

# Uso de técnicas de reconhecimento de padrões para a detecção de logotipos em imagens

Yuri Tavares dos Passos

**Resumo**—Este artigo descreve um método para detecção de logotipos em imagens usando a metodologia adotada pelos sistemas de reconhecimento de padrões. Para extração de características foi usado PCA no resultado da DFT de janelas  $3 \times 3$ . Para agrupamento foi usado o *k-means*. Os resultados obtidos não foram satisfatórios.

## I. INTRODUÇÃO

Sistemas de detecção de logotipos visam detectar e localizar em imagens logotipos específicos. As imagens de busca podem ser quadros de vídeo, imagens em site de internet ou em um banco de dados. Grandes empresas demandam sistemas desse tipo pois desejam certificar-se de maneira rápida e eficiente que uma emissora tenha cumprido seu contrato de exibição de sua logomarca em sua programação. Estas empresas também necessitam de sistema de detecção de logotipos para averiguar se houve exibição inadvertida de suas logomarcas por outros meios.

A maior dificuldade em detectar logotipos está nas diversas formas em que o logotipo pode aparecer numa imagem. Além de transformações afins que o logotipo sofre na imagem (rotação, escala), existe o problema da iluminação, que pode ocorrer com maior ou menor intensidade em todo o logotipo ou parcialmente. Também observa-se que um mesmo logotipo pode aparecer em cores diferentes. A Figura 1 apresenta exemplos destes casos.

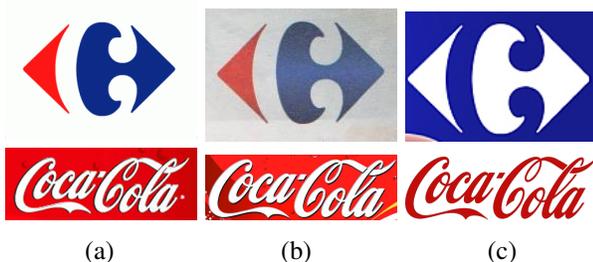


Figura 1. Exemplos de logotipo: (a) Imagem original. (b) Imagem com diferente iluminação. (c) Imagem com cores diferentes

## II. TRABALHOS RELACIONADOS

Detecção de logotipos é uma aplicação de detecção de objetos em imagens. Este tópico é muito pesquisado, pois cada um dos métodos existentes são eficientes na detecção de grupos de objetos específicos, não existindo um método que detecte objetos de um modo geral. Devido à variedade de grupos de objetos existentes, a quantidade de métodos na literatura é alta e cada um deles utilizam técnicas diferentes.

Em [1], é descrito um método para localizar logotipos em vídeos e automaticamente eliminá-los. Inicialmente, as fronteiras do logo é detectada usando um limiar de distância entre os quadros do vídeo, refinados mais tarde aplicando-se uma comparação entre o tamanho de suas bordas em cada quadro. Em seguida um classificador Bayesiano localiza fragmentos de logo (denominados *logolets*). Neste classificador, são usados vetores de características formados pelos valores RGB de pedaços do logo de tamanho  $12 \times 12$  pixels. Após detectado o logotipo, é utilizada a técnica *inpainting* para remover o logotipo. Neste trabalho, a taxa de reconhecimento foi de 87,2% para 226 logotipos.

No trabalho de Kovar *et al* [2], é apresentada uma abordagem para detectar logotipo em vídeos desportivos. Inicialmente são usados detectores de bordas seguidos dos operadores morfológicos erosão e dilatação para remover discontinuidades. Um agrupamento é utilizado para determinar quais grupos conexos são mais relevantes. Em seguida, é usada uma classificação destes grupos mais relevantes, em relação com logotipos armazenados numa base.

Bagdanov *et al* [3] descreve um sistema para detectar e recuperar logotipos em vídeos desportivos. É utilizado pontos de características SIFT para compactar os logotipos. A classificação é feita casando-se o conjunto SIFT da imagem com o conjunto do logotipo a ser procurado. A taxa de reconhecimento foi em torno de 85%.

Em [4], são utilizadas duas técnicas para a detecção de logotipos: Histograma do Gradiente de Bordas (EGH - *Edge Gradient Histogram*) e Histograma de Cor para Bordas de Co-ocorrência (CECH - *Co-ocurrence Edge Color Histogram*). Também são usadas informações de transição de cor entre pares de pixels, a partir da decomposição wavelet deste par.

Schietse *et al* [5] apresenta alguns desafios da pesquisa em busca por logotipos e são apresentadas práticas que devem ser executadas para facilitar o desenvolvimento de tais sistemas.

O presente trabalho visa utilizar o arcabouço de sistemas de reconhecimento de padrões, assim como foi usado em [6] para localização de placas de automóveis, mas com o desafio de tornar o sistema invariante a rotação. Naquele trabalho, é usado um pipeline de reconhecimento de padrões para a geração de uma imagem que classifica os padrões da imagem de entrada. O mapa de classificação gerado pode ser usado para localização de regiões em que podem existir objetos semelhantes ao logotipo. Para isso, usa-se uma amostra do logotipo e aplica-se o mesmo algoritmo de geração do mapa de padrões. A detecção passa a ser a comparação de dois mapas discretizados para um conjunto de poucos elementos.

### III. METODOLOGIA

Neste trabalho será apresentado um método cujo objetivo é verificar se um dado logotipo aparece em uma imagem. Portanto, este método possui dois argumentos: o logotipo a ser buscado e a imagem para busca. A solução para este problema de detecção de logotipos em imagens será baseado no uso de técnicas de reconhecimento de padrões. Um sistema de reconhecimento de padrões estatístico (ou sistema RP) pode ser dividido em duas etapas: a etapa de treinamento e a etapa de teste. Cada uma dessas etapas é dividida em três módulos como mostra a Figura 2.

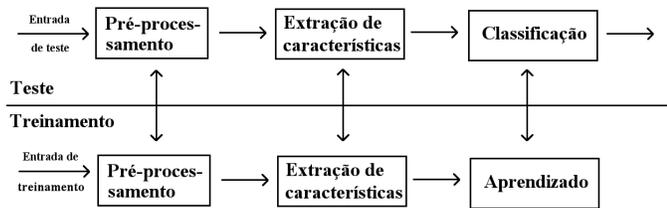


Figura 2. Esquema de um sistema de reconhecimento de padrões, adaptado de [7]

Na Figura 2 nota-se que as duas fases possuem uma *entrada*. Tanto para a etapa de teste quanto para etapa de treinamento, esta entrada corresponde a qualquer tipo de informação dentro da qual deseja-se reconhecer algum padrão. Exemplos de entradas são imagens, textos ou linhas de uma tabela e exemplos de padrões que podem ser identificados nestas informações são, respectivamente, grupos de objetos salientes numa imagem, palavras que se repetem e grupos de informações mais importantes numa tabela. Formalmente, a entrada é um vetor de dimensão  $d$ , isto é, um vetor com  $d$  características referentes a informação. Esta entrada é conhecida como *vetor de características*.

A etapa de treinamento corresponde a fase onde as entradas devem ser apresentados ao sistema de reconhecimento de padrões para que ele possa aprender o comportamento dos padrões. Nesta fase, o sistema associará esta entrada a um grupo ou classe.

A etapa de teste compreende a fase em que as entradas devem ser associadas a um grupo, aprendido na etapa de treinamento. É nessa fase que o aprendizado do sistema RP é testado, com intuito de verificar a eficiência deste aprendizado. Após a etapa de teste, alguns sistemas podem executar novamente a etapa de treinamento, pois o aprendizado pode não ter sido satisfatório e, portanto, uma apresentação de novos vetores de características pode ser útil para melhorar a classificação da etapa de teste.

As duas etapas são divididas em três blocos que representam módulos que devem ser executados num sistema de reconhecimento de padrões. As setas bidirecionais no esquema da Figura 2 ligando pares destes blocos em etapas diferentes significa que o método usado em um bloco numa etapa deve ser semelhante ao utilizado na outra etapa.

O primeiro módulo para ambas as etapas é o *pré-processamento*. Neste módulo, o vetor de características passa

por uma transformação que serve para aumentar ou diminuir a dispersão das informações, não alterando a relevância individual de cada vetor.

O módulo seguinte é a *extração de características*. Neste bloco, as características mais relevantes provenientes do vetor de entrada é detectada e extraída. Formalmente falando, o vetor de características contendo  $d$  dimensões, advindo do pré-processamento (ou não, caso tenha-se optado por não usar pré-processamento), será transformado em um vetor com  $m$  dimensões, tal que  $m < d$ . O propósito principal deste módulo é diminuir a dimensão do vetor de características, preservando as informações mais relevantes, para que o módulo seguinte possa trabalhar com mais eficiência. Tanto na etapa de teste quanto na de treinamento, os métodos utilizados para a extração de características devem ser semelhantes.

Para a etapa de treinamento, o módulo seguinte é o aprendizado e para a etapa de teste, é a classificação. O aprendizado refere-se a capacidade do sistema reconhecedor de padrões assimilar o padrão associado a um conjunto de vetores. Após assimilado, o sistema será capaz de relacionar um vetor de características qualquer a este padrão. O papel da classificação é relacionar um vetor de características qualquer a um padrão que foi concebido pela fase de aprendizado. Em outras palavras, a fase de aprendizado irá associar um conjunto de vetores a uma classe e, após aprendidas as classes, a fase de classificação deverá associar uma dessas classes a um vetor dado como entrada. O intercâmbio indicado pelas setas bidirecionais no esquema da Figura 2 indica justamente essa dependência entre as classes aprendidas pelo sistema e as classes usadas na classificação.

Baseando-se no modelo de um sistema RP, três problemas emergem. Estes problemas dizem respeito à qual técnica ou algoritmo deverá ser usada para: (i) pré-processar a entrada; (ii) extrair as características; e (iii) aprender e classificar os padrões. Cada problema necessitará de diferentes algoritmos para alcançar o seu objetivo. Neste trabalho serão apresentadas as técnicas utilizadas para cada uma destas três fases com o intuito de reconhecer logotipos em imagens.

#### A. Pré-processamento e extração de características

Como pré-processamento, as imagens coloridas serão convertidas em imagens de nível de cinza. Após esta etapa, algumas amostras do logotipo deverão ser apresentadas para o algoritmo, que irá retirar informações a seu respeito. Após coletadas estas informações, o algoritmo irá verificar se o logotipo dado se encontra na imagem.

Considere que são fornecidas  $M$  imagens diferentes do logotipo. O número de linhas e colunas para cada imagem é diferente entre si. Para cada imagem do logotipo deverá ser retiradas janelas de tamanho  $N \times N$  pixels, com salto de  $N$  pixels para esquerda e para baixo. Para cada janela será aplicada o logaritmo do módulo da transformada de Fourier Discreta (DFT). Este resultado é linearizado para formar um vetor de tamanho  $N^2$ . A DFT é usada devido ao fato da visão humana ser sensível a mudanças de frequência [8]. Para o conjunto de vetores será aplicado a PCA (*Principal Component Analysis*) [9] para encontrar os autovetores mais

relevantes (ordenados decrescentemente pelos autovalores). As informações obtidas dos logotipos estarão sintetizadas nestes autovetores. O uso do PCA é justificado pela compactação da informação que pode ser obtida, diminuindo a dimensão dos vetores. Além disso, o objetivo do uso do PCA é descrever quais são os pesos mais importantes atribuídos para cada subimagem do logotipo.

O método descrito acima para obtenção dos autovetores mais importantes será replicada para as mesmas imagens, mas dessa vez rotacionadas em  $\theta$  graus. Assim, serão gerados autovetores para a base original (isto é, a base rotacionada de 0 graus) e a base rotacionada de  $\theta$  graus. O intuito deste trabalho é testar se existe alguma relação entre os autovetores para as imagens de logotipo fornecidas e os autovetores das imagens rotacionadas. Caso exista uma maneira de relacionar estes autovetores, então não será necessário repetir este processo para imagens rotacionadas.

### B. Aprendizado e Classificação

Os vetores resultantes da compactação por PCA serão aplicados no algoritmo k-means. Com isso,  $K$  grupos poderão ser gerados. Com isso, a fase de classificação terá o papel de reduzir blocos de subimagem de tamanho  $N \times N$  para apenas um valor entre 1 e  $K$ .

Este processo de classificação é feito tanto na imagem para busca quanto no logotipo. A fase de detecção será realizada casando-se o resultado da classificação do logotipo com o resultado da classificação para alguma subimagem da imagem de busca.

## IV. RESULTADOS E DISCUSSÕES

Como não existe uma base de imagens consolidada para o teste de sistemas de detecção de logotipos, foram obtidas 100 imagens em formato JPEG contendo o logo da Coca-Cola<sup>®</sup>. Dentre os logotipos pesquisados, as imagens da Coca-Cola eram mais variadas quanto à posição, escala, rotação e iluminação. Foram escolhidas para treinamento sete imagens contendo somente o logo, todas com resoluções diferentes entre si (Figura 3).

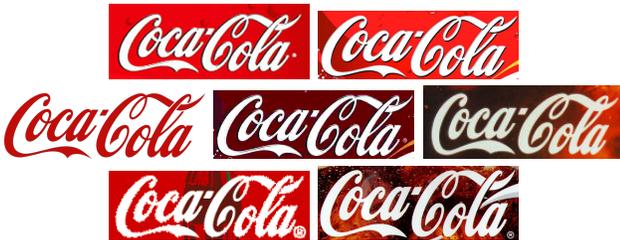


Figura 3. Imagens para treinamento

Para estas sete imagens foram extraídas janelas  $N \times N$  com saltos de tamanho  $N$  para esquerda e para baixo. Foram usados os seguintes valores para  $N$ : 3, 4, 8, 9, 16, 27. Para cada subimagem extraída, foi aplicado o logaritmo da magnitude da sua transformada de Fourier. O resultado foi convertido para um vetor de tamanho  $N^2$ . Para cada  $N$ , foram extraídos autovetores para a base com as imagens rotacionadas com os

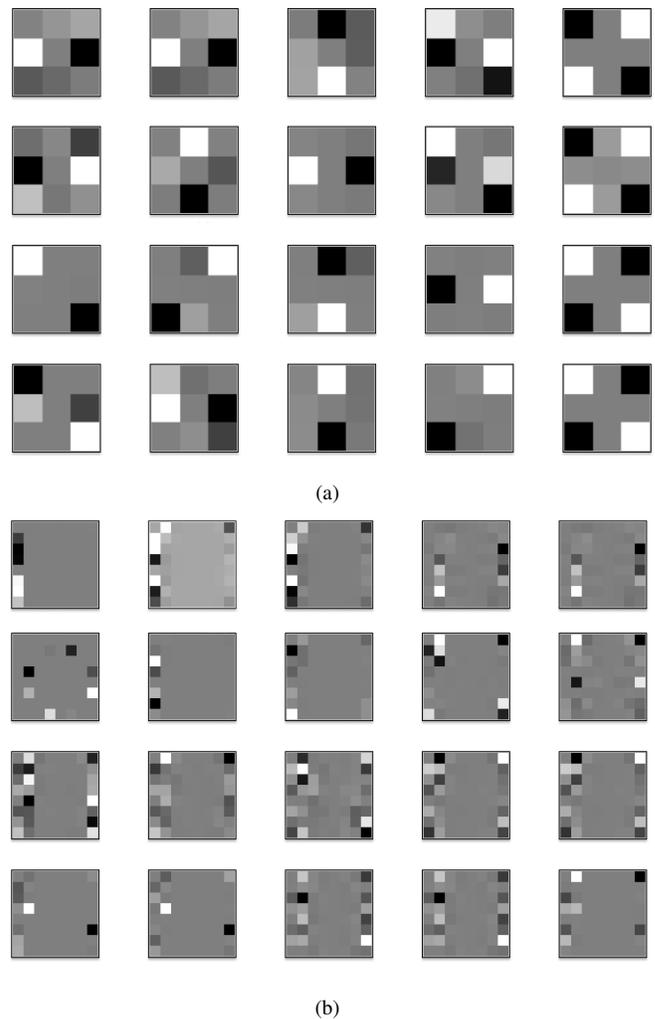


Figura 4. Autovetores extraídos de subimagens de tamanho  $N = 3$  em (a) e  $N = 8$  em (b). Para as duas imagens, cada linha representa os autovetores para a base rotacionada de  $\theta = 0, 45, 90, 180$ . Cada coluna representa um autovetor, sendo que o mais à esquerda é o que possui o maior autovalor e o mais à direita o que possui menor valor

seguintes valores de  $\theta$  (em graus): 0, 45, 90, 180. Isto totalizou 24 conjuntos de vetores diferentes distribuídos para cada  $N$  e  $\theta$ . Para cada um destes conjuntos, foram extraídos os 5 autovetores mais importantes, usando o método descrito na Seção III-A porque a soma dos 5 primeiros autovalores representava mais de 95% da soma de todos os autovalores.

Alguns autovetores resultantes apresentavam componentes imaginários. No entanto, considerando a razão da soma dos módulos dos valores imaginários sobre a dos valores reais, este valor era da ordem de  $10^{-6}$ . Em vista disso, foi considerado apenas os valores reais.

Analisando os autovetores como imagens de tamanho  $N \times N$ , foi verificado que não existe um padrão entre autovetores para um  $N$  fixo e as bases de imagens rotacionadas com  $\theta$  variável. Para  $N = 3$ , os autovetores parecem sofrer rotação, mas o ângulo de rotação não é o mesmo aplicada para a base de imagens que foi oriundo. A Figura 4 apresenta os 5 autovetores para  $N = 3$  e  $N = 8$ , para cada  $\theta$ .

Para testar o k-means, foram considerados apenas os 2

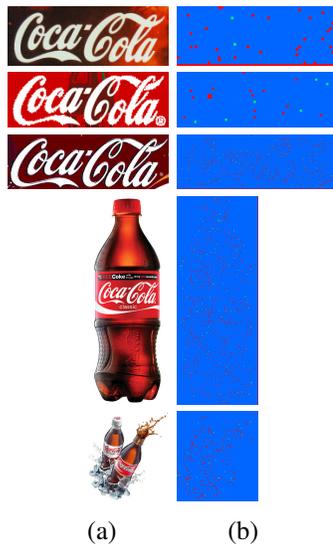


Figura 5. Resultado da classificação. (a) Imagem Original, (b) Imagem contendo os grupos de cada janela extraída.

primeiros autovetores para  $N = 3$ , visto que seus autovalores garantiam 90% da soma de todos os autovalores. O k-means foi usado na união dos conjuntos de vetores gerados de cada base de imagens, isto é, da base para  $\theta = 0, 45, 90$  e  $180$ . Para  $K = 5$ , os centróides encontrados foram:

$$\begin{array}{ll} -5.952785 \cdot 10^{-13} & -3.518813 \cdot 10^{-13} \\ 5.451757 \cdot 10^{-11} & 2.618009 \cdot 10^{-11} \\ 1.912260 \cdot 10^{-12} & 7.270050 \cdot 10^{-13} \\ 1.211157 \cdot 10^{-13} & -1.782356 \cdot 10^{-14} \\ -8.835791 \cdot 10^{-12} & -3.506291 \cdot 10^{-12} \end{array}$$

Na fase de teste, foram utilizadas as imagens dos rótulos com  $\theta = 0$ . Para as imagens de teste, utilizou-se o processo de janelamento e extração de característica descrito na Seção III-A com  $N = 3$ . Os vetores extraídos eram multiplicados aos 2 autovetores mais importantes, diminuindo sua dimensão de  $N^2$  para 2. O vetor resultante era comparado com os 5 centróides acima. O grupo associado a este vetor era igual a  $i \in \{1, 2, 3, 4, 5\}$ , isto é,  $i$  é a linha correspondente ao centróide de menor distância. Uma cor distinta foi associada para cada grupo. A Figura 5 mostra o resultado do agrupamento para 5 imagens contendo o logo. Percebe-se que a classificação não foi boa o suficiente para caracterizar o logotipo nas imagens. Os grupos do k-means ficaram muito próximos devido aos baixos valores obtidos. Isto influenciou na escolha dos grupos, fazendo que os grupos vencedores fossem apenas dois grupos.

## V. CONCLUSÕES

O método descrito neste artigo visa detectar logotipos em imagens usando a metodologia adotada pelos sistemas de reconhecimento de padrões. Para extração de características foi usado PCA no resultado da DFT de janelas extraídas de imagens de logotipo. Para agrupamento foi usado o *k-means*. Para extração de características foram testadas janelas  $3 \times 3$ ,  $4 \times 4$ ,  $8 \times 8$ ,  $9 \times 9$ ,  $16 \times 16$  e  $27 \times 27$ . Também foi testado

se os autovetores treinados para diferentes rotações refletiram este comportamento. Este comportamento não foi notado.

Os resultados obtidos não foram satisfatórios. Os grupos do k-means eram muito próximos devido aos baixos valores obtidos. Isto influenciou na escolha dos grupos, fazendo que os grupos vencedores fossem apenas dois grupos.

Outro fator que influenciou nos testes foi o fato das imagens encontradas estarem em formato JPEG. A compressão JPEG elimina informações do domínio da frequência. Isto pode ter afetado na PCA para os vetores extraídos.

## REFERÊNCIAS

- [1] W.-Q. Yan, J. Wang, and M. S. Kankanhalli, "Automatic video logo detection and removal," *Multimedia Systems*, vol. 10, no. 5, pp. 379–391, 2005.
- [2] B. Kovar and A. Hanjalic, "Logo detection and classification in a sport video: video indexing for sponsorship revenue control," in *Proc. SPIE*, vol. 4676, San Jose, CA, USA, 2001.
- [3] A. D. Bagdanov, L. Ballan, M. Bertini, and A. D. Bimbo, "Trademark matching and retrieval in sports video databases," in *MIR'07*, Augsburg, Bavaria, Germany, 2007.
- [4] A. Hesson and D. Androustos, "Logo and trademark detection in images using color wavelet co-occurrence histograms," in *ICASSP*, Las Vegas, NV, 2008.
- [5] J. Schietse, J. P. Eakins, and R. C. Veltkamp, "Practice and challenges in trademark image retrieval," in *CIVR'07*, Amsterdam, The Netherlands, 2007.
- [6] Y. T. dos Passos, "Uma abordagem bio-inspirada para o problema de reconhecimento de placas de licença," Monografia de graduação, Universidade Federal de Sergipe, 2008.
- [7] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.
- [8] L. K. Cormack, "Computational models of early human vision," in *Handbook of Image And Video Processing*. Academic Press, 2000, pp. 271–288.
- [9] J. Shlens, "A tutorial on principal components analysis," Disponível em <http://www.sn1.salk.edu/shlens/pub/notes/pca.pdf>, 2005, Última visita em 24 de maio de 2008.