinga e Mills, 2011], outro em ata de conferência internacional, e.g., [Boutilier et al., 1995], outro em livro, e.g., [Bellifemine et al., 2007], ou apenas em capítulo de livro, e.g., [Wooldridge, 2000], ou pode ainda incluído numa coleção, e.g., [Howard e Matheson, 1984] e há também a hipótese de ser uma “publicação de proveniência diversa”, como no caso em que o “o sítio na Internet” é a principal forma de publicação, e.g., [Python3.2.3, 2012] e, por fim, a publicação pode ser um relatório técnico, e.g., [Marin, 2006].

Para conseguir lidar de forma adequada com as referências é importante construir um acervo e ter um mecanismo para geração automática (e correta) das referências que vai fazendo ao longo do texto.

Atualmente, as publicações têm também informação sobre o modo como devem ser corretamente citadas; em geral essa informação segue o formato `BIBTEX`.

Para fazer referência a um trabalho é necessário seguir as boas regras (sintáticas) para uma citação correta, mas isso não é suficiente; falta a “semântica”. Ou seja, é também preciso descrever o essencial do trabalho que está a citar. É necessário explicar esse trabalho e enquadrá-lo, no texto, de modo a tornar clara a relação entre esse trabalho e o seu projeto.

Capítulo 3

Modelo Proposto

O modelo deve ser independente de linguagens de programação ou bibliotecas, mas devem ser fornecidos argumentos que sustentem a escolha ou adoção das tecnologias utilizadas. O modelo proposto é composto por: requisitos; fundamentos; abordagem; implementação.

A figura que se segue refere-se ao processo de obtenção do *dataset*. Para além de tabelas pode também precisar de apresentar figuras. Por exemplo, a figura 5.1 descreve ...

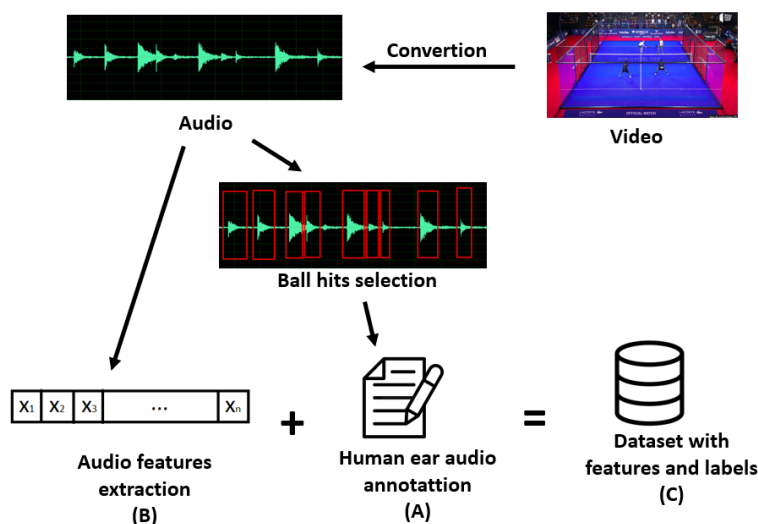


Figura 3.1: Processo de obtenção do dataset.

Aqui mostra um caminho que inicia com requisitos (cf., secção 3.1), passa pela aplicação dos fundamentos (cf., secção 3.2) e continua até conseguir

transmitir uma visão clara e um formalismo com nível de detalhe adequado a um leitor que tenha um perfil (competência técnica) idêntico ao seu.

Recorra, sempre que possível, a exemplos ilustrativos da utilização do seu modelo. Esses exemplos devem ajudar o leitor a compreender os aspectos mais específicos do seu trabalho.

O modelo aqui proposto deve ser (tanto quanto possível) independente de tecnologias concretas (e.g., linguagens de programação ou bibliotecas). No entanto deve fornecer os argumentos que contribuam para justificar uma posterior escolha (adoção) de tecnologias.

3.1 Requisitos

Aqui o essencial (e se aplicável) dos requisitos funcionais, não funcionais e modelo de casos de utilização. Aqui deve também apresentar matriz para decisão sobre prioridade dos casos de utilização (se aplicável) ...

Deve apresentar de forma “moderada” o resultado da fase avaliação de requisitos. A informação de maior detalhe (e.g., diagramas UML demasiado detalhados) deve ser colocada em apêndice.

3.1.1 Requisitos Funcionais

(Requisitos Funcionais aqui)

3.1.2 Requisitos Não Funcionais

(Requisitos Não Funcionais aqui)

3.1.3 Casos de Utilização

(Casos de utilização aqui)

3.2 Fundamentos

Aqui o sustento formal (teórico / tecnológico) do trabalho realizado ...

3.3 Abordagem

Aqui explique as formulações, os métodos, os algoritmos e outros contributos que desenvolveu e que considera centrais ao seu trabalho.

Aqui precisa de abordar tudo o que contribui para diferenciar o seu trabalho e que (na sua opinião) deve ser evidenciado e explicado de modo claro.

Lembre-se que a apresentação de um (ou mais) **exemplo(s) simples** é muito importante para que o leitor consiga seguir e compreender o seu trabalho.

Tenha em atenção que um exemplo acompanhado por figuras ilustrativas será certamente analisado (de modo cuidado) pelos leitores do seu trabalho.

Capítulo 4

Implementação do Modelo

Implementação do modelo aqui ...; pode precisar de referir o capítulo 3 ...

Aqui identifica as opções teóricas e justifica as dependências tecnológicas assumidas neste projeto. Descreva com rigor formal e detalhe adequado e faça evidência de tudo o que foi proposto e desenvolvido especificamente no contexto deste projeto. Aqui a ênfase está naquilo que foi de facto concretizado neste projeto.

O leitor quer detalhes de concretização. Ele já está enquadrado no tema (cf., capítulo 2), já conhece os aspetos mais abstratos da sua proposta (cf., capítulo 3) e agora precisa de entender os detalhes para conseguir também interpretar as validações e testes que posteriormente (cf., capítulo 5) lhe irá apresentar.

Capítulo 5

Validação e Testes

Validação e testes aqui ...; pode precisar de referir o capítulo 3 ou alguma das suas secções, e.g., a secção 3.2 ...

Pode precisar de apresentar tabelas. Por exemplo, a tabela 5.1 apresenta os dados obtidos na experiência ...

c_1	c_2	c_3	$\sum_{i=1} c_i$
1	2	3	6
1.1	2.2	3.3	6.6

Tabela 5.1: Uma tabela

Para além de tabelas pode também precisar de apresentar figuras. Por exemplo, a figura 5.1 descreve ...



Figura 5.1: Uma figura

Atenção. Todas as tabelas e figuras, e.g., diagramas, imagens ilustrativas da aplicação em funcionamento, têm que ser devidamente enquadradas no texto antes de serem apresentadas e esse enquadramento inclui uma explicação da imagem apresentada e eventuais conclusões (interpretações) a tirar dessa imagem.

Capítulo 6

Conclusões e Trabalho Futuro

(Em termos de trabalho futuro, poderia abordar a otimização do sistema desenvolvido para se poder adaptar ao tipo de ambiente em que o vídeo é gravado.

O até otimizar o sistema para conseguir através da duração do evento, detetar de que lado do campo ocorrem os eventos)

Conclusões e trabalho futuro aqui ...

Quais as principais mensagens a transmitir ao leitor deste trabalho? O leitor está certamente interessado nos temas aqui abordados. Em geral procurará, neste projeto, pistas para algum outro objetivo. Assim, é muito importante que o leitor perceba rapidamente a relação entre este trabalho e o seu próprio (do leitor) objetivo.

Aqui é o local próprio para condensar a experiência adquirida neste projeto e apresentá-la a outros (futuros leitores).

O pressuposto é o de que este projeto é um “elemento vivo” que recorreu a outros elementos (cf., capítulo 2) para ser construído e que poderá servir de suporte à construção de futuros projetos.

6.0.1 Conclusões

(Conclusões aqui)

6.0.2 Trabalho Futuro

(Trabalho Futuro aqui)

Apêndice A

Um Detalhe Adicional

O “apêndice” utiliza-se para descrever aspectos que tendo sido desenvolvidos pelo autor constituem um complemento ao que já foi apresentado no corpo principal do documento.

Neste documento utilize o apêndice para explicar o processo usado na **gestão das versões** que foram sendo construídas ao longo do desenvolvimento do trabalho.

É especialmente importante explicar o objetivo de cada ramo (“branch”) definido no projeto (ou apenas dos ramos mais importantes) e indicar quais os ramos que participaram numa junção (“merge”).

É também importante explicar qual a arquitetura usada para interligar os vários repositórios (e.g., Git, GitHub, DropBox, GoogleDrive) que contêm as várias versões (e respetivos ramos) do projeto.

Notar a diferença essencial entre “apêndice” e “anexo”. O “apêndice” é um texto (ou documento) que descreve trabalho desenvolvido pelo autor (e.g., do relatório, monografia, tese). O “anexo” é um texto (ou documento) sobre trabalho que não foi desenvolvido pelo autor.

Para simplificar vamos apenas considerar a noção de “apêndice”. No entanto, pode sempre adicionar os anexos que entender como adequados.

Apêndice B

Outro Detalhe Adicional

Escrever aqui o detalhe adicional que melhor explique outro aspecto (diferente do que está no apêndice A) descrito no corpo principal do documento

...

Bibliografia

- [Bellifemine et al., 2007] Bellifemine, F. L., Caire, G., e Greenwood, D. (2007). *Developing Multi-Agent Systems with JADE*. Wiley Series in Agent Technology. Wiley.
- [Boutilier et al., 1995] Boutilier, C., Dearden, R., e Goldszmidt, M. (1995). Exploiting structure in policy construction. In *Proceedings of the IJCAI-95*, p. 1104–1111.
- [Elzinga e Mills, 2011] Elzinga, K. e Mills, D. (2011). The lerner index of monopoly power: Origins and uses. *American Economic Review: Papers & Proceedings*, 101(3).
- [Howard e Matheson, 1984] Howard, R. e Matheson, J. (1984). Influence diagrams. In *Readings on the Principles and Applications of Decision Analysis*, volume 2, p. 721–762. Strategic Decision Group, Menlo Park, CA.
- [Marin, 2006] Marin, D. (2006). A formalization of RDF (applications de la logique á la sémantique du Web). Technical report, Dept. Computer Science, Ecole Polytechnique, Universidad de Chile, TR/DCC-2006-8.
- [Python3.2.3, 2012] Python3.2.3 (2012). Python programming language. <http://docs.python.org/py3k/>.
- [Wooldridge, 2000] Wooldridge, M. (2000). *Reasoning About Rational Agents*, cap.: Implementing Rational Agents. The MIT Press.