

RECONHECIMENTO DE IMAGENS BIDIMENSIONAIS UTILIZANDO REDES NEURAIS ARTIFICIAIS

Guy Perelmuter, Enrique Vinicio Carrera E., Marley Vellasco e Marco Aurélio Pacheco

ICA: Núcleo de Pesquisa em Inteligência Computacional Aplicada
Departamento de Engenharia Elétrica
Pontifícia Universidade Católica do Rio de Janeiro
Cx. Postal 38063 - CEP 22453-900
E-mail: ica@ele.puc-rio.br

Abstract. Artificial vision or computer vision may be defined as the field that studies the methods used by an artificial system to “perceive” the information contained in a certain image. Neural networks, due to their ability to extract information from complex sets of data, have been largely applied in computer vision for pattern recognition and pattern classification. This article presents an “intelligent classifier” based on neural networks. The complete image recognition/classification system (pre-processor + classifier) is invariant to translation, rotation and sizing of the analysed object. The system’s structure is quite flexible, allowing it to be applied to virtually any set of bidimensional images.

1. INTRODUÇÃO

Considerável esforço tem sido dedicado à solução do problema de reconhecimento e caracterização de objetos presentes em uma imagem. Até a década de 70, predominou o uso de técnicas óticas de processamento [Casasent-Psalti (1976)]. A partir do início dos anos 80, com os avanços verificados na microeletrônica e no desenvolvimento de arquiteturas paralelas de processamento, as técnicas digitais passaram a ser mais empregadas. Atualmente, o amadurecimento das técnicas computacionais inteligentes, como Sistemas Especialistas, Lógica Nebulosa, Redes Neurais e Algoritmos Genéticos, tem permitido novas abordagens para esse problema.

As técnicas inteligentes fazem uso metafórico dos conceitos, princípios e mecanismos fundamentais dos sistemas naturais, procurando capturar, tanto na teoria como na prática, os algoritmos encontrados na Natureza. A complexidade do problema de reconhecimento e de classificação de imagens, que dificilmente pode ser abordado em termos algorítmicos, tem tornado o uso dessas técnicas cada vez mais freqüente, especialmente as Redes Neurais Artificiais.

As *Redes Neurais Artificiais (Artificial Neural Networks)* são dispositivos não-lineares, inspirados na funcionalidade dos neurônios biológicos, aplicados no reconhecimento de padrões, na otimização e na previsão de sistemas complexos [Treleaven-Vellasco-Pacheco (1989)]. A habilidade

em formar mapeamentos não-lineares torna as redes neurais prósperas nessas aplicações.

Redes neurais são compostas por diversas unidades computacionais paralelas, interconectadas parcial ou totalmente. Cada uma dessas unidades (“neurônios artificiais”) efetua um certo número de operações simples e transmite seus resultados às unidades vizinhas com as quais possui conexão. Através de um processo de treinamento, as redes neurais passam a ser capazes de reconhecer padrões, mesmo que os dados utilizados nesse treinamento sejam não-lineares, incompletos ou até mesmo contraditórios. A habilidade de manipular estes dados imprecisos faz com que as redes neurais sejam extremamente eficazes em tarefas onde especialistas não estão à disposição ou um conjunto de regras não pode ser facilmente formulado. No caso da visão por computador, tanto a dificuldade de estabelecer algoritmos de reconhecimento quanto a natureza dos dados com os quais se trabalha indicam as redes neurais como uma solução promissora.

A implementação de um sistema de visão artificial baseado em redes neurais faz uso de suas características intrínsecas, isto é, da sua capacidade de extrair padrões de conjuntos de dados complexos. Neste trabalho foram efetuados testes com 2 conjuntos distintos de objetos, obtendo-se resultados bastante satisfatórios, principalmente quando comparados com aqueles obtidos através de técnicas convencionais.

O presente trabalho baseia-se nas idéias propostas por Emílio C. Acocella na sua pesquisa de mestrado, "Extração de Invariâncias em Processamento de Imagem Aplicado a Visão Computacional" [Acocella (1993)], realizada no Instituto Militar de Engenharia do Rio de Janeiro. Acocella desenvolveu um classificador de imagens baseado em modelos probabilísticos, cuja principal limitação é o longo e exaustivo processo de análise que os elementos a serem classificados devem sofrer. Tal dificuldade limita bastante a flexibilidade do sistema. Para contornar estas limitações, o Núcleo de Pesquisa em Inteligência Computacional Aplicada (ICA) desenvolveu um classificador inteligente baseado em redes neurais, que, além de eliminar este exaustivo processo de extração de características, facilita o processo de adaptação (quando novas peças forem incluídas no conjunto a ser reconhecido), mantendo as vantagens de ser invariante à translação, à rotação e ao escalonamento das imagens a serem classificadas. O modelo de rede neural utilizado nesta aplicação é o Back-Propagation [Wasserman (1989)], [Rumelhart - McClelland (1986)], implementado na linguagem de programação C em estações de trabalho SUN.

Neste artigo são apresentados os conceitos e as características da visão artificial, bem como algumas das propostas anteriores feitas na área (seção 2). Na seção 3 é descrita a solução proposta e na seção 4 são mostrados os resultados obtidos, indicando o desempenho alcançado pela rede neural. Finalmente, na seção 5, são apresentadas as conclusões e os planos para futuras extensões do classificador inteligente.

2. VISÃO ARTIFICIAL

A visão artificial é efetuada através de um conjunto de transformações, algumas delas em paralelo, que permitem a extração dos aspectos invariantes das imagens, a despeito da variabilidade do ambiente no qual elas estejam situadas. Tais invariâncias possibilitam o reconhecimento ou a caracterização da imagem, permitindo a interação do sistema de visão com as mesmas. Pode-se dizer então que a visão artificial procura "perceber" a informação contida numa imagem com o propósito de classificação, caracterização e/ou reconstrução da mesma.

Diversas denominações têm sido dadas a este campo multidisciplinar que aproveita os conhecimentos relacionados com Processamento de Sinais, Inteligência Computacional, Neurofisiologia e outros, sendo que os mais utilizados são: visão computacional, visão cibernética, visão de máquina,

visão robótica e visão por computador. De uma forma geral, todos estes nomes indicam o processo de cognição de uma imagem, que pertence à área do Reconhecimento e Análise de Padrões.

São inúmeras as aplicações da visão artificial. Em qualquer campo onde se faça necessário o reconhecimento de padrões visuais, ela pode ser utilizada: Robótica, Medicina, Cartografia, Meteorologia, Astronomia, etc.

Diversos projetos já foram desenvolvidos procurando construir sistemas visuais artificiais funcionais, sendo possível perceber que a maioria dos sistemas implementados (inclusive [Acocella (1993)]) segue a mesma estrutura básica: *Pré-processamento, Extração de Características e Classificação* (Figura 1). A seguir, é apresentada a funcionalidade de cada uma dessas etapas.

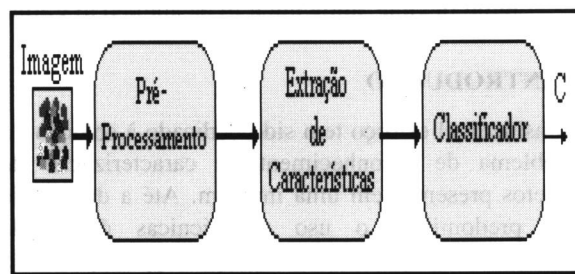


Figura 1. Estrutura Básica de um Sistema de Visão Artificial

2.1 Pré-Processamento

A etapa de pré-processamento executa uma série de transformações na imagem de modo a extrair suas características intrínsecas. Também existente nos sistemas visuais biológicos [Leow (1994)], o pré-processamento parte da matriz de *pixels* (chamado de "imagem bruta"), obtida através de métodos que transformam os sinais luminosos em sinais elétricos. Inicialmente, quando uma imagem é capturada, há sempre informações pertencentes à cena que não interessam ao sistema, como, por exemplo, o fundo contra o qual a imagem se encontra. É preciso retirar da imagem os elementos desnecessários, aplicando-se um algoritmo de detecção de bordas. Retira-se, também, a influência que a fonte de luz tem sobre a imagem, num processamento conhecido como homomórfico [Lim (1990)]. Posteriormente, o sistema de coordenadas também pode ser alterado. Representando-se a imagem através de coordenadas polares (obtendo-se os coeficientes harmônicos circulares e calculando-se a Transformada de Mellin [Wu-Stark (1986)]), o sistema torna-se invariante a operações como rotação, translação e escalonamento.

Depois de aplicar esse conjunto de transformações locais, sustentadas pela Teoria de Processamento de Sinais, é obtida a “imagem intrínseca”. Esta representação da imagem original depende fortemente do tipo de operações que foram efetuadas sobre a imagem. A imagem, que era inicialmente representada por uma matriz de *pixels*, passa a ser representada por uma matriz de coeficientes numéricos complexos.

2.2 Extração de Características

Uma vez que as imagens intrínsecas de todas as classes de imagens que o sistema de visão artificial se propõe a classificar são obtidas, é utilizada uma ferramenta auxiliar do Classificador, conhecida como Extrator de Características.

O Extrator de Características compara, de forma seqüencial, as representações das imagens intrínsecas e, através de métodos estatísticos, determina quais os coeficientes dessas imagens que apresentam maiores diferenças entre si. Estes serão os coeficientes que o Classificador irá utilizar para tentar classificar uma imagem qualquer. Este módulo é fundamental para os sistemas onde o Classificador é desenvolvido com técnicas convencionais de análise. Isso porque o conjunto de dados resultantes do pré-processamento da imagem é, em geral, muito grande para ser tratado de forma eficiente pelos métodos tradicionais. Por exemplo, no trabalho de Acocella, apenas 40 coeficientes, de um conjunto de cerca de 65.000, são utilizados na classificação.

2.3 Classificador

Os classificadores de imagens mais comuns estão baseados em técnicas de decisão Bayesiana [Toussaint (1974)]. A partir dos coeficientes selecionados pelo extrator de características, cada representação da imagem intrínseca é analisada e comparada com os padrões que o sistema conhece, de forma que alguma decisão a respeito da classe da imagem possa ser tomada.

No trabalho de Acocella, cinco peças mecânicas (Figura 2) foram subdivididas em duas sub-classes de duas e três peças cada uma. O autor fez uma análise exaustiva de centenas de imagens dessas cinco peças, extraíndo os quarenta coeficientes mais significativos para o classificador. Durante os testes foram apresentadas diversas imagens dessas peças, juntamente com a informação de à qual sub-classe a peça pertencia, obtendo-se uma classificação correta para todas as imagens.

Conforme será demonstrado na seção 3, com o uso de redes neurais não existe mais a necessidade de separar as peças em duas sub-classes distintas, nem de fazer uma inspeção nos coeficientes de cada imagem para selecionar os mais relevantes. As características intrínsecas das redes neurais artificiais, capazes de extrair informações a partir de um grupo complexo de dados, facilitam a execução desse processamento.

3. O CLASSIFICADOR INTELIGENTE

O sistema de visão artificial implementado neste trabalho tem por objetivo avaliar o desempenho das redes neurais artificiais quando comparadas aos modelos tradicionais de classificação. A meta é fornecer um sistema inteligente que supere os principais problemas do classificador convencional descrito na seção anterior: o exaustivo processamento de dados, associado à extração das características das imagens, e a pouca flexibilidade do sistema, que precisa ser modificado para cada novo conjunto de objetos a serem classificados.

O novo sistema possui apenas duas etapas: o Pré-Processamento e a Classificação. A etapa de Pré-Processamento, baseada no sistema desenvolvido em [Acocella (1993)], é mostrada esquematicamente na Figura 3. Ela parte de uma imagem bruta em formato *raster* de 256x256 *pixels* com 256 níveis de cinza. Nesta imagem é aplicado, inicialmente, um filtro de mediana [Lim (1990)], que procura atenuar o ruído que pode existir na imagem capturada. Posteriormente, é feita a detecção das bordas da imagem através do Algoritmo de Sobel [Lim (1990)], com a finalidade de se obter a posição do objeto no campo de visão do sistema, reduzindo as dimensões da imagem a ser processada.

Paralelamente, aplica-se à imagem bruta um processamento homomórfico por multiplicação [Lim (1990)], que permite a eliminação das distorções causadas pelos diferentes ângulos de incidência dos raios luminosos sobre a imagem. Tal efeito ocorre devido às diferenças entre as fontes de iluminação que são utilizadas na captura da imagem. Após esta etapa, e levando-se em conta os resultados obtidos no processo de detecção de bordas, é feita a transformação da imagem para coordenadas polares, utilizando-se uma discretização com 128 raios e 512 ângulos. Com isto, são obtidos os coeficientes harmônicos circulares da imagem [Wu, Ronald- Stark (1986)], que representam a imagem de forma invariante com relação à rotação.

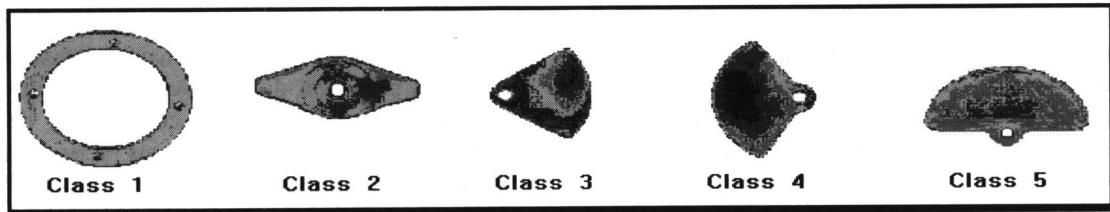


Figura 2 - Peças Mecânicas usadas para os testes com o Classificador

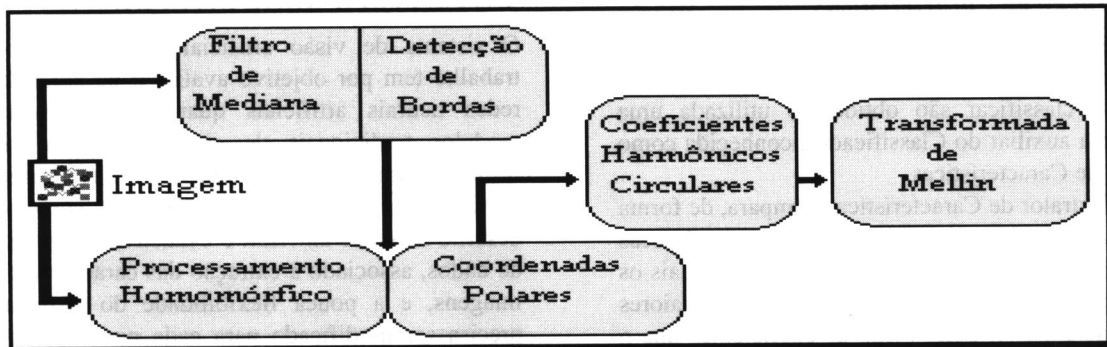


Figura 3. Pré-processamento da Imagem Bruta

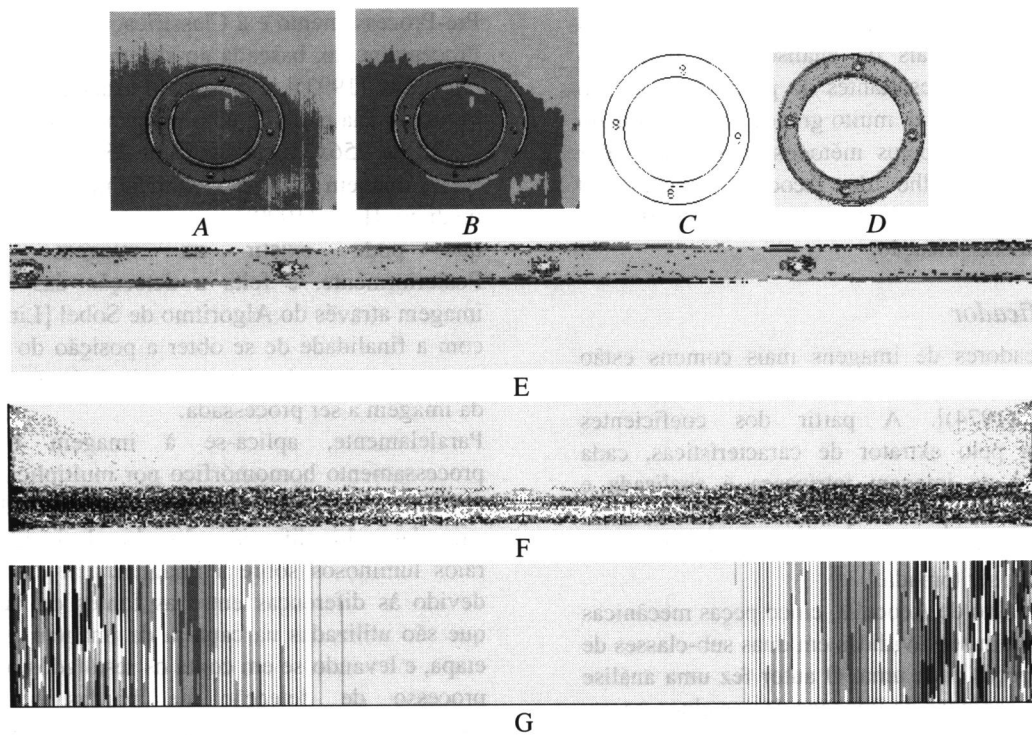


Figura 4 - Etapas do Pré-Processamento: A-Imagem Original, B-Filtro de Mediana, C-Detecção de Bordas, D-Processamento Homomórfico, E-Cordenadas Polares, F-Coefficientes Harmônicos Circulares e G-Transformada de Mellin.

Para terminar o pré-processamento, é gerada uma representação interna da imagem através da Transformada de Mellin [Wu-Stark (1986)], que é aplicada sobre o conjunto dos

coeficientes harmônicos circulares. O resultado dessa operação é a imagem intrínseca, que se traduz em uma matriz de 64 K coeficientes complexos e que representa a imagem bruta

inicial com invariância com relação ao escalonamento e à rotação. A Figura 4 mostra a imagem obtida após cada etapa do pré-processamento.

Uma vez que a representação da imagem intrínseca de todos os padrões é obtida, os métodos tradicionais geralmente empregam uma seleção dos valores mais significativos da matriz de coeficientes, de acordo com o conjunto de objetos a serem classificados. A partir dessa escolha é construído um classificador específico para tais imagens. Neste sistema convencional, qualquer imagem capturada é processada e seus coeficientes mais relevantes, obtidos pelo extrator de características, são comparados com os coeficientes dos modelos do classificador.

A mudança proposta neste trabalho é a de substituir o extrator de características, eliminando-se o exaustivo processo de pré-seleção dos valores mais significativos. Graças às suas características intrínsecas, a rede neural atuará como extratora de características (além de classificadora), selecionando os coeficientes mais significativos, independentemente dos conjuntos de imagens que forem apresentados. Isso permite a construção de um sistema classificador de propósito geral com um campo de aplicação muito mais abrangente.

As entradas da rede serão, portanto, os coeficientes obtidos após a Transformada de Mellin. A primeira operação que é realizada sobre os valores complexos resultantes dessa transformada é a obtenção de seus módulos. De modo a reduzir o número de coeficientes a um nível que pudesse ser tratado pelo equipamento disponível, foi realizada uma análise bastante simples da matriz.

Foi observado que, para o conjunto de peças mecânicas usadas nos testes do classificador, a magnitude relativa dos coeficientes correspondentes a um raio distante mais de 64 unidades do centro da imagem é muito pequena. Portanto, descartamos a metade superior da matriz. Além disso, com base na simetria parcial apresentada pela matriz (Figura 4G), descartamos o seu lado direito.

A partir desta análise superficial, pode-se reduzir a matriz de entrada original para apenas 16 K números reais. Estes 16384 valores serão as entradas de uma rede neural que utiliza o modelo Back-Propagation [Wasserman (1989)], [Rumelhart-McClelland (1986)], com uma camada escondida de 100

neurônios e com uma camada de saída com cinco elementos de processamento (um para cada classe de objetos a ser classificada). Esta rede funciona como um classificador neural que decide a classe à qual pertence uma imagem apresentada (Figura 5). O sistema completo, incluindo uma pequena interface gráfica utilizada para treinamento e testes, encontra-se rodando em estações de trabalho SUN.

Uma das extensões a serem implementadas no sistema é o desenvolvimento de um extrator inteligente de características (uma função de avaliação de um algoritmo genético, por exemplo). Com essa seleção mais criteriosa dos 16 K coeficientes, a rede neural terá melhores condições para realizar a classificação, uma vez que os dados apresentados irão conter as informações mais relevantes sobre a imagem.

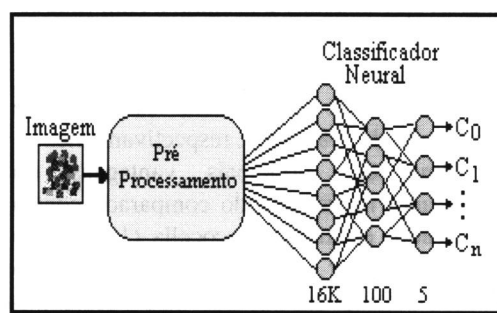


Figura 5. Sistema Neural para Visão Artificial

4. RESULTADOS

Conforme mencionado anteriormente, o classificador desenvolvido por Acocella foi capaz de classificar diferentes imagens das cinco peças mecânicas exibidas na Figura 2, baseado nos 40 coeficientes mais significativos obtidos pelo extrator de características. Para se ter uma idéia de como um classificador neural iria se comportar, utilizou-se, como primeiro teste, uma rede neural treinada com os mesmos 40 coeficientes obtidos pela etapa de Extração das Características. A rede utiliza o algoritmo Back-Propagation com uma camada escondida de 15 processadores e cinco elementos na saída. O treinamento foi feito utilizando-se três imagens-padrão de cada uma das cinco peças mecânicas.

Classe do Objeto	Nível de Ruído(%)	Maior Saída para Precisão de 1%	Maior Saída para Precisão de 10%
1	25	0.97	0.85
2	20	0.99	0.95
3	20	0.99	0.92
4	25	0.95	0.83
5	15	0.99	0.93

Tabela 1: Imagens Classificadas utilizando-se os 40 coeficientes do Extrator

Tipo de Imagem	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5
Rotacionada	0.97	0.91	0.94	0.86	0.75
Escalonada & Rotacionada	0.95	0.74	0.71	0.52	0.80
Escalonada & Transladada	0.96	0.92	0.83	0.73	0.97

Tabela 2: Imagens Classificadas utilizando 16K coeficientes

A Tabela 1 apresenta os resultados obtidos pelo classificador inteligente para imagens apresentadas sem ruído e com ruído até 25%. Os resultados nas colunas à direita correspondem aos processadores com maior valor na saída para uma precisão de treinamento de 1% e 10%, respectivamente.

Uma das principais vantagens do classificador neural quando comparado com o classificador estatístico [Acocella (1993)] é o fato de não ser necessária uma separação preliminar das peças em duas-sub-classes. As peças foram apresentadas para a rede sem nenhum tipo de pré-classificação.

No segundo teste com o classificador inteligente o processo de extração de características foi totalmente eliminado. Uma rede com 16 K processadores de entrada, uma camada escondida com 100 elementos e uma camada de saída com cinco elementos (Figura 5) foi treinada com três amostras de cada classe através do algoritmo Back-Propagation. Após 191 iterações, obteve-se a tolerância

desejada de 10%. Durante os testes, três novas imagens de cada classe foram apresentadas à rede neural. A Tabela 2 apresenta a saída do processador correspondente à classe em cada situação. Novamente, todas as peças foram classificadas corretamente. A coluna "Tipo de Imagem" descreve a imagem com relação à Figura 2.

O classificador inteligente também foi testado com um conjunto diferente de objetos, conforme pode ser visto na Figura 6. Quatro exemplos de cada classe foram apresentados à rede neural durante o treinamento. Após 127 iterações, a tolerância desejada de 10% foi obtida. Três novas imagens de cada classe foram apresentadas à rede neural, e todas elas foram classificadas corretamente. A Tabela 3 mostra o valor do estado de ativação do processador de saída correspondente a cada situação. Como no exemplo anterior, a coluna "Tipo de Imagem" especifica a posição da imagem em relação à Figura 6.

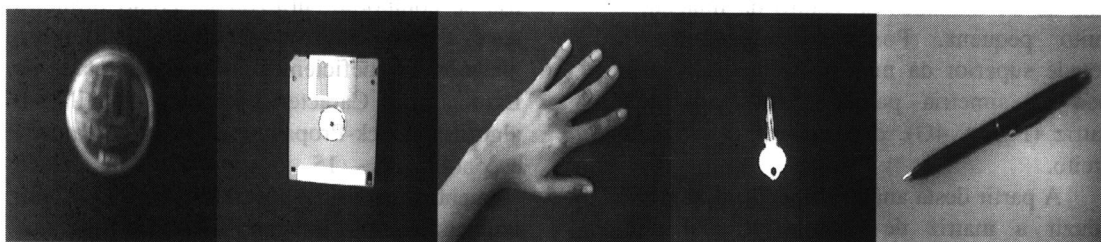


Figura 6 - Segundo conjunto de padrões usado nos testes do classificador inteligente

Tipo de Imagem	Moedas	Disquetes	Mãos	Chaves	Canetas
Rotacionada	0.94	0.84	0.94	0.96	0.99
Rotacionada & Transladada 1	0.66	0.72	0.76	0.65	0.98
Rotacionada & Transladada 2	0.78	0.79	0.75	0.72	0.95

Tabela 3 - Imagens Classificadas com 16K coeficientes

5. CONCLUSÕES

Através do desenvolvimento do classificador inteligente, as bases para um sistema geral de reconhecimento de imagens foram identificadas. Nesse classificador, o pré-processamento gera os coeficientes que são utilizados no processo de treinamento da rede neural. Tais coeficientes permitem a diferenciação entre diversas imagens e fornecem invariância com relação à rotação, escalonamento e translação. O classificador inteligente possui três vantagens principais em relação aos sistemas tradicionais: (1) nenhuma pré-classificação do objeto a ser identificado é necessária, (2) não há necessidade de um módulo extrator de características e (3) o classificador inteligente é um sistema de reconhecimento de propósito geral, capaz de ser aplicado a virtualmente qualquer conjunto de objetos bidimensionais.

Foi observado, entretanto, que as condições de captura da imagem são extremamente críticas para o desempenho do classificador no que diz respeito à detecção de bordas. Deste modo, um sistema melhor de captura, com iluminação adequada, deve ser implementado.

O desempenho do classificador pode também ser melhorado se o número de coeficientes apresentados à rede neural puder ser elevado ou se um método for desenvolvido para otimizar a escolha destes. Através de uma função de avaliação apropriada, poderemos detectar os coeficientes mais relevantes através de um algoritmo genético, apresentando-os à rede neural.

Um outro passo importante é o desenvolvimento de métodos para paralelizar o processamento, de forma que os tempos de treinamento e de reconhecimento sejam reduzidos e que este último possa ser efetuado em tempo real. Tal paralelização é possível, visto que diversas operações efetuadas no pré-processamento são completamente independentes entre si.

AGRADECIMENTOS

Os autores deste trabalho gostariam de agradecer ao Sr. Emílio Acocella pelo fornecimento do material necessário para o desenvolvimento do classificador inteligente.

REFERÊNCIAS

- David Casasent e Demetri Psalti**, "Applied Optics", (1976), 15, 7, 1795-1799.
- Yan-Neng Hsu et Alii**, "Applied Optics", (1982), 21, 22, 4012-4015.
- P. Treleaven, M. Vellasco, M. Pacheco**, "VLSI Architectures for Neural Networks", IEEE Micro, vol.9, number 6, pp.8-27, December 1989.
- E. C. Acocella**, "Extração de Invariâncias em Processamento de Imagem Aplicado a Visão Computacional", Tese de Mestrado IME, Rio de Janeiro, 1993.
- Philip D. Wasserman**, "Neural Computing: Theory and Practice", VNR, New York, 1989.
- D. Rumelhart e J.L. McClelland**, "Parallel Distributed Processing: Explorations in the Microstructure of Cognition", Cambridge, Massachusetts, MIT Press, 1986.
- W.K. Leow**, "Visor: Learning Visual Schemas in Neural Networks for Object Recognition and Scene Analysis", Tese de Doutorado - Universidade do Texas, Austin, 1994.
- Jae S. Lim**, "Two-Dimensional Signal and Image Processing", Prentice-Hall, Englewood Cliffs, NJ, 1990.
- Rong Hwang Wu e Henry Stark**, "Proceedings of the Eighth International Conference on Pattern Recognition", pp. 92-94, 1986.
- Godfried T. Toussaint**, "IEEE Transactions Information Theory", IT-20, pp.472-479, 1974.
- Ronald Wu e Henry Stark**, "Pattern Recognition in Practice II", pp. 401-410, 1986.

