

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE ELETRICIDADE

CARLOS MAGNO SOUSA JUNIOR

COMPRESSÃO DE IMAGENS UTILIZANDO ANÁLISE DE
COMPONENTES INDEPENDENTES

São Luís - MA

2007

CARLOS MAGNO SOUSA JUNIOR

COMPRESSÃO DE IMAGENS UTILIZANDO ANÁLISE DE
COMPONENTES INDEPENDENTES

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Eletricidade da UFMA, como requisito para a obtenção parcial do grau de MESTRE em Engenharia de Eletricidade.

Orientador: Prof. Allan Kardec Duailibe Barros Filho

(Orientador)

São Luís - MA

2007

Sousa Junior, Carlos Magno

Compressão de imagens utilizando análise de componentes independentes / Carlos Magno Sousa Junior. - 2007.

54f

Impresso por computador (fotocópia).

Orientador: Allan Kardec Duailibe Barros Filho.

Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia de Eletricidade, Universidade Federal do Maranhão, São Luís, 2007.

1. Compressão de dados. 2. Componentes independentes - Análise.
3. Compressão de imagens. I. Orientador: Barros Filho, Allan Kardec Duailibe. II. Título.

CDU : 004.627

CARLOS MAGNO SOUSA JUNIOR

COMPRESSÃO DE IMAGENS UTILIZANDO ANÁLISE DE COMPONENTES INDEPENDENTES

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Eletricidade da UFMA, como requisito para a obtenção parcial do grau de MESTRE em Engenharia de Eletricidade.

Aprovada em

BANCA EXAMINADORA

Prof. Allan Kardec Duailibe Barros Filho

(Orientador)

Prof. Dráulio Barros de Araújo

(Membro da Banca Examinadora)

Prof. João Viana da Fonseca Neto

(Membro da Banca Examinadora)

Ao meu pai, que à sua maneira sempre torceu pelo sucesso de cada um de seus filhos.

À minha mãe, que sempre torceu pelas vitórias de seus filhos, para que suas escolhas os fizessem felizes.

A meus irmãos, pelo companheirismo em todos os momentos.

Resumo

A redundância é um assunto antigo em pesquisa sobre compressão de dados. Os métodos de compressão de dados que usam estatísticas foram recentemente influenciados pelas pesquisas em neurociência. Neste trabalho, propomos um sistema de compressão de imagem baseado no conceito de codificação eficiente derivado dos modelos de processamento da informação neural. O desempenho do sistema é comparado aos resultados da transformada discreta cosseno (DCT) e análise de componentes principais (PCA) com a mesma taxa de compressão (CR). A avaliação através das medidas objetiva e visual mostrou que o sistema proposto apresentou menos distorções, tais como artefatos de blocos do que a DCT e PCA.

PALAVRAS-CHAVE: Análise de componentes independentes, codificação eficiente, análise de componentes principais.

Abstract

Redundancy is an old issue in data compression research. Compression methods that use statistics have been heavily influenced by neuroscience research. In this work, we propose an image compression system based on the *efficient coding* concept derived from neural information processing models. The system performance is compared with discrete cosine transform (DCT) and principal components analysis (PCA) results at several compression ratios (CR). Evaluation through both objective measurements and visual inspection showed that the proposed system is more robust to distortions such as blocking artifacts than DCT and PCA.

KEYWORDS: Independent component analysis, efficient coding, principal components analysis.

Agradecimentos

Agradeço a Deus por ter me permitido chegar ao fim desta etapa da minha vida.

Aos meus pais, Carlos Magno e Elmorane de Maria, pelo apoio incondicional e por todos os esforços que efetuaram para a realização dos meus objetivos.

Aos meus irmãos Mario Elias e Mariane Pires, por terem sempre e invariavelmente me incentivado a perseverar nos estudos.

Ao Prof. Allan Kardec Barros, por seu empenho em construir um bom laboratório num cenário de recursos escassos e objeções políticas, além da excelente orientação científica.

Expresso minha gratidão aos amigos do Laboratório de Processamento da Informação Biológica - PIB André Cavalcante, Aryfrance Rocha, Denner Guilhon, Daniel Duarte, Deusdete de Sousa, Ewaldo Santana, Enio, Jaciara, Marcio Santos, Ricardo Robson, Raniere Machado.

Aos amigos do Núcleo de Energias Alternativas - NEA Bartolomeu Ferreira, Cledson de Jesus, Shigeaki Leite, Walbermark Marques.

Aos amigos Fausto Lucena, Ivan Júnior, Jaciani Pereira, Lúcio Campos, Ruttyellem, Silvangela entre outros.

*“Embora ninguém possa voltar atrás e
fazer um novo começo, qualquer um pode
começar agora e fazer um novo fim”*

Chico Xavier

Contents

List of Figures		viii
List of Tables		x
Lista de Abreviações		xi
1 Introdução		1
1.1 Organização do Trabalho		2
2 Codificação Eficiente		4
3 Análise de Componentes Independentes		6
3.1 História		6
3.2 Definições		7
3.3 Estimação de Componentes Independentes		10
3.3.1 Estimação Através de Maximização de Não-Gaussianidade		10
3.3.2 Negentropia como Medida de Não-Gaussianidade		11
3.4 Análise de Componentes Principais		12
3.5 Comparação entre PCA e ICA		14
3.6 Representação ou Codificação de Imagens por ICA		14
3.6.1 Comparação entre DCT, PCA e ICA		16
4 O Método Proposto		18
4.1 Codificador Eficiente		19
4.1.1 Estimação das Funções Base Através de ICA		19

	vii
4.1.2 Estimação dos Coeficientes da Projeção	21
4.1.3 Seleção das Funções Base	23
5 Resultados e Discussão	24
5.1 Base de Dados de Imagens Utilizada	24
5.2 Condições de Teste	24
5.3 Medida de Distorção	25
5.4 Resultados	26
5.5 Discussão	35
6 Conclusões e Comentários Finais	37
Referências	38

List of Figures

2.1	(a) Exemplo de como o olho humano recebe a informação visual através do globo ocular, a transmite ao nervo óptico e a repassa para a região posterior do cérebro, denominada de córtex visual primário (CVP). No córtex visual primário, as células têm campos receptivos que tentamos mimicar; (b) exemplo de uma imagem e como ela seria “dividida” ou codificada nos campos receptivos de V1, no CVP.	4
3.1	Amostra de uma imagem como combinação linear de suas características. . . .	8
3.2	Amostra de uma imagem disposta como combinação linear das funções base do sinal. (a) Amostras da imagem. (b) Componentes s_n . (c) Funções base.	8
4.1	Diagrama de blocos do codificador de imagem. Primeiramente, a imagem original é transformada e logo após quantizada	18
4.2	Sistema proposto de compressão de imagem. O sistema consiste em duas fases: (A) aprendizagem e (B) a fase de projeção, sendo $[x_1, x_2, \dots, x_n]$ as amostras de treinamento $[\phi_1, \phi_2, \dots, \phi_n]$ as funções base, $[w_1, w_2, \dots, w_n]$ os coeficientes de projeção, \hat{y} a imagem reconstruída, y a imagem original e ε o erro médio quadrático.	19
4.3	Combinador linear de múltiplas entradas.	21
5.1	Subespaço de ICA apreendido a partir de imagens de faces.	27
5.2	Subespaço de PCA apreendido a partir de imagens de faces.	27
5.3	Subespaço da DCT.	27
5.4	WPRD para a reconstrução da imagem “Lena”, utilizando cinco <i>bits</i> no processo de quantização e um número variável de funções base.	29

5.5	WPRD médio para a reconstrução de cinco imagens de faces da AR Face Database não utilizadas na fase de treinamento. Foram usados cinco <i>bits</i> no processo de quantização e um número variável de funções base	30
5.6	WPRD para a reconstrução da imagem “Boat” utilizando cinco <i>bits</i> no processo de quantização e um número variável de funções base.	31
5.7	Imagens da Lena, quantizadas a cinco <i>bits</i> e reconstruídas a partir DCT, PCA e ICA, com seus respectivos valores do PQS e WPRD sendo na primeira coluna utilizadas 4 funções base e na segunda 6 funções base.	32
5.8	Imagens de face, quantizadas a cinco <i>bits</i> e reconstruídas a partir da DCT, PCA e ICA com 4 funções base e de cada lado os respectivos zoom.	33
5.9	Imagem de paisagem(<i>boat</i>), quantizadas a cinco <i>bits</i> e reconstruídas a partir da DCT, PCA e ICA com seus respectivos valores do PQS e WPRD, sendo na primeira coluna utilizadas 4 funções base e na segunda 6 funções base.	34

List of Tables

3.1 Relação entre ICA, PCA e DCT. 17

Lista de Abreviações

BSS	<i>Blind Source Separation</i>
CR	<i>Compression Ratio</i>
CVP	<i>Córtex Visual Primário</i>
DCT	<i>Discrete Cosine Transform</i>
DWT	<i>Discrete Wavelet Transform</i>
ICA	<i>Independent Component Analysis</i>
JPEG	<i>Joint Photographic Experts Group</i>
KLT	<i>Karhunen-Loève Transform</i>
MSE	<i>Mean-Square Error</i>
MOS	<i>Mean Opinion Score</i>
MPEG	<i>Moving Picture Experts Group</i>
PCA	<i>Principal Components Cnalysis</i>
SVH	<i>Sistema visual Humano</i>
WPRD	<i>Weighted Percent Root-Mean-Square Diffrence</i>

1 Introdução

A redundância é um assunto antigo em pesquisa de compressão de dados [KORTAMN (1967)]. O termo *compressão de dados* refere-se ao processo de redução da quantidade de dados necessária para representar uma certa quantidade de informação [GONZALES e WOODS (2000)]. Uma distinção clara deve ser feita entre *dados e informação*. Dados são os meios pelos quais a informação é conduzida. Por exemplo, duas pessoas podem relatar a mesma história, uma contando de maneira longa, a outra de maneira curta. Aqui, a informação de interesse é a história, e as palavras são os dados usados para relatar a informação. Se dois indivíduos usam quantidades de palavras diferentes para relatar a mesma história, pelo menos um deles usa dados irrelevantes ou simplesmente repete. Dizemos assim que há redundância de dados. Assim, o objetivo da compressão de dados é reduzir a redundância dos dados com o menor impacto nos dados descomprimidos.

O princípio da redução de redundância pode ser analisado nas formas determinística e estatística [KORTAMN (1967), BARROS (2002)]. Na primeira, a redundância é definida como amostras de dados que podem ser deduzidas através de modelos determinísticos tais como filtragem. No segundo, a redução de redundância efetua a transformação dos dados em uma representação eficiente, de acordo com o critério de independência estatística [KORTAMN (1967)].

Há um grande número de métodos determinísticos usados para a compressão de imagens. Por exemplo, o método bem conhecido JPEG emprega a Transformada Discreta de Cosseno (DCT) [N. AHMED (1974), WALLACE(1991)] para codificar imagens. A DCT é similar à Transformada de Fourier, que converte dados em componentes de frequência. Porém, ao contrário da Transformada Fourier, as componentes são coeficientes reais definidos como o produto interno entre a imagem a ser codificada e as funções base da DCT, estas funções base são cossenos ortogonais. Outro método que surgiu a partir do JPEG é o JPEG2000 [GONZALES, WOODS e EDDINS (2004)], sendo que esse utilizar *Discrete Wavelet Transform* (DWT), esse método não será analisado no presente trabalho.

Embora a DCT seja fácil de implementar e computacionalmente rápida, sujeita-

se a algumas desvantagens relacionadas à Transformada de Fourier. O fenômeno de Gibbs [GIBBS (1898)] é um exemplo dessa deficiência, onde as imagens aparecem com bordas apagadas. Um outro problema da DCT está nos “artefatos de blocos” [COUDOUX (1997)]. Estes artefatos são devidos ao processamento independente dos blocos nas imagens e podem ser entendidos como descontinuidades de luminância entre os limites dos blocos.

Por outro lado, os métodos estatísticos tiveram grandes influências através dos modelos neurais para o processamento de informação [BARROS (2002)]. Estudos em neurociência sugeriram que uma população de neurônios processa informações dos estímulos de acordo com o conceito de “codificação eficiente” [SIMONCELLI (2001)]. Sob este conceito, as respostas do neurônio são mutuamente estatisticamente independentes, o que significa que não há nenhuma “informação redundante” entre a população.

O critério estatístico de independência pode ser explorado por dois métodos: análise de componentes principais (PCA) e análise de componentes independentes (ICA). PCA utiliza estatística de segunda ordem enquanto ICA usa estatística de alta ordem para obter uma codificação eficiente. Por exemplo, PCA é empregado em diversos sistemas de compressão de imagens, a fim de reduzir a dimensão dos dados [DONY (1995), S. COSTA (2001)]. Sua deficiência é que ela se baseia em estatística de segunda ordem, que podem fornecer somente representações eficientes para dados gaussianos e imagens são normalmente dados não gaussianos [SIMONCELLI (2001)]. Para contornar este problema, novos sistemas de compressão usam ICA para codificar as imagens [FERREIRA (2006)].

A finalidade deste trabalho é desenvolver um sistema de compressão de imagens baseado no conceito de codificação eficiente através da projeção das imagens nos subespaços apreendidos por ICA, onde o código eficiente é dado pelos respectivos coeficientes de projeção.

1.1 Organização do Trabalho

Neste capítulo, foi apresentada a relevância da redução de redundância na compressão de imagens. Neste sentido, mostrou a DCT e PCA como um método a ser melhorado. Para isso, propõe-se utilizar a técnica de codificação eficiente. O restante do trabalho está organizado da seguinte forma:

No **capítulo 2**, é apresentado o modelamento do olho humano (como ele recebe e

processa a informação visual através do globo ocular) a partir do codificação eficiente.

No **capítulo 3**, são apresentadas a teoria e os termos técnicos sobre análise de componentes independentes.

No **capítulo 4**, descreve-se o método proposto, seu diagrama de blocos e a aplicação de análise independente em processamento de imagens.

No **capítulo 5**, são apresentados os resultados e as discussões.

Finalmente, o **capítulo 6** as *Conclusões e comentários finais* resume os resultados obtidos e as principais conclusões. Identificam-se pontos de expansão para futuros trabalhos .

2 Codificação Eficiente

Uma das perguntas mais importantes na área de neurociência é como o cérebro humano recebe e processa as informações sensoriais, i.e., táteis, visuais ou auditivas. Sabe-se que há regiões específicas que processam informações, tais como o córtex visual primário na região posterior do cérebro Figura 2.1. Esses resultados da literatura mostram que se pode modelar através de ferramentas estatísticas e redes neuronais artificiais as respostas, por exemplo, dos campos receptivos do córtex visual primário (área V1).

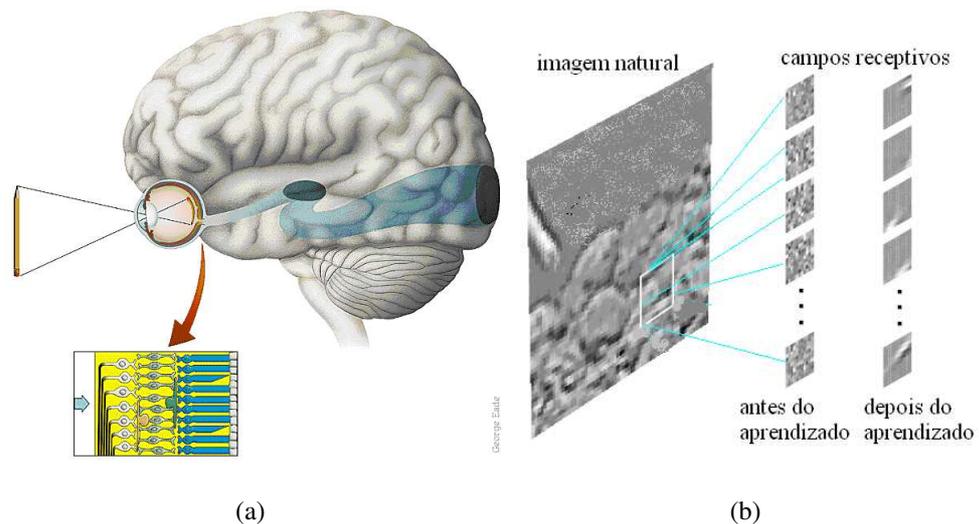


Figure 2.1: (a) Exemplo de como o olho humano recebe a informação visual através do globo ocular, a transmite ao nervo óptico e a repassa para a região posterior do cérebro, denominada de córtex visual primário (CVP). No córtex visual primário, as células têm campos receptivos que tentamos mimicar; (b) exemplo de uma imagem e como ela seria “dividida” ou codificada nos campos receptivos de V1, no CVP.

É importante que se saiba como os neurônios são excitados ou inibidos e que tipo de estratégias são utilizadas para tal atividade. No nosso caso, estamos interessados em modelar esse processamento. Supõe-se que o princípio da codificação eficiente é umas das estratégias utilizadas pelo cérebro. Este conceito foi proposto por Horace Barlow, em 1961, como um modelo teórico para a codificação das informações sensoriais pelo sistema nervoso [BARLOW (1961)]. Sua hipótese surgiu da busca de responder como os impulsos elétricos nervosos que transmitem as informações dos sistemas sensoriais, formariam um código que transmitisse de forma eficiente essas informações. Para ele, um modelo eficiente seria aquele

que minimizasse a quantidade de impulsos nervosos para transmitir a informação desejada.

No desenvolvimento de sua teoria, Barlow foi inspirado por conceitos da Teoria da Informação. Ele abordou os caminhos percorridos pelas informações sensoriais através dos neurônios como sendo canais de comunicação e aplicou conceitos como capacidade do canal e redundância da informação. Então, a codificação eficiente das informações sensoriais, que atravessassem tais canais, aconteceria de forma a maximizar a capacidade do canal e reduzir a redundância na informação transmitida.

Para visualizarmos um modelo baseado nesse princípio de codificação eficiente, imaginemos um sinal no tempo $x(t)$, em uma janela de tempo N , que seja codificado em um conjunto de M respostas, $a_1(t), a_2(t), \dots, a_M(t)$. O objetivo da codificação eficiente é então encontrar um conjunto de filtros $h_1(t), h_2(t), \dots, h_M(t)$ que maximize a independência estatística das respostas, ou seja, que reduza a redundância mútua. A resposta de um filtro particular i é:

$$a_i(t) = \sum_{\tau=0}^{N-1} x(\tau)h(t - \tau). \quad (2.1)$$

Baseando-se no conceito de Barlow de codificação eficiente para informações sensoriais, podemos inferir, portanto, que esse modelo deve ser bem adaptado à codificação do som pelo sistema auditivo. Presume-se que o sistema auditivo humano seja adaptado a codificar eficientemente os sons que são, de alguma forma, relevantes para a sobrevivência da espécie; que são basicamente sons naturais do ambiente, vocalizações de animais e a fala. Michael Lewicki, em 2002 [LEWICK (2002)], mostrou que a codificação eficiente de um conjunto que envolva essas três classes de sons origina filtros cujas respostas se aproximam das respostas medidas das fibras do nervo auditivo. É importante que se tenha em mente que, tendo esse modelo, precisa-se de técnicas matemáticas para encontrar os filtros da codificação. Uma forma proposta para encontrar esse filtros é através da Análise de Componentes Independentes (ICA). Essa técnica visa encontrar uma representação linear de dados não gaussianos, de modo a obter componentes com máxima independência estatística [HYVÄRINEN & OJA (2000)] e dessa forma associarmos essas representações lineares como filtros da codificação eficiente. No modelo de ICA, é comum se chamar o conjunto de filtros de funções base.

3 Análise de Componentes Independentes

O presente capítulo introduz o conceito da análise de componentes independentes (Independent Component Analysis-ICA), que fornece o suporte teórico.

3.1 História

Durante muitos anos, foi tida como impossível a idéia de recuperação de sinais combinados no processo de transmissão, a partir somente de diferentes combinações obtidas e observadas num certo número limitado de sensores. Apenas em 1985, através do trabalho de Héroult, Jutten e Ans [HERAULT *et al* (1985)], começou-se a delinear a área denominada na época de separação cega de fontes (Blind Source Separation-BSS), mostrando a possibilidade de solução de tal problema através do uso de estrutura não-lineares. A partir deste momento, muito se avançou nas ferramentas utilizadas e pela comunidade de processamento de sinais.

A ICA foi desenvolvida no contexto de separação cega de fontes (Blind Source Separation-BSS). Uma das aplicações da técnica de ICA é a análise ou separação de fontes estatisticamente independentes a partir de um determinado modelo de mistura das fontes originais [CASIMIRO (2004)]. A estrutura geral de análise de componentes independentes foi introduzida por Jean Héroult e Christian Jutten em 1986, sendo mais claramente proposta por Pierre Comon em 1994.

Em 1995, Tony Bell e Terry Sejnowski introduziram um algoritmo rápido e eficiente de ICA baseado no infomax, um princípio criado por Ralph Linsker em 1992. Em 1997, Shunichi Amari melhorou este algoritmo usando o gradiente natural, o que foi descoberto de forma independente por Jean Francois Cardoso. Entretanto, o algoritmo original com não-linearidades sigmóides era apenas apropriado para fontes super-Gaussianas. Te-Won Lee e Mark Girolami desenvolveram uma versão ampliada e eficiente de algoritmo de ICA baseado no infomax, próprio para sinais não-Gaussianos em geral [LEE (1998)].

Diversas aproximações diferentes foram feitas em relação à análise de separação

cega de fontes, incluindo os métodos da probabilidade máxima, os métodos de Bussgang com ênfase em cumulantes, o método de procura da projeção e métodos de negentropia (entropia negativa). Todos eles são relacionados com a estrutura do Infomax [LEE (1998)]. Por isso, um grande número de pesquisadores que se dedicaram ao estudo de ICA, vindos de várias áreas, convergiram para um conjunto comum de princípios e de algoritmos.

Considere-se, por exemplo, os registros elétricos das atividades do cérebro, dados por um eletroencefalograma (EEG), consistindo dos potenciais elétricos em diferentes pontos do escalpo. Esses potenciais são supostamente gerados pela mistura de componentes implícitas da atividade cerebral. Essa situação é semelhante ao problema de *cocktail party*, pois desejamos encontrar as componentes originais da atividade cerebral, mas podemos apenas observar as misturas das componentes. ICA pode revelar informações interessantes sobre a atividade cerebral a partir das suas componentes independentes.

Outra aplicação importante de ICA é a extração de características. Um problema fundamental em processamento de imagem é encontrar a representação apropriada para imagem. Um exemplo é a classificação de lesões em mamografias digitais [LEITE (2006)]. A representação de dados pode ser feita através de transformações lineares, logo é bastante útil estimar tal transformação dos próprios dados, pois nesse caso a transformação seria adaptada ao tipo de dado processado [HYVÄRINEN & OJA (2000), HYVÄRINEN *et al* (2001)].

3.2 Definições

Considere-se que sejam observadas n misturas lineares x_1, \dots, x_n de um sinal, modeladas como combinação linear de n funções base

$$x_i = \phi_1 s_1 + \phi_2 s_2 + \dots + \phi_n s_n \quad (3.1)$$

e que cada amostra x_i , assim como cada componente s_i seja uma variável aleatória. Em análise funcional e suas aplicações, um espaço de funções pode ser visto como um espaço vetorial de infinitas dimensões, cujos vetores bases são funções e não vetores. Isso significa que cada função no espaço de funções pode ser representada como combinação linear das funções base [LEE (1998)]. As Figuras 3.1 e 3.2 ilustram o proposto acima.

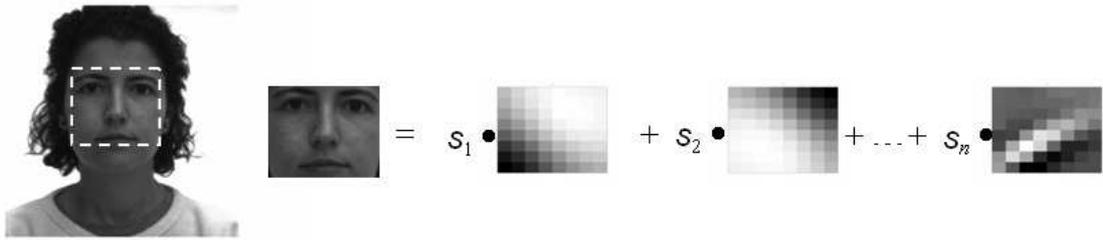


Figure 3.1: Amostra de uma imagem como combinação linear de suas características.

Sem perda de generalidade, supomos que tanto as variáveis originadas das amostras x_i da imagem quanto aquelas das componentes s_i têm média zero. Por conveniência, será usada a notação vetorial em vez de somas, como aquelas vistas em (3.1), utilizando letras minúsculas em negrito para matrizes e não negrito para vetores colunas.

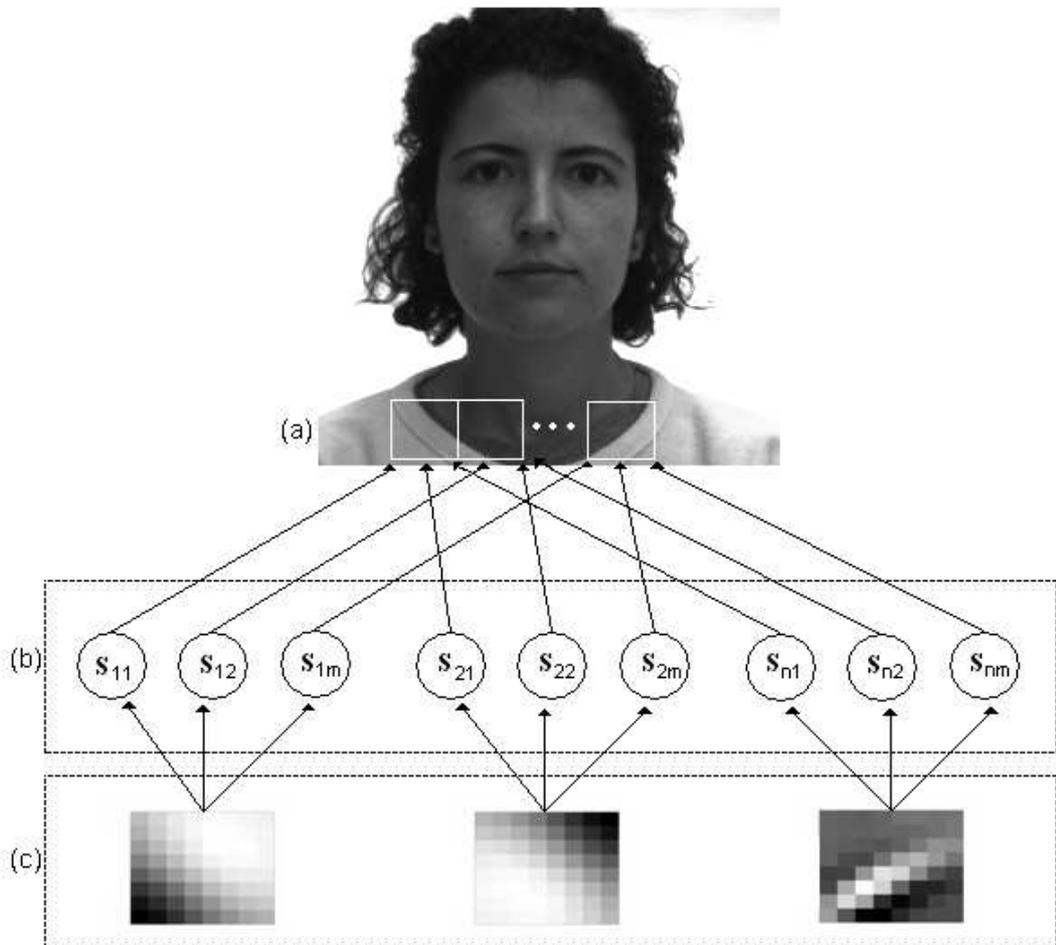


Figure 3.2: Amostra de uma imagem disposta como combinação linear das funções base do sinal. (a) Amostras da imagem. (b) Componentes s_n . (c) Funções base.

Dessa maneira similar, podemos reescrever (3.1) da seguintes forma:

$$\mathbf{x} = \Phi \mathbf{s}. \quad (3.2)$$

O modelo estatístico de (3.2) é chamado de modelo de análise de componentes independentes. É preciso estimar tanto a matriz de componentes independentes \mathbf{s} quanto a matriz de funções base Φ , que também é desconhecida, pois tudo o que se observa são as amostras do sinal \mathbf{x} .

Para tanto, é preciso fazer suposições tão gerais quanto possível. Portanto, supomos que [HYVÄRINEN *et al* (2001)]:

- a) As componentes s_n são estatisticamente independentes;
- b) Elas têm distribuições não-gaussianas;
- c) Por motivos de simplicidade, a matriz Φ é quadrada.

O modelo de ICA apresenta algumas ambigüidades em relação às componentes independentes. São elas:

1. Não se pode determinar suas variâncias (energias);
2. Não se pode determinar a ordem em que serão estimadas.

Ambas derivam do fato de \mathbf{s} e Φ serem desconhecidas. No item 1, existe ambigüidade já que qualquer escalar α_n multiplicando uma das fontes s_n pode ser cancelado dividindo-se a coluna ϕ_n correspondente pelo mesmo escalar α_n , ou seja

$$x_i = \sum_n \left(\frac{1}{\alpha_n} \phi_n \right) (s_n \alpha_n). \quad (3.3)$$

Daí também ocorre a ambigüidade de sinal, pois é possível multiplicar uma componente por -1 . Já no item 2, a ambigüidade ocorre devido à possibilidade de se alterar livremente a ordem dos termos em (3.1), denominando qualquer componente como a primeira.

3.3 Estimação de Componentes Independentes

3.3.1 Estimação Através de Maximização de Não-Gaussianidade

A não-gaussianidade é um elemento chave para a estimação das componentes independentes no modelo de ICA, pois a matriz Φ não é identificável quando as componentes independentes têm distribuição gaussiana [ROBERT (2006)]. Assumimos que \mathbf{x} seja uma das amostras do sinal, conforme a Figura 3.2, distribuído de acordo com o modelo de ICA em (3.2) e que todas as componentes independentes \mathbf{s} têm distribuições iguais. Para estimar as componentes independentes, basta encontrar as combinações lineares corretas de \mathbf{x}_i , de modo que

$$\mathbf{s} = \Phi^{-1}\mathbf{x}. \quad (3.4)$$

Suponha, então, uma combinação linear qualquer dos vetores x_i dada por

$$y = \mathbf{b}^T \mathbf{x} \quad (3.5)$$

como $\mathbf{x} = \mathbf{A}\mathbf{s}$, pode-se escrever:

$$y = \sum_i b_i x_i \quad (3.6a)$$

$$= \mathbf{b}^T \Phi \mathbf{s}, \quad (3.6b)$$

onde \mathbf{b} deve ser determinado. A partir de (3.6b), podemos observar que y é uma combinação linear de s_i , com coeficientes dados por $\mathbf{q} = \mathbf{b}^T \Phi$. Logo, obtemos

$$y = \mathbf{q}^T \mathbf{s} \quad (3.7a)$$

$$= \sum_i q_i s_i. \quad (3.7b)$$

Se \mathbf{b} corresponder a uma das linhas da inversa de Φ , então y será uma das componentes independentes e, nesse caso, apenas um dos elementos de \mathbf{q} será igual a 1, enquanto todos os outros serão iguais a zero.

Lembre-se, porém, que apenas o vetor de mistura \mathbf{x} é conhecido e portanto o vetor b não pode ser determinado exatamente. Precisa-se encontrar um estimador que forneça uma aproximação de b .

Uma forma de determinar b é variar os coeficientes em q e então verificar como a distribuição de $y = q^T \mathbf{s}$ muda. De acordo com o Teorema do Limite Central [PAPOULIS & PILLAI (2002)], a soma de duas variáveis aleatórias independentes é mais gaussiana que as variáveis originais [PAPOULIS & PILLAI (2002)], $y = q^T \mathbf{s}$ normalmente é mais gaussiana que qualquer uma das s_i e menos gaussiana quando se iguala a uma das s_i . Nesse caso, apenas um dos elementos q_i de q é diferente de zero [HYVÄRINEN *et al* (2001)].

Como, na prática, os valores de q são desconhecidos e sabemos, através de (3.5) e (3.7a), que

$$b^T \mathbf{x} = q^T \mathbf{s}, \quad (3.8)$$

podemos variar b e observar a distribuição de $b^T \mathbf{x}$. Portanto, podemos tomar, como b , um vetor que maximiza a não-gaussianidade de $b^T \mathbf{x}$, sendo que esse vetor necessariamente corresponde a $q = \Phi^T \mathbf{s}$, vetor esse que possui apenas uma de suas componentes diferente de zero. Isso significa que y em (3.5) é igual a uma das componentes independentes. Logo, a maximização da não-gaussianidade de $b^T \mathbf{x}$ permite encontrar uma das componentes. Resta, então, apenas encontrar uma medida para gaussianidade da distribuição de y .

3.3.2 Negentropia como Medida de Não-Gaussianidade

A negentropia é uma importante medida de não-gaussianidade. A definição de entropia [HAYKIN (2001), PAPOULIS & PILLAI (2002), VITERBI & OMURA (1979)] pode ser generalizada para vetores e variáveis aleatórias contínuas, vindo a ser chamada entropia diferencial. A entropia de uma variável aleatória está relacionada à quantidade de informação que essa variável contém. A entropia será maior quanto mais imprevisível for a variável. Tomando uma variável aleatória y , cuja função densidade de probabilidade é $f(y)$, temos a entropia diferencial dada por

$$H(y) = - \int f(y) \log f(y). \quad (3.9)$$

Da Teoria da Informação, sabe-se que uma variável gaussiana tem a maior

entropia entre todas as variáveis aleatórias de igual variância [HYVÄRINEN *et al* (2001), PAPOULIS & PILLAI (2002)]. Isso quer dizer que uma versão modificada da entropia diferencial pode ser usada como medida de não-gaussianidade. Essa medida é chamada negentropia, definida por

$$J(y) = H(y_{gauss}) - H(y), \quad (3.10)$$

sendo y_{gauss} uma variável aleatória de mesma matriz de covariância que y . A negentropia sempre é não-negativa, tem valor igual a zero se e somente se y tem distribuição gaussiana invariante para transformações lineares inversíveis.

Em contraste às suas qualidades como medida de não-gaussianidade, a negentropia é de difícil estimação. Por isso, é necessária a utilização de aproximações usando, por exemplo, momentos de alta ordem. Logo

$$J(\mathbf{y}) \approx \frac{1}{12} E \{y^3\}^2 + \frac{1}{48} \text{kurt}(y)^2 \quad (3.11)$$

sendo $\text{kurt}(y)$, ou seja, a kurtose de y , definida como o momento de quarta ordem da variável aleatória y , expresso por

$$\text{kurt}(y) = E \{y^4\} - 3(E \{y^2\})^2. \quad (3.12)$$

3.4 Análise de Componentes Principais

A análise de componentes principais (PCA) [HYVÄRINEN *et al* (2001), JAIN (1989), LEE (1998), PAJUNEN (1989)] é uma técnica estatística poderosa que pode ser utilizada para estudar correlações entre dados, ou seja, determinar as direções principais dos mesmos. Entende-se como direções principais o conjunto de vetores ortogonais sobre os quais os dados apresentam maior variância. O primeiro vetor representa a direção de máxima variância, o segundo vetor também está disposto segundo a direção de máxima variância sob a condição de ser ortogonal ao primeiro, e assim sucessivamente para os restante dos vetores.

A PCA também é chamada de transformada de Karhunen-Loève. Essa transformada tem a distinção de ser a transformação linear ótima para manter o subespaço que tem a maior variância. Entretanto, o custo computacional dessa transformação é muito grande. Ao contrário de outras transformações lineares, KLT não tem um número fixo de funções base.

O objetivo desta técnica é então encontrar o melhor conjunto de imagens-bases, tal

que as coordenadas da imagem (que correspondem aos coeficientes PCA) na nova base são não correlacionadas. [HYVÄRINEN *et al* (2001)].

Considere a combinação linear dos elementos de \mathbf{x}

$$y_1 = \sum_{k=1}^n w_{k1} x_k = \mathbf{w}_1^T \mathbf{x}. \quad (3.13)$$

Os termos w_{11}, \dots, w_{n1} são coeficientes escalares do vetor n -dimensional \mathbf{w}_1 . O fator y_1 é chamado de primeira componente principal de \mathbf{x} , caso a variância de y_1 seja a maior possível. Já que a variância depende tanto da norma quanto da orientação do vetor de coeficientes \mathbf{w}_1 e cresce sem limites à medida que a norma cresce, é imposta a restrição de que a norma de \mathbf{w}_1 seja igual a 1. Assim, procuramos o vetor de coeficientes \mathbf{w}_1 que maximize o critério

$$J_1^{PCA}(\mathbf{w}_1) = E[y_1^2] = E[(\mathbf{w}_1^T \mathbf{x})^2] \quad (3.14a)$$

$$= \mathbf{w}_1^T E[\mathbf{x}\mathbf{x}^T] \mathbf{w}_1 \quad (3.14b)$$

$$= \mathbf{w}_1^T \mathbf{C}_x \mathbf{w}_1, \quad (3.14c)$$

sendo que $\|\mathbf{w}\| = 1$ e definindo

$$\mathbf{C}_x = E[\mathbf{x}\mathbf{x}^T]. \quad (3.15)$$

A solução para o problema de PCA é dada em termos dos autovetores v_1, \dots, v_n de \mathbf{C}_x , ordenados de maneira que os autovalores d_1, \dots, d_n satisfaçam $d_1 \geq d_2 \geq \dots \geq d_n$. Assim a solução de (3.14c) é dada por

$$w_1 = v_1. \quad (3.16)$$

Deste modo, a primeira componente principal de \mathbf{x} é $y_1 = v_1^T \mathbf{x}$. O critério definido em (3.14) pode ser generalizado para m componentes, sendo m qualquer número entre 1 e n . Igualmente, para a m -ésima componente principal $y_m = w_m^T \mathbf{x}$, temos como solução $w_m = v_m$ e, portanto, $y_m = v_m^T \mathbf{x}$.

Seja um conjunto de m funções base ortonormais w_1, \dots, w_m tal que o erro médio quadrático entre \mathbf{x} e sua projeção seja mínimo. Então

$$J_{MSE}^{PCA} = E \left[\left\| \mathbf{x} - \sum_{i=1}^m (w_i^T \mathbf{x}) w_i \right\|^2 \right] \quad (3.17)$$

devido à ortogonalidade dos vetores w_i , (3.17) pode ser reescrito como

$$J_{MSE}^{PCA} = E [\|\mathbf{x}\|^2] - E \left[\sum_{i=1}^m (w_i^T \mathbf{x})^2 \right] \quad (3.18a)$$

$$= \text{tr}(\mathbf{C}_x) - \sum_{i=1}^m w_i^T \mathbf{C}_x w_i \quad (3.18b)$$

$$= \sum_{i=m+1}^n d_i, \quad (3.18c)$$

ou seja, o valor do MSE é igual à soma dos autovalores descartados v_{m+1}, \dots, v_n .

Uma importante aplicação de KLT é a compressão de dados, onde o vetor \mathbf{x} é o sinal original que é aproximado pela expansão truncada

$$\hat{\mathbf{x}} = \sum_{i=1}^m y_i e_i. \quad (3.19)$$

A partir de (3.18c) sabemos que o erro diminui à medida que mais termos são incluídos (3.19), até que se torne nulo quando $m = n$ ou todas as componentes principais são incluídas. O problema consiste em escolher m de forma que haja um compromisso entre erro e taxa de compressão.

3.5 Comparação entre PCA e ICA

PCA obtém vetores (direções) ortogonais para representação dos dados. O primeiro vetor estimado corresponde à direção de maior variância, o segundo vetor é ortogonal a este e assim sucessivamente. No caso de ICA, não existem restrições quanto à ortogonalidade dos vetores no processo de estimação. Dessa forma, dois vetores podem ser não ortogonais desde que sejam estatisticamente independentes.

3.6 Representação ou Codificação de Imagens por ICA

Nesta seção, são descritas as técnicas de representação ou codificação de imagens através do método da análise de componentes independentes. Uma imagem pode ser representada por uma função $I(x,y)$ de intensidade luminosa no ponto (x,y) percebido pelo Sistema Visual Humano (SVH). É uma quantidade finita, real e não-negativa

[GONZALES e WOODS (2000)]. Cada ponto da imagem é chamado de *pixel* (*picture element*).

A representação de imagens baseia-se em combinações lineares discretas dos dados observados. Considere que $I(x,y)$ denota o valor do *pixel* na escala de cinza, com coordenadas (x,y) de uma imagem com diferentes níveis na escala de cor de cinza. Segundo vários modelos de processamento de imagem, $I(x,y)$ é expresso como a combinação linear de funções base $a_i(x, y)$:

$$I(x,y) = \sum_{i=1}^n a_i(x, y) s_i. \quad (3.20)$$

onde s_i denota as variáveis independentes que não são diretamente observáveis, diferentes para cada imagem $I(x,y)$.

É possível considerar uma representação matricial, conhecida por modelo estatístico linear:

$$\mathbf{x} = \mathbf{A}\mathbf{s}. \quad (3.21)$$

bastando, para tal, associar os valores de todos os *pixel* num vetor coluna $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$.

Este modelo descreve os dados observados \mathbf{x} como sendo gerados através de um processo de mistura representado pela matriz \mathbf{A} que depende das variáveis independentes \mathbf{s} . Nesta representação \mathbf{x} é um vetor de dimensão n , \mathbf{A} é uma matriz quadrada de dimensão $n \times n$ e \mathbf{s} é um vetor de dimensão n .

Na maioria dos problemas, a matriz \mathbf{A} , designada por matriz de mistura, é desconhecida. É possível contornar esta objeção utilizando técnicas estatísticas que permitem estimar os coeficientes da matriz \mathbf{A} . Através do processo de separação é possível obter as variáveis independentes \mathbf{s} :

$$\mathbf{s} = \mathbf{W}\mathbf{x}, \quad (3.22)$$

onde a matriz \mathbf{W} se designa por matriz de separação ou matriz de pesos.

Como se dispõe apenas de informação relativa aos dados observados \mathbf{x} , a matriz de mistura \mathbf{A} , que transforma as componentes independentes \mathbf{s} na imagem original \mathbf{x} , é reconhecida, o que força a necessidade de estimar as componentes independentes. Caso a matriz \mathbf{A} fosse conhecida e inversível, seria fácil encontrar as componentes independentes, recorrendo à equação (3.21), bastando considerar $\mathbf{A} = \mathbf{W}^{-1}$

Assim, os algoritmos sugeridos para esta técnica têm por finalidade descrever um processo de estimação das componentes independentes. Este processo depende da estimação da matriz \mathbf{W} que representa o processo de mistura, por forma a obter-se:

$$\hat{\mathbf{s}} = \hat{\mathbf{W}}\mathbf{x}. \quad (3.23)$$

onde $\hat{\mathbf{s}}$ denota estimativa das componentes independentes \mathbf{s} .

3.6.1 Comparação entre DCT, PCA e ICA

Na codificação, a imagem é codificada previamente por uma transformada. Tipicamente adota-se essa representação para imagens por causa da grande redundância, ou seja, a grande correlação entre os *pixels* formadores da imagem [GOULART (2004)], e através dos coeficientes de projeção do sinal sobre cada função base, o processo de codificação por transformada reduz a correlação. Nesta seção, relaciona-se a ICA com as outras transformadas utilizadas.

A PCA (KLT), comentada em seções anteriores, é uma transformada baseada em estatísticas de segunda ordem. É obtida no sentido de compactação de energia dos coeficientes obtidos.

A DCT é sem dúvida, a mais popular das transformadas no mundo das comunicações de imagem, sendo usada nas normas de codificação JPEG, MPEG-1, MPEG-2 e MPEG-4, entre outros o que comprova o seu elevado desempenho. Por esse motivo, comparamos o método proposto com essa transformada. A Tabela 3.1 apresenta algumas propriedades entre as transformadas que iremos trabalhar. Outros comentários sobre a DCT poderão ser encontrados em [SALOMON (2004)].

Table 3.1: Relação entre ICA, PCA e DCT.

ICA	Utiliza estatística de ordem superior à segunda Maximiza a independência estatística entre os dados Não é ortogonal
PCA	Utiliza estatística de segunda ordem Descorrelaciona os dados Compactação de energia
DCT	Aproxima-se de PCA para imagens altamente correlacionadas [JAIN (1989)] Decomposição nas componentes de frequência Resolução na frequência

4 O Método Proposto

Neste capítulo, apresentaremos o método proposto. Os codificadores de imagens utilizam transformação [JAIN (1989), NELSON (1995)], esquematizado na Figura 4.1, constituído por dois blocos funcionais: transformada e quantização. O método proposto baseia-se nesse diagrama, o que será mais bem detalhado a seguir:

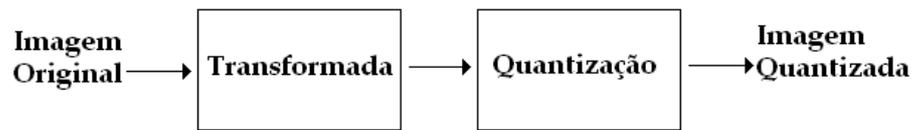


Figure 4.1: Diagrama de blocos do codificador de imagem. Primeiramente, a imagem original é transformada e logo após quantizada

Suponha que uma imagem seja dividida em blocos $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$ com tamanho $m \times n$. Agora, supomos que cada bloco y_i possa ser reconstruído como uma combinação linear de vetores de um subespaço estocástico aprendido, $\Phi = [\phi_1, \phi_2, \dots, \phi_n]^T$, onde ϕ_i é chamada de função base. Este processo pode ser escrito como

$$\hat{y}_i = w_1\phi_1 + w_2\phi_2 + \dots + w_n\phi_n, \quad (4.1)$$

onde \hat{y}_i é a versão reconstruída do bloco y_i e cada w_i é uma componente da projeção da imagem na i -ésima função base do subespaço Φ .

Para aprender o subespaço Φ e estimar os coeficientes de projeção w_i , propomos um sistema de compressão de imagem mostrado na Figura 4.2, que é baseado no conceito de codificação eficiente. As seguintes subseções fornecem uma explanação da estrutura de nosso modelo.

4.1 Codificador Eficiente

Suponhamos que uma imagem está codificada pelos coeficientes de projeção $\mathbf{w} = [w_1, w_2, \dots, w_n]^T$ em (4.1). O objetivo da codificação eficiente é estimar um subespaço Φ que reduz a dependência estatística mútua entre coeficientes w_i .

A estimação do subespaço Φ pode ser realizada por PCA ou ICA. O primeiro supõe que as componentes sejam não correlacionadas, enquanto o segundo supõe que as componentes sejam mutuamente estatisticamente independente. Neste trabalho, usamos essa última aproximação.

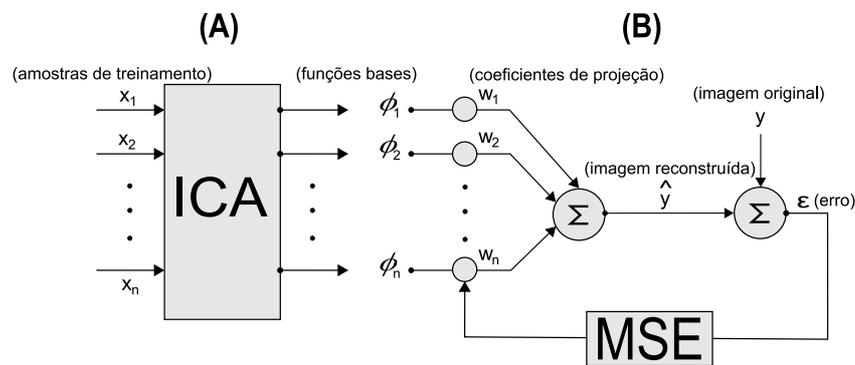


Figure 4.2: Sistema proposto de compressão de imagem. O sistema consiste em duas fases: (A) aprendizagem e (B) a fase de projeção, sendo $[x_1, x_2, \dots, x_n]$ as amostras de treinamento $[\phi_1, \phi_2, \dots, \phi_n]$ as funções base, $[w_1, w_2, \dots, w_n]$ os coeficientes de projeção, \hat{y} a imagem reconstruída, y a imagem original e ϵ o erro médio quadrático.

O sistema consiste em duas fases: aprendizagem e projeção. Na primeira fase (A), usamos a análise de componentes independentes para estimar o subespaço. Na fase de projeção (B), estimamos os coeficientes de projeções utilizando o erro médio quadrático.

4.1.1 Estimação das Funções Base Através de ICA

Supondo que temos um conjunto de k observações variáveis aleatórias x_1, \dots, x_k (segmentos, de comprimento n , de uma imagem usado como entrada do treinamento), modelados como uma combinação linear de n variáveis aleatórias a_1, \dots, a_n , tais que

$$x_i = \phi_{i1}a_1 + \phi_{i2}a_2 + \dots + \phi_{in}a_n \quad i = 1, \dots, k, \quad (4.2)$$

sendo que ϕ_{in} são coeficientes reais. Definimos \mathbf{x} , Φ e \mathbf{A} como

$$\mathbf{x} = [x_1, \dots, x_k] \quad (4.3)$$

$$\Phi = \begin{bmatrix} \phi_{11} & \cdots & \phi_{1n} \\ \vdots & & \vdots \\ \phi_{k1} & \cdots & \phi_{kn} \end{bmatrix} \quad (4.4)$$

$$\mathbf{A} = [a_1, \dots, a_n]. \quad (4.5)$$

usando (4.3) - (4.5), podemos reescrever (4.2) como

$$\mathbf{x} = \Phi^T \mathbf{A}. \quad (4.6)$$

Através de análise de componentes independentes, podemos determinar Φ de forma que os a_i sejam mutuamente estatisticamente independentes. Aqui, o algoritmo utilizado foi o *FastICA*, que trabalha com a maximização da estatística de quarta-ordem, descrito em [HYVÄRINEN *et al* (2001)] pelos seguintes passos:

1. Centralizar os dados para tornar sua média nula;
2. Branquear os dados, resultando em z ;
3. Escolher p , o número de componentes independentes a estimar;
4. Escolher os valores iniciais para b_i , $i = 1, \dots, p$, cada um de norma unitária. Ortogonalizar a matriz B como no passo 6 abaixo;
5. Para cada $i = 1, \dots, p$, seja

$$b_i \leftarrow E \{ z g(b_i^T z) \} - E \{ g'(b_i^T \mathbf{z}) \} w,$$

sendo g definido;

6. Ortogonalizar simetricamente a matriz $B = (b_1, \dots, b_p)^T$ fazendo

$$B \leftarrow (BB^T)^{-1/2} B.$$

7. Caso o algoritmo não convirja, repetir o passo 5.

Para o algoritmo proposto, foi utilizado $p = n$. Após a estimação de \mathbf{B} , podemos facilmente obter Φ .

4.1.2 Estimação dos Coeficientes da Projeção

A fase da projeção consiste em estimar os coeficientes \mathbf{w} utilizando um combinador linear, cuja estrutura é mostrada na Figura 4.3. Os termos ϕ_i , y_i , \hat{y}_i e \mathbf{w} correspondem respectivamente a funções base, imagem original, imagem reconstruída e os coeficientes da projeção. O sinal estimado \hat{y}_i pode ser expresso por

$$\hat{y}_i = \sum_{j=0}^n w_j \phi_j \quad (4.7a)$$

$$= \mathbf{w}^T \Phi \quad (4.7b)$$

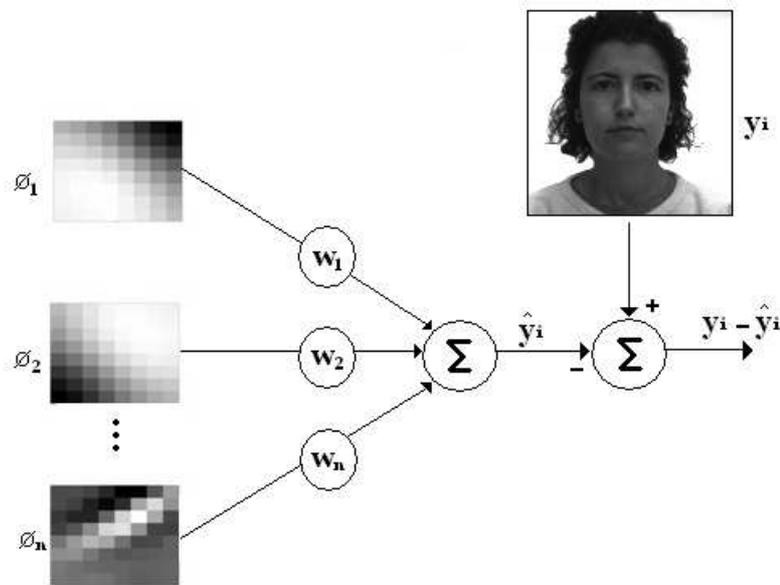


Figure 4.3: Combinador linear de múltiplas entradas.

Definindo o erro de estimação como sendo a diferença entre o sinal desejado e o sinal estimado, e considerando (4.7a)-(4.7b), obtemos

$$\varepsilon_i = y_i - \hat{y}_i \quad (4.8a)$$

$$= y_i - \mathbf{w}^T \Phi \quad (4.8b)$$

supondo que ε_i , y_i e ϕ_i sejam estatisticamente estacionários, tomamos o valor esperado de ε_i^2 ,

ou seja:

$$E [\boldsymbol{\varepsilon}_i^2] = E [(y_i - \mathbf{w}^T \Phi)^2] \quad (4.9a)$$

$$= E [y_i^2 - \mathbf{w}^T \Phi \Phi^T \mathbf{w} - 2y_i \Phi^T \mathbf{w}] \quad (4.9b)$$

$$= E [y_i^2] - \mathbf{w}^T E [\Phi \Phi^T] \mathbf{w} - 2E [y_i \Phi^T] \mathbf{w} \quad (4.9c)$$

Podemos definir

$$\mathbf{R} = E [\Phi \Phi^T] \quad (4.10)$$

$$\mathbf{P} = E [y_i \Phi^T], \quad (4.11)$$

sendo \mathbf{R} a matriz de autocorrelação de entrada e \mathbf{P} a matriz de correlação cruzada entre a entrada e o sinal desejado. Assim, designando o erro médio quadrático em (4.9c) por ξ e reescrevendo essa equação a partir de (4.10) e (4.11)

$$MSE \triangleq \xi = E [\boldsymbol{\varepsilon}_i^2] \quad (4.12a)$$

$$= E [y_i^2] - \mathbf{w}^T \mathbf{R} \mathbf{w} - 2\mathbf{P}^T \mathbf{w}. \quad (4.12b)$$

Desejamos encontrar a matriz de coeficientes $w_m = [w_1, \dots, w_n]^T$ que resulte em um valor mínimo para o MSE. Uma forma de encontrar tais valores é através dos métodos gradientes. O gradiente do MSE, definido por $\nabla(\xi)$, pode ser obtido pela diferenciação de (4.12b) [HAYKIN (1996), PRINCIPE et al (2000), WIDROW & STEARNS (1985)], ou seja

$$\nabla(\xi) \triangleq \frac{\partial \xi}{\partial \mathbf{w}} = \left[\frac{\partial \xi}{\partial w_1} \quad \frac{\partial \xi}{\partial w_2} \quad \dots \quad \frac{\partial \xi}{\partial w_n} \right]^T \quad (4.13a)$$

$$= 2\mathbf{R}\mathbf{w} - 2\mathbf{P}. \quad (4.13b)$$

O erro médio quadrático mínimo é obtido quando a matriz de coeficientes tem valor ótimo \mathbf{w}^* , onde o gradiente $\nabla(\xi)$ é igual a zero:

$$\nabla(\xi) = 0 = 2\mathbf{R}\mathbf{w}^* - 2\mathbf{P} \quad (4.14a)$$

$$\mathbf{w}^* = \mathbf{R}^{-1}\mathbf{P}. \quad (4.14b)$$

Para tanto, assumimos que \mathbf{R} seja não-singular. A expressão (4.14b) é chamada de equação de Wiener-Hopf.

4.1.3 Seleção das Funções Base

Para selecionar um número m de funções base de um determinado subespaço $\Phi = [\phi_1, \phi_2, \dots, \phi_n]^T$ de dimensão n , onde $m < n$, utiliza-se um processo de busca em deflação de forma que o MSE entre a imagem reconstruída \hat{y} definida e a imagem original y seja minimizado.

O algoritmo de busca pode ser resumido nos seguintes passos:

- **1.** Inicializar um subespaço vazio Ψ que irá conter as m funções base selecionadas a partir de Φ .
- **2.** Fazer $k = 1 : n$
- **4.** Encontrar a imagem reconstruída \hat{y}_k projetando y no subespaço formado por $[\Psi, \phi_k]$ através da Eq.(4.14b).
- **5.** Após $k=n$, selecionar a função base ϕ_k de acordo com o critério abaixo:

$$\phi_k = \operatorname{argmin}(MSE(\hat{y}_k - y)) \quad (4.15)$$

- **6.** Atualizar o subespaço Ψ inserindo a função base ϕ_k selecionada no passo anterior:
 $\Psi \leftarrow [\Psi, \phi_k]$.
- **7.** Retirar a função ϕ_k do subespaço original Φ fazendo com que $n = n - 1$.
- **8.** Se a dimensão de Ψ igual a m parar. Senão, voltar ao passo 2.

5 Resultados e Discussão

Neste capítulo apresentam-se as condições de teste para os dados de entrada do codificador. A partir dos dados obtidos, foram analisadas as medidas de distorção, a fim de se comparar o método proposto com outros algoritmos.

5.1 Base de Dados de Imagens Utilizada

Foram utilizadas 400 imagens retiradas da base de dados de imagens *AR Face Database* [BASE (1998)]. Essa base inclui 4.000 imagens coloridas de face de 126 pessoas, entre homens e mulheres. São imagens frontais, com diferentes expressões faciais, condições de iluminação e acessórios (óculos de sol e cicatrizes). Não há restrições quanto a roupas, óculos de grau, maquiagem e corte de cabelo.

5.2 Condições de Teste

Para a classe de face de imagem que utilizamos, pretendemos verificar a capacidade de reconstrução das funções base de PCA, DCT e ICA(método proposto) nas imagens, verificando a diferença entre a original e a descomprimida para uma determinada taxa de compressão.

A escolha das amostras sobre as imagens foram feitas na forma de blocos quadrados de dimensão $N \times N$, convertidos em vetores de N^2 elementos. Para escolha da dimensão dos blocos, alguns fatores foram considerados. O valor de N deve:

- Ser suficientemente grande para exibir a estrutura, ou seja, informação visual;
- Ser tal que o cálculo computacional não se torne pesado;
- Ser levado em conta aqueles utilizados nos métodos de compressão já existente [SALOMON (2004)];

Diante desses fatores utilizou-se o tamanho de $N=8$ por ser o tamanho de bloco comumente utilizado pelos algoritmos de compressão. Os blocos foram escolhidos aleatoriamente sobre as imagens, sem restrições de sobreposições.

5.3 Medida de Distorção

As técnicas normalmente utilizadas em compressão de imagem para análise do grau de degradação da imagem original com a imagem descomprimida são bons indicadores para erros dispersos na imagem. A ponderação da diferença percentual da raiz média quadrática (*Weighted Percent Root-Mean-Square Diffrence* - WPRD) é um critério para medir o grau de fidelidade da imagem reconstruída. O WPRD é encontrado através da filtragem da frequência espacial das imagens ($\hat{y}_i - y_i$), através da junção de sensibilidade da função de contraste (CF) [JAMES (1974)]. A filtragem tem como objetivo, combinar a influência do erro de acordo com a sensibilidade do sistema visual humano (SVH) para uma determinadas frequência. Os WPRD de um dado bloco y_i são definidos por Eq. (5.1), sendo definido como:

$$WPRD = \sqrt{\frac{E[CF(\hat{y}_i - y_i)^2]}{E[y_i^2]}}. \quad (5.1)$$

onde \hat{y}_i e y_i são a imagem original e a descomprimida. Mas o WPRD não é adequado para avaliar erros estruturados, os quais contribuem muito mais para a degradação visual das imagens [M. MIYAHARA (1998)]. O PQS está em concordância com a escala MOS (*mean opinion score*, da avaliação subjetiva perceptiva descrita na recomendação do ITU-R (*International telecommunication union - radiocommunications*) [INTERNATIONAL T. U. (1998)]. Os seguintes valores podem ser atribuídos ao se comparar as imagens original e reconstruída: 5-imperceptível, 4-perceptível mas não perturbante, 3-ligeiramente perturbante, 2-perturbante e 1-muito perturbante. O PQS combina cinco fatores de distorção visual:

- F_1 - Diferença entre a imagem original e a codificada, ponderada de acordo com a frequência, tal como definida pelo CCIR [CCIR. (1982)] (designação anterior do ITU-R);
- F_2 - Diferença entre a imagem original e a codificada, ponderada de acordo com um modelo de percepção visual do SVH;

- F_3 - Descontinuidades nas vizinhanças horizontal e vertical de blocos adjacentes;
- F_4 - Erros com correlação espacial;
- F_5 - Erros na proximidade de transições de elevado contraste.

Os primeiros dois fatores contabilizam perturbações e erros aleatórios globais, enquanto os três últimos se referem às perturbações e erros estruturados e localizados. O PQS é calculado através de

$$PQS = b_0 + \sum_{j=1}^J b_j Z_j, \quad (5.2)$$

em que $b_0 = 5.797$ e b_j são os coeficientes da regressão parcial, calculados através de MRA (*multiple regression analysis*) [M. KENDALL (1975)] de forma a verificar a gama de valores da escala MOS e Z_j são os coeficientes de representação, sobre as componentes principais calculadas a partir dos cinco fatores de distorção indicados.

O PQS é definido de acordo com a escala MOS, toma valores no intervalo 1, ..., 5. Há publicações [S. GRAPS. (2001)] que utilizam valores 0, ..., 5.797. No presente trabalho, devido as taxas de compressão, obtivemos valores negativos de PQS em alguns testes. Utilizou-se o PQS por considerar que traduz a qualidade visual das imagens.

5.4 Resultados

Na fase de aprendizagem, foram utilizadas 400 imagens monocromáticas, com resolução de 256x256 (8 *bit/pixel*), sendo 200 faces masculinas e 200 femininas. Para construir o vetor de treinamento para ICA, 1000 blocos de 8x8 foram extraídos de cada imagem sem nenhuma restrição de sobreposição. Cada bloco da imagem foi transformado um vetor de 64 *pixels*. As Figuras, 5.2 e 5.3 mostram exemplos de subespaços com dimensão $n = 64$ aprendidos a partir de imagens de faces usando ICA e PCA. A Figura 5.1 mostra o subespaço da DCT para comparação.

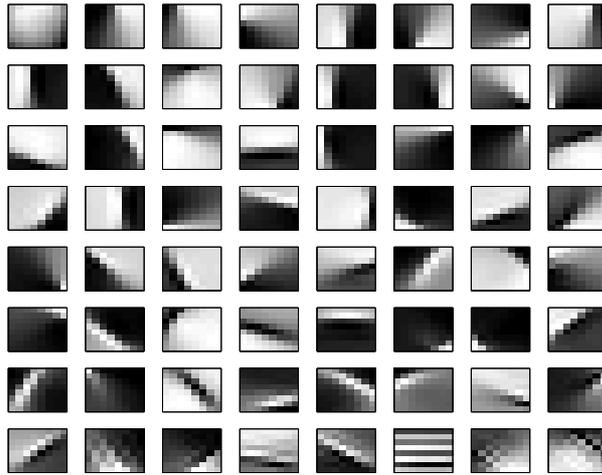


Figure 5.1: Subespaço de ICA apreendido a partir de imagens de faces.

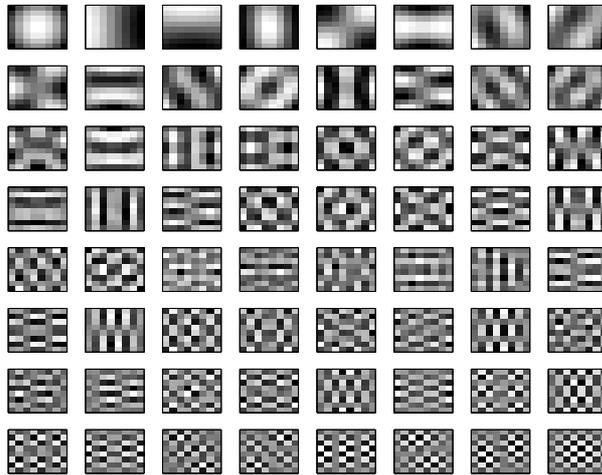


Figure 5.2: Subespaço de PCA apreendido a partir de imagens de faces.

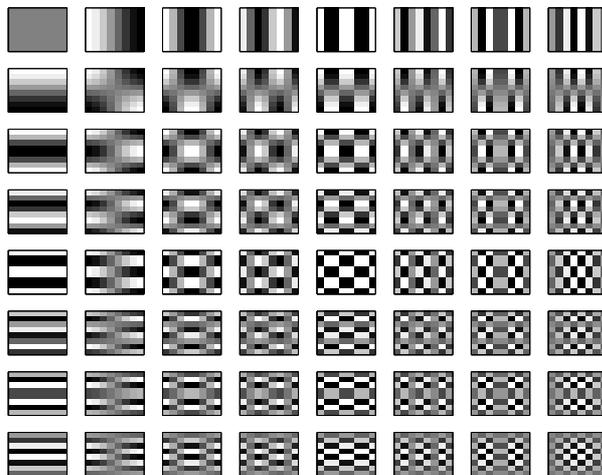


Figure 5.3: Subespaço da DCT.

Na fase de projeção, os coeficientes w foram quantizados usando 5 *bits*, para todos os resultados. A quantização desses coeficientes foi realizada através do algoritmo *Lloyd* [MAX (1960), LLOYD (1982)].

A quantidade de compressão obtida pela codificação, foi calculada pela taxa de compressão (CR) (do inglês, *compression ratio*), definido por

$$CR = \frac{\overline{M}}{\overline{L}} \quad (5.3)$$

sendo \overline{M} o comprimento da imagem original e \overline{L} , o da imagem comprimida.

Como o interesse está em estudar as perdas introduzidas pelo processo de codificação, não usamos nenhum método sem perdas adicionais, tais como a codificação de Huffman [HUFFMAN (1952)]. Para avaliar o desempenho do sistema proposto, comprimimos algumas imagens usando nosso modelo, PCA e DCT cujos coeficientes foram quantizados com os mesmos *bits* e para uma mesma taxa de compressão, para compararmos os resultados. Essas imagens não foram usadas na fase de aprendizagem do nosso método.

As Figuras 5.4, 5.5(b) e 5.6(c) mostram as médias do WPRD obtidas aumentando o número de funções base usadas no processo de reconstrução, através da base de face, para nosso modelo, PCA e DCT. Na Figura, 5.4 reconstruímos a imagem da Lena. Utilizamos essa imagem por ser uma imagem padrão em processamento de imagem e por ter outros detalhes fora da classe de face. Para a Figura 5.5, utilizamos cinco imagens reconstruídas de face para a mesma base e variamos as funções base. Na Figura 5.6, utilizamos a imagem de *boat*, com as mesmas variações das funções base utilizadas nas outras imagens, com o objetivo de busca característica não encontradas em face. Na próxima seção, comentaremos mais a respeito dessa imagem, projetada nesse tipo de classe(face).

Para análise subjetiva, a Figura 5.7 apresenta as imagens reconstruídas da Lena, com 4 e 6 funções base a partir da DCT, PCA e nosso método, e para Figura 5.8 reconstruímos as imagens de face. Os resultados foram analisados através do PQS e WPRD para uma mesma taxa de compressão, observamos o desempenho da DCT e PCA e a influência das funções base aumentando o número de funções base.

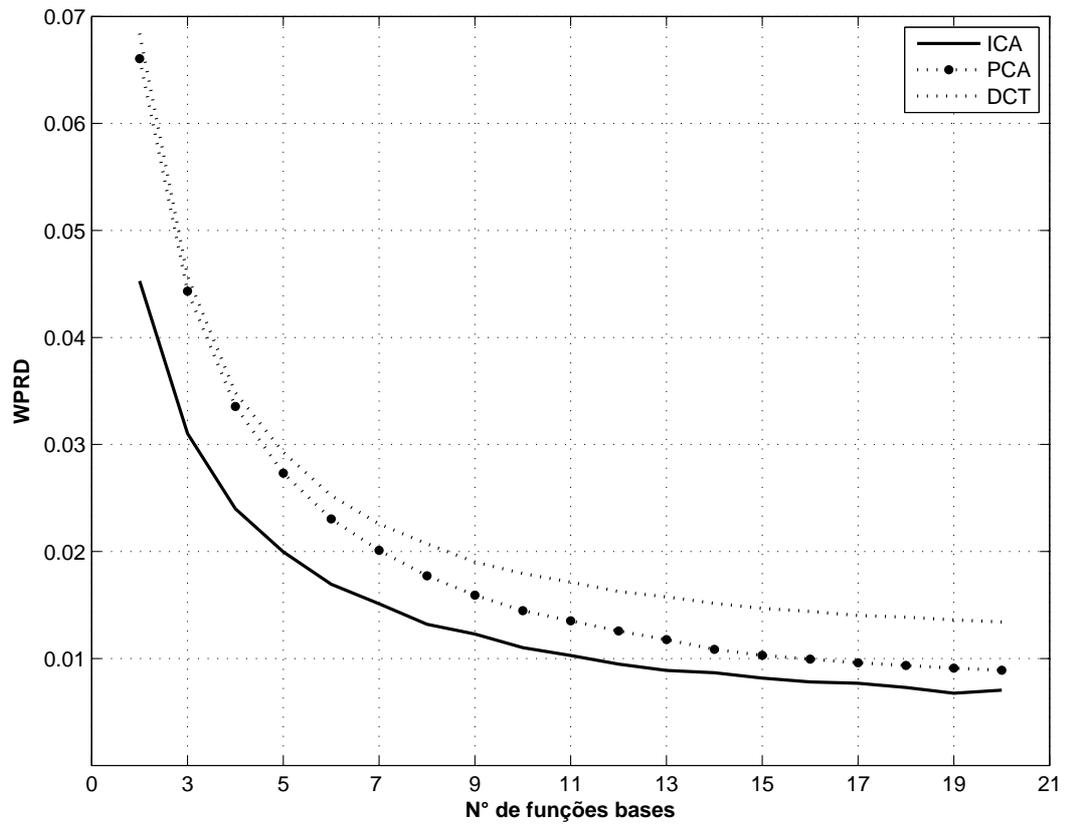


Figure 5.4: WPRD para a reconstrução da imagem “Lena”, utilizando cinco *bits* no processo de quantização e um número variável de funções base.

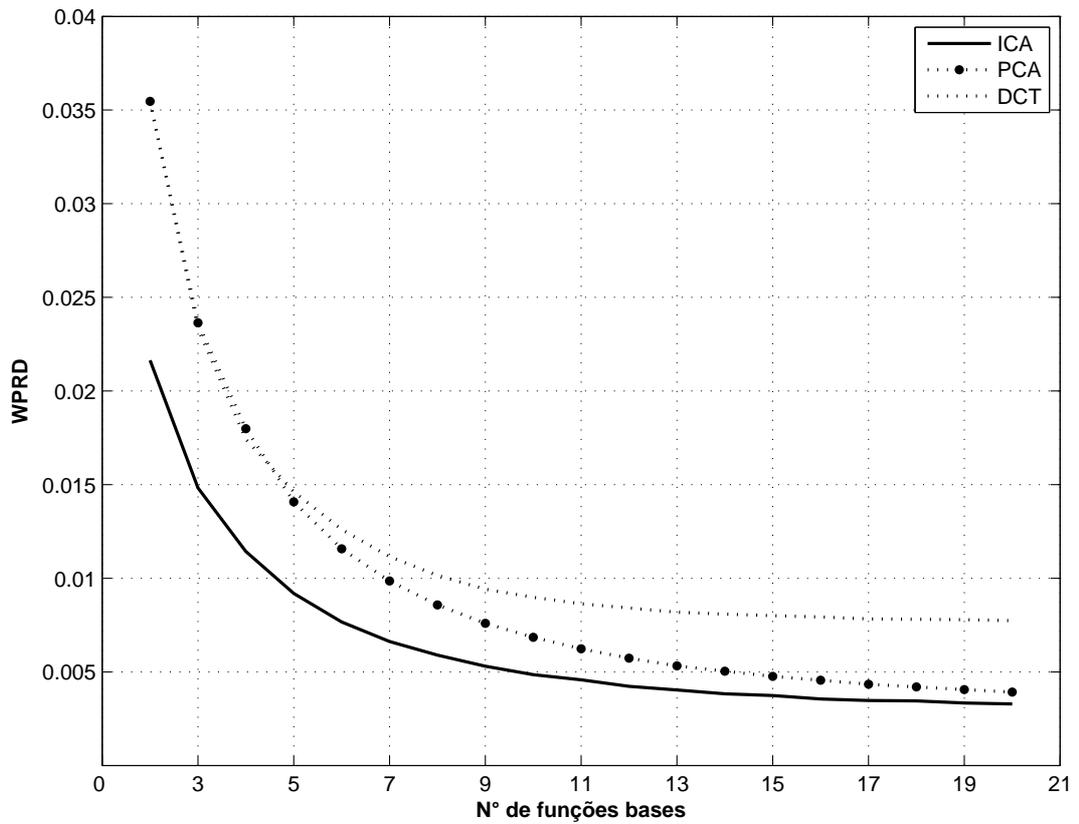


Figure 5.5: WPRD médio para a reconstrução de cinco imagens de faces da AR Face Database não utilizadas na fase de treinamento. Foram usados cinco *bits* no processo de quantização e um número variável de funções base

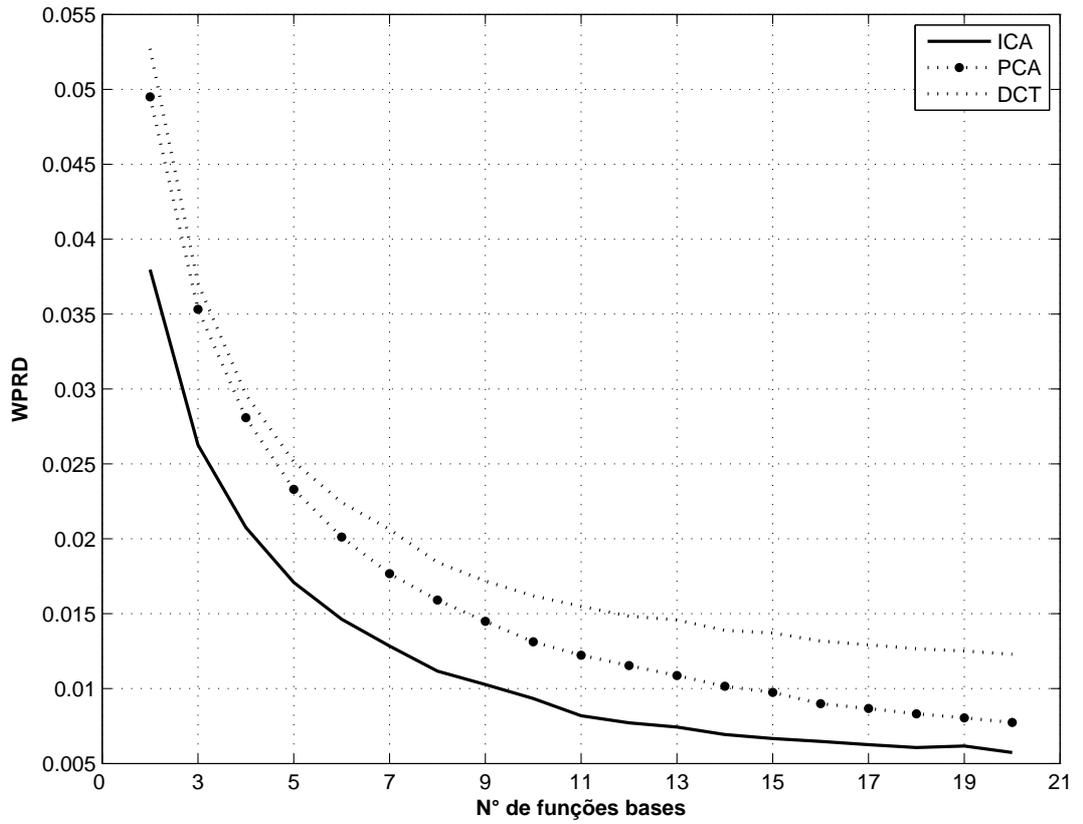


Figure 5.6: WPRD para a reconstrução da imagem “Boat” utilizando cinco *bits* no processo de quantização e um número variável de funções base.



Figure 5.7: Imagens da Lena, quantizadas a cinco *bits* e reconstruídas a partir DCT, PCA e ICA, com seus respectivos valores do PQS e WPRD sendo na primeira coluna utilizadas 4 funções base e na segunda 6 funções base.

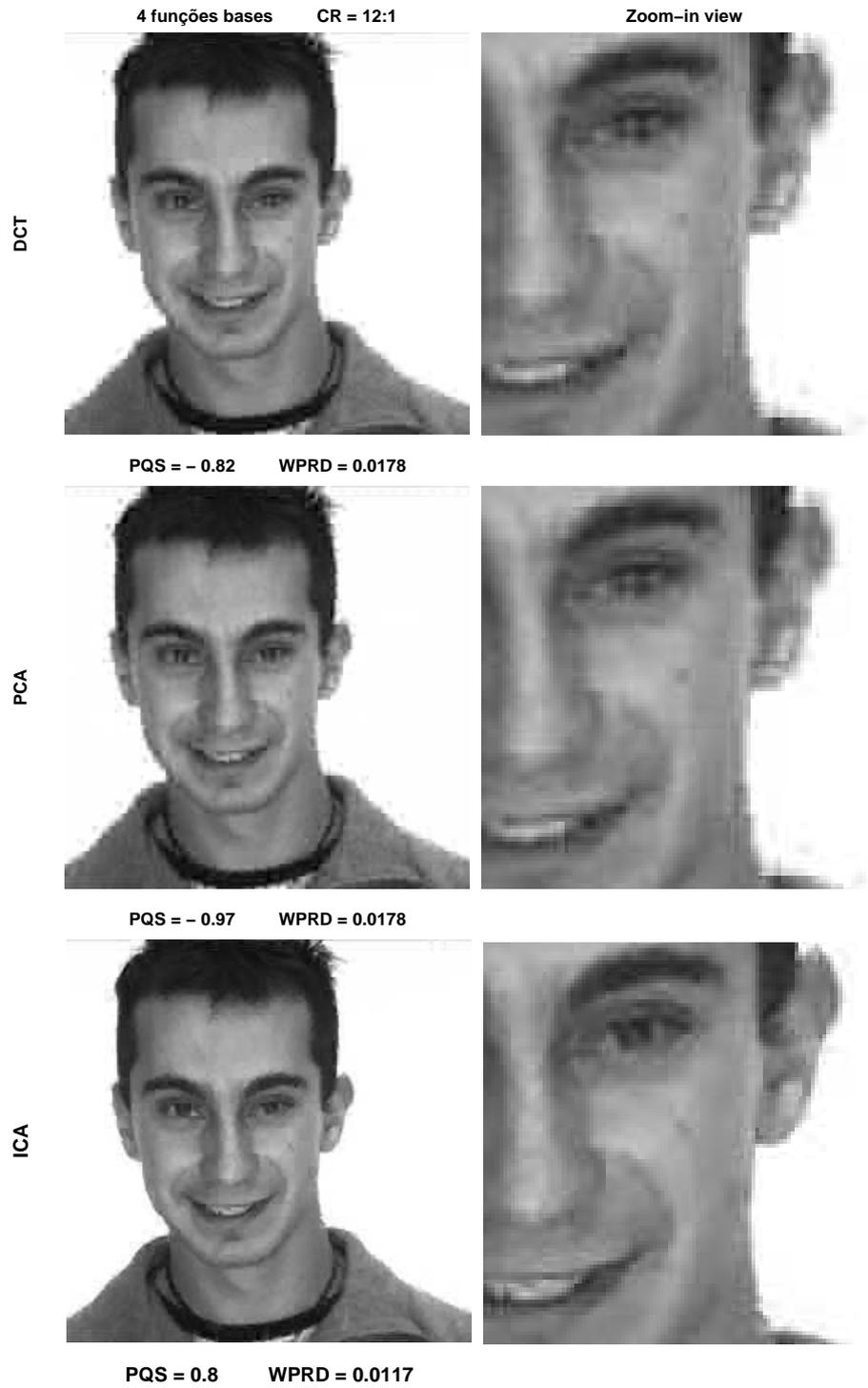


Figure 5.8: Imagens de face, quantizadas a cinco *bits* e reconstruídas a partir da DCT, PCA e ICA com 4 funções base e de cada lado os respectivos zoom.

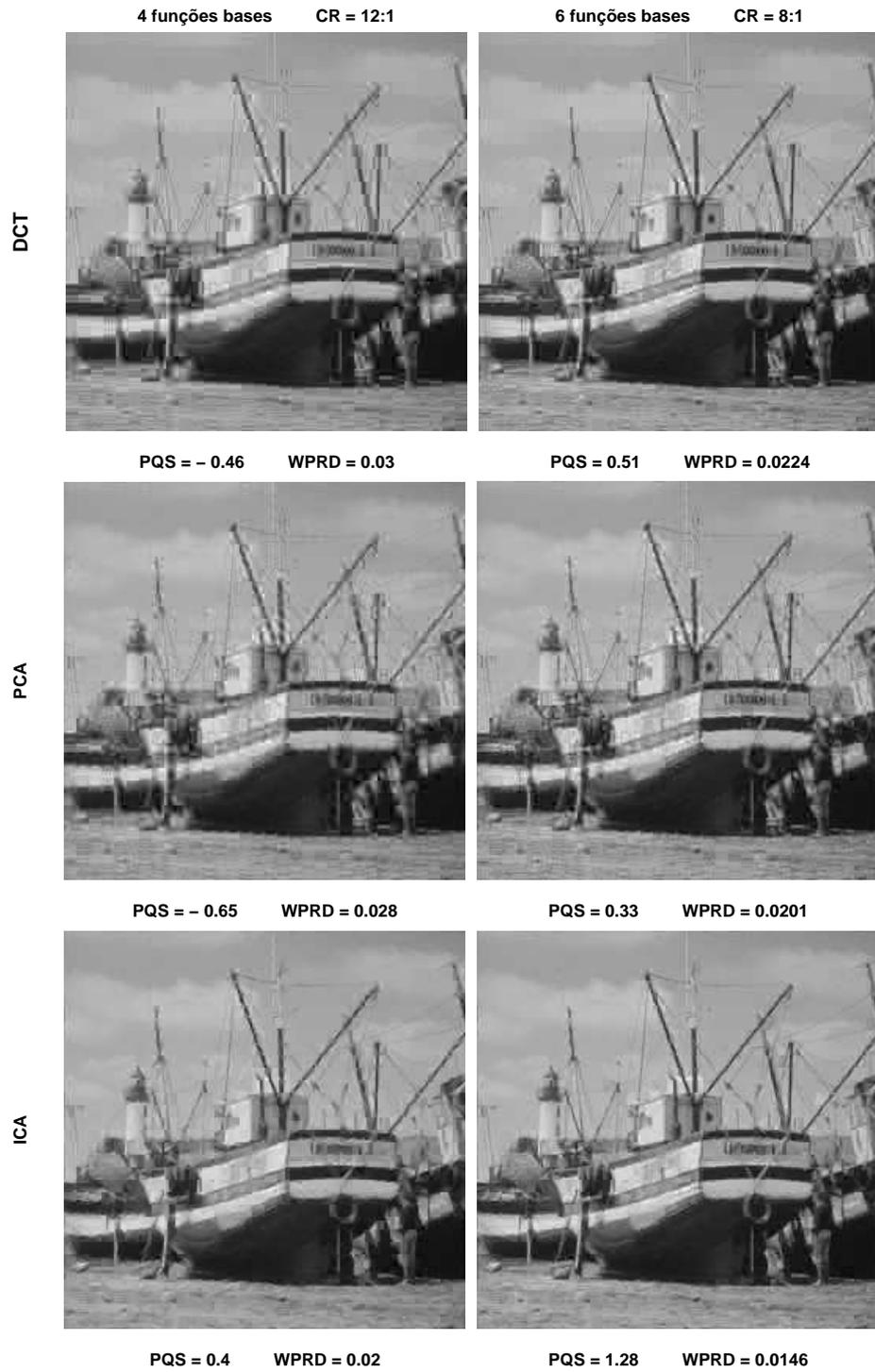


Figure 5.9: Imagem de paisagem(*boat*), quantizadas a cinco *bits* e reconstruídas a partir da DCT, PCA e ICA com seus respectivos valores do PQS e WPRD, sendo na primeira coluna utilizadas 4 funções base e na segunda 6 funções base.

5.5 Discussão

O modelo proposto tem uma fundamentação razoável: codificação eficiente. Conseqüentemente, este modelo deve introduzir menos distorções perceptíveis para os seres humanos que os métodos que não são baseados em modelos neurais, como a DCT ou métodos baseados em estatística de segunda ordem, como PCA.

Este fato pode ser parcialmente confirmado pela análise dos resultados de WPRD, mostrados na Figura 5.5, para cinco imagens reconstruídas usando o modelo proposto, PCA e DCT. Há diversos pontos interessantes a destacar. Primeiramente, conforme as Figuras 5.4, 5.5 e 5.6 o WPRD calculado mostra que nosso modelo introduz menos erros do que a DCT e PCA, a respeito de uma medida objetiva.

Para confirmar inteiramente que o método fornece as imagens reconstruídas, mais adaptadas aos seres humanos do que a DCT e PCA em termos de percepção visual, nós devemos discutir um ponto importante: a presença de *ringing* e de artefatos de bloco [COUDOUX (1997)]. Artefatos de *ringing* são produzidos pelo fenômeno de *gibbs* [GIBBS (1898)]. Este fenômeno pode ser compreendido como grandes oscilações na imagem, como descontinuidades nas bordas, o que podemos perceber na Figura 5.8 nas imagens com *zoom*. Estas oscilações produzem um efeito de suavização dos detalhes na imagem, o que pode ser inaceitável para aplicações como a compressão de imagem médica [STRIJM (1991)]. Observando as imagens reconstruídas nas Figuras 5.7 e 5.8, podemos ver que as bordas, que definem detalhes da imagem foram melhor preservadas no nosso modelo do que em DCT e PCA. Certamente, o *zoom* da imagem de face mostra que na região da orelha do indivíduo do nosso modelo e nos olhos da Lena as bordas possuem um contorno melhor do que a DCT e PCA.

O problema dos artefatos de “*ringing*” podem ser piores em conjunto com os artefatos de blocos. No modelo proposto, DCT e PCA, as imagens foram processadas (reconstruídas) em blocos sem sobreposição, que são transformados e quantizados independentemente. Conseqüentemente, a descontinuidade de luminância pode ser introduzida entre os limites de bloco, o que é altamente perceptível por seres humanos quando poucos coeficientes são usados no processo de reconstrução. Assim, ou o método de compressão é intrinsecamente robusto a estas distorção ou métodos adicionais podem ser requeridos para resolver este problema [COUDOUX (1997)], aumentando a complexidade do sistema de

compressão.

Finalmente, devemos discutir um ponto essencial de sistemas de compressão: generalidade. A imagem a ser codificada é geralmente desconhecida. E os filtros estimados por ICA são adaptados à classe de dados específica do treinamento. Dessa forma, o subespaço pode perder a generalidade e fornecer imagens reconstruídas “pobres”, para uma outra classe.

Entretanto, apesar do fato de que imagens de face foram usadas para treinamento, podemos observar claramente nas imagens reconstruídas que os blocos que não fazem parte das faces, foram reconstruídos tão bem como ou ainda melhor do que os que são, como mostrada na Figura 5.9. Nós atribuímos este comportamento à relação entre a complexidade (quantidade de detalhes) da classe do treinamento e a complexidade da imagem a ser codificada. Por exemplo, na Figura 5.7 os blocos do fundo atrás da Lena que têm menos detalhes, receberam menos distorções perceptíveis do que qualquer outro bloco. Ou seja, nossos resultados sugerem que qualquer outra classe de imagem possa ser projetada nos subespaços apreendidos a partir de classes com complexidade mais elevada.

6 Conclusões e Comentários Finais

Nós propusemos um sistema de compressão de imagem baseado no conceito de codificação eficiente. O método de codificação foi baseado em transformação e quantização. O nosso sistema compreende duas fases: aprendizagem e projeção. Na fase de aprendizagem, nós usamos a análise componentes independentes (ICA), para aprender um subespaço que maximiza a informação dos coeficientes de representação da imagem. Na fase de projeção, encontramos o código que é a projeção da imagem no subespaço. A projeção foi realizada através da minimização do erro médio quadrado entre a imagem original e reconstruída. O sistema foi comparado com as transformações PCA e DCT utilizada no JPEG através das medidas objetivas e a inspeção visual.

A análise dos resultados mostrou que nosso modelo é mais robusto às distorções perceptíveis para seres humanos para uma mesma taxa de compressão.

A partir do que foi proposto, podemos destacar os seguintes pontos de expansão que consideramos serem contribuições originais para trabalhos futuros.

- Implementar um padrão de compressão através do uso de codificadores sem perda como Huffman e codificação entrópica.
- Implementar o sistema proposto em outros tipos de sinais. Por exemplo compressão de ECG.
- Estudo da seleção de busca das funções base para classificações de lesões em mamografia.

Este trabalho deu origem ao artigo aceito na 15ª Conferência de Análise de Imagens Escandinava (Scandinavian Conference on Image Analysis - 2007), sob o título “Image Compression by Efficient Coding”.

Referências

- [BARROS (2002)] BARROS, A. K, Chichocki, A.: Neural Coding by Redundancy Reduction and Correlation, Proc. of the VII Brazilian Symposium on Neural Networks (SBRN), (IEEE), 2002.
- [BASE (1998)] Martinez, A.M.; Benavente, R.: The AR Face Database. CVC Technical Report, Vol. 24, 1998.
- [BARLOW (1961)] Barlow, H. Possible principles underlying the transformation of sensory messages in Sensory Communication. MIT Press, 1961.
- [CCIR. (1982)] CCIR. Transmission performance of television circuits designed for use in international connections, pl-38. Technical Report 567-1, CCIR, Geneva, 1982.
- [COUDOUX (1997)] COUDOUX, F. X.; Gazalet, M. G.; Corlay, P.; Rouvaen, J. M.: A Perceptual Approach to the Reduction of Blocking Effect in DCT-Coded Images, Journal of visual Communication and Image Representation. Vol. 8, No. 4, pp. 327-337, 1997.
- [CASIMIRO (2004)] CASIMIRO, Charles Cavalcante. Sobre Separação Cega de Fontes: Proposições e Análise de Estratégias para Processamento Multi-Usuário. Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação Departamento de Comunicações, Tese de Doutorado, Abril de 2004.
- [DONY (1995)] DONY, R. D., Haykin, S., Proc. of IEEE Neural Network Approaches to Image Compression, vol. 83, NO. 2, pp 288-303, 1995.
- [FERREIRA (2006)] FERREIRA, A. J, Figueirodo, M. A. T.: On the use of independent component analysis for image compression, Signal Procession: Image Communication, Vol. 21, pp. 378-389, 2006
- [WALLACE(1991)] G. K. WALLACE.: The JPEG still-picture compression standard, Commun. ACM. vol.34, pp. 3044, 1991.
- [GIBBS (1898)] FOURIER SERIES. Nature, vol. 59, 1898.
- [GONZALES, WOODS e EDDINS (2004)] GONZALES, R. C.;WOODS, R. E.;EDDINS S. L. Digital Image Processing. Ed. Pearson Prentice Hall, 2004.
- [GONZALES e WOODS (2000)] GONZALES, R. C.;WOODS, R. E. Processamento de Imagens Digitais. Editora Edgard Blucher, 2000.
- [GOULART (2004)] GOULART, Humberto R., Codificação de Imagem Usando Transformada Cosseno Discreta. Tese Mestrado, Universidade Federal do Mato Grosso do Sul, 25 de fevereiro de 2004.
- [HYVÄRINEN & OJA (2000)] HYVÄRINEN, Aapo; OJA, Erkki. Independent Component Analysis: Algorithms and Applications. Neural Networks, 2000.
- [HYVÄRINEN *et al* (2001)] HYVÄRINEN, Aapo; KARHUNEN, Juha; OJA, Erkki. Independent Component Analysis. Nova York: John Wiley & Sons. 2001. 481 p.
- [HAYKIN (2001)] HAYKIN, Simon. Communication Systems. 4.ed. Nova York: John Wiley & Sons. 2001. 816 p.
- [HERAULT *et al* (1985)] HERAULT, J., JUTTEN, C., & ANS, B. Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In Actes du Xème Colloque GRETSI(p. 1017-1022). Nice, France. Citado nas pág(s): 4, 16, 54.

- [HAYKIN (1996)] HAYKIN, Simon. Adaptive Filter Theory. 3.ed. Englewood Cliffs: Prentice-Hall. 1996. 989 p.
- [HUFFMAN (1952)] HUFFMAN, D.A.: A method for the construction of minimum-redundancy codes, Proceedings of the I.R.E., 1952 pp 1098-1102.
- [INTERNATIONAL T. U. (1998)] INTERNATIONAL TELECOMMUNICATION UNION, Methods for the subjective assessment of the quality of television pictures. Technical Reprot 500-7, ITU-R, 1998.
- [JAIN (1989)] A. Jain. Fundamentals of Digital Image Processing. Prentice Hall, New Jersey, 3rd edition, 1989.
- [JAMES (1974)] J. L. Mannos, D. J. Sakrison.: The Effects of a Visual Fidelity Criterion on the Encoding of images, IEEE Trans. on information Theory, vol. it-20, pp 525-536, 1974.
- [KORTAMN (1967)] C. M. Kortamn.: Redumdancy reduction-A Practical Method of Data Compression, Proc. of The IEEE Vol. 55, no. 3, pp. 223-226, 1967.
- [LEE (1998)] LEE, T-W., Independent Component Analysis Theory and Applications, Kluwer, 1998.
- [LEE (1999)] LEE, T-W., Girolami, M. & Sejnowski, T. J. Independent Component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources. Neural Computation, 11(2): pp 417-441, 1999.
- [LEWICK (2002)] M.S. Efficient coding of natural sounds. Nature: Neuroscience, pp. 356-363, 2002.
- [LEITE (2006)] LEITE, Luciana, Análise de Componentes independentes aplicada à Identificação de Regiões Lesionadas em Mamogramas. Tese Mestrado, Universidade Federal do Rio de Janeiro, Abril de 2005.
- [LLOYD (1982)] S. Lloyd. Least squares quantization in PCM, IEEE Trans. on Information Theory, Vol IT-28, March, 1982 pp 127-135.
- [MCKEOWN (1998)] MCKEOWN, M., Makeig, S., Brown, S., Jung, T. P., Kindermann S., Bell, A. J., Iragui, V. & Sejnowski, T. Blind separation of functional magnetic resonance imaging (Fmri) data. Human Brain Mapping, 6(5-6):368-372, 1998.
- [MAX (1960)] MAX, J. Quantizing for Minimum Distortion, IRE Transactions on Information Theory, Vol IT-6, March, 1960 pp 7-12.
- [M. KENDALL (1975)] M. KENDALL. Multivariate Analysis. Charles Griffin, 1975.
- [M. MIYAHARA (1998)] M. Miyahara, K. Kotani and V. Algazi. Objective picture quality scale (PQS) for image coding. IEEE Trans. on communications, 46(9): 1215-1226, Setember 1998.
- [N. AHMED (1974)] N. AHMED, T. Natarajan, K. R. Rao.: Discrete Cosine Transform, IEEE Trans. Computer, pp 3044 1991.
- [NELSON (1995)] M. Nelson, J. L. Gailly, The Data Compression Book. M & T Books, New York, 2nd edition, November 1995.
- [PAPOULIS & PILLAI (2002)] PAPOULIS, Athanasios; PILLAI, S. Unnikrishna. Probability, Random Variables and Stochastic Processes. 4 e.d. Nova York: McGraw-Hill. 2002. 852 p.
- [PAJUNEN (1989)] P. PAJUNEN. Extensions of Linear Independent Component Analysis: Neural and information-Theoretic Methods. PhD Thesis, Helsinki University of Techology, Helsinki, Finland, May 1998.
- [PRINCIPE et al (2000)] PRINCIPE, José C.; EULIANO, Neil R.; LEFEBVRE, W. Curt. Neural and Adaptive Systems. Nova York:John Wiley & Sons. 2000. 672 p.

- [ROBERT (2006)] ROBERT, D. Rodrigues Guilhaon, Compressão e Sinais de Eletrocardiograma Utilizando Análise de Componentes Independentes. Tese Mestrado, Universidade Federal do Maranhão, 2006.
- [SIMONCELLI (2001)] SIMONCELLI, E. P., Olshausen B. A.: Natural Image Statistics and Neural Representation, Annu. Rev. Neurosci. pp. 1193-216 1995.
- [S. COSTA (2001)] S. COSTA, S. Fiori.: Image Compression Using Principal Component Neural Networks. Image and Vision Computing, Vol. 19, pp. 649-668, 2001.
- [STRIJM (1991)] STRIJM, J., Cosman, P.C.: Medical image compression with lossless regions of interest. Signal Processing, Vol. 59, pp. 155-171, 1991.
- [S. GRAPS. (2001)] S. Grgic, M. Mrak, and M. Grgic. Comparison of JPEG image coders. In Proc. of the 3rd international symposium on Video Processing and Multimedia Communications (VIPromCom-2001), pages 79-85, Zadar Croatia, 2001.
- [SALOMON (2004)] SALOMON, David. Data Compression. Nova York: Springer-Verlag. 2004. 289 p.
- [VITERBI & OMURA (1979)] VITERBI, Andrew J.; OMURA, Jim K. Principles of Digital Communication and Coding. Tokyo: McGraw-Hill. 1979 560 p.
- [WIDROW & STEARNS (1985)] WIDROW, Bernard; STEARNS, Samuel D. Adaptive Signal Processing. Nova Jersey: Prentice-Hall. 1985. 528 p.