



Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Engenharia de Software

# Aprimoramento de sistemas de busca por imagens usando modelos de realimentação de relevâncias

Autor: Gustavo Lopes de Brito  
Orientador: Prof. Dr. Glauco Vitor Pedrosa

Brasília, DF  
2019



Gustavo Lopes de Brito

# **Aprimoramento de sistemas de busca por imagens usando modelos de realimentação de relevâncias**

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Universidade de Brasília - UnB

Faculdade UnB Gama - FGA

Orientador: Prof. Dr. Glauco Vitor Pedrosa

Brasília, DF

2019

---

Gustavo Lopes de Brito

Aprimoramento de sistemas de busca por imagens usando modelos de realimentação de relevâncias/ Gustavo Lopes de Brito. – Brasília, DF, 2019-  
57 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Glauco Vitor Pedrosa

Trabalho de Conclusão de Curso – Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA , 2019.

1. Banco de Dados. 2. Visão Computacional. I. Prof. Dr. Glauco Vitor Pedrosa.  
II. Universidade de Brasília. III. Faculdade UnB Gama. IV. Aprimoramento de  
sistemas de busca por imagens usando modelos de realimentação de relevâncias

CDU 02:141:005.6

---

Gustavo Lopes de Brito

## **Aprimoramento de sistemas de busca por imagens usando modelos de realimentação de relevâncias**

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Trabalho aprovado. Brasília, DF, 03 de julho de 2019:

---

**Prof. Dr. Glauco Vitor Pedrosa**  
Orientador

---

**Prof. Dr. Nilton Correia da Silva**  
Convidado 1

---

**Prof. Dr. Marcus Vinícius Chaffim  
Costa**  
Convidado 2

Brasília, DF  
2019

# Resumo

A recuperação de imagens refere-se à tarefa de buscar por imagens similares em uma base de dados a partir de outra imagem fornecida como consulta. Tal tarefa é realizada por sistemas denominados CBIR (do inglês *Content-Based Image Retrieval*). Tais sistemas de recuperação geram um *ranking* com as imagens mais similares, a partir de uma imagem-consulta, usando algum critério de similaridade baseado nas características intrínsecas relacionadas ao conteúdo visual da imagem, tais como: cor, forma e/ou textura. Porém, o uso de características extraídas automaticamente da imagem pode não ser o suficiente para associar o conteúdo de uma imagem ao seu significado, o que se reflete nos resultados de uma busca, afetando a satisfação do usuário do sistema. Este trabalho busca mitigar esse problema por meio do uso de modelos de Realimentação de Relevância, motivado pelos bons resultados relatados na literatura referentes ao uso de tais técnicas. A proposta desses modelos é utilizar a interação com o usuário a fim de aprimorar os resultados da busca. A ideia central do modelo é que o usuário forneça informações sobre quais imagens ele considera serem relevantes e/ou não-relevantes a partir de uma consulta inicial e, então, utilizar essa informação para aprimorar o *ranking* gerado pelo sistema. Este trabalho apresenta alguns conceitos e trabalhos relacionados com o tema em questão, bem como alguns testes experimentais realizados com o objetivo de comparar diferentes técnicas propostas na literatura e um método proposto que combina múltiplos centros de consulta e agregação de rankings para realizar a realimentação de relevância, esta possui desempenho melhor que a Movimentação de Objeto Consulta (QPM, do inglês *Query Point Movement*) e desempenho similar a técnica *Relevance Feedback Multiple Projection* em relação à precisão das consultas.

**Palavras-chaves:** Recuperação de Informação, Visão Computacional, Inteligência Artificial, Processamento Digital de Imagens, Busca por Similaridade, Realimentação de Relevâncias.

# Abstract

Image retrieval refers to the task of searching for similar images in a database from the image sent in a query. Such task is performed by systems called CBIR (*Content-Based Image Retrieval*). Such retrieval systems generate a ranking with the most similar images, by using a query image, they make use of some criterion of similarity based on the intrinsic characteristics related to the visual content of the image, such as: color, shape and/or texture. However, the use of automatically extracted features is not yet effective enough to associate the content of an image with its meaning, which is reflected in the results of a search, affecting the user satisfaction of the system. The current project seeks to mitigate this problem through the use of Relevance Feedback models, motivated by good results reported in the literature that report the use of these techniques. One of the proposals of these models is to make use of user interaction to improve the results of a search. The idea is that an user should inform the images that they consider relevant and/or non-relevant based on an initial query, and then use this information to improve the ranking generated by the system. This work presents some concepts and works related to the subject in question, as well as some experimental tests performed with the objective of comparing different techniques proposed in the literature and a new method proposed to perform relevance feedback with better performance than Query Point movement, and similar performance to Relevance Feedback Multiple Projection.

**Key-words:** information retrieval, computer vision, artificial intelligence, image processing, similarity search, relevance feedback.

# Lista de ilustrações

Figura 1 – Exemplo de consulta realizada pelo Google Images. . . . .	13
Figura 2 – Arquitetura básica de um sistema CBIR . . . . .	16
Figura 3 – Representação do subespaço de cores do modelo RGB (GONZALEZ; WOODS, 2006). . . . .	17
Figura 4 – Imagens visualmente diferentes, porém com o mesmo histograma das imagens em escala de cinza. . . . .	21
Figura 5 – Processo de Realimentação de Relevâncias em um Sistemas de Busca por Imagens Similares . . . . .	25
Figura 6 – Movimentação de um objeto de consulta usando o modelo de Query Point Movement iterativo . . . . .	26
Figura 7 – Representação dos múltiplos centros de consulta (ORTEGA-BINDERBERGER; MEHROTRA, 2003) . . . . .	27
Figura 8 – Passos de construções das projeções (MARQUES, 2006). . . . .	29
Figura 9 – Passos da técnica Relevance Multiple Point Projection (MARQUES; TRAINA, 2006) . . . . .	30
Figura 10 – Página inicial onde o usuário pode selecionar a imagem-consulta. . . . .	34
Figura 11 – Exemplo do resultado obtido usando a imagem-consulta da Fig. 10 . . . . .	34
Figura 12 – Nova consulta utilizando uma imagem da classe FLOR. . . . .	35
Figura 13 – Resultados recuperados pela consulta representada na Fig. 12 . . . . .	35
Figura 14 – Diagrama de Atividades do sistema desenvolvido . . . . .	36
Figura 15 – A técnica proposta realiza uma nova consulta para cada imagem seleci- onada como relevante pelo usuário. Os rankings gerados por essas novas consultas são agregados para gerar um novo ranking para a imagem- consulta original. . . . .	38
Figura 16 – Imagens de cada classe da base <i>Corel1000</i> (LI LIN FENG, 2016) . . . . .	40
Figura 17 – Imagens da base <i>OxfordFlowers</i> (PIAO et al., 2016) . . . . .	41
Figura 18 – Precisão da primeira consulta sobre cada classe da base COREL1000. . . . .	42
Figura 19 – Precisão média da primeira consulta sobre a base <i>Oxford Flowers</i> . . . . .	42
Figura 20 – Precisão média por iteração na base COREL1000 analisando o <i>top-10</i> . . . . .	43
Figura 21 – Precisão média por iteração na base COREL1000 analisando o <i>top-20</i> . . . . .	44
Figura 22 – Precisão média por iteração na base <i>Oxford Flowers</i> analisando o <i>top-10</i> . . . . .	45
Figura 23 – Precisão média por iteração na base <i>Oxford Flowers</i> analisando o <i>top-20</i> . . . . .	45
Figura 24 – MAP na primeira e última iteração na base COREL1000 analisando o <i>top-20</i> . . . . .	46
Figura 25 – MAP na primeira e última iteração na base COREL1000 analisando o <i>top-20</i> . . . . .	47

Figura 26 – MAP na primeira e última iteração na base <i>Oxford Flowers</i> analisando o <i>top-10</i> . . . . .	47
Figura 27 – MAP na primeira e última iteração na base <i>Oxford Flowers</i> analisando o <i>top-20</i> . . . . .	48
Figura 28 – Mapeamento espacial das imagens da base Corel100 utilizando histograma de escala de cinza e projetadas pelo t-SNE (melhor visualizado no formato digital) . . . . .	56
Figura 29 – Mapeamento espacial das imagens da base OxfordFlowers utilizando histograma de escala de cinza e projetadas pelo t-SNE (melhor visualizado no formato digital) . . . . .	57



# Lista de tabelas

Tabela 1 – Tabela com funções para conversão de <i>pixels</i> no modelo RGB para escala de cinza. . . . .	18
Tabela 2 – Exemplo da agregação de <i>ranking</i> utilizada pela técnica combSUM. . .	31
Tabela 3 – Matriz de cálculo do vetor de médias . . . . .	39
Tabela 4 – <i>Mean Average Precision</i> da base COREL1000 analisando o <i>top-10</i> e <i>top-20</i> . . . . .	48
Tabela 5 – <i>Mean Average Precision</i> da base <i>Oxford Flowers</i> analisando o <i>top-10</i> e <i>top-20</i> . . . . .	49

# Lista de abreviaturas e siglas

CBIR	<i>Content-Based Image Retrieval</i> (Recuperação de imagens baseado em conteúdo)
QPM	<i>Query Point Movement</i> (Movimentação de Objeto Consulta)
AP	<i>Average Precision</i> (Precisão Média)
MAP	<i>Mean Average Precision</i> (Precisão Média)
RF	<i>Relevance Feedback</i> (Realimentação de Relevâncias)
RFP	<i>Relevance Feedback Projection</i> (Projeção de Realimentação de Relevâncias)
RFMP	<i>Relevance Feedback Projection</i> (Múltiplas Projeções de Realimentação de Relevâncias)

# Sumário

	<b>Introdução</b> . . . . .	<b>12</b>
<b>1</b>	<b>RECUPERAÇÃO DE IMAGENS BASEADO EM CONTEÚDO</b> . . . . .	<b>15</b>
1.1	Fluxo do funcionamento de um sistema CBIR . . . . .	15
1.2	Extração de Características . . . . .	16
1.2.1	Cores . . . . .	17
1.2.1.1	Conversão de imagem colorida para imagem em tons de cinza . . . . .	18
1.2.1.2	Formas . . . . .	18
1.2.1.3	Textura . . . . .	19
1.2.2	Descrição de imagens usando Histograma de Cor . . . . .	19
1.3	Medidas de Similaridade . . . . .	20
1.3.1	Funções de distância de Minkowski . . . . .	21
1.4	Avaliação do Desempenho de Sistemas CBIR . . . . .	22
1.4.1	<i>Precisão</i> . . . . .	22
1.4.2	<i>Mean Average Precision (MAP)</i> . . . . .	23
<b>2</b>	<b>REALIMENTAÇÃO DE RELEVÂNCIAS</b> . . . . .	<b>24</b>
2.1	Funcionamento . . . . .	24
2.2	Algoritmos de Realimentação de Relevâncias . . . . .	24
2.2.1	Movimentação do Objeto de Consulta (QPM) . . . . .	25
2.2.2	Múltiplos centros de consulta . . . . .	26
2.2.2.1	Relevance Feedback Projection (RFP) . . . . .	27
2.2.2.2	Relevance Feedback Multiple Projection (RFMP) . . . . .	29
2.2.3	Agregação de Rankings . . . . .	30
<b>3</b>	<b>ATIVIDADES DESENVOLVIDAS</b> . . . . .	<b>32</b>
3.1	Detalhes Técnicos de Implementação do Sistema . . . . .	32
3.1.1	Metodologia de Desenvolvimento do Software . . . . .	32
3.1.2	Ferramentas utilizadas . . . . .	32
3.1.2.1	Lógica do sistema de consultas por conteúdo . . . . .	33
3.1.2.2	Interface gráfica . . . . .	33
3.1.3	Fluxo de funcionamento do sistema . . . . .	34
3.2	Extração de Características e Medidas de Similaridade . . . . .	35
3.3	Técnicas de Realimentação de Relevância Implementadas . . . . .	36
3.3.1	Movimentação de Objeto Consulta (QPM) . . . . .	37
3.3.2	Múltiplos centros de consulta . . . . .	37

3.4	Método Proposto: Realimentação de Relevâncias com Agregação de <i>Rankings</i> (RFRA) . . . . .	38
4	RESULTADOS EXPERIMENTAIS . . . . .	40
4.1	Bases de imagens . . . . .	40
4.2	Metodologia para a Avaliação de Desempenho . . . . .	40
4.3	Resultados Obtidos . . . . .	41
4.3.1	Sem Realimentação de Relevâncias . . . . .	41
4.3.2	Com Realimentação de Relevância . . . . .	43
5	CONSIDERAÇÕES FINAIS . . . . .	50
5.1	Trabalhos futuros . . . . .	50
	REFERÊNCIAS . . . . .	52
	APÊNDICES . . . . .	55
	APÊNDICE A – CORRELAÇÃO DAS IMAGENS DA BASE DE DADOS . . . . .	56

# Introdução

## Contextualização

Em meio à inovação tecnológica, a quantidade de dados gerados através dos meios eletrônicos tem aumentado significativamente nas últimas décadas. Isso se deve, principalmente, ao desenvolvimento tecnológico que possibilitou um avanço dos sistemas de armazenamento de dados e a capacidade destes em armazenar diferentes tipos de conteúdo, tais como áudios, imagens e vídeos. Diante do surgimento de grandes bases de dados, surge a necessidade da construção de sistemas computacionais capazes de realizar consultas (recuperar informações) a partir de um dado fornecido pelo usuário, seja este dado estruturado (tabelas, textos) ou não-estruturado (imagens, vídeos, áudios).

Para realizar a busca por imagens similares, surgiram os sistemas de Recuperação de Imagens Baseado em Conteúdo (CBIR, do inglês *Content Based Image Retrieval*) (TORRES; FALCAO, 2006a). Esses sistemas são capazes de buscar por imagens similares, a partir de uma imagem-consulta fornecida pelo usuário, usando apenas as informações referentes ao conteúdo visual da imagem. A medicina é uma das áreas mais beneficiadas por tais sistemas (PILEVAR, 2011). Em um ambiente clínico, a possibilidade de recuperar e comparar exames, a partir de imagens, usando suas características visuais intrínsecas (cor, forma e/ou textura), permite que exames similares possam ser facilmente recuperados e, assim, aumentam a confiança do especialista médico no diagnóstico.

Os sistemas CBIR realizam, de maneira automática, a extração das características das imagens e cálculos para determinar a similaridade de uma imagem-consulta e as imagens presentes em uma base de dados. O Google Images <sup>1</sup> é um exemplo de sistema que realiza tal tarefa: ele permite que o usuário forneça como consulta uma imagem e, em seguida, busca por todas as imagens na web que são similares à esta informação fornecida. A Figura 1 mostra um exemplo de consulta realizada pelo Google Images.

A eficácia dos sistemas CBIR depende diretamente das características que são extraídas da imagem. Pode-se, por exemplo, extrair informações de cores, texturas e sobre as formas dos objetos presentes no conteúdo da imagem. Na literatura existem diversas técnicas propostas para realizar essa tarefa, porém com uma única consulta pode não ser possível capturar a expectativa do usuário e, como consequência, o sistema poderá retornar resultados não-relevantes, ou seja, imagens não-similares. Por isso, o uso de técnicas de realimentação de relevância tem sido importante para aprimorar a eficácia desses sistemas (MANNING; RAGHAVAN; SCHÜTZE, 2008). Essas técnicas tendem a melho-

---

<sup>1</sup> <https://www.google.com.br/images>

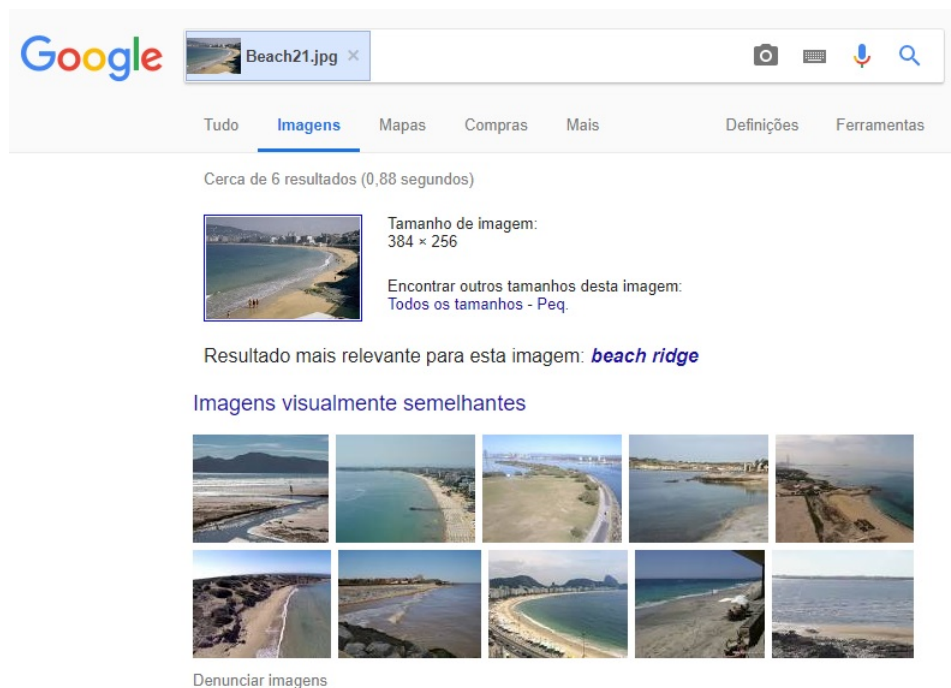


Figura 1 – Exemplo de consulta realizada pelo Google Images.

rar o resultado da busca, pelo fato de que o usuário consegue determinar qual imagem é relevante ou não (a partir de uma consulta inicial) e, desta forma o sistema pode realizar uma nova consulta que atenda às suas expectativas, ou seja, a busca é otimizada com relação à perspectiva do usuário do sistema.

## Objetivos

O objetivo geral desse trabalho foi desenvolver um sistema computacional capaz de realizar a busca por imagens similares em uma base de dados, a partir de uma imagem fornecida como consulta pelo usuário.

Para que o sistema tivesse desempenho eficaz, a proposta foi utilizar técnicas de realimentação de relevância para melhorar o poder de consulta do sistema. Essas técnicas são bastante utilizadas na literatura de recuperação de informação e consistem em obter do usuário um *feedback* a partir de um resultado inicial e, assim, otimizar as buscas a partir da perspectiva do usuário do sistema.

## Objetivos específicos

Para atingir o objetivo geral deste trabalho, alguns objetivos específicos foram definidos:

- Construção de um protótipo de sistema CBIR;

- Implementação de técnicas de realimentação de relevância sobre o protótipo CBIR desenvolvido;
- Comparação das técnicas de realimentação de relevâncias implementadas.
- Analisar o desempenho das técnicas usando diferentes bases de dados.

## Organização do trabalho

A organização do texto deste trabalho está definido da seguinte forma:

- O Capítulo 1 fornece algumas informações conceituais sobre a área de recuperação de imagens baseada em conteúdo, explicitando os principais passos dessa atividade;
- O Capítulo 2 traz informações sobre técnicas de realimentação de relevância e como estas são utilizadas para a melhora do desempenho de um sistema CBIR;
- O Capítulo 3 apresenta os materiais e métodos utilizados para a confecção do trabalho proposto;
- O Capítulo 4 trata dos resultados gerais do trabalho, em que são apresentadas as métricas coletadas sobre as técnicas aplicadas no sistema CBIR;
- O Capítulo 5 traz conclusões acerca do trabalho executado, e possíveis trabalhos futuros.

# 1 Recuperação de imagens baseado em conteúdo

A recuperação de imagens baseada em conteúdo é uma técnica de busca que permite a recuperação de imagens similares, em uma base de dados, a partir de outra imagem fornecida como consulta pelo usuário, usando apenas as informações do conteúdo visual da imagem, sem nenhuma informação de texto associada.

A recuperação de imagens é realizada por sistemas denominados na literatura como CBIR (*Content Based Image Retrieval*). Segundo Marques (2006), tais sistemas demandam a integração das áreas de processamento de imagens e banco de dados, já que são necessários conhecimentos sobre: extração de características de imagens, desenvolvimento de um modelo eficiente de armazenamento para permitir a indexação na busca de imagens por conteúdo, além de permitir consultas por similaridade.

Neste capítulo são apresentados os principais conceitos e definições sobre recuperação de imagens. Serão discutidos aspectos teóricos e computacionais sobre o tema, bem como medidas de avaliação de desempenho de sistemas que realizam a busca por conteúdo.

## 1.1 Fluxo do funcionamento de um sistema CBIR

Quando o usuário fornece uma imagem para busca, o sistema CBIR primeiramente extrai as características visuais dessa imagem e as representa na forma de um vetor de características. Esse vetor é, então, comparado com os vetores de características das imagens que estão no banco de dados. Essa comparação consiste em aplicar uma função de distância, a fim de obter um valor de “dissimilaridade”. O sistema faz o ranqueamento das imagens de acordo com o valor dessa medida de dissimilaridade, de modo que as imagens mais similares (com menor distância) apareçam nas primeiras posições do *ranking*.

A Figura 2 ilustra o fluxo básico do funcionamento de um sistema CBIR. Esses sistemas possuem normalmente duas etapas: uma etapa de extração das características que irá gerar um vetor de características associado a cada imagem, e outra etapa que realiza a recuperação de imagens similares em uma base de dados a partir do cálculo da dissimilaridade entre os vetores de características extraídos na etapa anterior (MISTRY; INGOLE, 2013).

De acordo com Mistry e Ingole (2013), as características mais comuns que podem ser extraídas de uma imagem são: cores, texturas e as formas (silhuetas) dos objetos



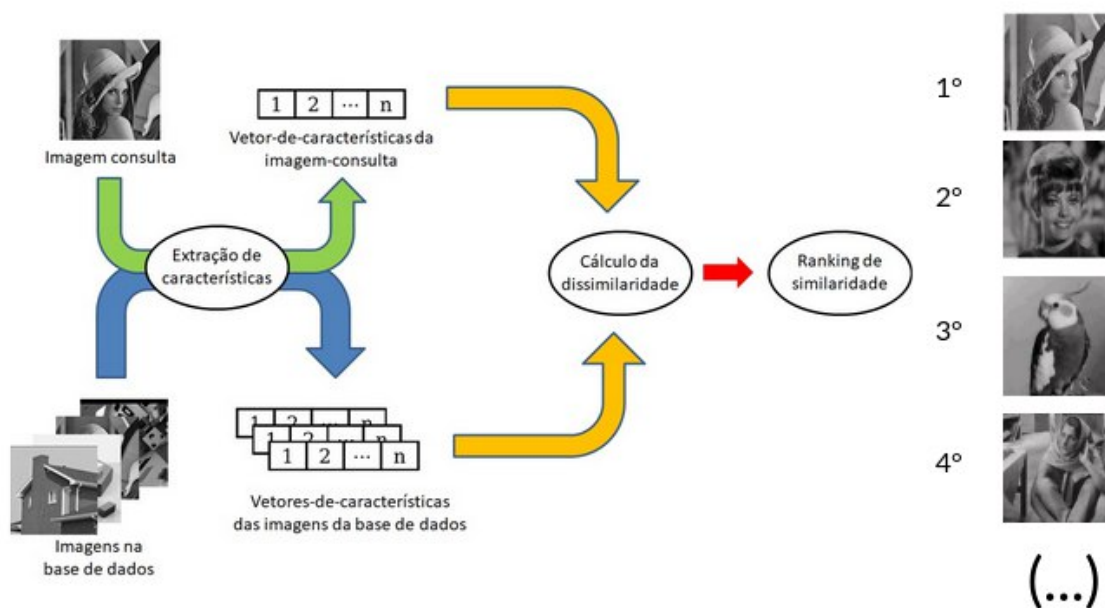


Figura 2 – Arquitetura básica de um sistema CBIR

contidos na imagem. Geralmente, utiliza-se a distância Euclidiana para calcular a diferença dos vetores de características entre a imagem fornecida para busca e das imagens da base de dados (TORRES; FALCAO, 2006b). Após o cálculo da distância entre cada par de vetores, há uma ordenação dos resultados em ordem crescente, retornando as imagens da base de dados no formato de um *ranking*. Uma imagem comparada com ela mesma sempre terá distância igual a *zero*.

## 1.2 Extração de Características

Para analisar a similaridade entre imagens, o primeiro passo é extrair o que há de mais representativo no conteúdo visual da imagem, de modo que sejam obtidas características que representem o conteúdo visual de uma imagem de maneira sucinta e que essas características extraídas possam ser discriminativas na análise da similaridade com outras imagens. Esse processo é definido como extração de características (ASLANDOGAN; YU, 1999).

O conteúdo visual de uma imagem pode ser caracterizado usando três características de baixo-nível da imagem, que são: distribuição de cores, elementos de textura e atributos de forma dos objetos presentes nas imagens, além de avaliar também o relacionamento entre eles (MARQUES, 2006). Estas informações são consideradas na literatura de características de baixo-nível.

### 1.2.1 Cores

A distribuição de cores é uma informação muito importante na imagem, já que pequenas alterações de movimento não afetam essa característica, além disso é uma das características mais importantes para se distinguir rapidamente conteúdos não-similares (MARQUES, 2006).

Existem diferentes modelos para representar as cores em imagens digitais. Um desses modelos de representação de cores é o modelo RGB, em que cada cor é representada pela combinação dos componentes vermelho (R do inglês *red*), verde (G do inglês *green*), azul (B do inglês *blue*) (GONZALEZ; WOODS, 2006). A Figura 3 ilustra o subespaço de cor, nela é possível observar que os extremos do cubo são as cores vermelho (1, 0, 0), verde (0, 1, 0) e azul (0, 0, 1), as cores secundárias magenta (1, 0, 1), ciano, (0, 1, 1), amarelo (1, 0, 1) e os valores de preto (0, 0, 0) na origem e branco (1, 1, 1) sendo o ponto mais distante da origem.

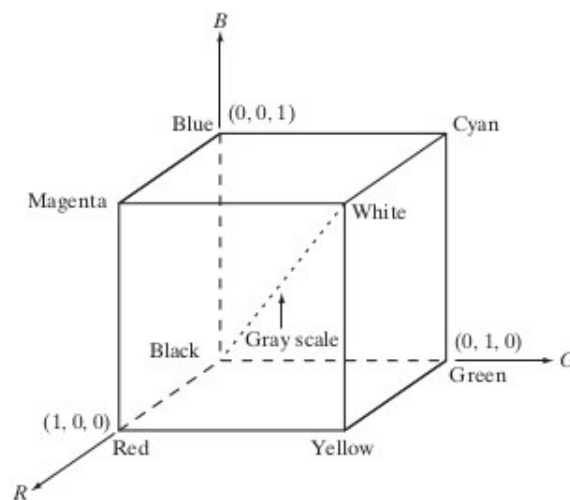


Figura 3 – Representação do subespaço de cores do modelo RGB (GONZALEZ; WOODS, 2006).

As imagens representadas no modelo RGB consistem em três matrizes, uma para cada componente, o número de *bits* para representar um *pixel* é chamando de profundidade de *bits*. Uma imagem de um componente RGB isolado possui um valor de profundidade de bits equivalente a 8, logo os valores variam no intervalo de  $[0, 255]$ . Uma imagem RGB contém uma profundidade de *bits* com valor igual a 24, logo o número de cores possíveis para a representação de um pixel chega ao valor de  $(2^8)^3 = 16.777.216$  possibilidades (GONZALEZ; WOODS, 2006).

Outros modelos como o CMY, que utiliza dos valores ciano (C, do inglês *Cyan*), magenta (M, do inglês *Magenta*) e amarelo (Y, do inglês *Yellow*), e o HSI, que utiliza os valores de matiz (H, do inglês *hue*), saturação (S, do inglês *saturation*), e intensidade (I,

do inglês *intensity*), são outras formas de representar as cores presentes em uma imagem.

### 1.2.1.1 Conversão de imagem colorida para imagem em tons de cinza

Matematicamente, pode-se definir uma conversão de imagens coloridas para escala de cinza como uma função  $G$  que recebe uma imagem  $R^{3mn}$  e produz como saída uma imagem monocromática  $R^{mn}$ . Em modelos de representação de cores nos quais cada canal é representado por 8 *bits*, com isso um *pixel* em uma matriz e em escala de cinza é representado por valores no intervalo de  $[0, 255]$  (MACÊDO; MELO; KELNER, 2015). Na escala de tons de cinza 0 representa o preto e 255 representa o branco. A imagem gerada pela aplicação da função de conversão gera uma matriz  $m \times n$  com valores variando entre 0 e 255.

A Tabela 1 representa alguns algoritmos que atendem a solução de conversão de imagens coloridas no modelo RGB para escala de cinza apresentados por Kanan e Cottrell (2012).

Nome da função	Fórmula
$G_{Intensity}$	$\frac{(R+G+B)}{3}$
$G_{Luminance}$	$0.3R + 0.59G + 0.11B$
$G_{Value}$	$max(R, G, B)$
$G_{Luster}$	$\frac{max(R,G,B)+min(R,G,B)}{2}$

Tabela 1 – Tabela com funções para conversão de *pixels* no modelo RGB para escala de cinza.

### 1.2.1.2 Formas

Formas se referem à silhueta dos objetos presentes na imagem. Para utilizar características de formas é necessário, primeiramente, segmentar (separar) os objetos (SILVA, 2011). A segmentação dos objetos presentes na imagem pode auxiliar na busca por conteúdo similares, de tal maneira que são determinadas imagens que compartilham a mesma coleção de objetos.

Para a extração de formas, geralmente é feito o processamento da imagem com um algoritmo de segmentação, de forma que o objeto de interesse fique isolado, e sejam extraídas características a partir desta área (PONCIANO-SILVA, 2009). É realizado um pré-processamento para a identificação de bordas, após isso a forma de um objeto pode ser caracterizada utilizando-se da área, curvatura, assinatura (sequência de distâncias para o centro da borda), entre outras características no segmento identificado (MARQUES, 2006).

### 1.2.1.3 Textura

A textura não é definida por *pixels* unitários, mas por regiões caracterizadas por terem padrões uniformes, ou seja, uma região que possui um grande número de *pixels* densamente visíveis e uniformemente dispostos (MARQUES, 2006). A textura possui qualidades como periodicidade e escala, e pode ser descrita em termos de direção, contraste e rigidez (TAMURA H., 1978).

Existem duas formas de analisar a textura de uma imagem, a abordagem estruturada e a abordagem estatística. A abordagem estatística caracteriza a variação de contraste, intensidade e direção, já a análise estrutural capta elementos na imagem, determina a forma e sua ligação com outros componentes a partir da posição capturada, de forma que são determinados número de elementos vizinhos, número de elementos em uma região, e se estes são representados homogeneamente ou em camadas (MARQUES, 2006).

### 1.2.2 Descrição de imagens usando Histograma de Cor

Uma maneira bem simples de descrever uma imagem é através do uso de Histograma de Cor, que representa a frequência com que as cores ocorrem numa imagem. Esse tipo de extração de característica se mostra eficaz para operações de processamento digital de imagens em tempo real (GONZALEZ; WOODS, 2006), já que são simples de serem computados.

Um histograma é calculado a partir de uma função discreta:

$$h(r_k) = n_k, \quad (1.1)$$

em que  $r_k$  representa o  $k$ -ésimo valor de intensidade e  $n_k$  é o número de *pixels* de uma imagem que apresentam aquela intensidade. Essa função varia do intervalo  $[0, L - 1]$ , tal que  $L$  é o maior valor de intensidade capturado nesta representação da imagem. (GONZALEZ; WOODS, 2006). Uma imagem de 8 bits, por exemplo, consegue representar valores inteiros entre 0 (preto) e 255 (branco). Neste caso  $L = 255$ , portanto o histograma seria um vetor de 256 posições.

De acordo Gonzalez e Woods (2006), uma prática comum ao se utilizar histogramas para representar uma imagem é normaliza-los de modo que os valores armazenados passam a ser a frequência ou probabilidade desse valor de intensidade ocorrer na imagem. A a função 1.2 representa o cálculo efetuado,

$$P_r(r_k) = \frac{n_k}{N}, \quad (1.2)$$

em que  $N$  representa o número total de *pixels* da imagem. Esta função retorna o valor de distribuição de probabilidade para um determinado valor  $r$ . O somatório do histograma normalizado resulta em um valor igual a 1.

De um modo geral, existem duas abordagens para se extrair as características visuais de uma imagem: as abordagens globais e as abordagens locais (MARQUES, 2006).

As técnicas de extração de características baseadas nas abordagens globais são aquelas que consideram a imagem como um todo, ou seja, não considera localmente as regiões da imagem (HASSABALLAH; ALI; ALSHAZLY, 2016). O histograma de cores é um exemplo de técnica de extração de características com abordagem global, já que ele considera as distribuições de cores considerando toda a estrutura da imagem. Porém, se o histograma for computado considerando apenas uma região bem específica da imagem, então a abordagem se torna local.

A vantagem das técnicas de extração de características globais é o baixo custo computacional, já que são métodos de execução rápidos, com poucos laços iterativos e sem a necessidade de segmentar regiões e/ou objetos na imagem. Porém, os descritores globais são muito genéricos, uma vez que estes podem falhar em identificar características visuais que possam ser importantes na análise da similaridade entre as imagens (ALKHAWLANI; ELMOGY; EL-BAKRY, 2015).

Por outro lado, as abordagens locais são capazes de caracterizar regiões de destaque de uma imagem. Algumas técnicas conseguem realizar essa caracterização de maneira invariável à mudança de iluminação ou ponto de vista. Com isso a imagem é representada em características locais de um conjunto de regiões da imagem, que são chamados de regiões de interesse (SILVA, 2011).

As abordagens locais consomem mais esforço computacional comparadas às técnicas de abordagens globais, já que uma imagem pode ter várias características locais, porém o resultado é mais discriminante em problemas de comparação e reconhecimento de objetos (HASSABALLAH; ALI; ALSHAZLY, 2016).

A Figura 4 ilustra um exemplo onde a abordagem global é desvantajosa em relação às abordagens locais. Nota-se que as três imagens são visualmente diferentes, porém apresentam o mesmo histograma das imagens em escala de cinza. Para evidenciar as diferenças entre as três imagens seria necessário utilizar alguma abordagem de caracterização local.

### 1.3 Medidas de Similaridade

A busca por dissimilaridade se refere à tarefa de buscar por elementos similares a partir de outro elemento-consulta. Dois elementos  $O_i$  e  $O_j$  são ditos similares quanto menor for o valor de dissimilaridade entre eles (PONCIANO-SILVA, 2009).

Para medir a dissimilaridade entre dois elementos, geralmente usa-se uma função de distância. Essa função de distância diz quão próximos dois elementos estão um do

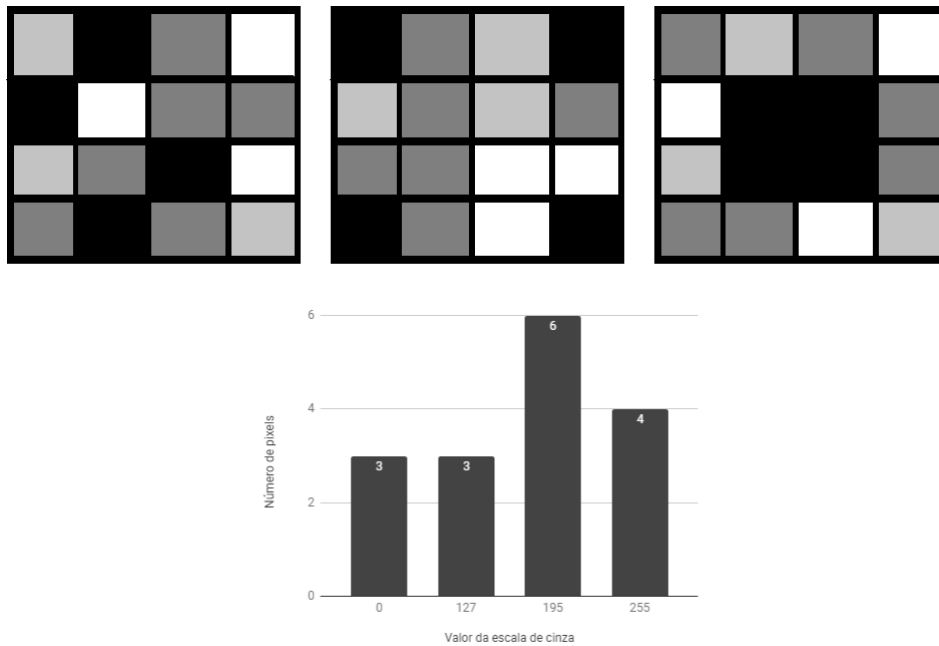


Figura 4 – Imagens visualmente diferentes, porém com o mesmo histograma das imagens em escala de cinza.

outro.

Uma função de distância  $d(O_i, O_j)$  é definida como **métrica** quando atende à três propriedades:

1. **Simetria:**  $d(O_1, O_2) = d(O_2, O_1)$
2. **Não-negatividade:**  $0 < d(O_1, O_2) < \infty$ , se  $O_1 \neq O_2$  e  $d(O_1, O_1) = 0$
3. **Desigualdade triangular:**  $d(O_1, O_2) \leq d(O_1, O_3) + d(O_3, O_2)$

em que,  $O_1, O_2, O_3$ , se referem a três elementos. Neste trabalho, esses elementos são os vetores de características extraídos das imagens.

A função de distância utilizada interfere diretamente na precisão do ranqueamento das imagens, logo deve-se escolhê-la ao levar em consideração as características utilizadas (MARQUES, 2006).

### 1.3.1 Funções de distância de Minkowski

A distância de Minkowski é uma métrica que inclui algumas outras métricas de distância de forma generalizada (KRUSKAL, 1964) e é amplamente utilizada em sistemas de recuperação de imagens.

Dados dois vetores de características  $A = (a_1, a_2, a_3, \dots, a_n)$  e  $B = (b_1, b_2, b_3, \dots, b_n)$ , a distância de Minkowski entre esses dois vetores é definida como:

$$L_P(A, B) = \sqrt[p]{\sum_{i=1}^n |a_i - b_i|^p}, \quad (1.3)$$

em que  $n$  representa o comprimento (dimensionalidade) dos vetores.

Ao variar o valor de  $p$  na Eq. 1.3 são obtidas as variações das equações pertencentes à família das funções de distâncias de Minkowski (PONCIANO-SILVA, 2009). Abaixo estão descritas as distâncias Manhattan e Euclidiana, que correspondem, respectivamente, à distância de Minkowski quando  $p = 1$  e  $p = 2$ :

- Distância de Manhattan: corresponde ao somatório do módulo das diferenças de cada elemento dos vetores:

$$L_1(A, B) = \sum_{i=1}^n |a_i - b_i| \quad (1.4)$$

- Distância Euclidiana: corresponde a raiz quadrada da soma das diferenças quadráticas de cada elementos dos vetores:

$$L_2(A, B) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (1.5)$$

## 1.4 Avaliação do Desempenho de Sistemas CBIR

Um sistema CBIR tem o intuito de recuperar imagens similares à imagem informada como consulta. A eficácia desse sistema deve ser mensurada de forma a verificar se as imagens recuperadas pelo sistema satisfazem a expectativa do usuário, ou seja, se são relevantes. A seguir serão descritas duas medidas.

### 1.4.1 Precisão

Uma maneira bem simples de calcular a eficácia de sistemas CBIR é calcular a precisão de uma consulta  $R$ , que é dada pela seguinte fórmula:

$$Precision(R) = \frac{x}{k} \quad (1.6)$$

em que  $x$  é a quantidade de imagens relevantes dentro de um conjunto com  $k$  imagens.

Para exemplificar, suponha que dentre as 20 primeiras imagens retornadas pelo sistema CBIR (top-20), apenas 5 imagens são realmente similares à imagem-consulta.

Então, para essa consulta a precisão foi de  $5/20 = 0,25$ , ou seja, 25%. Se dentro deste *top-20* todas as imagens são relevantes, então a precisão da consulta seria de 100%.

### 1.4.2 Mean Average Precision (MAP)

A medida *Precision* apresentada na seção anterior é utilizada apenas para medir a eficácia de uma consulta. Para medir a eficácia considerando um conjunto maior de consultas, então utiliza-se a *Mean Average Precision* (MAP). (MANNING; RAGHAVAN; SCHÜTZE, 2008).

De acordo com Manning, Raghavan e Schütze (2008), para uma única consulta a *Average Precision* (AP) é a média da precisão obtida para o conjunto dos  $k$  documentos similares obtidos após a recuperação de cada documento relevante. Considere que  $q_j \in Q$  é o conjunto de consultas realizadas até chegar a um resultado considerado satisfatório, sendo  $q_j$  igual a  $\{d_1, \dots, d_{m_j}\}$ , e que  $R_{jk}$  é o conjunto de resultados retornados até que o documento  $d_k$  esteja presente na consulta, então a fórmula do *MAP* se dá por

$$MAP(Q) = \frac{1}{N} \sum_{j=1}^N \frac{1}{m_j} \sum_{k=1}^{m_j} Precision(R_{jk}), \quad (1.7)$$

em que  $N$  representa o tamanho do conjunto  $Q$  e  $m_j$  é o número de iterações de consulta.



## 2 Realimentação de Relevâncias

Segundo [Manning, Raghavan e Schütze \(2008\)](#), a proposta das técnicas de Realimentação de Relevância (RF, do inglês *Relevance Feedback*) consiste em envolver o uso da opinião do usuário sobre os resultados de uma consulta, com o intuito de compreender o desejo do usuário e, assim, tornar o sistema mais próximo do resultado esperado pelo usuário.

### 2.1 Funcionamento

As etapas básicas do funcionamento do processo de realimentação de relevância em um sistema de busca por imagens estão esquematizadas pela Figura 5. A sequência desses passos é definida como:

1. O usuário informa um objeto (uma imagem) para consulta;
2. O sistema retorna um primeiro conjunto de imagens julgadas como similares a partir da característica de baixo nível identificada automaticamente pelo sistema;
3. O usuário informa para o sistema quais imagens são relevantes e não-relevantes considerando sua perspectiva;
4. Uma nova busca é realizada com as informações do *feedback* do usuário na etapa anterior;
5. O sistema retorna um novo ranking de imagens a partir da consulta modificada;
6. Se o usuário julgar o resultado retornado satisfatório ele encerra a consulta, se não ele retorna novamente para o passo 3.

Este processo é iterativo nas etapas 3 a 5. A expectativa é que a cada iteração o sistema retorne melhores resultados, porém é altamente recomendado que o processo termine em poucas iterações. Uma vantagem desta técnica é a possibilidade de o usuário expressar a necessidade de consulta sem conhecer a fundo os detalhes de representação da imagem adotados pelo sistema ([SILVA, 2011](#)).

### 2.2 Algoritmos de Realimentação de Relevâncias

De uma maneira geral, os algoritmos para realizar a realimentação de relevância envolvem a criação de pseudo-objetos a partir do *feedback* fornecido pelo usuário e a

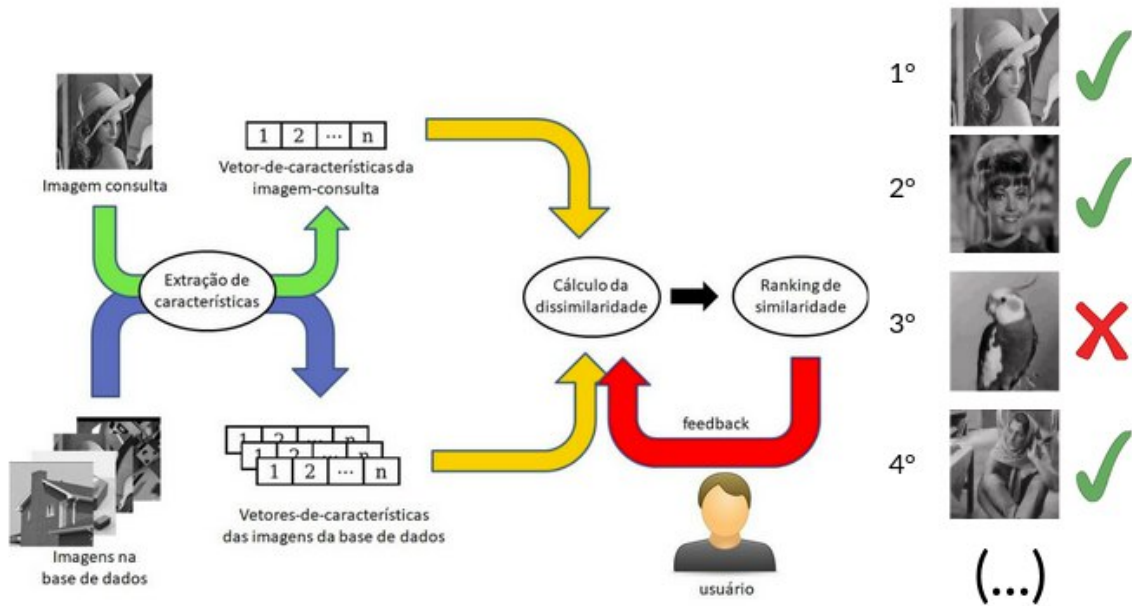


Figura 5 – Processo de Realimentação de Relevâncias em um Sistemas de Busca por Imagens Similares

utilização destes novos objetos nas consultas subsequentes (MARQUES, 2006). A seguir, serão detalhadas três técnicas de realimentação de relevâncias encontradas na literatura:

- Movimentação do Objeto de Consulta (QPM, do inglês *Query Point Movement*)
- Múltiplos Centros de Consulta
- Agregação de Rankings

### 2.2.1 Movimentação do Objeto de Consulta (QPM)

Uma estratégia apresentada por Rocchio (1971), é a de usar a perspectiva do usuário para adequar uma consulta, ou seja, o usuário informa quais documentos recuperados são relevantes e não-relevantes, com isso o *ponto de consulta* é modificado para se aproximar dos resultados relevantes e distanciar-se dos não-relevantes. Tal técnica é nomeada como Movimentação do Objeto de Consulta (QPM, do inglês *Query Point Movement*) (MARQUES, 2006).

A fórmula apresentada por Rocchio (1971) descreve um processo de QPM e é descrita por

$$\bar{Q}_1 = \bar{Q}_0 + \frac{1}{n_1} \sum_{i=1}^{n_1} \bar{R}_i - \frac{1}{n_2} \sum_{i=1}^{n_2} \bar{S}_i, \quad (2.1)$$

em que  $\bar{Q}_0$  é o vetor da imagem de consulta,  $n_1$  é o número de imagens relevantes informados pelo usuário e  $n_2$  o número de imagens não relevantes,  $R_i$  e  $S_i$  representam os

$i$ -ésimos vetores de características das imagens relevantes e não relevantes informadas, respectivamente. O novo vetor  $\bar{Q}_1$  é o que será utilizado na próxima consulta.

Rocchio (1971) cita que o processo pode ser iterativo e a cada novo vetor de consulta gerado, o usuário pode alimentar o algoritmo com um novo conjunto de imagens relevantes e não relevantes retornados em  $Q_i$ . Considerando que a fórmula apresentada por ele define uma função  $f(\bar{Q}, R, S)$ , uma expressão geral para esse processo iterativo se dá por

$$\bar{Q}_{i+1} = f(\bar{Q}_i, R_i, S_i). \quad (2.2)$$

Uma modificação geralmente adotada na fórmula é a atribuição de pesos à imagem informada, e às imagens relevantes e não relevantes informadas pelo usuário (MARQUES, 2006). Assim, a expressão seria

$$\bar{Q}_1 = \alpha \bar{Q}_0 + \frac{\beta}{n_1} \sum_{i=1}^{n_1} \bar{R}_i - \frac{\gamma}{n_2} \sum_{i=1}^{n_2} \bar{S}_i, \quad (2.3)$$

em que os novos elementos  $\alpha$ ,  $\beta$  e  $\gamma$  são os pesos escolhidos. Pelo fato de o número de imagens não relevantes normalmente ser maior que as relevantes, a constante  $\gamma$  normalmente é menor que  $\beta$ .

A Figura 6 apresenta o processo descrito por tal função iterativa. O passo em (a) representa a primeira consulta de um usuário, (b) representa o primeiro apontamento das imagens relevantes (+) e não relevantes (-), (c) representa mais uma iteração aonde a maioria dos resultados retornados são considerados relevantes.

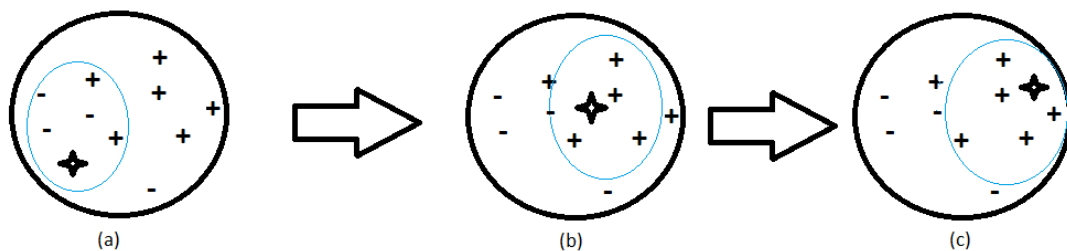


Figura 6 – Movimentação de um objeto de consulta usando o modelo de Query Point Movement iterativo

### 2.2.2 Múltiplos centros de consulta

Informar mais de um objeto-consulta pode melhorar o resultado retornado em uma consulta. Essa abordagem é conhecida como *multi-point approaches*: o usuário fornece várias imagens-consulta e, então, as dissimilaridades são computadas considerando todas as imagens fornecidas como consulta. Assim, as distâncias individuais de cada imagem são agregadas em uma distância geral, diferentemente da consulta simples em que apenas uma distância é utilizada na consulta (ORTEGA-BINDERBERGER; MEHROTRA, 2003).

Segundo Marques e Traina (2006), o uso de várias imagens como consulta fornece visões diferentes do que o usuário deseja recuperar e, assim, há uma melhora no desempenho entre o que o usuário deseja e o que o sistema retorna como resultado para o usuário. A Figura 7 mostra que ao utilizar a proposta de múltiplos centros de consultas e uma função de distância que agrega as distâncias individuais, a quantidade de objetos fornecidos e as posições destes influenciam no formato da consulta no espaço de características (ORTEGA-BINDERBERGER; MEHROTRA, 2003).

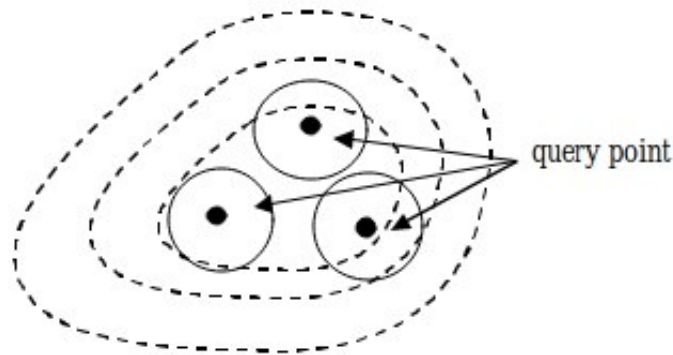


Figura 7 – Representação dos múltiplos centros de consulta (ORTEGA-BINDERBERGER; MEHROTRA, 2003)

A técnica *Relevance Feedback Multiple Projection* (RFMP), proposta por Marques e Traina (2006), utiliza a ideia de Múltiplos Centros de Consulta em seu algoritmo e a criação de um objeto modificado a partir da técnica *Relevance Feedback Projection* (RFP).

#### 2.2.2.1 Relevance Feedback Projection (RFP)

A técnica RFP consiste na criação de um objeto modificado através do *feedback* do usuário. Marques (2006) cita que esta técnica analisa cada atributo do vetor separadamente, dividindo-o em três faixas, sendo essas faixas superior e inferior para os objetos não-relevantes e a faixa central para os objetos relevantes. O cálculo da faixa de relevantes se dá pela equação

$$\delta = \max R_i - \min R_i, \quad (2.4)$$

em que  $\max R_i$  e  $\min R_i$  representam o maior e menor valor da feature  $i$  em todos objetos relevantes retornados.

A faixa de relevantes contém o intervalo que caracteriza o grau da similaridade que o usuário espera atingir (MARQUES, 2006).

As faixas superiores e inferiores  $f_{upper}$  e  $f_{lower}$  atuam como uma pequena tolerância que limitam o intervalo da faixa de relevantes. O cálculo dessas faixas se dá pelas equações

$$f_{upper} = \max R_i + \alpha * \delta \quad (2.5)$$

e

$$f_{lower} = \min R_i - \alpha * \delta, \quad (2.6)$$

em que  $\alpha$  é um valor que ajusta o agrupamento das imagens relevantes no espaço de características. Se as características extraídas agrupam boa parte dos objetos relevantes então  $\alpha$  assume um valor menor, porém se estes estiverem mais espalhados no espaço de características adota-se um valor de  $\alpha$  maior (MARQUES, 2006). O valor adotado de  $\alpha$  em Marques (2006) é de 2,3.

Para o cálculo das projeções é preciso gerar  $p_1$  e  $p_2$ , estes definem o início das projeções. Estes pontos projetados utilizam a realimentação positiva da consulta. Com base na realimentação negativa, procura-se projetar dois valores no intervalo de relevantes, o tamanho das faixas inferiores e superiores de objetos não relevantes é representado por  $t_1$  e  $t_2$ .

O cálculo das projeções se dá pelas equações

$$p_1 = \min R_i - \delta, \quad (2.7)$$

$$p_2 = \max R_i + \delta, \quad (2.8)$$

$$t_1 = |\min S_i - p_1|, \quad (2.9)$$

$$t_2 = |p_2 - \max S_i|, \quad (2.10)$$

$$projP_1 = \min R_i + t_1, \quad (2.11)$$

$$projP_2 = \max R_i - t_2. \quad (2.12)$$

Os valores de  $projP_1$  e  $projP_2$  são utilizados para gerar o valor de nova característica projetada. Se  $\min S_i$  for menor que o limite inferior da faixa de objetos relevantes,  $projP_1$  passará a assumir o valor de  $\min R_i$ . De maneira análoga o mesmo vale para  $projP_2$ , caso  $\max S_i$  seja menor que o valor da faixa superior de objetos relevantes, o valor de  $projP_2$  passa a ser  $\max R_i$  (MARQUES, 2006).

Observe na Figura 8 as etapas de projeções dado uma característica de diferentes imagens.

O novo valor, dada uma característica  $i$ , é calculado através da equação

$$newfeature_i = \frac{ProjP_1 + ProjP_2 + avg'}{3}, \quad (2.13)$$

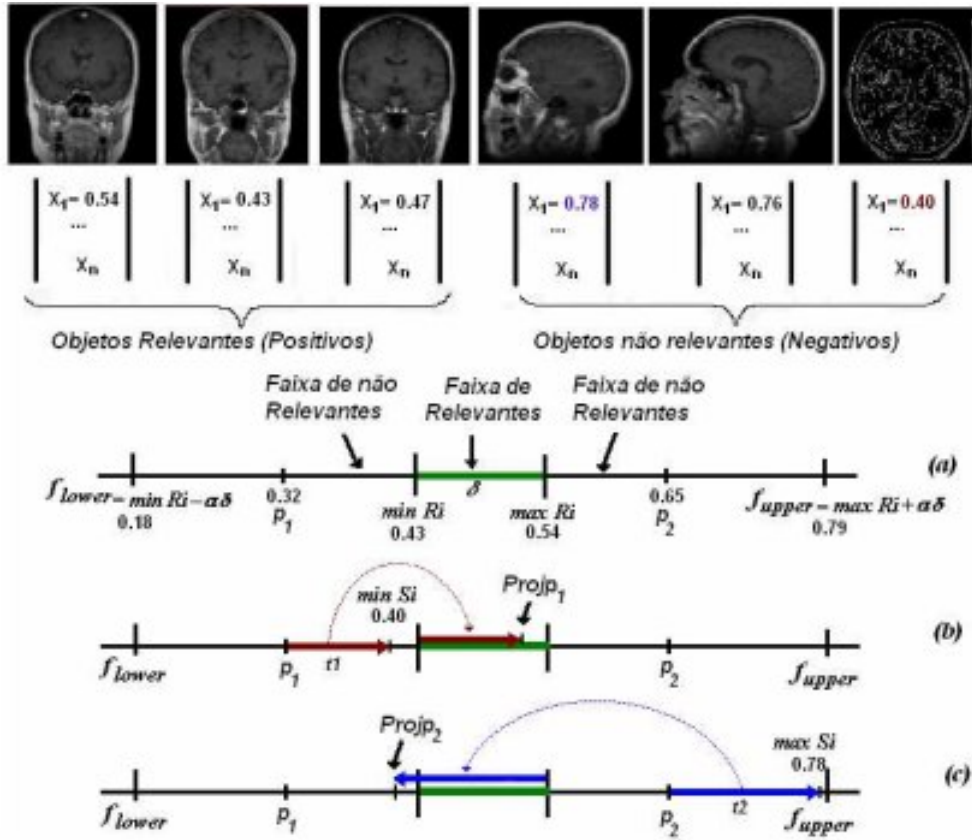


Figura 8 – Passos de construções das projeções (MARQUES, 2006).

em que  $avg'$  é definido como a média ponderada considerando os pesos e distâncias dos elementos relevantes pela expressão

$$avg' = \frac{\sum_{j=1}^r \left[ feature_i \left( \frac{\sum_j dist_j}{dist_j} \right) \left( \frac{weight_j}{\sum_j weight_j} \right) \right]}{2}, \quad (2.14)$$

em que  $r$  é o número de imagens relevantes, e as variáveis  $feature_i$ ,  $weight_j$  e  $dist_j$  representam o valor do atributo, o peso do objeto e a distância do objeto ao centro de consulta (MARQUES; TRAINA, 2006).

### 2.2.2.2 Relevance Feedback Multiple Projection (RFMP)

Nesta técnica proposta em Marques (2006), o usuário fornece uma imagem-consulta e, a partir de um *ranking* inicial, ele seleciona as imagens relevantes. A técnica RFMP utiliza o vetor de características das imagens relevantes e um novo vetor adicional denominado de *phantom object* gerado pela técnica RFP como centros de consulta para uma nova consulta. Essa nova consulta considera a distância de cada imagem para cada um dos centros de consultas.

A Figura 9 representa os passos da técnica RFMP: dado um objeto de consulta  $Q$  e  $r$  objetos relevantes selecionadas pelo usuário a partir de um ranking inicial, o algoritmo primeiramente computa um novo objeto (*phantom object*); após isso, o algoritmo realiza um busca usando esses elementos, gerando  $r + 1$  rankings denotados por:  $O_1, O_2, O_3, \dots, O_{r+1}$ .

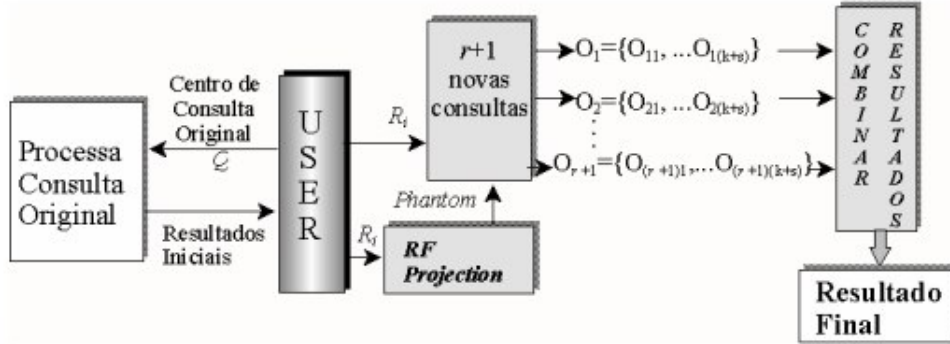


Figura 9 – Passos da técnica Relevance Multiple Point Projection (MARQUES; TRAINA, 2006)

A união dos conjuntos retornados pela consulta de um objeto  $Q$  é representado por  $O$ . Para que o resultado seja ordenado, calcula-se a distância entre cada descritor do conjunto resposta  $O_i$  e dos descritores do centro de consulta correspondente (MARQUES; TRAINA, 2006). Uma nova distância é computada considerando os pesos informados pelo usuário e a distância original entre cada objeto e o centro de consulta inicial (BIGIO, 2008). A partir dessa nova distância ( $d'$ ) os objetos com o menor valor de  $d'$  são retornados. O cálculo da distância combinada é representado por

$$d' = \frac{d \times weight_i}{\sum_i weight_i} + \frac{d \times \sum_i dist_i}{dist_i}, \quad (2.15)$$

em que  $d$  é a distância do objeto da nova consulta ao centro de consulta,  $weight_i$  representa o peso dado pelo usuário, e  $dist_i$  é a distância original do objeto relevante ao centro de consulta original  $Q$  (MARQUES; TRAINA, 2006).

### 2.2.3 Agregação de Rankings

Agregação de Ranking consiste em obter um único ranking a partir da combinação de diferentes rankings gerados pelo mesmo conjunto de dados. Em outras palavras, suponha dois rankings  $R_1 = \{B, D, E, F, G, A, C\}$  e  $R_2 = \{A, C, D, E, F, G, B\}$ . Estes rankings não precisam, necessariamente, serem gerados a partir do mesmo objeto de consulta. A proposta das técnicas baseadas em Agregação de Rankings é gerar um novo ranking  $R^*$  a partir da combinação dos rankings  $R_1$  e  $R_2$ . Pedronette e Torres (2011) propõe um

algoritmo capaz de ordenar *rankings* retornados de diferentes consultas de sistemas CBIR de forma a gerar um consenso, o que mostra a aplicabilidade deste conceito nesse tipo de sistema.

As técnicas que realizam a Agregação de Rankings podem ser divididas em duas abordagens: os métodos baseados em *score* e os baseados em ordem de *ranking*. Nos métodos baseados nos *scores*, o novo ranking é obtido a partir da distância de cada elemento ao objeto consulta, enquanto que os métodos baseados na ordem do *ranking*, analisam apenas a posição dos elementos no *ranking* (RENDA; STRACCIA, 2003).

A técnica combSUM, apresentada em Shaw e Fox (1994), é uma maneira simples de realizar a agregação de rankings usando a abordagem baseada em *scores*. Para ilustrar o funcionamento dessa técnica, considere um conjunto formado pelos elementos {A, B, C}. A Tabela 2.2.3 mostra três rankings  $R_1$ ,  $R_2$  e  $R_3$  gerados com essas três imagens. O objeto consulta utilizado para gerar esses rankings não é uma informação relevante. Os valores numéricos correspondem à distância de cada imagem naquele ranking. O ranking agregado  $R^*$ , que está exibido na última coluna, representa a soma das distâncias de cada imagem em cada ranking. Para este exemplo, o elemento mais similar seria o B, pois obteve a menor distância nos três rankings analisados.

#	$R_1$	$R_2$	$R_3$	$R^*$
A	0	0,1	0,8	<b>0,9</b>
B	0,1	0	0,6	<b>0,7</b>
C	0,3	0,9	0,4	<b>1,6</b>

Tabela 2 – Exemplo da agregação de *ranking* utilizada pela técnica combSUM.

As técnicas de Múltiplos Centros de Consulta (apresentadas na seção anterior) quando combinadas com a técnica de Agregação de Ranking, possibilitam gerar um novo ranking mais próximo da expectativa do usuário. Neste caso, os rankings poderiam ser obtidos usando as imagens relevantes selecionados pelo usuário e combinados usando a técnica combSUM.



## 3 Atividades Desenvolvidas

Esse capítulo apresenta as atividades desenvolvidas ao longo do período de realização deste trabalho.

Inicialmente foi desenvolvido um protótipo de CBIR para a realização de testes comparativos. Os detalhes técnicos da construção desse sistema serão descritos na seção 3.1. Em seguida, na seção 3.2, serão apresentadas o descritor e medida de similaridade adotadas para realizar a busca por imagens similares pelo sistema desenvolvido. Na seção 3.3 serão descritas as abordagens de realimentação de relevâncias utilizadas na realização dos testes comparativos.

Por fim, na seção 3.4 será apresentada uma nova proposta de algoritmo que foi desenvolvida para realizar a realimentação de relevâncias. Essa nova abordagem, denominada de **Realimentação de Relevâncias com Agregação de *Rankings* (RFRA)** é uma versão combinada das técnicas de Agregação de Rankings e Múltiplos Centros de Consultas.

### 3.1 Detalhes Técnicos de Implementação do Sistema

#### 3.1.1 Metodologia de Desenvolvimento do Software

Segundo (SOMMERVILLE, 2010), metodologia de *software* é o conjunto de práticas recomendadas para o desenvolvimento de um *software*.

No *Kanban* há um quadro que são anexadas tarefas e estas passam por estados de desenvolvimento, com isso o trabalho fica dividido em partes, deixa transparente o desenvolvimento do produto e usa controle de cronograma (KNIBERG, 2010).

Foi adotada a metodologia *Kanban* pela simplicidade desta e pela equipe de desenvolvimento consistir apenas de um desenvolvedor. O *Kanban* utilizado consiste em três estados, *To Do*, *Doing* e *Done*, estes são responsáveis por manter a rastreabilidade do que já foi produzido e o que está pendente.

#### 3.1.2 Ferramentas utilizadas

As ferramentas utilizadas se dividem em duas funcionalidades distintas: a lógica do sistema de consultas por conteúdo e a interface gráfica do sistema. A seguir cada seção abordará uma explicação geral e a ferramenta escolhida para desenvolvimento destas duas funcionalidades.

### 3.1.2.1 Lógica do sistema de consultas por conteúdo

Esta funcionalidade remete a toda regra de negócio envolvida na consulta baseada em conteúdo. É nela que são realizados cálculos, coleta de características e retorno do *ranking* de imagens.

Foi escolhida a linguagem de programação *Python* (v.3.6.3) para o desenvolvimento desta etapa. A linguagem *Python* é de alto nível, interpretada e possui um modelo de desenvolvimento *open-source* gerenciado pela organização *Python Software Foundation*. A escolha dessa ferramenta se dá pelo fato da facilidade em acesso a documentação de tudo que será utilizado para a construção do sistema CBIR, além da simplicidade na realização dos cálculos necessários para realização da busca por similaridade.

### 3.1.2.2 Interface gráfica

Para que a interação do usuário com a ferramenta seja de forma mais atrativa, foi construída uma interface onde pode-se fornecer a imagem para consulta e a partir disso iniciar o processo de realimentação de relevâncias até que o interesse deste seja alcançado. Existem diferentes ferramentas para a construção de telas, e algumas já possuem integração direta com o Python de forma *offline*. Dentre estas as mais conhecidas são o **QT**, disponibilizado por meio do *wrapper* *PyQt*, e a biblioteca **Kivy**, aonde ambas são utilizadas para desenvolvimentos de softwares multiplataforma.

Uma alternativa a estas é a biblioteca emergente **Electron**, a qual cria interfaces gráficas *desktop* a partir de componentes de desenvolvimento Web, logo as telas de aplicações nesse framework utilizam da linguagem de marcação *Hypertext Markup Language* (HTML), além de *Javascript* e seus frameworks derivados para construção de lógicas internas das telas.

A tecnologia *Electron* foi escolhida para o desenvolvimento da interface de usuário do sistema CBIR, pela facilidade da implementação de *templates* que atendam à solução proposta, por se tratar de componentes que utilizam das mesmas propriedades de uma aplicação web.

Na arquitetura do *Electron*, existe um processo principal que utiliza da tecnologia *Node* para criação de uma tela baseada no *browser Chromium*, esse processo é responsável pela inicialização dos *Renderer Processes*, que são os processos criados para renderização de cada página construída ([ELECTRONJS, 2013](#)).

A Figura 10 apresenta a página em que o usuário realiza o *upload* da imagem de consulta para que o programa realize o processamento e a busca na base de imagens, nesta também podem ser selecionados a base de busca e a técnica de realimentação de relevância a ser aplicada. A Figura 11 demonstra os resultados retornados em relação ao exemplo de imagem consulta presente na Figura 10.

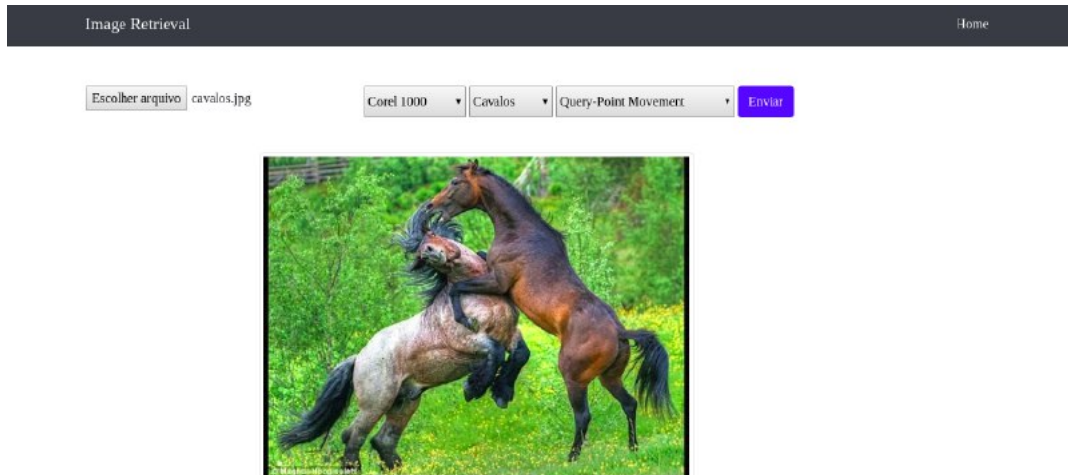


Figura 10 – Página inicial onde o usuário pode selecionar a imagem-consulta.

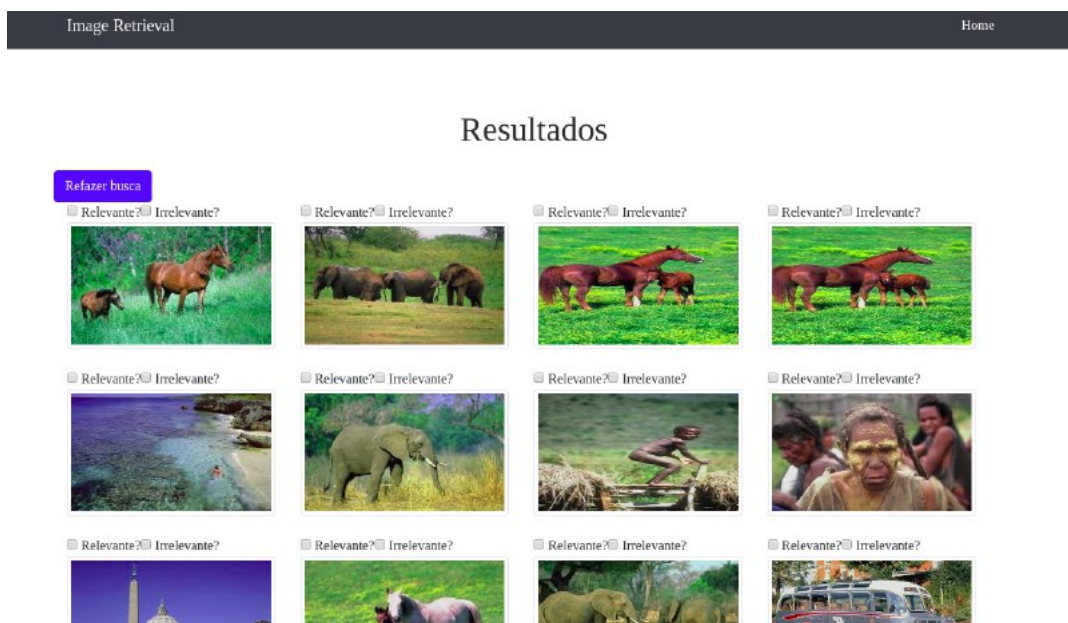


Figura 11 – Exemplo do resultado obtido usando a imagem-consulta da Fig. 10

As Figuras 12 e 13 mostram outra consulta utilizando outra classe presente na base de dados, é possível notar que os resultados não trazem tantas imagens relevantes.

### 3.1.3 Fluxo de funcionamento do sistema

O sistema desenvolvido possui o fluxo tradicional de funcionamento de um sistema CBIR, incluindo a etapa de realimentação de relevâncias, descrita no Capítulo 1.

O diagrama de atividades representado pela figura 14 ilustra todas as interações entre usuário, a interface (*front-end*) e o controle (*back-end*). As atividades de 1 a 8

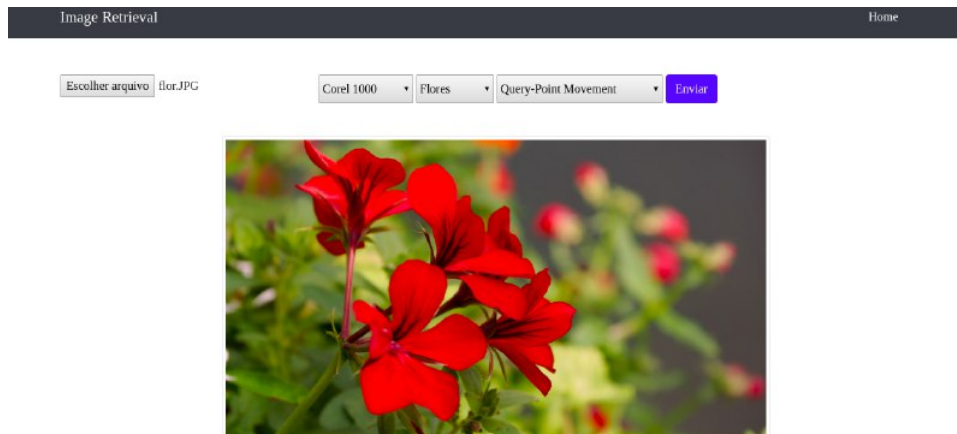


Figura 12 – Nova consulta utilizando uma imagem da classe FLOR.

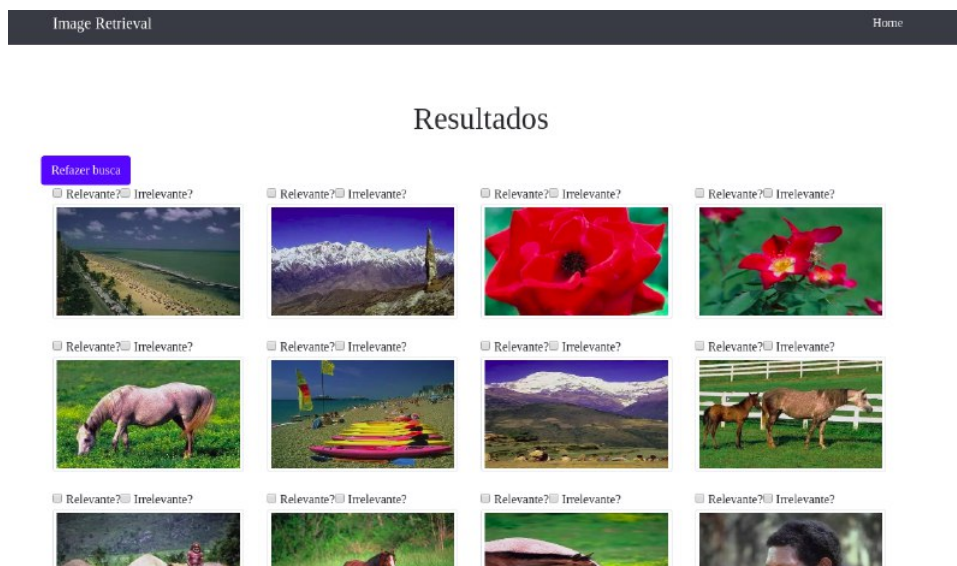


Figura 13 – Resultados recuperados pela consulta representada na Fig. 12

constituem etapas do sistema CBIR comum, e o *loop* que envolve as atividades 9 a 12 representa as etapas de realimentação de relevâncias.

### 3.2 Extração de Características e Medidas de Similaridade

Na parte de extração de características foram utilizadas as características de cores das imagens em nível de cinza. Para isso, as imagens coloridas (modelo RGB) foram convertidas para a escala de cinza usando a média dos valores de cada canal de cor. Assim, o vetor de características gerado a partir da extração de características é dado pelo o histograma em escala de cinza da imagem, que possui 256 posições e foi normalizado entre 0 e 1 para torná-la invariante à escala.

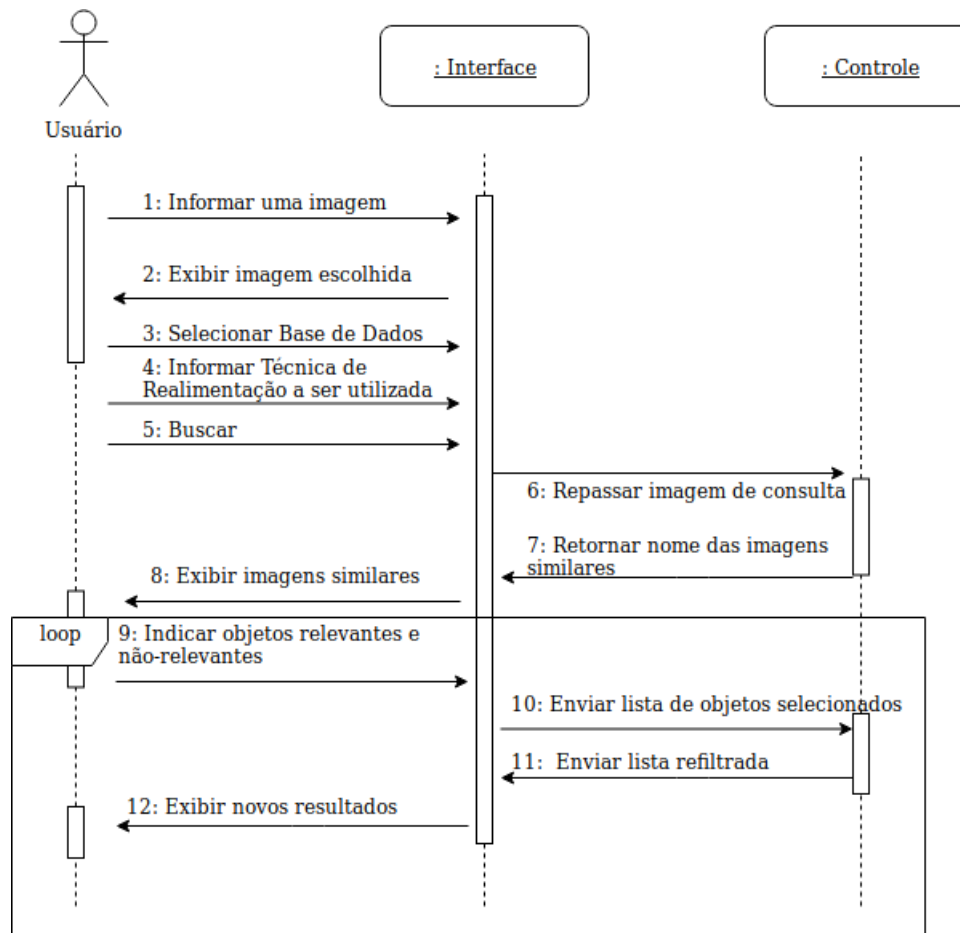


Figura 14 – Diagrama de Atividades do sistema desenvolvido

Para o cálculo da dissimilaridade, foi adotado a distância de Manhattan, descrita na equação 1.4, pelo fato desta função exigir menos operações do que a distância Euclidiana, e assim diminuir uma parte do tempo de processamento.

A interface do sistema desenvolvido retornar um *ranking* com as 10 imagens julgadas mais similares ao objeto consulta. Abaixo de cada imagem, existe um campo para o usuário informar se a imagem do ranking é relevante (ou não) e assim alimentar os algoritmos de realimentação de relevâncias.

### 3.3 Técnicas de Realimentação de Relevância Implementadas

Para a realização dos testes comparativos, foram utilizadas duas abordagens de realimentação de relevâncias:

- Movimentação do Objeto de Consulta
- Múltiplos Centros de Consulta

### 3.3.1 Movimentação de Objeto Consulta (QPM)

Para a implementação desta abordagem foi utilizado a técnica de Rocchio dada pela equação 2.1. O processo de realimentação dessa técnica é iterativo e, em cada iteração, o objeto de consulta  $Q_i$  a ser movimentado é modificado a partir do *feedback* do usuário sobre as imagens retornadas da consulta e se torna  $Q_{i+1}$ , que será utilizado na próxima etapa de realimentação.

Para essa técnica, o usuário deve definir, em cada iteração as imagens que ele considera relevantes e não-relevantes.

### 3.3.2 Múltiplos centros de consulta

A implementação da técnica explicada na seção 2.2.2 consiste em duas etapas de implementação, já que esta, além de realizar consultas para cada objeto considerado relevante para o usuário, também realiza uma consulta extra com um *phantom object* gerado pela técnica RFP, explicado na seção 2.2.2.1.

Primeiramente foi implementado a técnica RFP para a geração de um objeto modificado a partir do *feedback* do usuário. Para isso foi necessário determinar alguns valores para a implementação do algoritmo.

Da mesma forma que a técnica de QPM, para esta não foi adotada a solicitação dos pesos de cada imagem para ser informado pelo usuário, porém cada imagem julgada relevante recebeu um peso igual a 1 para a realização dos cálculos. A variável  $\alpha$  utilizada para determinar os limites dos valores ( $f_{upper}$ ,  $f_{lower}$ ) recebeu um valor fixo de 2,3 da mesma forma que os experimentos de Marques (2006).

Também foi gerada uma adaptação a equação para calcular o valor de uma característica  $i$  pertencente ao objeto RFP, a equação 2.14 tem um problema no caso da distância entre o objeto consulta e uma imagem da base ser igual a 0, levando a uma divisão por 0. A adaptação utilizada é a mesma equação apresentada em Bigio (2008) e está representada pela equação 3.1.

$$avg' = \frac{\sum_{j=1}^r \left[ feature_i \left( \frac{dist_j}{\sum_j dist_j} \right) \left( \frac{\sum_j weight_j}{weight_j} \right) \right]}{2} \quad (3.1)$$

Desta forma a técnica RFP foi implementada e o novo objeto gerado pôde ser utilizado para a realização de uma consulta, de forma que este resultado também fosse usado na distância combinada.

O cálculo da distância combinada utilizou a equação 2.15 e para este eram considerados apenas os objetos que tinham a distância diferente de 0 em relação ao objeto

consulta, para evitar uma divisão por zero.

### 3.4 Método Proposto: Realimentação de Relevâncias com Agregação de *Rankings* (RFRA)

Baseada na técnica de múltiplos centros de distância apresentada na seção anterior, foi proposta uma nova maneira de combinar as técnicas de Múltiplos Centros de Consulta e Agregação de Ranking para a combinação dos resultados retornados de cada consulta realizada. A ideia da nova proposta desenvolvida consiste em realizar uma nova busca usando como imagens-consultas as imagens selecionadas como relevantes pelo usuário em cada iteração.

O exemplo da Figura 15 auxilia na explicação da técnica desenvolvida: suponha que uma base de imagens seja composta por 10 imagens representadas pelas letras: A, B, C, D, E, F, G, H, I, J. Usando a imagem A como consulta, o sistema retorna no primeiro ranking as imagens B, C, D, E e F no top-5. Dentre essas 5 imagens, o usuário seleciona como relevantes apenas as imagens B, D e F. O sistema então realiza uma nova busca usando, individualmente, cada uma dessas três imagens como consulta e gerando, portanto, três novos *rankings*. A Tabela 3 mostra os valores das distâncias de cada imagem da base com relação às imagens-consulta B, D, F. Um novo ranking será gerado considerando a média das distâncias de cada imagem em cada ranking. Para este exemplo, o novo top-5 do ranking seria: D, F, B, I e J.

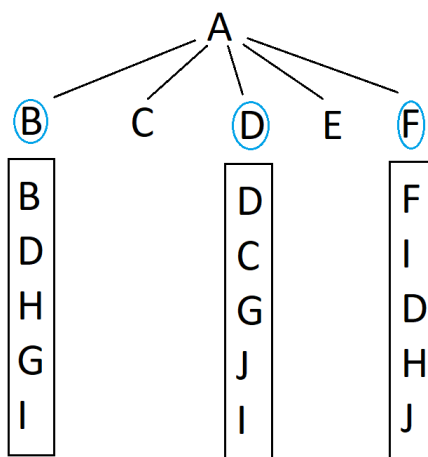


Figura 15 – A técnica proposta realiza uma nova consulta para cada imagem selecionada como relevante pelo usuário. Os rankings gerados por essas novas consultas são agregados para gerar um novo ranking para a imagem-consulta original.

A técnica escolhida para agregação dos rankings é uma variante da técnica comb-SUM apresentada na seção 2.2.3 do capítulo anterior. Na técnica proposta foi utilizada a média das distâncias das imagens em cada ranking ao invés do somatório.

#	<b>B</b>	<b>D</b>	<b>F</b>	<b>MÉDIA</b>
<b>B</b>	0	0,15	0,8	<b>0,31</b>
<b>D</b>	0,1	0	0,4	<b>0,16</b>
<b>F</b>	0,8	0,4	0	<b>0,4</b>
<b>H</b>	0,3	0,9	0,45	<b>0,55</b>
<b>G</b>	0,55	0,4	0,6	<b>0,51</b>
<b>I</b>	0,6	0,5	0,2	<b>0,43</b>
<b>C</b>	0,8	0,2	0,8	<b>0,6</b>
<b>J</b>	0,6	0,45	0,55	<b>0,53</b>

Tabela 3 – Matriz de cálculo do vetor de médias

Porém, para utilizar a técnica proposta, é necessário que o usuário selecione em cada iteração, pelo menos uma imagem como relevante. Caso isso não ocorra, não é possível gerar um novo ranking e a técnica não é capaz de evoluir. Para resolver este problema, foi adotado a abordagem de QPM, ou seja, se o usuário não selecionar nenhuma imagem relevante, o sistema aplica automaticamente a abordagem QPM para gerar um novo ranking, até que o usuário selecione uma imagem relevante e seja possível utilizar a técnica proposta.



## 4 Resultados Experimentais

Este capítulo apresenta os resultados dos testes experimentais realizados para avaliar o desempenho das técnicas de Realimentação de Relevâncias que foram apresentadas na seção 3.3 do capítulo anterior.

### 4.1 Bases de imagens

Para a realização dos testes experimentais foram utilizadas duas bases de imagens públicas: COREL1000 o Oxford Flowers.

A base de imagens COREL1000 contém 1000 imagens, divididas em 10 classes com 100 imagens em cada classe. Essas 10 classes são formadas por imagens de cenas naturais divididas nos seguintes grupos: *Africa*, *Cavalo*, *Elefante*, *Dinossauro*, *Montanha*, *Praia*, *Comida*, *Construções*, *Ônibus* e *Flores*. A Figura 16 ilustra alguma das imagens de cada categoria presente no banco de imagens.

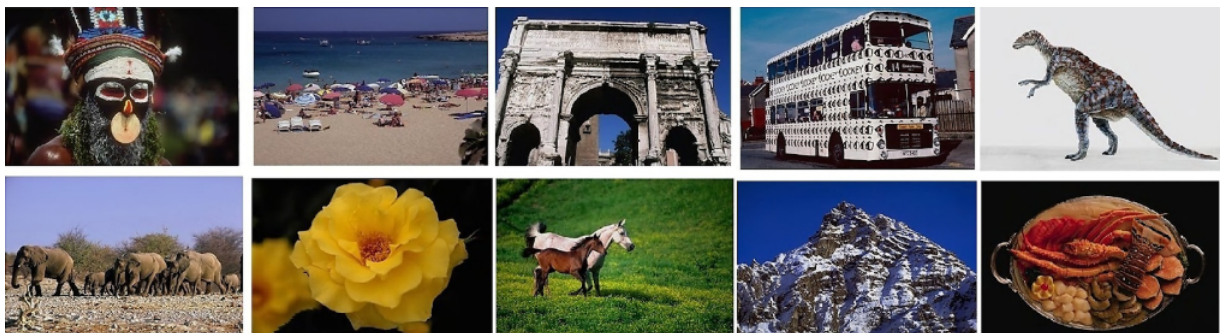


Figura 16 – Imagens de cada classe da base *Corel1000* (LI LIN FENG, 2016)

A base *Oxford Flowers* Nilsback e Zisserman (2006) contém 1360 imagens, divididas em 17 classes de flores diferentes. Cada classe possui igualmente 80 imagens. As classes da base são *Buttercup*, *Colt's Foot*, *Daffodil*, *Daisy*, *Dandelion*, *Fritillary*, *Iris*, *Pansy*, *Sunflower*, *Windflowers*, *Snowdrop*, *Lily Valley*, *Bluebell*, *Crocus*, *Tigerlily*, *Tulip*, *Cowslip*. A Figura 17 representa algumas imagens dessa base de dados.

### 4.2 Metodologia para a Avaliação de Desempenho

Para avaliar o desempenho dos testes experimentais individualmente as imagens de cada uma das bases foram utilizadas como imagem-consulta no sistema CBIR desenvolvido e a busca foi efetuada considerando as imagens da base da qual a imagem-consulta foi



Figura 17 – Imagens da base *OxfordFlowers* (PIAO et al., 2016)

extraída. Uma imagem retornada pela busca é considerada relevante se pertencer à mesma classe da imagem-consulta.

Considerando os experimentos sobre as técnicas de realimentação de relevâncias, foram executadas quatro iterações de realimentação na busca: o primeiro ranking gerado se refere à busca tradicional, os outros rankings foram obtidos pela aplicação das técnicas de realimentação de relevância.

Para medir o desempenho do sistema, a métrica considerada foi a precisão média por iteração em cada técnica de realimentação de relevância, analisando o *top* – 10 e *top* – 20 das imagens retornadas. A métrica MAP, explicada na seção 1.4.2, também foi utilizada para comparar a precisão geral entre as técnicas.

## 4.3 Resultados Obtidos

### 4.3.1 Sem Realimentação de Relevâncias

Uma primeira avaliação do sistema foi realizada sem utilizar nenhuma abordagem de realimentação de relevâncias.

O gráfico da Figura 18 apresenta a precisão média de cada classe numa primeira consulta sobre a base COREL1000 analisando as 10 imagens mais similares. É possível notar que há classes com resultados bastante satisfatórios como a classe DINOSSAURO (99% de precisão). Isso se deve ao fato de que o histograma dessas imagens serem bastante similares entre si e bem diferentes dos histogramas das imagens das outras classes.

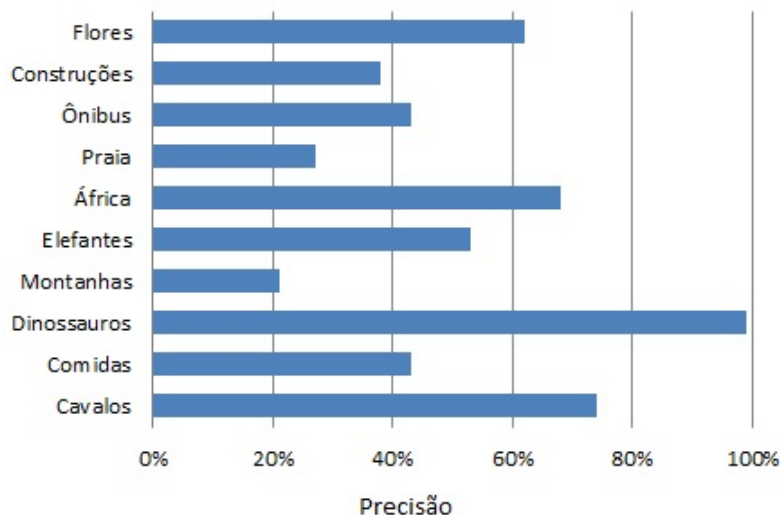


Figura 18 – Precisão da primeira consulta sobre cada classe da base COREL1000.

Por outro lado, considerando a base *Oxford Flowers*, a primeira avaliação (sem utilizar o processo de realimentação de relevância) mostra valores bem baixos de precisão. A Figura 19 apresenta os valores de precisão para cada classe quando analisados os 10 objetos mais similares ao objeto-consulta. O melhor resultado foi para a classe *Daisy* com 42% de precisão. Isso se deve ao fato de que o histograma de escala de cinza não é um bom descritor para esta base, pois existem muitas flores de classes diferentes com as mesmas intensidades em tons de cinza, o que leva a um mapeamento parecido de intensidade para os *pixels* quando passados para o nível de cinza.

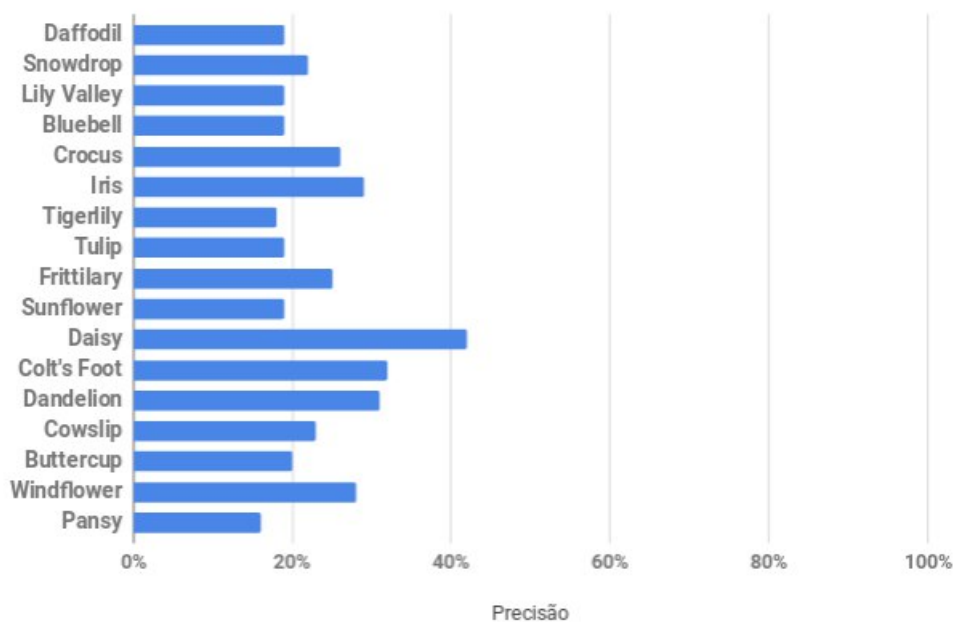


Figura 19 – Precisão média da primeira consulta sobre a base *Oxford Flowers*.

Em resumo, nota-se que ao utilizar apenas o histograma de intensidades em escala de cinza nem sempre é possível obter um resultado satisfatório, ou seja, usando apenas o histograma não é possível obter uma busca por imagens similares com boa precisão nos resultados. Por isso, neste caso, espera-se que a realimentação de relevância seja uma maneira de melhorar a precisão dos resultados.

### 4.3.2 Com Realimentação de Relevância

As Figuras 20 e 21 mostram os valores de precisão considerando, respectivamente, o top-10 e o top-20 da base COREL1000. A 1ª iteração é a precisão da consulta inicial (sem a realimentação de relevâncias) e as quatro iterações subsequentes representam os resultados com a aplicação de cada técnica de Realimentação de Relevância. Nota-se que de fato os valores de precisão aumentam a cada iteração usando as técnicas de Realimentação de Relevâncias.

Comparando o desempenho de cada uma das técnicas de Realimentação implementadas, a técnica QPM obteve o pior desempenho. A técnica proposta RFRA se saiu melhor quando analisado o top-10. Por outro lado, a técnica RFMP obteve melhor resultado considerando o top-20.

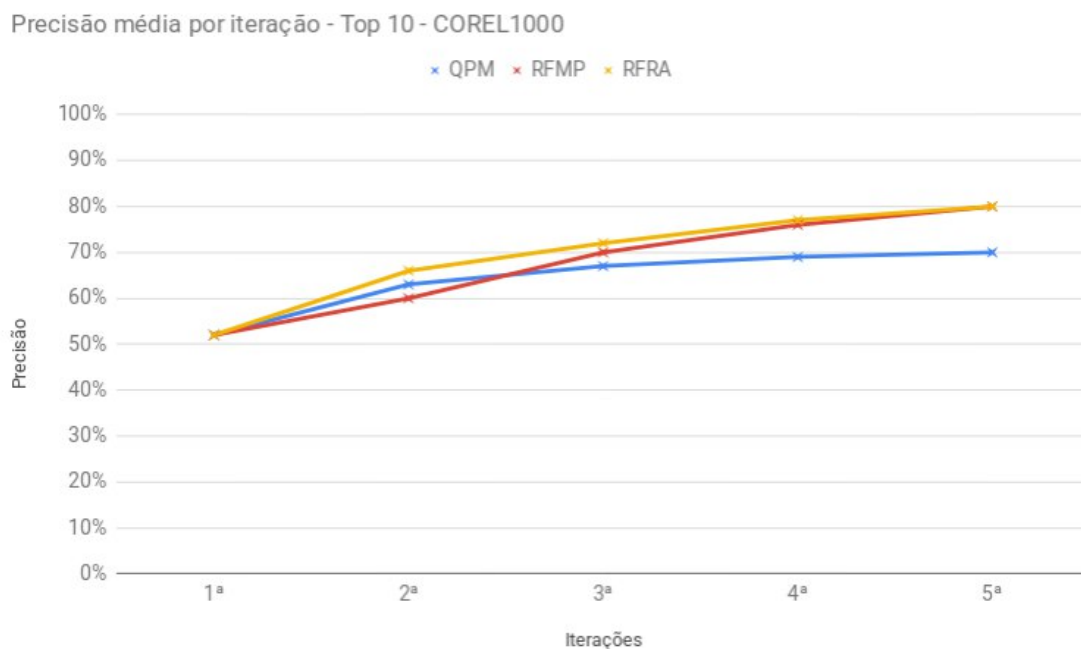


Figura 20 – Precisão média por iteração na base COREL1000 analisando o *top-10*.

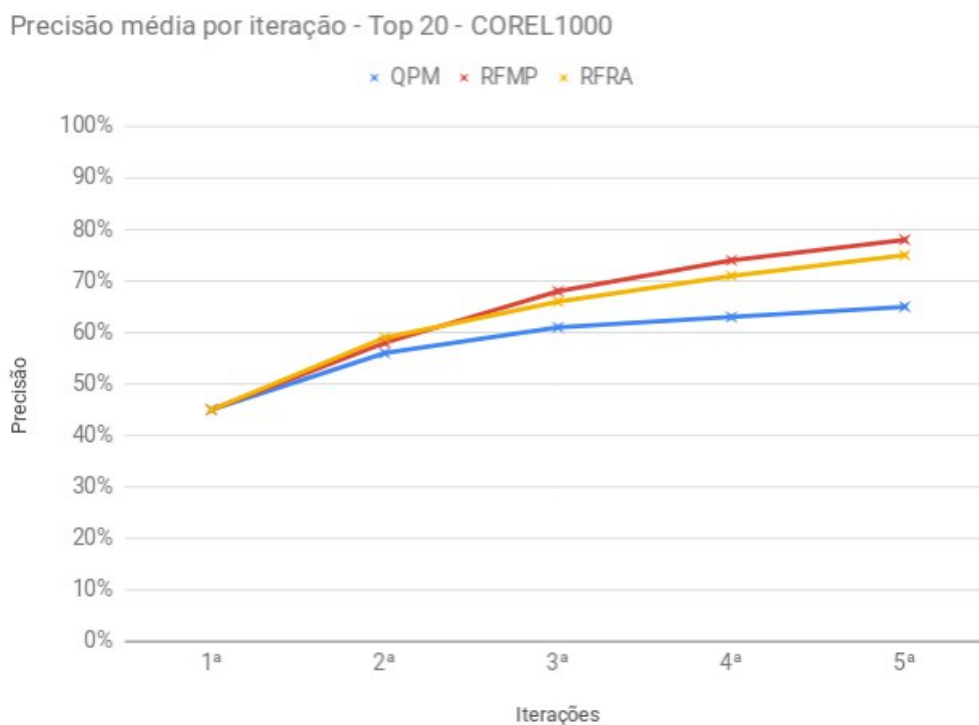


Figura 21 – Precisão média por iteração na base COREL1000 analisando o *top-20*.

As Figuras 22 e 23 mostram os valores de precisão considerando o top-10 e o top-20 da base *Oxford Flowers*. Nota-se o mesmo comportamento obtido com relação à análise da base COREL1000: os valores de precisão aumentam quando aplicado as técnicas de Realimentação de Relevância. A técnica QPM também apresentou o pior desempenho dentre as técnicas analisadas. A técnica proposta RFRA apresentou a maior precisão quando analisado o top-10, por outro lado, a técnica RFMP apresentou a maior precisão quando analisado o top-20.

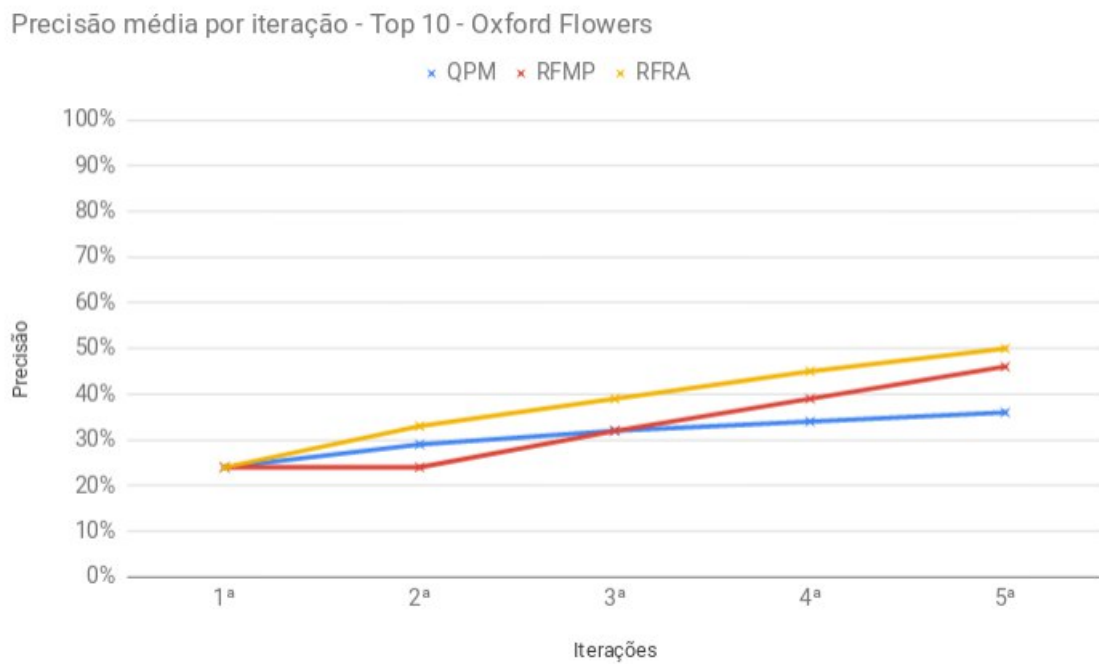


Figura 22 – Precisão média por iteração na base *Oxford Flowers* analisando o *top-10*.

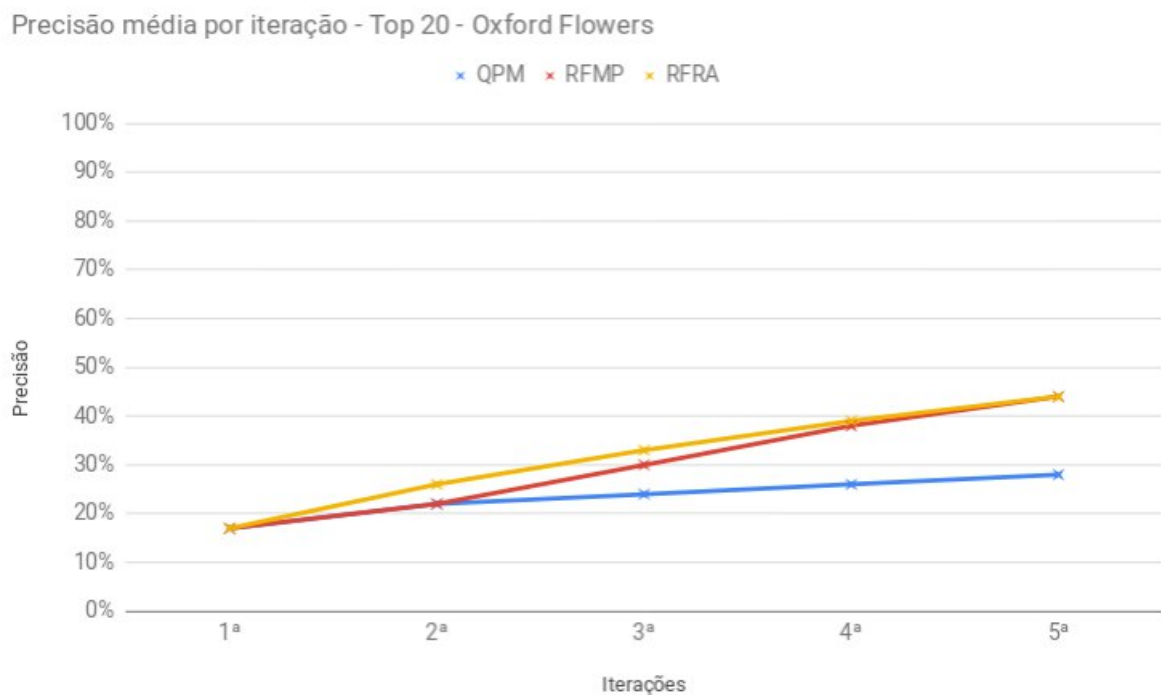


Figura 23 – Precisão média por iteração na base *Oxford Flowers* analisando o *top-20*.

Destaca-se que a precisão média por iteração das técnicas RFRA e RFMP estão bem similares, já que estas duas técnicas possuem o funcionamento similar: que é o uso das

imagens relevantes como centros de novas consulta. Porém a técnica RFMP apresentou um desempenho relativamente superior em cada iteração quando mais imagens são analisadas. Isso indica que a técnica RFMP apresenta desempenho maior quando mais informações forem coletadas pelo usuário.

As Figuras 24 e 25 apresentam as diferenças dos valores de MAP entre cada técnica na quinta iteração em relação a primeira consulta do sistema na base COREL1000, e as figuras 26 e 27 representam estes valores na base *Oxford Flowers*. É possível notar que todas técnicas apresentam impacto significativo em relação a primeira consulta em ambas as bases, mostrando a efetividade dos modelos de realimentação de relevâncias, porém as técnicas de múltiplos centros de consulta se provaram mais eficientes em trazer resultados relevantes.

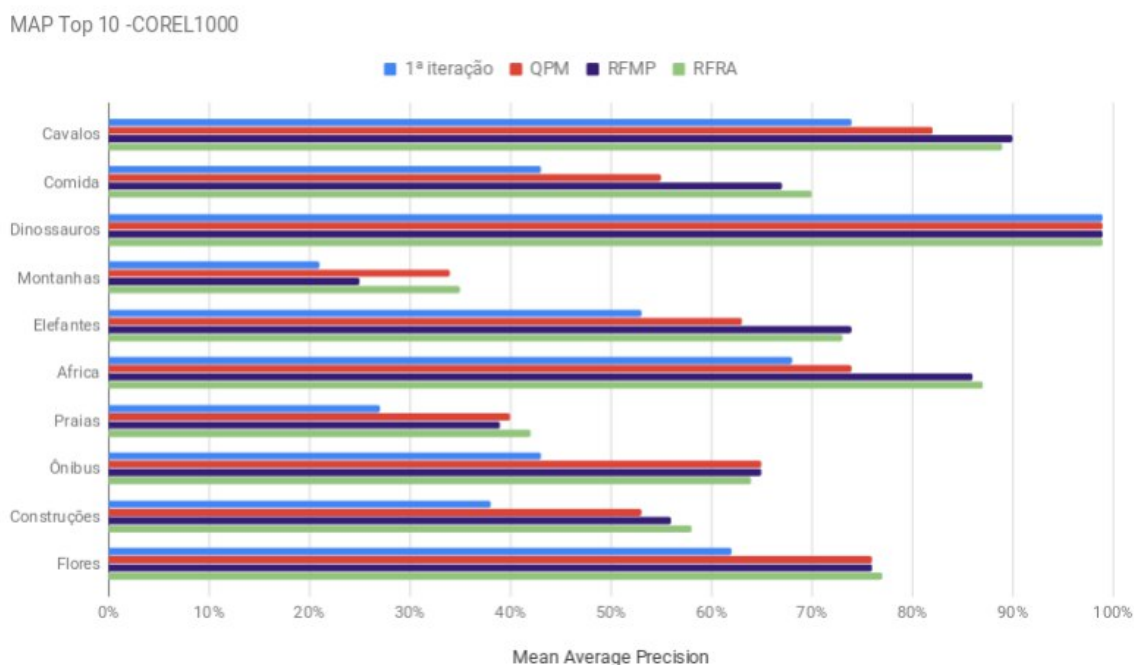


Figura 24 – MAP na primeira e última iteração na base COREL1000 analisando o *top-20*.

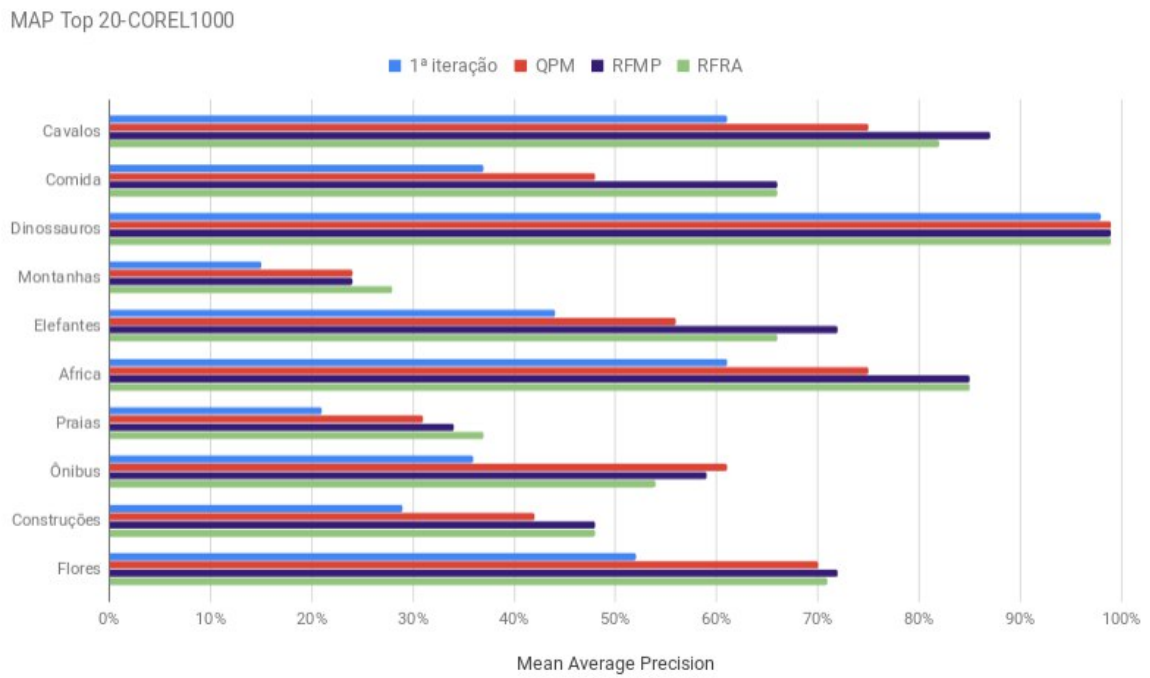


Figura 25 – MAP na primeira e última iteração na base COREL1000 analisando o *top-20*.

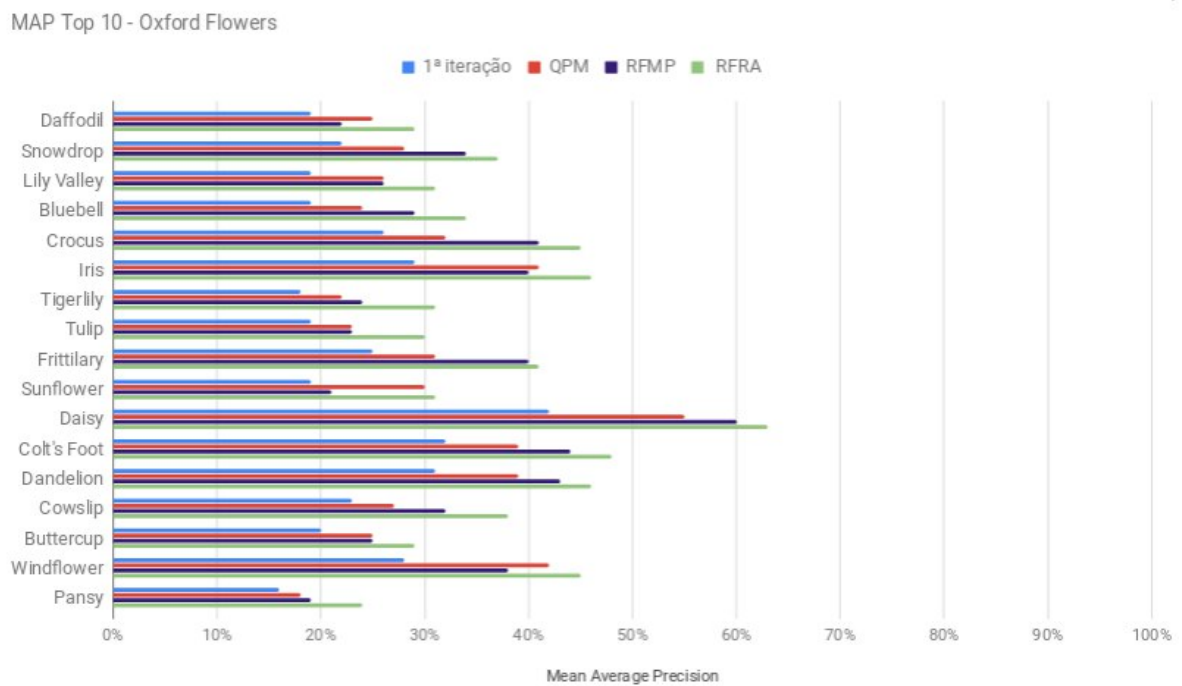


Figura 26 – MAP na primeira e última iteração na base *Oxford Flowers* analisando o *top-10*.



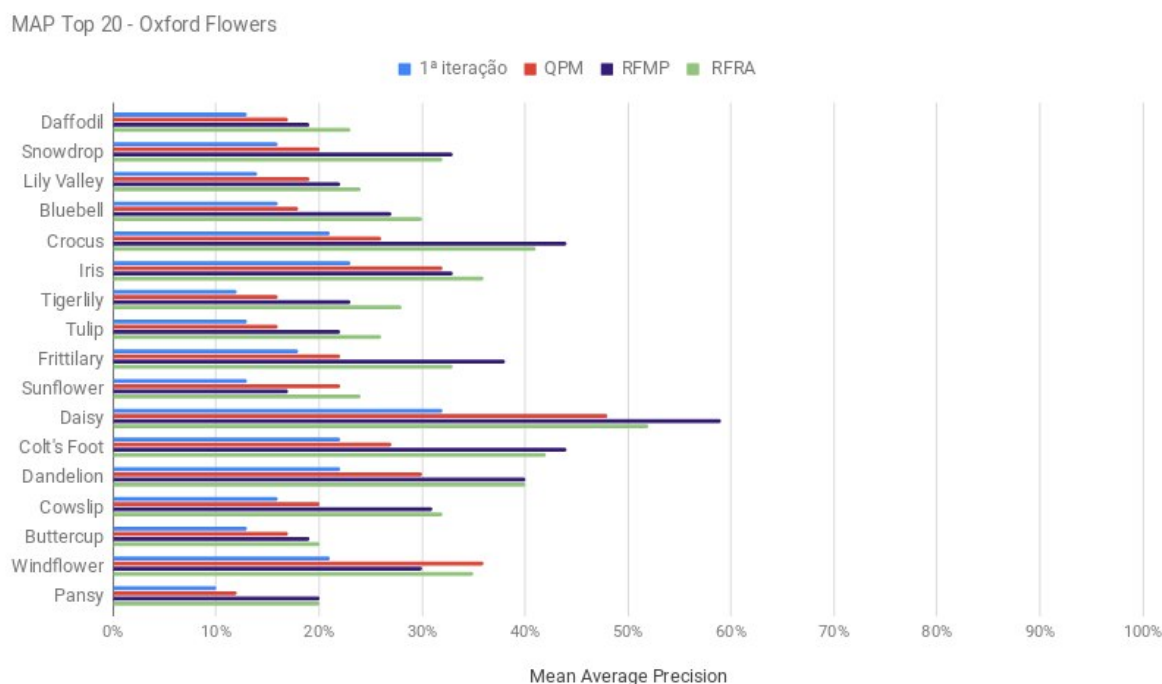


Figura 27 – MAP na primeira e última iteração na base *Oxford Flowers* analisando o *top-20*.

As Tabelas 4 e 5 mostram os resultados de cada base detalhando cada classe separadamente e analisando top-10 e top-20. No desempenho geral a técnica QPM foi a que retornou menos resultados relevantes, porém nota-se que em algumas classes esta técnica obteve desempenho próximo das técnicas RFMP e RFRA. Na base COREL1000, a técnica RFRA foi a que obteve desempenho melhor em relação ao top-10 e o RFMP se adequou melhor em relação ao top-20. Já na base *Oxford Flowers* a técnica RFRA foi a que obteve melhor MAP em relação as outras técnicas.

CLASSE	<i>top-10</i>			<i>top-20</i>		
	QPM	RFMP	RFRA	QPM	RFMP	RFRA
CAVALOS	0,82	0,90	0,89	0,75	0,87	0,82
COMIDA	0,55	0,67	0,70	0,48	0,66	0,66
DINOSSAUROS	0,99	0,99	0,99	0,99	0,99	0,99
MONTANHAS	0,34	0,25	0,35	0,24	0,24	0,28
ELEFANTES	0,63	0,74	0,73	0,56	0,72	0,66
AFRICA	0,74	0,86	0,87	0,75	0,85	0,85
PRAIAS	0,40	0,39	0,42	0,31	0,34	0,37
ÔNIBUS	0,65	0,65	0,64	0,61	0,59	0,54
CONSTRUÇÕES	0,53	0,56	0,58	0,42	0,48	0,47
FLORES	0,76	0,76	0,77	0,70	0,72	0,71
MAP GERAL	0,64	0,68	0,70	0,58	0,65	0,63

Tabela 4 – Mean Average Precision da base COREL1000 analisando o *top-10* e *top-20*.

CLASSE	<i>top-10</i>			<i>top-20</i>		
	QPM	RFMP	RFRA	QPM	RFMP	RFRA
DAFFODIL	0,25	0,22	0,29	0,17	0,19	0,23
SNOWDROP	0,28	0,34	0,37	0,20	0,33	0,32
LILY VALLEY	0,26	0,26	0,31	0,19	0,22	0,24
BLUEBELL	0,24	0,29	0,34	0,18	0,27	0,30
CROCUS	0,32	0,41	0,45	0,26	0,44	0,41
IRIS	0,41	0,40	0,46	0,32	0,33	0,36
TIGERLILY	0,22	0,24	0,31	0,16	0,23	0,28
TULIP	0,23	0,23	0,30	0,16	0,22	0,26
FRITILLARY	0,31	0,40	0,41	0,22	0,38	0,33
SUNFLOWER	0,30	0,21	0,31	0,22	0,17	0,24
DAISY	0,55	0,60	0,63	0,48	0,59	0,52
COLT'S FOOT	0,39	0,44	0,48	0,27	0,44	0,42
DANDELION	0,39	0,43	0,46	0,30	0,40	0,40
COWSLIP	0,27	0,32	0,38	0,20	0,31	0,32
BUTTERCUP	0,25	0,25	0,29	0,17	0,19	0,20
WINDFLOWER	0,42	0,38	0,45	0,36	0,30	0,35
PANSY	0,18	0,19	0,24	0,12	0,20	0,20
MAP GERAL	0,31	0,33	0,38	0,23	0,30	0,32

Tabela 5 – Mean Average Precision da base Oxford Flowers analisando o *top-10* e *top-20*.

## 5 Considerações finais

Este trabalho abordou o contexto de sistemas de recuperação de imagens baseado em conteúdo, de forma a elencar algumas técnicas que pudessem melhorar o desempenho destes sistemas. Neste caso, foi utilizado a abordagem de Realimentação de Relevâncias, em que o usuário fornece um *feedback* sobre um ranking inicial gerado pelo sistema.

Para demonstrar o potencial das técnicas de Realimentação de Relevâncias no aprimoramento dos sistemas de busca por imagens similares, foram realizados alguns experimentos a partir da construção de um sistema simples, capaz de extrair características de uma imagem e medir a similaridade entre uma imagem-consulta e outras imagens de uma base de dados.

Foram implementadas três técnicas de realimentação de relevância sobre o protótipo de busca desenvolvido, sendo uma destas técnicas uma nova proposta que foi desenvolvida a partir dos conhecimentos adquiridos ao decorrer do trabalho. Estas três técnicas foram analisadas usando duas bases de imagens diferentes e comparadas utilizando as métricas usualmente empregadas na análise de sistemas de recuperação de informação.

Pelos testes experimentais foi possível notar que, de fato, as técnicas de Realimentação de Relevâncias são abordagens capazes de melhorar a precisão das buscas, retornando imagens mais próximas do esperado pelo usuário a cada iteração, independente da base de dados e características extraídas.

### 5.1 Trabalhos futuros

Os trabalhos futuros elencados a seguir envolvem o aprimoramento do que foi construído no decorrer deste trabalho, de forma a melhorar o desempenho das técnicas de Realimentação de Relevâncias implementadas.

- Utilizar diferentes métodos de conversão para a escala de cinza.
- Trabalhar diferentes abordagens de extração de características em cima das bases utilizadas no desenvolvimento do trabalho: checar o impacto de diferentes descritores de imagens, de forma a avaliar se estes trazem informações mais ricas sobre uma imagem e com isso diferenciar melhor as imagens das bases.
- Coletar e analisar as métricas utilizando a distância euclidiana, de forma a gerar comparação com as métricas geradas neste trabalho que utilizou a distância de Manhattan.

- 
- Implementar a ideia de pesos das imagens a partir da entrada do usuário, ou seja, o usuário informa o peso das imagens e com isso as técnicas QPM e RFMP terão mais uma variável de impacto para a determinação da nova consulta;
  - Aprimorar a técnica RFRA: a técnica RFRA se mostrou bastante eficiente, porém somente quando são retornadas imagens relevantes. A proposta para aprimorar esta técnica seria utilizar as imagens não-relevantes de alguma forma no algoritmo proposto;
  - Realizar mais testes experimentais usando outras bases de imagens para avaliar a eficiência das técnicas de Realimentação de Relevâncias em outros contextos.

## Referências

- ALKHAWLANI, M.; ELMOGY, M.; EL-BAKRY, H. Content-based image retrieval using local features descriptors and bag-of-visual words. *International Journal of Advanced Computer Science and Applications*, 2015. Citado na página 20.
- ASLANDOGAN, Y. A.; YU, C. T. Techniques and systems for image and video retrieval. *IEEE Transactions on Knowledge and Data Engineering*, 1999. Citado na página 16.
- BIGIO, G. J. Desenvolvimento de um método para refinamento de respostas baseados em relevance feedback para refinar consultas por similaridade. 2008. Citado 2 vezes nas páginas 30 e 37.
- ELECTRONJS. *Arquitetura de Aplicativos Electron*. 2013. Disponível em: <<https://electronjs.org/docs/tutorial/application-architecture>>. Acesso em: 18 nov. 2018. Citado na página 33.
- GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006. ISBN 013168728X. Citado 3 vezes nas páginas 6, 17 e 19.
- HASSABALLAH, M.; ALI, A.; ALSHAZLY, H. Image features detection, description and matching. v. 630, p. 11–45, 2016. Citado na página 20.
- KANAN, C.; COTTRELL, G. Color-to-grayscale: Does the method matter in image recognition? *PloS one*, v. 7, p. e29740, 01 2012. Citado na página 18.
- KNIBERG, H. *Kanban and Scrum - Making the Most of Both*. [S.l.]: Lulu.com, 2010. ISBN 0557138329, 9780557138326. Citado na página 32.
- KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika* vol.29, p. 1–27, 1964. Citado na página 21.
- LI LIN FENG, L. Y. J. W. S. L. L. Fusion framework for color image retrieval based on bag-of-words model and color local haar binary patterns. *Journal of Electronic Imaging*, v. 25, p. 25 – 25 – 17, 2016. Disponível em: <<https://doi.org/10.1117/1.JEI.25.2.023022>>. Citado 2 vezes nas páginas 6 e 40.
- MACÊDO, S.; MELO, G.; KELNER, J. A comparative study of grayscale conversion techniques applied to sift descriptors. In: . [S.l.: s.n.], 2015. Citado na página 18.
- MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press, 2008. ISBN 978-0-521-86571-5. Disponível em: <<http://nlp.stanford.edu/IR-book/information-retrieval-book.html>>. Acesso em: 07 nov. 2018. Citado 3 vezes nas páginas 12, 23 e 24.
- MARQUES, J. Realimentação de relevância para recuperação por conteúdo de imagens médicas visando diminuir a descontinuidade semântica. USP - São Carlos, 2006. Citado 14 vezes nas páginas 6, 15, 16, 17, 18, 19, 20, 21, 25, 26, 27, 28, 29 e 37.

- MARQUES, J.; TRAINA, A. Realimentação de relevância: Integração do conhecimento do especialista com a recuperação de imagens por conteúdo. 11 2006. Citado 4 vezes nas páginas 6, 27, 29 e 30.
- MISTRY, Y.; INGOLE, D. D. Survey on content based image retrieval systems. Prof. Ram Meghe Institute of Technology & Research, 2013. Citado na página 15.
- NILSBACK, M.-E.; ZISSERMAN, A. A visual vocabulary for flower classification. In: *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*. Washington, DC, USA: IEEE Computer Society, 2006. (CVPR '06), p. 1447–1454. ISBN 0-7695-2597-0. Disponível em: <<http://dx.doi.org/10.1109/CVPR.2006.42>>. Citado na página 40.
- ORTEGA-BINDERBERGER, M.; MEHROTRA, S. Relevance feedback in multimedia databases. In: *In Handbook of Video Databases: Design and Applications*. [S.l.]: CRC Press, 2003. Citado 3 vezes nas páginas 6, 26 e 27.
- PEDRONETTE, D. C. G.; TORRES, R. da S. Image re-ranking and rank aggregation based on similarity of ranked lists. In: *CAIP*. [S.l.: s.n.], 2011. Citado na página 30.
- PIAO, X. et al. Block-diagonal sparse representation by learning a linear combination dictionary for recognition. 01 2016. Citado 2 vezes nas páginas 6 e 41.
- PILEVAR, A. H. Cbmir: Content-based image retrieval algorithm for medical image databases. *Journal of medical signals and sensors*, n. 1, 2011. Citado na página 12.
- PONCIANO-SILVA, M. Processamento de consultas por similaridade em imagens médicas visando à recuperação perceptual guiada pelo usuário. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2009. Acesso em: 04 nov. 2018. Citado 3 vezes nas páginas 18, 20 e 22.
- RENDA, M. E.; STRACCIA, U. Web metasearch: Rank vs. score based rank aggregation methods. In: . [S.l.: s.n.], 2003. p. 841–846. Citado na página 31.
- ROCCHIO, J. J. Relevance feedback in information retrieval. In: SALTON, G. (Ed.). *The Smart retrieval system - experiments in automatic document processing*. [S.l.]: Englewood Cliffs, NJ: Prentice-Hall, 1971. p. 313–323. Citado 2 vezes nas páginas 25 e 26.
- SHAW, J. A.; FOX, E. A. Combination of multiple searches. In: *THE SECOND TEXT RETRIEVAL CONFERENCE (TREC-2)*. [S.l.: s.n.], 1994. p. 243–252. Citado na página 31.
- SILVA, A. T. da. Recuperação de imagens por conteúdo baseada em realimentação de relevância e classificador por floresta de caminhos ótimos. Faculdade de Engenharia Elétrica e de Computação, UNICAMP, 2011. Citado 3 vezes nas páginas 18, 20 e 24.
- SOMMERVILLE, I. *Software Engineering*. 9th. ed. USA: Addison-Wesley Publishing Company, 2010. ISBN 0137035152, 9780137035151. Citado na página 32.
- TAMURA H., M. S. Y. T. Textural features corresponding to visual perception. *IEEE Trans on Systems, Man and Cybernetics* 8, 1978. Citado na página 19.

TORRES, R. da S.; FALCAO, A. X. Content-based image retrieval: Theory and applications. *RITA*, v. 13, n. 2, p. 161–185, 2006. Disponível em: <<http://dblp.uni-trier.de/db/journals/rita/rita13.html#TorresF06>>. Citado na página 12.

TORRES, R. S.; FALCAO, A. X. Recuperação de imagens baseada em conteúdo. Instituto de Computação, UNICAMP, 2006. Citado na página 16.

## Apêndices



# APÊNDICE A – Correlação das imagens da base de dados

Para uma maior compreensão de como as imagens estão distribuídas na base de dados, foram calculados e armazenados os histogramas de cada imagem na escala de cinza. Estes são utilizados como entrada em um método de redução de dimensionalidade, o t-SNE (*t-Distributed Stochastic Neighbor Embedding*), que neste caso é alimentado com os vetores de características das imagens. O t-SNE exibe os objetos de forma que objetos similares sejam projetados como pontos próximos. O algoritmo t-SNE utilizado está implementado na biblioteca *sklearn*.

A Figura 28 apresenta o mapeamento das 1000 imagens existentes na base *Corel1000*. É possível notar que há dois grandes grupos formados ao se utilizar apenas o histograma de cor como extrator de características. Um desses grupos é formado por quase que exclusivamente pelas imagens da classe DINOSSAURO. As imagens desta classe são formadas por figuras ilustrativas e possui muitos *pixels* com mesma intensidade, por esse motivo é bem nítida a distância desta classe com as outras classes da base.



Figura 28 – Mapeamento espacial das imagens da base Corel100 utilizando histograma de escala de cinza e projetadas pelo t-SNE (melhor visualizado no formato digital)

A Figura 29 representa o mesmo algoritmo aplicado sobre a base de imagens *Oxford Flowers*. É possível observar que nesta base, apesar de possuir 17 classes de flores, as imagens de diferentes classes estão mais juntas utilizando apenas esse descritor.

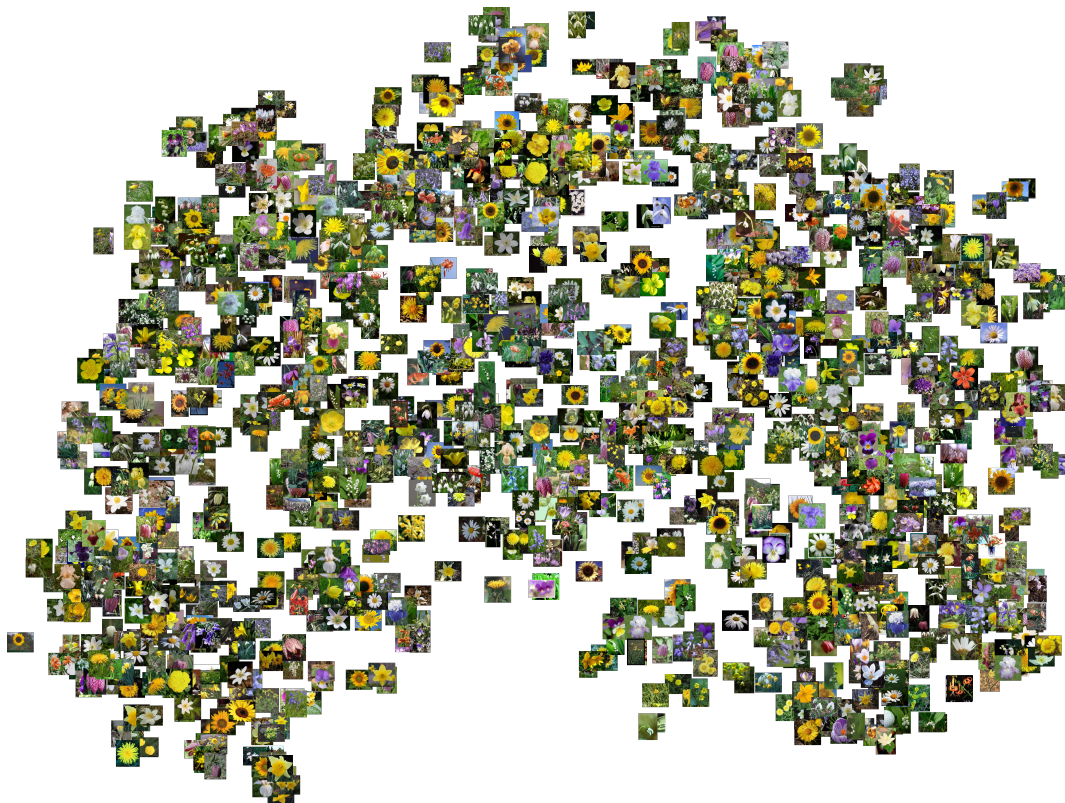


Figura 29 – Mapeamento espacial das imagens da base OxfordFlowers utilizando histograma de escala de cinza e projetadas pelo t-SNE (melhor visualizado no formato digital)