

Recuperação de Imagens 2D Utilizando Técnicas de Inteligência Artificial

Túlio C. Bender, Fernando Santos Osório

Universidade do Vale do Rio dos Sinos – Unisinos

Programa Interdisciplinar de Pós-Graduação
em Computação Aplicada

Av Unisinos, 950

São Leopoldo, RS – Brasil

tulio@euler.unisinos.br

osorio@exatas.unisinos.br

ABSTRACT

This work presents an approach based on Artificial Intelligence technics for the task of recognition and retrieval of images in the domain of 2D images (bitmaps) stored in images databases. The main aspects are the use of machine learning and image processing techniques (color manipulation) to identify similarities in relation to the image's color patterns.

KEYWORDS image retrieval, machine learning, artificial neural networks.

RESUMO

Este trabalho apresenta uma abordagem baseada em técnicas de Inteligência Artificial voltada para tarefas de reconhecimento e recuperação de imagens 2D (bitmaps) em base de dados de imagens. As principais características deste trabalho são o emprego de técnicas de machine learning e de processamento de imagens, que

permitem manipular cores e identificar semelhanças em relação aos padrões de cores de imagens.

PALAVRAS-CHAVE recuperação de imagens, machine learning, redes neurais artificiais.

1 Introdução

Atualmente com o aumento considerável de bases de dados de imagem e vídeo devido a inovações tecnológicas tornou-se necessário manter algum tipo de mecanismo de indexação para posterior recuperação de dados. Executar esta indexação manualmente é uma tarefa sensível a interpretações subjetivas, é laboriosa e sujeita a erros (1), (4), (3), (5). Métodos automáticos de indexação e recuperação baseados em algum tipo de característica (cor, forma, textura ou região de uma imagem) são interessantes neste contexto porque reduzem a intervenção humana. A idéia de recuperação de imagens por exemplos é uma extensão da idéia de recuperação de páginas web que contenham as palavras-chave especificadas na pesquisa (por exemplo Google, Altavista e outros). No caso de imagens, busca-se recuperar numa base de dados de imagens aquelas que contenham o elemento gráfico procurado como por exemplo uma determinada cor, textura ou forma. As técnicas de recuperação de imagens vem sendo desenvolvidas desde 1970 principalmente pelas áreas de gerenciamento de banco de dados (indexação textual) e visão computacional (3). No método baseado em indexação textual existe um DBMS que gerencia a informação semântica anexada por uma pessoa a imagem. Já nos *Content Based Image Retrieval* (CBIR) procura-se extrair automaticamente das imagens características que as identifiquem. A partir destas características da imagem-exemplo procura-se por

características similares nas imagens disponíveis numa base de pesquisa. As melhores similaridades são apresentadas como resultado da pesquisa. Um exemplo de um sistema CBIR é o IKONA (2), que está disponível na Internet para demonstrações.

2 Abordagem Usual para Recuperação de Imagens

A abordagem utilizada nos sistemas CBIR pode ser dividida em dois subsistemas: geração da base de dados de imagens de pesquisa e o subsistema de pesquisa nesta base. No subsistema de geração da base de dados são aplicados os métodos de extração de características (*feature extraction*) nas imagens da base e o resultado é armazenado nesta base para posterior comparação. O subsistema de pesquisa aplica os mesmos métodos de extração de características na imagem-exemplo fornecida ao sistema pelo usuário e depois aplica métodos de comparação (*matching*) de características entre as características da imagem de pesquisa e as armazenadas na base de dados de imagens. Na Figura 1 temos um exemplo de um esquema típico de CBIR atual.

Podemos classificar as Bases de Dados de Imagens (BDI) quanto ao seu conteúdo gráfico como:

- Base de Dados Homogêneas: ou de domínio específico, são BDIs nas quais as imagens que a compõem pertencem a uma mesma categoria, elas possuem uma similaridade implícita entre as imagens como, por exemplo, imagens de impressões digitais, retina ou faces.
- Base de Dados Heterogêneas: nestas bases não existe necessariamente uma similaridade implícita entre as imagens que compõem a base. A similaridade neste caso é subjetiva, dependendo do usuário, do contexto e do tipo de aplicação. Diferentes usuários podem executar pesquisas nesta base com objetivos muito diversificados. Neste caso a obtenção de uma "assinatura" que caracterize a base não é trivial.

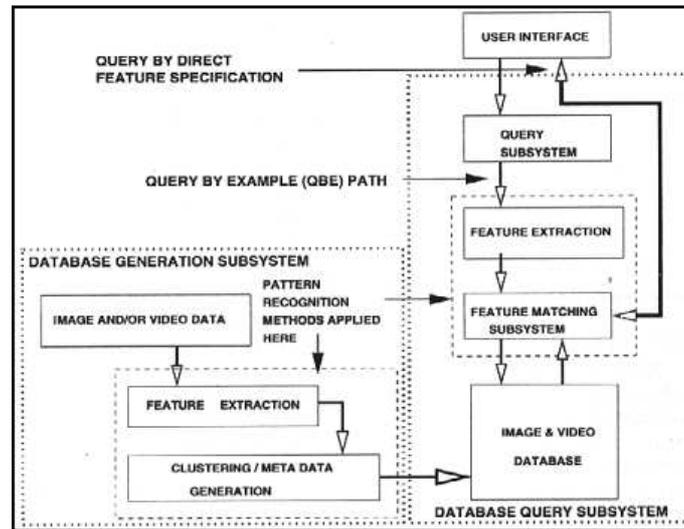


Figura 1. Esquema típico de um CBIR atual, extraído de Antani 02.

O objetivo da extração de características é obter um tipo de "assinatura", uma transformação do sinal gráfico no domínio espacial para outro domínio mais apropriado (domínio da frequência, por exemplo), da imagem-exemplo e das imagens que estão na base de exemplos. Então se aplica uma função de comparação obtendo-se uma medida de similaridade que será utilizada na escolha das melhores imagens. É usual que esta extração de características ocorra em duas fases:

- obter um vetor de características que identifique algumas das propriedades da imagem e armazenar este vetor em um banco de dados de características;
- a partir da imagem de exemplo, calcular o seu vetor de características e comparar com os vetores armazenados no banco de dados recuperando os que possuem os melhores índices de similaridades.

Os métodos CBIR dependem inteiramente desta extração de carac-

terísticas. As características mais comumente utilizadas são: cor, textura e forma, bem como uma combinação dessas características pode ser utilizada. No domínio visual de características, estas podem ser classificadas em: características gerais e de domínio específico. Nas características gerais enquadram-se cor, textura e forma e nas de domínio específico podemos ter, por exemplo, características referentes a impressões digitais e faces humanas.

A comparação das características está baseada numa função que retorna a diferença entre as características extraídas da imagem de exemplo com as características de cada imagem da base de procura, portanto **executando uma comparação um-a-um**. As imagens com as menores diferenças são então retornadas ao usuário da pesquisa como sendo as melhores candidatas. A escolha dos métodos de extração e comparação utilizados, que melhor se adaptam a uma determinada base de dados, é normalmente uma responsabilidade de quem projeta e prepara a base de imagens a ser pesquisada.

Um exemplo de uma pesquisa (onde foi selecionada a última imagem da tela) no sistema Ikona é apresentado na Figura 2 e o resultado desta pesquisa na Figura 3. Uma apresentação mais detalhada sobre CBIRs, tipos de características, extração e comparação de características pode ser obtido em (6).

3 Aplicação da Inteligência Artificial

Neste trabalho estamos aplicando técnicas de Inteligência Artificial (IA) para executar o reconhecimento e recuperação de imagens 2D a partir de características extraídas automaticamente das imagens. A aplicação faz uso de técnicas de machine learning da IA com aprendizado supervisionado, onde optou-se pelo uso de Redes Neurais Artificiais (RNA) do tipo MLP com arquitetura *Cascade Correlation*. A escolha desta técnica é interessante pela capacidade das RNAs em aprender a identificar padrões a partir de bases de exemplos e pela sua capacidade de obter uma generalização do conhecimento aprendido. Através do uso de RNAs se pretende obter

uma generalização do(s) modelo(s) de imagem procurado. Inicialmente é realizada uma extração de características (principalmente referentes à cor) de uma coleção de imagens-exemplo pertencentes a uma mesma classe (por exemplo diferentes tipos de latas de coca-cola), que servirão de base para o processo de aprendizado e posteriormente de busca de imagens em bases de dados. A generalização obtida pela rede neural pode ser entendida como uma "assinatura" que caracteriza aquela coleção de imagens-exemplo. Deste modo não estamos mais executando uma comparação uma-a-um mas sim uma **comparação de um-com-N**, o que é um diferencial em relação às técnicas atuais para recuperação de imagens. Esta assinatura constitui-se dos pesos sinápticos que são ajustados durante o treinamento de uma RNA. Cada neurônio da RNA determina um hiperplano de separação no espaço de padrões. Um padrão situado em um lado deste plano pertence a uma classe e, se estiver do outro lado, a outra classe. Conforme o treinamento da RNA estes hiperplanos formam superfícies de classificação que podem ser abertas ou fechadas em torno de um grupo de padrões que representam uma classe.

3.1 Superfícies de Separação em RNAs MLP

Os pesos sinápticos de cada neurônio de uma RNA do tipo MLP ajustados durante o treinamento determinam superfícies de separação no espaço de padrões. A partir destas superfícies a RNA pode classificar um padrão como pertencente a uma classe ou outra. Conforme os padrões de treinamento utilizados e a quantidade de neurônios, estas superfícies de separação podem formar uma região fechada em torno das classes que se deseja aprender ou uma superfície aberta. No caso do treinamento resultar numa superfície de separação fechada, a RNA poderá tanto classificar corretamente um padrão quanto executar uma rejeição confiável de um padrão que não seja conhecido para esta rede. Para exemplificar, treinou-se uma RNA MLP com a topologia 2-5-1 (duas entradas, cinco neurônios na camada oculta e um neurônio de saída) utilizando o algoritmo de aprendizagem backprop. Na Figura 4 tem-se um exemplo de treinamento de uma RNA que resultou

numa superfície de separação aberta em torno dos padrões correspondentes a classe Nuvem1. Neste caso, esta RNA não pode diferenciar corretamente um padrão pertencente a classe Nuvem1 de um pertencente a Nuvem0. Com a adição aleatória de padrões pertencentes a classe Nuvem0 e um novo treinamento, pode-se perceber na Figura 5 que estes novos padrões permitiram um ajuste dos pesos a determinar planos de separação que resultam em uma superfície fechada em torno dos padrões Nuvem1. Assim, um padrão que esteja fora desta região é classificado como pertencente a classe Nuvem0. Bem como padrões que nunca foram apresentados no treinamento podem ser rejeitados. A partir desta observação

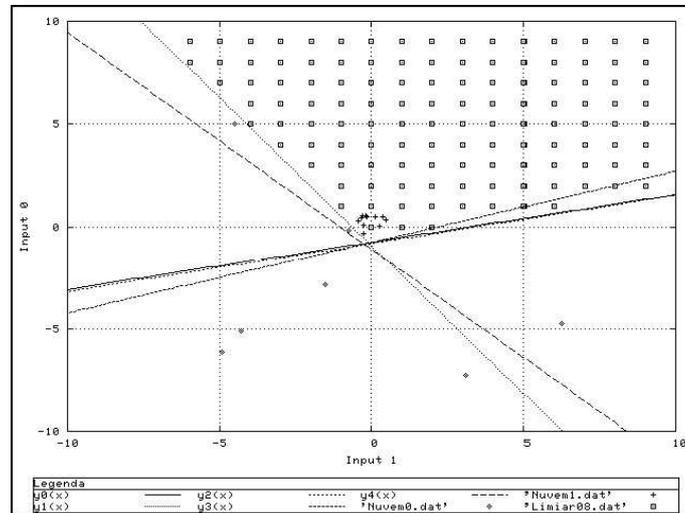


Figura 4. Exemplo de superfície aberta de separação.

de que contra-exemplos contribuem para o aprendizado, Bender e Osório (7) propõem dois métodos, Regiões de Imagens Falso Positivo (RIFP) e Nuvem de Não-Classe (NNC), para melhorar a obtenção de superfícies fechadas de separação para classificação e recuperação de imagens em cores. No contexto de classificação e recuperação de imagens, isto significa reduzir a ocorrência de imagens falso-positivo (imagens classificadas erroneamente).

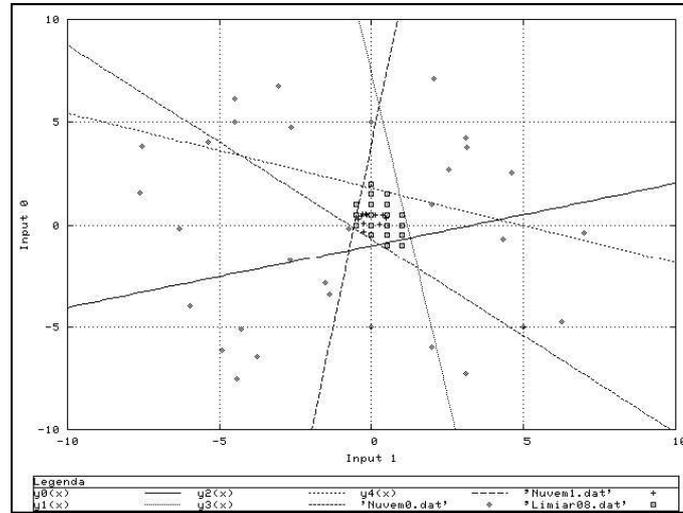


Figura 5. Exemplo de superfície fechada de separação.

3.2 Métodos RIFP e NNC

O método RIFP consiste de um treinamento inicial da RNA com exemplos de classe (o que se quer que a RNA aprenda a classificar) e de não-classe (por exemplo, imagens de fundo) e etapas de refinamento do treinamento pela inclusão de exemplos falso-positivos rotulados como não-classe. Durante uma rodada de classificação com a BDI, capturam-se algumas regiões falso-positivo de imagens e inclui-se na base de treinamento com o rótulo de não-classe (i.e. como contra-exemplo). Executa-se então um novo treinamento da RNA, repetindo-se o processo até se obter um resultado satisfatório.

Já o método NNC visa reduzir a necessidade de iteração gerando automaticamente exemplos de não-classe. Este método produz uma "nuvem" esparsa de exemplos não-classe que são adicionados aos exemplos de treinamento. Uma descrição mais detalhada sobre os métodos e superfícies de separação é apresentada em (6).

4 Experimentos

Para a experimentação destes dois métodos propostos foram preparadas três BDIs. A primeira BDI consiste de padrões de imagens geradas em programa de edição gráfica com cinco cores e inseridos em imagens diversas. O objetivo desta BDI é verificar aplicação dos métodos em um ambiente controlado. A segunda BDI utiliza imagens fotográficas de objetos obtidas em condições controladas (fundo preto, iluminação uniforme). Estas imagens estão disponíveis na Internet e foram utilizadas como benchmark. A terceira base constitui-se de imagens fotográficas obtidas sem nenhum controle (diversos tipo de iluminação, escala e ambientes diversos) e constitui-se de uma base para verificar a performance dos métodos em situação real.

Três métodos para extração de características a partir da cor foram utilizados e dois sistemas de cores (RGB e HSV). O primeiro método consiste em se aplicar uma redução para n cores com o algoritmo *Median Cut* e obter o histograma da paleta de cores. Este método resulta em um vetor de características de tamanho n . O segundo método utiliza uma clusterização do tipo *k-Means* para encontrar n cores-protótipo (centróides) que representam aquela imagem, o que resulta num vetor de característica de tamanho $n*3$. O terceiro método consiste em aplicar uma redução para n cores com o algoritmo *Median Cut* obtendo o histograma da paleta de cores acrescido das triplas das componentes da cor. Este método resulta num vetor de características de tamanho $n*4$.

A etapa de extração de características utiliza "janelas de varredura", usualmente um tamanho de 64x64 pixels que é deslocada por toda a imagem em incrementos normalmente de 32x32 pixels. Para cada uma destas janelas (que corresponde a uma janela de dimensões iguais utilizada para as imagens de treinamento da RNA) obtém-se o vetor de características ao qual se aplica um dos métodos de extração descritos. Este vetor será a entrada da RNA durante a classificação.

A Figura 6 apresenta o resultado da classificação para lata de Coca-Cola com o treinamento inicial da RNA. Note-se a grande quantidade de regiões falso-positivo (FP) presente na imagem.

Após duas iterações do método RIFP (Figura 7) e do método NNC (Figura 8) é possível se perceber uma redução de regiões FP considerável.



Figura 6. Exemplo de classificação com treino inicial da RNA para latas de Coca-Cola.

5 Conclusões e Perspectivas

A tarefa de recuperação e classificação de imagens executada de maneira automática possui grande importância atualmente devido a grande quantidade de imagens armazenadas, muitas vezes de forma desordenada, em meios eletrônicos. Portanto sistemas que são capazes de recuperar uma imagem a partir de um exemplo possuem grande importância para permitir um rápido acesso ao conteúdo visual procurado. Dentro da grande área da visão computacional estes métodos formam a base para tarefas mais desenvolvidas no campo de visão de máquina. Seu estudo, portanto consiste em tema atual de pesquisa e possuem aplicações práticas imediatas em áreas tão diversas quanto localização de ob-



Figura 7. Exemplo de classificação após duas iterações de RIFP para latas de Coca-Cola.



Figura 8. Exemplo de classificação após duas iterações de NNC para latas de Coca-Cola.

jetos numa cena, levantamento de ocorrência de logomarcas na mídia, recuperação de imagens com conteúdo visual semelhante para posterior organização até aplicações em robótica.

Buscando uma abordagem diferenciada para a tarefa de recuperação e classificação de imagens propomos o uso de aprendizado de máquina para não mais executar uma comparação entre uma imagem de exemplo contra uma imagem da base de imagens, mas sim obter um "modelo" que identifique o objeto procurado a partir de uma coleção destes objetos utilizando a cor como principal característica. Para tanto se optou por utilizar RNAs do tipo MLP com arquitetura Cascade Correlation para a tarefa de classificação.

Como perspectivas para este trabalho estão a utilização de outro tipo de RNA, notadamente redes do tipo RBF, com o mesmo pré-processamento para verificar o seu resultado na ocorrência de regiões de FP e imagens FP recuperadas e a aplicação de técnicas de pré-processamento mais elaboradas.

Referências

- [1] ANTANI, Sameer; KASTURI, Rangachar; RAMESH, Jain. *A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video*, Pattern Recognition, Amsterdã, Países Baixos, Elsevier Science B.V, **35**(2002)945–965.
- [2] IKONA. Irnia / imedia project research team. <http://www-rocq.inria.fr/imedia/ikona>, Out 2002.
- [3] RUI, Yong; HUANG, Thomas S.; CHANG, Shih-Fu. *Image retrieval: Past, present, and future*, International Symposium on Multimedia Information Processing, Taipei, Dez 1997.
- [4] SEBE, Nicu; LEW, Michael S. *Color-based retrieval*, Pattern Recognition Letters, Amsterdã, Países Baixos, Elsevier Science B.V, **22**(2001)223–230.
- [5] SMITH, John R.; CHANG, Shih-Fu. Tools and techniques

- for color image retrieval. *IS&T/SPIE In Symposium on Electronic Imaging: Science and Technology - Storage & Retrieval for Image and Video Databases IV, San Jose, CA, 2670, Fev 1996.*
- [6] BENDER, Túlio C. Classificação e Recuperação de Imagens por Cor Utilizando Técnicas de Inteligência Artificial. *Universidade do Vale do Rio dos Sinos - Programa Interdisciplinar de Pós-Graduação em Computação Aplicada - Dissertação de Mestrado, Jun 2003.*
- [7] BENDER, Túlio C.; OSÓRIO, Fernando S. Reconhecimento e Recuperação de Imagens Utilizando Redes Neurais Artificiais do Tipo MLP. *IV Encontro Nacional de Inteligência Artificial, Campinas, SP, Ago 2003.*