COPPE/UFRJ

RWISARD: UM MODELO DE REDE NEURAL SEM PESO PARA RECONHECIMENTO E CLASSIFICAÇÃO DE IMAGENS EM ESCALA DE CINZA

Leandro Almeida de Araújo

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientadores: Ricardo de Cordeiro Farias
Felipe Maia Galvão França

Rio de Janeiro
Maio de 2011

RWISARD: UM MODELO DE REDE NEURAL SEM PESO PARA RECONHECIMENTO E CLASSIFICAÇÃO DE IMAGENS EM ESCALA DE CINZA

Leandro Almeida de Araújo

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Examinada por:

Prof. Ricardo de CordeiroFarias, Ph.D.

Prof. Felipe Maia Galvão França, Ph.D.

Prof. Alvaro Luiz Gayoso Azeredo Coutinho, D.Sc.

Prof. Webe João Mansur, Ph.D.

Prof. Paulo Batista Gonçalves, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
MAIO DE 2011

Araújo, Leandro Almeida de

RWiSARD: Um Modelo de Rede Neural Sem Peso para Reconhecimento e Classificação de Imagens em Escala de Cinza/Leandro Almeida de Araújo. – Rio de Janeiro: UFRJ/COPPE, 2011.

XI, 63 p.: il.; 29, 7cm.

Orientadores: Ricardo de Cordeiro Farias
Felipe Maia Galvão França

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2011.

Referências Bibliográficas: p. 62 – 63.

1. Redes Neurais. 2. Reconhecimento de Imagens. 3. Reconhecimento Facial. 4. Redes Neurais Sem Peso. I. Farias, Ricardo de Cordeiro *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

Se você ficar pensando muito
sobre como fazer alguma coisa,
você jamais irá realizar essa
coisa.
Bruce Lee

Agradecimentos

Agradeço a toda a minha família, em especial meus pais, que por toda a minha vida nunca faltaram com o suporte. Eu não poderia ter tido melhores. Não importa onde quer que eu termine, tenho certeza de que, graças a eles, ao menos eu tive um bom começo.

Ao meu professor, orientador e amigo Ricardo Farias, que mesmo antes de me enveredar por essa empreitada já vinha me apoiando. Sem seu o apoio e orientação, nada disso seria possível. Ao professor Felipe França, cujas opiniões, sugestões e críticas foram fundamentais para dar a essa pesquisa o rumo certo. Não poderia ter chegado em melhor hora.

Aos meus amigos do Laboratório de Computação Gráfica do PESC/COPPE, com quem tive a honra de partilhar da mesma sina. A sua participação nessa pesquisa está em todo o apoio e em todo o aconselhamento que me ajudaram a levar esse trabalho a diante. É muito bom saber que fiz parte desse time.

E por fim, à minha garota, Lúcia Helena, a quem eu amo muito, cujo incentivo me serviu de inspiração para continuar, e cuja paciência lhe serviu de inspiração para continuar comigo.

À todos, muito obrigado.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

RWISARD: UM MODELO DE REDE NEURAL SEM PESO PARA RECONHECIMENTO E CLASSIFICAÇÃO DE IMAGENS EM ESCALA DE CINZA

Leandro Almeida de Araújo

Maio/2011

Orientadores: Ricardo de Cordeiro Farias
Felipe Maia Galvão França

Programa: Engenharia de Sistemas e Computação

As histórias dos campos da Visão Computacional e das Redes Neurais tem muito em comum, uma vez que o desenvolvimento de um eventualmente implica no aprimoramento do outro. As Redes Neurais Sem Peso, também conhecidas como Redes Neurais Baseadas em RAM, em especial, tem sido desenvolvidas por todos esses anos quase que exclusivamente visando aplicações na área de Reconhecimento de Imagens. Mesmo assim, estas aplicações tem sofrido com a falta de suporte apropriado para imagens compostas de pixels em escala de cinza. Algumas tentativas de se prover o modelo de tal capacidade tem sido feitas, embora geralmente com pouco ou nenhum sucesso. O modelo apresentado neste trabalho consegue oferecer , de forma eficaz, a capacidade para reconhecimento de imagens em escala de cinza, ao mesmo tempo que mantém a mesma eficiência computacional que fez das Redes Neurais Sem Peso um modelo muito bem sucedido.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

RWISARD: A WEIGHTLESS NEURAL NETWORK MODEL FOR GREYSCALE IMAGE RECOGNITION AND CLASSIFICATION

Leandro Almeida de Araújo

May/2011

Advisors: Ricardo de Cordeiro Farias
Felipe Maia Galvão França

Department: Systems Engineering and Computer Science

The histories of the fields of Computer Vision and Neural Networks share a lot in common, as the development of one eventually incurred in further improvement of the other. Weightless, or RAM based neural networks, in particular, have been developed for all these years almost solely due to their applications in the Image Recognition area. Yet, these applications have suffered a lack of support for images composed by greyscale pixels. Some attempts have been made in order to provide the model with such capability, but with limited or no success, so far. The model presented in this work manages to provide effective greyscale image recognition functionality, at same time achieving the same computational efficiency that made weightless neural networks a well succeeded model.

Sumário

Lista de Figuras

Capítulo 1

Introdução

1.1 Motivação

Desde o conceito foi proposto pela primeira vez, a cerca de seis décadas atrás, a comunidade científica e a Indústria tem testemunhado o surgimento de um número vasto de modelos de redes neurais e seus derivados. Tendo como foco situações onde a ausência de um modelo matemático consistente impõe limites óbvios para soluções computacionais, as redes neurais representam uma alternativa muito bem sucedida na elucidação de problemas onde habilidades similares à cognição, à percepção e ao aprendizado humanos se fazem necessárias.

Naturalmente, o campo da Visão Computacional é repleto de tais situações, uma vez que o ramo como um todo é, em si, uma tentativa de emular um sentido inerentemente humano e de trazer ao Computador a capacidade de tomar decisões e colher informação a partir da visão. Dessa forma, não é de se surpreender que a história relativamente recente da evolução de ambas as áreas chega muitas vezes a se confundir dada a grande gama de problemas e aplicações das quais compartilham entre si.

Aspectos físicos tais como incidência da luz, a geometria e a textura das superfícies retratadas nas imagens, a enorme variabilidade de formas e de contornos capturados, a própria maneira com os objetos e formas desviam ou inibem a luz, traduzem-se na imagem capturada pela máquina em variações de maior ou menor intensidade na luminância e no tom de seus pixels, a enriquecendo de valiosas informações acerca dos artefatos observados.

O emprego de redes neurais no processamento de imagens cujos pixels assumem diversos valores dentro da escala de cinza tem constituído ao longo de todos estes anos em um grande desafio para a área, haja vista a extensa cobertura da literatura para esse assunto. Diversos modelos de redes neurais e técnicas de pré-processamento e extração de características foram desenvolvidos na tentativa de se tornar as redes

neurais tão eficazes no tratamento de imagens em escalas de cinza quanto o são quando utilizadas com padrões de caráter binário.

Neste trabalho é apresentado um novo modelo de rede neural sem peso que procura preencher algumas lacunas deixadas em aberto no emprego dessa categoria de rede no problema de se classificar imagens em escala de cinza.

1.2 Objetivos

Redes neurais sem peso, também conhecidas como Redes Neurais Baseadas em RAM, são modelos extremamente eficazes e de custo operacional razoavelmente baixo. Dada a natureza do seu funcionamento, apresentam, por outro lado, a desvantagem de operar com versões binarizadas das imagens de interesse, o que em alguns casos pode implicar em uma menor eficácia nos processos de reconhecimento e detecção de padrões devido à perda de informação.

Esse texto apresenta um modelo que estende o funcionamento das redes baseadas em RAM, bem como também suporta com evidências experimentais as vantagens que seu emprego traz para processos de visão computacional que envolvem imagens em escala de cinza.

1.2.1 reconhecimento de padrões

Reconhecimento de padrões compreende todos os processos envolvidos na tentativa de se associar os elementos de um dado conjunto, com um segundo conjunto de rótulos, lhes conferindo uma identidade como classe. Tal atividade pode consistir em relacionar dados de uma amostra a um conjunto previamente definido de classes - classificação - ou em encontrar uma conveniente partição dessa amostra em grupos de elementos com características em comum - clusterização.

Quando artefatos do mundo real tem sua imagem capturada por meios físicos, diversas características do ambiente tais com relação a iluminação, presença de outros artefatos em cena ou características atmosféricas podem afetar a representação desses objetos de forma a dificultar ou até a comprometer a possibilidade de se extrair informação da imagem, para que tal informação seja utilizada pelos processos de reconhecimento de padrões.

O modelo proposto oferece uma oportunidade de se trazer os efeitos dessas influências na imagem para dentro da operação da rede neural, tornando possível se estabelecer estratégias mais consistentes para extração de características dessas imagens.

1.2.2 reconhecimento de faces

Dado o enorme interesse por sistemas de identificação automática de pessoas, assim como a igualmente grande complexidade dos problemas encontrados na implementação destes sistemas, Reconhecimento Facial se tornou um fétil nicho de em seu próprio direito. Ao longo dos anos, diversas abordagens foram apresentadas, com maior ou menor sucesso, mas de um modo geral diversos problemas persistem e o campo ainda se mostra ávido por inovações.

Dentre os motivos que tornam esse problema particularmente difícil, está a sua Não-Linearidade. Toda abordagem de cunho matemático que procura elucidar esse problema via análise subespacial se vê face a imensa dimensionalidade dos espaços envolvidos. Mesmo tentar distinguir subespaços que representem rostos humanos representa um grande problema. Processos baseados em PCA e decomposição espectral (TURK [1]) reduzem tremendamente o problema para espaços de dimensões menores, mas ainda assim faces representam variedades topológicas de caráter ainda fortemente não-linear e não-convexo, o que ainda torna em grande parte inviável a aplicação de métodos analíticos na sua solução.

Ainda quanto a capacidade de distinguir e identificar diferentes indivíduos através da imagem de seus rostos, existe uma dificuldade inerente que surge do fato de que variações acarretadas por mudança do ângulo do observador com relação ao rosto, ou mesmo mudanças da iluminação incidente, podem ser ainda maiores do que as variações causadas na imagem quando se troca o rosto de um indivíduo por outro, como visto em MOSES [2]. Em outras palavras, um rosto é mais parecido com o rosto de outra pessoa quando preservados o ângulo de visão e a iluminação, do que é parecido com ele mesmo quando se alteram esses parâmetros.

O modelo aqui proposto procura oferecer uma maior resiliênica na abordagem desses problemas trazendo alguma habilidade adicional aos sistemas vigentes para se avaliar a semelhança entre imagens.

1.3 Contribuições Relevantes

Além da proposta de um novo modelo de Rede Neural Sem Peso capaz de assimilar e classificar imagens em escala de cinza sem a perda de informação causada pela binarização, está entre as contribuições relevantes desse trabalho a proposta de uma métrica para se aferir similaridade entre imagens, baseada no funcionamento desse novo modelo de rede neural, bem como um estudo sobre possíveis novos métodos de detecção e alinhamento de faces em imagens baseados nesse novo modelo. Cada uma dessas contribuições são amparadas por evidências levantadas experimentalmente.

Capítulo 2

Trabalhos anteriores

2.1 Redes Neurais

2.1.1 Neurônios reais X Neurônios artificiais

Qualquer célula em um ser vivo é capaz de provocar ou responder a estímulos químicos. Em seres pluricelulares, esses estímulos químicos, ou sinais químicos, são os responsáveis pela comunicação intercelular que permite ao organismo realizar as mais diversas funções em variados graus de complexidade, como por exemplo a respiração e o metabolismo, muitas vezes envolvendo mais de um tecido vivo no processo. As células neuronais, ou neurônios, como normalmente são chamadas, compõem toda uma classe de células que se especializou na comunicação intercelular e que desempenha essa tarefa num patamar de sofisticação muito além de qualquer comparação com qualquer outro tipo de célula, sendo o mais importante dos elementos formadores do Sistema Nervoso. Um exemplo de neurônio pode ser visto na figura 2.1.a. Neurônios são capazes de transmitir sinais de natureza eletroquímica em altas velocidades e a relativamente grandes distâncias, podendo cobrir toda a extensão do organismo. São também capazes de se organizar em redes extremamente intricadas afim de processar informações e controlar processos extremamente elaborados. A maneira como um novo dado é processado, a resposta a ser dada mediante um estímulo ou a maneira como uma informação é armazenada na rede neural depende não só da natureza do funcionamento de cada neurônio individualmente, mas também da forma como essa rede está estruturada e de como os elos entre os seus neurônios são estabelecidos.

Tais elos, chamados de Sinapses, são estabelecidos através dos dentritos e do axônio dos neurônios 2.1.b. Os dentritos são o principal sítio para os terminais sinápticos oriundos do axônio de outro neurônio. Seu axônio, por sua vez, é o responsável por levar o sinal eletroquímico até a próxima conexão sináptica através de um mecanismo chamado Potencial de Ação (PURVES [3]). Os sinais são então

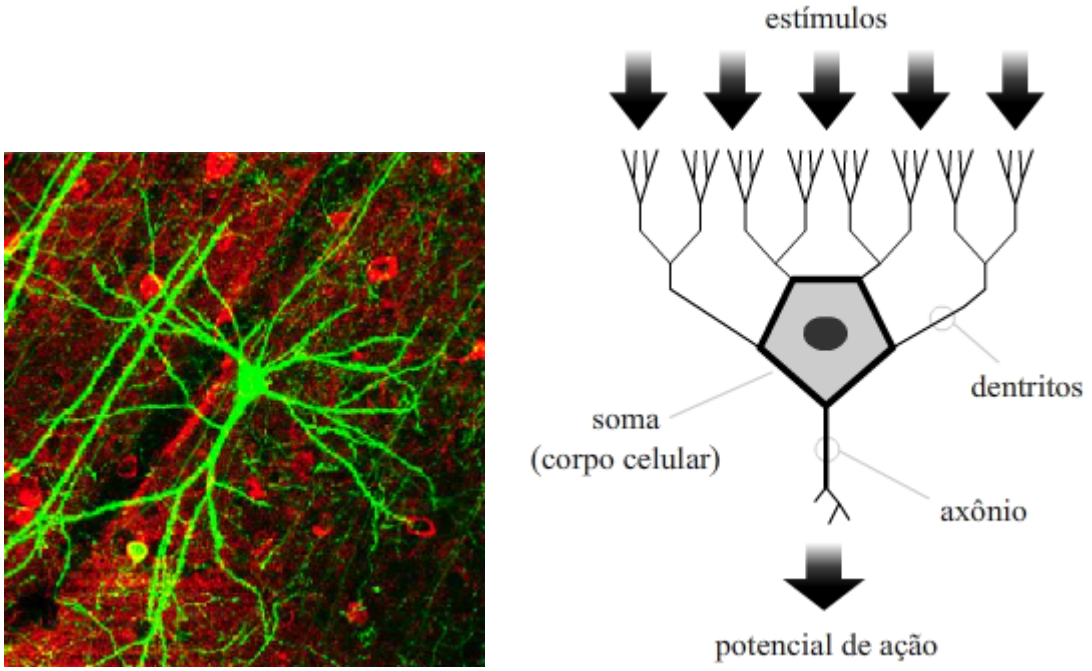

Figura 2.1: a) Neurônio; b) Diagrama esquemático do Neurônio biológico;

transmitidos do axônio de um neurônio até o dentrito de outra, ou até uma outra célula-alvo, através dos neurotransmissores. As células neuronais apresentam dentritos em diferentes quantidades e arranjos, podendo variar desde absolutamente nenhum dentrito até algumas dezenas de milhares de dentritos por célula, dependendo do papel que esta célula desempenha no sistema nervoso. Ao longo de todo ciclo de vida de um ser vivo dotado de sistema nervoso, novas sinapses são estabelecidas o tempo todo, mudando a estrutura do sistema nervoso. A capacidade de assimilar um novo dado, ou de prover uma nova resposta a um determinado estímulo, vem necessariamente dessa notável plasticidade dos circuitos neurais. Axônios e dentritos crescem mediante estímulos químicos capazes de determinar quais serão as novas sinapses a serem formadas de forma a se obter maior eficiência sináptica. Tal mecanismo confere capacidade de aprendizado e memória à rede neural biológica.

A idéia de se simular o comportamento dos circuitos neurais presentes nos seres vivos através de constructos computacionais constitui um dos fundamentos da escola conectivista, no campo da Inteligência Artificial, e foi proposta pela primeira vez por McCulloch e Pitts [4], através de um modelo matemático extremamente simplificado do neurônio. Assim como um neurônio real, o neurônio artificial tem seu comportamento modificado mediante a apresentação de estímulos bem definidos, o que o força a responder de maneira apropriada quando posteriormemente for apresentado a estímulos semelhantes. Em outras palavras, ao ser treinado, o neurônio artificial é capaz de aprender como ele deve responder a determinados estímulos.

Este modelo de neurônio, hoje chamado de Neurônio de McCulloch-Pitts 2.2, é

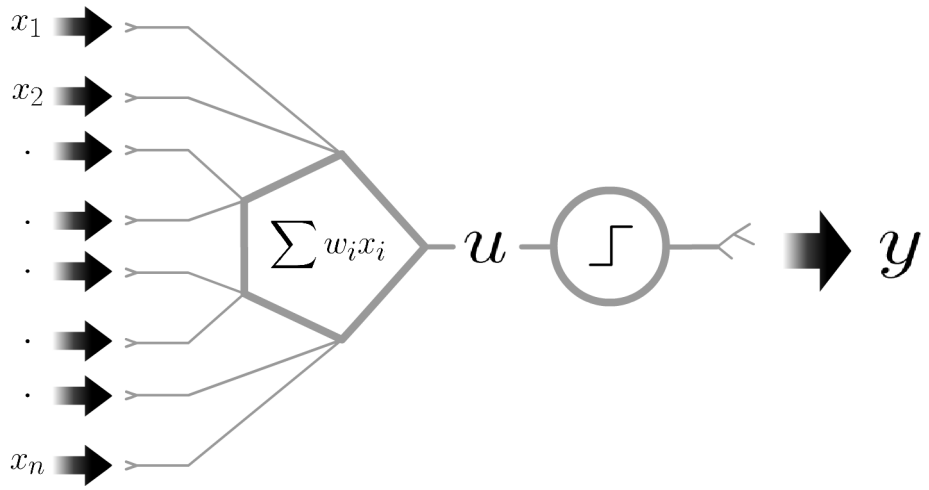

Figura 2.2: Diagrama esquemático do Neurônio de McCulloch-Pitts.

o protótipo das redes neurais baseadas em pesos. O processo de aprendizado neste neurônio procura emular a habilidade das células neuronais de se reestruturarem para acomodar um novo aprendizado mediante o aumento da eficiência sináptica. Formalmente, este neurônio emite um sinal binário y mediante a entrada de um padrão que consiste em um vetor de sinais representados por valores escalares $x = \{x_1, \cdots, x_n\}$. O sinal y é determinado pela seguinte função:

$$y = \begin{cases} 1 & \text{se } u \geq 0 \\ 0 & \text{se } u < 0 \end{cases}$$

conhecida como Função de Ativação, onde

$$u = \sum_{i=1}^{n} w_i x_i$$

O vetor de escalares $w = \{w_1, \cdots, w_n\}$ consiste nos pesos que determinam como a rede deve responder mediante a apresentaçãos dos valores de x. O treinamento dessa rede é realizado se ajustando os pesos w através de um método de otimização baseado em gradiente descendente, até que a rede emita o sinal esperado para o padrão apresentado.

2.1.2 classificação das redes neurais artificiais

Dependendo da natureza da aplicação, o modelo de rede neural a ser empregado deverá se encaixar em uma dentre duas categorias distintas. Em situações onde as características que determinam a que classe um dado padrão não sejam conhecidas a priori, ou onde se procura se conhecer como os padrões se organizam em partições, ou em termos mais gerais, quando não se pode afirmar a princípio a que classe

um dado padrão, será empregado um modelo de rede neural de Aprendizado Não-Supervisionado. Padrões são apresentados aos neurônios dessa categoria, durante a fase de treinamento, sem qualquer informação a respeito de como esses padrões deverão ser classificados ou reconhecidos pela rede. Essa categoria de redes neurais encontra um campo fértil para aplicações nas áreas de Análise de Cluster e Data Mining. Os modelos de Mapas Auto-Organizáveis (SOM) e os baseados na Teoria da Ressonância Adaptativa (ART) constituem os exemplos mais típicos dessa categoria. AutoWiSARD [5] é um modelo derivado da rede WiSARD voltado ao Aprendizado Não-Supervisionado. Já quando se trata de situações onde as classes nas quais os padrões devem se enquadrar são bem definidas, ou quando os padrões empregados no treinamento pertençam a grupos bem definidos e se deseja que novos padrões sejam classificados conforme a sua semelhança aos elementos desses grupos, algum modelo da categoria das redes neurais de aprendizado supervisionado poderá ser utilizado. O treinamento desse tipo de rede consiste na apresentação de padrões cujas classes são bem conhecidas até que a rede neural seja capaz de identificar essas classes com satisfatório índice de acertos. Exemplos dessa classe de redes neurais são o Perceptron, ADALINE, o modelo WiSARD e seu derivado RWiSARD, que é o objeto desse trabalho.

Quanto aos processos internos relacionados ao treinamento e classificação de padrões, os modelos se dividem entre Redes Neurais com Peso e Redes Neurais Sem Peso. As redes neurais baseadas em peso procuram emular o processo de aprendizado dos neurônios biológicos através de uma simplificação formal da otimização da eficiência sináptica, que nessas células consiste na formação organizada de novas sinapses. Nas redes neurais artificiais baseadas em peso, essa otimização de eficiência sináptica é obtida por meio da atualização dos pesos de suas sinapses, a exemplo do que ocorre no modelo MLP e nas redes ART. O modelo ART3 chega mesmo ao ponto de simular os processos eletroquímicos de mudança de concentração iônica observados em células neuronais afim de se obter uma performance similar ao de seu análogo vivo. Assim como ocorre nos neurônios biológicos, onde o processo de aprendizado consome uma considerável parcela de tempo se comparado a sua rápida transmissão de sinais, modelos de redes neurais artificiais que procuram imitar o seu comportamento também exibem aprendizado relativamente lento. Modelos de redes neurais artificiais sem peso, por outro lado, apresentam capacidade para serem treinadas e de classificar padrões sem a preocupação de simular o comportamento dos neurônios biológicos. Ao invés de tentarem implementar algum modelo matemático que emule o funcionamento interno dos neurônios vivos, processam e registram os próprios padrões utilizados em seu treinamento em um banco de memória de forma a empregar a informação oriunda desses padrões no reconhecimento e classifica??ão de novos padrões. Sendo seu aprendizado consistindo quase que integralmente de aces-

sos à memória, conforme será visto mais adiante, tendem a exibir um aprendizado muito mais rápido que as redes neurais com peso.

2.1.3 O Método N-tuple

O método N-tuple (N-uplas) é o fundamento por trás de todos os modelos de redes neurais baseadas em RAM. Apresentado pela primeira vez por Bledsoe e Browning [6], propõe que o problema de aprendizado e reconhecimento de uma imagem seja formulado através de funções lógicas. A imagem é particionada em grupamentos ordenados de tamanho fixo de pixels - a Upla, propriamente dita, e o aprendizado se dá ao se construir fórmulas lógicas conjuntivas para cada uma das uplas da imagem, de maneira que essas fórmulas, aplicadas aos pixels da upla, retornem verdadeiro para cada uma das uplas. O exemplo a seguir, extraído de [7] ilustra o processo. A figura 2.3.b apresenta o mapa de bits para a letra "I", enquanto que na figura 2.3.c se vê o mapa de bits para a letra "T". A figura 2.3.a apresenta as variáveis das funções lógicas associadas aos pixels da imagem.

A	B	C
D	E	F
G	H	I

1	1	1
0	1	0
1	1	1

1	1	1
0	1	0
0	1	0

Figura 2.3: a) Mapeamento das variáveis; b) Binarização da letra "I"; c) Binarização da letra "T";

Neste exemplo, o Classificador N-tuple emprega um tamanho fixo de 3 pixels por upla (3-tuple). Assim, ao ser treinado com o padrão referente à letra "I", ele assimila a seguinte fórmula:

$$R_I = A \cdot B \cdot C + \overline{D} \cdot E \cdot \overline{F} + G \cdot H \cdot I$$

Já o aprendizado da letra "T"se dá pela construção da fórmula:

$$R_T = A \cdot B \cdot C + \overline{D} \cdot E \cdot \overline{F} + \overline{G} \cdot H \cdot \overline{I}$$

A ser assimilada por um segundo classificador responsável pela classe dos caracteres "T". Uma vez que o classificador N-tuple tenha sido treinado, as fórmulas por ele assimiladas podem então ser aplicadas no reconhecimento de imagens de caracteres para atribuir valores numéricos inteiros a cada padrão apresentado, valores esses que serão comparados com os atribuidos pelos outros classificadores afim de se

decidir a qual classe pertence o padrão. No próximo exemplo, as fórmulas assimiladas para as classes "T"e "I"serão utilizadas para se decidir a qual classe pertence o padrão exibido na figura 2.4.

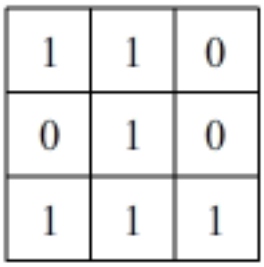

Figura 2.4: Padrão a ser classificado.

Mantendo-se o mesmo mapeamento utilizado no treinamento, as fórmulas lógicas de ambos classificadores são aplicadas aos pixels da imagem apresentada. Assim, para o classificador associado à classe "I"se obtém:

$$R_I = 1 \cdot 1 \cdot 0 + \overline{0} \cdot 1 \cdot \overline{0} + 1 \cdot 1 \cdot 1 = 2$$

Ao passo que para o classificador associado à classe "T", a respectiva fórmula resulta em:

$$R_T = 1 \cdot 1 \cdot 0 + \overline{0} \cdot 1 \cdot \overline{0} + \overline{1} \cdot 1 \cdot \overline{1} = 1$$

Os valores retornados pelas fórmulas presentes nos classificadores são então comparados entre si. A classe vencedora é então aquela cuja equação associada retorna o maior valor dentre todos. No exemplo apresentado, o maior valor é $R_I = 2$, o que significa que o padrão em questão pertence à classe "I".

2.2 A WiSARD

À época em que o método n-tuple foi desenvolvido, a tarefa de registrar as operações conjuntivas extraídas das diversas uplas da imagem, durante o treinamento, e depois resgatar essas fórmulas de maneira rápida e eficiente ao se classificar uma outra imagem representava um entrave tecnológico à aplicação do método em situações reais. Somente mais tarde, nos anos 80, Wllkie, Stoneham e Aleksander apresentaram uma solução ao mesmo tempo simples, e extremamente elegante, que não apenas viabilizou as aplicações práticas desse método como também permitiu que ele pudesse ser implementado a partir de componentes de baixo custo em um dispositivo capaz de operar em tempo real. Essa solução consiste em empregar um decodificador 1-para-n para que a partir de um código binário que representasse a n-upla em questão,

se escolha dentre n caminhos possíveis aquele que represente a fórmula conjuntiva relativa a esta upla. O diagrama na figura 2.5 ilustra esse conceito.

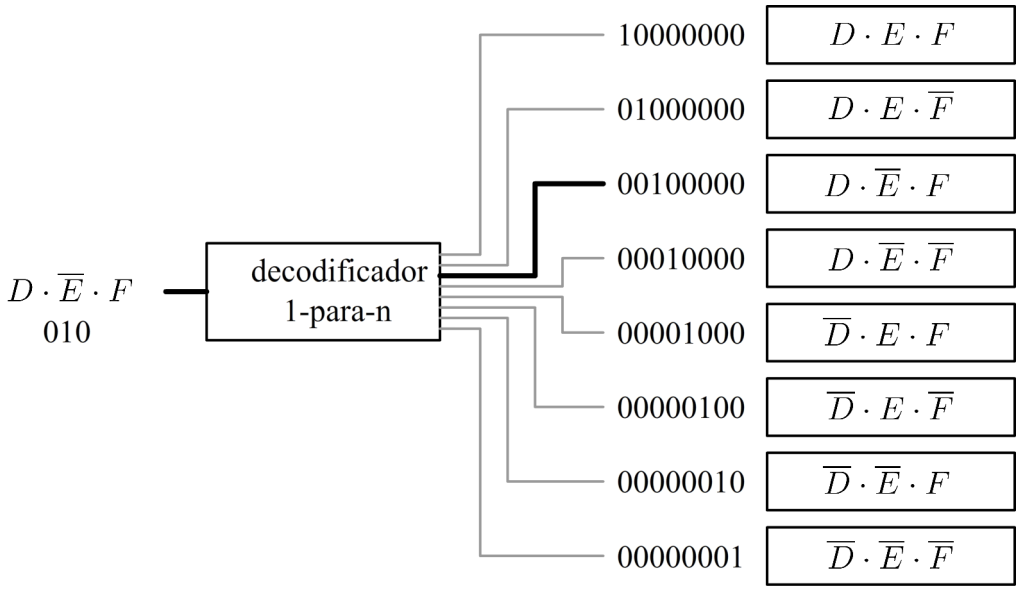

Figura 2.5: O estágio de decodificação de Wllkie, Stoneham e Aleksander.

Assim, ao invés de se registrar e depois procurar verificar a existência desta fórmula na memória do neurônio, todas as fórmulas conjuntivas possíveis, para esta upla, estariam representadas no neurônio e poderiam ser instantaneamente acessadas assim que a upla fosse decodificada. Bastaria ao sistema verificar se a fórmula conjuntiva identificada pela upla foi assimilada pelo neurônio durante o treinamento ou não. Esse registro é feito através de um sinal de um bit associado a cada uma dessas fórmulas. Todo esse conceito é naturalmente implementado através da utilização de memórias do tipo RAM (Random Access Memory).

2.2.1 Neurônio básico da rede WiSARD

Uma imagem, ou uma região de interesse da imagem, composta por $n_(pixels)$ é particionada em conjuntos de pixels, cada um com um número fixo n_{upla} de pixels. Essa partição é formada aleatoriamente, mas uma vez criada, é fixa por todo o ciclo de vida da rede neural. a figura 2.6 mostra um exemplo de como uma imagem é particionada em n-uplas de pixels.

Cada n-upla de pixels é associada, no neurônio, a um nó de memória RAM, onde cada endereço aponta para um valor de 1 bit apenas, como é visto na figura 2.7. Em cada neurônio presente na rede, todas as n-uplas da imagem estão representadas cada qual com seu respectivo nó RAM (figura 2.8) , sendo que cada um desses nós RAM possui um espaço de endereçamento exclusivo.

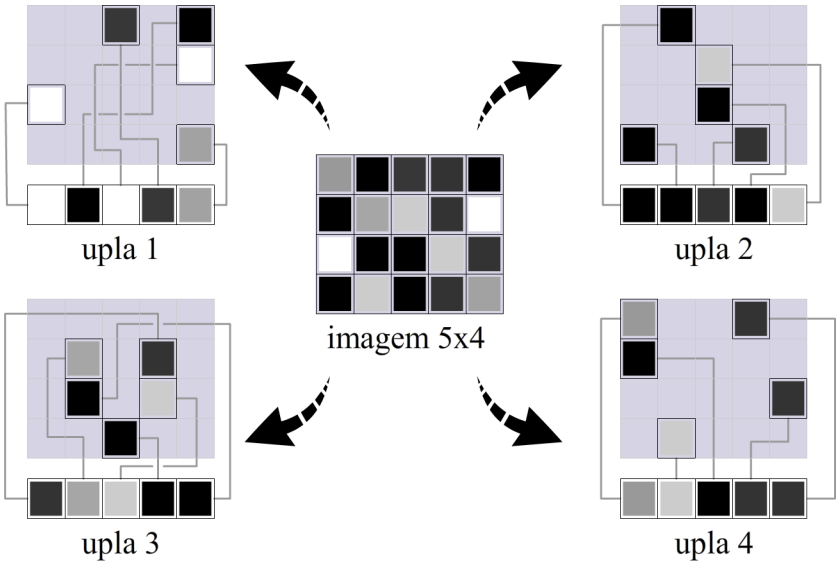

Figura 2.6: Partição dos pixels de uma imagem em n-uplas de pixels.

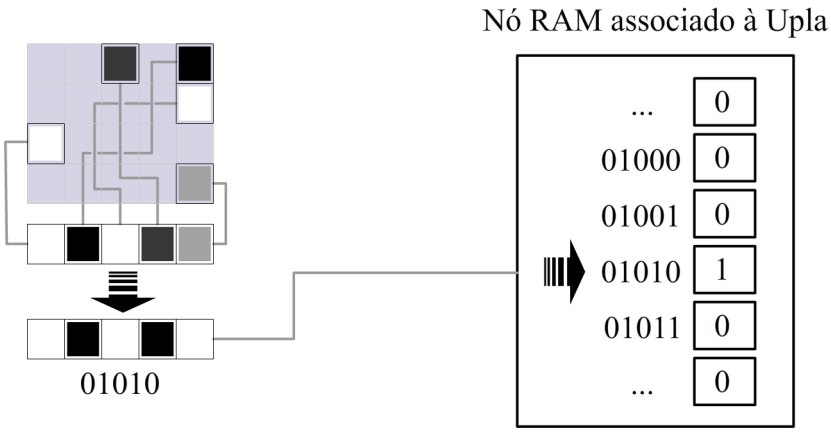

Figura 2.7: Nó RAM associado a uma n-upla.

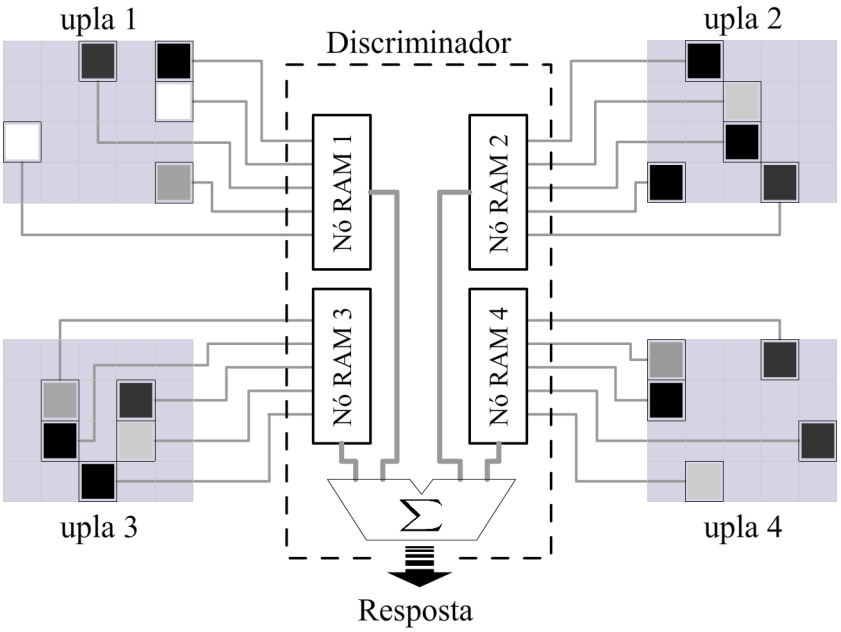

Figura 2.8: Arquitetura do neurônio WiSARD.

2.2.2 Treinamento

Um limiar é aplicado a cada n-upla de pixels convertendo essa n-upla em um vetor binário de n_{upla} bits. Esse vetor por sua vez é usado como um endereço para o nó RAM associado a essa n-upla. O valor de um dígito binário para o qual esse endereço aponta é ajustado para "1". Esse processo é executado em cada nó RAM pertencente ao neurônio que corresponde à classe a ser treinada.

2.2.3 Classificação

Endereços para nós RAM são extraídos das diversas n-uplas de pixels com compõem a partição da imagem da mesma maneira como é feito no caso do treinamento. Um contador em cada neurônio é zerado no início da operação de classificação de uma imagem e, a cada n-upla onde o endereço gerado aponta para um "1", este contador é incrementado. O valor final desse contador constitui a resposta do neurônio à imagem apresentada. Vence o neurônio que apresentar a resposta mais alta, ou seja, o neurônio cujo contador tiver a soma mais alta. Esse processo é equivalente à operação de classificação do classificador n-tuple descrito anteriormente.

2.2.4 Complexidade de tempo e de memória da rede WiSARD

Sejam n_{upla} o número de pixels com os quais a upla é definida, n_{pixels} o total de pixels na imagem, t_{bin} e t_{acesso} constantes que representam respectivamente o tempo

de binarização da upla e o tempo de acesso a algum endereço da RAM, a seguinte expressão determina o tempo gasto durante o treinamento da rede neural WiSARD com uma dada imagem:

$$\frac{n_{pixels}}{n_{upla}} \times (t_{bin} + t_{acesso}) \in O\left(n_{pixels}\right)$$

Ou seja, o treinamento tem complexidade operacional de ordem linear. Aqui, n_{upla}, apesar de variar conforme a aplicação, será sempre um valor insignificante diante de n_{pixels}, e por isso considera-se que um fator constante arbitrariamente grande possa representar um pior caso. Da mesma forma pode ser calculada a ordem da complexidade da operação de classificação de uma imagem. Aqui, $n_{classes}$ é o número de classes, ou de outra forma, o número de neurônios presentes na rede, enquanto que t_{somar} e $t_{comparar}$ são respectivamente o tempo dispensado para se somar as repostas entre duas uplas e o tempo exigido para se comparar as respostas entre dois neurônios, ambos parâmetros sendo constantes:

$$n_{classes} \times (\frac{n_{pixels}}{n_{upla}} \times (t_{bin} + t_{acesso}) + \frac{n_{pixels}}{n_{upla}} \times t_{somar}) + n_{classes} \times t_{comparar}$$

Para fins práticos, $n_{classes}$ será sempre um número de magnitude menor que n_{pixels}, qualquer que seja a natureza da aplicação. Na verdade, é razoável admitir que é possível se estipular uma constante k_1 arbitrariamente grande para que sirva de cota para um pior caso para o número de classes. A expressão assim se torna:

$$k_1 \times (k_2 \times n_{pixels} + k_3) \in O\left(n_{pixels}\right)$$

Indicando que essa operação também tem comportamento de ordem linear. Quando à demanda por memória, a complexidade por ser expressa pela seguinte fórmula:

$$n_{classes} \times (2^{n_{upla}} \times 1_{bit} \times \frac{n_{pixels}}{n_{upla}}) \in O\left(n_{pixels}\right)$$

Apesar do comportamento linear com relação a n_{pixels}, deve ser observado que o termo $2^{n_{upla}}$ cresce exponencialmente com o valor de n_{upla}, o que compromete a viabilidade desse modelo para uplas muito grandes. Experimentos práticos no entanto mostraram que uplas de no máximo 16 pixels são geralmente suficientes para qualquer aplicação.

2.3 Extensões do modelo WiSARD

2.3.1 DRASiW e as Imagens Mentais

O modelo DRASiW descrito em GRIECO [8] foi proposto com a intenção de se empregar a frequência de acesso aos endereços do Nó RAM de forma a melhorar a eficácia da rede WiSARD. Esse modelo estende a arquitetura do neurônio WiSARD acomodando um contador por endereço de memória no lugar do sinalizador de um único bit. A cada vez que um endereço de memória é acessado, durante o treinamento, esse contador é incrementado.

Como um estágio adicional, um valor mínimo de frequência é determinado de forma a se maximizar a eficácia no reconhecimento de imagens de um conjunto de controle. Durante a classificação de um padrão, um endereço aplicado a um nó RAM de uma upla só retorna "1"quando o valor apontado pelo endereço é igual ou maior do que a frequência mímina determinada.

Figura 2.9: Imagens mentais de um neurônio treinado com imagens de uma boca;

O emprego de contadores de frequência no lugar de um dígito binário permite se extrair dos neurônio uma imagem que retrata o seu conteúdo, ou melhor dizendo, a classe de padrões que esse neurônio reconhece. Essas imagens, chamadas Imagens Mentais, são construídas a partir dos valores de frequência estocados nos endereços dos nós RAM das uplas, que por sua vez são exibidos em sobreposição. A aplicação de um valor de frequência mínimo funciona como uma espécie de limiar de operação para essa rede. A figura 2.9 exibe a imagem mental de uma boca, com valores crescentes para a frequência mínima de operação.

2.3.2 Células de Minchinton

A proposta das Células de Minchinton, apresentadas em AUSTIN [7] é gerar vetores binários para os nós RAM WiSARD de forma menos sensível a variações de luminosidade entre as imagens, possivelmente contornando problemas causados pela alteração de ângulo de iluminação ou sombra sobre artefatos presentes nas imagens. Isso é feito se substituindo o processo usual de binarização pela aplicação de um limiar por um processo que leva em conta a relação entre os pixels de uma upla. Na literatura especializada são encontrados dois tipos de Células de Minchinton: o Tipo 0, onde a condição para que o bit B_i do endereço gerado seja "1"é $P_i > P_j$, sendo i e j duas posições diferentes dentro da upla, e Tipo 1, onde essa condição é

redefinida como $P_i - P_{i+1} > P_j - P_{j+1}$. Dos dois tipos mencionados, apenas o Tipo 0 teve algum relato de emprego com sucesso.

2.3.3 Extensões para Imagens em Escala de Cinza

Com o tempo, foram surgindo algumas tentativas de extensão do modelo WiSARD para tratamento de imagens em escala de cinza. Mesmo contando com uma escassa literatura, alguns exemplos podem ser citados.

Um Classificador N-tuple Contínuo é apresentado em LUCAS [9], o qual implica no neurônio armazenar os vetores com valores escalares correspondetes às uplas com pixels em escala de cinza nos nós RAM durante o treinamento, e usar uma métrica como Distância de Manhattan para avaliar a resposta das uplas, um processo que guarda semelhanças com o método abordado no presente trabalho. O método dessa forma apresenta o inconveniente de se ter que manter em memória cada padrão em escala de cinza apresentado, o que compromete a performance a cada padrão que é apresentado a upla para o treinamento. Como uma alternativa, é apresentado também apresentado um algoritmo que compila os vetores escalares assimilados para um nó RAM binário. Ainda assim, além do grande consumo de memória, esse modelo consome maior tempo durante o treinamento.

Uma outra proposta, que por sua vez não oferece uma estratégia baseada em métrica tal como a anterior e o modelo RWiSARD apresentado nesse texto, pode ser visto em [10]. Neste modelo, cada n-upla u de pixels tem seus valores ordenados em uma nova upla u_{ord}. O mapeamento entre as posições de u e u_{ord} é binariamente codificado para um endereço válido do nó RAM da upla. Como o número de estados necessários para representar todas as configurações que uma n-upla pode assumir tem ordem $O(!n)$, é apresentado ainda um algoritmo de redução de estados que, ao custo de uma sensível perda de informação, separa os valores de cinza em ρ intervalos (intervalos de limiarização) e reduz o número de estados necessários para $\rho^n - (\rho - 1)^n$, o que leva ainda a uma demanda de memória que pode tornar o uso desse método inviável em algumas aplicações prática. Isso talvez justifique a ausência de melhores relatos sobre sua eficácia.

Capítulo 3

A RWiSARD

3.1 Processamento de Imagens em Escala de Cinza

Redes neurais baseadas no método n-tuple são notoriamente eficazes e apresentam uma performance excelente em termos de velocidade de aprendizado e reconhecimento, quando comparadas a outros modelos de redes neurais. Porém, assim como a maioria dos modelos vigentes, as redes neurais sem peso, salvo algumas propostas mais recentes, tem sido restritas ao processamento de imagens binárias. Em uma grande gama de aplicações, tal imposição não chega a tornar seu emprego inviável, havendo aliás casos onde as características relevantes da imagem podem ser delimitadas por um limiar bem definido, o que torna essa restrição uma vantagem. Já em outras situações, as características pelas quais se procura reconhecer e classificar uma dada imagem podem estar implícitas às variações de tom entre os pixels da imagem, podem depender mais da diferença entre os valores dos pixels do que de seus valores absolutos, ou podem ainda essas características estar associadas a padrões em tons de cinza que não podem ser trivialmente capturado pela determinação de um limiar de binarização qualquer que seja. Nesses casos, onde a aplicação de um limiar resulta quase sempre em perda de informação relevante, o emprego de redes neurais limitadas a imagens binárias pode ser comprometido até o ponto da ineficácia. O emprego do método n-tuple de forma direta ao problema de processamento de imagens em escala de cinza levaria ao óbvio engodo de se lidar com espaços de endereçamento da ordem de q^n, quando o intervalo de valores possíveis de cinza é quantizado em q níveis.

O modelo descrito neste capítulo estende a proposta do emprego de limiares para binarização de uplas de valores escalares de forma a registrar a relação entre os pixels que compõem a upla. Como ainda se trata de um processo de binarização, possui a mesma complexidade operacional que o modelo WiSARD clássico. Utilizando o valor

dos limiares empregados no registro e na classificação de padrões, captura informação adicional que concerne à natureza da imagem em escala de cinza original. Dessa forma, este modelo se mostra ao mesmo tempo viável e eficaz no processamento de imagens em escala de cinza.

Devido a essa característica desse modelo de empregar os intervalos de valores que refletem a relação entre os pixels da upla, essa rede é chamada RWiSARD, uma abreviação para Range WiSARD (WiSARD por Intervalos).

3.1.1 Uma métrica para a similaridade entre n-uplas de níveis de cinza

O modelo WiSARD atribui uma resposta de caráter binário (igual - não igual) a uma upla como forma de se responder à seguinte pergunta: é essa n-upla de pixels similar à alguma upla, correspondente às mesmas posições dos pixels, apresentada durante o treinamento? Dentro dessa perspectiva, o valor do somatório dessas respostas, retornado pelo neurônio, pode ser encarado como uma medida da similaridade da imagem apresentada como um todo, com relação à classe de imagens empregadas no treinamento desse neurônio. O modelo RWiSARD procura estender tal métrica retornando um valor real normalizado para cada upla de pixels ao invés de apenas sinalizar se essa upla foi assimilada ou não pelo neurônio durante o treinamento. Essa nova métrica para a medição da similaridade torna possível ao modelo identificar relações entre os pixels de uma imagem cuja relevância foi ignorada pelo modelo binário clássico. Alguns resultados práticos disso serão apresentados no próximo capítulo.

3.2 Arquitetura do Neurônio RWiSARD

3.2.1 Layout da Memória da N-upla

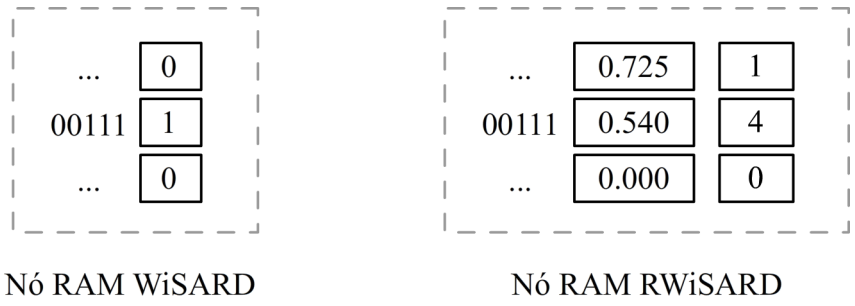

Figura 3.1: Nó de memória RAM RWiSARD comparada ao nó equivalente WiSARD

A organização do nó RAM RWiSARD difere da encontrada no modelo WiSARD

em dois pontos. O primeiro é que, ao invés de abrigar apenas um valor de 1 bit, o nó RAM abriga para cada endereço um valor escalar, e o tamanho ocupado por este valor pode variar conforme a implementação, afetando a precisão das operações da rede em troca de eficiência de memória. O dado escalar pode também variar entre um número representado em ponto flutuante ou em ponto fixo. O segundo ponto é a presença de um contador para cada endereço de memória. Dessa forma, o nó RAM da rede RWiSARD representa aumento de tamanho em um fator constante se comparado ao nó equivalente WiSARD.

3.2.2 Decomposição de N-uplas de Valores Escalares

Ao invés de um limiar único fixado para a imagem como um todo, o método adotado pela rede RWiSARD consiste em utilizar cada valor de cinza contido na upla em questão para binarizar toda a upla. O produto desse processo não é mais um único vetor binário a servir de endereço para acesso ao nó RAM da upla. O produto será um conjunto de vetores binários, um para cada elemento da upla, gerado pela limiarização dos valores da upla através do valor desse elemento. Dessa forma, quando um padrão é apresentado para essa upla quer seja para treinamento, quer seja para reconhecimento, caso a upla tenha tamanho N, N vetores binários serão gerados e assim ocorrerão N acessos ao nó RAM associado a essa upla. Os endereços produzidos pela decomposição de uma upla pertencem todos a um mesmo espaço de endereçamento.

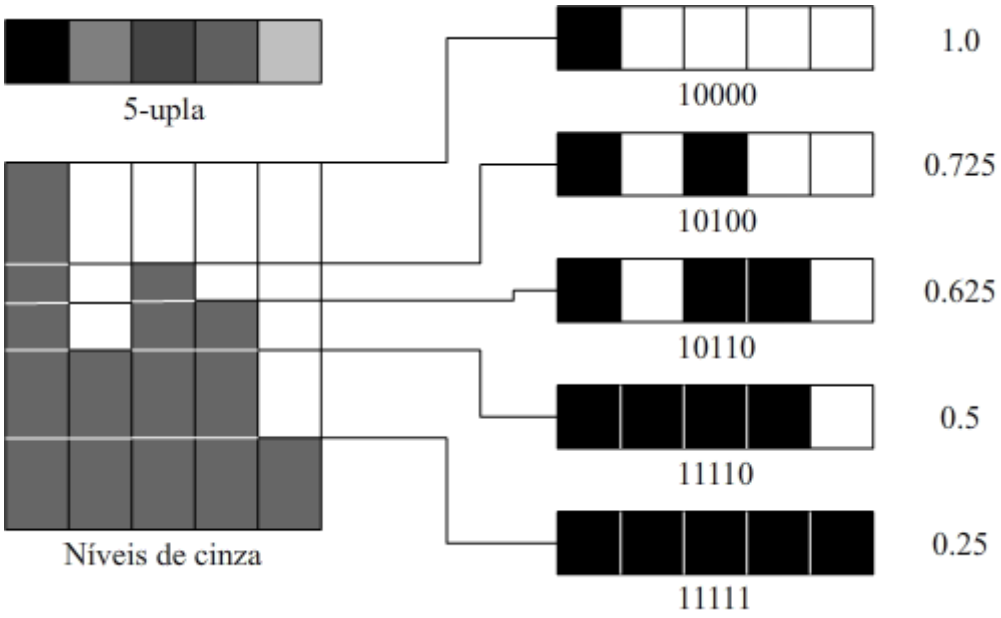

Figura 3.2: Decomposição de uma upla de pixels em escala de cinza.

A figura 3.2 ilustra o método conforme aplicado para uma 5-upla de pixels.

Formalmente, uma n-upla de valores escalares $C = \{C_i \mid i \in [0, n-1]\}$ ao ser decomposta por este método gera um conjunto de n-uplas binárias $B = \{B_i \mid i \in [0, n-1]\}$, onde $B_i = \{b_{\ i,j} \mid i, j \in [0, n-1]\}$. Cada bit $b_{\ i,j}$ é por sua vez definido como:

$$b_{\ i,j} = \begin{cases} 0 & \text{se } C_j < C_i \\ 1 & \text{se } C_j \geq C_i \end{cases}$$

Uma questão que naturalmente surge em face dessa definição, seria quanto ao que ocorre quando mais de um pixel da upla possui o mesmo valor de cinza. Este método gera n uplas binárias para qualquer upla de valores de cinza de tamanho n, mesmo que haja uplas binárias repetidas. Isso implica no fato de que durante um único ciclo de treinamento ou de reconhecimento nessa upla, podem ser realizados até n acessos a um mesmo endereço do nó RAM. Ainda que isso pareça uma falha de projeto, mais adiante ficará claro que esse comportamento não apenas era esperado como também necessário para o funcionamento da rede.

3.3 Treinamento

Conforme mencionado anteriormente, a cada vez que uma n-upla de pixels em tons de cinza é apresentada à rede para treinamento, são realizadas n operações de escritas ao respectivo nó RAM. Uma vez assimilados pela rede, os vetores binários resultantes da composição não guardam nenhuma relação entre si, ou seja, mais do que simplesmente aprender um padrão na forma de upla de valores de cinza, o que a rede realmente faz é aprender os vários padrões binários que o compõem. Mais adiante será demonstrado como a combinação entre esses padrões binários assimilados provê maior generalização ao mecanismo de reconhecimento de padrões. Ao passo que a ativação de um bit apenas é suficiente pra indicar a assimilação de um dado padrão binário no modelo WiSARD original, a rede RWiSARD é alimentada com os níveis de cinza que serviram de limiar de binarização na decomposição da upla de pixels em escala de cinza. Cada vetor gerado pela decomposição é empregado como um endereço do nó RAM associado à upla em questão, e nesse endereço está mantida a média aritmética de todos os valores de limiar que geraram esse mesmo endereço.

3.3.1 Contadores e Média Aritmética Móvel

Conforme visto na análise da arquitetura do neurônio RWiSARD, cada endereço do nó RAM da upla aponta para uma estrutura com dois campos: o valor da média ponderada para esse endereço, e o contador de operações de atualização. Sendo

inicializado com 0, a cada vez que esse endereço sofre uma atualização esse contador é incrementado. Durante o treinamento, esse contador cumpre com a primeira de suas duas finalidades participando do algoritmo de atualização da média móvel do endereço, descrito a seguir:

início

para cada $i \in \{1 \cdots n\}$ **faça**

$média[B_i] \Leftarrow \frac{média[B_i] \times contador[B_i]}{contador[B_i]+1}$;

$contador[B_i] \Leftarrow contador[B_i] + 1$;

fim

fim

Algoritmo 1: Atualiação do endereço do nó RAM RWiSARD

Um resultado interessante desse algoritmo de atualização, é que o conteúdo mantido em um dado endereço não é um retrato exato de nenhum dos padrões apresentados durante o treinamento, e sim um valor representativo para a classe como um todo. Um dos efeitos desse resultado será visto mais adiante. A figura 3.3 mostra um nó RAM sofrendo uma atualização em seus endereços em decorrência do treinamento da rede com uma upla de pixels.

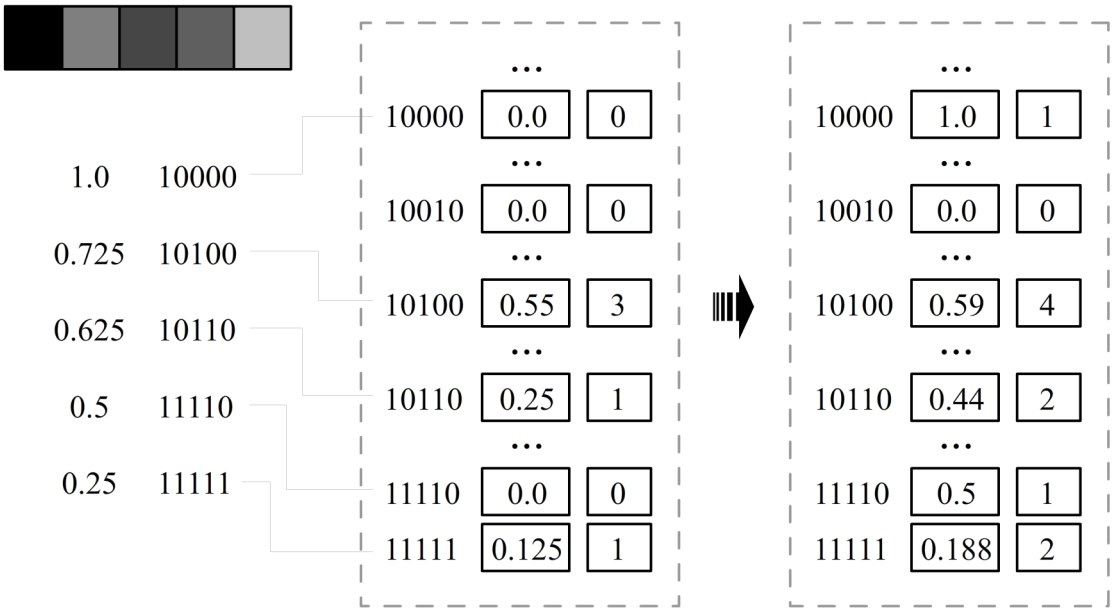

Figura 3.3: Atualização do nó RAM de uma 5-upla durante o treinamento.

3.4 Classificação

3.4.1 Diferenças para a rede WiSARD

Conforme visto, a rede WiSARD binária clássica responde a uma n-upla de pixels com 1, caso o vetor binário resultante da binarização desses pixels coincida com um endereço de memória onde esteja guardado o valor 1, e 0, no caso contrário. O neurônio que apresentar o maior somatório de todas essas respostas é o vencedor. No caso da rede RWiSARD, o neurônio responde com um valor real normalizado para o intervalo $[0, 1]$, indicando o quanto esse valor é distante da classe de padrões reconhecida por esse neurônio. Uma resposta igual a 0.0 corresponde a uma upla de pixels pertinente a essa classe reconhecida pelo neurônio. Uma resposta igual a 1.0 representa um padrão desconhecido para o neurônio. Valores entre 0.0 e 1.0 indicam o quanto o neurônio considera o padrão apresentado similar, ou próximo, da classe que ele reconhece. O neurônio que apresentar o menor somatório dessas respostas aponta para a classe identificada para a imagem, ao contrário do que a rede WiSARD faz.

3.4.2 Cálculo da resposta da rede RWiSARD

Na rede WiSARD binária o bit apontado pelo endereço gerado indica se o respectivo vetor binário foi assimilado ou não. No caso da rede RWiSARD, essa função é desempenhada pelo contador. Quando um endereço aponta para um contador com valor 0, o vetor binário correspondente não foi assimilado durante o treinamento. Nessa situação, um valor de 1.0 (maior distância possível do conteúdo da RAM para um dado vetor binário) é adicionado à resposta da upla como penalidade. Dessa forma, n-uplas de pixels que, quando decompostas, retornam vetores binários desconhecidos pela RAM, tenderão a apresentar uma resposta de maior valor, sinalizando uma maior distância para a classe reconhecida para o neurônio. O algoritmo de cálculo

da resposta para a upla é apresentado a seguir:

Entrada:

valores em escala de cinza dos pixels $P = \{P_1 \cdots P_n\}$;

endereços extraídos da n-upla $B = \{B_1 \cdots B_n\}$.

Saída: Resposta da n-upla

início

$somatório = 0.0$;

para cada $i \in \{1 \cdots n\}$ **faça**

se $contador \geq 0$ **então**

$somatório \leftarrow somatório + |média[B_i] - P_i|$;

senão

$somatório \leftarrow somatório + 1.0$;

fim

fim

retorna $\frac{somatório}{n}$;

fim

Algoritmo 2: Calcular resposta para a n-upla

O somatório final dessas respostas, como já foi mencionado, constitui a resposta do neurônio para a imagem apresentada. a figura 3.4 ilustra esse processo, mostrando como o nó RAM exibido na figura 3.3 responde a uma upla de pixels com valores ligeiramente diferentes daquela que foi usada durante o treinamento.

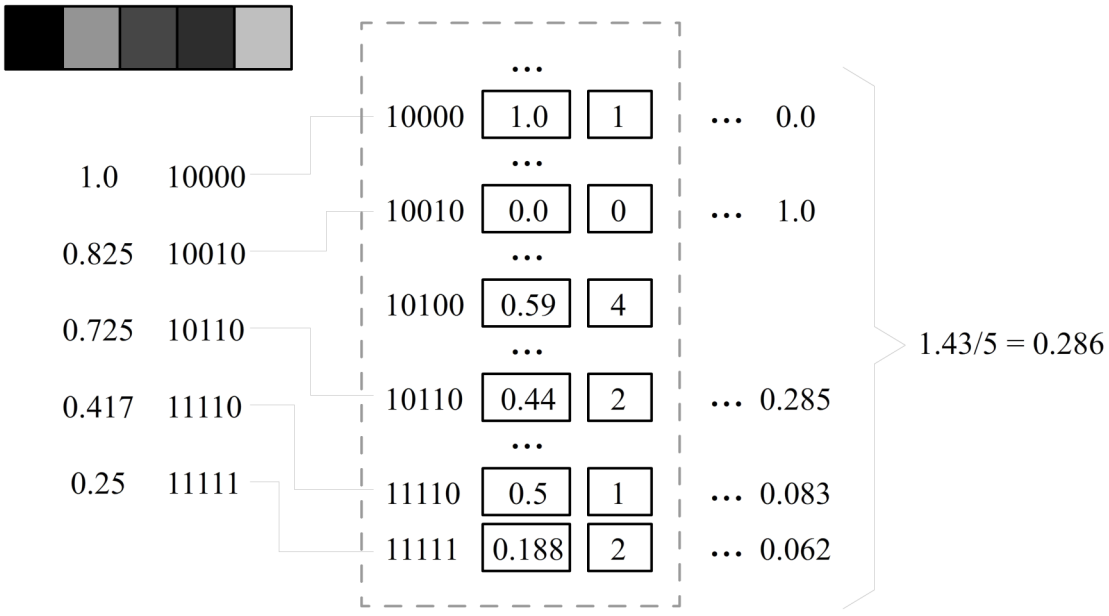

Figura 3.4: Cálculo da resposta da rede RWiSARD diante de uma 5-upla de pixels em escala de cinza.

3.5 Complexidade de Tempo e de Memória

Ainda que o valor de um bit da estrutura apontada pelos endereços tenha sido substituído por dois valores, um escalar e um inteiro, e que a cada operação em uma n-upla sejam realizadas n operações de acesso e escrita, se for o caso, no lugar de apenas uma, o comportamento assintótico da rede RWiSARD permanece o mesmo que a rede WiSARD original. Os novos valores que entram no cálculo de sua complexidade de memória e operacional são valores constantes, o que permite que a mesma análise seja estendida para esse modelo. A complexidade de treinamento pode ser expressa como:

$$\frac{n_{pixels}}{n_{upla}} \times (n_{upla} \times t_{bin} + n_{upla} \times t_{acesso}) = n_{pixels} \times (t_{bin} + t_{acesso}) =$$

$$k \times n_{pixels} \ \in \ O\left(n_{pixels}\right)$$

enquanto que a complexidade da operação de classificação é expressa por:

$$n_{classes} \times (\frac{n_{pixels}}{n_{upla}} \times (n_{upla} \times t_{bin} + n_{upla} \times t_{acesso}) + \frac{n_{pixels}}{n_{upla}} \times t_{somar}) + n_{classes} \times t_{comparar} \simeq$$

$$k_1 \times (k_2 \times n_{pixels} + k_3) \ \in \ O\left(n_{pixels}\right)$$

Quanto à complexidade de memória por sua vez, é descrita como:

$$n_{classes} \times (2^{n_{upla}} \times (s_{média} + s_{contador}) \times \frac{n_{pixels}}{n_{upla}}) \ \in \ O\left(n_{pixels}\right)$$

onde pode ser visto que a mudança na estrutura apontada pelo endereço aparece na expressão como um valor constante.

A mesma observação quanto ao valor de n_{upla} feita para o modelo WiSARD clássico permanece válida aqui, já que não há motivos práticos para uso de quantidades de pixesl na upla arbitrariamente alto.

Enfim, o modelo RWiSARD representa, sim, uma demanda maior de memória e de processamento, mas essa demanda é maior do que a da rede WiSARD apenas em um fator constante, e em termos práticos, um fator pequeno. Tal qual o modelo WiSARD clássico, esse modelo permanece uma solução viável e eficiente para aplicações reais, em contraste com algumas soluções propostas para tratamento de imagens em escala de cinza.

Capítulo 4

Experimentos e Resultados

4.1 Objetivo dos Experimentos

Afim de servir como base para comparação entre as performances da rede WiSARD binária clássica e o modelo RWiSARD, os experimentos mencionados a seguir se limitaram a empregar apenas essas redes neurais em uma abordagem direta, sem emprego de métodos mais elaborados como redução de dimensionalidade, análise de componentes principais, emprego de descritores de cena ou de objetos, modelos topográficos de rostos ou expressões faciais ou aplicação de cascatas de características. O dado consumido pelas redes neurais em si consiste nos próprios pixels das imagens, que não passaram por qualquer processo que não seja um estágio de normalização e, no caso da aplicação em redes binárias, do estágio de binarização exigido. O propósito da adoção dessa abordagem direta, em detrimento da utilização de métodos mais elaborados que poderiam proporcionar alguma melhoria nos resultados, é evitar que a performance de um ou outro modelo seja favorecida ou prejudicada pelo método sem que o devido estudo dessa influência fosse aqui exposto. Também não cabe aqui neste trabalho uma discussão mais detalhada a cerca do emprego do modelo aqui proposto, a rede RWiSARD, em conjunto com tais métodos mais elaborados, o que fica então sugerido como trabalho futuro.

Dois bancos de faces distintos foram empregados nesses experimentos, cada um se mostrando mais adequado ao seu respectivo experimento do que o outro por valorizar aspectos diferentes dos problemas de reconhecimento e detecção de faces.

4.2 Reconhecimento Facial

Esse experimento consistiu em treinar duas redes neurais, uma baseada no modelo WiSARD clássico e a outra, no modelo RWiSARD, com rostos de diferentes pessoas e depois, com um conjunto diferente de imagens, empregar estas redes para identificar

os indivíduos nas imagens. A comparação entre as redes é baseada na razão entre o número de imagens apresentadas para identificação e o número de imagens onde os indivíduos foram corretamente identificados.

4.2.1 Método

Ambas as redes apresentam a mesma topologia. Uma única camada de neurônios onde cada neurônio fica responsável por um indivíduo diferente. O vetor de entrada dos neurônios consiste na própria imagem em si, na forma como ela é disponibilizada pelo banco de faces empregado. A razão de acertos das redes neurais é levantada conforme o número de indivíduos diferentes, ou seja, o número de neurônios, presentes na rede, este número variando de 2 até 40 indivíduos diferentes. Para cada número

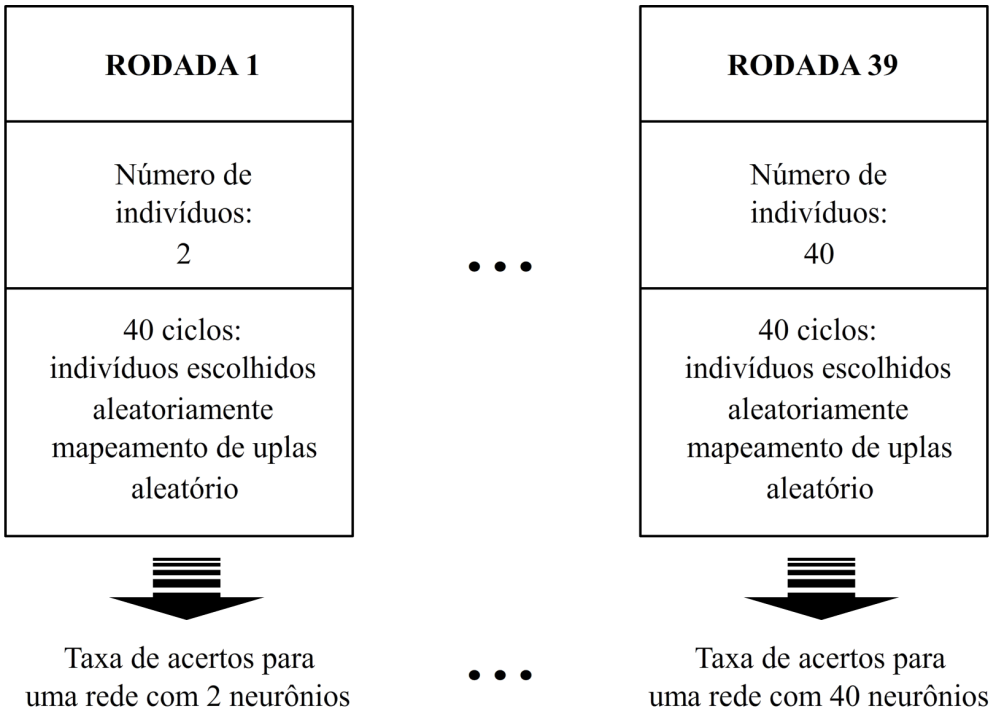

Figura 4.1: Diagrama do experimento de reconhecimento de faces.

de indivíduos, são realizados 40 ciclos de treinamento e classificação, a cada ciclo se empregando um conjunto distinto de indivíduos, escolhidos de forma aleatória a partir do banco de faces, além de um mapeamento aleatório de uplas diferentes. É importante se observar que mesmo no caso em que apenas um conjunto de 40 indivíduos diferentes é possível a partir do banco de faces empregados, ainda assim serão realizados 40 ciclos de treinamento e classificação afim de que se use um mapeamento de uplas diferente. Essa medida foi tomada para que a influência do

mapeamento aleatório e dos tipos das faces dos indivíduos empregados seja minimizada. Ao fim de cada rodada, a razão de acertos respectiva é tomada a partir da média de acertos para todos os 40 ciclos da rodada. Por fim a performance de cada rede é avaliada contra o número de neurônios empregados na rede. Todo esse processo é ilustrado pelo diagrama na figura 4.1.

4.2.2 Banco de faces utilizado

O banco de faces empregado neste experimento foi o Banco de Faces ORL, disponibilizado gratuitamente pelos Laboratórios AT&T de Cambridge. Este banco de dados apresenta 10 imagens diferentes para cada um dos 40 indivíduos disponíveis. Cada imagem tem 92x112 pixels em 256 níveis de cinza. Cada imagem apresenta o rosto em diferentes ângulos e apresentando diferentes expressões. Foram mantidos o mesmo fundo para todas as fotos, e o mesmo ângulo de iluminação. De um modo geral todas as imagens de rostos são delimitadas pela região que compreende o rosto, incluindo o cabelo do indivíduo. Estão misturadas imagens de pessoas com e sem óculos, exibindo a boca aberta ou não, como também em alguns momentos olhos abertos ou não. Algumas imagens do banco de faces ORL são exibidos na figura 4.2 como exemplo.

Figura 4.2: Exemplos de imagens faciais extraídos do banco de faces ORL dos Laboratórios AT&T de Cambridge.

4.2.3 Configuração das redes neurais e dos ensaios

Após uma série de testes, constatou-se que ambas as redes neurais exibiram maior índice de acertos, para este problema em especial, ao se empregar uplas de 3 pixels (3-uplas). Levando-se em conta a influência do mapeamento de pixels para upla, que é como de praxe feito aleatoriamente, na eficácia da operação de ambos os modelos, onde alguns mapeamentos aparentemente privilegiam uma rede em detrimento de outra, e como ainda não há um estudo que mostre quais mapeamentos são melhores tanto para a RWiSARD como para a WiSARD binária, a cada instância do teste foi empregado necessariamente o mesmo mapeamento de pixels para ambas as redes binária e RWiSARD. Tendo sido realizado 40 destes ciclos para cada taxa de acerto levantada, a cada ciclo sendo gerado um novo mapeamento aleatório, minimiza-se a influência de cada mapeamento na operação das redes neurais. As imagens são submetidas à rede RWiSARD sem nenhum estágio prévio de processamento ou segmentação. Isso também quer dizer que nenhuma normalização foi utilizada. Quanto a rede WiSARD binária, o limiar adotado para binarização das imagens foi a média da luminância de todos os pixels da imagem. Quanto ao conjunto de imagens faciais utilizados, havendo a disponibilidade de 10 imagens faciais diferentes por indivíduos, optou-se por empregar 2 dessas imagens como conjunto de treinamento, escolhidas aleatoriamente dentre as 10, e deixar as 8 imagens restantes como conjunto de teste de identificação. Apesar de parecer uma abordagem muito radical quando comparada ao usual 10-fold cross validation, há de se levar em conta que o espaço gerado pelas variações combinadas de expressão facial, incid??ncia de iluminação, posição do observador e outros será sempre maior em várias escalas de magnitude do que qualquer conjunto de treinamento que possa ser providenciado. Assim, um conjunto de faces de treinamento reduzido em comparação ao conjunto de testes representaria uma situação mais próxima de um cenário de aplicação real. Um outro motivo para se tomar essa decisão, foi o fato de que menores conjuntos de treinamento naturalmente causam um aumento da incidência de erros. Aumentando o risco de identificação errônea da face de um indivíduo se acentua a diferença de performance em número de acertos entre os dois modelos de rede, mesmo para os casos em que um número reduzido de indivíduos foi empregado.

4.2.4 Resultados

O gráfico na figura 4.3 exibe a performance dos dois modelos de redes neurais ao longo das 39 rodadas de testes realizados. A rede RWiSARD apresenta vantagem sobre a rede WiSARD binária clássica mesmo com apenas 2 neurônios e essa vantagem tende para uma diferença pontual de 12% à medida que o número de neurônios na rede se aproxima de 40. Ainda que se espere um comportamento monotômico para

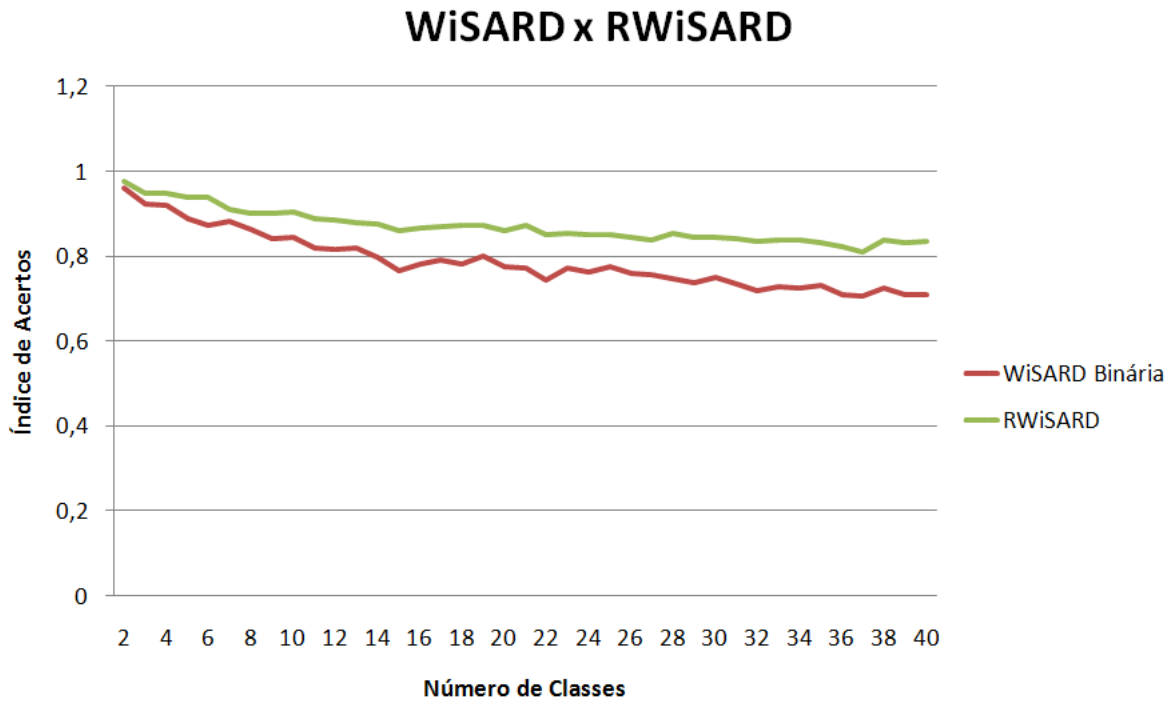

Figura 4.3: Número de Indivíduos X Taxa de Acertos das redes WiSARD e RWiSARD.

estas curvas, a ocorrência de oscilações onde se vê algumas vezes resultados melhores para números de neurônios maiores, como no intervalo entre 15 e 19 neurônios é explicada pelo fato de que 40 instâncias de teste para cada rodada com x está longe de exaurir todas as $\binom{40}{x}$ combinações possíveis para o experimento. Por exemplo, para 4 indivíduos diferentes, seriam necessários 91390 ensaios de teste para se obter a taxa de acertos exata para esse número de indivíduos. Como algumas combinações de indivíduos tendem a favorecer os resultados de ambas as redes, mais do que outras, é de se esperar que ocorram oscilações como as apresentadas pelas curvas. Ainda assim, numa escala que vai de 0.0 a 1.0, oscilações menores do que 0.03 pontos se mostram de pouca relevância para o objetivo desse teste. Pode se concluir que, pelo menos para o banco de faces empregado, a rede RWiSARD apresenta uma performance notavelmente superior à da rede WiSARD binária clássica.

A figura 4.4.a exibe a imagem de uma mulher que foi erroneamente identificada pela rede WiSARD binária durante o experimento. O exemplo ilustra como a limitação em se processar apenas imagens binárias afeta o reconhecimento de imagens através da rede WiSARD. As imagens em 4.4.b e 4.4.c mostram respectivamente as respostas dos neurônios RWiSARD e WiSARD que foram treinados com imagens dessa mulher. Como uma forma de se avaliar os resultados visualmente, as imagens correspondendo às respostas das redes sofreram uma conversão de cor, onde o tom vermelho representa valor de resposta mais alto e tons de amarelo a verde,

Figura 4.4: a) Imagem a ser identificada; b) Resposta da rede RWiSARD; c) Resposta da rede WiSARD para o neurônio correto.

valores de resposta mais baixos. A resposta da rede WiSARD binária clássica, em particular, corresponde à uma imagem binária, uma vez que cada uma de suas uplas responde apenas com 1 ou 0. Ao ser convertido pelo mesmo processo que a resposta pela rede RWiSARD, a rede WiSARD exibe pixel vermelho para as uplas que tiveram resposta igual a 1.

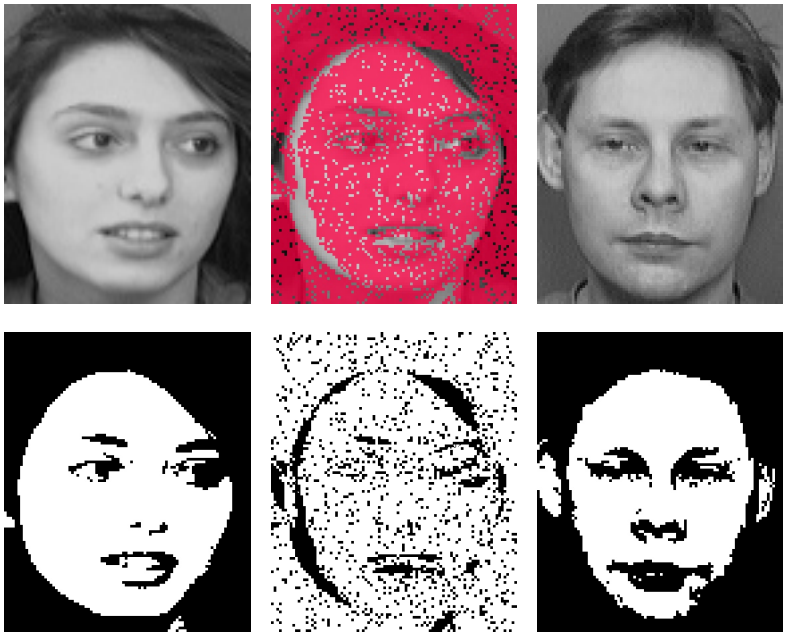

Figura 4.5: a) Imagem a ser identificada; b) Resposta da rede WiSARD para o neurônio vencedor; c) Indivíduo identificado pela rede WiSARD. d) Imagem binarizada; e) A resposta do neurônio WiSARD vencedor; f) Indivíduo identificado pela rede WiSARD, binarizado;

O indivíduo que foi erroneamente identificado como sendo o apresentado pela imagem em 4.5.a é exibido na figura 4.5.c. A figura 4.5.b mostra a resposta gerada pelo neurônio vencendor, treinado com imagens do indivíduo mostrado em 4.5.a. As imagens nas figuras 4.5.d e 4.5.f exibem respectivamente a imagem original e a do indivíduo que foi identificado, binarizadas. Finalmente, a imagem exibida na figura 4.5.e mostra a mesma resposta do neurônio vencedor que foi exibida em

4.5.b, desta vez tal como foi obtida a partir do neurônio, sem conversão de cores ou sobreposição à imagem original.

Figura 4.6: a) resposta do neurônio WiSARD vencedor; b) Operação OU-Exclusiva entre a imagem a ser identificada binarizada e a imagem binária do indivíduo identificado pela WiSARD;

A figura 4.6.b mostra a imagem resultante da aplicação da operação OU-Exclusiva entre as imagens binarizadas da imagem originalmente submetida à rede para identificação, e da imagem do indivíduo que foi equivocadamente identificado pela rede WiSARD binária. Essa operação, formalmente definida como $x_{i\,j} = a_{i\,j} \oplus \overline{b_{i\,j}}$ retorna como pixels brancos os pontos onde os pixels das duas imagens assumem o mesmo valor 1 ou 0. Na figura 4.6.a é novamente reproduzida a imagem da resposta do neurônio vencedor. É notória a semelhança entre essas duas imagens, a menos de alguns poucos pixels. Um dado mapeamento aleatório de pixels de uplas pode levar a uma grande incidência de uplas cujos pixels se encontram todos nas mesmas posições em que alguns dos pixels exibidos como brancos na figura 4.6.b, e essa incidência pode levar o neurônio errado a apresentar uma resposta total maior do que a do neurônio supostamente correto. Essa situação ilustra um sintoma do modelo clássico provocado pela perda de informação devida ao processo de binarização da imagem.

4.3 Detecção de Características Faciais

Já foi demonstrado neste trabalho como a rede RWiSARD realiza reconhecimento e classificação de padrões medindo a semelhança entre padrões em escala de cinza atráves de um modelo matemático simples que procura emular a habilidade humana de atribuir graus de similaridade entre imagens diversas. O experimento anteriormente descrito mostra a vantagem que esse método traz sobre o modelo WiSARD clássico em um cenário de reconhecimento de imagens. O próximo experimento a ser descrito visa lançar m??o dessa habilidade para identificar dentro de uma imagem, a região que apresenta a maior semelhança a uma outra imagem anteriormente apresentada como protótipo de busca para a rede neural. O objetivo dessa experiência

é mostrar a viabilidade de se empregar a rede RWiSARD em um sistema eficaz de detecção de características ou em um sistema mais complexo para detecção de faces, e tentar avaliar quais os limites para eficácia desse modelo, tendo sempre a rede WiSARD binária clássica como parâmetro de comparação.

Nenhum estágio como extração de características ou uso de descritores foi empregado como forma de pré-processamento de imagens, bem como não foi utilizado nenhum modelo de detecção facial baseado em biometria. Os dados com os quais a rede é alimentada consiste mais uma vez apenas nas próprias imagens em si. O método empregado na detecção de características se baseia puramente na semelhança visual entre a imagem-modelo e a imagem apresentada para comparação, e em como a rede RWiSARD emula essa percepção de semelhança visual.

Estando longe do escopo deste trabalho um estudo a respeito da influência que o emprego de técnicas de detecção de características mais complexos possa ter sobre a comparação entre as redes RWiSARD e WiSARD binária, foi adotado um método mais simples e direto neste experimento. Um sistema de detecção de características completo e robusto poderia ser construído através da aplicação de redes RWiSARD em conjunto com tais métodos mais sofisticados, e uma breve discussão sobre essa possibilidade fica relegada ao próximo capítulo.

4.3.1 Método

Tomando-se a saída de um neurônio da rede como métrica, é possível medir o quanto uma imagem apresentada como alvo para a rede neural se parece com as imagens originalmente apresentadas à rede durante o treinamento desse neurônio, a grosso modo emulado a habilidade humana de conferir graus de semelhança visual entre diferentes imagens. Esta habilidade naturalmente será empregada como forma de se estabelecer nexo entre imagens de um mesmo objeto sendo observado em diferentes circunstâncias tais como posição relativa ao observador - ângulo e distância - iluminação incidente, presença de obstáculos, etc.

Em uma situação ideal, o valor da saída do neurônio se aproximaria do valor extremo para esse neurônio, que seria o valor mais alto possível para um neurônio WiSARD, igual a quantidade de uplas presentes na rede, ou 0 no caso de um neurônio RWiSARD, quanto maior a semelhança entre a imagem alvo e a imagem original empregada no treinamento, daqui em diante chamada Protótipo. Assim, no espaço de busca que consiste em toda a imagem onde a tal característica será procurada, o ponto x onde a saída $S(x)$ alcança um valor extremo, o valor máximo para toda a imagem no caso da rede WiSARD, ou o valor mínimo no caso da rede RWiSARD, para toda a imagem é sinalizado como o ponto onde a imagem delimitada pela janela de busca mais se assemelha a imagem protótipo. Os gráficos na figura 4.7

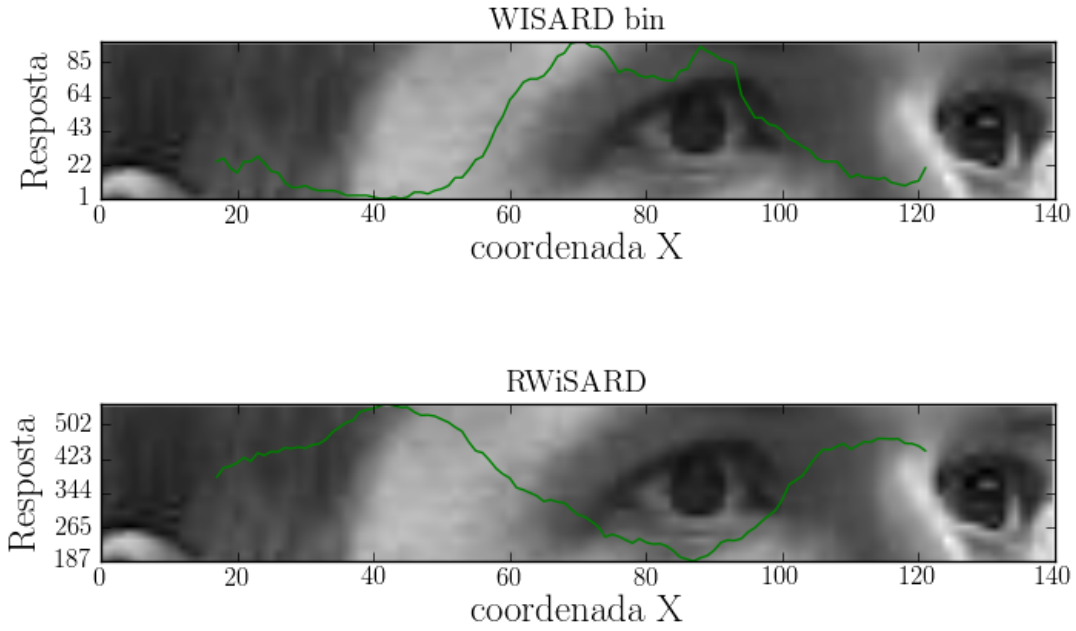

Figura 4.7: Gráficos das resposta da rede WiSARD binária e RWiSARD.

mostram as saídas dos neurônios das rede WiSARD binária e RWiSARD, enquanto a janela de busca se desloca no eixo horizontal da imagem. Ambos os neurônios foram treinados com a imagem do olho direito do indivíduo. Como era esperado, ambos os gráficos exibem a saída do neurônio alcançando o valor extremo quando a janela está centralizada no olho direito do indivíduo.

Já a figura 4.8 mostra um ensaio de busca para o olho direito do indivíduo. O retângulo azul da figura 4.8.a delimita a imagem que servirá de protótipo para o olho direito. A figura 4.8.b exibe uma imagem do mesmo indivíduo, com o rosto em um ângulo diferente da 4.8.a. O retângulo vermelho aponta a posição registrada como ground-truth para essa imagem. Já o retângulo verde mostra a localização apontada pela rede RWiSARD. Finalmente, a figura 4.8.c mostra a saída do neurônio codificada em uma escala de cores que vai de azul sinalizando a resposta mais baixa, passando pelo verde até o vermelho, que indica a resposta mais alta. É importante aqui se lembrar que a resposta da rede RWiSARD se torna mais alta à medida que a sua saída se aproxima de zero.

Este experimento é uma tentativa de se avaliar até que ponto se pode chegar partindo de uma assunção tão simples e aparentemente ingênua como essa, e o quanto esse método pode se beneficiar da adoção do modelo RWiSARD como neurônio de busca.

Foram utilizados como protótipos de busca os olhos direito e esquerdo, separadamente, a região entre os olhos, o nariz e a boca. As imagens de protótipo de cada uma

Figura 4.8: a) Olho direito como protótipo de busca; b) Olho direito detectado em um ângulo diferente pela rede RWiSARD; c) Saída do neurônio codificada por cores.

Figura 4.9: Face em Tilt: 0, Pan: 0. Regiões empregadas como protótipos para busca

dessas características foram extraídas de uma única imagem, referente ao indivíduo quando ele olha diretamente para o observador. A figura 4.9 exibe essa imagem bem como todas as características empregadas no experimento em destaque. Deve ser observado que o tamanho da janela que delimita a região que serve de protótipo também determina o tamanho da janela de busca, constante por todo o processo em que varre a imagem alvo.

Um detalhe importante a ser observado é que como as imagens usadas no treinamento das redes neurais como as imagens que serão posteriormente processadas pelos neurônios possuem a mesma dimensão, ou que pelo menos o número de pixels dos quais são compostas as uplas e que serão lidos das imagens é constante por todo o ciclo de vida do neurônio. Isso causa uma primeira limitação visível ao método, já que ambas a janela delimitatória do protótipo e a janela de busca deverão ter as mesmas dimensões. Isso evidentemente restringe o quanto o alvo a ser comparado pela rede pode ser maior ou menor que o protótipo.

Como já comentado anteriormente nesse trabalho no capítulo 1, a enorme diversidade de situações em que a imagem de um rosto pode ser capturada impõe uma séria restrição a qualquer tentativa de se estabelecer um conjuto de exemplares que represente de maneira adequada o universo de padrões referentes a imagens de um rosto. Mesmo a gama de imagens diferentes que podem ser geradas a partir de um mesmo rosto será sempre muito maior do que qualquer conjunto de amostras que seja possível se empregar como base para o treinamento, e isso é um problema mesmo com a aplicação de métodos baseados em modelos biométricos ou orientados à redução da dimensionalidade do espaço de busca, como os que empregam PCA. Tal fato levou à decisão de se realizar este experimento com o menor número de protótipos para treino possível - apenas uma imagem - como forma de se melhor avaliar a robustez dessa abordagem dentro de um cenário tão restrito.

A métrica usada aqui para avaliação dessa rosbustez então será a quantidade de imagens do rosto em ângulos diferentes, onde a característica a ser buscada é efetivamente encontrada, a partir de uma única imagem do rosto do indivíduo. Ground-Truth será fornecido ao sistema informando sobre a posição da característica na imagem-alvo. Deve ser levado em conta o fato de que a posição da janela fornecida pelo Ground-Truth foi em primeiro lugar estabelecida pela observação e pelo julgamento de um agente humano. Dessa forma, essa informação deve ser tomada como uma aproximação, e uma razoável margem de tolerância deve ser considerada para se avaliar a efetividade da resposta da rede neural a partir do ground-truth fornecido. De um modo geral, foi adotada uma área de 4x4 pixels em torno do ground-truth como margem de tolerância para a indicação de acerto.

Sob a restrição de se usar apenas uma única imagem como protótipo, uma estratégia de treinamento deve ser adotada para permitir que a rede neural extrapole

por si só informação suficiente para que ela seja capaz de reconhecer a característica sob diferentes ângulos. Não se contando com um modelo geométrico ou topológico da característica em questão, não é possível se prever como a imagem da característica projetada no plano do observador irá se deformar diante de mesmo as menores alterações quanto ao ângulo do observador. Além das deformações de caráter não linear da imagem observada, novos ângulos podem tornar algumas regiões do rosto visíveis, introduzindo novas superfícies à imagem observada, bem como também podem ocultar outras.

Tentar prever quais as novas informações que serão trazidas pelas imagens durante a busca, e ainda tentar treinar a rede neural a partir dessa estimação, sem um modelo tridimensional, fotométrico ou de qualquer natureza que fosse, se mostra um exercício de futilidade e está longe da proposta desse experimento. Por outro lado, mesmo diante das mais severas deformações alterações de ângulo, expressão e mesmo de iluminação, diversas áreas da superfície do rosto estarão presentes tanto na imagem protótipo como na imagem de busca, ainda que deslocadas e sob pequenas deformações no plano da imagem. Do ponto de vista da rede neural, a cada upla cujos pixels correspondam a aproximadamente a mesma região, ou ainda a uma região da imagem suficientemente parecida com a da imagem original, se disparará uma resposta positiva indicando o reconhecimento daquela região. A estratégia adotada então evitará fazer assunções a respeito das alterações no plano da imagem, e se baseará apenas no grau de coincidência entre regiões da imagem ainda que sob algum deslocamento ou deformidade.

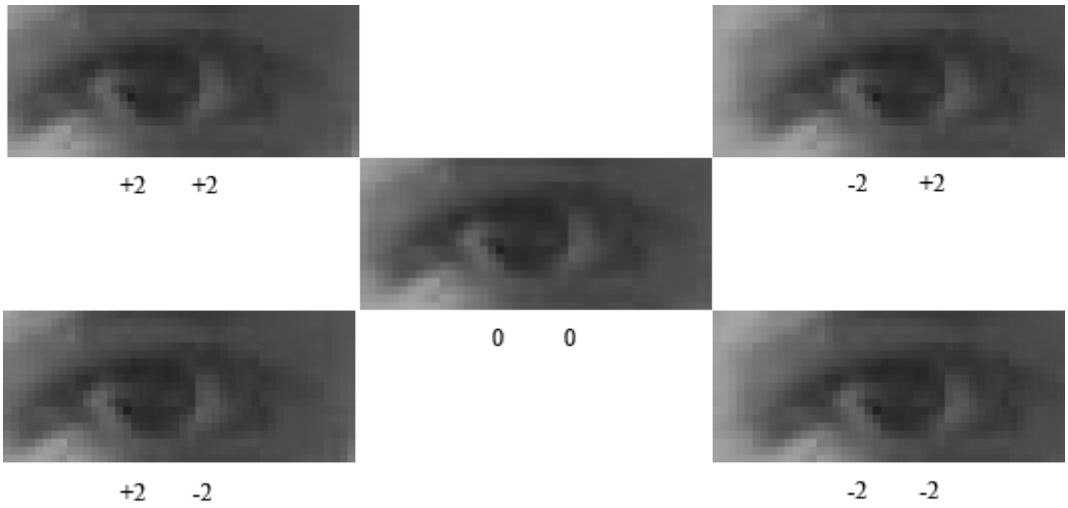

Figura 4.10: Imagens resultantes do deslocamento do protótipo.

Modelos de redes neurais baseados em classificadores n-tuple já fragmentam imagens em pequenos conjuntos não necessariamente contínuos de pixel por natureza. Esses conjuntos podem ser treinados com padrões de pixel independentes e depois recombinados de forma a reconhecer novas imagens diferentes das originais. A es-

tratégia tomada tira proveito disso para reconhecer uma mesma característica sob circunstâncias diferentes. Para que padrões de pixels inerentes a regiões da imagem que sofreu deslocamento sejam reconhecidos, a imagem protótipo é passada para a rede neural durante o treinamento com diversos deslocamentos. A rede neural vai então assimilar o mesmo padrão de pixels empregando diferentes uplas ou combinações delas. Devido a independência do funcionamento das uplas, padrões que sofreram diferentes deslocamentos ainda tem as chances de serem reconhecidos pela rede ampliadas por este método.

A figura 4.10 mostra a imagem prototípica do olho direito sob o efeito de vários deslocamentos. O método adotado e descrito nesse trabalho para gerar o conjunto de imagens de treinamento foi delineado a partir de resultados práticos. Outras estratégias para geração de um conjunto de treinamento a partir de uma única imagem são possíveis, e não há nenhum estudo ainda mostrando qual seria a melhor estratégia possível. Uma discussão sobre possíveis estratégias de treinamento fica então relegada ao próximo capítulo.

A estratégia aqui apresentada adotou deslocamentos verticais e horizontais diretamente proporcionais à altura e à largura respectivamente. Neste experimento foi empiricamente determinado o uso de um fator de 10% das respectivas dimensões para o deslocamento. De um modo geral, valores muito maiores ou menores que esse levaram a resultados menos satisfatórios. É importante observar que essa estratégia não leva em conta a natureza as dimensões da característica de interesse em si. Apenas a dimensão da janela foi usada no cálculo. Isso significa que o cálculo dos deslocamentos foi empregado sem alteração para cada uma das características de interesse. Um estudo mais aprofundado pode vir a revelar alguns fatores inerentes aos objetos de interesse nas imagens que possam se mostrar relevantes para alguma abordagem mais robusta, no futuro. O deslocamento foi aplicado em ambos os sentidos nos eixos ortogonais, bem como nos eixos diagonais.

Uma das maiores diferenças entre o modelo WiSARD clássico e o modelo RWiSARD está no conteúdo das memórias da upla. Ao passo que a rede WiSARD se limita a sinalizar um dado padrão binário como reconhecido ou não atrás do valor de um único bit, a rede RWiSARD mantém um valor escalar correspondendo à média móvel do limiar com o qual o padrão referente àquele endereço foi extraído a cada nova imagem apresentada. Obviamente o valor dessa média móvel irá se deslocar na medida em novo padrões forem apresentados, possivelmente se tornando cada vez mais distante do valor de um dos elementos do conjunto de treinamento. Este aspecto da rede RWiSARD leva à necessidade de se tomar um cuidado especial quanto a forma como o treinamento irá afetar o conteúdo do nó RAM da upla e consequentemente, a maneira como a rede irá responder aos padrões a ela apresentados mais tarde.

Em aplicações como reconhecimento de padrões, tais como o caso do experimento anterior, esse deslocamento da média móvel propicia uma maior capacidade de generalização da rede neural. No caso em que se procura identificar de todas as imagens apresentadas quais mais se assemelham a um protótipo fixo, gerar deslocamentos e treinar a rede com estes deslocamentos pode significar afastar os valores registrados pelas uplas dos valores originais, ou seja, dos valores que teriam sido registrados caso apenas a imagem protótipo, sem deslocamento, fosse apresentada à rede. O efeito colateral disso é a rede acabar com uma capacidade de generalização maior que a esperada, e ainda sinalizar com uma resposta de maior similaridade para imagens visualmente bem diferentes do protótipo original, comprometendo a eficácia do modelo.

Afim de sanar esse efeito colateral, seria ideal manter o valor correspondente à média móvel de cada endereço o mais próximo possível do valor gerado pela imagem original. Para tanto, a rede poderia ser treinada repetidas vezes com o protótipo, mantendo os valores da média móvel o mais próximo possível dos valores gerados pela imagem do protótipo. Como foi comprovado durante os ensaios, tal abordagem leva a um empobrecimento do treinamento proporcionado pelas imagens de deslocamento. Em um aprimoramento dessa abordagem, as imagens deslocadas também são repetidas vezes apresentadas à rede para o treinamento, de forma a não ter a sua contribuição para o treinamento tão empobrecida pela repetição do treinamento com o protótipo original. Finalmente, o desenvolvimento dessa abordagem levou a adotar uma quantidade de repetições de treinamento maior para as imagens mais próximas - menor deslocamento - da imagem original do protótipo, reduzindo o número de repetições para as imagens mais deslocadas com relação ao protótipo. Na prática, isso levou a se empregar, nos experimentos, o seguinte cálculo para o número de repetições para cada índice de deslocamento:

$$nR = \left\lfloor \left(max(d) - \sqrt{dx^2 + dy^2} \right)^3 \right\rfloor$$

Onde $max(d)$ representa a maior distância possível de uma imagem deslocada do protótipo original.

O gráfico exibido na figura 4.11 mostra o número de repetições por deslocamento empregado no experimento. Algo importante a ser notado é que para alguns deslocamentos mais distantes, o número de repetições calculado será igual a zero, ou ainda menor que zero, caso ao invés de $max(d)$ seja adotado um valor de corte para o deslocamento. As imagens cujos deslocamentos levaram a um número de repetições menor ou igual a zero serão descartadas do conjunto de treinamento da rede.

Mais uma vez, a fórmula empregada foi fruto de um trabalho experimental, e

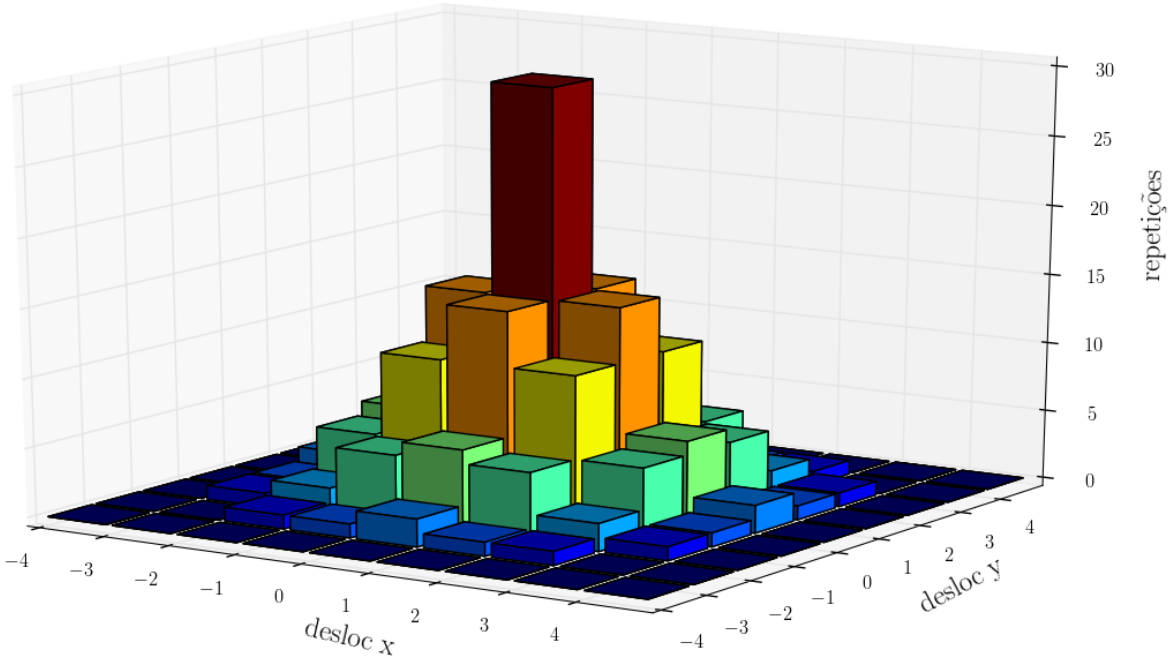

Figura 4.11: Gráfico Deslocamento X vs. Deslocamento Y vs. Número de Repetições durante o treinamento.

não há nenhum estudo que a indique como a melhor formulação possível. Tampouco foi levado em conta qualquer particularidade da imagem prototípica em si para se derivar uma fórmula mais apropriada para cada caso.

Figura 4.12: a) Imagem mental da boca, sem o uso de repetições; b) Protótipo original da boca; c) Imagem mental da boca, treinada com repetições.

Ilustrando o resultado dessa abordagem, a figura 4.12 exibe o impacto do uso de repetições no treinamento. a figura 4.12.c apresenta uma imagem visivelmente mais próxima da imagem original que serviu de protótipo para a boca. Essa é a imagem mental extraída do neurônio que foi treinado com o método de cálculo de repetições empregados. Esse método de repetições pode então ser encarado como uma forma de se manter um controle sobre a capacidade de generalização da rede, de forma a manter essa generalização ampla o bastante pra englobar algumas das diversas formas que a característica de interesse pode tomar em uma imagem, ao mesmo tempo que restrita o bastante para evitar que imagens que não exibam a

mesma característica não sejam apontadas como similares, ao menos não com uma frequência tão grande que torne este método inviável.

Treinar a rede neural repetidas vezes para a imagem protótipo e cada um de seus deslocamentos pode tornar o processo de treinamento significativamente mais lento, sem que nenhuma informação realmente nova seja inserida a cada treinamento. Esse problema é trivialmente resolvido ao se introduzir o número de repetições ao cálculo que leva à atualização da média móvel, no nó RAM da upla. Dessa forma, seja m_c o valor presente no nó RAM juntamente com o contador C, x o valor com o qual a upla será atualizada e r o número de repetições calculado para o treinamento, o emprego da fórmula:

$$m_{C+r} = \frac{m_C + rx}{C + r}$$

Surtirá o mesmo efeito que treinar a rede neural r vezes com o mesmo valor x. Em seguida, o contador é atualizado como $C = C + r$. Assim a rede é treinada por cada imagem em um único passo.

4.3.2 Banco de faces utilizado

A possibilidade de se gerar resultados de natureza quantitativa para esse experimento, e a relevância desses resultados, dependem diretamente do emprego de um banco de faces que cujas imagens apresentassem um comportamento ligeiramente monotômico. Diante de uma situação em que a variação de um único parâmetro no espaço do objeto pode levar a uma grande diferença entre as imagens desse mesmo objeto antes e depois da variação desse parâmetro, a utilização de um conjunto de imagens que exibam o mesmo objeto sob aproximadamente as mesmas condições a menos de um ou dois parâmetros se revela uma forma razoável de se medir a eficiência dos métodos testados, uma vez associada cada imagem ao índice de variação desses parâmetros. O banco de faces escolhido para este experimento foi o Head Pose Image Database, produzido pelo laboratório GRAVIR, associado ao INRIA, em virtude da pesquisa descrita em [11]. Este banco de dados apresenta 15 indivíduos, cada um em 2 séries, cada série por sua vez contendo 93 imagens, somando um total de 2790 imagens. Para cada série, as 93 imagens representam o rosto de um mesmo indivíduo, com inclinações verticais (tilt) tanto quanto horizontais (pan) variando no intervalo de -90 a 90 graus. Demais características tais como iluminação ambiente, expressão facial e distância do observador permanecem constantes por toda a série.

O emprego de uma dessas séries durante cada ensaio permite que os resultados obtidos para cada uma das imagens empregadas no ensaio sejam trivialmente indexadas segundo seus valores de tilt e pan, a exemplo do que pode ser visto na figura 4.13. Alguns exemplos das várias imagens contidas em uma série referente

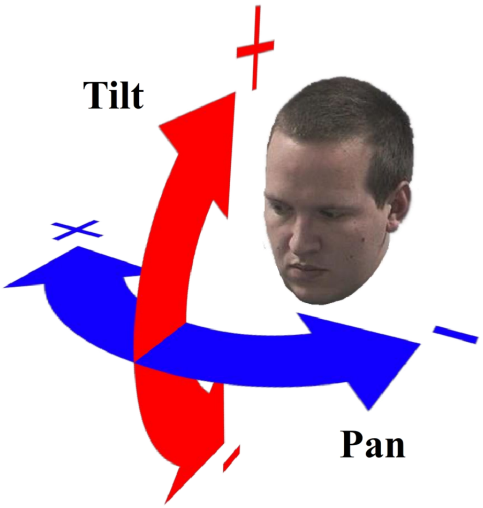

Figura 4.13: Imagens do banco de faces indexadas por Tilt e Pan.

a um indivíduo podem ser vistos na figura 4.14. Quanto à qualidade das imagens deste banco de faces, todas as imagens tem as mesmas dimensões, 384x288 pixels, e apesar de serem imagens coloridas, para o experimento foram convertidas em escala de luminância, afim de serem processadas pelas redes neurais.

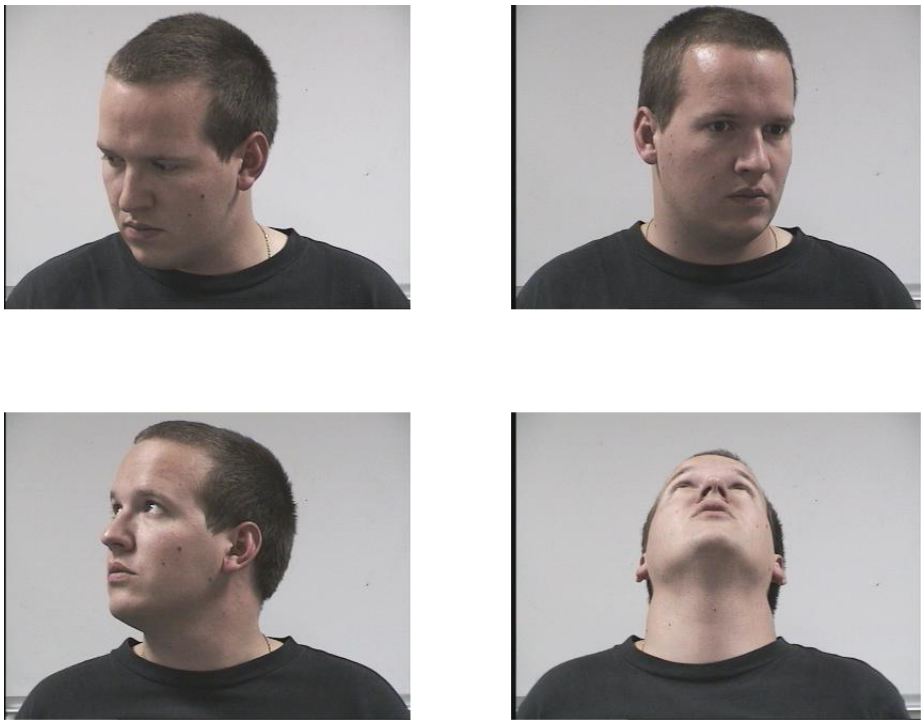

Figura 4.14: Exemplos extraídos do banco de faces utilizado.

Todos ensaios realizados neste experimento extraíram seus protótipos da imagem correspondente ao rosto em $tilt = 0$ e $pan = 0$ (o indivíduo olhando diretamente para o observador).

4.3.3 Resultados

A cada vez que o método foi aplicado a uma imagem de busca, foram registrados três valores: R_{max}, o valor da resposta máxima do neurônio para essa imagem; R_j, o valor da resposta máxima do neurônio para uma janela que se encontrasse a menos de 4 pixels de distância do Ground Truth para essa imagem de busca, e R_{min}, o valor da resposta mínima do neurônio para essa imagem. É assumido que sempre que R_j assume o mesmo valor que R_{max}, o sistema apontou o corretamente o Ground Truth como a posição encontrada da característica facial na imagem de busca. Então uma primeira métrica da performance seria a quantidade de imagens, dentre todas as imagens de busca que foram empregadas no ensaio, para as quais $R_j = R_{max}$.

Além da quantidade de imagens nas quais o ground truth foi corretamente identificado como a posição da característica facial buscada, é de grande relevância se registrar também o quão próximo está o valor de resposta da janela do valor máximo de resposta. Em inúmeras situações, a área quadrada de 4x4 pixels de tolerância em torno do ground truth não será o suficiente para capturar o valor de resposta máxima ainda que esse valor tenha ocorrido razoavelmente próximo ao ground truth, o qual, vale a pena frisar, é fixado meramente por um observador humano, sem nenhuma outra referência mais precisa.

Outra possibilidade é de que ainda que o resultado apontado neurônio não seja suficiente próximo do ground truth, ou seja, ainda que a rede tenha apontado equivocadamente para outra região da imagem, a região referente ao ground truth possa ter surtido uma resposta consideravelmente próxima à resposta máxima. Uma abordagem complementar a essa técnica poderia tirar proveito desse fato adotando um limiar a partir do qual valores de resposta altos o bastante delimitam o espaço de busca a umas poucas regiões nas quais há grande possibilidade de se encontrar a característica buscada. Partindo-se dessa premissa, essa métrica se mostra igualmente importante para se avaliar a eficácia dos modelos aqui discutidos.

A resposta máxima do neurônio varia muito de imagem para imagem, e tende a cair quanto maior a diferença de tilt e pan entre a imagem de busca e a imagem protótipo, já que imagem de busca também tende a ficar mais diferente da imagem protótipo de um modo geral. Dessa forma os valores são normalizados dentro do intervalo entre a maior e a menor resposta para cada imagem de busca através da fórmula:

$$v_j = \frac{(R_j - R_{min})}{(R_{max} - R_{min})}$$

Os resultados exibidos a seguir tomam v_j como escala de performance, ao lado da quantidade de imagens para as quais $v_j = 1$. Cada ponto nos gráficos exibidos a seguir corresponde a uma imagem de busca do conjunto usado para o experimento,

posicionado no plano conforme os valores de tilt e pan para a sua respectiva imagem. Pontos pretos representam imagens onde o ground truth não foi indicado com a característica encontrada, ou seja, $R_j < R_{max}$, enquanto pontos brancos indicam que a imagem teve seu ground truth corretamente identificado como a característica procurada naquela imagem.

De um modo geral, se considera o modelo que alcançou os melhores aquele que encontrou a característica no ground truth em um número maior de imagens de busca (maior quantidade de pontos brancos no gráfico), bem como aquele que apresentou maior número de pontos com valores próximos a 1.0.

Olhos direito e esquerdo

As figuras 4.15 e 4.16 trazem algumas imagens que foram encontradas durante os ensaios de busca realizados para o olho direito e o esquerdo, respectivamente. Também são mostrados as posições no gráfico tilt x pan referentes a essas imagens resultantes.

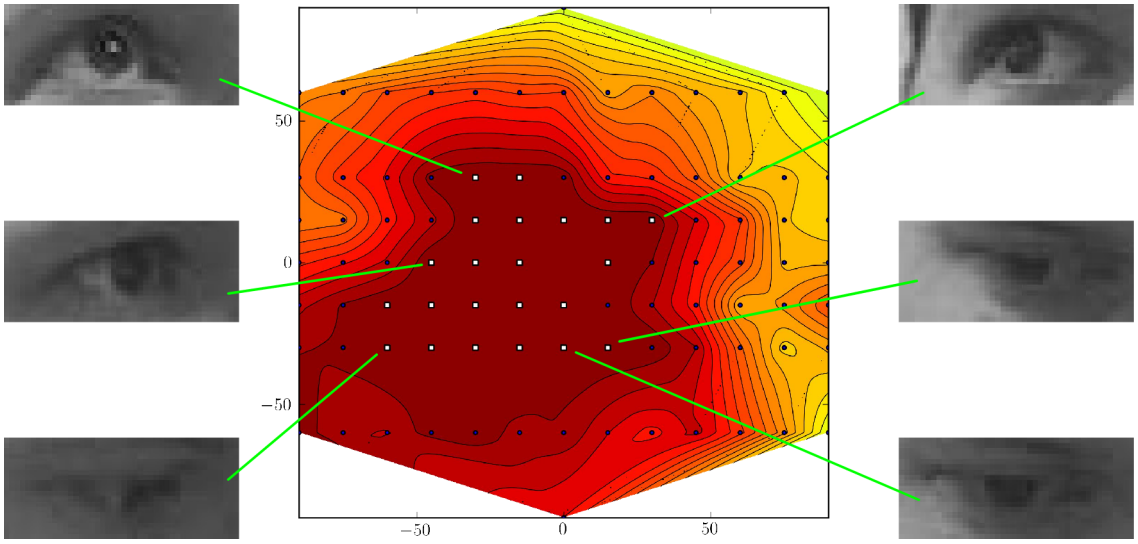

Figura 4.15: Algumas imagens de resposta para o olho direito.

Uma das imagens resultantes em destaque na 4.16 ilustra uma situação em que a área de tolerância de 4x4 pixels não teria sido suficiente para reconhecer um resultado que poderia ser tomado como aceitável. Na figura 4.17.c é mostrada a imagem referente ao ground truth para o rosto em $tilt = -15$ e $pan = +15$, enquanto que na figura 4.17.d é mostrada a imagem que foi apontada pela rede como o olho esquerdo. Mais uma vez é importante frisar que as especificações adotadas para estes ensaios não foram adaptados para nenhum caso particular como por exemplo as peculiaridades de cada estrutura presente no rosto, o que induziria a uma abordagem baseada em biometria, o que está fora do escopo desse trabalho.

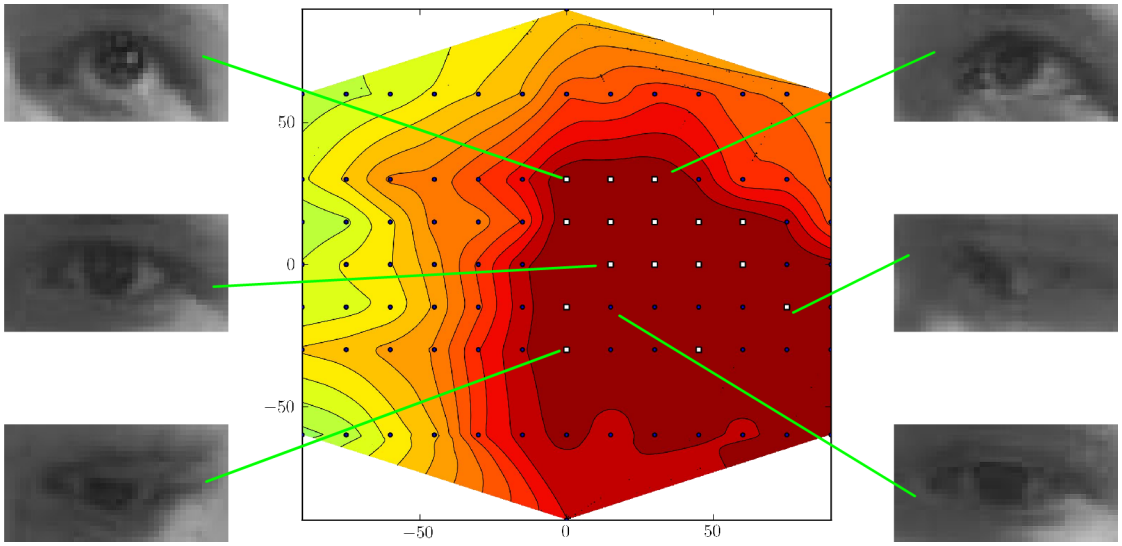

Figura 4.16: Imagens de resposta para o olho esquerdo.

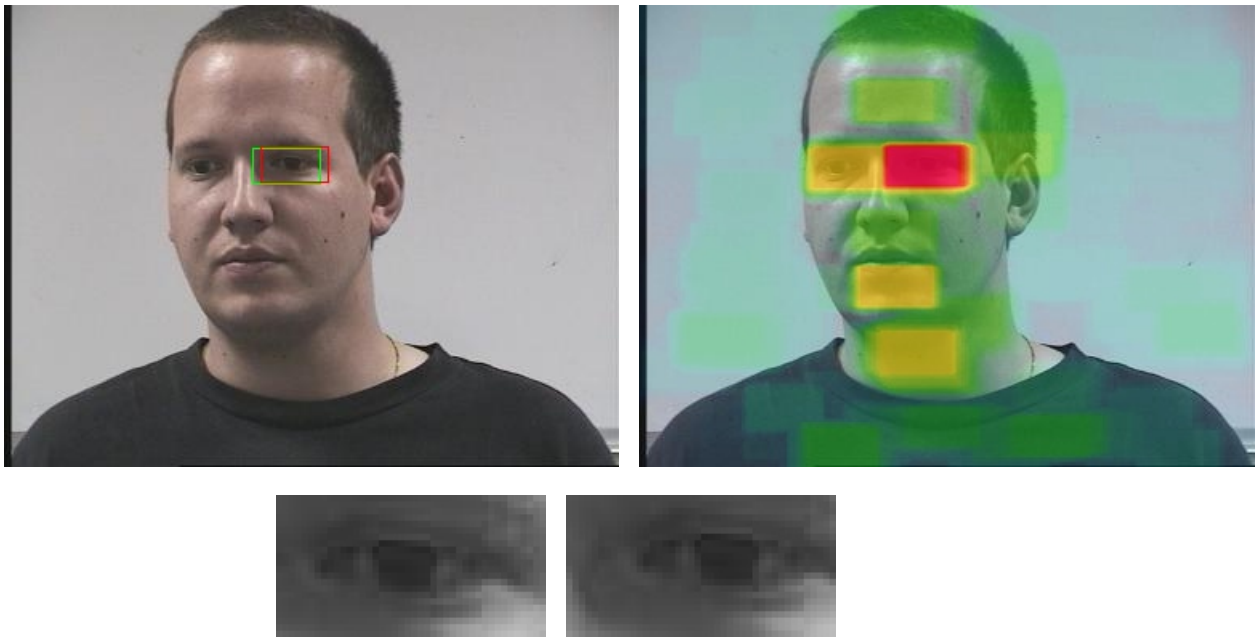

Figura 4.17: a) A janela de resultado para o rosto em $tilt = -15$, $pan = +15$; b) Resposta do neurônio RWiSARD; c) Imagem da janela respectiva ao ground truth; d) Imagem encontrada pela rede RWiSARD.

As figuras 4.18 e 4.19 mostram respectivamente o gráfico de resposta das redes RWiSARD e WiSARD binária para a busca pelo olho direito. Os valores são apresentados no gráfico codificados em cores, e tanto a escala de valores como a codificação em cores empregada são os mesmos para ambos os gráficos. Os valores dos gráficos são interpolados linearmente a partir dos valores registrados nos pontos que representam as imagens usadas nos testes.

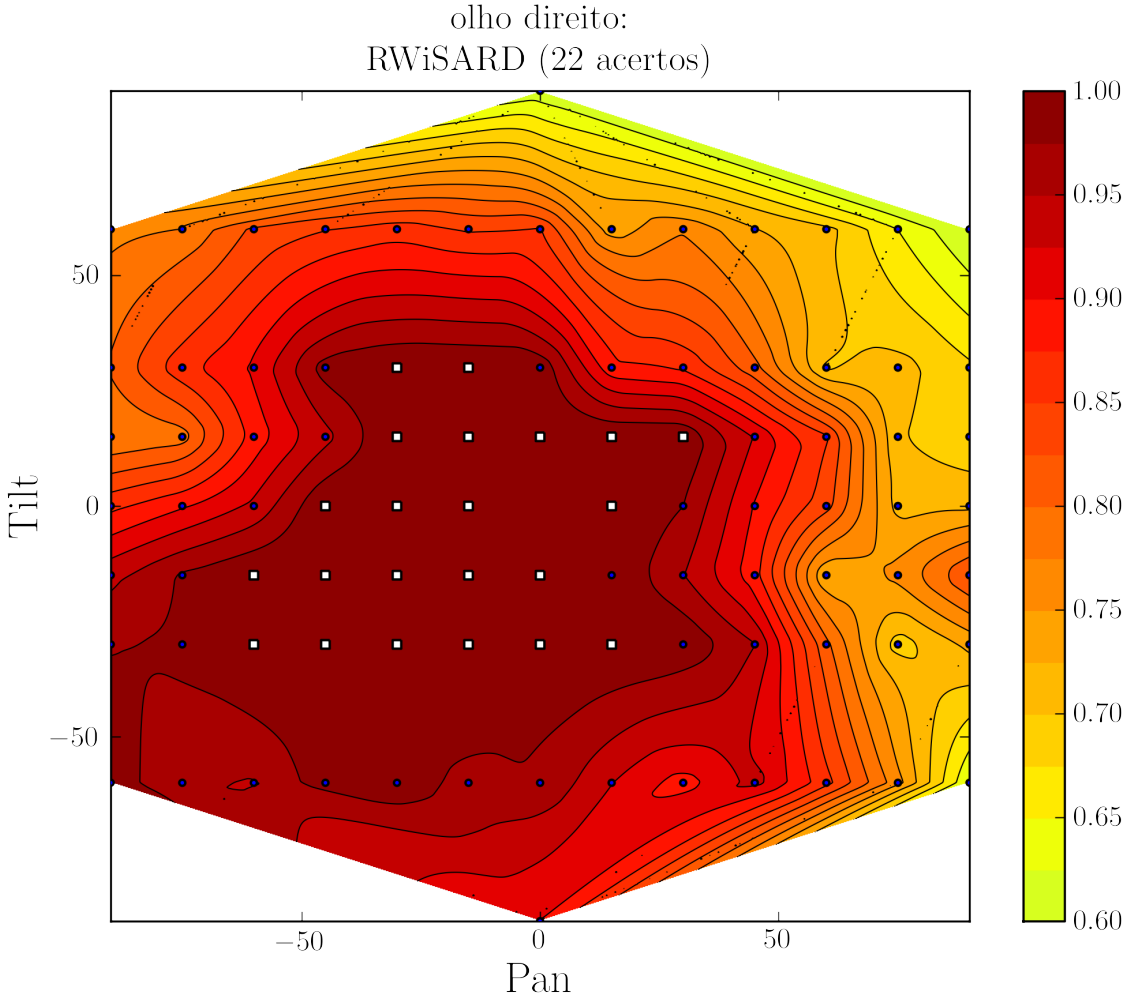

Figura 4.18: Gráfico de respostas para o olho direito - RWiSARD.

Fica evidente a larga vantagem no emprego da rede RWiSARD em detrimento da rede WiSARD binária ao se observar o número de imagens de teste onde foram encontrados os ground truth. Isso também indica que a rede RWiSARD apresenta resultados satisfatórios mesmo diante de ângulo do rosto.

Um segundo aspecto relevante é o valor registrado para as demais imagens empregadas nos ensaios. A maioria das imagens obteve como resposta relativa para os seus ground truth $v_j \geq 0,85$ no caso da rede RWiSARD. No caso da rede WiSARD binária, a maioria dos ground truth obteve uma resposta abaixo de 0,75, o que representa valores bem abaixo do resultado máximo encontrado em cada imagem de

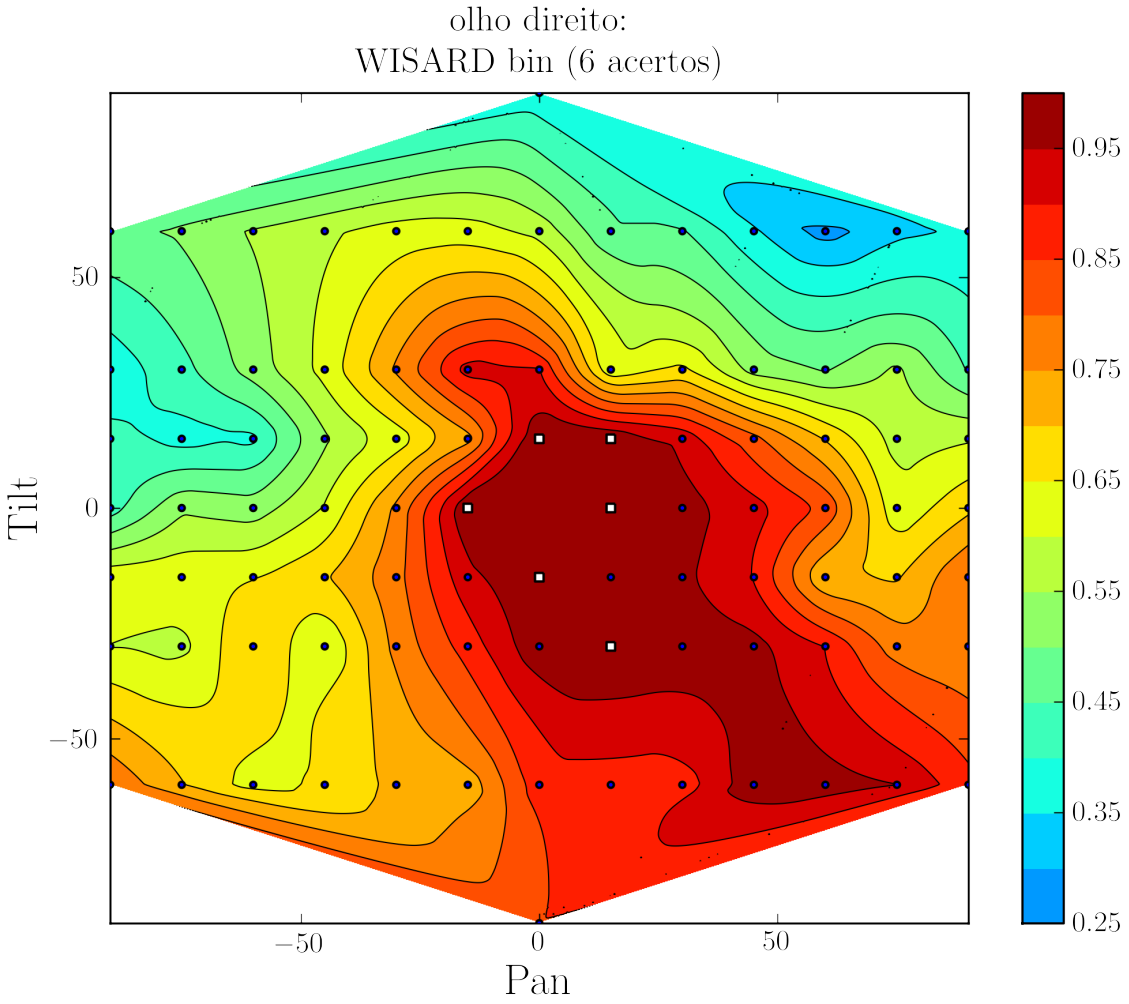

Figura 4.19: Gráfico de respostas para o olho direito -WiSARD binária.

busca.

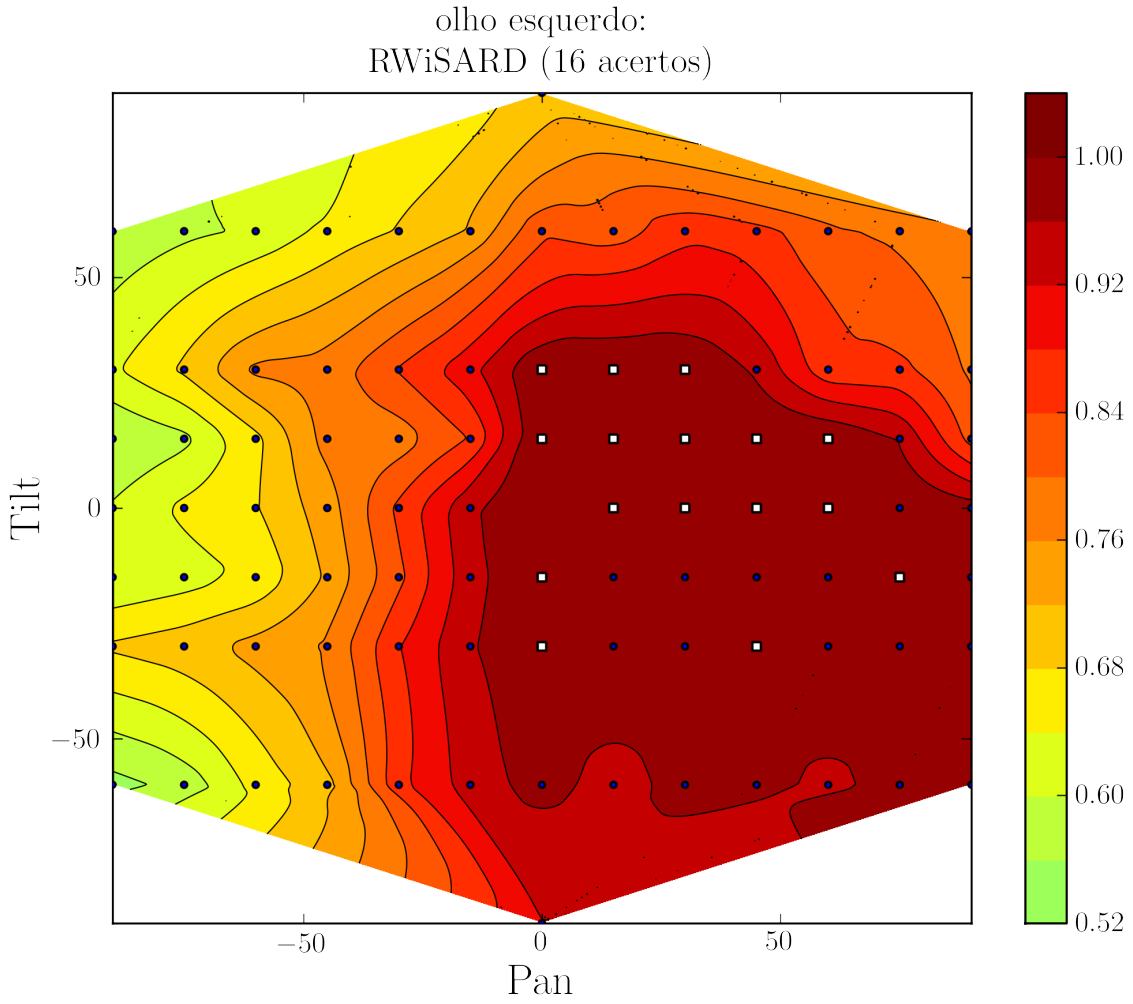

Figura 4.20: Gráfico de respostas para o olho esquerdo - RWiSARD.

Os gráficos das figuras 4.20 e 4.21 mostram as respostas das redes RWiSARD e WiSARD binárias para o olho esquerdo.Mais uma vez a rede RWiSARD se mostra mais eficaz, encontrando corretamente o olho esquerdo em pelo menos 16 imagens contra 8 da rede WiSARD binária.

Entre os olhos

A figura 4.22 exibe imagens referentes ao ground truth da região localizada entre os olhos (ponte nasal), para algumas das imagens de busca.

As figuras 4.23 e 4.24 mostram os gráficos para as respostas das redes RWiSARD e WiSARD binárias respectivamente para a busca pela região entre os olhos. A rede RWiSARD logrou determinar a posição dessa estrutura facial em 41 das imagens de busca, enquanto a rede WiSARD binária só encontrou de fato as mesmas regiões em 13 imagens de busca.

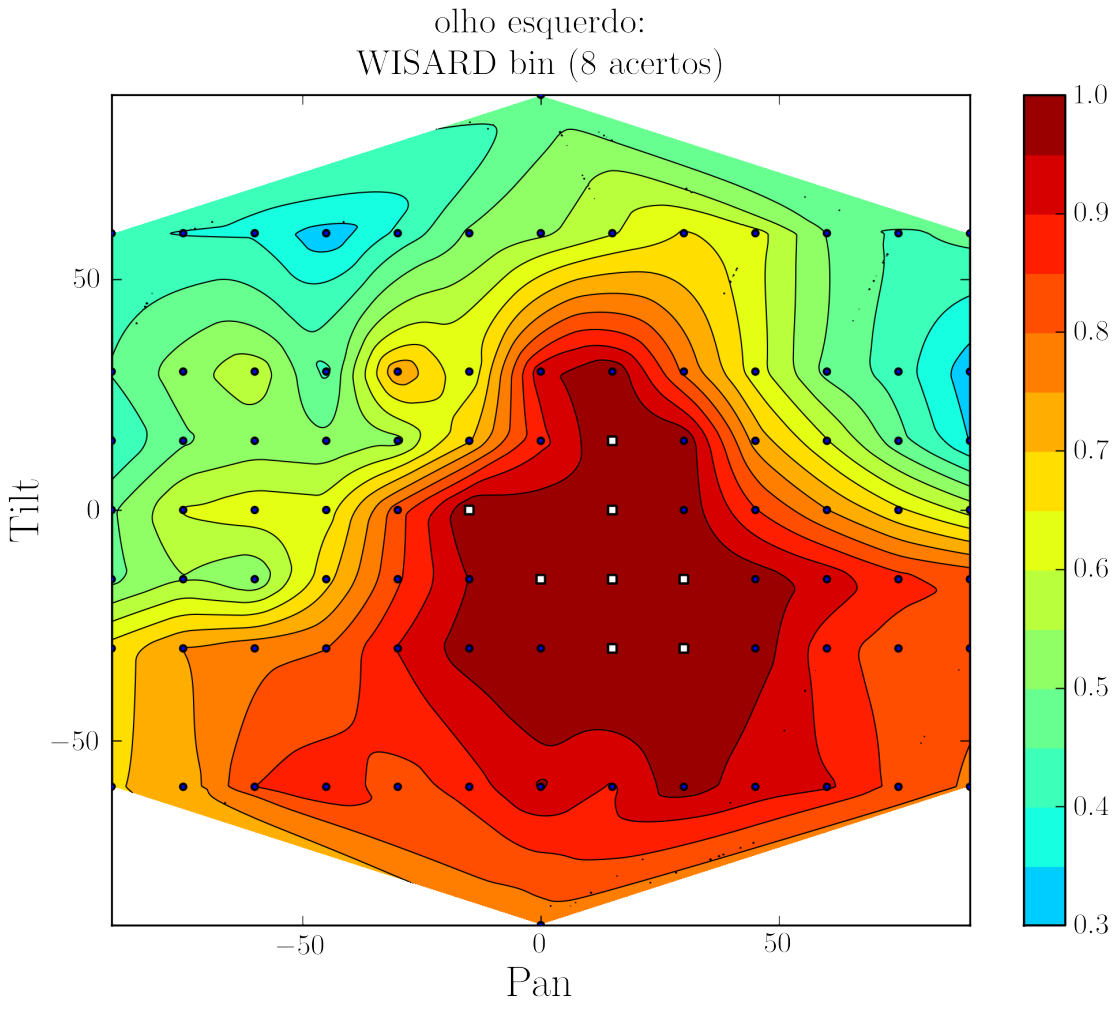

Figura 4.21: Gráfico de respostas para o olho esquerdo -WiSARD binária.

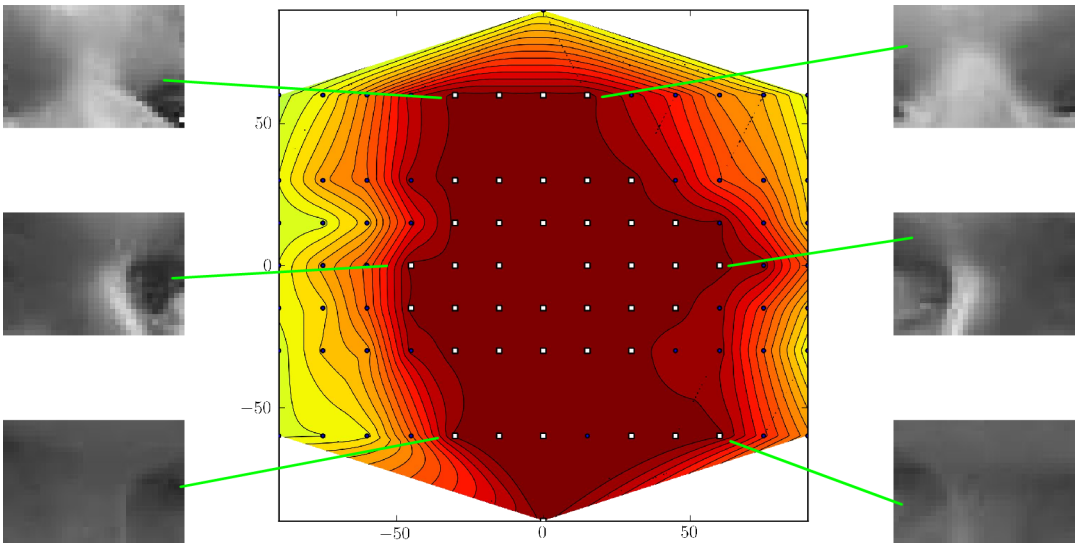

Figura 4.22: Exemplos extraídos do banco de faces utilizado.

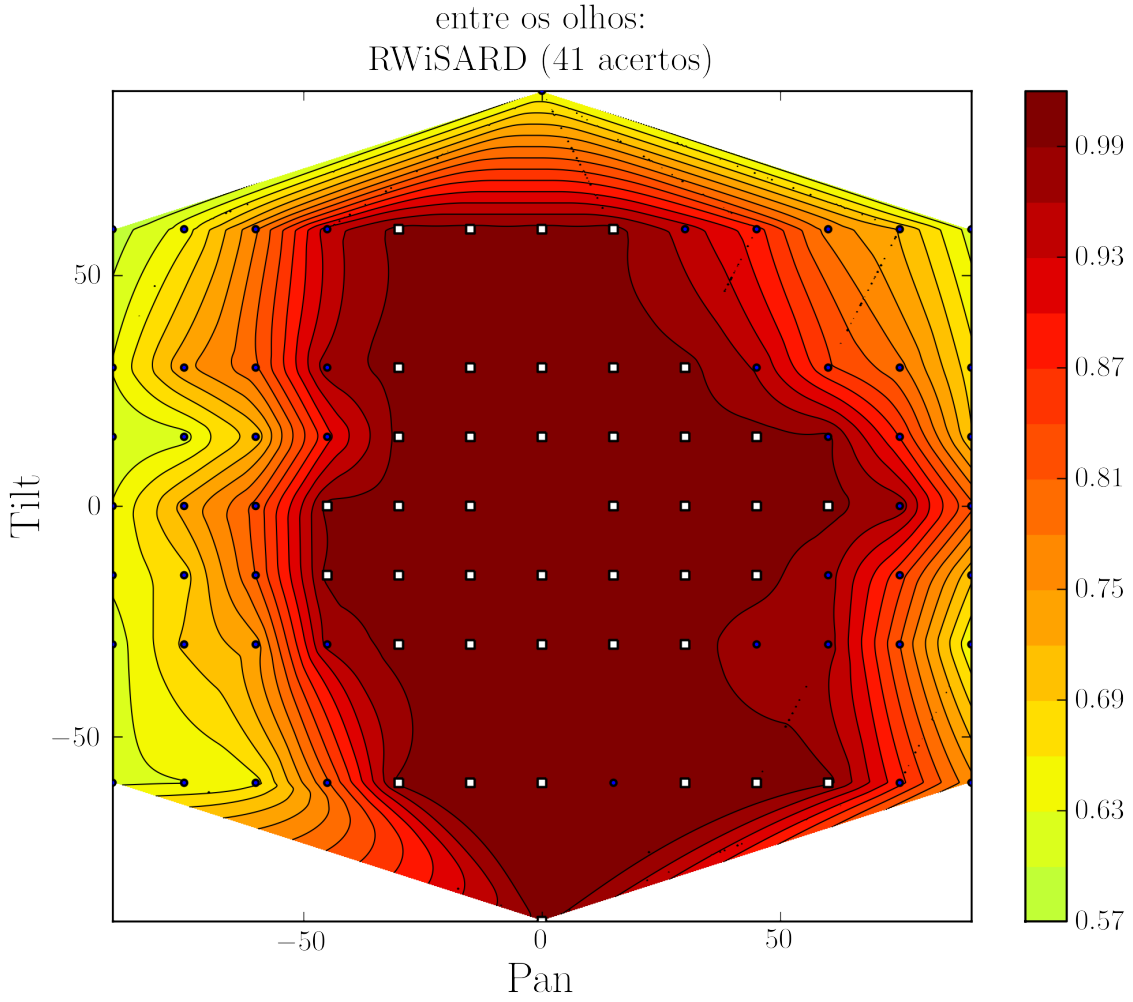

Figura 4.23: Exemplos extraídos do banco de faces utilizado.

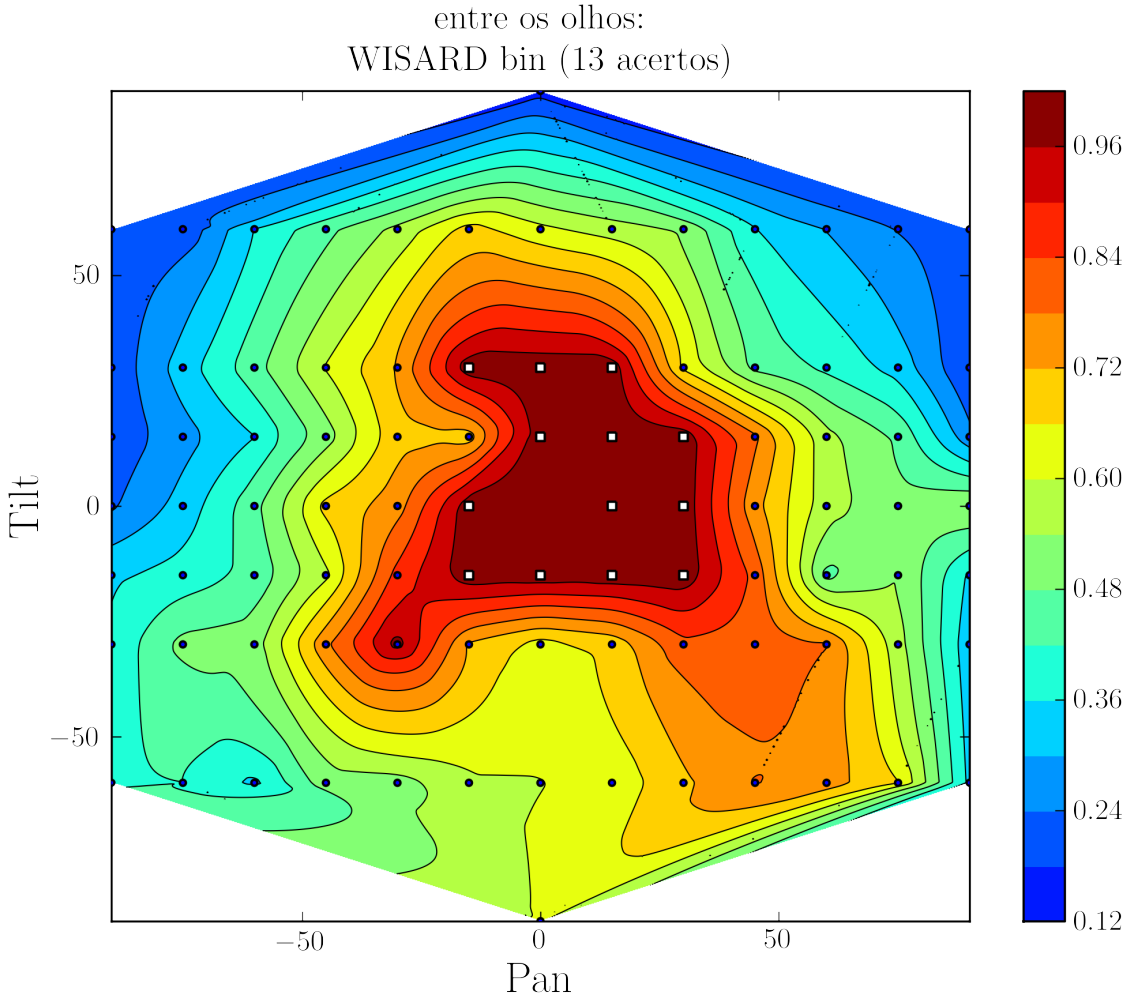

Figura 4.24: Exemplos extraídos do banco de faces utilizado.

Nariz

A figura 4.25 destaca algumas das imagens encontradas pela aplicação do método na busca do nariz. Um aspecto particular dessa estrutura é o grande número de superfícies que ficam ocultas sob diversos ângulos de visão diferentes dada a natureza côncava de sua geometria.

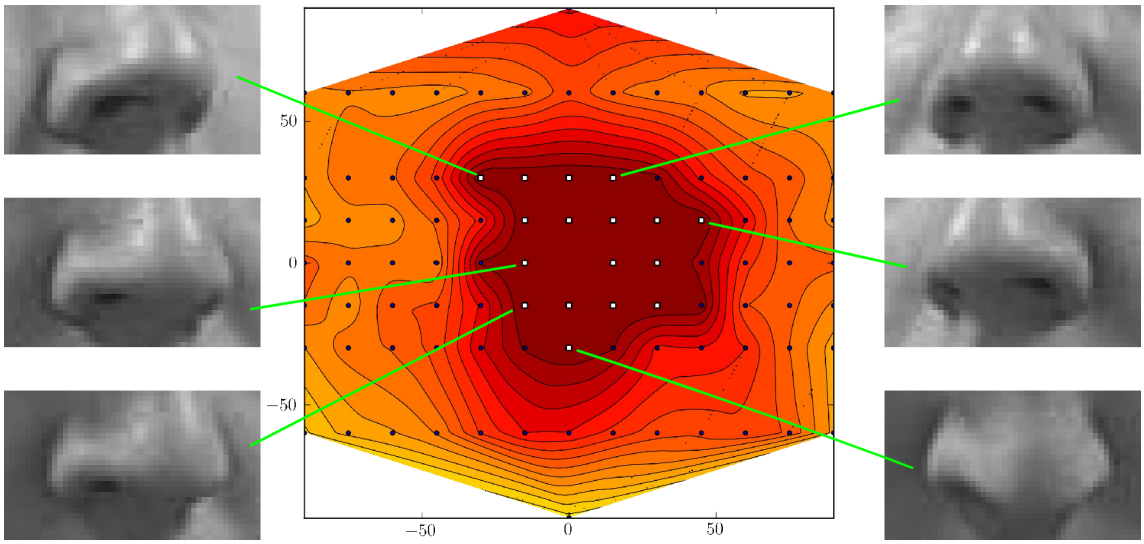

Figura 4.25: Exemplos extraídos do banco de faces utilizado.

As figuras 4.26 e 4.27 mostram, respectivamente os gráficos de resposta das redes RWiSARD e WiSARD binária, para a busca pelo nariz.

Boca

A última das estruturas faciais que foram objetos desse experimentos é a boca. Diferente do nariz e da ponte nasal, e provavelmente em maior escala do que os olhos, a boca é um dos principais responsáveis pela capacidade humana de esboçar expressões. Certamente é a característica do rosto de maior mobilidade e elasticidade, se revelando um dos principais desafios dos processos de reconhecimento facial e de expressões. Para este experimento, procurou se limitar a gama de variações da imagem do rosto à sua inclinação (tilt e pan), então o tempo todo foram empregadas imagens do rosto em uma expressão neutra, e devido a isso a boca se mostra aqui como uma estrutura estática, permanentemente fechada. Alguns exemplos das imagens encontradas para a boca podem ser vistos na figura 4.28, onde também é mostrada a localização das respectivas imagens de busca no plano do gráfico $tilt \times pan$.

Enfim, nas figuras 4.29 e 4.30 podem ser vistos respectivamente os gráficos de resposta para as redes RWiSARD e WiSARD binária. Mais uma vez fica evidente a maior eficácia da rede RWiSARD na resolução desse problema, tendo ela encontrado

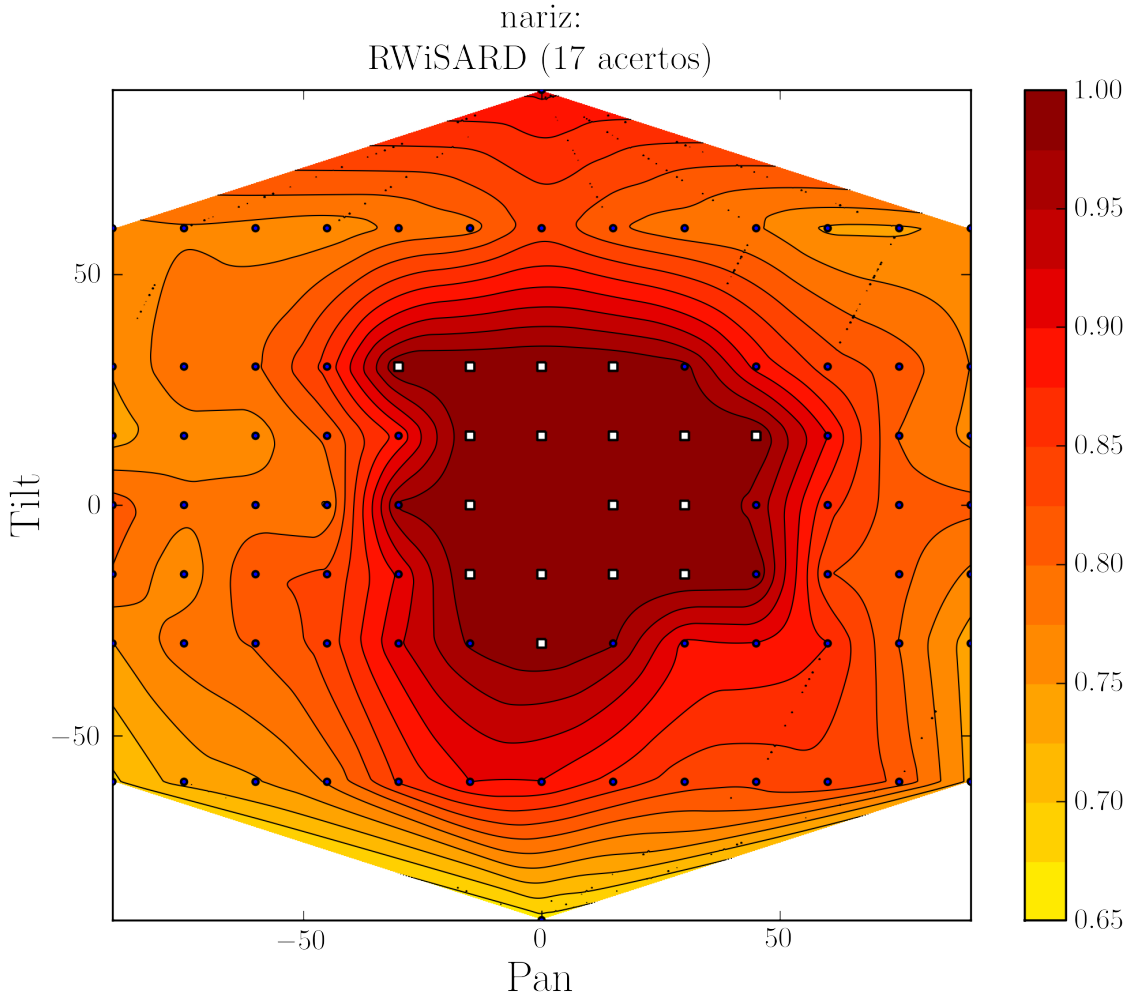

Figura 4.26: Exemplos extraídos do banco de faces utilizado.

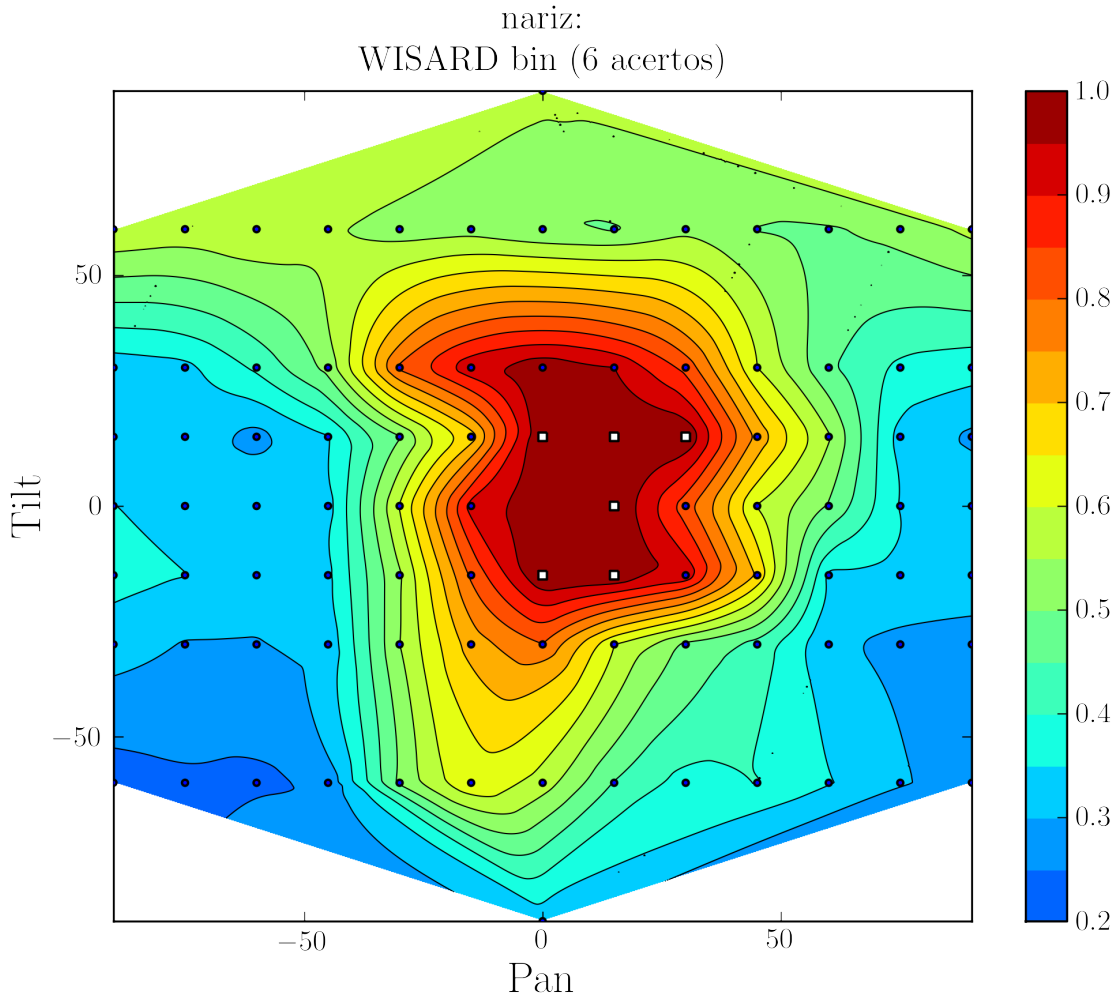

Figura 4.27: Exemplos extraídos do banco de faces utilizado.

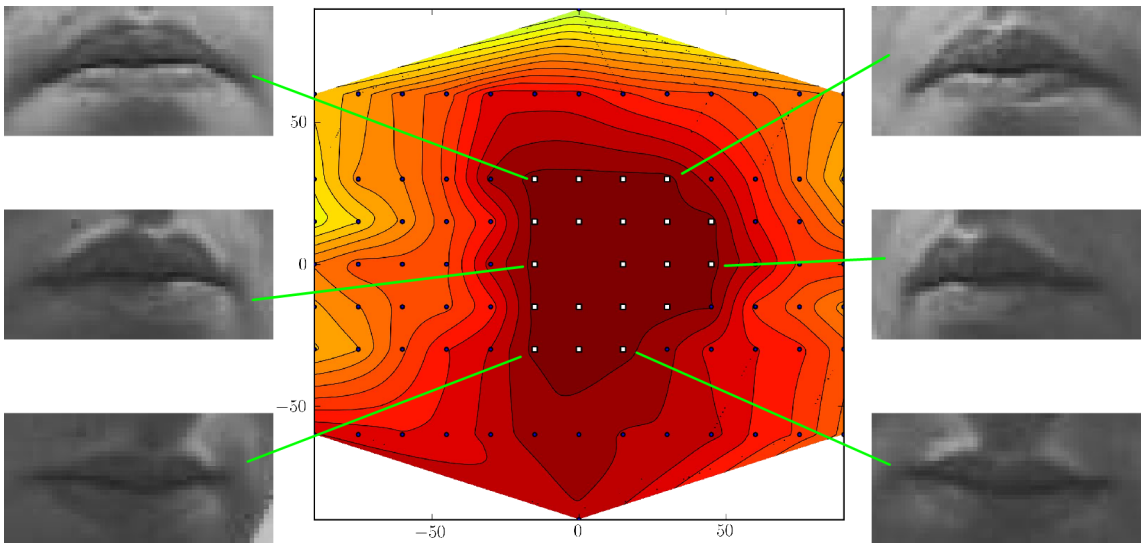

Figura 4.28: Exemplos extraídos do banco de faces utilizado.

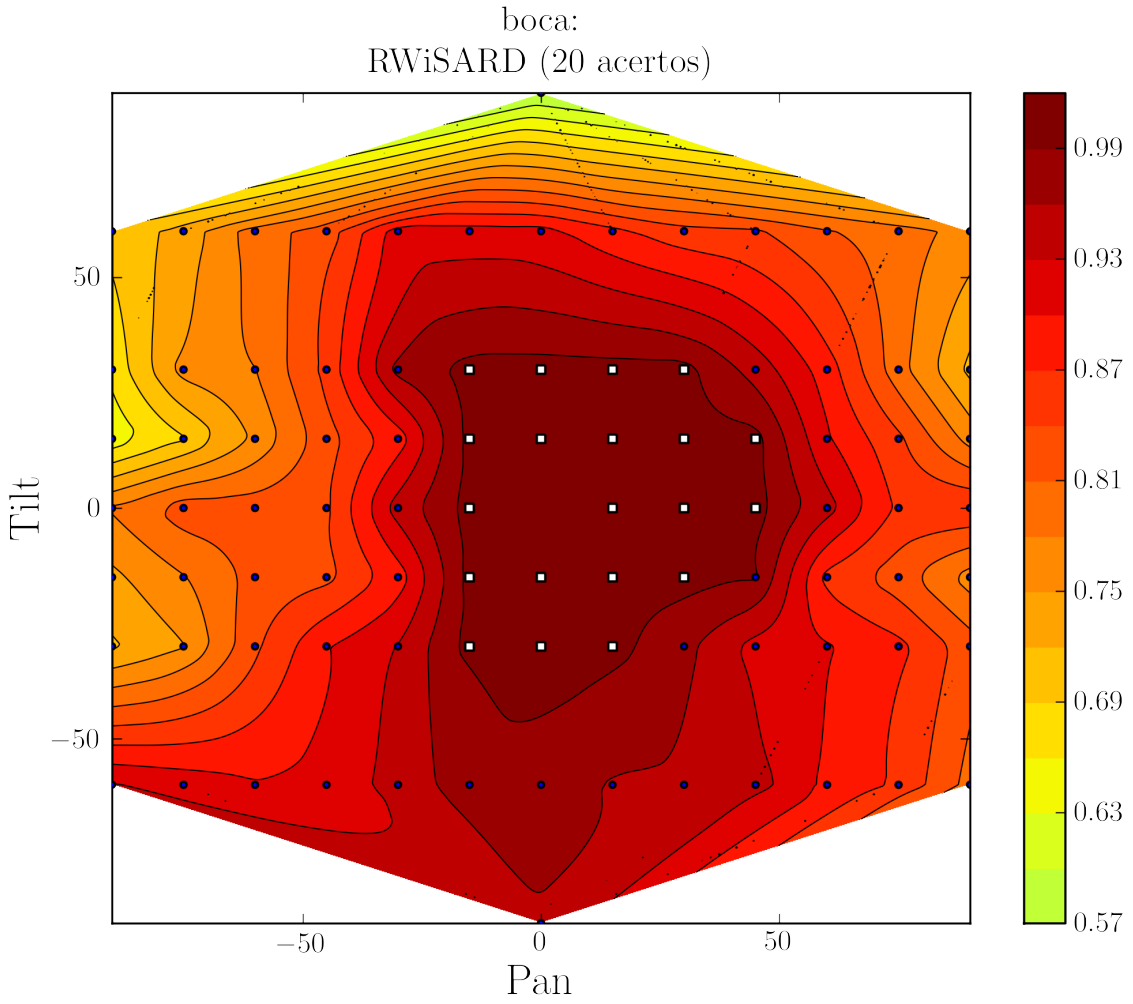

Figura 4.29: Exemplos extraídos do banco de faces utilizado.

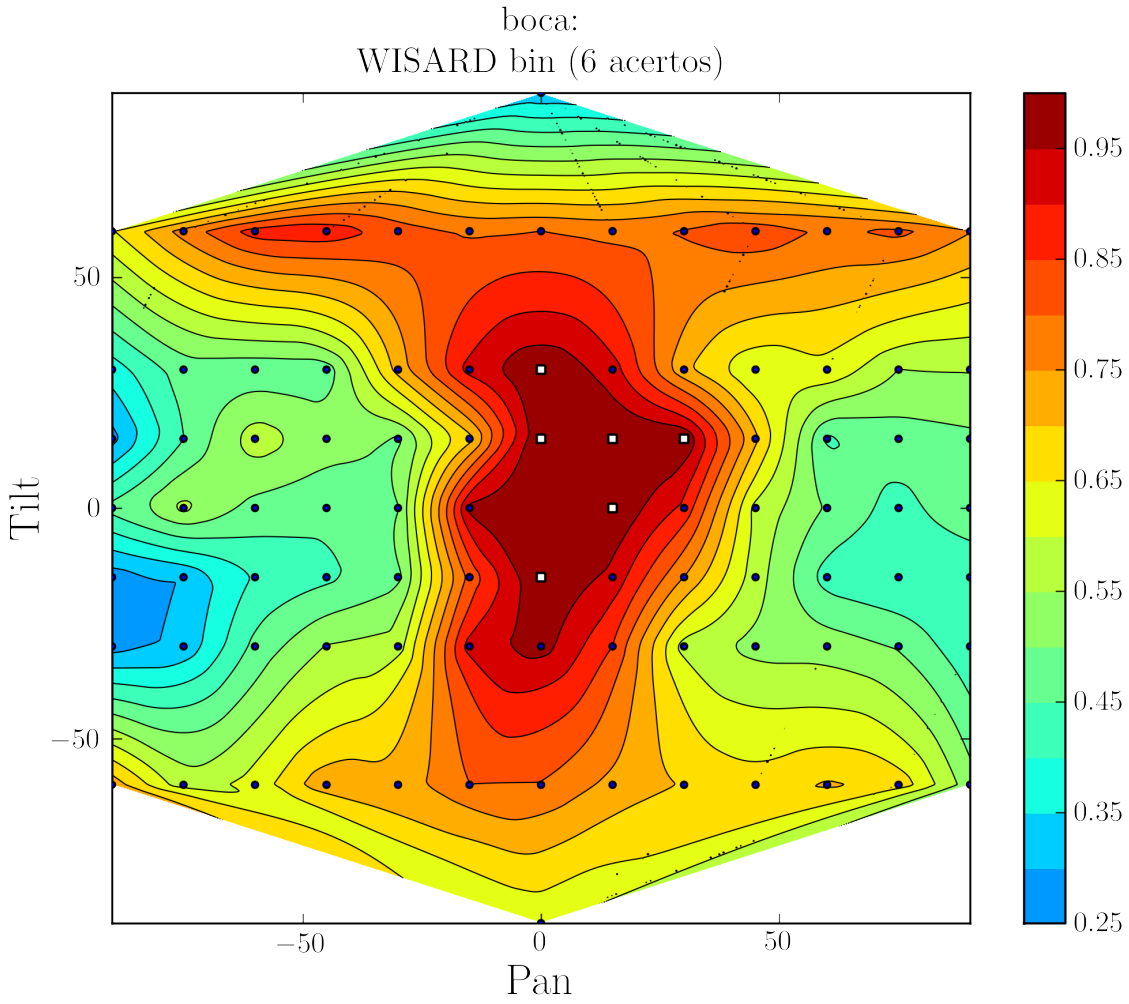

Figura 4.30: Exemplos extraídos do banco de faces utilizado.

de maneira correta a boca em 20 imagens de busca, contra apenas 6 da rede binária.

4.4 Análise dos Resultados

4.4.1 Crítica

Conforme apresentado no gráfico na figura 4.3 apresenta uma relevante margem de 12% de vantagem entre a rede RWiSARD e a rede WiSARD binária, em seu emprego. O processo adotado neste experimento não leva em condição nenhum mapeamento ou restrição de ordem biométrica, ou seja, apenas a semelhança entre as imagens é de fato levada em conta.

A tabela 4.1 sumariza os resultados exibidos para o experimento de detecção de características faciais. Para cada item exibido, os resultados para a rede RWiSARD e para a rede WiSARD binária são mostrados, respectivamente, separados por hífen. A primeira coluna, N^o de acertos, indica a quantidade de ensaios onde o Ground Truth foi identificado corretamente pelo algoritmo. A segunda coluna exibe a porcentagem de casos em a resposta do neurônio para o Ground Truth foi maior que 99%, da resposta mais elevada. Uma abordagem mais robusta para a detecção de característica pode levar em conta tal resultado, por exemplo, para limtar o espaço de busca para regiões que conseguiram melhor resposta. A terceira e última coluna mostra a porcentagem de casos em que a resposta para o Ground Truth foi por outro lado mais baixa que 50% da resposta mais alta. Do contrário das duas outras colunas, o valor maior é considerado o pior valor. Essa tabela também deixa evidente a vantagem em se empregar a rede RWiSARD em tal aplicação.

4.4.2 Limitações do modelo

Em alguns ensaios realizados, a rede RWiSARD identificou algunas imagens como similares ao protótipo apresentado quando visualmente suas aparências não poderiam ser classificadas como menos do que destoantes. Ainda que seja esperada uma limitação desse modelo artificial de reconhecimento quando comparado à mais elaborada capacidade humana de perceber padrões e imagens similares, alguns desses

Tabela 4.1: Detecção de Características Faciais (RWiSARD - WiSARD)

Estrutura facial	N^o de acertos	$r_{GT} > 99\%$	$r_{GT} < 50\%$
Olho Direito	22 - 6	0.50 - 0.22	0.00 - 0.17
Olho Esquerdo	16 - 8	0.47 - 0.22	0.00 - 0.22
Entre os olhos	41 - 13	0.57 - 0.15	0.00 - 0.39
Nariz	17 - 6	0.30 - 0.12	0.00 - 0.54
Boca	20 - 6	0.48 - 0.13	0.00 - 0.34

episódios de confusão por parte da rede podem ter sua causa rastreada até a estrutura matemática desse modelo.

Conforme visto anteriormente o valor de resposta de um neurônio RWiSARD diante de uma n-upla de valores de luminância depende do total das respostas isoladas para cada vetor binário. O valor de resposta para cada vetor binário por sua vez depende de que o endereço correspondente tenha sido atualizado durante o treinamento, e, se for o caso, da diferença entre o limiar que gerou esse vetor binário e o conteúdo guardado por este endereço.

Como uma consequência direta do método de decomposição em vetores binários, o vetor $b_n = (1, 1, 1 \cdots 1)$ sempre será gerado pela decomposição de qualquer n-upla, e terá como limiar o menor valor dentre todos os pixels da n-upla. Então sempre haverá registro no endereço apontado por $(1, 1, 1 \cdots 1)$ em decorrência da decomposição dos padrões em níveis de cinza aprensentados à rede durante o treinamento, bem como qualquer padrão apresentado posteriormente para classificação e reconhecimento sempre irá produzir um vetor binário $b_n = (1, 1, 1 \cdots 1)$ quando decomposto.

Um efeito colateral deste fato, é de que não importa com quais padrões de níveis de cinza a rede RWiSARD tenha sido treinada - uma upla que tenha todos os pixels com o mesmo valor irá produzir apenas o vetor $b_n = (1, 1, 1 \cdots 1)$, e assim sempre encontrará o endereço correspondente com o contador diferente de zero. A resposta da rede RWiSARD a essa upla de pixels de mesmo valor dependerá apenas do valor desses pixels, e do valor registrado no endereço $(1, 1, 1 \cdots 1)$ do nó RAM da upla. Essa resposta sempre será alta quando a diferença entre esses valores for pequena.

Figura 4.31: Imagem usada no treinamento.

O exemplo a seguir ilustra o quanto esse efeito colateral pode ser prejudicial à operação da rede. Um neurôrio foi treinado apenas com uma 5-upla de pixels com os valores $(0.125, 0.25, 0.25, 0.725, 0.125)$. Essa upla representa um mapeamento aleatório dos pixels da imagem na figura 4.31. A seguir, ao mesmo neurônio são apresentadas para classificação as imagens cujas respectivas uplas são: $((0.125, 0.125, 0.125, 0.125, 0.125)$, $(0.136, 0.25, 0.25, 0.725, 0.125)$, $(0.125, 0.25, 0.25, 0.725, 0.125)$ e $(0.125, 0.125, 0.125, 0.725, 0.125)$. O neurônio respondeu a essas imagens com, respectivamente: 0.0, 1.0, 0.0, 0.0. Essas imagens são exibidas na figura 4.32, juntamente com suas respostas codificadas com cores: azul para respostas mais baixas (ou seja, mais distantes de 0.0), até vermelho para as mais altas (mais próximas de 0.0) passando por amarelo.

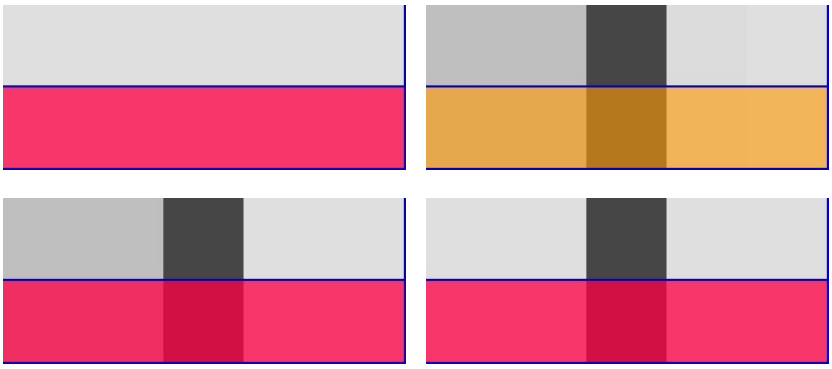

Figura 4.32: Imagens usadas para classificação acompanhadas das respostas do neurônio;

A imagem à esquerda na primeira fileira da figura 4.32, cujos pixels tem todos eles o mesmo valor 0.125, gera apenas o vetor binário $(1, 1, 1, 1, 1)$, associado com o valor 0.125. O neurônio em questão tem neste registrado neste endereço o valor 0.125. Logo a resposta para essa imagem é 0, a mais alta possível. Visualmente essa imagem não guarda qualquer semelhança com a imagem usada para o treinamento a não ser pelo valor de seus pixels de valor mais baixo. Por outro lado a imagem à direita, na segunda fileira, ao ter sua upla $(0.136, 0.25, 0.25, 0.725, 0.125)$ decomposta, gera, entre outros vetores binários, o vetor $(1, 1, 1, 1, 0)$, que aponta para um endereço não atualizado durante o treinamento, e em decorrência disso é penalizado com uma resposta mais baixa do que a imagem da upla $((0.125, 0.125, 0.125, 0.125, 0.125)$. Quer seja por uma inspeção visual, quer seja através de uma comparação entre seus pixels um a um, é notório que a imagem que sofreu maior penalidade tem uma similaridade muito maior com a imagem de treinamento do que a imagem que tem todos os pixels com o mesmo valor. E ainda assim foi apontado pelo neurônio RWiSARD como mais distante, devido a uma pequena diferença entre os valores de um de seus pixels e sua contraparte na imagem de treinamento.

Tal sintoma do modelo, em imagens maiores, com diversas uplas, pode levar igualmente a resultados equivocados.

Capítulo 5

Conclusões

5.1 Objetivos e Resultados

No primeiro capítulo foi citado como objetivo apresentar um modelo de rede neural baseado no modelo WiSARD, que o complementasse provendo capacidade de lidar diretamente com imagens em escala de tons de cinza de forma a melhorar seus resultados em aplicações que envolvem esse tipo de imagem. Os resultados apresentados pelos experimentos descritos nesse trabalho exibem números significativos que indicam que a qualidade da resposta oferecida por esta rede em tais situações melhorou substancialmente. Esse acréscimo de qualidade proporcionado pelo modelo RWiSARD pode e será observado mesmo quando aplicado em diferentes bancos de imagens, sempre que a natureza da aplicação em si prover a oportunidade para seu emprego.

Diante de um primeiro contato com esse modelo alguém poderia indagar se, diante das vantagens oferecidas, poderia sempre se substituir integralmente redes clássicas WiSARD pelo novo modelo, com lucros. A reposta é Não.

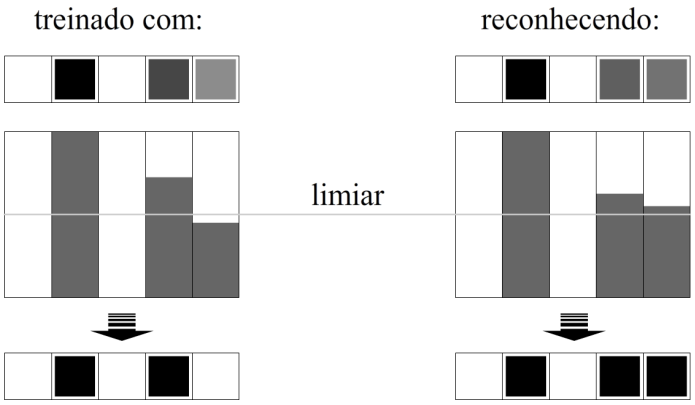

Figura 5.1: Imagens similares, vetores binários diferentes.

Os experimentos aqui descritos e detalhados tem aqui o traço em comum de

serem todos exemplos de aplicações onde o objeto se trata de imagens em vários tons de cinza, e onde a aplicação de um único limiar arbitrário ou fixo não seria capaz de capturar de forma adequada toda a informação necessária para uma precisa classificação dos dados inerentes à imagem. Em um panorama como esse, ainda que seja possível se empregar redes WiSARD binárias dispostas em camadas com diferentes limiares, ainda ocorrerá perda de informação relevante pela rigidez imposta pela binarização da imagem. Um exemplo prático de uma situações onde isso ocorre pode ser visto na figura 5.1.

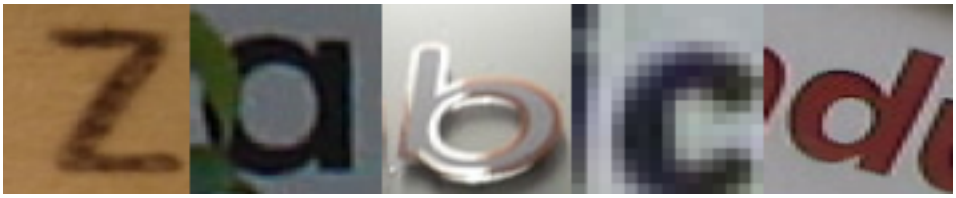

Figura 5.2: Imagens de caracter extraídas de fotografia.

Já em uma situação diferente, como por exemplo no caso de um sistema de reconhecimento de caracteres, a aplicação direta de RWiSARD nas imagens de caracteres pode levar à introdução de informação que seria normalmente considerada ruído. Tal ruído teria sido naturalmente eliminado atráves da aplicação de um limiar bem definido, e toda a informação relevante para o correto reconhecimento dos caracteres pode ser traduzida em pixels pretos e pixels brancos. Mesmo em casos como os mostrados na figura 5.2, extraídos do banco de imagens naturais de caracteres descrito em DeCAMPOS [12], onde as imagens não são originalmente binárias, o emprego do modelo RWiSARD, além de não trazer qualquer vantagem adicional, permite que o ruído seja assimilado como parte do padrão, incorrendo em maior índice de erros na classificação.

Ainda assim, nada impede que métodos adaptativos para se determinar o melhor limar a ser utilizado, ou como delimitar regiões de interessem em imagens ou ainda um método para se ajustar filtros de forma a se obter a melhor imagem para posterior processamento, permitam combinar redes RWiSARD e WiSARD em um mesmo sistema, em etapas complementares. Em uma circunstância em que sejam mapeadas características da imagem onde o uso de limiares sejam mais convenientes, um método híbrido que combine o uso de redes RWiSARD e WiSARD clássica também pode levar a resultados muito mais expressivos do que os resultados proporcionados por qualquer uma das duas redes sozinha.

De qualquer forma, diante dos resultados apresentados nesse trabalho, é razoável se assumir que seu objetivo foi cumprido.

5.2 Bancos de Imagens empregados

O banco de faces empregado no experimento de reconhecimento facial é um banco conhecido na pesquisa em Visão Computacional, tendo sido utilizado como banco de provas para alguns algoritmos conhecidos. Ainda assim, não há um consenso sobre se esse, ou qualquer outro banco de faces disponível, é o mais indicado para servir de base de comparação entre os diversos algoritmos e métodos de reconhecimento disponíveis. De um modo geral, a forma como os rostos dos indivíduos são retratados, as condições das imagens, a proporção da imagem efetivamente ocupada pelo rosto, além de aspectos relacionados a condições que podem ser encaradas como circunstanciais para essas imagens, tais como a presença ou não de óculos, o penteado, maquiagem, barba ou bigode entre outros, além de variáveis inerentes às pessoas presentes nas imagens, tais como etnia, cor da pele, etc. geram um espaço de possibilidades para as imagens de faces que nenhum banco de faces jamais irá cobrir. Isso permite que se afirme que os resultados colhidos de qualquer experimento, com qualquer técnica possível, seja em grande parte uma propriedade também do banco de faces empregado. Um outro banco de faces, com fotografias tiradas em circunstâncias diferentes, de pessoas diferentes, pode surtir um resultado consideravelmente diferente.

Dessa forma, em nenhuma etapa do experimento qualquer característica da imagem que não fosse o tamanho da imagem excerceu maior influência quanto às decisões tomadas. Nenhum processo de segmentação ou extração de características foi empregado, sacrificando um provável melhor aproveitamento da informação inerente às imagens em prol de uma comparação mais justa entre o modelo clássico e o proposto, bem como também em prol de uma maior generalidade dos resultados obtidos em si.

Quanto ao experimento de detecção de estruturas faciais, o banco de faces empregado foi a melhor opção disponível dada a escassez de bancos de imagens voltados pra esse tipo de teste. Com imagens devidamente rotuladas, e apresentando rostos que mudem de angulo segundo um padrão regular, permitiu fazer uma análise satisfatória dos algoritmos diante de mudanças de ângulo bem comportadas, facilitando inclusive indexar os resultados em seguida.

Ainda assim, esse banco de dados não aborda o problema da variação de ângulo ou intensidade da iluminação sobre os rostos. Ao tempo em que esse trabalho foi escrito, nenhum banco de faces foi encontrado que abordasse esse aspecto do problema de uma forma tão organizada e conveniente quanto o banco de dados usado no experimento o fez para o problema da mudança de ângulo do rosto.

5.3 Trabalhos Futuros e Oportunidades

No capítulo 4 foi levantado um problema colateral do modelo RWiSARD que pesa negativamente em alguns de seus resultados. Alguma informação ainda é perdida quando a n-upla é convertida em n uplas binárias e nenhuma informação que relacione essas uplas entre si é fornecida ao neurônio, o que tornaria, por si só, todo o modelo inviável. Uma possível solução poderia consistir em adicionar mais um campo à estrutura de dados mantida nos endereços do Nó RAM, talvez mantendo informação sobre a participação da respectiva upla binária na upla em escala de cinza original, e em introduzir esse novo campo no cálculo de resposta da upla. Em teoria, tal alteração corrigiria qualquer atribuição "injusta"na aferição da similaridade e surtiria automaticamente em grande melhoria dos resultados.

Quanto à implementação do modelo, devido ao modo como as respostas das uplas são calculadas de forma independentemente, é possível se implementar o modelo em arquiteturas trivialmente paralelas, tais como arquiteturas voltadas para GPGPU. Dada a sua natureza matematicamente simples e linear, é perfeitamente plausível sua implementação direta em um microcircuito dedicado, talvez em FPGA (Field Programmable Gate Array), a um baixo custo e com promessas de alta performance, permitindo sua aplicação em sistemas de tempo real, mesmo as que requerem baixa latência.

Enfim, abordagens mais robustas do que as apresentadas nesses experimentos podem lançar mão do modelo RWiSARD combinado com outras estratégias bem sucedidas em visão computacional, como é o caso do método Viola-Jones ([13]). Uma versão do meta-algoritmo Adaboost com as modificações apresentadas nesse trabalho poderia se beneficiar da velocidade de treinamento e classificação oferecida pelo modelo RWiSARD, usando este modelo como base para os "Classificadores Fracos"do comitê de classificadores. Uma versão alterada do modelo RWiSARD permitiria ainda atribuir peso às uplas, acomodando assim o "boosting por características"descrito no artigo. Dado o sucesso do método Viola-Jones e sua notável versatilidade, esta também se figura como uma promissora oportunidade para futuros trabalhos.

As sugestões que foram mencionadas acima ainda só terão a sua validade confirmada mediante futuros trabalhos, mas ao menos servem para deixar claro que a pesquisa relacionada a esse novo modelo está longe de se encerrar nesse trabalho, bem como estão longe de se esgotar aqui, todas as possibilidades para a rede neural RWiSARD. Espera-se que mentes mais brilhantes e de imaginação mais fértil venham a dar a devida continuidade à pesquisa em torno da qual girou esta tese, permitindo assim que esta possa cumprir com o seu real objetivo, que não é nada mais do que começar algo novo.

Referências Bibliográficas

[1] TURK, M., PENTLAND, A. "Eigenfaces for recognition", *J. Cognitive Neuroscience*, v. 3, pp. 71–86, January 1991. ISSN: 0898-929X. doi: 10.1162/jocn.1991.3.1.71. Disponível em: ¡http://portal.acm.org/citation.cfm?id=1326887.1326894¿.

[2] ADINI, Y., MOSES, Y., ULLMAN, S. "Face Recognition: The Problem of Compensating for Changes in Illumination Direction", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 19, pp. 721–732, 1997. ISSN: 0162-8828. doi: http://doi.ieeecomputersociety.org/10.1109/34.598229.

[3] PURVES, D. "Neuroscience", *Scholarpedia*, v. 4, n. 8, pp. 7204, 2009. Disponível em: ¡http://www.scholarpedia.org/article/Neuroscience¿.

[4] MCCULLOCH, W., PITTS, W. "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biology*, v. 5, pp. 115–133, 1943. ISSN: 0092-8240. Disponível em: ¡http://dx.doi.org/10.1007/BF02478259¿. 10.1007/BF02478259.

[5] WICKERT, I., FRANCA, F. M. G. "AUTOWISARD: Unsupervised Modes for the WISARD". , 2001.

[6] BLEDSOE, W. W., BROWNING, I. "Pattern Recognition and Reading By Machine", *Managing Requirements Knowledge, International Workshop on*, v. 0, pp. 225, 1959. doi: http://doi.ieeecomputersociety.org/10.1109/AFIPS.1959.88.

[7] AUSTIN, J. *RAM-Based Neural Networks*. River Edge, NJ, USA, World Scientific Publishing Co., Inc., 1998. ISBN: 9810232535.

[8] GRIECO, B. P., LIMA, P. M., GREGORIO, M. D., et al. "Producing pattern examples from mental images", *Neurocomputing*, v. 73, n. 7-9, pp. 1057 – 1064, 2010. ISSN: 0925-2312. doi: DOI:10.1016/j.neucom.2009.11.015. Disponível em: ¡http://www.sciencedirect.com/science/article/

B6V10-4Y6S7JF-1/2/417af53c8e068cdff872ce8e44678b95¿. Advances in Computational Intelligence and Learning - 17th European Symposium on Artificial Neural Networks 2009, 17th European Symposium on Artificial Neural Networks 2009.

[9] LUCAS, S. M. "Face Recognition with the continuous n-tuple classifier". In: *In Proceedings of the British Machine Vision Conference*, pp. 222–231, 1997.

[10] AUSTIN, J. "Grey Scale N Tuple Processing". In: *In Pattern Recognition: 4th International Conference*, pp. 110–120. Springer-Verlag, 1988.

[11] GOURIER, N., HALL, D., CROWLEY, J. L. "Estimating Face Orientation from Robust Detection of Salient Facial Features". In: *Proceedings of Pointing 2004, ICPR, International Workshop on Visual Observation of Deictic Gestures*, 2004.

[12] DE CAMPOS, T. E., BABU, B. R., VARMA, M. "Character recognition in natural images". In: *Proceedings of the International Conference on Computer Vision Theory and Applications, Lisbon, Portugal*, February 2009.

[13] VIOLA, P. A., JONES, M. J. "Robust Real-Time Face Detection". In: *ICCV*, p. 747, 2001.