UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

VAGNER MENDONÇA GONÇALVES

**Recuperação de vídeos médicos baseada em conteúdo utilizando extratores
de características visuais e sonoras**

São Paulo

2017

VAGNER MENDONÇA GONÇALVES

**Recuperação de vídeos médicos baseada em conteúdo utilizando extratores
de características visuais e sonoras**

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação.

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 12 de dezembro de 2016. A versão original encontra-se em acervo reservado na Biblioteca da EACH-USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Área de Concentração:
Metodologia e Técnicas da Computação

Orientador:
Profa. Dra. Fátima de Lourdes dos Santos
Nunes Marques

São Paulo

2017

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

CATALOGAÇÃO-NA-PUBLICAÇÃO

(Universidade de São Paulo. Escola de Artes, Ciências e Humanidades. Biblioteca)

Gonçalves, Vagner Mendonça

Recuperação de vídeos médicos baseada em conteúdo utilizando
extratores de características visuais e sonoras / Vagner Mendonça
Gonçalves ; orientadora, Fátima de Lourdes dos Santos Nunes
Marques. – São Paulo, 2017

99 f. : il.

Dissertação (Mestrado em Ciências) - Programa de Pós-
Graduação em Sistemas de Informação, Escola de Artes, Ciências
e Humanidades, Universidade de São Paulo em 2016

Versão corrigida

1. Informática médica. 2. Recuperação da informação. 3.
Vídeo. 4. Diagnóstico por computador. I. Marques, Fátima de
Lourdes dos Santos Nunes, orient. II. Título

CDD 22.ed.— 610.0285

Dissertação de autoria de Vagner Mendonça Gonçalves, sob o título **“Recuperação de vídeos médicos baseada em conteúdo utilizando extratores de características visuais e sonoras”**, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação, na área de concentração Metodologia e Técnicas da Computação, aprovada em 12 de dezembro de 2016 pela comissão julgadora constituída pelos doutores:

Prof. Dr. Profa. Dra. Fátima de Lourdes dos Santos Nunes Marques

Universidade de São Paulo

Presidente

Prof. Dr. Aparecido Nilceu Marana

Universidade Estadual Paulista Júlio de Mesquita Filho

Prof. Dr. Helton Hideraldo Bísaro

Universidade de São Paulo

*À minha mãe, Cida, amiga, confidente e companheira, cuja memória será para mim a
mais bela das recordações.*

AGRADECIMENTOS

Agradeço, primeiramente, a Deus por me inspirar e dar forças para chegar até aqui, até mesmo quando eu acreditei que não conseguiria.

Agradeço à Nossa Senhora que, com seu exemplo de humildade e fidelidade, me inspirou a prosseguir na caminhada, mesmo diante das dificuldades.

Agradeço à minha mãe, Maria Aparecida Mendonça Gonçalves, exemplo de luta e perseverança, que não me viu concluir este trabalho, mas que teve, tem e sempre terá papel fundamental em cada passo que eu dou.

Agradeço à minha orientadora, Profa. Dra. Fátima de Lourdes dos Santos Nunes Marques, primeiramente por acreditar que eu conseguiria vencer esse desafio, mesmo quando o caminhar estava difícil e duvidoso. Agradeço também por todo o apoio e orientação, desde a graduação, que me auxiliaram a chegar até aqui.

Agradeço à minha namorada, Camila Ericka Andrade de Melo, por todo o amor, carinho e apoio incondicionais, bem como pelas palavras de conforto e motivação que tiveram papel fundamental na minha caminhada.

Agradeço ao meu pai, João Costa Gonçalves, e ao meu irmão, Carlos Junior Mendonça Gonçalves, que estão sempre ao meu lado, partilhando os bons momentos e também os difíceis.

Agradeço aos colegas Rafael Alves Paes de Oliveira (LabES/ICMC-USP) e Leila Cristina Carneiro Bergamasco (LApIS/EACH-USP) pelo precioso apoio que me concederam durante a condução deste trabalho.

Agradeço ao Prof. Dr. Márcio Eduardo Delamaro (LabES/ICMC-USP) pelo precioso apoio e parceria na condução dos projetos de pesquisa que me trouxeram até aqui.

Agradeço à minha tia, Ana Cristina da Silva Leão, por todo o apoio e amizade que tem dedicado à minha família, em especial, nos últimos anos; sem eles, com certeza, a caminhada até aqui seria muito mais difícil.

A todos o meu **muito obrigado!**

*“As pessoas mais felizes não têm as melhores coisas. Elas sabem fazer o melhor das
oportunidades que aparecem em seus caminhos.”*
(Clarice Lispector)

RESUMO

GONÇALVES, Vagner Mendonça. **Recuperação de vídeos médicos baseada em conteúdo utilizando extratores de características visuais e sonoras**. 2017. 99 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2016. Versão corrigida.

A evolução dos dispositivos de armazenamento e das redes de computadores permitiram que os vídeos digitais assumissem um importante papel no desenvolvimento de sistemas de informação multimídia. Com a finalidade de aproveitar todo o potencial dos vídeos digitais no desenvolvimento desses sistemas, técnicas automatizadas eficientes para análise, interpretação e recuperação são necessárias. A recuperação de vídeos baseada em conteúdo (CBVR, do inglês *content-based video retrieval*) permite o processamento e a análise do conteúdo de vídeos digitais visando à extração de informações relevantes que viabilizem indexação e recuperação. Trabalhos científicos têm proposto a aplicação de CBVR em bases de vídeos médicos a fim de proporcionar diferentes contribuições como diagnóstico auxiliado por computador, suporte à tomada de decisão e disponibilização de bases de vídeos para utilização em treinamento e educação médica. Em geral, características visuais são as principais informações utilizadas no contexto de CBVR aplicada em vídeos médicos. No entanto, muitos diagnósticos são realizados por meio da análise dos sons produzidos em diferentes estruturas e órgãos do corpo humano. Um exemplo é o diagnóstico cardíaco que, além de exames de imagem como ecocardiografia e ressonância magnética, também pode empregar a análise dos sons provenientes do coração por meio da auscultação. O objetivo deste trabalho consistiu em aplicar e avaliar extratores de características de som em conjunto com extratores de características visuais para viabilizar CBVR e, então, inferir se a abordagem resultou em ganhos com relação ao desempenho de recuperação quando comparada à utilização apenas das características visuais. Vídeos médicos constituíram nosso principal interesse, porém o trabalho considerou também vídeos não relacionados à área médica para a validação da abordagem. Justifica-se o objetivo, pois a análise do som, visando a obter descritores relevantes para melhorar os resultados de recuperação, ainda é pouco explorada na literatura científica. Essa afirmação foi evidenciada com a condução de uma revisão sistemática sobre o tema. Dois conjuntos de experimentos foram conduzidos visando a validar a abordagem de CBVR mencionada. O primeiro conjunto de experimentos foi aplicado sobre uma base de vídeos sintetizados para validação da abordagem. Já o segundo, foi aplicado em uma base de vídeos construídos utilizando-se imagens provenientes de exames de ressonância magnética em conjunto com sons provenientes de auscultação do coração. Os resultados foram analisados utilizando-se as métricas de revocação e precisão, bem como o gráfico que as relaciona. Demonstrou-se que a abordagem é promissora por meio da melhora significativa dos resultados de recuperação nos diferentes cenários de combinação entre características visuais e sonoras experimentados.

Palavras-chave: Recuperação de vídeos baseada em conteúdo. CBVR. Diagnóstico auxiliado por computador. Vídeos médicos. Imagens médicas.

ABSTRACT

GONÇALVES, Vagner Mendonça. **Content-based medical video retrieval using visual and sound feature extractors**. 2017. 99 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2016. Corrected version.

Advance of storage devices and computer networks has contributed to digital videos assume an important role in the development of multimedia information systems. In order to take advantage of the full potential of digital videos in the development of these systems, it is necessary the development of efficient techniques for automated data analysis, interpretation and retrieval. Content-based video retrieval (CBVR) allows processing and analysis of content in digital videos to extract relevant information and enable indexing and retrieval. Scientific studies have proposed the application of CBVR in medical video databases in order to provide different contributions like computer-aided diagnosis, decision-making support or availability of video databases for use in medical training and education. In general, visual characteristics are the main information used in the context of CBVR applied in medical videos. However, many diagnoses are performed by analysing the sounds produced in different structures and organs of the human body. An example is the cardiac diagnosis which, in addition to images generated by echocardiography and magnetic resonance imaging, for example, may also employ the analysis of sounds from the heart by means of auscultation. The objective of this work was evaluating combination between audio signal and visual features to enable CBVR and investigating how much this approach can improve retrieval results comparing to using only visual features. Medical videos are the main data of interest in this work, but video segments not related to the medical field were also used to validate the approach. The objectives of this work are justifiable because audio signal analysis, in order to get relevant descriptors to improve retrieval results, is still little explored in the scientific literature. This statement was evidenced by results of a systematic review. Two experiment sets were conducted to validate the CBVR approach described. The first experiment set was applied to a synthetic images database specially built to validate the approach, while the second experiment was applied to a database composed of digital videos created from magnetic resonance imaging and heart sounds from auscultation. Results were analyzed using the recall and precision metrics, as well as the graph which relates these metrics. Results showed that this approach is promising due the significantly improvement obtained in retrieval results to different scenarios of combination between visual and audio signal features.

Keywords: Content-based video retrieval. CBVR. Computer-aided diagnosis. Medical videos. Medical images.

LISTA DE FIGURAS

Figura 1 – Ilustração de um processo de aquisição e digitalização de uma imagem.	21
Figura 2 – Tipos de vizinhança de um pixel: (a) vizinhança-4; (b) vizinhança-8.	22
Figura 3 – Absorção de luz pelos cones R, G e B no olho humano como função do comprimento de onda.	23
Figura 4 – Representação esquemática do modelo RGB. Pontos ao longo da diagonal principal representam os níveis de cinza, do preto na origem até o branco no ponto (1, 1, 1).	24
Figura 5 – Passos fundamentais em processamento de imagens digitais.	25
Figura 6 – Representação de uma onda sonora.	27
Figura 7 – Representação do processo de digitalização de uma onda sonora, com amostras obtidas a intervalos regulares.	28
Figura 8 – A dimensionalidade de imagens e vídeos.	30
Figura 9 – Esquema simplificado de um sistema de CBR.	31
Figura 10 – Consulta por similaridade: abrangência.	33
Figura 11 – Consulta por similaridade: k vizinhos mais próximos.	33
Figura 12 – Representação de um gráfico de revocação <i>versus</i> precisão contendo uma curva ideal e uma curva realista.	35
Figura 13 – Esquema de circulação do sangue.	37
Figura 14 – Anatomia do coração humano.	38
Figura 15 – Regiões auscultatórias.	40
Figura 16 – Forma de onda do som de batimentos cardíacos com a presença dos sons S1 e S2.	40
Figura 17 – Forma de onda do som de batimentos cardíacos com a presença dos sons S1, S2 e S3.	41
Figura 18 – Forma de onda do som de batimentos cardíacos com a presença dos sons S1, S2 e S4.	42
Figura 19 – Evolução da quantidade de óbitos no Brasil por doenças do aparelho circulatório entre os anos de 2005 e 2014.	43
Figura 20 – Processo de captação de imagens por meio de Ressonância Magnética Nuclear.	44

Figura 21 – Diagrama de fluxo resumizando a etapa de seleção de estudos.	51
Figura 22 – Fases do projeto de pesquisa apresentado nesta dissertação.	61
Figura 23 – Exemplo de quadro representando uma fatia do exame de RMN de um paciente com diagnóstico de ICC.	66
Figura 24 – Exemplo de imagem retratando o ventrículo esquerdo segmentado por meio da marcação do endocárdio.	67
Figura 25 – Ilustração da rotina de cálculo do vetor de distâncias para obtenção da assinatura de forma.	68
Figura 26 – Modelo de consultas por similaridade aplicado para a condução dos experimentos.	70
Figura 27 – Gráfico de revocação <i>versus</i> precisão das consultas que utilizaram apenas os extratores sonoros para recuperar casos com ausência de ICC.	71
Figura 28 – Gráfico de revocação <i>versus</i> precisão das consultas que utilizaram apenas os extratores de forma para recuperar casos com presença de ICC.	72
Figura 29 – Gráfico de revocação <i>versus</i> precisão das consultas que utilizaram apenas as características de forma para recuperar casos com ausência de ICC.	75
Figura 30 – Gráfico de revocação <i>versus</i> precisão das consultas que utilizaram apenas as características de forma para recuperar casos com presença de ICC.	75
Figura 31 – Gráfico de revocação <i>versus</i> precisão das consultas que utilizaram combinação entre características de forma e de som para recuperar casos com ausência de ICC.	76
Figura 32 – Gráfico de revocação <i>versus</i> precisão das consultas que utilizaram combinação entre características de forma e de som para recuperar casos com presença de ICC.	76

LISTA DE TABELAS

Tabela 1 – Exemplos de características de imagens digitais.	26
Tabela 2 – Composições de termos utilizadas nas buscas.	49
Tabela 3 – Números relativos às buscas realizadas em cada base consultada. . . .	50
Tabela 4 – Trabalhos incluídos e principais tópicos de interesse para a revisão sistemática - parte 1/2.	58
Tabela 5 – Trabalhos incluídos e principais tópicos de interesse para a revisão sistemática - parte 2/2.	59
Tabela 6 – Características fisiológicas associadas aos sons do coração utilizados. .	65
Tabela 7 – Valores de AUC calculados para as consultas que visaram a recuperação de casos com ausência de ICC.	77
Tabela 8 – Valores de AUC calculados para as consultas que visaram a recuperação de casos com presença de ICC.	77

LISTA DE ABREVIATURAS E SIGLAS

AUC	Area under the curve (área sob a curva)
CAD	Computer-aided diagnosis (diagnóstico auxiliado por computador)
CBIS	Congresso Brasileiro de Informática em Saúde
CBR	Content-based retrieval (recuperação baseada em conteúdo)
CBIR	Content-based image retrieval (recuperação de imagens baseada em conteúdo)
CBVR	Content-based video retrieval (recuperação de vídeos baseada em conteúdo)
CMY	Cyan, magenta, yellow (ciano, magenta, amarelo)
CMYK	Cyan, magenta, yellow, black (ciano, magenta, amarelo, preto)
HSI	Hue, saturation, intensity (matiz, saturação, intensidade)
ICC	Insuficiência cardíaca congestiva
InCor	Instituto do Coração da Faculdade de Medicina da Universidade de São Paulo
LApIS	Laboratório de Aplicações de Informática em Saúde
MFCC	Mel-frequency cepstral coefficients (coeficientes cepstrais de frequência Mel)
OCR	Optical character recognition (reconhecimento ótico de caracteres)
RGB	Red, green, blue (vermelho, verde, azul)
RMN	Ressonância magnética nuclear
RS	Revisão sistemática
TTS	Text-to-speech (síntese de voz e texto-fala)

SUMÁRIO

1	Introdução	15
1.1	Considerações iniciais	15
1.2	Motivação	16
1.3	Problema, questão de pesquisa e objetivos	17
1.4	Organização do documento	18
2	Aspectos conceituais	20
2.1	Considerações iniciais	20
2.2	Imagens digitais	20
2.2.1	Imagem colorida	22
2.2.2	Processamento de imagens digitais	23
2.2.3	Extração de características	25
2.3	Sinais sonoros	26
2.3.1	Processamento de som digital	27
2.4	Vídeos digitais	29
2.5	Recuperação baseada em conteúdo	30
2.5.1	Métricas de desempenho de recuperação em sistemas de CBR	34
2.6	Diagnóstico auxiliado por computador	34
2.7	O coração humano	36
2.7.1	Sistema circulatório humano e anatomia do coração	36
2.7.2	Auscul ta cardíaca	39
2.7.3	Insuficiência cardíaca congestiva	42
2.7.4	Ressonância magnética nuclear	43
2.8	Considerações finais	44
3	Revisão sistemática	46
3.1	Considerações iniciais	46
3.2	Planejamento e condução	46
3.2.1	Planejamento	47
3.2.2	Condução	50

3.3	Resultados e discussões	50
3.3.1	Trabalhos incluídos e considerações sobre os resultados . .	50
3.3.2	Aplicações de interesse e modalidades de vídeos	52
3.3.3	Métodos e técnicas utilizados nos trabalhos identificados .	54
3.4	Considerações finais	55
4	Metodologia	60
4.1	Considerações iniciais	60
4.2	Fases do projeto de pesquisa	60
4.3	Materiais e métodos empregados para a condução de experimentos . .	64
4.3.1	Base de vídeos	64
4.3.2	Segmentação do ventrículo esquerdo	65
4.3.3	Extratores de características	66
4.3.4	Função de similaridade	69
4.3.5	Metodologia para a execução das consultas	70
4.4	Considerações finais	72
5	Resultados experimentais com exames cardiológicos	74
5.1	Considerações iniciais	74
5.2	Resultados e discussões	74
5.3	Considerações finais	78
6	Conclusões	79
6.1	Trabalhos futuros	81
	Referências¹	82
	APÊNDICE A – Protocolo de revisão sistemática	88
	APÊNDICE B – Artigo científico produzido baseado em re- sultados de experimentos em uma base de vídeos sintéticos	92

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

CAPÍTULO 1 – INTRODUÇÃO

1.1 CONSIDERAÇÕES INICIAIS

A evolução dos dispositivos de armazenamento e das redes de computadores permitiram que os vídeos digitais assumissem um importante papel no desenvolvimento de sistemas de informação multimídia. Diferentes modalidades de vídeos são empregadas em áreas como educação, cultura, segurança, entretenimento e medicina.

Com a finalidade de aproveitar todo o potencial dos dados multimídia, tais como os vídeos digitais, no desenvolvimento de sistemas de informação são necessárias a investigação e a utilização de técnicas automatizadas eficientes para análise, interpretação e recuperação desses dados. No entanto, a interação com grandes quantidades de dados multimídia é uma tarefa significativamente desafiadora (GRIERSON; CORNEY; HATCHER, 2015). Técnicas como a recuperação baseada em conteúdo (CBR, do inglês *content-based retrieval*), que envolve a extração e a comparação entre características de imagens, vídeos e sons, têm sido exploradas em trabalhos da área.

O princípio básico dos sistemas que utilizam CBR é pesquisar em um banco de dados uma determinada quantidade de objetos (imagens, vídeos, sons) similares a um objeto estabelecido como referência de consulta, de acordo com um ou mais critérios definidos. Os critérios de similaridade são obtidos por meio da extração de características, geralmente relacionadas a cores, texturas, formas, movimentos e som (DEMENTHON; DOERMANN, 2003; DATTA et al., 2008; ABDULLAH, 2009).

O conjunto de características que descreve um objeto forma o seu vetor de características. A comparação entre os vetores de características de dois objetos permite definir o nível de similaridade entre eles. Em uma arquitetura básica, uma função de similaridade pode ser utilizada para se calcular uma distância entre os vetores de dois objetos. No entanto, abordagens mais sofisticadas como, por exemplo, a definição de modelos por meio do treinamento de classificadores têm sido amplamente investigadas na literatura científica da área (DATTA et al., 2008).

CBR tem sido explorada nas últimas décadas em trabalhos científicos para o desenvolvimento dos mais diversos tipos de aplicações (DATTA et al., 2008). Diversos trabalhos têm se dedicado ao desenvolvimento de técnicas que sejam úteis, por exemplo, no manuseio e na utilização da grande quantidade de imagens digitais disponíveis atualmente.

Uma das áreas na qual essas técnicas se tornam úteis é a área médica, considerando-se a necessidade de desenvolvimento de sistemas de diagnóstico auxiliado por computador (CAD, do inglês *computer-aided diagnosis* ou *computer-assisted diagnosis*).

Sistemas de CAD consistem em sistemas de informação com a finalidade de fornecer uma segunda opinião aos especialistas da área médica (médicos, radiologistas), contribuindo na tomada de decisão a respeito de um diagnóstico (DOI, 2007). Esses sistemas computam a saída a partir de informações provenientes de diversas fontes de dados, sendo que as imagens médicas de diversas modalidades são as principais delas.

Além das imagens médicas, diferentes modalidades de vídeos digitais são empregadas em tarefas que vão desde o diagnóstico (exames de ultrassonografia e endoscopia, por exemplo), passando pelo monitoramento de procedimentos e cirurgias, até tarefas de auxílio à educação médica.

A recuperação de vídeos baseada em conteúdo (CBVR, do inglês *content-based video retrieval*) é uma variação de CBR que permite o processamento e a análise do conteúdo de vídeos digitais visando à extração de informações relevantes que viabilizem indexação e recuperação. Como se pode verificar em Zhang et al. (1997) e Su et al. (2010), abordagens de CBVR têm sido exploradas na literatura científica. A aplicação de técnicas de CBVR tem potencial para auxiliar na resolução de problemas relacionados ao CAD como se verifica nas abordagens apresentadas em André et al. (2011b) e Quellec et al. (2014). Além disso, CBVR pode ser de grande utilidade na área de segurança, como se verifica em DeMenthon e Doermann (2003).

1.2 MOTIVAÇÃO

Diversos estudos na área de computação e sistemas de informação dedicam esforços para contribuir com o auxílio ao diagnóstico. Diferentes desafios surgem dentro desta área tão complexa. Um exemplo é a identificação de regiões suspeitas em imagens médicas. Tal tarefa é de grande relevância, pois auxilia radiologistas a aumentarem a sua precisão nos diagnósticos, chamando a atenção para regiões muitas vezes despercebidas.

Com o constante desenvolvimento e a maior utilização de técnicas de diagnóstico por imagem, o volume de vídeos provenientes de exames médicos e de monitoramento de procedimentos cirúrgicos cresceu consideravelmente nos últimos anos (ANDRÉ et al.,

2011a; HAN et al., 2013; QUELLEC et al., 2014). Diante desse cenário, cresceu também a relevância de ferramentas automatizadas para auxiliar os profissionais da área médica na análise desses vídeos, fornecendo suporte ao diagnóstico e à avaliação de procedimentos realizados nos pacientes.

CBVR pode ser, então, aplicada nesse contexto para diversas finalidades. Como exemplos, pode-se mencionar:

- suporte à tomada de decisão ao possibilitar que o médico consulte casos semelhantes a um dado caso atualmente em análise;
- definição de bases de vídeos para treinamento e educação médica nas quais os usuários podem recuperar vídeos com base em segmentos de vídeos de referência ou descrição das características visuais ou sonoras de interesse;
- classificação de anomalias por meio da identificação de casos semelhantes.

Uma revisão sistemática conduzida e apresentada na presente dissertação (capítulo 3) evidenciou que a aplicação do conceito de CBVR na área médica é recente e foi pouco explorada. Logo, o processamento, a análise e a recuperação de vídeos médicos constitui uma oportunidade de pesquisa que pode gerar contribuições significativas.

1.3 PROBLEMA, QUESTÃO DE PESQUISA E OBJETIVOS

O **problema** de pesquisa abordado neste trabalho de mestrado divide-se em dois eixos. O primeiro eixo se refere à investigação e à aplicação de métodos e técnicas visando a melhorar o desempenho de recuperação em sistemas de CBVR. Já o segundo eixo, refere-se à investigação de abordagens de CBVR para viabilizar CAD.

A abordagem proposta consiste na análise combinada entre características sonoras (extraídas de um sinal sonoro digital) e características visuais (extraídas do conteúdo visual de uma imagem ou de uma sequência de imagens) para viabilizar ganhos com relação ao desempenho de recuperação. Com a condução de um estudo de caso utilizando-se uma base de dados provenientes de exames cardiológicos, foi estudada também a viabilidade da aplicação dessa tecnologia no desenvolvimento de sistemas de CAD para a área da Cardiologia.

O **objetivo** principal do trabalho consistiu em aplicar e avaliar extratores de características sonoras em conjunto com extratores de características visuais para viabilizar

recuperação de vídeos médicos baseada em conteúdo e, então, inferir se a abordagem resultou em ganhos com relação ao desempenho de recuperação quando comparada à utilização apenas das características visuais.

De acordo com os estudos teóricos exploratório e sistemático conduzidos, a maior parte dos trabalhos que envolveu CBVR aplicou descritores relacionados a características visuais espaciais e temporais dos objetos representados, tais como formas, cores e movimentos. A análise do som, visando a obter descritores relevantes para melhorar os resultados de recuperação, ainda é pouco explorada na literatura científica. Um trabalho correlato ao nosso é o estudo apresentado por Abdullah (2009) que descreveu uma abordagem de combinação entre características visuais e características sonoras para a identificação de ações humanas em vídeos. O estudo envolveu o processamento de segmentos de vídeos provenientes de filmes e seriados. Os experimentos evidenciaram a obtenção de melhores resultados de recuperação com a combinação das características visuais e sonoras em comparação com a utilização apenas das características visuais.

A partir das considerações tecidas, enuncia-se a seguinte **questão de pesquisa** a ser respondida durante o presente trabalho: “O emprego de características sonoras, em adição a características visuais, pode melhorar o nível de descrição de vídeos e, conseqüentemente, melhorar os resultados de recuperação por conteúdo em vídeos no contexto médico?”

Foram estabelecidos como objetivos específicos do trabalho:

- realizar um estudo teórico sobre os conceitos e técnicas relacionados ao tema do trabalho por meio de revisões de literatura exploratória e sistemática;
- definir uma metodologia experimental que forneça subsídios para se responder à questão de pesquisa enunciada;
- aplicar um conjunto significativo de extratores de características visuais e sonoras viabilizando a descrição dos objetos (vídeos) de interesse;
- aplicar a abordagem de combinação entre características sonoras e características visuais em um estudo de caso, utilizando-se resultados reais de exames médicos.

1.4 ORGANIZAÇÃO DO DOCUMENTO

Além deste capítulo introdutório, esta dissertação está organizada da seguinte forma:

- **Capítulo 2 - Aspectos conceituais:** apresenta os principais conceitos e técnicas que embasam este trabalho de Mestrado.
- **Capítulo 3 - Revisão sistemática:** apresenta, detalha e discute a condução e os resultados da revisão bibliográfica sistemática conduzida, cujo objetivo foi levantar o estado da arte com relação à aplicação de CBVR em sistemas de informação voltados para a área médica.
- **Capítulo 4 - Metodologia:** apresenta a metodologia aplicada para a condução do projeto de pesquisa referente a este trabalho de mestrado.
- **Capítulo 5 - Estudo de caso com exames cardiológicos:** apresenta um estudo de caso planejado e conduzido utilizando-se dados provenientes de exames cardiológicos; detalha e discute os objetivos, materiais e métodos, bem como os resultados obtidos.
- **Capítulo 6 - Conclusões:** apresenta as conclusões do trabalho, bem como sugestões de trabalhos futuros.

CAPÍTULO 2 – ASPECTOS CONCEITUAIS

2.1 CONSIDERAÇÕES INICIAIS

Este capítulo introduz conceitos, técnicas e ferramentas utilizados para a condução da pesquisa referente a este trabalho de Mestrado. Inicialmente, são definidos os conceitos que embasam os artefatos abordados no presente trabalho: imagens digitais, sinais sonoros e vídeos digitais, detalhados nas seções 2.2, 2.3 e 2.4, respectivamente. Em seguida, na seção 2.5, são apresentadas a definição e as principais características de sistemas de CBR. Na seção 2.6, por sua vez, são apresentados o conceito e as características de sistemas de CAD, bem como os seus principais objetivos. Por fim, na seção 2.7, são introduzidos os principais conceitos referentes ao coração humano, à insuficiência cardíaca congestiva e aos exames cardiológicos que embasaram o estudo de caso conduzido neste projeto de pesquisa.

2.2 IMAGENS DIGITAIS

A maior parte das imagens de interesse quando nos referimos ao processamento de imagens digitais é formada pela combinação entre uma fonte de iluminação e a reflexão ou absorção da energia proveniente desta fonte pelos elementos da cena capturada (GONZALEZ; WOODS, 2008). A captura de uma imagem, em geral, envolve a utilização de um sensor cuja saída é composta por sinais contínuos. Uma imagem digital é o resultado da transformação desse sinal contínuo capturado do ambiente em um sinal discreto passível de ser armazenado e processado por meio de computadores (BOVIK, 2000). A esse processo de discretização, denominamos **digitalização**.

No domínio espacial, uma imagem digital $f(x, y)$ pode ser representada por meio de uma matriz bidimensional com M linhas e N colunas, em que o par (x, y) representa as coordenadas espaciais discretas da imagem e $f(x, y)$ representa a intensidade (brilho ou cor) obtida no par de coordenadas (x, y) . Considerando-se $x = 0, 1, 2, \dots, M - 1$ e

$y = 1, 2, \dots, N - 1$, a representação de $f(x, y)$, partindo-se do ponto inicial da imagem $f(0, 0)$, é apresentada na equação 1. (GONZALEZ; WOODS, 2008).

$$f(x, y) = \begin{bmatrix} f(0, 0) & f(0, 1) & \dots & f(0, N - 1) \\ f(1, 0) & f(1, 1) & \dots & f(1, N - 1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M - 1, 0) & f(M - 1, 1) & \dots & f(M - 1, N - 1) \end{bmatrix} \quad (1)$$

A obtenção de uma imagem digital a partir dos sinais contínuos capturados por meio do sensor é possível por meio dos processos de **amostragem** e **quantização**. Considerando-se a imagem $f(s, t)$ de duas variáveis contínuas, s e t , obtemos a imagem digital correspondente, $f(x, y)$, com a digitalização das coordenadas (processo de amostragem) e, em seguida, com a digitalização da intensidade de brilho em cada ponto (processo de quantização) (GONZALEZ; WOODS, 2008).

Uma ilustração do processo de aquisição e digitalização de uma imagem é apresentada na figura 1. O sistema de imageamento, composto pelo sensor, captura o elemento da cena por meio da sensibilidade ao efeito da energia que incide sobre o mesmo. Internamente, o elemento é projetado sobre um plano de imagem. Então, por meio de processos de amostragem e quantização, a imagem é obtida.

Figura 1 – Ilustração de um processo de aquisição e digitalização de uma imagem.

Cada ponto de uma imagem digital é denominado **pixel**. A **vizinhança** de um pixel em uma imagem digital determina como ele se relaciona com os pixels ao seu entorno no domínio espacial. Essa relação de vizinhança entre os pixels é importante, pois embasa muitos algoritmos de processamento de imagens (GONZALEZ; WOODS, 2008; PEDRINI; SCHWARTZ, 2008). Na figura 2 são apresentados os dois principais tipos de vizinhança de um pixel aplicados em algoritmos de processamento de imagens digitais (PEDRINI; SCHWARTZ, 2008).

Figura 2 – Tipos de vizinhança de um pixel: (a) vizinhança-4; (b) vizinhança-8.

Fonte: Pedrini e Schwartz (2008)

2.2.1 Imagem colorida

Existem muitas teorias postuladas para se explicar a visão colorida humana. A mais clássica diz que, em um olho humano normal, cores são percebidas como combinações não-lineares de comprimentos de onda longos, médios e curtos (BOVIK, 2000; PRATT, 2007). Grosseiramente, isto corresponde às três cores primárias utilizadas em sistemas padrão de imageamento: vermelho (R, do inglês *red*), verde (G, do inglês *green*) e azul (B, do inglês *blue*) (BOVIK, 2000).

O olho humano possui células especiais, os **cones**, responsáveis pela absorção da luz e reconhecimento das cores. Basicamente, são três os tipos de cones, cada um responsável pela percepção de uma faixa de comprimentos de onda e, conseqüentemente, de uma cor primária R, G ou B (GONZALEZ; WOODS, 2008). Na figura 3 é apresentado um gráfico que relaciona a absorção da luz pelos cones R, G e B em função do comprimento de onda.

A representação das cores em uma imagem digital é realizada por meio de um **modelo de cor**. Um modelo de cor é uma especificação de um sistema de coordenadas

Figura 3 – Absorção de luz pelos cones R, G e B no olho humano como função do comprimento de onda.

Fonte: Gonzalez e Woods (2008)

e um subespaço dentro do sistema no qual cada cor é representada por um único ponto (GONZALEZ; WOODS, 2008). Uma representação esquemática do sistema de coordenadas do modelo RGB (do inglês *red, green, blue*), principal utilizado em monitores coloridos e grande parte das câmeras de vídeo, é apresentada na figura 4.

Além do modelo de cor RGB, outros modelos são também comumente utilizados como, por exemplo, os modelos CMY (do inglês *cyan, magenta, yellow*) e CMYK (do inglês *cyan, magenta, yellow, black*), para impressão de cores; e o modelo HSI (do inglês *hue, saturation, intensity*), baseado nas componentes matiz, saturação e intensidade, que representam bem o modo como os humanos descrevem e interpretam cores (GONZALEZ; WOODS, 2008).

2.2.2 Processamento de imagens digitais

São muitas as motivações que justificam o emprego de técnicas de processamento de imagens digitais. Girod e Wetzstein (2016) apresentam as principais, bem como exemplos, conforme listagem a seguir:

Figura 4 – Representação esquemática do modelo RGB. Pontos ao longo da diagonal principal representam os níveis de cinza, do preto na origem até o branco no ponto $(1, 1, 1)$.

Fonte: Gonzalez e Woods (2008)

- **capturar uma imagem:** enquadramento correto, reconstrução de imagens a partir de projeções;
- **preparar para exibição ou impressão:** ajuste do tamanho da imagem, mapeamento de cores, correções da luminosidade;
- **facilitar armazenamento e transmissão:** armazenamento eficiente em câmeras digitais, envio de imagens do espaço;
- **melhorar e restaurar imagens:** retoque de imagens pessoais, melhoramento de cor para aplicações de biometria e segurança;
- **extrair informações das imagens:** leitura de código de barras, reconhecimento de padrões e características.

De forma resumida, melhorar a informação visual para a interpretação humana e processar dados de cenas para percepção automática por meio de máquinas são as principais áreas de aplicação que justificam o interesse em métodos de processamento de imagens digitais (GONZALEZ; WOODS, 2000).

Na figura 5 é apresentado um esquema com os passos fundamentais em processamento de imagens digitais. Uma aplicação, em geral, atende a um domínio de problema. Sendo assim, uma **base de conhecimento** relacionada ao domínio do problema é parte integrante do sistema e guia todos os processos que ocorrem desde a sua entrada até a sua saída (GONZALEZ; WOODS, 2000).

Figura 5 – Passos fundamentais em processamento de imagens digitais.

Fonte: Gonzalez e Woods (2000)

Após a **aquisição** da imagem, a mesma deve ser pré-processada visando a melhorar a mesma para um aumento das chances de sucesso dos processos posteriores. Realizado o **pré-processamento**, em geral, tem-se uma etapa de **segmentação**, isto é, identificação das partes de uma imagem e seleção das regiões ou objetos de interesse. Identificada a região ou objeto de interesse, extrai-se características dessa região a fim de se **representá-la e descrevê-la**, possibilitando, assim, o processo de **reconhecimento e interpretação** que, em geral, dará o resultado do sistema de processamento de imagens (GONZALEZ; WOODS, 2000).

2.2.3 Extração de características

A extração de características de imagens digitais é uma etapa fundamental em sistemas de recuperação de imagens baseada em conteúdo (CBIR, do inglês *content-based image retrieval*). CBIR é uma variação de CBR que tem sido investigada, aprimorada e aplicada em diversos estudos científicos ao longo das últimas décadas (DATTA et al., 2008).

Diversas características visuais do conteúdo de uma imagem podem ser extraídas com a utilização de diferentes técnicas. Tais características podem ser relacionadas ao espaço ocupado por uma região de interesse da imagem, à sua forma, às cores representadas na imagem ou a texturas. Características de textura representam informações relevantes sobre a granularidade e padrões repetitivos de superfícies em uma imagem (WANG; ZHANG; YANG, 2014). Exemplos de características de cada categoria são apresentados na tabela 1 juntamente com referências nas quais é possível obter maiores detalhes sobre as mesmas.

Tabela 1 – Exemplos de características de imagens digitais.

Categoria	Características	Referências
Região	Área, perímetro, largura, altura e centro de massa.	Nunes et al. (2015)
Forma	Assinatura de forma (regularidade da borda).	Delamaro, Nunes e Oliveira (2013)
Cor	Média, desvio padrão, assimetria, energia e entropia com base no histograma de cor.	Sergyán (2008)
Textura	Matriz de coocorrência de níveis de cinza; modelo baseado em campo aleatório de Markov; modelo autoregressivo simultâneo.	Wang, Zhang e Yang (2014)

Fonte: Vagner Mendonça Gonçalves, 2017

Neste trabalho de pesquisa, características de região, forma e cor foram empregadas para descrever as regiões de interesse em imagens provenientes dos segmentos de vídeos processados.

2.3 SINAIS SONOROS

O **som** é qualquer vibração do ar (variação de pressão) que possa ser detectada pelo ouvido humano (MURGEL, 2007). Sua **origem** se dá, em geral, por vibrações de objetos materiais (HEWITT, 2002). O som se propaga por um meio material sem que ocorra o transporte de matéria de um ponto a outro, o que indica a sua **natureza ondulatória** (NUSSENZVEIG, 2004).

O som pode ser representado como uma função do tempo, em que ambos os valores de amplitude do som e de tempo são contínuos na natureza (BOSI; GOLDBERG, 2003). Uma representação básica de onda sonora é apresentada na figura 6 (GIL et al., 2014).

Figura 6 – Representação de uma onda sonora.

Fonte: Gil et al. (2014)

Para permitir que o som seja processado por computadores é necessária a sua digitalização, isto é, a conversão do sinal contínuo para um sinal discreto, por meio da amostragem do sinal. Para isso, a amplitude da onda sonora é dividida em intervalos precisos de tempo e, assim, obtém-se uma discretização da onda em relação ao tempo. Uma ilustração da digitalização de uma onda sonora é apresentada na figura 7. Cada valor de amplitude obtido, denominado **amostra**, precisa ter uma precisão finita de bits para ser processado. O processo responsável por definir essa precisão é denominado quantização (BOSI; GOLDBERG, 2003).

Uma propriedade básica do som digital é a **frequência de amostragem**, que representa o número de amostras obtidas por segundo durante o processo de amostragem e é medida em Hertz (Hz) (BOSI; GOLDBERG, 2003).

2.3.1 Processamento de som digital

O processamento de sons digitais tem sido aplicado em diversos problemas, tais como segmentação, reconhecimento de discurso, recuperação de informação em músicas e reconhecimento de sons para propósitos gerais e ambientais (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010).

Figura 7 – Representação do processo de digitalização de uma onda sonora, com amostras obtidas a intervalos regulares.

Fonte: Gil et al. (2014)

Em alto nível, sons são tradicionalmente descritos por propriedades como altura, volume, duração e timbre, sendo as três primeiras percepções psicológicas bem definidas e eficazmente mensuráveis, enquanto que o timbre é um atributo mais complexo (WOLD et al., 1996).

A **altura** de um som musical, por exemplo, corresponde à sensação que permite distinguir entre sons mais graves e mais agudos (NUSSENZVEIG, 2004), estando relacionada à frequência. O **volume** é uma sensação fisiológica relacionada à capacidade do ouvido sentir algumas frequências melhores do que outras (HEWITT, 2002); ele permite a ordenação de diferentes sons em uma escala que vai do baixo ao alto. **Duração** é o tempo entre o início e o fim de um sinal sonoro de interesse (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010). O **timbre**, por fim, é uma propriedade que permite distinguir dois sons de mesma altura e volume; por exemplo, a distinção entre a nota musical “lá” emitida por um piano, violino, flauta ou pela voz humana (NUSSENZVEIG, 2004).

Em geral, características utilizadas em aplicações de processamento de sons digitais descrevem aspectos das propriedades do som mencionadas. Mitrović, Zeppelzauer e Breiteneder (2010) propuseram uma taxonomia para classificação de características sonoras aplicáveis em diferentes domínios de problemas relacionados ao processamento e recuperação de sons digitais. No presente trabalho, utilizou-se características provenientes de três domínios com base nessa taxonomia. São eles:

- **Domínio temporal:** representa as mudanças do sinal sonoro ao longo do tempo, ou seja, a forma da onda.

- **Domínio de frequência:** revela a distribuição espectral de um sinal e permite, por exemplo, análises de estrutura harmônica, largura de banda e tonalidade.
- **Domínio Cepstral:** representação obtida tomando-se a transformada de Fourier do logaritmo da magnitude do espectro; é um modo de se computar uma aproximação da forma do espectro. Em geral, características cepstrais obtêm informações sobre timbre e têm sido aplicadas em reconhecimento de discurso e biometria por som (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010).

As características de sinais sonoros empregadas no presente projeto de pesquisa são apresentadas na subseção 4.3.3, dentro do contexto da descrição do estudo de caso conduzido.

2.4 VÍDEOS DIGITAIS

Vídeos consistem em uma sequência de estímulos visuais (imagens) que, transmitidos aos nossos olhos a uma taxa rápida o suficiente, nos fornecem a percepção de objetos em movimento (SANTOS; MORIMOTO, 2008). Um **vídeo digital** é o resultado da conversão de cenas, representadas por sinais contínuos, para um formato passível de armazenamento e processamento via computador (BOVIK, 2000). O conceito de digitalização aplicado a vídeos é o mesmo empregado em imagens e sons, apresentados nas seções anteriores.

Enquanto imagens são representadas no espaço bidimensional, vídeos possuem uma terceira dimensão, o tempo. Na figura 8 é apresentada uma representação esquemática das dimensionalidades de uma imagem digital e de uma sequência de vídeo digital.

Em um vídeo digital destacam-se dois componentes importantes para fins de processamento: o **quadro** e a **tomada**. Um quadro corresponde a uma imagem estática do vídeo. Uma tomada, por sua vez, é uma sequência de quadros que representam uma ação contínua no tempo e no espaço (SANTOS; MORIMOTO, 2008).

Um outro conceito importante é o conceito de **cena** que consiste na combinação de tomadas. O conjunto de cenas compõe o vídeo final. O conceito de tomada é a mais fina granularização referente à unidade de ação e têm sido utilizada em rotinas de segmentação e sumarização de vídeos (SANTOS; MORIMOTO, 2008).

Quadros podem ser utilizados para sumarizar o conteúdo de trechos de vídeo em termos de cores, texturas e formas. A utilização de um único quadro-chave possibilita de

Figura 8 – A dimensionalidade de imagens e vídeos.

Fonte: Bovik (2000)

forma simplificada a indexação de trechos de vídeo por meio de tais características. As tomadas, por sua vez, são adequadas para a extração de características de movimento em trechos de vídeos (SANTOS; MORIMOTO, 2008).

Um dos problemas de grande importância para fins de processamento, indexação e análise de imagens é a segmentação do vídeo em tomadas. A transição entre tomadas fornece informações essenciais para a identificação de mudança de contexto em um vídeo. A transição entre tomadas pode ser obtida por meio de simples concatenação entre quadros ou por meio da edição e inclusão de efeitos de transição entre as tomadas (SANTOS; MORIMOTO, 2008).

2.5 RECUPERAÇÃO BASEADA EM CONTEÚDO

Nesta dissertação, utilizou-se o termo **dado complexo** para definir imagens, sons e vídeos. Essa terminologia foi adotada visando a facilitar as referências aos dados de consulta quando o contexto refere-se à CBR de forma genérica.

Um sistema de CBR funciona seguindo, basicamente, os seguintes passos. Dada uma base de dados complexos (imagens, sons, vídeos), deve-se extrair e indexar n características escolhidas de cada dado complexo, compondo um **vetor de características** para cada um deles. Um usuário fornece, então, um **dado de consulta** (modelo) da qual serão

extraídas as mesmas n características que formarão o seu vetor de características. O vetor de características do dado modelo será, então, comparado aos vetores de cada dado complexo presente na base. O sistema deve reconhecer os dados da base mais semelhantes ao dado de consulta medindo a similaridade entre os vetores de características e retornando, ao final, um conjunto ordenado finito de dados mais semelhantes. Na figura 9 é apresentado um esquema simplificado de um sistema de CBR.

Figura 9 – Esquema simplificado de um sistema de CBR.

Fonte: Nunes e Delamaro (2010)

Extraír as características dos dados complexos é uma etapa essencial. Os **extratores de características** são algoritmos que utilizam técnicas de processamento de sinais para representar numericamente as propriedades de um dado complexo. Isto possibilita gerar vetores de características para cada dado e, posteriormente, compará-los.

Extraídas as características dos dados complexos, o passo seguinte consiste em compará-las por meio de alguma **função de similaridade**. Embora algoritmos mais complexos (redes neurais artificiais, redes probabilísticas, máquinas de vetores de suporte, entre outros) possam ser empregados para esta finalidade, a comparação mais simples é feita por meio de funções de similaridade. Uma função de similaridade é, em geral, baseada em alguma distância métrica entre vetores e pode ser utilizada para comparar vetores de características (NUNES; DELAMARO, 2010). Dados $X = (x_1, x_2, x_3, \dots, x_n)$, $Y = (y_1, y_2, y_3, \dots, y_n)$ e $Z = (z_1, z_2, z_3, \dots, z_n)$, três vetores de dimensão n , uma distância

métrica ou função de distância ou, simplesmente, função de similaridade Dis deve respeitar as seguintes propriedades (GONZALEZ; WOODS, 2000; LI; HOU; LI, 2006):

- **não-negatividade:** $Dis(X, Y) \geq 0$;
- **identidade:** $Dis(X, Y) = 0$ se e somente se $X = Y$;
- **simetria:** $Dis(X, Y) = Dis(Y, X)$, e
- **desigualdade triangular:** $Dis(X, Z) \leq Dis(X, Y) + Dis(Y, Z)$.

Desse modo, quanto mais próxima de zero for a distância entre dois vetores de características, mais semelhantes os dados serão para o sistema de CBR.

A família de distâncias métricas *Minkowski* ou L_p é composta pelas funções de similaridade mais utilizadas em trabalhos que envolvem CBR. São distâncias tradicionalmente utilizadas, porém, muitas vezes, escolhidas empiricamente (BUGATTI; TRAINA; TRAINA-JR, 2008). Considerando-se A e B , os vetores de características de dois dados complexos, e n , o número de atributos (extratores de características) utilizados para a realização das comparações, a forma geral das distâncias da família *Minkowski* é apresentada na equação 2. Quando $p = 2$, temos a distância **Euclidiana**.

$$Dis_{L_p}(A, B) = \sqrt[p]{\sum_{i=0}^{n-1} |a_i - b_i|^p} \quad (2)$$

A partir do momento em que um usuário fornece um dado modelo para um sistema de CBR e recebe como retorno do programa um conjunto dos dados mais semelhantes, afirma-se que foi realizada uma **consulta por similaridade**. Em CBR, dois tipos de consultas por similaridade são mais comuns: **consultas por abrangência** e **consultas pelos k -vizinhos mais próximos** (BUGATTI, 2008). As consultas por abrangência envolvem definir um limiar t que significa retornar todas os dados presentes no banco de dados cuja distância com o dado modelo seja menor ou igual a t .

Já as consultas pelos k -vizinhos mais próximos significam fornecer ao sistema de CBR um valor k , tal que o sistema retorne para o usuário os k dados presentes no banco mais semelhantes ao dado modelo (BUGATTI, 2008). Na figura 10 é apresentado um esquema de consulta por abrangência. Já na figura 11 é apresentado um esquema de consulta pelos k -vizinhos mais próximos com $k = 7$. Os elementos cinzas de cada consulta representam os objetos modelos, enquanto os elementos pretos representam os objetos recuperados.

Figura 10 – Consulta por similaridade: abrangência.

Fonte: Bugatti (2008)

Figura 11 – Consulta por similaridade: k vizinhos mais próximos.

Fonte: Bugatti (2008)

As **estruturas de indexação**, elementos fortemente relacionados aos bancos de dados, possuem influência considerável em consultas por similaridade. Recuperar dados complexos utilizando técnicas de CBR requer a comparação de vetores de alta dimensionalidade, diretamente relacionada com a quantidade de características extraídas para a pesquisa. Assim, é necessária a otimização do desempenho aplicando-se estruturas de indexação adequadas, envolvendo pesquisas nas áreas de banco de dados e estruturas de dados (NUNES; DELAMARO, 2010).

2.5.1 Métricas de desempenho de recuperação em sistemas de CBR

A maior parte dos trabalhos que envolvem CBR utilizam as métricas de **precisão** e **revocação** para avaliar o desempenho das técnicas aplicadas. Precisão refere-se à porcentagem de objetos recuperados que são relevantes à consulta. Revocação, por sua vez, refere-se à porcentagem de todos os objetos relevantes na base de dados que foram recuperados (DATTA et al., 2008).

Para uma consulta no contexto de CBR, define-se: VP, a quantidade de dados recuperados que realmente são similares ao dado modelo (verdadeiros positivos); FP, a quantidade de dados recuperados que não são similares ao dado modelo (falsos positivos); e FN, a quantidade de dados não recuperados da base, mas que são similares ao dado modelo (falsos negativos). As métricas de precisão e revocação são, então, definidas conforme apresentado nas equações 3 e 4, respectivamente.

$$\text{Precisão} = \frac{VP}{VP + FP} (\times 100\%) \quad (3)$$

$$\text{Revocação} = \frac{VP}{VP + FN} (\times 100\%) \quad (4)$$

O gráfico de **revocação versus precisão** permite analisar a variação da precisão com o aumento da revocação. O resultado ideal seria 100% de precisão para qualquer valor de revocação. A área delimitada pela curva do gráfico e pelo eixo das abscissas, a área sob a curva (AUC, do inglês *area under the curve*), é uma métrica importante para a avaliação de desempenho de abordagens de CBR. Quanto mais próximo de 1 for o valor da AUC, melhor é o desempenho da abordagem de CBR. Na figura 12 é apresentado um exemplo de gráfico de revocação *versus* precisão contendo uma curva ideal (azul), representando o melhor desempenho possível ($AUC = 1$), e uma curva realista (vermelha), representando um desempenho inferior ($AUC < 1$).

2.6 DIAGNÓSTICO AUXILIADO POR COMPUTADOR

Os sistemas de CAD consistem em sistemas computacionais com a finalidade de fornecer uma “segunda opinião” ao especialista da área médica, contribuindo na tomada de decisão a respeito de um diagnóstico (DOI, 2007). Esses sistemas constróem uma saída

Figura 12 – Representação de um gráfico de revocação *versus* precisão contendo uma curva ideal e uma curva realista.

Fonte: Vagner Mendonça Gonçalves, 2017

a partir de informações provenientes de diversas fontes, sendo que as imagens médicas de diversas modalidades são as principais delas.

Segundo Van-Ginneken et al. (2010), CAD se tornou a área de pesquisa mais ativa dentro do contexto de análise de imagens médicas. Seu objetivo é enfatizado por Doi (2006), consistindo em melhorar a acurácia do diagnóstico e a consistência da interpretação de tais dados. Hoeks et al. (2011) vão além ao afirmarem que o desenvolvimento de técnicas de CAD é necessário para se conseguir diagnósticos rápidos e reproduzíveis em grandes quantidades de dados complexos.

O início da pesquisa desta categoria de sistemas ocorreu nas décadas de 1960 e 1970, quando ainda imaginava-se construir um diagnóstico sem a presença de radiologistas. Na década de 1980, surgiram vários grupos de pesquisa no mundo com o objetivo de desenvolver sistemas de CAD para finalidades diversas, tendo principalmente as imagens provenientes de Raios-X como entrada de dados. No final da década de 90, alguns grupos avançaram no desenvolvimento, incluindo imagens de outras modalidades, como Ressonância Magnética Nuclear e Ultrassom. Nessa década, os primeiros sistemas de CAD comerciais começaram a ser aprovados nos Estados Unidos para emprego na rotina médica, como o *ImageChecker*, comercializado pela empresa *R2 Technology*, de Los Altos, Califórnia (BROWN, 2001; GIGER, 1999).

Segundo Doi (2007), no desenvolvimento de esquemas de CAD, é necessário empregar o processamento de imagens e técnicas para o processamento de informação para uma análise computacional quantitativa e, além disso, compreender a relevância médica das imagens com base nas suas características.

De forma geral, um sistema de CAD tem uma ou mais imagens médicas digitais como fonte de dados e possui algumas fases bem definidas até chegar-se ao resultado final, compreendendo:

- aquisição de imagem digital;
- pré-processamento;
- extração do sinal;
- extração de atributos;
- classificação.

As pesquisas científicas recentes têm investigado técnicas de CAD aplicáveis no auxílio ao diagnóstico de diferentes patologias como, por exemplo, câncer de mama (CHENG et al., 2010; OLIVER et al., 2010), câncer de próstata (HOEKS et al., 2011; DOYLE et al., 2012; PENG et al., 2013), câncer de pulmão (MESSAY; HARDIE; ROGERS, 2010; ARMATO-III et al., 2011) e doenças oculares (ABRAMOFF; GARVIN; SONKA, 2010; NIEMEIJER et al., 2010). É grande a diversidade de abordagens empregadas em pesquisas científicas da área, bem como a quantidade de desafios. Automatizar a análise de dados complexos, extrair informações relevantes de grandes bases de dados médicos, diminuir a taxa de erro e fornecer ferramentas eficazes e eficientes à comunidade médica são apenas alguns exemplos dos desafios que os pesquisadores da área têm pela frente.

2.7 O CORAÇÃO HUMANO

2.7.1 Sistema circulatório humano e anatomia do coração

O sistema circulatório humano é um dos mais complexos anatomicamente, sendo composto pelo sistema vascular sanguíneo e pelo sistema vascular linfático (MONTANARI, 2016). O **sistema vascular sanguíneo** tem a função de transportar o sangue pelos tecidos, levando oxigênio, nutrientes, hormônios, fatores de coagulação, células de defesa e calor (MONTANARI, 2016). Compõem o sistema vascular sanguíneo o coração, as artérias,

os capilares e as veias. Na figura 13 é apresentada uma representação esquemática do sistema vascular sanguíneo humano, destacando-se o coração, artérias (em vermelho), veias (em azul) e capilares (em branco).

Figura 13 – Esquema de circulação do sangue.

Fonte: Spalteholz e Spanner (2006)

O **sistema vascular linfático** tem a função de recolher a linfa, líquido tecidual gerado em nível dos capilares e das vênulas (pequenos vasos sanguíneos), devolvendo-a ao sangue nas grandes veias próximo ao coração. A linfa permite a circulação e o transporte de linfócitos, imunoglobulinas e lipídios (MONTANARI, 2016).

O **coração** é um órgão muscular que se contrai e relaxa ritmicamente, bombeando o sangue. Ele compreende quatro câmaras: dois átrios e dois ventrículos, cujas contrações sucessivas são responsáveis por manter a circulação do sangue. Os movimentos de contração e relaxamento dos átrios são denominadas sístole atrial e diástole atrial, respectivamente (SPALTEHOLZ; SPANNER, 2006).

Os dois ventrículos impelem o sangue através de dois troncos arteriais (aorta e tronco pulmonar) e de suas ramificações menores, artérias e arteríolas, até as redes capilares, onde se efetua a troca de substâncias. O sangue, então, retorna através de minúsculas raízes venosas, as vênulas, que confluem para veias cada vez maiores; estas, por sua vez, alcançam o átrio direito, através das duas veias cavas e, o esquerdo, pelas quatro veias pulmonares (SPALTEHOLZ; SPANNER, 2006).

Outros componentes consistem nas valvas, responsáveis por auxiliar no bombeamento do sangue não permitindo que o mesmo retorne para a estrutura anterior, e o músculo papilar, que colabora com o funcionamento das valvas (NETTER, 2011). Uma ilustração da anatomia do coração humano é apresentada na figura 14 (NETTER, 2011).

Figura 14 – Anatomia do coração humano.

Fonte: Netter (2011)

O coração é revestido por quatro camadas de tecido: endocárdio, miocárdio, epicárdio e pericárdio. O endocárdio é a camada mais interna, composta por tecido conjuntivo; fina e lisa, ela permite que o sangue circule facilmente nessa estrutura. O miocárdio é uma camada intermediária, sendo a mais espessa do coração; composta por tecido muscular, permite a contração involuntária do coração. O epicárdio é uma camada mais externa formada também por tecido conjuntivo e bastante fina. O pericárdio é a camada mais externa que reveste o coração com um tecido conjuntivo resistente e inelástico (SPENCE, 1991).

2.7.2 Ausculta cardíaca

A técnica de ausculta cardíaca é possível por meio da utilização do estetoscópio. O estetoscópio é uma ferramenta importante e eficaz para o exame físico do coração (VOIN et al., 2016). Combinado com o conhecimento fisiológico e anatômico necessário, torna-se uma ferramenta viável para a distinção entre a maior parte das doenças cardíacas (CHIZNER, 2008).

O estetoscópio conduz até o aparelho auditivo as vibrações das estruturas cardíacas e vasculares que atingem a superfície do tórax. Para a realização da ausculta cardíaca, o profissional posiciona o aparelho, em geral, em uma das regiões auscultatórias do paciente (aórtica, mitral, pulmonar, tricúspide entre outras) (PAZIN-FILHO; SCHMIDT; MACIEL, 2004). Na figura 15 é apresentada uma ilustração com indicação das regiões auscultatórias.

Os sons (ou ruídos) cardíacos básicos são transitórios, de curta duração e sua propagação até a superfície do tórax depende do local de origem e da intensidade da vibração (PAZIN-FILHO; SCHMIDT; MACIEL, 2004). Os mais importantes a serem identificados são os primeiro e segundo sons cardíacos (VOIN et al., 2016).

O primeiro som cardíaco (S1) ocorre como consequência da tensão e desaceleração abrupta da valva mitral durante o seu fechamento, que delimita o início da sístole. As valvas semilunares, por sua vez, são submetidas à tensão durante seu fechamento; essa tensão determina uma abrupta desaceleração do sangue e do movimento valvar, originando vibrações que caracterizam o segundo som cardíaco (S2) (PAZIN-FILHO; SCHMIDT; MACIEL, 2004).

Figura 15 – Regiões auscultatórias.

Fonte: Pazin-Filho, Schmidt e Maciel (2004)

S1 e S2 consistem em ruídos cardíacos naturais que ocorrem em todos os corações humanos normais. S1 marca o início da sístole ventricular, enquanto que S2 marca o início da diástole ventricular. Na figura 16 é apresentada uma representação gráfica da forma de onda do som de batimentos cardíacos com a presença dos sons S1 e S2.

Figura 16 – Forma de onda do som de batimentos cardíacos com a presença dos sons S1 e S2.

Fonte: Littmann (2016)

Outros ruídos cardíacos, geralmente associados a anomalias/doenças, também podem ser identificados por meio de auscultação. São os casos do terceiro e quarto sons cardíacos.

O terceiro som cardíaco (S3) ocorre concomitantemente com a fase de enchimento rápido ventricular do ciclo cardíaco, durante a qual ocorre a maior parte do enchimento diastólico do ventrículo. Pode ser proveniente tanto do ventrículo direito como do esquerdo. Acredita-se que ele se origine devido à súbita limitação da expansão da parede ventricular. Uma das situações clínicas nas quais esse ruído pode aparecer trata-se da presença de anormalidades estruturais de complacência e volume nos ventrículos, tal como se verifica em casos de insuficiência cardíaca (PAZIN-FILHO; SCHMIDT; MACIEL, 2004). Na figura 17 é apresentada uma representação gráfica da forma de onda do som de batimentos cardíacos com a presença dos sons S1, S2 e S3.

Figura 17 – Forma de onda do som de batimentos cardíacos com a presença dos sons S1, S2 e S3.

Fonte: Littmann (2016)

S3 pode ser audível em batimentos cardíacos de crianças e adultos jovens não representando, necessariamente, o indício de uma patologia. Existem também evidências da audibilidade e normalidade deste som em mulheres com menos de 40 anos de idade (PAZIN-FILHO; SCHMIDT; MACIEL, 2004).

O quarto som cardíaco (S4) pode estar relacionado com vibrações da parede ventricular, secundárias à expansão volumétrica dessa cavidade produzida pela contração atrial. É frequentemente detectada em situações clínicas de hipertensão arterial sistêmica ou pulmonar, estenose aórtica ou pulmonar, miocardiopatia hipertrófica e doença isquêmica

do coração (que também pode levar à insuficiência cardíaca) (PAZIN-FILHO; SCHMIDT; MACIEL, 2004). Na figura 18 é apresentada uma representação gráfica da forma de onda do som de batimentos cardíacos com a presença dos sons S1, S2 e S4.

Figura 18 – Forma de onda do som de batimentos cardíacos com a presença dos sons S1, S2 e S4.

Fonte: Littmann (2016)

2.7.3 Insuficiência cardíaca congestiva

Doenças cardiovasculares foram responsáveis por mais de 800 mil mortes nos Estados Unidos em 2014 (BENJAMIN et al., 2017). No Brasil, também em 2014, foram mais de 340 mil mortes por doenças do aparelho circulatório das quais quase 27 mil foram decorrentes de insuficiência cardíaca (MINISTÉRIO DA SAÚDE, 2017).

No gráfico apresentado na figura 19 é possível observar a evolução da quantidade de óbitos no Brasil decorrentes de doenças do aparelho circulatório entre os anos de 2005 e 2014.

A insuficiência cardíaca congestiva (ICC) é uma doença que se refere à incapacidade do coração de bombear o sangue necessário para as funções metabólicas. Ela pode surgir no estágio final de diferentes cardiopatias como, por exemplo, a hipertensão e a isquemia cardíaca.

A ICC pode ser classificada em duas variações: Esquerda ou Direita. Em outras palavras, ICC pode estar presente em apenas um lado do coração. No entanto, é comum

Figura 19 – Evolução da quantidade de óbitos no Brasil por doenças do aparelho circulatório entre os anos de 2005 e 2014.

Fonte: Vagner Mendonça Gonçalves, 2017, baseado em dados provenientes do MINISTÉRIO DA SAÚDE (2017)

que a insuficiência presente de um lado do coração resulte em um esforço excessivo do outro lado, levando a uma ICC global (BARRETTO et al., 2002).

A ICC Esquerda é a mais comum, podendo ser dividida em insuficiência sistólica ou diastólica. Na primeira, o ventrículo esquerdo está comprometido e há insuficiência do bombeamento de sangue. Na insuficiência diastólica, o ventrículo também está comprometido e o coração, que está contraído devido a sístole, não consegue relaxar, impedindo que o ventrículo receba a quantidade esperada de sangue e, consequentemente, bombeando menos sangue do que o necessário (BARRETTO et al., 2002).

2.7.4 Ressonância magnética nuclear

A ressonância magnética nuclear (RMN) é uma modalidade de exame por imagem que busca medir o magnetismo da distribuição do hidrogênio na estrutura analisada. Para tanto, o equipamento utilizado possui bobinas que captam e emitem radiação (SUETENS, 2009). Na figura 20 é apresentada uma ilustração do processo de captação de imagens por meio de RMN.

Figura 20 – Processo de captação de imagens por meio de Ressonância Magnética Nuclear.

Fonte: Imaging (2012 apud BERGAMASCO, 2013)

O exame gera uma sequência de imagens (fatias) da estrutura sob análise. O exame é preciso e sem a necessidade de preparação prévia, tendo como única restrição a impossibilidade de uso interno de objetos metálicos como implantes metálicos muito antigos ou marca-passos (SUETENS, 2009).

No contexto cardíaco, a RMN é muito precisa e permite a quantificação de estruturas cardiológicas sem a utilização de métodos invasivos. Para a realização do exame de RMN, é preciso levar em consideração os três planos anatômicos: coronal, sagital e axial. Nos planos coronal e axial é possível analisar as quatro câmaras, enquanto que, no plano sagital, é possível analisar apenas duas câmaras por imagem (BERGAMASCO, 2013).

2.8 CONSIDERAÇÕES FINAIS

No presente capítulo foram apresentados os conceitos, técnicas e ferramentas aplicados para o desenvolvimento deste trabalho de Mestrado. Esta parte da dissertação é fruto de uma análise exploratória, bem como do aprofundamento dos estudos nos conceitos

e tecnologias que foram necessários para o planejamento, para a condução e para a análise dos resultados dos experimentos.

CAPÍTULO 3 – REVISÃO SISTEMÁTICA

3.1 CONSIDERAÇÕES INICIAIS

Este capítulo apresenta a condução e os resultados de uma revisão sistemática (RS) que teve como objetivo conhecer o estado da arte com relação à aplicação de CBVR no contexto de imagens e vídeos médicos.

Além desta seção introdutória, o capítulo está organizado da seguinte forma: na seção 3.2 são apresentados conceitos sobre RS, o protocolo utilizado e o processo de condução da revisão; na seção 3.3 são apresentados e discutidos os resultados obtidos; e, por fim, na seção 3.4 são apresentadas as considerações finais.

3.2 PLANEJAMENTO E CONDUÇÃO

A RS é uma metodologia rigorosa de pesquisa bibliográfica que visa a identificar estudos primários e secundários relacionados a um determinado tema de pesquisa. Ela permite avaliar e interpretar toda a pesquisa relevante desenvolvida sobre uma questão particular ou sobre um tópico de interesse (KITCHENHAM, 2004).

Segundo Kitchenham (2004) e Biolchini et al. (2007), uma RS é conduzida em três fases bem definidas: **Planejamento de Revisão**, **Execução de Revisão** e **Análise de Resultados**. Na fase de planejamento se define um protocolo no qual são especificadas a questão de pesquisa e a metodologia que será empregada na execução da revisão. Integrarão esse protocolo os objetivos da RS, as fontes de consulta, os critérios de inclusão e de exclusão de estudos primários, as palavras-chave e outros tópicos de interesse para a pesquisa bibliográfica.

Na fase de execução, a pesquisa bibliográfica propriamente dita é realizada. É nesta etapa que ocorrem a seleção dos estudos primários, de acordo com os critérios de inclusão e de exclusão definidos, e a extração e a síntese de dados dos trabalhos incluídos. Por fim, na fase de análise de resultados, são realizadas a análise e a documentação dos resultados e das conclusões obtidas por meio deles.

Uma importante diferença entre a RS e a revisão de literatura comum é o fato de que o estabelecimento de um protocolo permite que a revisão seja passível de reprodução. Isso significa que outros pesquisadores podem reproduzir o mesmo protocolo, sendo também

capazes de avaliar os métodos empregados para o caso em questão (BIOLCHINI et al., 2007). As subseções seguintes descrevem cada uma dessas fases da RS realizada.

3.2.1 Planejamento

A RS conduzida e aqui apresentada teve como objetivo principal conhecer o estado da arte com relação à aplicação de CBVR no contexto de vídeos médicos. Foi interesse o levantamento de aplicações, técnicas utilizadas e possíveis desafios relacionados à área. Nesse contexto, foram definidas as seguintes questões de pesquisa:

- Quais são os interesses atuais na aplicação de CBVR no contexto de vídeos médicos?
- Quais são as técnicas e metodologias aplicadas para CBVR no contexto de vídeos médicos?

Uma análise exploratória sobre o tema de interesse foi previamente conduzida utilizando-se, principalmente, a ferramenta Google Acadêmico¹. Esse primeiro contato com o tema forneceu indícios de que a sua exploração pela comunidade científica é recente. Foram recuperados trabalhos que envolveram o processamento de vídeos médicos ao longo da década de 2000, porém trabalhos que abordaram CBVR foram observados apenas no final da década de 2000 e início da década de 2010. Com esse resultado, optou-se por, na condução da RS, não limitar o período de tempo para recuperação de trabalhos sobre o tema de interesse.

A análise exploratória norteou a seleção das fontes de consulta e a definição das palavras-chave utilizadas na RS. Foram consultadas as bases que tradicionalmente publicam artigos sobre o tema, de acordo com a experiência de trabalhos anteriores do grupo de pesquisa e os dados colhidos na análise exploratória preliminar.

As seguintes bases, listadas em ordem na qual foram consultadas, foram selecionadas:

- *PubMed*: base de citações e resumos de artigos científicos principalmente sobre tecnologia aplicada na área médica.
- *PubMed Central*: base de artigos científicos completos principalmente sobre tecnologia aplicada na área médica.

¹ Google Acadêmico. Disponível em: <http://scholar.google.com.br>.

- *IEEE Xplore Digital Library*: base de artigos sobre engenharia e tecnologia publicados em periódicos e eventos científicos organizados ou apoiados pelo Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE).
- *ACM Digital Library*: base de artigos científicos e citações de artigos de periódicos ou eventos científicos organizados pela *Association for Computing Machinery* (ACM).
- *Scopus*: Base de citações e resumos da literatura revisada por pares (livros, periódicos e eventos científicos) das mais diversas áreas do conhecimento.

Para a realização das consultas nas bases selecionadas foi utilizada a seguinte composição de termos:

```
("video retrieval" OR CBVR)
AND
("computer-aided diagnosis" OR "computer-assisted diagnosis" OR
"medical image" OR "medical imaging" OR surgery)
```

O protocolo utilizado na presente RS é apresentado em detalhes no Apêndice A.

Os termos deveriam estar presentes no título, no resumo, nas palavras-chave ou nos tópicos de indexação de um trabalho para que o mesmo fosse recuperado. Essas opções foram definidas por meio de ferramentas de busca avançada disponíveis nas bases.

Apenas nas bases PubMed e PubMed Central, que não disponibilizam recursos avançados nas ferramentas de consulta, foi realizada uma busca simples utilizando o termo "video retrieval". Dessa forma, foi possível a recuperação de todos os trabalhos dessas bases que mencionam o termo-chave.

Na tabela 2 são apresentadas as composições de termos utilizadas para cada uma das bases consultadas.

A fim de selecionar somente trabalhos relevantes para o tema de estudo foram definidos critérios de inclusão e exclusão. Foram incluídos somente trabalhos que atendessem ao critério de inclusão e nenhum dos critérios de exclusão. O **critério de inclusão** definido foi:

- (a) apresentar ou abordar aplicações e metodologias para o uso de Recuperação de Vídeos Baseada em Conteúdo no contexto de imagens e vídeos médicos.

Tabela 2 – Composições de termos utilizadas nas buscas.

Base	Ferramenta de Busca	String de Busca
PubMed	Search	"video retrieval"
PubMed Central	Search	"video retrieval"
IEEE	Command Search	(("Document Title":"video retrieval"OR "Abstract":"video retrieval"OR "Author Keywords":"video retrieval"OR "Index Terms":"video retrieval"OR Topic:"video retrieval"OR "Document Title":CBVR OR "Abstract":CBVR OR "Author Keywords":CBVR OR "Index Terms":CBVR OR Topic:CBVR) AND ("Document Title":"computer-aided diagnosis"OR "Abstract":"computer-aided diagnosis"OR "Author Keywords":"computer-aided diagnosis"OR "Index Terms":"computer-aided diagnosis"OR Topic:"computer-aided diagnosis"OR "Document Title":"computer-assisted diagnosis"OR "Abstract":"computer-assisted diagnosis"OR "Author Keywords":"computer-assisted diagnosis"OR "Index Terms":"computer-assisted diagnosis"OR Topic:"computer-assisted diagnosis"OR "Document Title":"medical image"OR "Abstract":"medical image"OR "Author Keywords":"medical image"OR "Index Terms":"medical image"OR Topic:"medical image"OR "Document Title":"medical imaging"OR "Abstract":"medical imaging"OR "Author Keywords":"medical imaging"OR "Index Terms":"medical imaging"OR Topic:"medical imaging"OR "Document Title":surgery OR "Abstract":surgery OR "Author Keywords":surgery OR "Index Terms":surgery OR Topic:surgery)
ACM	Advanced Search	("video retrieval"or Keywords:"video retrieval"or CBVR or Keywords:CBVR) and ("computer-aided diagnosis"or Keywords:"computer-aided diagnosis"or "computer-assisted diagnosis"or Keywords:"computer-assisted diagnosis"or "medical image"or Keywords:"medical image"or "medical imaging"or Keywords:"medical imaging"or surgery or Keywords:surgery)
Scopus	Search	("video retrieval"OR CBVR) AND ("computer-aided diagnosis"OR "computer-assisted diagnosis"OR "medical image"OR "medical imaging"OR surgery)

Fonte: Gonçalves e Nunes (2016)

Já os **critérios de exclusão** definidos foram os seguintes:

- (b) não abordar aplicações ou metodologias para o uso de Recuperação de Vídeos Baseada em Conteúdo no contexto de imagens e vídeos médicos;
- (c) ser semelhante, em conteúdo e resultados, a outro(s) trabalho(s) do(s) mesmo(s) autor(es) que foi(ram) também recuperado(s) na revisão sistemática (esse é o caso de trabalhos recuperados em mais de uma base consultada e publicações de mesmos autores com mesmo conteúdo em veículos distintos);

- (d) não estar disponível integralmente nas bases consultadas ou em alguma outra base acessível por meio de ferramentas eletrônicas como, por exemplo, o Google Acadêmico.

3.2.2 Condução

As buscas foram realizadas durante o primeiro semestre de 2014, entre os meses de março e julho. Todas as buscas foram revisadas em julho de 2014. Foram recuperados, no total, 154 trabalhos científicos distintos. Ao todo, foram incluídos apenas 8 trabalhos (5,19%). Toda a etapa de condução da RS foi devidamente documentada tomando-se como base os modelos e formulários propostos em Kitchenham (2004) e Biolchini et al. (2007).

Na tabela 3 são apresentados os números relativos às buscas realizadas em cada base consultada. Vale explicar que na base *IEEE Xplore* foram recuperados alguns resultados que não são trabalhos científicos, mas capas e sumários de livros de conferências que também são indexados na base individualmente e, por casamento com a composição de termos de busca utilizada, acabaram também sendo recuperados.

Tabela 3 – Números relativos às buscas realizadas em cada base consultada.

Base	Retornados	Duplicatas	Não são artigos	Analizados	Incluídos
PubMed	24	0	0	24	3
PubMed Central	33	3	0	30	1
IEEE	27	3	10	14	3
ACM	71	0	0	71	1
Scopus	37	22	0	15	0
TOTAL	192	28	10	154	8

Fonte: Vagner Mendonça Gonçalves, 2017

Na figura 21 é apresentado um diagrama de fluxo, baseado em Liberati et al. (2009), que sumariza a seleção de estudos realizada.

Na próxima seção são apresentados e discutidos os resultados obtidos por meio desta RS.

3.3 RESULTADOS E DISCUSSÕES

3.3.1 Trabalhos incluídos e considerações sobre os resultados

Nas tabelas 4 e 5 são apresentados os trabalhos incluídos, bem como os tópicos extraídos de cada um que são de maior interesse para a presente RS.

Figura 21 – Diagrama de fluxo sumarizando a etapa de seleção de estudos.

Fonte: Gonçalves e Nunes (2016) baseados no modelo proposto em Liberati et al. (2009)

É interessante observar que a quantidade de trabalhos recuperados é pequena frente à vasta literatura científica disponível envolvendo a aplicação de tecnologia computacional na área de auxílio ao diagnóstico. Diante desse cenário, a hipótese previamente levantada de que o conceito de CBVR foi pouco explorado no contexto médico é reforçada. Além disso, trata-se de um assunto recente dado que a maior parte dos trabalhos incluídos foram publicados a partir da década de 2000.

Tal resultado evidencia que a exploração de técnicas para o processamento, análise e recuperação de vídeos médicos é uma oportunidade de pesquisa que pode gerar contribuições significativas nos próximos anos.

3.3.2 Aplicações de interesse e modalidades de vídeos

Técnicas endoscópicas

Dos trabalhos incluídos, pode-se observar que o interesse mais frequente está na análise, descrição e recuperação de estruturas em vídeos provenientes de técnicas endoscópicas.

André et al. (2012) possuem forte interesse no auxílio ao diagnóstico em vídeos provenientes de endomicroscopia. A **endomicroscopia confocal** é uma técnica que permite a análise em tempo real, e em alta resolução (nível microscópico), do tecido epitelial do trato gastrointestinal (ANDRÉ et al., 2012b).

Esses pesquisadores têm apresentado estudos para o desenvolvimento de uma ferramenta que utiliza CBVR no auxílio ao diagnóstico do câncer colorretal. Seus trabalhos propõem diferentes objetivos que juntos tendem a desenvolver tal ferramenta. Os trabalhos citados têm explorado técnicas de recuperação de vídeos baseadas em características visuais (ANDRÉ et al., 2011b), combinação entre recuperação baseada em conteúdo e anotação semântica (ANDRÉ et al., 2012), e uma abordagem que utiliza recuperação de vídeos de endomicroscopia para a classificação de pólipos colorretais (ANDRÉ et al., 2012b).

Todos esses trabalhos envolvem a extração de características de baixo nível para a descrição dos vídeos. No entanto, uma preocupação dos pesquisadores é o **gap semântico**, isto é, a diferença entre o significado atribuído a um vídeo por meio de suas características de baixo nível (cores, texturas, formas, movimentos) e o real significado da situação ali representada. Todas as abordagens apresentadas propõem formas de agregar algum critério semântico à CBVR para diminuir o problema citado.

Chattopadhyay et al. (2008) também apresentaram uma abordagem de recuperação baseada em conteúdo aplicada em vídeos provenientes de uma técnica endoscópica, a **laparoscopia**. Laparoscopia é uma técnica cirúrgica minimamente invasiva (SCHUESSLER et al., 1997). É aplicada para visualizar e operar estruturas e órgãos intra-abdominais por meio de incisão e introdução de um dispositivo chamado laparoscópio. O interesse dos pesquisadores foi recuperar os quadros do vídeo que contêm o instrumento utilizado no procedimento e representado em uma imagem fornecida como referência. Segundo os pesquisadores, tal tarefa é de interesse para que os fabricantes dos instrumentos empregados

no procedimento possam analisar como ocorre a utilização dos seus produtos, evitando-se ter que varrer todo o vídeo para encontrar os trechos de interesse.

No trabalho de Cao et al. (2004) é apresentada uma abordagem baseada em recuperação por conteúdo para a detecção de intervenções de terapia ou biópsia em vídeos provenientes de **colonoscopia**. A colonoscopia é uma técnica endoscópica que permite a inspeção de todo o cólon e a realização de operações terapêuticas como a remoção de pólipos (CAO et al., 2004). Os objetivos da abordagem vão desde facilitar a posterior análise de complicações devido a procedimentos realizados por meio da colonoscopia, até o desenvolvimento de um sistema de CBVR capaz de apoiar educação e pesquisa endoscópica.

Vídeos de monitoramento cirúrgico

Quelleg et al. (2014) tiveram como objetivo o desenvolvimento de uma ferramenta capaz de pesquisar em tempo real sequências de vídeos similares a uma sequência de vídeo gravada durante um procedimento de cirurgia oftálmica. Foram investigados pelos pesquisadores os casos específicos de cirurgia da membrana epirretiniana e cirurgia de catarata. A técnica envolve a identificação da tarefa cirúrgica que está sendo executada no segmento de vídeo processado. Com essa abordagem, o cirurgião poderá ser alertado e auxiliado na tomada de decisão frente a uma situação atípica ou de risco durante o procedimento.

O *gap* semântico também foi uma preocupação de Quelleg et al. (2014). Os pesquisadores utilizaram uma abordagem de ponderação de características para melhorar a correlação entre as características de baixo nível e conceitos semânticos relacionados às tarefas cirúrgicas passíveis de identificação.

No trabalho de Fan et al. (2007) é apresentada uma abordagem visando a anotação automática de vídeos de monitoramento de cirurgias voltados para a educação médica. Trata-se de uma abordagem mista de CBVR e recuperação baseada em palavras-chave, na qual a descrição do conteúdo dos vídeos por meio de extratores é utilizada na atribuição de rótulos aos mesmos. O usuário, por sua vez, utiliza palavras-chave que remetem aos rótulos atribuídos para buscar os vídeos de interesse.

Ecocardiografia

Uma abordagem para a recuperação de vídeos provenientes de exames de ecocardiografia foi apresentada por Syeda-Mahmood, Beymer e Amir (2009). Segundo os pesquisadores, essa modalidade de vídeos é uma importante fonte de informação para o auxílio ao diagnóstico cardíaco, sendo capaz de retratar forma e movimentos do coração em diferentes ângulos.

A técnica de extração de características apresentada pelos pesquisadores é, no entanto, bem diferente das demais analisadas. São extraídas características obtidas a partir da interpretação de textos destacados ao longo do segmento de vídeo. Os textos indicam medições realizadas durante o exame. Para tanto, é utilizado um motor de reconhecimento ótico de caracteres (OCR, do inglês *optical character recognition*). As medidas identificadas a partir dos textos são utilizadas na composição dos vetores de características, que, por sua vez, permitem a medição da similaridade entre os vídeos.

3.3.3 Métodos e técnicas utilizados nos trabalhos identificados

Os trabalhos recuperados e incluídos nesta RS mostram diferentes abordagens de extração de características para descrição do conteúdo dos vídeos, além de abordagens para o treinamento de classificadores para a obtenção de modelos úteis na tarefa de recuperação de vídeos semelhantes e, também, diferentes abordagens de recuperação propriamente dita.

André et al. (2011b) utilizaram abordagens para a definição de assinaturas visuais para a descrição de vídeos de endomicroscopia confocal. Trata-se de uma adaptação do método apresentado originalmente por Sivic e Zisserman (2006), denominado *Bag-of-Visual-Words*.

Para o cômputo dessas assinaturas são utilizadas abordagens de divisão dos quadros do vídeo em regiões e agrupamento dessas regiões de forma a identificar conceitos representados nas imagens (ANDRÉ et al., 2011b). A descrição desses conceitos utilizando-se ferramental matemático e estatístico permite a comparação entre as assinaturas de vídeos distintos.

Refletindo-se sobre essa abordagem de identificação de regiões atreladas a conceitos em imagens ou vídeos, pode-se perceber que a mesma é de fato interessante para trabalhos

com o objetivo de diagnóstico, pois pode auxiliar na identificação de anomalias na imagem ou vídeo explorado.

No trabalho que envolve vídeos de monitoramento de cirurgias oftálmicas, de Quellec et al. (2014), verifica-se o emprego de características de cor e textura, extraídas da transformada *Wavelet*, e de movimento, extraídas utilizando-se o conceito de fluxo óptico. No caso do monitoramento cirúrgico, características descritivas de movimento são bem interessantes, talvez até mais que as características tradicionais de cor, forma e textura. Isso porque descrever movimentos em vídeos dessa modalidade pode ser muito útil na identificação de ações que ocorreram no procedimento filmado e que merecem a atenção do especialista ou cirurgião.

A maior parte dos trabalhos incluídos apresentaram abordagens para aumentar a relevância semântica dos resultados de recuperação, diminuindo, assim, o problema do *gap* semântico. Para essas abordagens, em geral, foram utilizadas técnicas de classificação, tais como *k-Nearest Neighbors* e máquinas de vetores de suporte, para a obtenção de modelos semanticamente eficientes utilizando-se como dados de treinamento vídeos rotulados por especialistas das áreas.

3.4 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados os resultados de uma RS que possibilitou o levantamento e a análise do estado da arte com relação à aplicação de CBVR no contexto de vídeos médicos. Foram analisados oito trabalhos que desenvolveram técnicas e abordagens em tal contexto. Os trabalhos foram recuperados por meio de consultas sistemáticas em cinco importantes bases de trabalhos científicos.

Os sistemas de informação têm, cada vez mais, incluído recursos multimídia buscando a concretização de diversos objetivos: resolver problemas complexos, descobrir informação em dados complexos, melhorar a interação com os usuários, entre outros. Muito se deve ao grande crescimento dos dados digitais disponíveis, fruto do avanço das tecnologias de comunicação e informação.

No contexto médico não é diferente. É grande a quantidade de dados digitais disponíveis e que possuem grande potencial para apoiar o desenvolvimento de tecnologias capazes de auxiliar os profissionais de saúde em tarefas como análise e interpretação,

auxílio ao diagnóstico e auxílio à educação médica. Para tanto, métodos eficientes para descrição, armazenamento e recuperação desses dados são necessários.

Todos os pesquisadores cujos trabalhos foram incluídos nesta RS até o momento acreditam no crescimento das aplicações e na capacidade que a área de recuperação baseada em conteúdo possuiu na área médica. Dispõem-se atualmente de grande ferramental matemático e estatístico para o processamento e interpretação de imagens e vídeos digitais, podendo-se empregar CBVR em diferentes modalidades de imagens para as mais diversas tarefas, quer seja no ramo do CAD, quer seja no monitoramento cirúrgico ou, ainda, na área de educação médica.

No que concerne aos desafios da área, os principais apontados pelos pesquisadores são: a redução do *gap* semântico nos diferentes níveis das aplicações de CBVR (modelagem dos dados, descrição, classificação e recuperação) e a seleção e redução de características para a descrição eficiente dos vídeos (lidar com o “problema da dimensionalidade”).

Apesar de o escopo da presente revisão estar delimitado na área médica, é importante mencionar que a aplicação das técnicas de CBVR tem potencial para auxiliar na resolução de problemas nas mais diversas áreas do conhecimento. Podemos citar como exemplos problemas nas áreas de segurança (identificação de pessoas ou objetos de interesse em vídeos), usabilidade de sistemas de informação (identificação de erros em sequências de ações do usuário, avaliação da usabilidade por meio de filmes retratando usuários interagindo com o sistema) ou até mesmo no estudo dos fenômenos naturais representados em vídeos.

Desse modo, a investigação das tendências de pesquisa nas diferentes áreas é importante e auxilia os pesquisadores, estimulando-os na investigação de problemas e na proposição de soluções. O trabalho de mestrado proposto pretende contribuir nesse sentido, tendo como foco a área médica, com sua vasta gama de problemas a serem investigados.

Os resultados da presente RS permitem concluir que a aplicação do conceito de CBVR na área médica é recente e foi pouco explorada. Logo, o processamento, a análise e a recuperação de vídeos médicos constituem uma oportunidade de pesquisa que pode gerar contribuições significativas nos próximos anos.

As aplicações de interesse nos trabalhos analisados consistem, em geral, na identificação e descrição de estruturas representadas nos vídeos para uma posterior indexação e recuperação. Os objetivos mais comuns são auxílio ao diagnóstico, auxílio na análise de

procedimentos, auxílio as cirurgiões na tomada de decisão e disponibilização de conteúdo facilmente recuperável para fins de auxílio à educação médica.

Tabela 4 – Trabalhos incluídos e principais tópicos de interesse para a revisão sistemática - parte 1/2.

Referência	Modalidade de Vídeo	Área do Corpo / Procedimento	Aplicação	Extração de Características	Abordagem de Recuperação
Quellec et al. (2014)	Vídeos de monitoramento cirúrgico	Olhos / Cirurgias de catarata e de membrana epirretiniana	Sistema para análise de sequências de vídeos e identificação de tarefas cirúrgicas em tempo-real.	Textura e cor por meio de análise Wavelet; movimento por meio de análise do fluxo óptico entre frames.	Consulta dos vizinhos mais próximos utilizando uma variação de k-d tree.
André et al. (2012b)	Endomicroscopia confocal	Região colorretal	Classificação de pólipos colorretais.	Assinatura de características visuais obtidas por meio de uma adaptação do método Bag-of-Visual-Words.	Extração das características; medição da similaridade entre os objetos descritos; treinamento de um classificador k-nearest neighbors.
André et al. (2012)	Endomicroscopia confocal	Região colorretal	Sistema de recuperação de vídeos que combina características visuais com conceitos semânticos, a fim de reduzir o <i>gap</i> .	Assinatura de características visuais obtidas por meio de uma adaptação do método Bag-of-Visual-Words.	Extração das características visuais; utilização de abordagem probabilística (Fisher) para estimar, dado um vetor de características visuais, se um dado conceito semântico está presente no segmento de vídeo.
André et al. (2011b)	Endomicroscopia confocal	Região colorretal	Sistema (atlas) de vídeos de endomicroscopia com suporte à recuperação por conteúdo visual.	Assinatura de características visuais obtidas por meio de uma adaptação do método Bag-of-Visual-Words.	Extração das características visuais; utilização de um classificador k-nearest neighbors para identificação dos vídeos mais semelhantes.
Syeda-Mahmood, Beyner e Amir (2009)	Ecocardiografia	Coração	Recuperação por meio da extração de características obtidas a partir de textos escritos no vídeo. Os textos indicam medições realizadas durante o exame.	Deteção dos frames somente textuais; remoção de qualquer conteúdo não-textual; aplicação de um motor OCR para extrair palavras nas regiões textuais; utilização das medidas identificadas dos textos para compor vetores de características.	Comparação por similaridade entre o vetor de um vídeo modelo e os vetores dos vídeos na base.

Fonte: Gonçalves e Nunes (2016)

Tabela 5 – Trabalhos incluídos e principais tópicos de interesse para a revisão sistemática - parte 2/2.

Referência	Modalidade de Vídeo	Área do Corpo / Procedimento	Aplicação	Extração de Características	Abordagem de Recuperação
Chattopadhyay et al. (2008)	Vídeos de procedimentos de laparoscopia	Região intra-abdominal / Laparoscopia	Abordagem de recuperação por conteúdo na qual, por meio de uma imagem de consulta, são recuperados os frames do vídeo que contém o instrumento representado na imagem modelo.	Extração de características por meio de histograma de cor, Transformada de Hough (linearidade da borda) e nível de compacidade do objeto de interesse na imagem.	Medição da similaridade entre as características da imagem modelo e as características de cada frame. Um limiar definido indica quando a similaridade é tal que o frame deve ser recuperado.
Fan et al. (2007)	Vídeos utilizados em educação médica	Cirurgias diversas gravadas	Anotação automática de vídeos voltados para a educação médica.	Identificação de objetos ao longo dos frames dos vídeos; caracterização dos objetos detectados por meio de características de baixo nível (cor e textura).	Utilização de <i>Support Vector Machine</i> para classificar os objetos identificados de forma hierárquica; atribuição de conceitos semânticos aos vídeos de acordo com o resultado da classificação; usuário busca por termos-chave que remetem às classes/conceitos anotados(os).
Cao et al. (2004)	Colonoscopia	Região colorretal	Abordagem baseada em recuperação por conteúdo para a detecção de intervenções de terapia ou biópsia no vídeo.	Identificação automatizada das intervenções ao longo do vídeo por meio da identificação de instrumentos utilizados pelo médico (descritos por características de forma baseadas em descritores Fourier); identificação de palavras que definem a parte do cólon em análise a as ações realizadas no exame (utilizam software que, dado o segmento de áudio, gera o texto pronunciado no discurso).	Seleção por parte do usuários de uma imagem única da intervenção detectada, uma cena selecionada da intervenção detectada ou de um trecho de discurso identificado no momento da intervenção. A escolha do usuário remete ao trecho do vídeo que contém a intervenção selecionada por meio de um dos três parâmetros.

Fonte: Gonçalves e Nunes (2016)

CAPÍTULO 4 – METODOLOGIA

4.1 CONSIDERAÇÕES INICIAIS

Este capítulo apresenta a metodologia empregada para a condução deste projeto de pesquisa. Na seção 4.2 são sumarizadas as fases envolvidas, desde a definição do tema, passando pelo estudo teórico e pela condução de experimentos, até a análise e sumarização de resultados. A organização utilizada para a execução das fases ao longo do projeto também é apresentada.

Na seção 4.3 são detalhados os materiais e métodos aplicados para a condução de experimentos visando a testar e a analisar os resultados obtidos com a aplicação da abordagem proposta de combinação entre características visuais e sonoras em vídeos provenientes de exames cardiológicos.

4.2 FASES DO PROJETO DE PESQUISA

O presente projeto de pesquisa foi conduzido de acordo com as fases apresentadas na figura 22 e detalhadas nos tópicos a seguir.

1. Definição do tema

A área de atuação e o tema inicial do trabalho foram definidos no primeiro semestre de curso. Inicialmente, foram realizadas diversas discussões entre orientador e orientando com relação às áreas de pesquisa desenvolvidas no Laboratório de Aplicações de Informática em Saúde (LApIS), no qual este trabalho foi conduzido.

Uma primeira análise exploratória sobre CBIR e CBVR foi realizada a fim de se levantar possíveis temas de interesse. Foram identificados trabalhos que começaram a investigar a composição entre características visuais e sonoras para melhorar a recuperação de dados multimídia por conteúdo. Sendo assim, levantou-se a hipótese de determinada técnica ter contribuição significativa na área de computação aplicada em saúde.

Figura 22 – Fases do projeto de pesquisa apresentado nesta dissertação.

Fonte: Vagner Mendonça Gonçalves, 2017

2. Análise exploratória

Foi realizada uma análise exploratória mais específica em busca de trabalhos que contribuíram com aplicações de CBVR no contexto de imagens e vídeos médicos. Não foram encontrados muitos trabalhos, sendo que os encontrados trabalharam, em geral, com o processamento de vídeos provenientes de exames de endoscopia digestiva.

Com o tema do trabalho em amadurecimento, o estudo teórico sobre diversos conceitos e técnicas úteis ao contexto do trabalho foi aprofundado. Todo o referencial teórico conceitual levantado nesta fase foi apresentado no capítulo 2.

3. Revisão sistemática

Uma revisão bibliográfica sistemática foi conduzida, definindo-se critérios controlados para a realização das buscas com base no conhecimento prévio obtido nas análises exploratórias. O objetivo foi levantar o estado da arte com relação à aplicação de CBVR

no contexto de imagens e vídeos médicos e, assim, compreender em quais lacunas da área de aplicação a abordagem poderia ser aplicada.

A revisão sistemática foi conduzida utilizando-se quatro importantes bases científicas das áreas de computação e tecnologia aplicada em saúde: *PudMed*, *IEEE Xplore Digital Library*, *ACM Digital Library* e *Scopus*. Os resultados da revisão evidenciaram que a aplicação do conceito de CBVR na área médica é recente e foi pouco explorada. Portanto, o processamento, a análise e a recuperação de vídeos médicos constituem uma oportunidade de pesquisa que podem gerar contribuições significativas. O detalhamento do planejamento e da condução da RS, bem como a análise e discussão dos resultados, podem ser conferidos no capítulo 3.

4. Definição da questão de pesquisa e dos objetivos

Após a realização de um estudo teórico aprofundado, bem como uma análise sistemática dos trabalhos correlatos, foi possível consolidar a questão de pesquisa e os objetivos do presente trabalho de mestrado, conforme detalhado no capítulo 1, seção 1.3.

5. Definição dos experimentos

Os experimentos conduzidos foram definidos visando a responder a questão de pesquisa enunciada e a atingir os objetivos propostos para o projeto de pesquisa. Inicialmente, foram conduzidos experimentos em uma base de vídeos sintéticos para fins de validação da abordagem proposta. Os vídeos sintéticos foram construídos combinando-se diferentes características de região, cor, movimento e som a fim de se testar e analisar, de forma geral, os resultados da combinação entre características visuais e sonoras em aplicações de CBVR. Esta base de vídeos e os experimentos conduzidos foram descritos em detalhes em um artigo científico que está em processo de submissão a veículo científico. A versão mais recente do artigo mencionado está disponível no apêndice B.

Em seguida, foi definido um estudo de caso visando a aplicar a abordagem de CBVR descrita em vídeos provenientes da área médica a fim de se testar e analisar a viabilidade da abordagem e o impacto da combinação entre características visuais e sonoras no desempenho de recuperação. A base de vídeos utilizada foi construída por meio de

dados provenientes de exames de RMN em conjunto com sinais sonoros provenientes de auscultação de sons do coração. Os resultados obtidos, bem como a análise e a discussão dos mesmos, são apresentados no capítulo 5.

6. Definição da metodologia de testes

Com os experimentos e o estudo de caso definidos, foi também definida uma metodologia de execução de experimentos e análise de resultados. A seção 4.3 deste capítulo apresenta os materiais e métodos aplicados na condução do estudo de caso com vídeos provenientes de exames cardiológicos.

7. Implementação

Foram executadas implementação e adaptação de extratores de características para a condução dos experimentos. Referente a características visuais foram implementados seis extratores (área, perímetro, largura, altura e coordenadas espaciais do centro de massa) e adaptado um extrator (assinatura de forma) do trabalho de Delamaro, Nunes e Oliveira (2013). Referente a características sonoras, foram utilizados três extratores: um extrator de fonemas da Língua Portuguesa, apresentado no trabalho de Oliveira et al. (2014); um extrator relacionado à taxa média de passagens pelo zero; e um extrator relacionado à frequência, obtida por meio do centróide espectral. Os dois últimos disponíveis no *framework jAudio* (MCENNIS et al., 2005).

Rotinas para a automatização dos experimentos também foram implementadas, incluindo-se leitura e decodificação de vídeos, imagens e arquivos de áudio; processamento dos dados decodificados com a utilização dos extratores de características; consultas por similaridade, utilizando-se os vetores de características construídos, bem como a distância Euclidiana; e cálculo das métricas de revocação e precisão para cada consulta.

A linguagem de programação **Java** foi utilizada por meio da distribuição **OpenJDK 7**. Os experimentos foram executados em ambiente **Linux (Ubuntu 14.04)** em um computador com 16GB de memória RAM e processador **Intel® Core™i5 vPro™**.

8. Execução de experimentos

Os experimentos foram executados por meio dos procedimentos metodológicos descritos na seção 4.3. Os resultados das diversas consultas realizadas foram sintetizados em termos de precisão e revocação (ver seção 2.5.1). Além disso, os gráficos de revocação *versus* precisão também foram plotados e as áreas sob a curva de cada consulta calculadas.

9. Análise e sumarização de resultados

Os resultados obtidos por meio da presente pesquisa foram analisados, discutidos e compilados nesta dissertação de mestrado. Além disso, artigos para publicação em veículos científicos da área foram escritos.

Até o momento foi publicado um artigo sobre a revisão sistemática realizada (GONÇALVES; NUNES, 2016) e está em processo de submissão o artigo apresentado no apêndice B.

4.3 MATERIAIS E MÉTODOS EMPREGADOS PARA A CONDUÇÃO DE EXPERIMENTOS

Nesta seção são apresentadas a metodologia empregada na condução dos experimentos referentes ao estudo de caso definido, bem como a base de dados, os extratores de características e a função de similaridade utilizados.

4.3.1 Base de vídeos

Uma base contendo 30 vídeos foi construída com a utilização de 30 conjuntos de exames provenientes de RMN do coração de diferentes pacientes, bem como sons provenientes de auscultação dos sons do coração.

Os exames de RMN foram fornecidos pelo Instituto do Coração da Faculdade de Medicina da Universidade de São Paulo (InCor) já com os respectivos diagnósticos, sendo 53% com presença de ICC e 47% com ausência de anomalias. 77% dos pacientes possuíam mais que 40 anos. Cada exame possui cerca de 45 fatias obtidas durante a fase de diástole.

Cada vídeo possui uma duração de três segundos e reproduz sequencialmente as fatias de um exame. As imagens utilizadas para a composição dos vídeos estão em escala de cinza, com resolução espacial de 256x256 pixels e resolução de contraste de 16 bits.

O som integrado a cada vídeo consiste na reprodução de um sinal proveniente de auscultação do coração cujas características são relacionadas à presença ou à ausência de ICC. Os sons foram obtidos do portal da empresa *3M™ Littmann® Stethoscopes* (LITTMANN, 2016), que os disponibiliza em seu portal para fins de treinamento em auscultação.

Na tabela 6 são apresentadas as características fisiológicas associadas aos sons utilizados. Sendo a maioria dos casos de pacientes com mais de 40 anos, assumiu-se a audibilidade de S3 nos batimentos cardíacos como indício da presença de ICC no paciente.

Tabela 6 – Características fisiológicas associadas aos sons do coração utilizados.

Som	Características fisiológicas	Diagnóstico relacionado
1	S1 e S2	Ausência de ICC
2	S1 e S2	Ausência de ICC
3	S2	Ausência de ICC
4	S3	Presença de ICC
5	S4	Presença de ICC
6	S3 e S4	Presença de ICC

Fonte: Vagner Mendonça Gonçalves, 2017

Os sons utilizados foram integrados aos casos provenientes de exames de RMN para a composição da base de vídeos. Cada caso foi associado a um som que tenha relação com o diagnóstico (ausência ou presença de ICC) nele representado. Os sons 1, 2 e 3 foram associados homogeneamente e aleatoriamente aos casos com ausência de ICC. Já os sons 4, 5 e 6 foram associados homogeneamente e aleatoriamente aos casos com presença de ICC.

4.3.2 Segmentação do ventrículo esquerdo

Para a condução dos experimentos de recuperação baseada em conteúdo, as características visuais foram extraídas do ventrículo esquerdo. A segmentação dos quadros processados levou em consideração o contorno do endocárdio, camada interna do ventrículo.

A segmentação foi realizada de forma manual por meio do *software Seg3D* (CBIC, 2012), conforme metodologia detalhada por Bergamasco (2013, p. 93). Na figura 23 é apresentado um exemplo de quadro retratando a fatia original com a marcação dos contornos do epicárdio (em verde) e do endocárdio (em vermelho).

Figura 23 – Exemplo de quadro representando uma fatia do exame de RMN de um paciente com diagnóstico de ICC.

Fonte: InCor, com marcações provenientes da metodologia de segmentação descrita em Bergamasco (2013, p. 93)

Na figura e 24 é apresentado um exemplo retratando a imagem do quadro da figura 23 segmentada por meio da metodologia referenciada.

4.3.3 Extratores de características

Para a condução dos experimentos referentes ao estudo de caso foram selecionadas características de forma e características sonoras. As características de forma extraídas fornecem dados sobre a regularidade da borda delimitada pelo contorno do endocárdio, sendo capazes de indicar deformações anormais do ventrículo durante a fase da diástole.

As características sonoras selecionadas, por sua vez, fornecem dados relativos à forma de onda e à frequência do sinal sonoro processado, sendo capazes de ressaltar as particularidades de cada som cardíaco (S1, S2, S3 ou S4) presente no sinal.

Figura 24 – Exemplo de imagem retratando o ventrículo esquerdo segmentado por meio da marcação do endocárdio.

Fonte: Vagner Mendonça Gonçalves, 2017, baseado na segmentação resultante da metodologia descrita em Bergamasco (2013, p. 93)

Assinatura de forma

O extrator de características de assinatura de forma utilizado neste trabalho foi adaptado da abordagem apresentada por Delamaro, Nunes e Oliveira (2013). Ele mede a regularidade da borda da região de interesse representada em uma imagem. Para tanto, é necessário, previamente, a execução da tarefa de segmentação para se isolar tal região.

O extrator de característica, primeiramente, calcula o centro de massa da região de interesse. Em seguida, toma intervalos em ângulos iguais calculando, para cada ângulo, a distância em pixels do centro de massa da região de interesse até a borda da mesma. Ao final desta execução, obtém-se, portanto, um vetor de distâncias do centro de massa até a borda. Por fim, o extrator calcula e retorna o desvio padrão com base nos valores do vetor de distâncias.

Na figura 25 é apresentada uma ilustração da rotina de cálculo do vetor de distâncias para obtenção da assinatura de forma (GONÇALVES, 2011). As linhas representam as distâncias entre o centro de massa e a borda da região de interesse. Esse exemplo mostra as distâncias calculadas no primeiro quadrante da região de interesse, considerando-se um intervalo de ângulos de dez graus.

Figura 25 – Ilustração da rotina de cálculo do vetor de distâncias para obtenção da assinatura de forma.

Fonte: Gonçalves (2011)

O extrator de características de assinatura de forma foi utilizado para se obter a variação da regularidade da borda da região segmentada do ventrículo esquerdo, delimitada pelo endocárdio, ao longo do vídeo. Para tanto, foram selecionados sete quadros de cada vídeo e, para cada um dos quadros, calculada a assinatura de forma da região de interesse.

Considerando-se as sete assinaturas obtidas para cada vídeo, as características utilizadas para se descrever cada um deles foram:

- assinatura de forma mínima;
- assinatura de forma máxima;
- média das assinaturas de forma;
- desvio padrão das assinaturas de forma.

Características sonoras

As características sonoras extraídas do sinal integrado a cada vídeo foram:

- taxa média de passagens pelo zero;
- frequência mais forte calculada por meio do centróide espectral.

A **passagem pelo zero** (em inglês, *zero crossing*) é uma característica básica do sinal sonoro, pertencente ao domínio temporal. Ela se caracteriza quando o sinal (no domínio do tempo) intercepta o eixo das abscissas. A **taxa de passagem pelo zero** é calculada por meio da contagem da quantidade de passagens pelo zero do sinal em uma determinada janela (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010).

O **centróide espectral** é uma aproximação comumente utilizada de **brilhância**. **Brilhância** caracteriza a distribuição espectral de frequências e descreve se um sinal é dominado por baixas ou altas frequências (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010). Trata-se de uma propriedade do domínio da frequência.

Nos experimentos conduzidos foram utilizadas as implementações disponíveis no *framework jAudio* (MCENNIS et al., 2005).

4.3.4 Função de similaridade

Utilizou-se a Distância Euclidiana (equação 5) como função de similaridade para se realizar as comparações entre os vetores de características. Conforme já discutido, essa distância é muito utilizada nos trabalhos que envolvem CBR (NUNES et al., 2015) e o valor zero calculado entre dois vetores de características representa máxima similaridade entre eles.

$$Dis_{L_2}(A, B) = \sqrt{\sum_{i=0}^{n-1} (a_i - b_i)^2} \quad (5)$$

É importante ressaltar que o objetivo principal dos experimentos conduzidos não contempla a obtenção de alto desempenho de recuperação com a arquitetura utilizada. Por esse motivo, optou-se por se utilizar neste estudo inicial uma arquitetura mais simples de CBR, composta por extratores e uma função de similaridade. Em trabalhos futuros, a

abordagem poderá ser testada e validada em arquiteturas mais complexas envolvendo, por exemplo, redes neurais ou probabilísticas.

4.3.5 Metodologia para a execução das consultas

Os experimentos conduzidos consistiram, primeiramente, na execução de consultas utilizando-se apenas as características de forma para descrever os vídeos. Em seguida, as mesmas consultas foram executadas utilizando-se as características de forma em conjunto com as características sonoras. Na figura 26 é apresentado o modelo de consultas por similaridade aplicado no presente trabalho.

Figura 26 – Modelo de consultas por similaridade aplicado para a condução dos experimentos.

Fonte: Vagner Mendonça Gonçalves, 2017

Para a discussão apresentada no capítulo 5, referente aos resultados obtidos com o estudo de caso, foram selecionadas cinco consultas em cada experimento utilizando-se um vídeo modelo representando um caso com presença de ICC e cinco consultas utilizando-se um vídeo modelo representando um caso com ausência de ICC. Para cada vídeo utilizado como modelo, considerou-se como resultados corretos de recuperação todo vídeo da base que represente o mesmo diagnóstico. Em outras palavras, para cada caso utilizado como modelo cujo diagnóstico é a presença de ICC, os resultados de recuperação

considerados corretos são os demais casos da base que também possuem o diagnóstico de ICC. Já para cada caso utilizado como modelo cujo diagnóstico é a ausência de ICC, os resultados de recuperação considerados corretos são os demais casos também possuem o diagnóstico de ausência de ICC.

Os resultados dos experimentos foram avaliados utilizando-se as métricas de revocação e precisão (subseção 2.5.1). Para cada resultado de consulta, traçamos o gráfico de revocação *versus* precisão. Para fins de comparação, uma curva de revocação *versus* precisão é melhor do que outra quando sua AUC é maior. Isto significa que, em uma curva melhor, a precisão diminui menos quando a revocação aumenta.

Experimentos com consultas utilizando-se apenas as características sonoras também foram executados. No entanto, dada a natureza controlada da base de vídeos utilizada no estudo de caso, foram obtidos nestes experimentos resultados praticamente ideais de recuperação. Os gráficos apresentados nas figuras 27 e 28 apresentam as curvas de revocação *versus* precisão para as consultas envolvendo vídeos modelos de casos normais e vídeos modelos com diagnóstico de ICC, respectivamente, utilizando-se apenas as características sonoras para compor os vetores de características.

Figura 27 – Gráfico de revocação *versus* precisão das consultas que utilizaram apenas os extratores sonoros para recuperar casos com ausência de ICC.

Fonte: Vagner Mendonça Gonçalves, 2017

É possível observar que, com exceção da consulta 4 dos casos de ICC, todas as demais consultas apresentam curva ideal no gráfico. A Consulta 4 também apresenta

Figura 28 – Gráfico de revocação *versus* precisão das consultas que utilizaram apenas os extratores de forma para recuperar casos com presença de ICC.

Fonte: Vagner Mendonça Gonçalves, 2017

um bom resultado de recuperação, mesmo não sendo o ideal. Isto poderia significar que somente as características sonoras já seriam suficientes para a aplicação de CBVR na base de vídeos utilizada. No entanto, não é possível tecer tal conclusão utilizando-se apenas os experimentos apresentados, conduzidos sobre uma base de vídeos construída de forma controlada.

De qualquer forma, um dos objetivos deste estudo é validar a hipótese de que, acrescentando as características sonoras ao vetor de características que já possuiu características visuais em sua composição, é possível obter melhores resultados de recuperação. A confirmação desta hipótese é relevante e pode ser útil em diferentes aplicações de CBVR, em especial, na área médica, em casos que o diagnóstico pode ser suportado tanto por características visuais obtidas em exames de imagens, como por características sonoras obtidas dos sons gerados pelo corpo humano.

4.4 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados os materiais e métodos empregados para a condução do trabalho referente a esta pesquisa de Mestrado. Foram sumarizadas todas as fases envolvidas no trabalho, desde a definição do tema, passando pelo estudo teórico e

pela condução de experimentos, até a análise e sumarização de resultados. A metodologia empregada para a condução de experimentos em um estudo de caso envolvendo vídeos provenientes de exames cardiológicos foi descrita em detalhes.

No próximo capítulo os resultados do estudo de caso conduzido são apresentados, detalhados e discutidos. Tais resultados permitiram responder a questão de pesquisa enunciada no capítulo introdutório e embasaram as conclusões do trabalho, descritas no capítulo 6

CAPÍTULO 5 – RESULTADOS EXPERIMENTAIS COM EXAMES CARDIOLÓGICOS

5.1 CONSIDERAÇÕES INICIAIS

Neste capítulo são apresentados e discutidos resultados experimentais obtidos com a utilização de uma base de vídeos construída com imagens e sons provenientes de exames cardiológicos. Os objetivos específicos dos experimentos conduzidos se referem à validação da abordagem de combinação entre características visuais e sonoras para melhorar os resultados de recuperação em um contexto real (prático) e ao estudo da viabilidade de aplicação da abordagem no contexto médico.

A metodologia empregada na condução dos experimentos, a base de dados, os extratores e a função de similaridade utilizados foram detalhados na seção 4.3; na seção 5.2 são apresentados e discutidos os resultados obtidos; e, por fim, na seção 5.3 são apresentadas as considerações finais referentes aos experimentos e seus resultados.

5.2 RESULTADOS E DISCUSSÕES

No primeiro experimento, considerou-se apenas as características de forma para compor os vetores de características de cada vídeo. Na figura 29 é apresentado o gráfico de revocação *versus* precisão obtido para cada consulta que utilizou como modelo um caso de paciente com ausência de ICC (denominado **normal**, neste contexto). Já na figura 30 é apresentado o gráfico de revocação *versus* precisão obtido para cada consulta utilizando como modelo um caso de paciente com presença de ICC.

Por meio da análise dos gráficos dos experimentos que utilizaram apenas as características de forma, observa-se que, para valores de revocação mais baixos (menores que 0,6, para os casos normais, e menores que 0,3, para os casos de ICC) existem consultas com bons valores de precisão e consultas com valores de precisão baixos, não sendo possível identificar um desempenho homogêneo. No entanto, para valores de revocação maiores, em ambos os casos, o desempenho das consultas se aproximam, variando na faixa aproximada de 40% a 60%.

Com a inclusão das características sonoras na composição dos vetores de características, combinado-as com as características de forma, foram obtidos os resultados

Figura 29 – Gráfico de revocação *versus* precisão das consultas que utilizaram apenas as características de forma para recuperar casos com ausência de ICC.

Fonte: Vagner Mendonça Gonçalves, 2017

Figura 30 – Gráfico de revocação *versus* precisão das consultas que utilizaram apenas as características de forma para recuperar casos com presença de ICC.

Fonte: Vagner Mendonça Gonçalves, 2017

apresentados na figura 31, para consultas que utilizaram um caso normal como modelo, e na figura 32, para consultas que utilizaram um caso de ICC como modelo. Por meio dos gráficos, observa-se que todas as consultas resultaram em uma melhora significativa do desempenho de recuperação. Esse resultado demonstra a exequibilidade da abordagem,

mostrando que as características sonoras tiveram influência positiva e determinante na obtenção de melhores resultados.

Figura 31 – Gráfico de revocação *versus* precisão das consultas que utilizaram combinação entre características de forma e de som para recuperar casos com ausência de ICC.

Fonte: Vagner Mendonça Gonçalves, 2017

Figura 32 – Gráfico de revocação *versus* precisão das consultas que utilizaram combinação entre características de forma e de som para recuperar casos com presença de ICC.

Fonte: Vagner Mendonça Gonçalves, 2017

Nas tabelas 7 e 8 são apresentadas os valores de AUC calculados referentes aos resultados de recuperação obtidos nos experimentos envolvendo características de forma e som. Os valores corroboram o resultado inferido por meio dos gráficos, mostrando que a inclusão das características sonoras melhorou o desempenho de recuperação para todas as consultas.

Tabela 7 – Valores de AUC calculados para as consultas que visaram a recuperação de casos com ausência de ICC.

	Consulta 1	Consulta 2	Consulta 3	Consulta 4	Consulta 5	AUC média
Forma	0,49	0,50	0,68	0,66	0,57	0,58
Forma e som	0,92	1,00	0,99	1,00	0,83	0,95

Fonte: Vagner Mendonça Gonçalves, 2017

Tabela 8 – Valores de AUC calculados para as consultas que visaram a recuperação de casos com presença de ICC.

	Consulta 1	Consulta 2	Consulta 3	Consulta 4	Consulta 5	AUC média
Forma	0,53	0,47	0,65	0,51	0,50	0,53
Forma e som	0,98	0,99	0,96	0,71	0,98	0,92

Fonte: Vagner Mendonça Gonçalves, 2017

Os resultados mostraram que, para todas as consultas realizadas, as características sonoras melhoraram a precisão de recuperação. Nos experimentos executados, obtiveram-se valores de precisão altos (maiores que 80%) para valores de revocação também altos (até, aproximadamente, 95%) na maioria das consultas quando se combinou as características de forma com as características sonoras para a composição dos vetores.

Considerando-se o valor médio de AUC, obteve-se um desempenho 63,8% maior quando se utilizou características de forma e de som em conjunto para se recuperar casos normais, em comparação com a utilização apenas dos extratores de forma. Para a recuperação de casos de ICC, o aumento do desempenho foi de 73,6%.

Os resultados alcançados no presente estudo de caso são importantes e mostraram que, para o contexto e o escopo deste trabalho de mestrado a questão de pesquisa levantada pode ser respondida positivamente. Certamente, esse é um estudo experimental inicial, porém seus resultados mostraram que vídeos médicos com informação sonora relevante podem ser analisados e recuperados por meio de abordagens mais eficientes quando se utiliza também características sonoras em sua descrição.

5.3 CONSIDERAÇÕES FINAIS

No presente capítulo foram apresentados os resultados de experimentos conduzidos em uma base de vídeos construída por meio de dados provenientes de exames cardiológicos de RMN e auscultação. Foram combinadas características de forma da região segmentada, representando o ventrículo esquerdo, bem como características extraídas dos sons do coração.

O objetivo foi mostrar a viabilidade da aplicação da análise de características sonoras em conjunto com características visuais para viabilizar recuperação de vídeos médicos baseada em conteúdo, bem como a obtenção de melhor desempenho de recuperação, comparando-se com a utilização de características visuais apenas.

Os resultados obtidos com os experimentos foram promissores, deixando indícios que permitem responder positivamente à questão de pesquisa dentro do escopo investigado neste trabalho de mestrado.

CAPÍTULO 6 – CONCLUSÕES

A recuperação de vídeos baseada em conteúdo permite o processamento e a análise do conteúdo de vídeos digitais visando à extração de informações relevantes que viabilizem indexação e recuperação. No contexto médico, CBVR pode ser aplicada para diversas finalidades como, por exemplo, diagnóstico auxiliado por computador, suporte à tomada de decisão e disponibilização de bases de vídeos para treinamento e educação médica.

O problema de pesquisa abordado no presente trabalho de mestrado se refere à investigação e à aplicação de métodos e técnicas visando a melhorar o desempenho de recuperação em sistemas de CBVR, bem como à investigação de técnicas de CBVR para viabilizar CAD.

A abordagem investigada consistiu na análise combinada entre características sonoras e características visuais para viabilizar ganhos com relação ao desempenho de recuperação visando a responder a questão de pesquisa enunciada.

O objetivo principal do trabalho consistiu em aplicar e avaliar extratores de características sonoras em conjunto com extratores de características visuais para viabilizar recuperação de vídeos médicos baseada em conteúdo e, então, inferir se a abordagem resultou em ganhos com relação ao desempenho de recuperação quando comparada à utilização apenas das características visuais.

O estudo teórico conduzido permitiu identificar que poucos são os trabalhos que consideraram características sonoras para a descrição de vídeos no contexto de CBVR. Tal resultado indica a relevância do trabalho conduzido, tornando-o contribuição significativa para a área.

Resultados aplicados ao domínio dos vídeos médicos, conforme se propôs o presente trabalho, foram obtidos utilizando-se uma base construída por meio de dados provenientes de exames cardiológicos de ressonância magnética nuclear e auscultação. Foram combinadas características de forma da região segmentada, representando o ventrículo esquerdo, bem como características extraídas dos sons do coração. Esses resultados foram promissores no que se refere ao aumento do desempenho de recuperação com a combinação entre características visuais e sonoras.

A questão de pesquisa enunciada para o presente trabalho de mestrado foi a seguinte: “O emprego de características sonoras, em adição a características visuais, pode

melhorar o nível de descrição de vídeos e, conseqüentemente, melhorar os resultados de recuperação por conteúdo em vídeos no contexto médico?”. Os resultados obtidos com a condução dos experimentos mostraram que, em todos os cenários, as características de som proporcionaram maior precisão durante a recuperação. Com base nestes resultados, a questão pode ser respondida positivamente dentro do escopo investigado.

Nos experimentos conduzidos utilizando-se a base de vídeos cardiológicos, considerando-se os valores médios de AUC, obteve-se um desempenho 63,8% maior quando se utilizou características de forma e de som em conjunto para se recuperar casos normais, em comparação com a utilização apenas dos extratores de forma. Para a recuperação de casos de ICC, o aumento do desempenho foi de 73,6%.

Uma limitação desta pesquisa consiste na execução de experimentos apenas em uma base de vídeos construída de forma controlada. Para a construção desta base de vídeos provenientes de exames cardiológicos, foi necessária a vinculação entre o exame de imagem e um sinal de áudio relacionado que, não necessariamente, é proveniente do mesmo coração.

Outras limitações da pesquisa consistem na condução de experimentos em apenas um estudo de caso aplicado à área da saúde e na aplicação de uma única arquitetura de CBVR para a execução das consultas.

Uma limitação da abordagem de CBVR investigada consiste na complexidade de implementação de extratores de características sonoras. O processamento de sinais sonoros é uma tarefa complexa e relacionar características desses sinais com as especificidades do caso a ser tratado também não é trivial, exigindo um conhecimento profundo do problema. Contudo, com o avanço do desenvolvimento de estudos nas áreas de processamento de sinais e CBR, é possível encontrar na literatura extratores que podem ser utilizados em diferentes aplicações e que foram disponibilizados por seus pesquisadores. É o caso do trabalho de Oliveira et al. (2014), que desenvolveram o extrator de fonemas utilizado em experimentos apresentados no artigo científico do apêndice B, bem como um *framework*, que utiliza conceitos de CBR para automatização de testes de programas com saídas complexas, que contém diversos extratores de características implementados e disponibilizados.

Melhorar os resultados de recuperação promovendo eficiência à abordagem também é um desafio no contexto de CBR. Selecionar características adequadas ao problema; evitar características correlacionadas (que implicam em processamento adicional desnecessário); e selecionar abordagens eficientes de medição de similaridade por conteúdo entre dados

multimídia são desafios complexos da área. Nunes et al. (2015), por exemplo, apresentaram um estudo comparativo de funções de similaridade com aplicações em CBIR. Eles contribuem com uma discussão com relação aos impactos da seleção de diferentes funções de similaridade para a tarefa de comparação entre vetores de características.

Trabalhos como o de Nunes et al. (2015), bem como este trabalho de mestrado contribuem com o desenvolvimento de novas abordagens visando a melhorar cada vez mais o desempenhos de aplicações de CBR, viabilizando a expansão de sua adoção em diferentes aplicações.

6.1 TRABALHOS FUTUROS

São sugestões de trabalhos futuros dentro do contexto abordado neste projeto de pesquisa:

- conduzir mais experimentos para confirmar os resultados obtidos com o experimento de vídeos médicos;
- estudar e realizar experimentos visando a investigar a viabilidade técnica e clínica da abordagem para aplicações em outros domínios de diagnóstico (endoscopia, ultrassonografia, ecocardiografia);
- aplicar arquiteturas mais sofisticadas de CBVR para a realização de consultas por similaridade, utilizando-se, por exemplo, técnicas de inteligência artificial, modelos probabilísticos e realimentação de relevância;
- estudar, investigar e propor extratores de características sonoras específicos para diferentes modalidades de vídeos médicos;
- empregar ou desenvolver métodos para verificação da correlação entre características a fim de usar somente aqueles que garantem seletividade;
- implementar um protótipo de sistema de CBVR para vídeos médicos utilizando características visuais e sonoras.

REFERÊNCIAS¹

- ABDULLAH, L. N. A framework for evaluating human action detection via multidimensional approach. In: *Proceedings of the Sixth International Conference on Computer Graphics, Imaging and Visualization*. [S.l.: s.n.], 2009. p. 186–190. Citado 2 vezes nas páginas 15 e 18.
- ABRAMOFF, M. D.; GARVIN, M. K.; SONKA, M. Retinal imaging and image analysis. *IEEE Reviews in Biomedical Engineering*, v. 3, p. 169–208, 2010. ISSN 1937-3333. Citado na página 36.
- ANDRÉ, B. et al. Learning semantic and visual similarity for endomicroscopy video retrieval. *IEEE Transactions on Medical Imaging*, v. 31, n. 6, p. 1276–1288, 2012. Citado 2 vezes nas páginas 52 e 58.
- ANDRÉ, B. et al. Retrieval evaluation and distance learning from perceived similarity between endomicroscopy videos. In: FICHTINGER, G.; MARTEL, A.; PETERS, T. (Ed.). *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2011*. [S.l.]: Springer Berlin Heidelberg, 2011, (Lecture Notes in Computer Science, v. 6893). p. 297–304. Citado na página 17.
- ANDRÉ, B. et al. A smart atlas for endomicroscopy using automated video retrieval. *Medical Image Analysis*, v. 15, n. 4, p. 460–476, 2011. Citado 4 vezes nas páginas 16, 52, 54 e 58.
- ANDRÉ, B. et al. Software for automated classification of probe-based confocal laser endomicroscopy videos of colorectal polyps. *World Journal of Gastroenterology*, v. 18, n. 39, p. 5560–5569, 2012b. Citado 2 vezes nas páginas 52 e 58.
- ARMATO-III, S. G. et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Medical Physics*, American Association of Physicists in Medicine, v. 38, n. 2, p. 915–931, 2011. ISSN 2473-4209. Disponível em: <http://dx.doi.org/10.1118/1.3528204>. Citado na página 36.
- BARRETTO, A. C. P. et al. Revisão das II diretrizes da Sociedade Brasileira de Cardiologia para o diagnóstico e tratamento da insuficiência cardíaca. *Arquivos Brasileiros de Cardiologia*, Scielo, v. 79, p. 1–30, 2002. Citado na página 43.
- BENJAMIN, E. J. et al. Heart disease and stroke statistics—2017 update: A report from the american heart association. *Circulation*, American Heart Association, Inc., 2017. ISSN 0009-7322. Disponível em: <http://circ.ahajournals.org/content/early/2017/01/25/CIR.0000000000000485>. Citado na página 42.
- BERGAMASCO, L. C. C. *Recuperação de imagens cardíacas tridimensionais por conteúdo*. Dissertação (Dissertação) — Universidade de São Paulo, São Paulo, 2013. Citado 4 vezes nas páginas 44, 65, 66 e 67.
- BIOLCHINI, J. C. d. A. et al. Scientific research ontology to support systematic review in software engineering. *Advanced Engineering Informatics*, v. 21, n. 2, p. 133–151, 2007. Citado 3 vezes nas páginas 46, 47 e 50.

¹ DE ACORDO COM A ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 6023.

BOSI, M.; GOLDBERG, R. E. *Introduction to Digital Audio Coding and Standards*. New York: Springer Science+Business Media, 2003. Citado na página 27.

BOVIK, A. C. Introduction to digital image and video processing. In: BOVIK, A. (Ed.). *Handbook of Image and Video Processing*. [S.l.]: Academic Press, 2000. p. 3–17. Citado 4 vezes nas páginas 20, 22, 29 e 30.

BROWN, D. G. The evaluation of computer-aided diagnosis systems: an FDA perspective. In: *Proceedings of the 30th Applied Imagery Pattern Recognition Workshop*. [S.l.: s.n.], 2001. p. 17–20. Citado na página 35.

BUGATTI, P. H. *Análise da Influência de Funções de Distância para o Processamento de Consultas por Similaridade em Recuperação de Imagens por Conteúdo*. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) — Instituto de Ciências Matemáticas e de Computação – USP, São Carlos, 2008. Citado 2 vezes nas páginas 32 e 33.

BUGATTI, P. H.; TRAINA, A. J. M.; TRAINA-JR, C. Assessing the best integration between distance-function and image-feature to answer similarity queries. In: *Proceedings of the 2008 ACM Symposium on Applied Computing*. [S.l.]: ACM, 2008. p. 1225–1230. Citado na página 32.

CAO, Y. et al. Parsing and browsing tools for colonoscopy videos. In: *Proceedings of the 12th Annual ACM International Conference on Multimedia*. Nova Iorque: ACM, 2004. p. 844–851. Citado 2 vezes nas páginas 53 e 59.

CBIC. Seg3D: volumetric image segmentation and visualization. Scientific Computing and Imaging Institute (SCI). Disponível em <http://www.seg3d.org>. 2012. Citado na página 65.

CHATTOPADHYAY, T. et al. An application for retrieval of frames from a laparoscopic surgical video based on image of query instrument. In: *Proceedings of the TENCON - IEEE Region 10 Conference*. [S.l.: s.n.], 2008. p. 1–5. Citado 2 vezes nas páginas 52 e 59.

CHENG, H. et al. Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognition*, v. 43, n. 1, p. 299–317, 2010. ISSN 0031-3203. Disponível em: <http://dx.doi.org/10.1016/j.patcog.2009.05.012>. Citado na página 36.

CHIZNER, M. A. Cardiac auscultation: Rediscovering the lost art. *Current Problems in Cardiology*, v. 33, n. 7, p. 326–408, jul. 2008. Citado na página 39.

DATTA, R. et al. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, ACM, Nova Iorque, v. 40, n. 2, p. 5:1–5:60, mai. 2008. Citado 3 vezes nas páginas 15, 25 e 34.

DELAMARO, M. E.; NUNES, F. L. S.; OLIVEIRA, R. A. P. Using concepts of content-based image retrieval to implement graphical testing oracles. *Software Testing, Verification and Reliability*, John Wiley e Sons, v. 23, n. 3, p. 171–198, 2013. Citado 3 vezes nas páginas 26, 63 e 67.

DEMENTHON, D.; DOERMANN, D. Video retrieval using spatio-temporal descriptors. In: *Proceedings of the eleventh ACM international conference on Multimedia*. [S.l.: s.n.], 2003. p. 508–517. Citado 2 vezes nas páginas 15 e 16.

DOI, K. Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology. *Physics in Medicine and Biology*, v. 51, n. 13, p. R5–R27, 2006. Citado na página 35.

DOI, K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, v. 31, n. 4-5, p. 198–211, 2007. Citado 3 vezes nas páginas 16, 34 e 36.

DOYLE, S. et al. A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. *IEEE Transactions on Biomedical Engineering*, v. 59, n. 5, p. 1205–1218, mai 2012. ISSN 0018-9294. Citado na página 36.

FAN, J. et al. Incorporating concept ontology for hierarchical video classification, annotation, and visualization. *IEEE Transactions on Multimedia*, v. 9, n. 5, p. 939–957, ago. 2007. Citado 2 vezes nas páginas 53 e 59.

GIGER, M. L. Overview of computer-aided diagnosis in breast imaging. In: DOI, K. et al. (Ed.). *Computer-Aided Diagnosis in Medical Imaging*. 1. ed. [S.l.]: Elsevier Science BV, 1999, (International Congress Series, v. 1182). p. 167–176. Citado na página 35.

GIL, V. N. et al. Extratores para oráculos de teste de sistemas texto-fala utilizando recuperação de áudio baseada em conteúdo. *Revista Eletrônica de Iniciação Científica*, n/a, n. n/a, p. n/a–n/a, 2014. Artigo aceito para publicação. Citado 2 vezes nas páginas 27 e 28.

GIROD, B.; WETZSTEIN, G. Digital image processing -introduction 1. Notas de Aula, EE368/CS232 - Digital Image Processing, Stanford University. 2016. Citado na página 23.

GONÇALVES, V. M. Avaliação de funções de similaridade em sistemas de CBIR. Relatório de Iniciação Científica apresentado à FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo. 2011. Citado na página 68.

GONÇALVES, V. M.; NUNES, F. L. S. A systematic review on content-based medical video retrieval. *Journal of Health Informatics*, v. 8, p. 799–808, 2016. Suplemento I – XV Congresso Brasileiro de Informática em Saúde - CBIS 2016. Citado 5 vezes nas páginas 49, 51, 58, 59 e 64.

GONZALEZ, R. C.; WOODS, R. E. *Processamento de Imagens Digitais*. 1. ed. São Paulo: Blucher, 2000. Citado 3 vezes nas páginas 24, 25 e 32.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. 3. ed. [S.l.]: Pearson, 2008. Citado 5 vezes nas páginas 20, 21, 22, 23 e 24.

GRIERSON, H. J.; CORNEY, J. R.; HATCHER, G. D. Using visual representations for the searching and browsing of large, complex, multimedia data sets. *International Journal of Information Management*, v. 35, n. 2, p. 244–252, 2015. Citado na página 15.

HAN, J. et al. Representing and retrieving video shots in human-centric brain imaging space. *IEEE Transactions on Image Processing*, v. 22, n. 7, p. 2723–2736, jul. 2013. Citado na página 17.

HEWITT, P. G. *Física conceitual*. Porto Alegre: Bookman, 2002. Citado 2 vezes nas páginas 26 e 28.

HOEKS, C. M. A. et al. Prostate cancer: Multiparametric mr imaging for detection, localization, and staging. *Radiology*, v. 261, n. 1, p. 46–66, 2011. PMID: 21931141. Disponível em: <http://dx.doi.org/10.1148/radiol.11091822>. Citado 2 vezes nas páginas 35 e 36.

IMAGING, B. How MRI works. 2012. Citado na página 44.

KITCHENHAM, B. A. *Procedures for Performing Systematic Reviews*. [S.l.], 2004. Citado 2 vezes nas páginas 46 e 50.

LI, Z.; HOU, K.; LI, H. Similarity measurement based on trigonometric function distance. In: *Proceedings of the First International Symposium on Pervasive Computing and Applications*. Ürümqi, China: [s.n.], 2006. p. 227–231. Citado na página 32.

LIBERATI, A. et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine*, Public Library of Science, v. 6, n. 7, p. e1000100–1–e1000100–28, jul. 2009. Citado 2 vezes nas páginas 50 e 51.

LITTMANN, S. Littmann heart and lung sound library. Disponível em: <http://www.littmann.ca/>. Acesso em: 10 de junho de 2016. 2016. Citado 4 vezes nas páginas 40, 41, 42 e 65.

MCENNIS, D. et al. jAudio: a feature extraction library. In: *Proceedings of the International Conference on Music Information Retrieval*. [S.l.: s.n.], 2005. p. 600–603. Citado 2 vezes nas páginas 63 e 69.

MESSAY, T.; HARDIE, R. C.; ROGERS, S. K. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Medical Image Analysis*, v. 14, n. 3, p. 390–406, 2010. ISSN 1361-8415. Disponível em: <http://dx.doi.org/10.1016/j.media.2010.02.004>. Citado na página 36.

MINISTÉRIO DA SAÚDE. DATASUS: Mortalidade - Brasil (em 2014), pela CID (I00 a I99). Disponível em: <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/obt10uf.def>. Acesso em: 26 de janeiro de 2017. 2017. Citado 2 vezes nas páginas 42 e 43.

MITROVIĆ, D.; ZEPPELZAUER, M.; BREITENEDER, C. Chapter 3 - features for content-based audio retrieval. In: ZELKOWITZ, M. V. (Ed.). *Advances in Computers: Improving the Web*. [S.l.]: Elsevier, 2010, (Advances in Computers, v. 78). p. 71–150. Citado 4 vezes nas páginas 27, 28, 29 e 69.

MONTANARI, T. *Histologia: texto, atlas e roteiro de aulas práticas*. 3. ed. Porto Alegre: Edição da Autora, 2016. Disponível em <http://www.ufrgs.br/livrodehisto>. Acesso em 21/09/2016. Citado 2 vezes nas páginas 36 e 38.

MURGEL, E. *Fundamentos de acústica ambiental*. São Paulo: Editora Senac São Paulo, 2007. Citado na página 26.

NETTER, F. H. *Atlas de Anatomia Humana*. 5. ed. Rio de Janeiro: Elsevier, 2011. Citado na página 38.

NIEMEIJER, M. et al. Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs. *IEEE Transactions on Medical Imaging*, v. 29, n. 1, p. 185–195, jan 2010. ISSN 0278-0062. Citado na página 36.

NUNES, F. L. S.; DELAMARO, M. E. Recuperação de imagens baseada em conteúdo e sua aplicação na área de saúde. In: FERNANDES, A. M. R.; WANGHAM, M. S. (Ed.). *Computer on the Beach 2010: Livro de Minicursos*. Florianópolis: Universidade do Vale do Itajaí, 2010. v. 1, p. 116–144. Citado 2 vezes nas páginas 31 e 33.

NUNES, F. L. S. et al. CBIR based testing oracles: An experimental evaluation of similarity functions. *International Journal of Software Engineering and Knowledge Engineering*, v. 25, n. 08, p. 1271–1306, 2015. Citado 3 vezes nas páginas 26, 69 e 81.

NUSSENZVEIG, H. M. *Curso de física básica, 2*. São Paulo: Edgard Blücher, 2004. Citado 2 vezes nas páginas 26 e 28.

OLIVEIRA, R. A. P. et al. An extensible framework to implement test oracle for non-testable programs. In: *Proceedings of the 26th International Conference on Software Engineering and Knowledge Engineering*. [S.l.: s.n.], 2014. p. 199–204. Citado 2 vezes nas páginas 63 e 80.

OLIVER, A. et al. A review of automatic mass detection and segmentation in mammographic images. *Medical Image Analysis*, v. 14, n. 2, p. 87–110, 2010. ISSN 1361-8415. Disponível em: <http://dx.doi.org/10.1016/j.media.2009.12.005>. Citado na página 36.

PAZIN-FILHO, A.; SCHMIDT, A.; MACIEL, B. C. Ausculta cardíaca: Bases fisiológicas-fisiopatológicas. *Medicina*, Ribeirão Preto, v. 37, p. 208–226, jul./dez. 2004. Citado 4 vezes nas páginas 39, 40, 41 e 42.

PEDRINI, H.; SCHWARTZ, W. R. *Análise de Imagens Digitais: princípios, algoritmos e aplicações*. São Paulo: Thomson Learning, 2008. Citado na página 22.

PENG, Y. et al. Quantitative analysis of multiparametric prostate mr images: Differentiation between prostate cancer and normal tissue and correlation with gleason score—a computer-aided diagnosis development study. *Radiology*, v. 267, n. 3, p. 787–796, 2013. Disponível em: <http://dx.doi.org/10.1148/radiol.13121454>. Citado na página 36.

PRATT, W. K. *Digital Image Processing: PIKS Scientific Inside*. 4. ed. [S.l.]: Wiley-Interscience, 2007. Citado na página 22.

QUELLEC, G. et al. Real-time recognition of surgical tasks in eye surgery videos. *Medical Image Analysis*, v. 18, n. 3, p. 579–590, 2014. Citado 5 vezes nas páginas 16, 17, 53, 55 e 58.

SANTOS, T. T.; MORIMOTO, C. H. Segmentação, indexação e recuperação de vídeo utilizando OpenCV. In: *Proceedings of the Fourth Workshop de Visão Computacional*. [S.l.: s.n.], 2008. p. 1–21. Citado 2 vezes nas páginas 29 e 30.

SCHUESSLER, W. W. et al. Laparoscopic radical prostatectomy: Initial short-term experience. *Urology*, v. 50, n. 6, p. 854–857, 1997. Citado na página 52.

SERGYÁN, S. Color histogram features based image classification in content-based image retrieval systems. In: *Proceedings of the 6th International Symposium on Applied Machine Intelligence and Informatics*. [S.l.: s.n.], 2008. p. 221–224. Citado na página 26.

- SIVIC, J.; ZISSERMAN, A. Video Google: Efficient visual search of videos. In: PONCE, J. et al. (Ed.). *Toward Category-Level Object Recognition*. [S.l.]: Springer Berlin Heidelberg, 2006, (Lecture Notes in Computer Science, v. 4170). p. 127–144. Citado na página 54.
- SPALTEHOLZ, W.; SPANNER, R. *Atlas de anatomia humana*. São Paulo: Roca, 2006. Citado 2 vezes nas páginas 37 e 38.
- SPENCE, A. P. *Anatomia Humana Básica*. 2. ed. São Paulo: Manole, 1991. Citado na página 39.
- SU, J.-H. et al. Effective content-based video retrieval using pattern-indexing and matching techniques. *Expert Systems with Applications*, v. 37, n. 7, p. 5068–5085, 2010. Citado na página 16.
- SUETENS, P. *Fundamentals of Medical Imaging*. 2. ed. New York: Combridge University Press, 2009. Citado 2 vezes nas páginas 43 e 44.
- SYEDA-MAHMOOD, T.; BEYMER, D.; AMIR, A. Disease-specific extraction of text from cardiac echo videos for decision support. In: *Proceedings of the 10th International Conference on Document Analysis and Recognition*. [S.l.: s.n.], 2009. p. 1290–1294. Citado 2 vezes nas páginas 54 e 58.
- VAN-GINNEKEN, B. et al. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. *Medical Image Analysis*, v. 14, n. 6, p. 707–722, 2010. Citado na página 35.
- VOIN, V. et al. Auscultation of the heart: The basics with anatomical correlation. *Clinical Anatomy*, p. n/a–n/a, 2016. Disponível em: <http://dx.doi.org/10.1002/ca.22780>. Citado na página 39.
- WANG, X.-Y.; ZHANG, B.-B.; YANG, H.-Y. Content-based image retrieval by integrating color and texture features. *Multimedia Tools and Applications*, v. 68, n. 3, p. 545–569, 2014. Citado na página 26.
- WOLD, E. et al. Content-based classification, search, and retrieval of audio. *IEEE MultiMedia*, IEEE Computer Society Press, Los Alamitos, CA, USA, v. 3, n. 3, p. 27–36, set 1996. Citado na página 28.
- ZHANG, H. J. et al. An integrated system for content-based video retrieval and browsing. *Pattern Recognition*, v. 30, n. 4, p. 643–658, 1997. Citado na página 16.

APÊNDICE A – PROTOCOLO DE REVISÃO SISTEMÁTICA

A.1 OBJETIVO

Conhecer o estado da arte com relação ao tema “aplicação de recuperação de vídeos baseada em conteúdo no contexto de vídeos médicos”. Pretende-se levantar aplicações, técnicas utilizadas e possíveis desafios relacionados à área.

A.2 QUESTÕES DE PESQUISA

- Quais os interesses atuais na aplicação de CBVR no contexto de vídeos médicos?
- Quais as técnicas e metodologias aplicadas para CBVR no contexto de vídeos médicos?

População: Trabalhos e projetos de pesquisa que explorem/apresentem técnicas e sistemas de diagnóstico auxiliado por computador.

Intervenção: Aplicações de recuperação de vídeos baseada em conteúdo no contexto de vídeos médicos; features e algoritmos utilizados neste tipo de tarefa.

Resultados Esperados: Entender o contexto atual da aplicação de recuperação de vídeos baseada em conteúdo no contexto de vídeos médicos, procurando destacar as principais abordagens utilizadas para a execução da tarefa e os desafios da área.

Aplicação: Pesquisadores interessados em análise automatizada (auxiliada por computador) de vídeos provenientes de exames médicos e em sistemas de auxílio ao diagnóstico.

A.3 CRITÉRIOS PARA A SELEÇÃO DAS FONTES

As fontes deverão estar disponíveis na internet, preferencialmente em bases de dados científicas reconhecidas na área de computação e suas aplicações.

A.4 MÉTODO DE BUSCA

Por meio de palavras-chave previamente definidas, buscas por artigos de periódicos ou anais de eventos científicos deverão ser realizadas nas bases de dados selecionadas. Serão considerados, preferencialmente, trabalhos publicados em inglês.

A.5 CRITÉRIOS DE QUALIDADE DOS ESTUDOS PRIMÁRIOS

Os seguintes critérios de qualidade (em ordem de relevância) serão aplicados nos trabalhos recuperados para determinar a relevância dos mesmos:

1. ser publicados em periódico ou anais de eventos científicos com revisão por pares;
2. apresentar e detalhar abordagens de recuperação de vídeos baseada em conteúdo no contexto de vídeos médicos e as técnicas utilizadas;
3. apresentar e detalhar experimentos e resultados, incluindo bases e quantidade de dados utilizados, métricas e critérios de avaliação.

A.6 BASES CIENTÍFICAS SELECIONADAS

- PubMed
- PubMed Central
- IEEE Xplore Digital Library
- ACM Digital Library
- Scopus

A.7 PALAVRAS-CHAVE

Para a realização das buscas, procurar-se-á utilizar a composição de termos apresentada abaixo de forma que tais termos-chave estejam presentes ou no título ou no resumo ou nas palavras-chave do trabalho indexado na base.

```

("video retrieval" OR CBVR)

AND

("computer-aided diagnosis" OR "computer-assisted diagnosis" OR
  "medical image" OR "medical imaging" OR surgery)

```

O objetivo é identificar trabalhos relevantes para a revisão proposta por meio de uma string de busca que seja reproduzível em todas as bases científicas consultadas.

A.8 CRITÉRIOS DE INCLUSÃO DE ESTUDOS PRIMÁRIOS

Serão incluídos trabalhos que atendem ao menos um dos critérios abaixo listados:

- (a) apresentar ou abordar aplicações e metodologias para o uso de Recuperação de Vídeos Baseada em Conteúdo no contexto de imagens e vídeos médicos.

A.9 CRITÉRIOS DE EXCLUSÃO DE ESTUDOS PRIMÁRIOS

Serão excluídos trabalhos:

- (b) que não abordam aplicações ou metodologias para o uso de Recuperação de Vídeos Baseada em Conteúdo no contexto de imagens e vídeos médicos;
- (c) semelhantes, em conteúdo e resultados, a outro(s) trabalho(s) do(s) mesmo(s) autor(es) que foi(ram) também recuperado(s) na revisão sistemática (esse é o caso de trabalhos recuperados em mais de uma base consultada e publicações de mesmos autores com mesmo conteúdo em veículos distintos);
- (d) não disponíveis integralmente nas bases consultadas ou em alguma outra base acessível por meio de ferramentas eletrônicas como, por exemplo, o Google Acadêmico.

A.10 PROCESSO DE SELEÇÃO DOS TRABALHOS

Os estudos primários recuperados serão avaliados de uma forma geral por meio dos critérios de inclusão e exclusão definidos. Desse modo, uma pré-avaliação (por meio do título e do resumo) será realizada a fim de se selecionar os trabalhos de interesse que

serão incluídos na revisão. Os textos selecionados na pré-avaliação deverão ser lidos na íntegra e avaliados rigorosamente de acordo com os mesmos critérios. Nessa etapa, possíveis trabalhos que não se mostrem interessantes, de acordo com os objetivos da revisão, ainda poderão ser excluídos. Os trabalhos de interesse serão, então, utilizados na elaboração de discussões que objetivam responder as questões definidas.

A.11 ESTRATÉGIA DE EXTRAÇÃO DE INFORMAÇÕES RELEVANTES

Serão extraídas informações relevantes tais como: modalidade de vídeo explorada no trabalho, área do corpo humano, base(s) de vídeos utilizada, aplicação discutida, classificação do(s) extrator(es) (cor, forma, textura etc) utilizados, método(s) de extração de características e de recuperação empregado(s), tipo(s) e quantidade(s) de vídeos utilizados nos testes, resultados dos testes, tendências indicadas pelos autores, trabalhos futuros e limitações.

A.12 SUMARIZAÇÃO DOS RESULTADOS

Com os resultados obtidos, deverá ser redigido um artigo que descreverá os resultados e conclusões obtidos por meio da revisão sistemática. Análises qualitativas e quantitativas, com relação aos trabalhos incluídos e suas conclusões, também deverão ser apresentadas no relatório.

**APÊNDICE B – ARTIGO CIENTÍFICO PRODUZIDO BASEADO EM
RESULTADOS DE EXPERIMENTOS EM UMA BASE DE VÍDEOS
SINTÉTICOS**

Combinação entre características de região, cor, movimento e som para melhorar recuperação de vídeos baseada em conteúdo

Vagner M. Gonçalves e Fátima L. S. Nunes

Laboratório de Aplicações de Informática em Saúde, Escola de Artes, Ciências e Humanidades

Universidade de São Paulo

São Paulo, Brasil

vagner.goncalves@usp.br, fatima.nunes@usp.br

Resumo—A evolução dos dispositivos de armazenamento e das redes de computadores contribuiu para que os vídeos digitais assumissem um importante papel no desenvolvimento de sistemas de informação multimídia. Se faz importante e necessário, portanto, o desenvolvimento de técnicas eficazes e eficientes de recuperação de vídeos em diferentes bases. Nosso objetivo neste trabalho consistiu em verificar se a combinação entre extratores de características sonoras de segmentos de vídeos e extratores de características visuais, relacionadas a região, cor e movimento de objetos representados, produzem maior precisão em recuperação de vídeos baseada em conteúdo quando comparada à utilização apenas de extratores de características visuais. Nós conduzimos um estudo de caso com a aplicação de um extrator de fonemas em segmentos de vídeo que retratam figuras geométricas em movimento e a reprodução de uma palavra da Língua Portuguesa. Os resultados dos experimentos indicaram uma melhora significativa na precisão da recuperação com a aplicação do extrator de fonemas em diferentes combinações com as demais características extraídas dos segmentos de vídeos.

Palavras-chave—recuperação de vídeos baseada em conteúdo; CBVR; características sonoras; processamento de vídeos

I. INTRODUÇÃO

A evolução dos dispositivos de armazenamento e das redes de computadores contribuíram para que os vídeos digitais assumissem um importante papel no desenvolvimento de sistemas de informação multimídia. Diferentes modalidades de vídeos são empregadas em áreas como educação, cultura, segurança, entretenimento e medicina.

Com a finalidade de aproveitar todo o potencial dos dados multimídia, tais como os vídeos digitais, no desenvolvimento de sistemas de informação são necessárias a investigação e a utilização de técnicas automatizadas eficientes para análise, interpretação e recuperação desses dados. No entanto, a interação com dados multimídia não é trivial [1]. Técnicas como a recuperação baseada em conteúdo (CBR, do inglês *content-based retrieval*), que envolve a extração e a comparação entre características de imagens, vídeos e sons, têm sido exploradas em trabalhos da área.

A recuperação de vídeos baseada em conteúdo (CBVR, do inglês *content-based video retrieval*) é uma modalidade de CBR que permite o processamento de imagens ao longo do tempo visando à extração de informações relevantes que viabilizem indexação e recuperação. Além disso, vídeos digitais possuem, em geral, um áudio vinculado cujo sinal também pode ser processado, permitindo a extração de características.

A maior parte dos trabalhos que envolveram CBVR aplicaram descritores relacionados a características espaciais e temporais dos objetos representados, tais como formas, cores e movimentos [2]. A análise do sinal sonoro, visando a obter descritores relevantes para melhorar os resultados de recuperação, ainda é pouco explorada na literatura científica. Abdullah [3] conduziu um estudo correlato ao nosso, apresentando uma abordagem de combinação entre características visuais e sonoras para a identificação de ações humanas em vídeos. No entanto, nosso trabalho contribui com um estudo comparativo entre diferentes combinações de características sonoras com outros tipos de características, tradicionalmente utilizadas em trabalhos da área.

Nosso objetivo neste trabalho consistiu em verificar se a aplicação de extratores de características de sinais sonoros produz melhores resultados de precisão em CBVR quando utilizados em diferentes combinações com extratores de características de região, cor e movimento de objetos representados nos vídeos.

Para avaliar viabilidade da abordagem, criamos uma base contendo 30 segmentos de vídeos na qual inserimos variações em termos de forma, tamanho, cor e movimento dos objetos retratados, bem como em termos de sinais sonoros. Dezesete características diferentes foram extraídas dos segmentos de vídeos e a distância Euclidiana foi utilizada para conduzir os experimentos.

Os resultados dos experimentos indicaram uma melhora significativa dos resultados de recuperação com a aplicação do extrator de características sonoras em diferentes combinações com as demais características extraídas dos segmentos de vídeos. A abordagem é, portanto, considerada promissora no contexto de CBVR para melhorar o desempenho de recuperação em aplicações reais em diversas áreas do conhecimento.

O presente artigo está organizado da seguinte forma. Na Seção II apresentamos os principais conceitos teóricos envolvidos no trabalho, bem como trabalhos correlatos. Na Seção III apresentamos a metodologia utilizada na condução dos experimentos, bem como a base de dados, os extratores e a função de similaridade utilizados. Na Seção IV apresentamos os resultados obtidos e discussões sobre eles. Por fim, na Seção V apresentamos nossas conclusões.

II. REFERENCIAL TEÓRICO

A. Recuperação baseada em conteúdo

O princípio básico de CBR é pesquisar em um banco de dados uma determinada quantidade de objetos (imagens, vídeos, sons) similares a um objeto fornecido como referência de consulta, de acordo com um ou mais critérios fornecidos.

Os critérios de similaridade são obtidos a partir da extração de características, geralmente relacionadas a cores, texturas e formas quando nos referimos a imagens. Quando nos referimos a vídeos, podemos adicionar características de movimento e sonoras. CBR tem sido aplicada em diferentes áreas. Para cada tipo de objeto multimídia, uma denominação diferente é adotada na literatura científica. Por exemplo, quando se trabalha com imagens o termo comumente adotado é recuperação de imagens baseada em conteúdo [4]; para vídeos, o termo é recuperação de vídeos baseada em conteúdo [5]; e, para sons, recuperação de som (ou áudio) baseada em conteúdo [6].

O conjunto de características que descreve um objeto forma o seu vetor de características. A comparação entre os vetores de características de dois objetos permite medir a similaridade entre eles. Em uma arquitetura básica, uma função de similaridade pode ser utilizada para se calcular uma distância entre os vetores de dois objetos. No entanto, abordagens mais sofisticadas como, por exemplo, a definição de modelos por meio do treinamento de classificadores, têm sido amplamente investigadas na literatura científica da área [7].

CBR tem sido explorada nas últimas décadas em trabalhos científicos para o desenvolvimento dos mais diversos tipos de aplicações. Diversos trabalhos têm apresentado técnicas úteis no manuseio e na utilização da grande quantidade de imagens digitais disponíveis atualmente [7].

Abordagens de CBVR também têm sido exploradas na literatura [1,8]. A aplicação de técnicas de CBVR tem potencial para auxiliar na resolução de problemas nas mais diversas áreas do conhecimento. Podemos citar como exemplos problemas nas áreas de segurança (identificação de pessoas ou objetos de interesse em vídeos) [9] e diagnóstico auxiliado por computador [10,11].

B. Trabalhos correlatos

Embora abordagens de CBVR tenham aplicações em diferentes áreas, a área médica é uma das mais beneficiadas, dado o interesse em análise, processamento e recuperação de vídeos médicos para auxílio ao diagnóstico e à tomada de decisões.

André e seus colaboradores [10,12,13] apresentaram estudos relacionados ao auxílio ao diagnóstico em vídeos provenientes de endomicroscopia. A endomicroscopia confocal é uma técnica que permite a análise em tempo real, e em alta resolução (nível microscópico), do tecido epitelial do trato gastrointestinal [13]. Esses pesquisadores desenvolveram um protótipo de ferramenta que utiliza CBVR no auxílio ao diagnóstico do câncer colorretal. Trabalhos do grupo exploraram técnicas de recuperação de vídeos baseadas em características visuais [10] e combinação entre recuperação baseada em conteúdo e anotação semântica [12]. Eles também

propuseram uma abordagem de recuperação de vídeos de endomicroscopia para a classificação de pólipos colorretais [13].

Quellec e seus colaboradores [11] desenvolveram uma ferramenta capaz de pesquisar, em tempo real, sequências de vídeos similares a uma sequência de vídeo gravada durante um procedimento de cirurgia oftálmica. Eles estudaram casos de cirurgia da membrana epirretiniana e de catarata. A técnica envolve a identificação da tarefa cirúrgica que está sendo executada no segmento de vídeo processado. Com essa abordagem, o cirurgião poderá ser alertado e auxiliado na tomada de decisão frente a uma situação atípica ou de risco durante o procedimento.

Outro objetivo explorado em trabalhos que envolvem CBVR é o reconhecimento de ações humanas em segmentos de vídeos. Essa tarefa pode ser de grande utilidade nas áreas de segurança e educação, por exemplo. DeMenthon e Doermann [9] utilizaram técnicas de CBVR baseadas em descritores de cor e movimento ao longo do tempo para o reconhecimento de ações em vídeos de segurança. Abdullah [13] apresentou uma abordagem de combinação de características visuais e sonoras para a identificação de ações humanas em vídeos. Os resultados demonstraram um aumento significativo do desempenho de recuperação, em termos de precisão e revocação, quando utilizada a abordagem combinada ao invés de somente características visuais ou somente características de som.

III. MATERIAIS E MÉTODOS

A. Base de dados

Nós criamos uma base contendo trinta segmentos de vídeos. Cada segmento de vídeo possui uma duração de três segundos e retrata uma figura geométrica colorida em movimento linear em um plano de fundo branco. Os quadros do vídeo foram processados como imagens RGB, com resolução espacial de 200x200 pixels e profundidade de cor de 32 bits por pixel. O sinal sonoro é proveniente do som da pronúncia de uma palavra selecionada da Língua Portuguesa obtida do Sistema Síntese de Voz e Texto-Fala (TTS, do inglês *text-to-speech*) CPqD Texto-Fala [14].

Cada figura geométrica foi associada a uma palavra em específico (Tabela I). A ideia foi representar a relação do som com uma característica importante do vídeo, no caso, a figura representada independentemente de tamanho, cor ou movimento.

TABELA I: RELAÇÃO DE FIGURAS GEOMÉTRICAS REPRESENTADAS NOS SEGMENTOS DE VÍDEOS COM SUAS RESPECTIVAS PALAVRAS RELACIONADAS.

Figura geométrica	Palavra
Círculo	Segunda
Losango	Novembro
Quadrado	Amanhã
Retângulo	Matutino
Triângulo	Semana

Inserimos variações na base em termos de forma, tamanho, cor, movimento dos objetos retratados e som, proporcionando as condições necessárias para a validação da técnica de CBVR utilizando-se diferentes combinações de características. Nas

Fig. 1 e Fig. 2 são apresentados exemplos de quadros de segmentos de vídeo da base construída para os experimentos.

Fig. 1: Exemplo de quadros do segmento de vídeo que retrata um círculo azul em movimento linear. Neste vídeo o objeto foi associado à palavra “segunda”.

Fig. 2: Exemplo de quadros do segmento de vídeo que retrata um losango vermelho em movimento linear. Neste vídeo o objeto foi associado à palavra “novembro”.

B. Características de região, cor e movimento

Neste estudo utilizamos extratores de características de imagens para extrair informações sobre a figura representada em cada segmento de vídeo. Área, perímetro, largura e altura (ambos em pixels) [15] foram obtidos por meio da análise de um quadro selecionado do segmento de vídeo. Um único quadro para essa análise justificou-se, pois a figura representada é sempre a mesma ao longo da reprodução.

O deslocamento horizontal e o deslocamento vertical, também em pixels, da figura representada foram calculados por meio da diferença entre as coordenadas do pixel centro de massa do primeiro e do último frames.

Foram calculados também para cada banda – R, G e B (do inglês: *red*, *blue* e *green*) – de um quadro selecionado a média do histograma de intensidades do canal, bem como o seu desvio padrão.

Com a aplicação desses extratores, obtivemos 12 valores numéricos que representam informações de região e cor da figura representada no segmento de vídeo, bem como o deslocamento dela ao longo da reprodução.

C. Características sonoras

As características sonoras foram extraídas por meio do processamento do sinal sonoro vinculado ao segmento de vídeo utilizando-se um extrator de fonemas da Língua Portuguesa. Esse extrator, apresentado por Oliveira e seus colaboradores [16], é capaz de determinar com precisão relevante a presença e o instante no qual um fonema do tipo Vogal (A [a/, /ɐ/], E [e/, /ɛ/], I [i/], O [o/, /ɔ/], e U [u/]) é pronunciado em um sinal sonoro.

Neste trabalho utilizamos o extrator para obter um valor numérico representando a presença de cada um dos cinco grupos de fonemas do tipo Vogal mencionados. Obtivemos, portanto, para cada segmento de vídeo, cinco valores numéricos que representam o sinal sonoro vinculado a ele.

Na Tabela II são apresentados os extratores utilizados agrupados por categoria. Nós investigamos combinações entre as diferentes categorias, visando a identificar a influência das características sonoras em cada combinação.

TABELA II: CARACTERÍSTICAS EXTRAÍDAS.

Categoria	Quantidade	Extratores
Região	4	Área, perímetro, largura e altura.
Cor	6	Média e desvio padrão do histograma de intensidades de cada banda.
Movimento	2	Deslocamento horizontal e deslocamento vertical do pixel centro de massa.
Som	5	Presença de cada um dos fonemas do tipo Vogal no sinal sonoro.

D. Função de similaridade

Aplicamos a distância Euclidiana como função de similaridade para realizar comparações entre vetores de características. Tal função é amplamente utilizada em trabalhos que envolvem CBR [15]. Na comparação entre dois vetores, o valor zero representa máxima similaridade entre eles.

E. Avaliação do desempenho

Os resultados dos experimentos foram avaliados utilizando-se as métricas de revocação e precisão, bem como o gráfico que as relaciona. Precisão refere-se à porcentagem de objetos recuperados que são relevantes à consulta. Revocação, por sua vez, refere-se à porcentagem de todos os objetos relevantes na base de dados que foram recuperados [7].

O gráfico de revocação *versus* precisão permite analisar a variação da precisão com o aumento da revocação. O resultado ideal seria máxima precisão (1,0) para qualquer valor de revocação maior que zero. A área delimitada pela curva do gráfico e pelo eixo das abscissas – área sob a curva (AUC, do inglês *area under the curve*) – é uma métrica importante para a avaliação de desempenho de abordagens de CBR. Quanto mais próximo de 1,0 for o valor da AUC, melhor é o desempenho da abordagem de CBR plotada no gráfico.

F. Metodologia de execução das consultas

Os experimentos conduzidos consistiram na execução de consultas combinando-se diferentes categorias de características. Na Tabela III são apresentadas as combinações utilizadas.

TABELA III: EXPERIMENTOS EXECUTADOS COMBINANDO-SE DIFERENTES CATEGORIAS DE CARACTERÍSTICAS.

Experimento	Combinação entre características
1	Região
2	Região + Som
3	Região + Movimento
4	Região + Movimento + Som
5	Região + Movimento + Cor
6	Região + Movimento + Cor + Som

A base de dados contém seis segmentos de vídeo para cada par [figura, palavra] distinto (Tabela I). Isto é, seis segmentos que representam a mesma figura em movimento e a reprodução da mesma palavra. As características que variam entre um segmento e outro são tamanho, cor e movimento da figura retratada.

Foram executadas cinco consultas em cada experimento. Cada consulta utilizou um segmento de vídeo representando um par [figura, palavra] distinto. Para cada segmento de vídeo utilizado como modelo, consideramos como resultados corretos de recuperação os cinco demais segmentos de vídeos que representam o mesmo par [figura, palavra].

Para cada resultado de consulta, traçamos o gráfico de revocação *versus* precisão. Para fins de comparação, uma curva de revocação *versus* precisão é melhor do que outra quando sua AUC é maior. Isto significa que, em uma curva melhor, a precisão diminui menos quando a revocação aumenta.

IV. RESULTADOS E DISCUSSÃO

Os experimentos executados visaram a comprovar que incluir características do sinal sonoro na análise dos segmentos de vídeos pode melhorar os resultados de recuperação, principalmente em casos em que a discriminação entre os segmentos de vídeos relevantes e os não relevantes é difícil considerando-se apenas as características dos objetos retratados. Os resultados dos experimentos discutidos abaixo ajudam a corroborar a nossa hipótese de trabalho.

A. Características de região e som

No primeiro experimento, consideramos apenas as características de região (área, perímetro, largura e altura) para compor os vetores de características de cada segmento de vídeo. Na Fig. 3 é apresentado o gráfico de revocação *versus* precisão obtido para cada consulta.

Fig. 3: Gráfico de revocação *versus* precisão das consultas que utilizaram apenas as características de região.

A análise do gráfico revela uma queda significativa da precisão com o aumento da revocação. Para valores de revocação maiores que 0,4, temos valores baixos de precisão.

Incluindo as características sonoras na composição dos vetores de características, combinando-as com as características

de região, obtivemos os resultados apresentados na Fig. 4.

Fig. 4: Gráfico de revocação *versus* precisão das consultas que utilizaram as características de região combinadas com as características sonoras.

Observamos que, com exceção da Consulta 4, todas as demais consultas resultaram em precisão máxima (100%) para valores de revocação maiores que zero. Esse resultado oferece indícios da exequibilidade da abordagem e demonstra que as características sonoras tiveram influência positiva e determinante na obtenção de resultados satisfatórios.

No caso da Consulta 4, mesmo com a baixa precisão para valores de revocação maiores que 0,4, o resultado ainda é melhor do que o observado considerando-se apenas características de região. Essa consulta utilizou como modelo um segmento de vídeo representando o par [retângulo, “matutino”]. As características de altura, largura e perímetro do retângulo representado são muito próximas das mesmas características calculadas em três segmentos de vídeo que retratam um triângulo. Esses segmentos que retratam os triângulos foram, então, recuperados na Consulta 4 e considerados mais relevantes de acordo com as distâncias calculadas, explicando-se, portanto, o resultado observado.

Na Tabela IV são apresentados os valores de AUC calculados para os experimentos que envolveram características de região e sonoras. Os valores corroboram o resultado inferido por meio dos gráficos: a inclusão das características sonoras melhorou o desempenho de recuperação para todas as consultas.

TABELA IV: VALORES DE AUC PARA OS EXPERIMENTOS QUE UTILIZARAM CARACTERÍSTICAS DE REGIÃO E SOM.

	C1	C2	C3	C4	C5	Média
Região	0,62	0,58	0,59	0,57	0,62	0,59
Região / Som	1,00	1,00	1,00	0,62	1,00	0,92

B. Características de região, movimento e som

No experimento seguinte combinamos características de região e movimento. Na base de dados construída, foram incluídas apenas duas variações de movimento, sendo que metade da base possui a primeira variação e a outra metade a segunda variação. Isto faz com que as características de

movimento se tornem um fator complicador para a discriminação entre segmentos de vídeos relevantes e segmentos de vídeo não relevantes, pois segmentos retratando figuras diferentes têm 50% de chances de possuírem os mesmos valores para as características de movimento.

Conforme podemos observar no gráfico da Fig. 5, não obtivemos melhora da precisão nas consultas quando comparado à utilização apenas das características de região (Fig. 3). Esse resultado se justifica, conforme mencionamos, na confusão proporcionada pelas características de movimento.

Fig. 5: Gráfico de revocação *versus* precisão das consultas que utilizaram as características de região combinadas com as características de movimento.

Combinando as características sonoras com as características de região e movimento, obtivemos um gráfico de revocação *versus* precisão melhor para todas as consultas executadas, conforme observamos na Fig. 6.

Fig. 6: Gráfico de revocação *versus* precisão das consultas que utilizaram a combinação entre as características de região, movimento e som.

Para valores de revocação maiores ou iguais a 0,6, obtivemos valores baixos de precisão, pois as características de movimento contribuem para dificultar a discriminação entre os segmentos de vídeos relevantes e os não relevantes, conforme mencionado anteriormente. Porém, também neste caso, as características sonoras melhoraram a precisão de recuperação

em todas as consultas. Esse resultado também é corroborado com a análise comparativa dos valores de AUC para os experimentos, conforme apresentado na Tabela V.

TABELA V: VALORES DE AUC PARA OS EXPERIMENTOS QUE UTILIZARAM CARACTERÍSTICAS DE REGIÃO, MOVIMENTO E SOM.

	C1	C2	C3	C4	C5	Média
Região / Movimento	0,62	0,58	0,59	0,57	0,62	0,59
Região / Movimento / Som	0,69	0,65	0,65	0,59	0,69	0,65

C. Características de região, movimento, cor e som

Combinamos, adicionalmente, as características de região, movimento e cor, obtendo o gráfico de revocação *versus* precisão apresentado na Fig. 7.

Fig. 7: Gráfico de revocação *versus* precisão das consultas que utilizaram a combinação entre as características de região, movimento e cor.

Assim como as características de movimento, as características de cor também funcionaram como fator complicador para a discriminação entre os segmentos de vídeos relevantes e os não relevantes. Na base de dados construída, as cores representadas (vermelho, verde e azul) foram distribuídas de forma homogênea entre as formas geométricas. Na Fig. 7 observamos, então, um gráfico pior do que o observado na Fig. 5 (combinação entre características de região e movimento).

Com a inclusão das características sonoras, os gráficos de revocação *versus* precisão melhoraram significativamente, conforme observamos na Fig. 8.

A Tabela VI mostra os valores de AUC referentes às consultas que utilizaram características de região, movimento, cor e som. Os resultados mostram que, também neste caso, as características sonoras proporcionaram melhora nos valores de precisão.

D. Discussão e aplicações

Os resultados demonstraram que, em todos os cenários, as características sonoras melhoraram a precisão de recuperação. Mesmo quando incluímos características visando a piorar a capacidade de discriminação entre os segmentos de vídeo relevantes e os não relevantes, a análise do sinal sonoro tendeu a melhorar os resultados de recuperação.

Fig. 8: Gráfico de revocação versus precisão das consultas que utilizaram a combinação entre as características de região, movimento, cor e som.

TABELA VI: VALORES DE AUC PARA OS EXPERIMENTOS QUE UTILIZARAM CARACTERÍSTICAS DE REGIÃO, MOVIMENTO, COR E SOM.

	C1	C2	C3	C4	C5	Média
Região / Movimento / Cor	0,33	0,28	0,27	0,23	0,32	0,29
Região / Movimento / Cor / Som	0,66	0,63	0,62	0,58	0,71	0,64

Considerando-se o valor de AUC médio, a combinação entre características que gerou melhores resultados é composta pelas características de região e som. O resultado médio desta combinação foi 55,9% melhor comparando-se com os resultados do experimento que utilizou apenas as características de região.

O maior aumento de desempenho de recuperação com a inclusão das características sonoras foi observado nos experimentos que utilizaram as características de região, movimento e cor. Nesse caso, a inclusão das características sonoras resultou em uma melhora de 120,7% em termos de desempenho de recuperação, considerando-se o valor médio de AUC.

Mesmo com nossos experimentos considerando situações controladas, estes resultados são importantes, pois indicam que a abordagem é promissora e pode ser útil na implementação de ferramentas de análise e recuperação de vídeos em diversas aplicações. Por exemplo, a área médica pode obter diferentes contribuições para ferramentas de auxílio ao diagnóstico em exames como ecocardiografia e ultrasonografia, nos quais estão presentes as componentes visuais e sonoras e ambas são relevantes para o diagnóstico.

Uma limitação desta abordagem consiste na complexidade de implementação de extratores de características, principalmente aqueles relacionados aos sinais sonoros. O processamento de sinais sonoros é uma tarefa complexa e relacionar características desses sinais com as especificidades do caso a ser tratado também não é trivial, exigindo um conhecimento profundo do problema. Contudo, o avanço de estudos relacionados à CBR e ao processamento de sinais tem

permitido encontrar na literatura extratores que podem ser utilizados em diferentes aplicações com até mesmo as implementações disponibilizadas por seus pesquisadores. É o caso do trabalho de Oliveira e seus colaboradores, que desenvolveram o extrator de fonemas utilizado no presente estudo, bem como um *framework*, que utiliza conceitos de CBR para automatização de testes de programas com saídas complexas, que contém diversos extratores de características implementados e disponibilizados [16].

Melhorar os resultados de recuperação promovendo eficiência à abordagem é um desafio em CBR. Selecionar características adequadas ao problema; evitar características correlacionadas (que implicam em processamento adicional desnecessário); e selecionar abordagens eficientes de medição de similaridade por conteúdo entre dados multimídia são desafios complexos da área. Nunes e seus colaboradores [15], por exemplo, apresentaram um estudo comparativo de funções de similaridade com aplicações em CBIR. Eles contribuíram com uma discussão com relação aos impactos da seleção de diferentes funções de similaridade para a tarefa de comparação entre vetores de características.

Trabalhos como o de Nunes e seus colaboradores [15], bem como o apresentado neste artigo auxiliam no desenvolvimento de novas abordagens visando a melhorar cada vez mais o desempenho de aplicações de CBR, viabilizando a expansão de sua adoção em diferentes aplicações.

V. CONCLUSÕES

No presente trabalho conduzimos experimentos de CBVR utilizando diferentes combinações entre características de região, movimento e cor dos objetos representados, bem como características dos sinais sonoros reproduzidos nos segmentos de vídeos. Nosso objetivo foi demonstrar que a análise das características sonoras podem melhorar significativamente os resultados de recuperação, mesmo quando a discriminação entre os vídeos relevantes e os não relevantes para uma consulta é difícil, dadas as demais características de baixo nível analisadas.

Os resultados demonstraram que as características sonoras tendem a melhorar a precisão da recuperação, mesmo quando incluímos características visando a piorar a capacidade de discriminação entre os segmentos de vídeo relevantes e os não relevantes.

Pretendemos, em trabalhos futuros, estudar a abordagem na análise e recuperação de vídeos médicos para auxílio ao diagnóstico, além de melhorar a abordagem por meio da verificação de outras formas de medir a similaridade.

Referências

- [1] H.J. Zhang, J. Wu, D. Zhong e S.W. Smoliar, "An integrated system for content-based video retrieval and browsing", *Pattern Recognition*, vol. 30, n. 4, p. 643-658, abr. 1997.
- [2] C. Chattopadhyay e S. Das, "Enhancing the MST-CSS representation using robust geometric features, for efficient content based video retrieval (CBVR)", *Proceedings of the 2012 IEEE International Symposium on Multimedia (ISM)*, 2012, p. 352-355.
- [3] L.N. Abdullah, "A framework for evaluating human action detection via multidimensional approach", *Proceedings of the Sixth International Conference on Computer Graphics, Imaging and Visualization*, 2009, p.

- [4] R. Datta, J. Li e J.Z. Wang, "Content-based image retrieval: approaches and trends of the new age", *Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval*, 2015, p. 253-262.
- [5] C. Gurrin, "Content-Based Video Retrieval", in L. Liu, and M.T. Özsu, "*Encyclopedia of database systems*", Boston: Springer US, 2009, p. 466-473.
- [6] M. Müller, "Content-based audio retrieval", in "*Fundamentals of music processing: audio, analysis, algorithms, applications*", Cham: Springer International Publishing, 2015. p. 355-413.
- [7] R. Datta, D. Joshi, J. Li e J.Z. Wang, "Image retrieval: ideas, influences, and trends of the new age", *ACM Computing Surveys*, vol. 40, n. 2, p. 5:1-5:60, mai. 2008.
- [8] J.-H. Su, Y.-T. Huang, H.-H. Yeh e V.S. Tseng, "Effective content-based video retrieval using pattern-indexing and matching techniques", *Expert Systems with Applications*, vol. 37, n. 7, p. 5068-5085, jul. 2010.
- [9] D. DeMenthon e D. Doermann, "Video retrieval using spatio-temporal descriptors", *Proceedings of the 11th ACM International Conference on Multimedia*, 2013, p. 508-517.
- [10] B. André, T. Vercauteren, A.M. Buchner, M.B. Wallace e N.A. Ayache, "A smart atlas for endomicroscopy using automated video retrieval", *Medical Image Analysis*, vol. 15, n. 4, p. 460-476, 2011.
- [11] G. Quellec, K. Charrière, M. Lamard, Z. Droueche, C. Roux, B. Cochener e G. Cazuguel, "Real-time recognition of surgical tasks in eye surgery videos", *Medical Image Analysis*, vol. 18, n. 3, p. 579-590, abr. 2014.
- [12] B. André, T. Vercauteren, A.M. Buchner, M.B. Wallace e N. Ayache, "Learning semantic and visual similarity for endomicroscopy video retrieval", *IEEE Transactions on Medical Imaging*, vol. 31, n. 6, p. 1276-1288, jun. 2012.
- [13] B. André, T. Vercauteren, A.M. Buchner, M. Krishna, N. Ayache e M.B. Wallace, "Software for automated classification of probe-based confocal laser endomicroscopy videos of colorectal polyps", *World Journal of Gastroenterology*, vol. 18, n. 39, p. 5560-5569, out. 2012.
- [14] "CPqD Texto-Fala", 2016. [Online]. Disponível em: <http://www.cpqd.com.br/textofala>.
- [15] F.L.S. Nunes, M.E. Delamaro, V.M. Gonçalves e M.S. Lauretto, "CBIR based testing oracles: an experimental evaluation of similarity functions", *International Journal of Software Engineering and Knowledge Engineering*, vol. 25, n. 8, p. 1271-1306, out. 2015.
- [16] R.A.P. Oliveira, A. Memon, V.N. Gil, F.L.S. Nunes e M.E. Delamaro, "An extensible framework to implement test oracle for non-testable programs", *Proceedings of the 26th International Conference on Software Engineering and Knowledge Engineering*, 2014, p. 199-204.