

Análise de Imagens e Visão Computacional

Prof. Henrique Batista da Silva

Introdução a Recuperação de Informação Visual

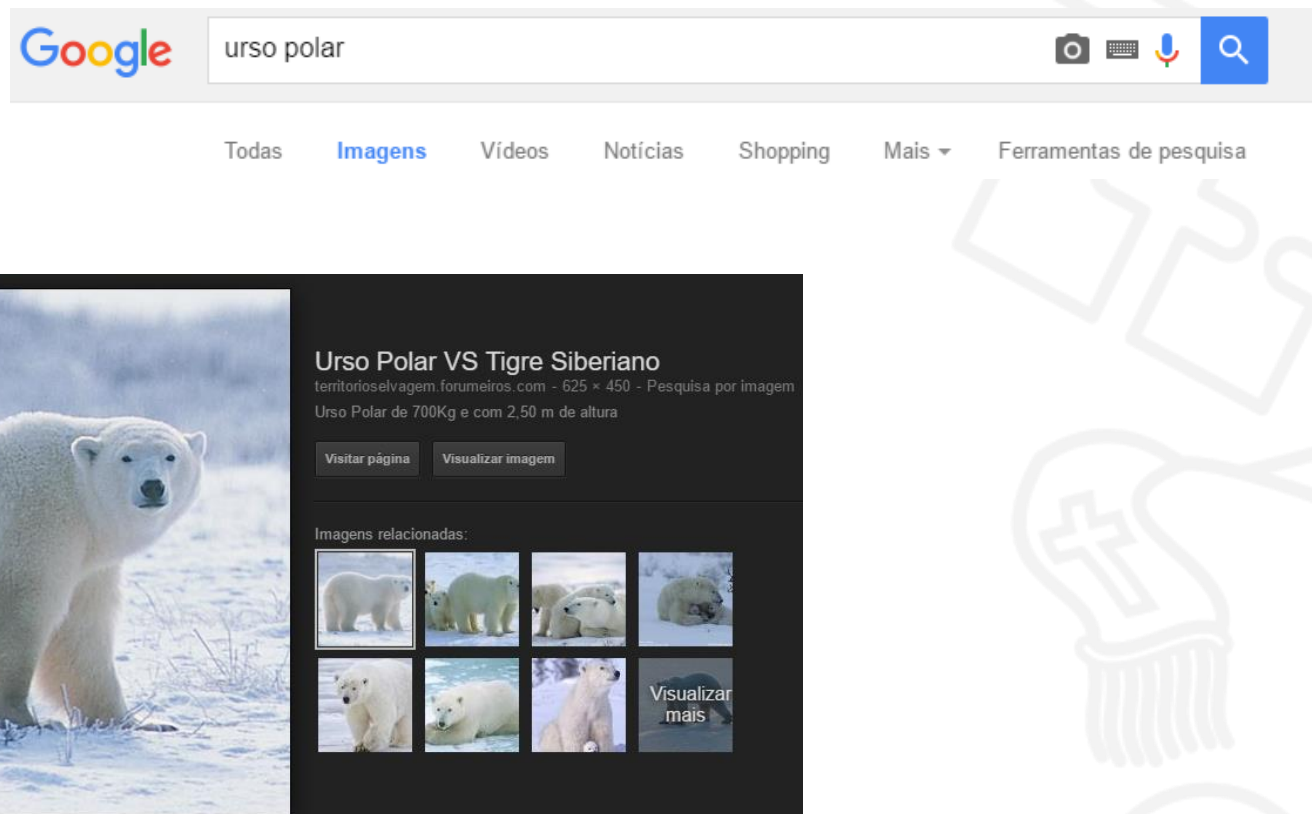
Introdução

A recuperação de informação por conteúdo textual é muito limitada.

Recuperação de informação visual (Web)

- Como é feito hoje?
- Indexação manual:
 - definição manual de palavras-chave que descrevem a imagem.
- Indexação automática:
 - utilização de: metadados semânticos associados à imagem; nome do arquivo da imagem; título do documento onde a imagem se encontra; texto próximo à imagem; etc.

Recuperação de informação visual (Web)



Fonte: Google Images

Recuperação de Imagem Baseada em Conteúdo - (CBIR)

Busca através das características primitivas da imagem

Representação e recuperação das imagens;

- através de suas características primitivas, como:
 - cor;
 - forma dos objetos;
 - relação espacial entre as cores e objetos da imagem; etc.

Recuperação de Imagem Baseada em Conteúdo - (CBIR)

Problema:

Recuperar imagens que casam;
total ou apenas parcialmente;
com uma imagem dada.

Recuperação de Imagem Baseada em Conteúdo - (CBIR)

- O casamento entre a imagem dada e a correspondente imagem recuperada pode ser classificado em:
 - casamento de imagens inteiras;
 - casamento de subpadrão:
 - procura por um subpadrão;
 - a imagem dada;
 - em outras imagens.
 - Exemplo:
 - busca de um objeto em imagens.

Descrição de Imagens

Descrição de imagens

Descritor representa a propriedade de uma imagem
Descritor deve ser invariante a rotação, escala, etc.

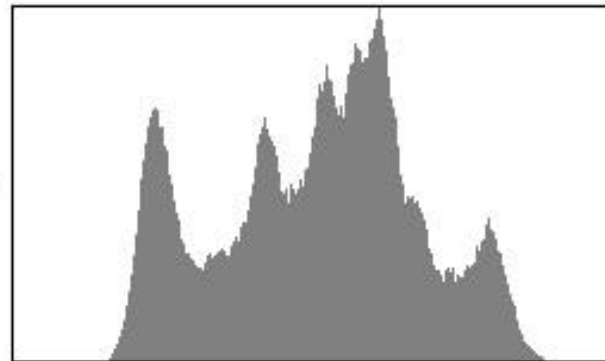
Image



Low-level
feature



Histogram

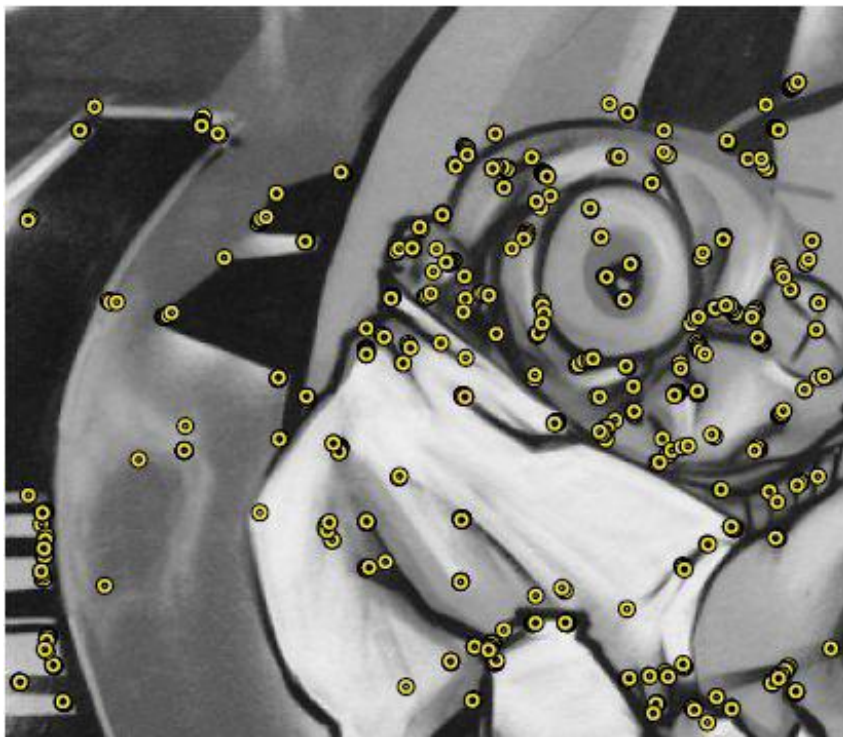


Fonte: Google Images

Descrição de imagens

Demo: cálculo de histograma

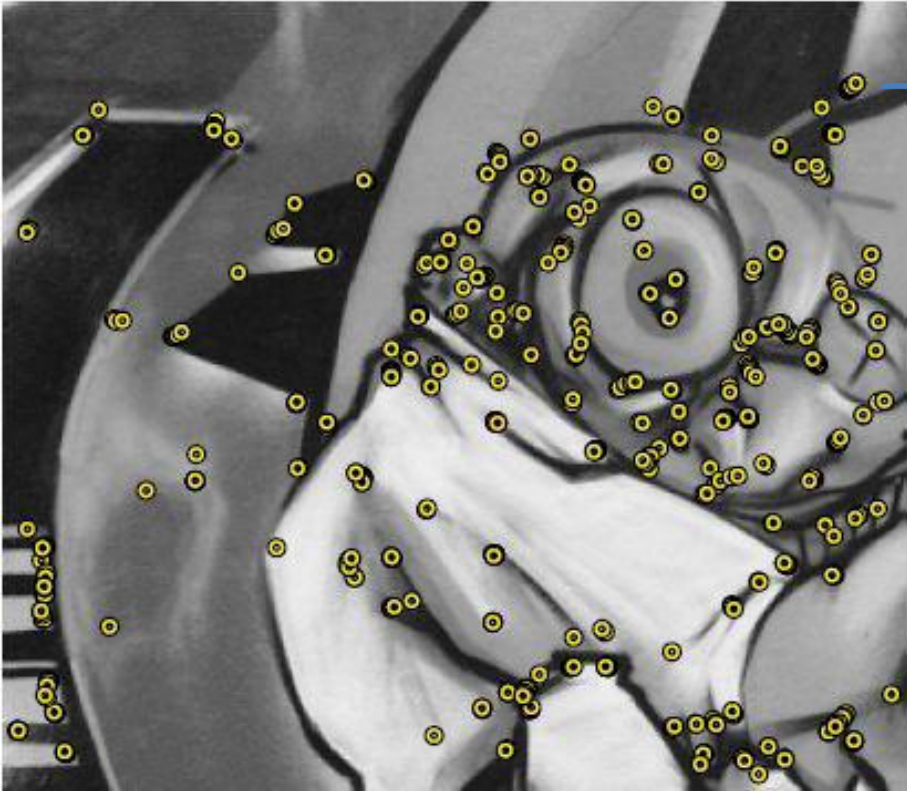
Descrição de imagens



Fonte: SIVIC, J (2006).

Features locais: Padrão que se difere da vizinhança local.
Pontos de interesse: cantos (corners), regiões, etc..

SIFT (Scale-invariant feature transform)



Fonte: SIVIC, J (2006).

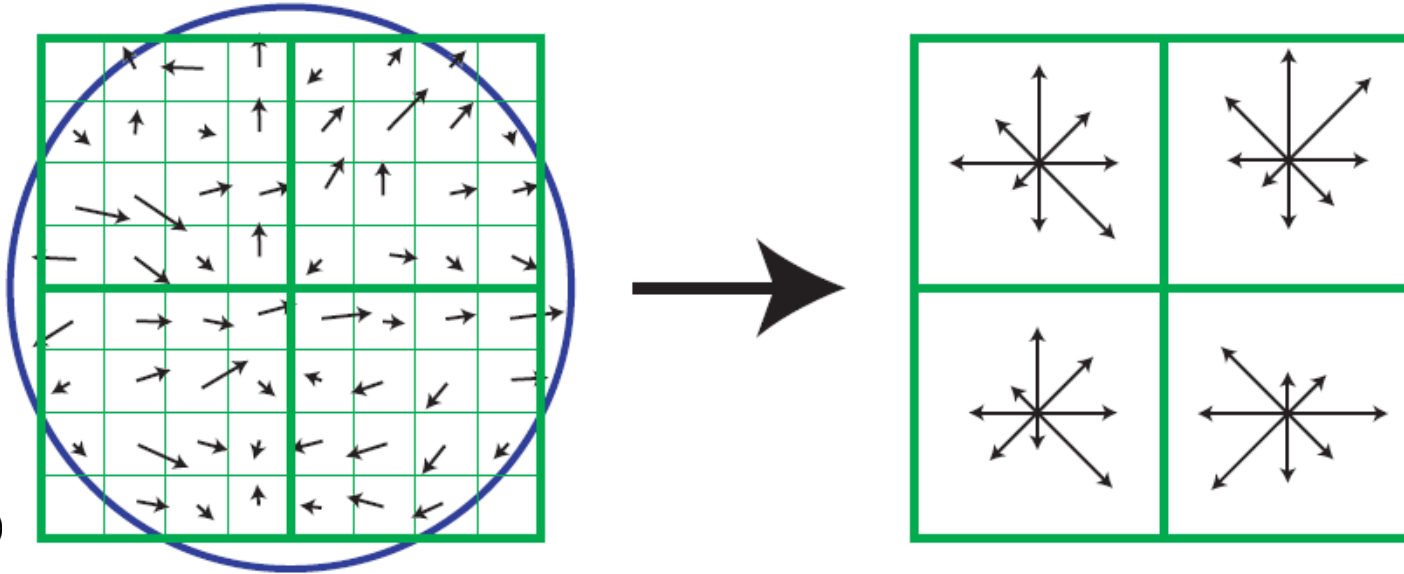
- Há técnicas melhores ...
- Uma das mais usadas: SIFT (Scale-invariant feature transform) proposta por Lowe (1999);

12 21 87 35 ... 14

SIFT (Scale-invariant feature transform)

- Descritor representa uma propriedade de um região.
- O descritor deve ser invariante a transformações.
- SIFT (Lowe, 1999).

(LOWE, 2004)



Representa cada região por um Histograma das orientações dos gradientes, correspondendo ao tamanho das setas no lado direito da figura, que é um vetor de 128 dimensões.

Detector Harris-Affine

O Detector de Harris-Affine é utilizado para detectar cantos na imagem (que são regiões de alta curvatura).



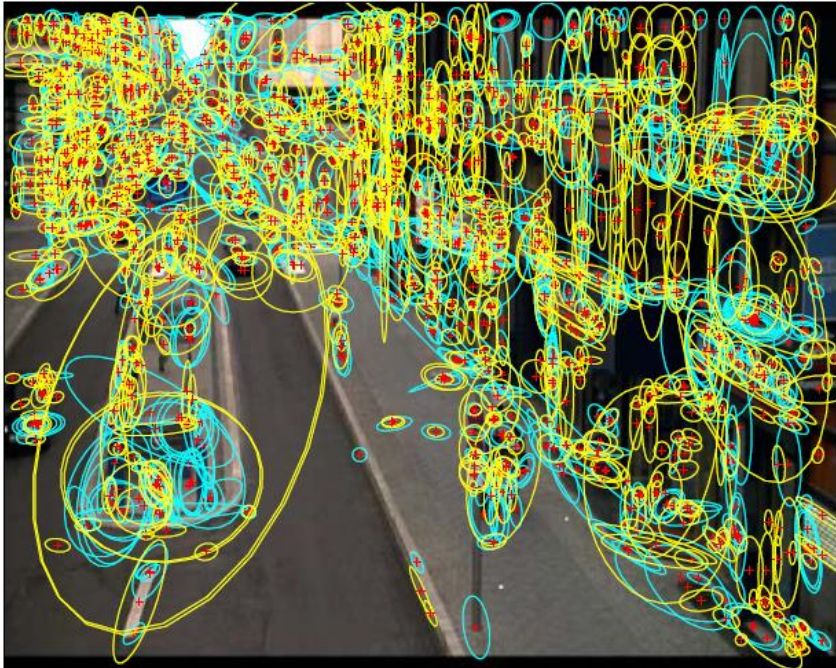
Fonte: SIVIC, J (2006).

Exemplo de detecção



Fonte: <https://www.robots.ox.ac.uk/~vgg/demo/>

Exemplo de detecção



Fonte: <https://www.robots.ox.ac.uk/~vgg/demo/>

Exemplo de detecção



Mesmo com pontos de vista diferentes, as mesmas regiões foram detectadas nas duas imagens

Fonte: <https://www.robots.ox.ac.uk/~vgg/demo/>

Exemplo de detecção



Fonte: <https://www.robots.ox.ac.uk/~vgg/demo/>

Exemplo de detecção (para localização)

Query



Results



Detecção e Extração de Features

Descrição de imagens

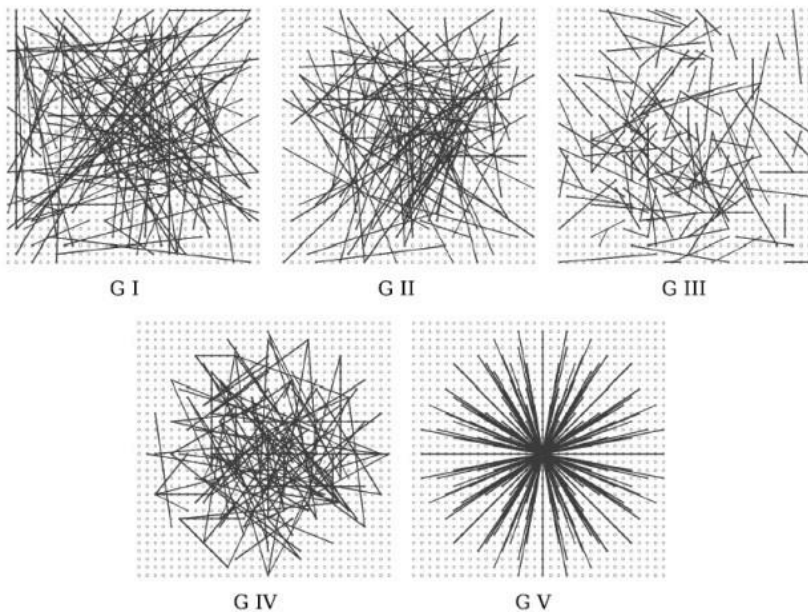
Métodos de descrição de características visuais, como o SIFT, não precisam da etapa de learning.

Descrição de imagens

Demo: ORB (Oriented FAST and Rotated BRIEF)

Descrição de imagens

BRIEF: para cada par de pontos, calcula o valor '1' ou '0' para cada distribuição de pares abaixo



$$\tau(\mathbf{p}; \mathbf{x}, \mathbf{y}) := \begin{cases} 1 & \text{if } \mathbf{p}(\mathbf{x}) < \mathbf{p}(\mathbf{y}) \\ 0 & \text{otherwise} \end{cases}$$

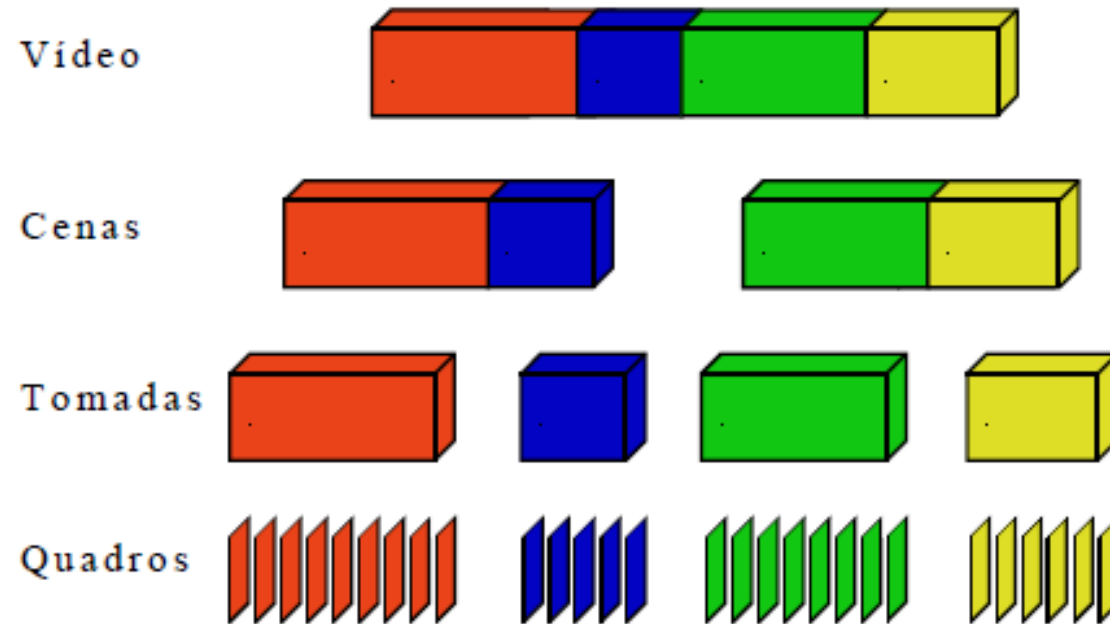
Fonte: https://www.researchgate.net/publication/221304115_BRIEF_Binary_Robust_Independent_Elementary_Features

Descrição de imagens

ORB (Oriented FAST and Rotated BRIEF) detecção e BRIEF para cálculo do descritor

Busca de Vídeo com Base em Conteúdo Visual

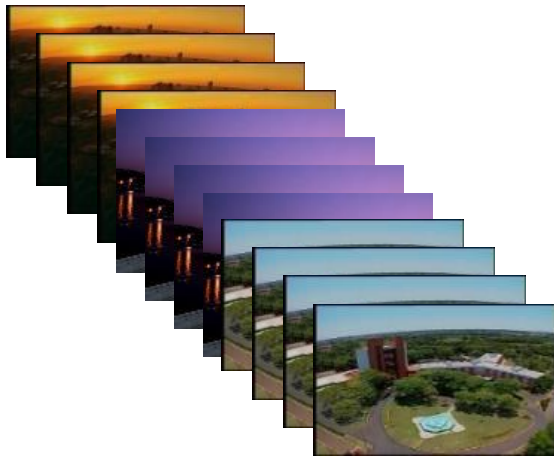
Vídeo Digital



Extraída de Ávila(2008)

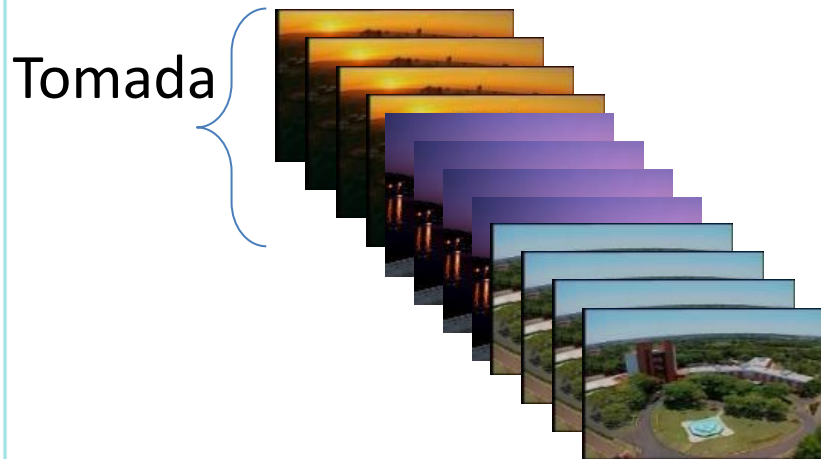
Vídeo Digital

- Vídeo é uma coleção de imagens (quadros) que são apresentados em uma unidade de tempo (ex.: 30 quadros por segundo)



Vídeo Digital

- Vídeo é uma coleção de imagens (quadros) que são apresentados em uma unidade de tempo (ex.: 30 quadros por segundo)



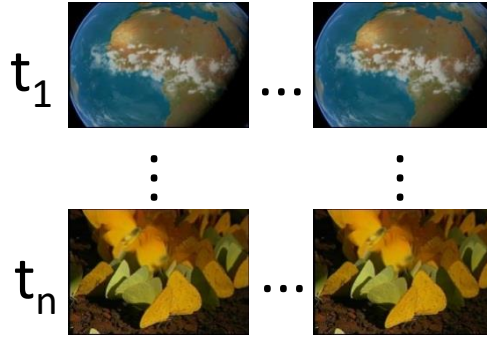
Fonte: KG do Patrocínio Jr, SJF Guimarães, HB da Silva, KJF de Souza. **An unified transition detection based on bipartite graph matching approach.** Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. 2010. p. 184-192.

Recuperação de vídeo

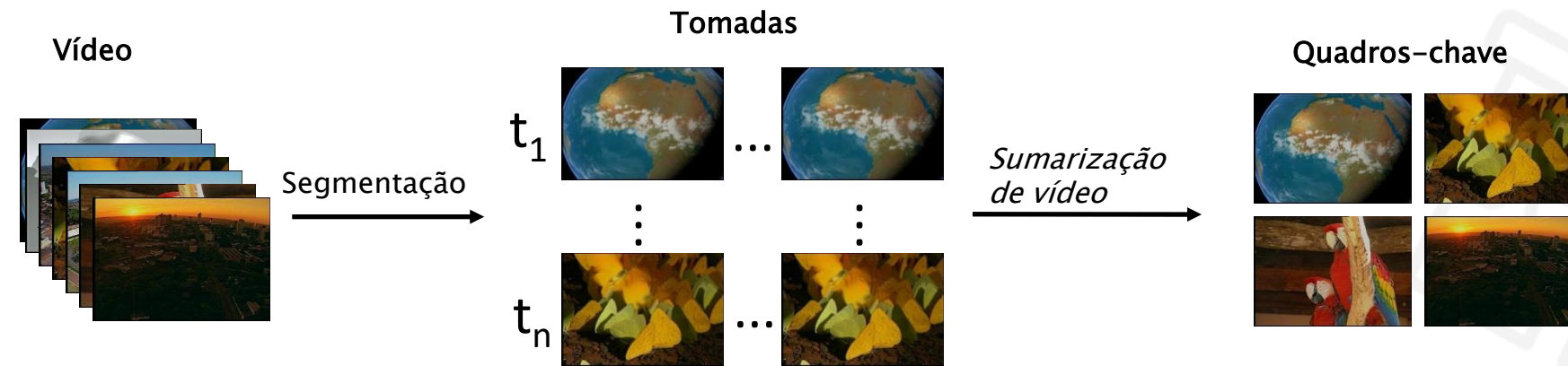
Vídeo

Segmentação

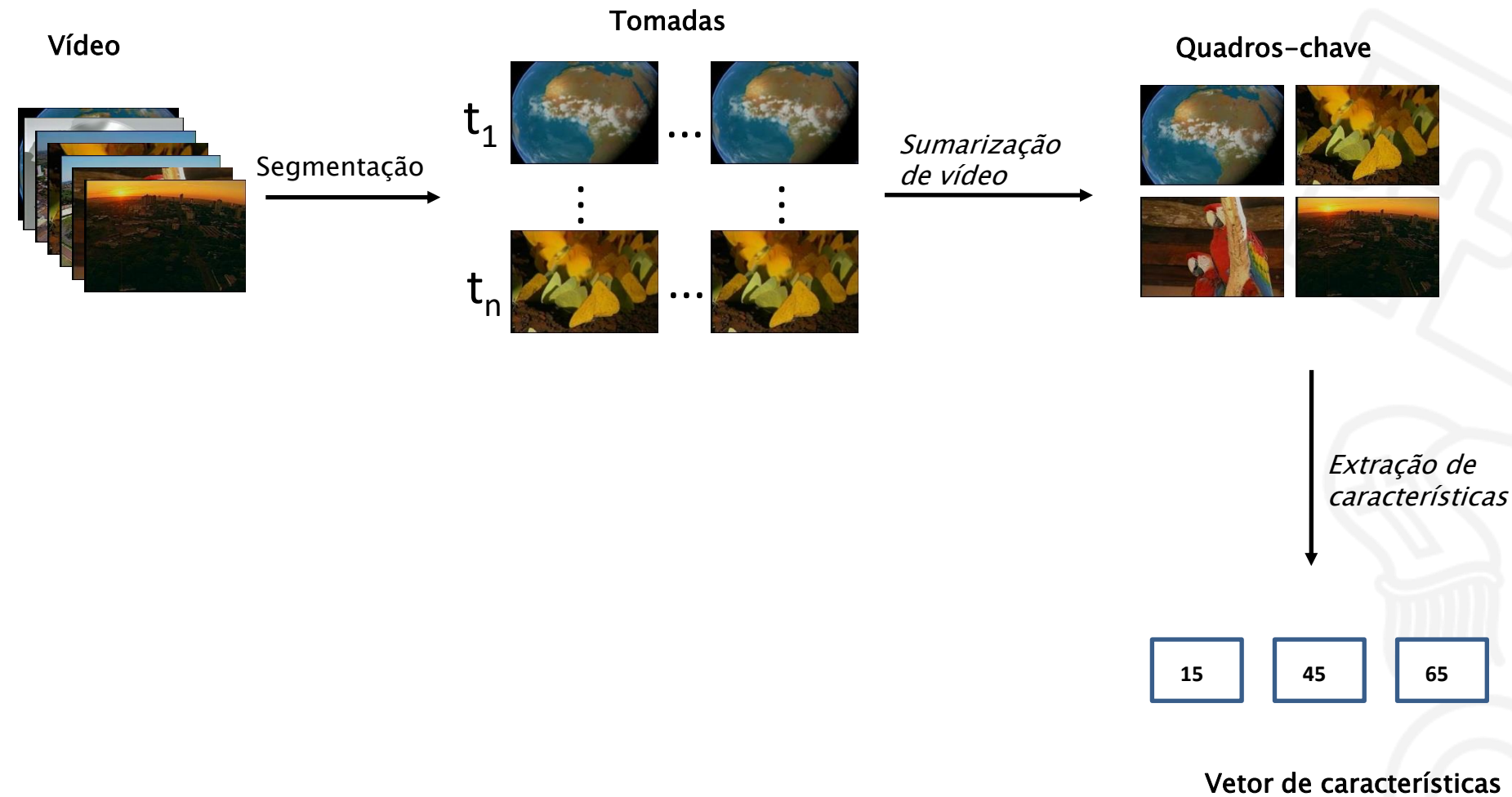
Tomadas



Recuperação de vídeo



Recuperação de vídeo

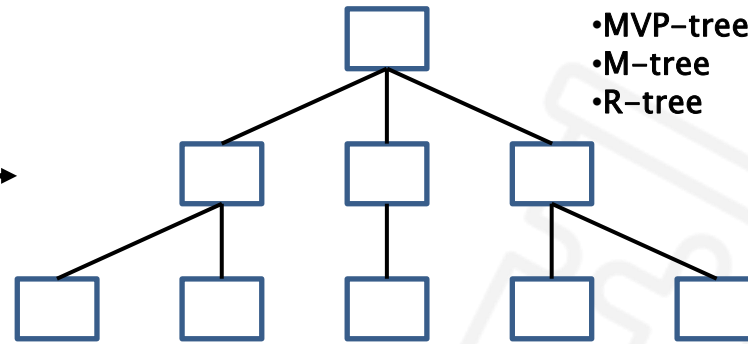


Recuperação de vídeo

Vetor de características
simplificado

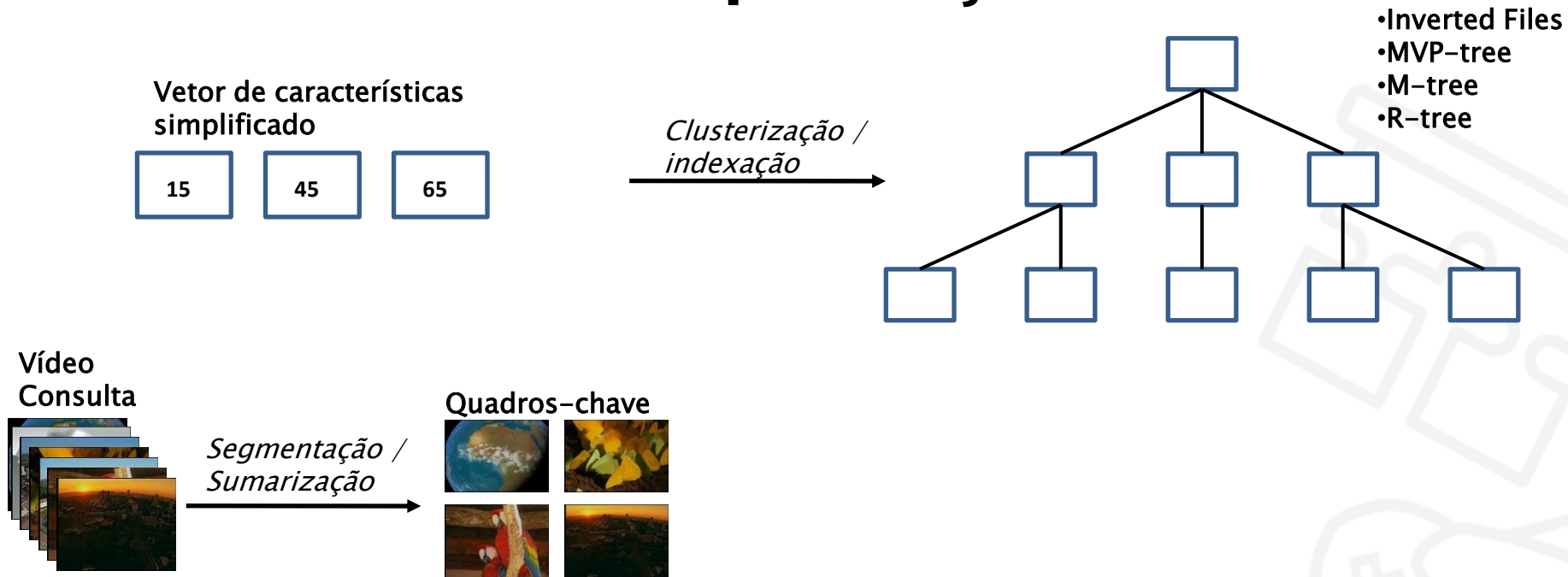


*Clusterização /
indexação*

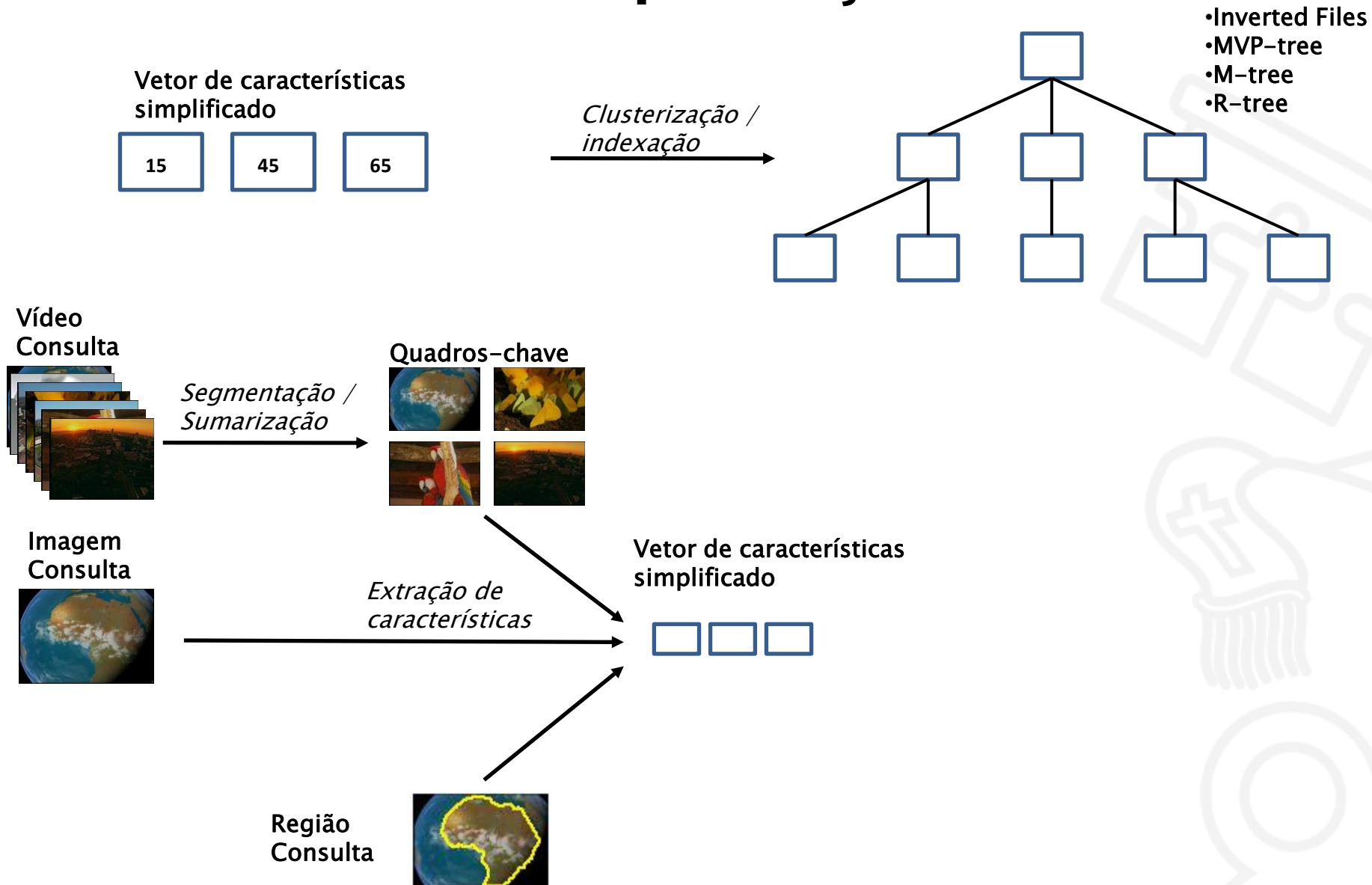


- Inverted Files
- MVP-tree
- M-tree
- R-tree

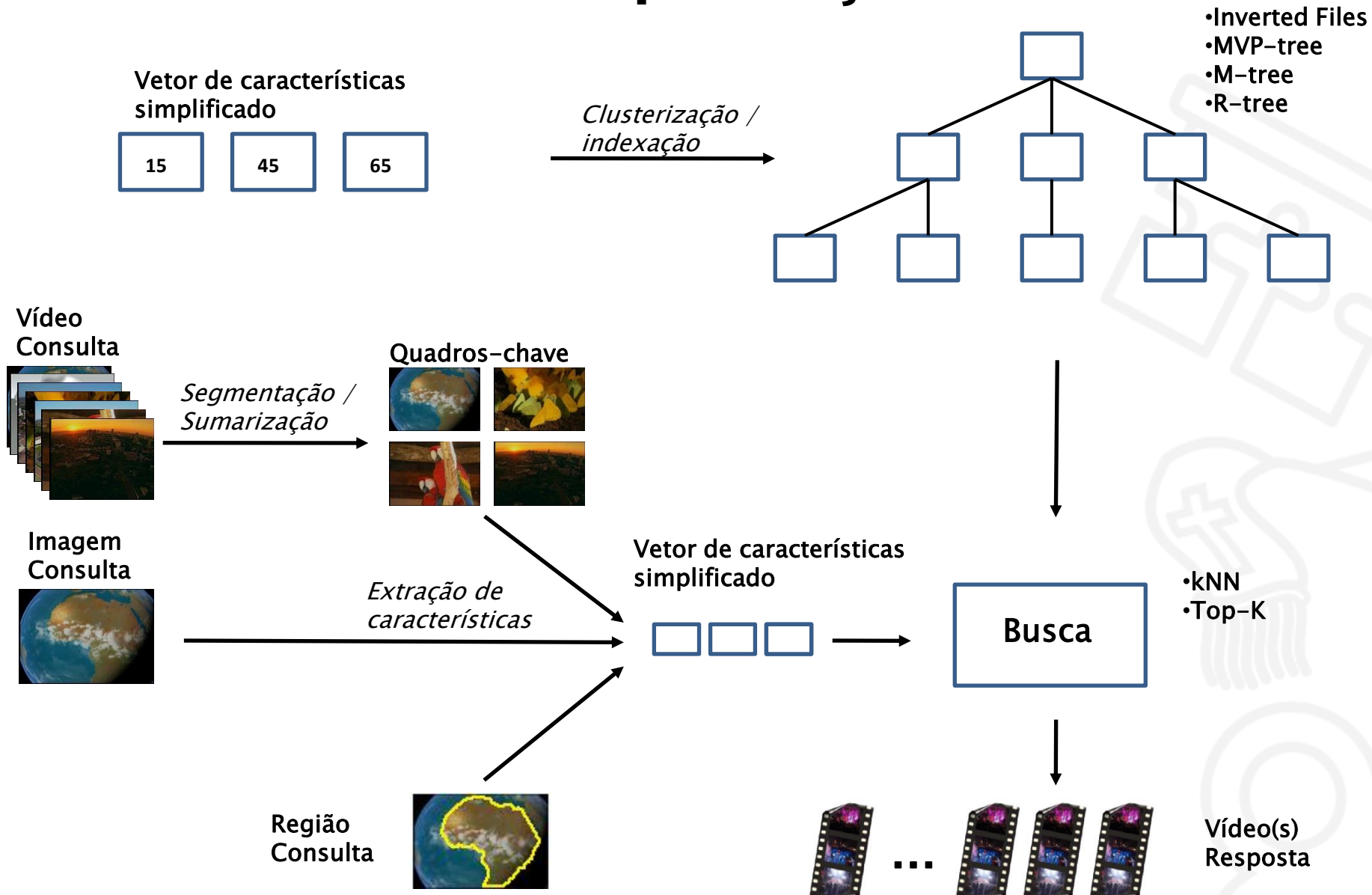
Recuperação de vídeo



Recuperação de vídeo

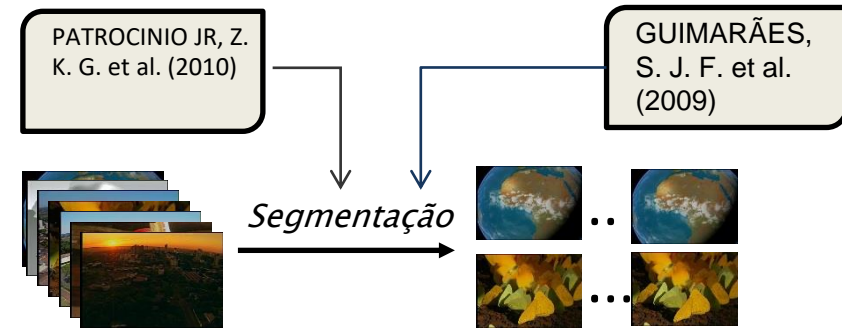


Recuperação de vídeo

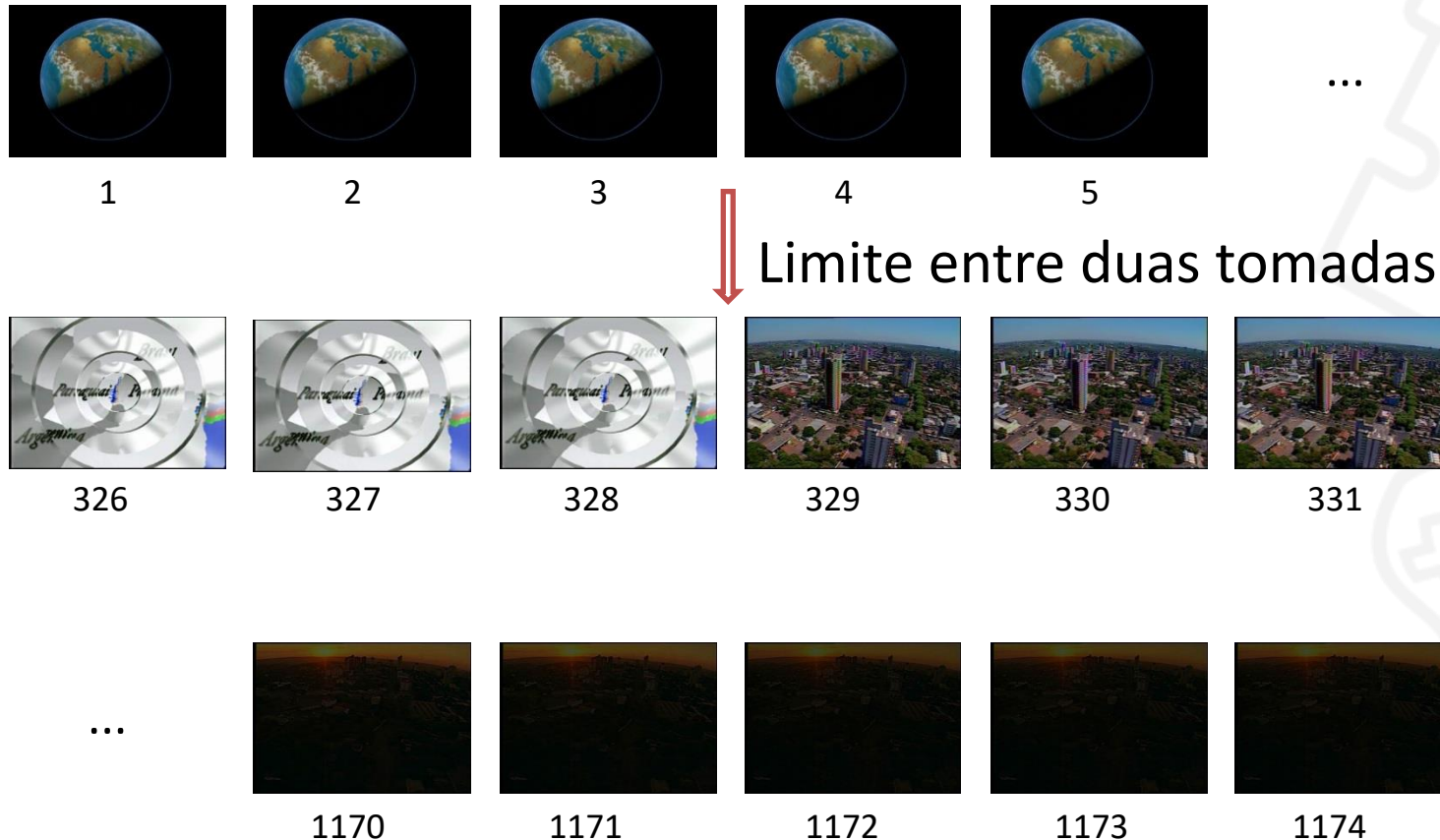


Operações para Recuperação de Vídeos

Recuperação de vídeo



Identificação de Transição Abrupta



Identificação de Transição Gradual



931



932



933



934



935



936



Início da transição



937



938



939



940



941



942



943



944



945



946



947

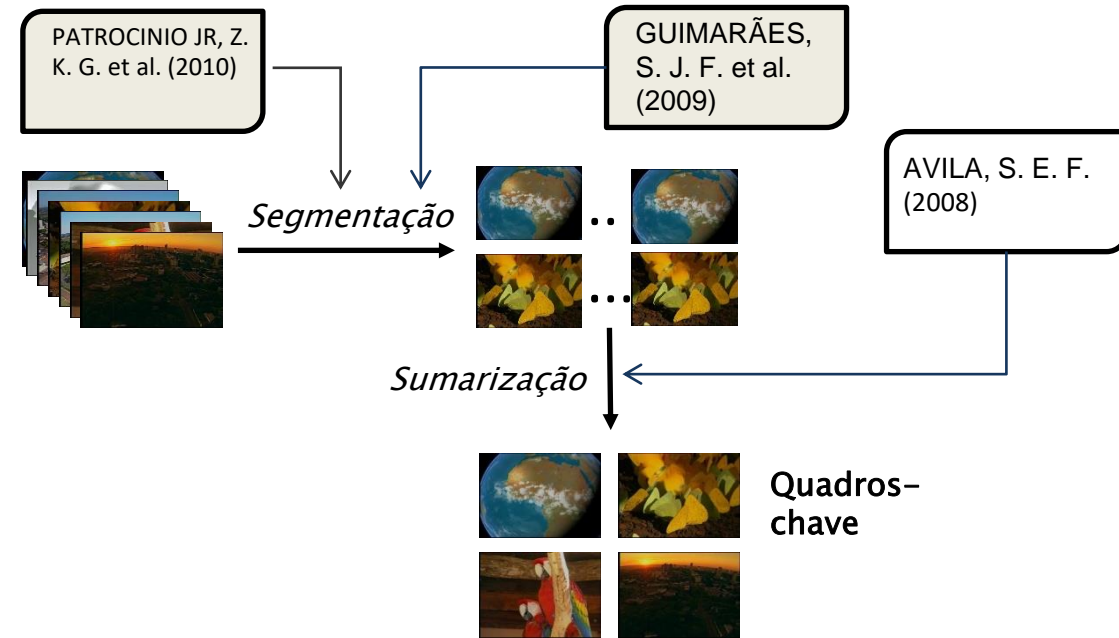


948



Fim da transição

Recuperação de vídeo



Seleção de Quadro-Chave



559



560



561



562



563

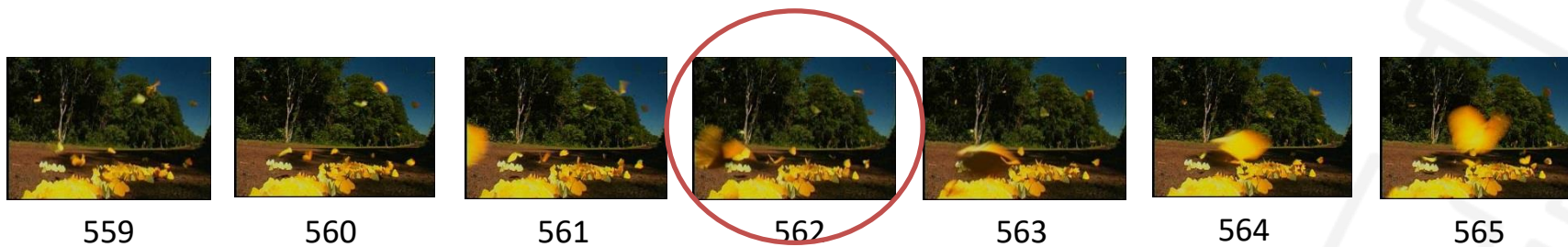


564



565

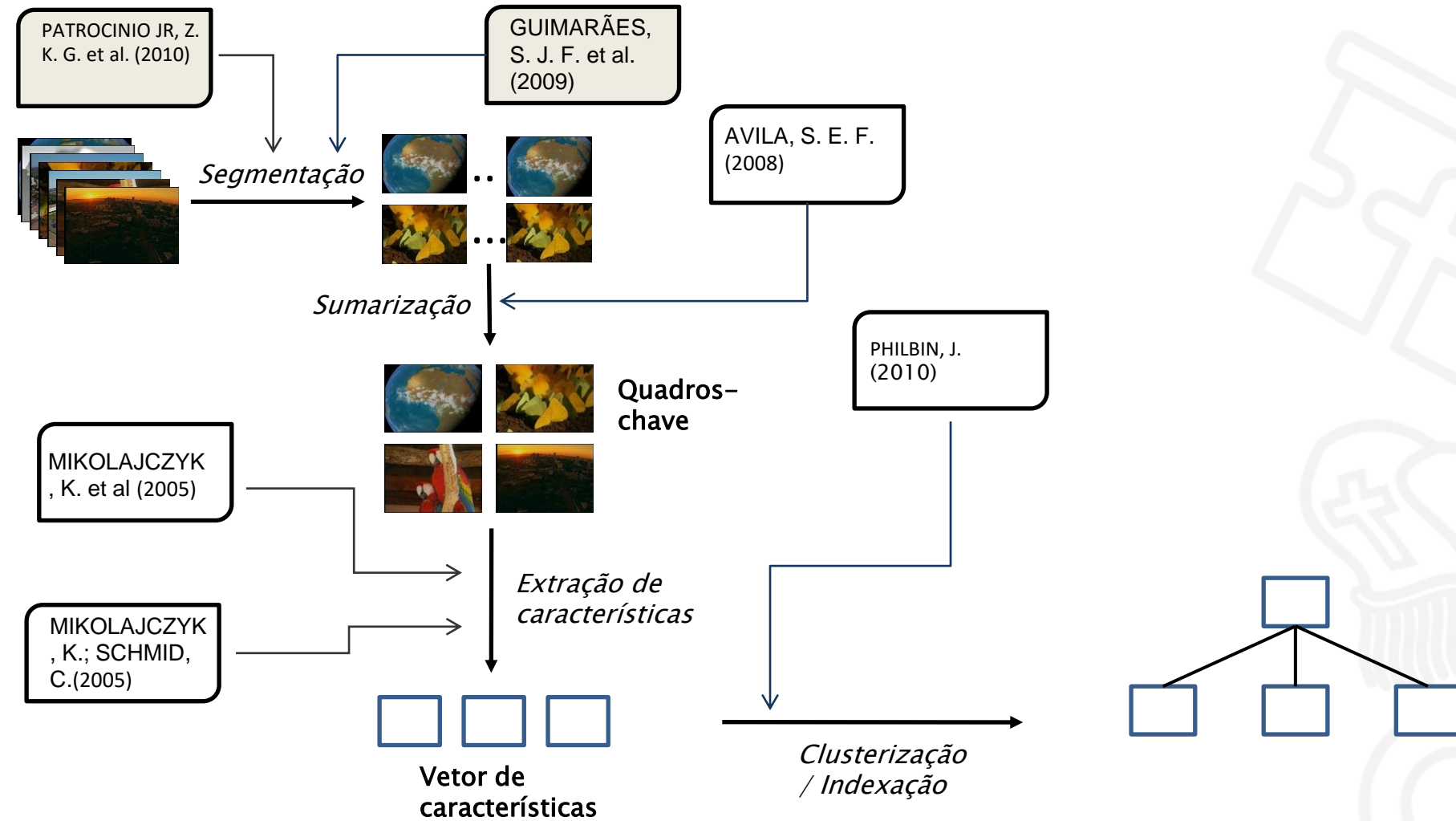
Seleção de Quadro-Chave



Seleção do quadro-chave



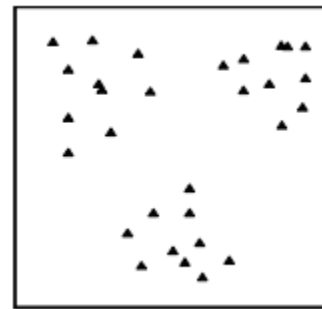
Recuperação de vídeo



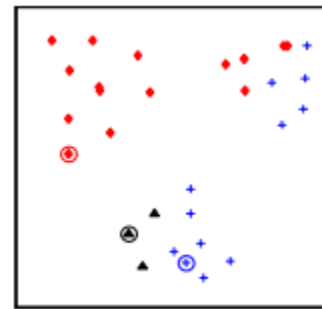
Recuperação de vídeo

- *Clusterização – K-means* (MacQueen, 1967)

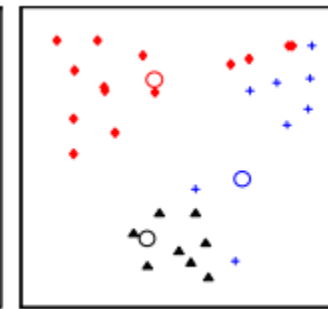
Clusterização para classificação dos descritores em classe de palavras visuais. Para um conjunto de pontos e um valor k , o objetivo é determinar k centroides (k cluster), e a partir de um processo iterativo, minimizar a distância de cada ponto para seu centroide. Gerando ao final do processo, um conjunto k de palavras visuais.



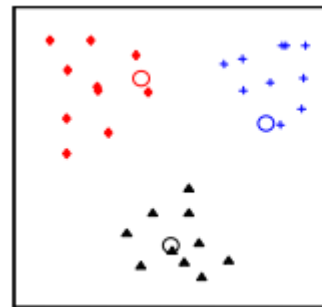
Pontos de entrada



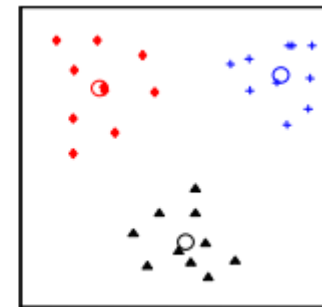
Seleção dos pontos sementes



Iteração 2



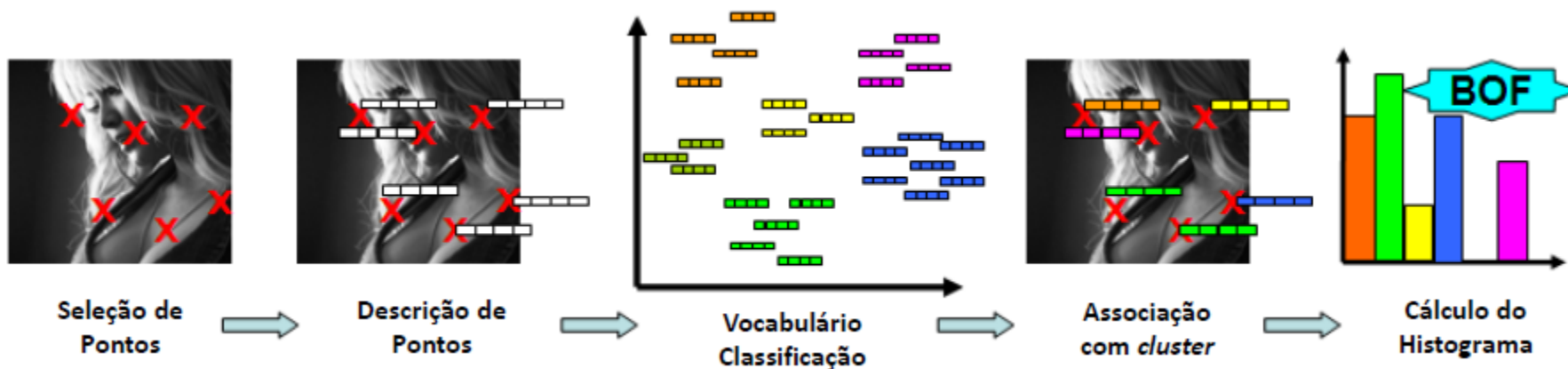
Iteração 3



Clusterização Final

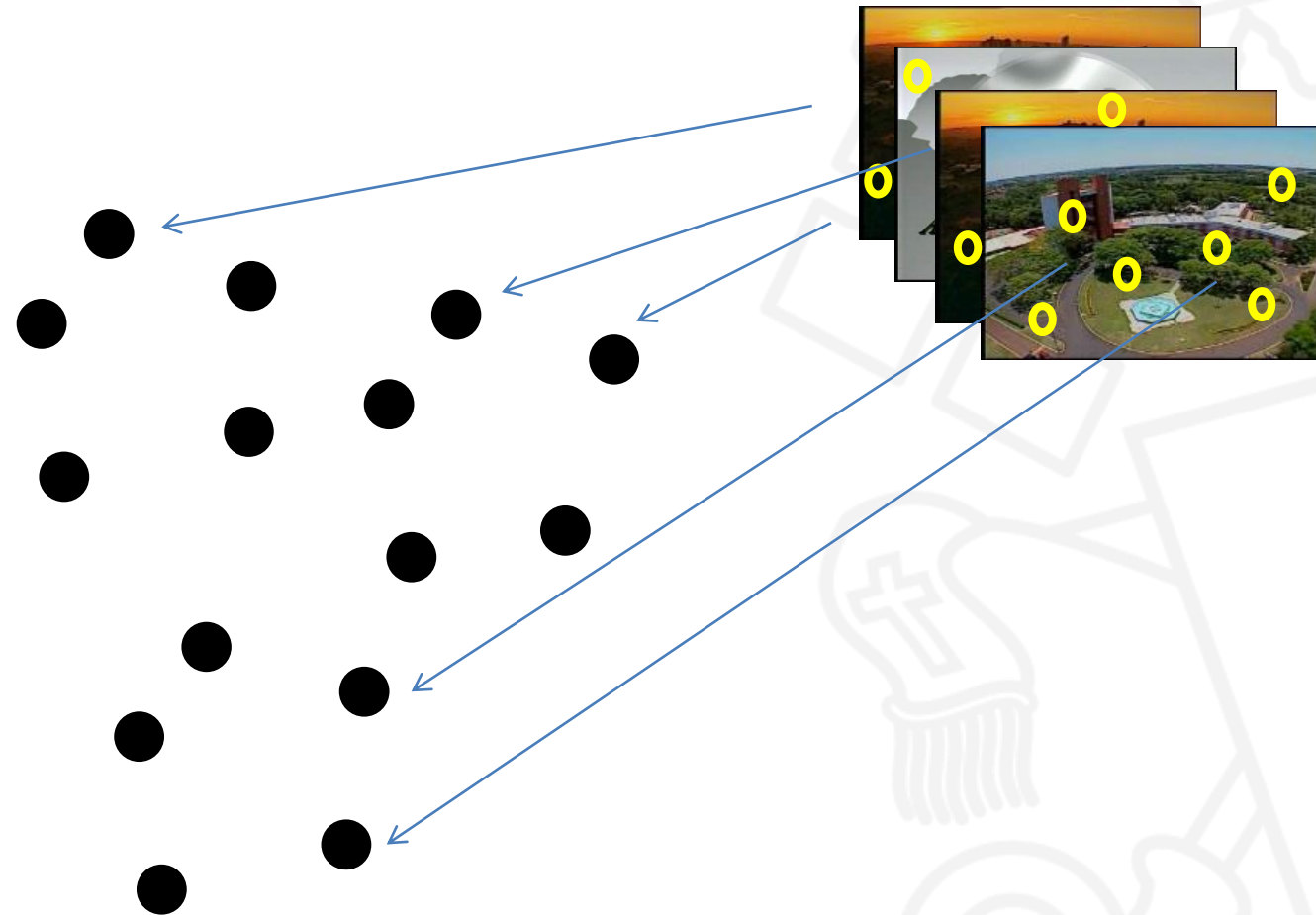
Construção do Vocabulário Visual

- *Bag of Feature*



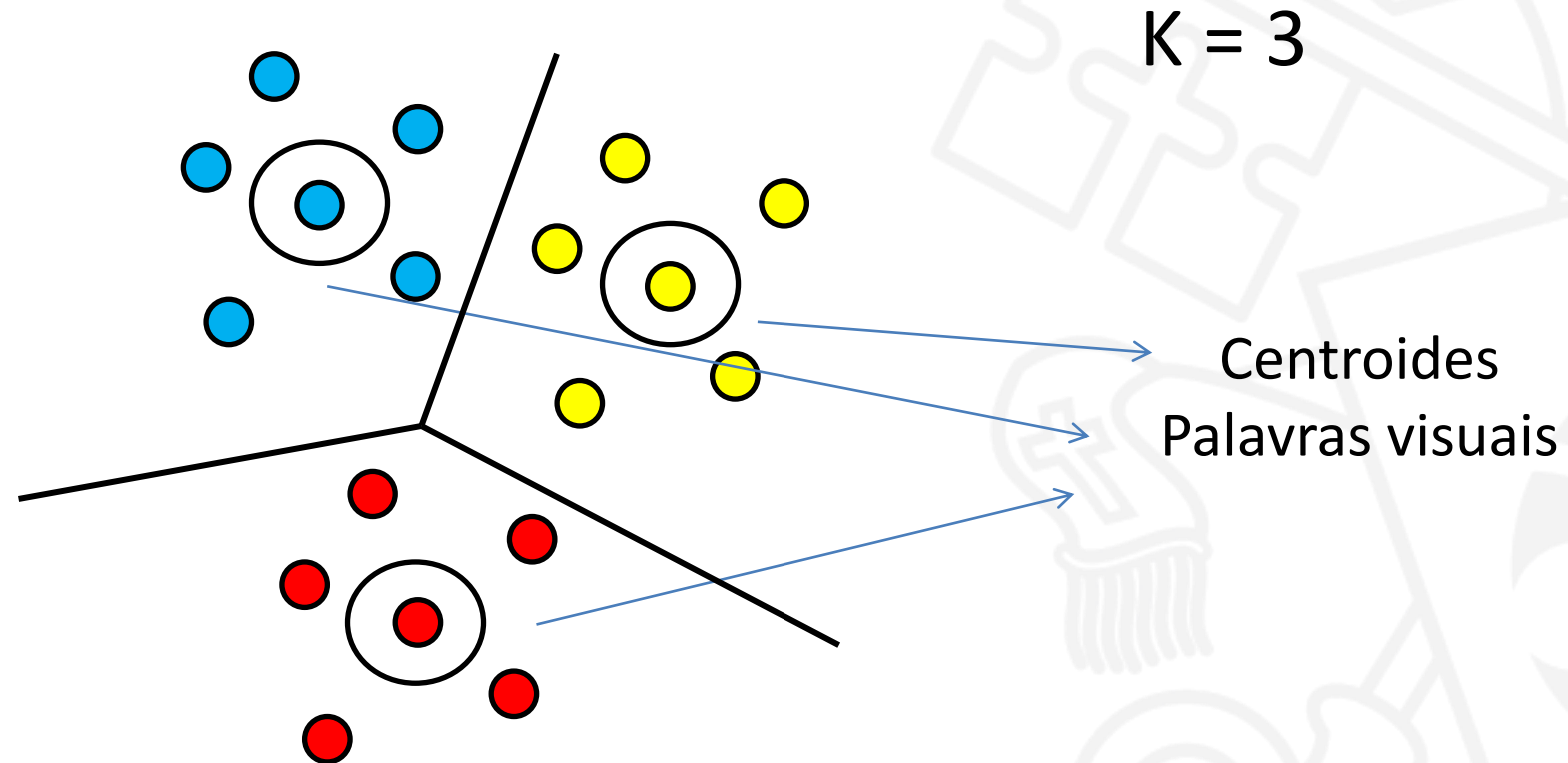
Adaptada de Lopes, Ávila e Peixoto (2009)

Construção do Vocabulário Visual (k-means)

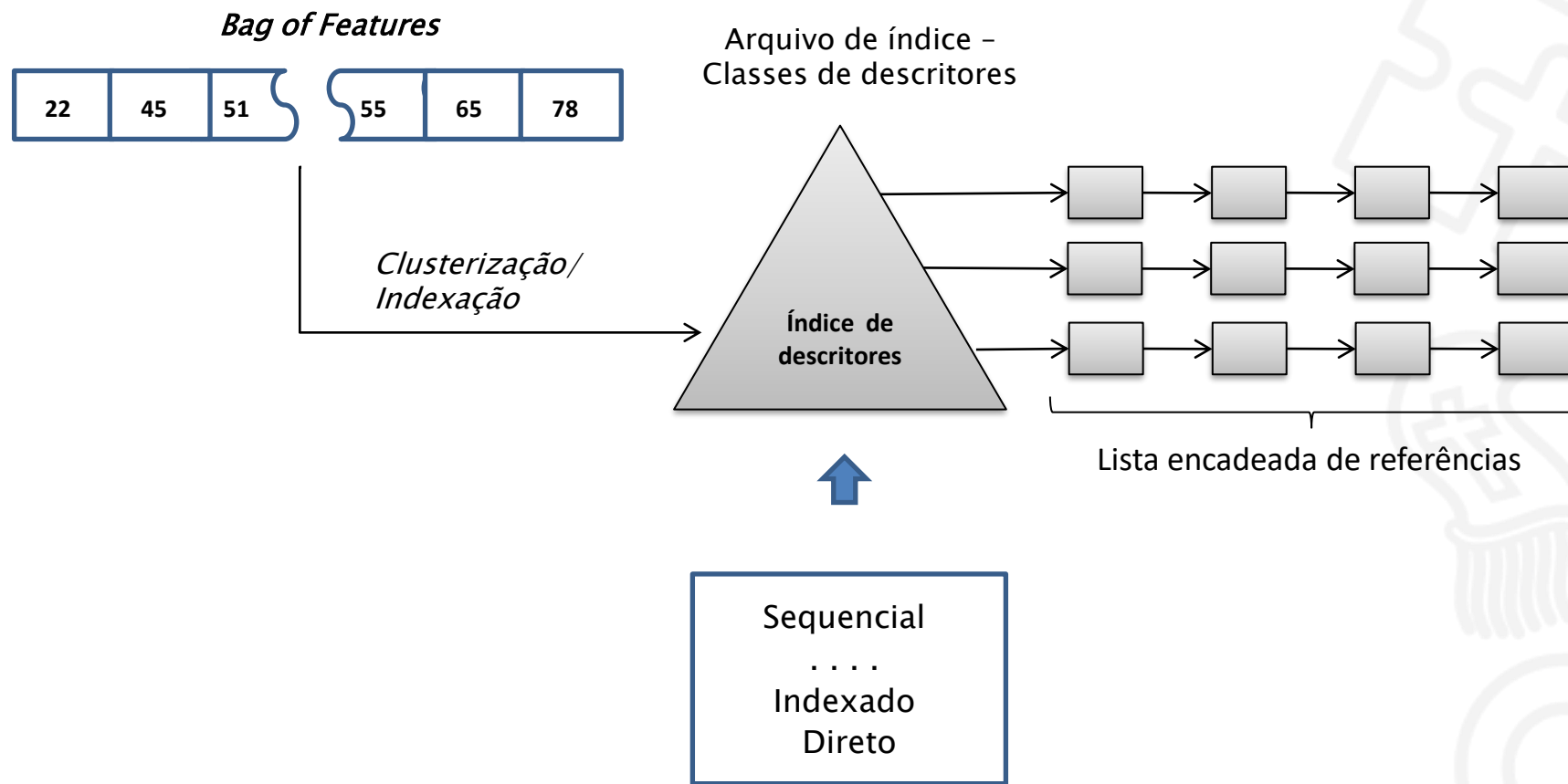


Josef Sivic and Andrew Zisserman. **Efficient Visual Search of Videos Cast as Text Retrieval.** IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, No. 4, April 2009

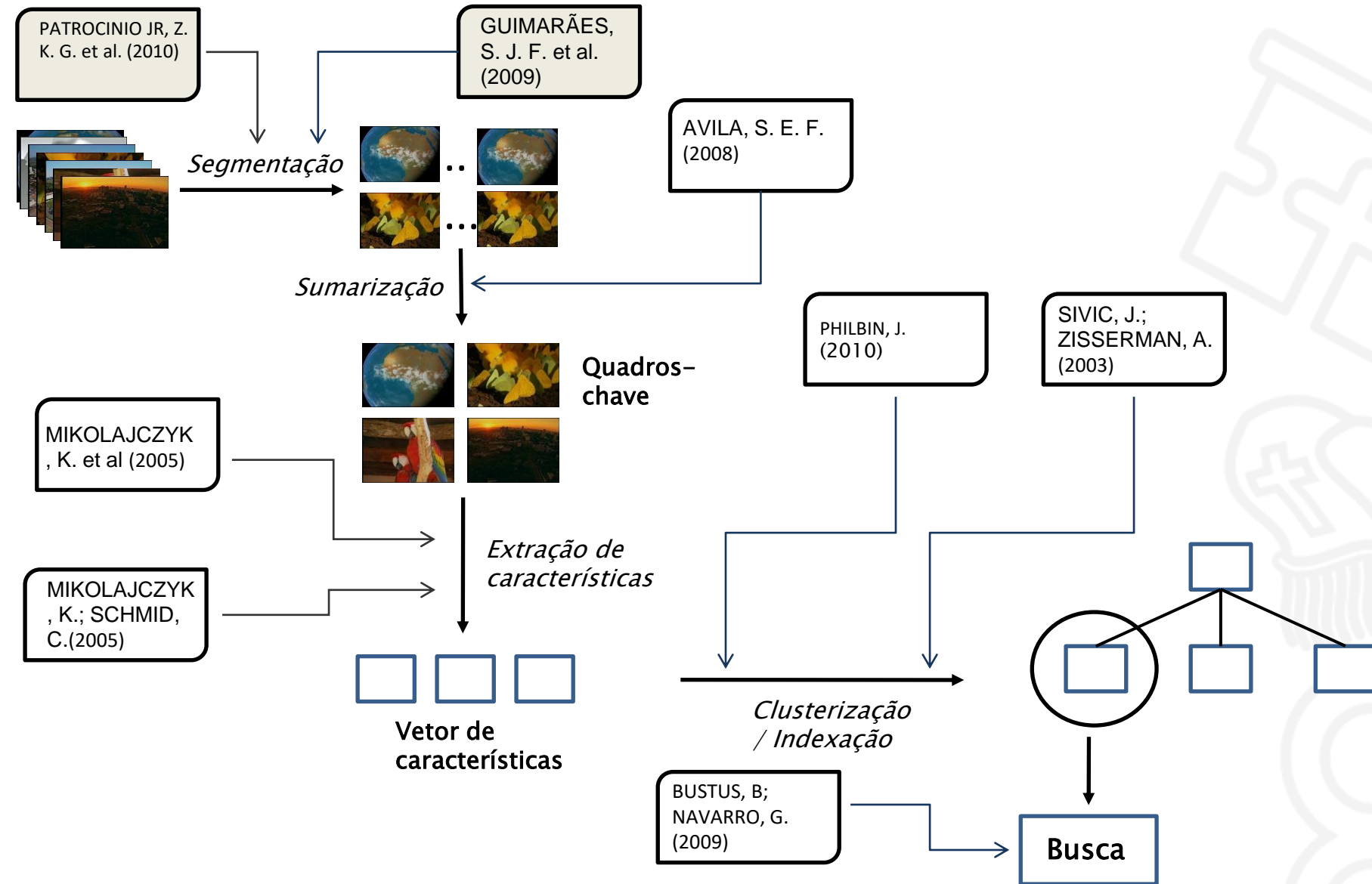
Construção do Vocabulário Visual (k-means)



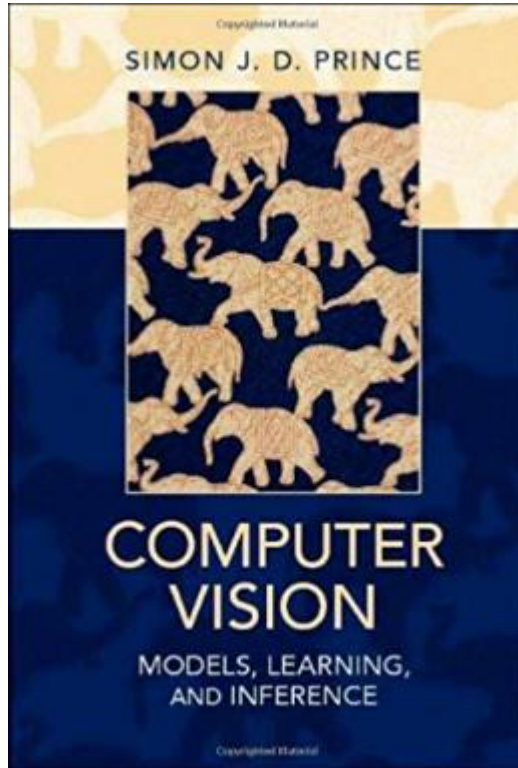
Arquivo Invertido



Recuperação de vídeo



Principais Referências



Simon J. D. Prince. **Computer Vision: Models, Learning, and Inference 1st Edition**