

# IDENTIFICAÇÃO DE EMOÇÕES A PARTIR DE EXPRESSÕES FACIAIS COM REDES NEURAIS SEM PESOS

FÁBIO SILVEIRA VIDAL\*<sup>†</sup>, HUGO CESAR DE CASTRO CARNEIRO<sup>‡</sup>, PAULO FERNANDO FERREIRA ROSA\*, FELIPE MAIA GALVÃO FRANÇA<sup>‡</sup>

*\*Instituto Militar de Engenharia  
Praça General Tibúrcio, 80 - Urca  
Rio de Janeiro, RJ, Brasil*

*†Instituto Federal do Tocantins  
Distrito Agroindustrial, BR 153, KM480  
Paraíso do Tocantins, TO, Brasil*

*‡Universidade Federal do Rio de Janeiro  
Av. Horácio Macedo, 2030, Prédio do Centro de Tecnologia, Bloco H - Cidade Universitária  
Rio de Janeiro, RJ, Brasil*

Email: vidalfs@gmail.com, hcesar@cos.ufrj.br, rpauloime@gmail.com, felipe@cos.ufrj.br

**Abstract**— This paper describes a study of neural networks without weights application for recognizing emotions. Task is performed from facial expressions image analysis. Eight situations were considered: anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. Experiments were made from TFEID (Taiwanese Facial Expression Image Database). Results shown that presented method is suitable to solve the recognizing emotions problem. Various training configurations were tested. Best returned hit rate above 95%.

**Keywords**— Computer Vision, Pattern Recognition, Emotion Recognition, Neural Network without Weights, Natural Interfaces

**Resumo**— Este trabalho descreve um estudo da aplicação de redes neurais sem pesos para o reconhecimento de emoções. A tarefa é feita a partir de análise de imagens de expressões faciais. Oito situações foram consideradas: raiva, desprezo, nojo, medo, felicidade, neutro, tristeza e surpresa. Os experimentos foram feitos a partir da base de dados TFEID (*Taiwanese Facial Expression Image Database*). Os resultados mostraram que a metodologia é adequada para a resolução do problema. Várias configurações de treinamento foram testadas. As melhores retornaram taxa de acerto acima de 95%.

**Keywords**— Visão Computacional, Reconhecimento de Padrões, Reconhecimento de Emoções, Redes Neurais sem Pesos, Interfaces Naturais

## 1 Introdução

O desenvolvimento tecnológico e o crescimento de poder de processamento proveem diferentes formas de interação com o computador. O reconhecimento automático de emoções viabiliza que uma máquina possa reagir de forma mais adequada ao estado emocional do usuário ou de pessoas em torno de uma máquina. Esta é uma forma de desenvolver interfaces naturais e pró-ativas, sendo uma alternativa ao uso do tradicional método de respostas a comandos por meio de hardwares como teclado, *mouse* ou *joysticks*.

O reconhecimento automático de expressões faciais vem ganhando importância para o desenvolvimento de interfaces interativas, podendo contribuir em diversas aplicações como: ambientes virtuais de aprendizagem, sistemas de segurança, casas inteligentes ou aplicações que visem acessibilidade a pessoas portadoras de necessidades especiais. O objetivo é fazer com que um sistema aja de acordo com reações de usuários ou de indivíduos a serem observados.

Uma solução de reconhecimento de expressões faciais pode passar por 3 (três) etapas: (1) loca-

lização da face; (2) segmentação de regiões de interesse; e (3) identificação da emoção. O escopo deste trabalho foca apenas a última etapa. Nos experimentos, as duas primeiras foram feitas de modo manual.

Vários métodos de reconhecimento de padrões podem ser adotados para identificar a emoção de uma pessoa, como Lógica Fuzzy, Redes Neurais Artificiais, *Support Vector Machine* (SVM), ou Redes Bayesianas. Neste trabalho, utilizou-se uma Rede Neural sem Pesos. Para tal escolha, considerou-se o processo de treinamento, o qual é mais rápido e mais simplificado que outros modelos de redes neurais, como o o MLP (*Multilayer Perceptron*).

Nas duas próximas seções, alguns trabalhos relacionados e opções de base de dados de imagens são referenciados. Em seguida, são descritos os processos de pré-processamento das imagens e o treinamento da rede neural sem pesos. Por fim, apresenta-se a validação e os resultados, e são feitas as considerações finais.

## 2 Trabalhos Relacionados

A descrição do estado da arte da análise automática de expressões faciais é dada em Pantic and Rothkrantz (2000). O contraste entre a facilidade dos seres humanos identificarem expressões e a dificuldade do desenvolvimento de uma solução computacional é discutido.

Uma descrição geral do problema de análise de expressões faciais é dada em Kanade et al. (2000). São abordados tópicos como: transição entre expressões; confiabilidade e validade de bases de dados de treinamento e de testes; orientação da cabeça; e complexidade da cena. Niedenthal et al. (2000) criou uma base de dados com 96 indivíduos. Os participantes foram induzidos a fazer a transição felicidade-neutro e tristeza-neutro. Oito deles foram descartados. A observação foi filmada. O método proposto viabilizou a identificação da transição entre os estados emocionais das pessoas.

Outra base de dados de imagens foi criada por Bartlett et al. (2003). O trabalho descreve um sistema para detecção de faces em uma *stream* de vídeo. Em tempo real, as faces são codificadas em sete dimensões (neutra, raiva, nojo, medo, alegria, tristeza e surpresa). É utilizada uma combinação das técnicas Adaboost e SVM para incrementar a acurácia do método.

Um método de aprendizado geneticamente inspirado é proposto em Yu and Bhanu (2006). O mesmo identifica características faciais de modo automático. A partir delas, uma SVM é treinada. O algoritmo funciona em imagens que não haviam sido processadas previamente. Porém, é sensível aos resultados da extração de características.

Zeng et al. (2006) explora a detecção de emoções durante uma conversa real. As emoções são classificadas apenas como positivas, negativas ou neutras. A base de dados foi adquirida a partir de entrevistas com pessoas.

Três variantes do modelo HMAX são propostas em Gu et al. (2009). O trabalho tratou com as situações de felicidade, tristeza, irritação, surpresa, nojo, medo e neutro. Utilizou-se duas bases de dados: TFEID (*Taiwanese Facial Expression Image Database*) e JAFFE (*Japanese Female Facial Expression*). Os experimentos foram feitos em duas situações: (a) utilizando a mesma base de dados para treinamento e validação; e (b) validação cruzada. A taxa de acerto para os modelos propostos foi superior ao HMAX padrão em ambos os casos.

A base de dados JAFFE foi utilizada em Jiang et al. (2012). O trabalho concentrou a taxa de acerto de 80,42% a 95,71%. Já Li and Oussalah (2010) trabalhou com a base de dados *Taiwanese and Indian Face Databases*. É feita a investigação de uma proposta que segue duas linhas. A primeira lida com PCA (*Principal Component Analysis*) e classificador KNN (*K-nearest Neigh-*

*bour*) e a segunda com NMF (*Negative Matrix Factorization*) e o algoritmo KNN. Ou (2012) é outro trabalho a utilizar KNN. As características são extraídas com *wavelets*. Segundo o autor, os resultados são excelentes.

Um estudo da generalização de um método de reconhecimento de emoções é feito por Zhang et al. (2011a). O método analisado tem desempenho ruim quando submetido a uma base de dados diferente daquela que foi treinado. As características foram extraídas com uso do Filtro de Gabor. O reconhecimento foi executado com SVM. A validação foi feita de duas formas: (a) com a mesma base de dados do treinamento; (b) utilizando outra base de dados. No primeiro caso, a taxa média de acerto foi de 85,7%, aproximadamente. No outro caso, a média foi de 60%.

## 3 Base de Dados

A produção de um banco de dados de imagens, para o propósito deste trabalho, é uma tarefa complexa. Isto acontece, principalmente, devido a fatores relacionados à aquisição de imagens, como: iluminação da cena, orientação da câmera e escala. No processo de captura das imagens é necessário que as pessoas façam expressões de modo deliberado. Este fato também provoca dificuldades, pois, em muitos casos, a expressão deliberada diverge da expressão espontânea. Estas divergências levam ao descarte de muitas amostras. Alguns dos ensaios são feitos com modelos profissionais e apoio técnico de psicólogos.

A base de dados JAFFE é utilizada em trabalhos como os de Zhao and Zhang (2012) e Jiang et al. (2012). A mesma disponibiliza um total de 213 imagens em tons de cinza. Nelas, foram observadas 10 modelos japonesas. Foram consideradas sete situações: alegria, tristeza, surpresa, raiva, nojo, medo e neutro.

O MUG (*Multimedia Understanding Group Facial Expression Database*) (Aifanti et al., 2010) disponibiliza imagens de 52 indivíduos em seis situações: raiva, nojo, medo, felicidade, tristeza e surpresa. Inicialmente, havia imagens de 86 pessoas. Algumas foram descartadas; pois as expressões feitas de modo deliberado não correspondiam a uma expressão feita de modo espontâneo.

Já a TFEID (Chen, 2007) é uma base de dados utilizada em trabalhos como os de Zhang et al. (2011b) e Gu et al. (2010). A mesma disponibiliza imagens de 34 modelos (18 homens e 16 mulheres), em 8 situações: raiva, desprezo, nojo, medo, felicidade, neutro, tristeza e surpresa. Algumas das amostras foram descartadas por não refletirem uma expressão espontânea. Vide Fig. 1.

Para a escolha da base de dados, considerouse: a qualidade das imagens; disponibilidade da base de dados; quantidade de amostras; e varie-

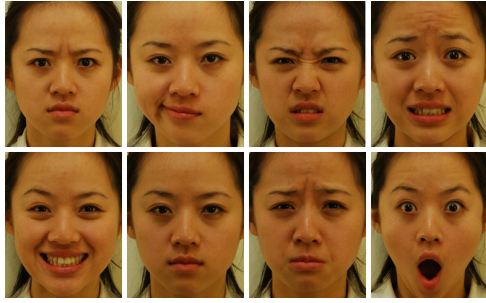


Figura 1: Exemplos de imagens do *Taiwanese Facial Expression Image Database (TFEID)*.

dade de situações. A TFEID foi considerada a melhor opção. Foram utilizadas trinta amostras de cada situação. Vinte para o conjunto de treinamento e dez para o conjunto de validação. Nas próximas seções, serão descritas as fases de pré-processamento e treinamento.

#### 4 Pré-processamento das Imagens

Antes de serem submetidas à rede neural, as imagens passaram por um processo de preparação. O mesmo aconteceu em duas fases: (a) segmentação; e (b) codificação.

A fase de segmentação extrai as três regiões de interesse da imagem: olho direito, nariz e boca. As mesmas possuem tamanho de  $200 \times 200$  pixels (olho direito),  $160 \times 120$  pixels (nariz) e  $240 \times 120$  pixels (boca) Vide Fig. 2.

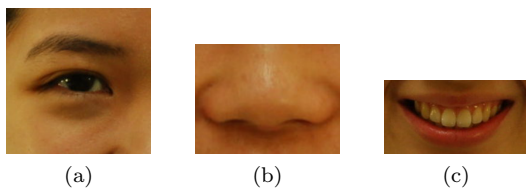


Figura 2: Exemplos de segmentos extraídos das imagens do TFEID.

O processo de codificação divide o segmento em subjanelas (vide Fig. 3) e cada uma delas é representada por um bit. Se uma subjanela possuir luminância média superior à de todo o segmento, recebe bit 1; caso contrário, recebe bit 0. Para o treinamento da rede neural, foram utilizadas as codificações listadas na Tab. 1. No caso da primeira codificação do segmento Olho Direito (OD-A), fora utilizada apenas  $\frac{1}{4}$  da imagem. Este detalhe pode ser observado na Fig. 4b.

#### 5 Treinamento da Rede Neural

O treinamento da rede neural foi feito individualmente para os segmentos boca, nariz e olho direito. Para cada um deles foi utilizada uma rede neural em particular. O treinamento obedeceu ao algoritmo de aprendizado não supervisio-

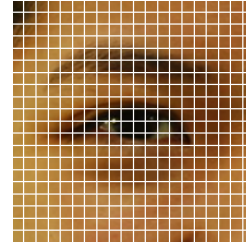


Figura 3: Segmento de imagem dividido em subjanelas de  $5 \times 5$  pixels, com distribuição  $20 \times 20$ .

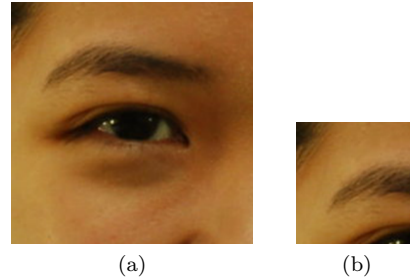


Figura 4: (a) Segmento olho direito; (b) Parte utilizada na codificação OD-A.

nado WiSARD (*Wilkes, Stonham and Aleksander Recognition Device*), descrito em alguns trabalho como os de Aleksander and Morton (1990), Aleksander (1998), Aleksander et al. (2009), Austin (1998) e Ludermir et al. (1999). O mesmo consiste da configuração de discriminadores, de modo a minimizar o erro na atividade de reconhecimento de padrões. Há um discriminador associado a cada classe de emoção.

Os discriminadores são constituídos por unidades de memória, chamadas RAMs. No início do treinamento, cada RAM é associada a três subjanelas de modo pseudoaleatório. Na fase de pré-processamento, um valor binário é atribuído a cada subjanela. As RAMs possuem oito posições de memória, iniciadas com o valor 0. Cada uma delas corresponde a uma possível combinação de valores atribuídos às subjanelas. Para todas as amostras utilizadas no treinamento, o respectivo discriminador tem suas RAMs atualizadas. Para tanto, incrementa-se o valor da posição de memória referente à combinação dos valores das subjanelas associadas à RAM. Vide Figura 5.

No processo de reconhecimento, o valor de ativação de uma RAM corresponde à posição de memória associada à combinação dos valores das suas subjanelas. Um discriminador retorna a soma dos valores de ativação das RAMs. Um padrão é atribuído à classe referente ao discriminador que retornou o maior valor.

A Tabela 1 mostra as configurações experimentadas durante o treinamento. Foram três para os segmentos olho direito (OD-A, OD-B e OD-C), nariz (N-A, N-B e N-C) e boca (B-A, B-B e B-C). Entre elas, se alterou o número de subjanelas, a

Tabela 1: Codificações experimentadas para cada classe durante o treinamento da rede neural sem pesos

Configuração	Segmento	Subjanelas	Distribuição	Bits	Repetições
OD-A	Olho Direito	5 x 5	20 x 20	400	400
OD-B	Olho Direito	10 x 10	20 x 20	400	400
OD-C	Olho Direito	5 x 5	40 x 40	1600	560
N-A	Nariz	20 x 20	8 x 6	48	48
N-B	Nariz	10 x 10	16 x 12	192	192
N-C	Nariz	5 x 5	32 x 24	768	768
B-A	Boca	24 x 24	5 x 10	50	50
B-B	Boca	12 x 12	10 x 20	200	200
B-C	Boca	6 x 6	20 x 40	800	800

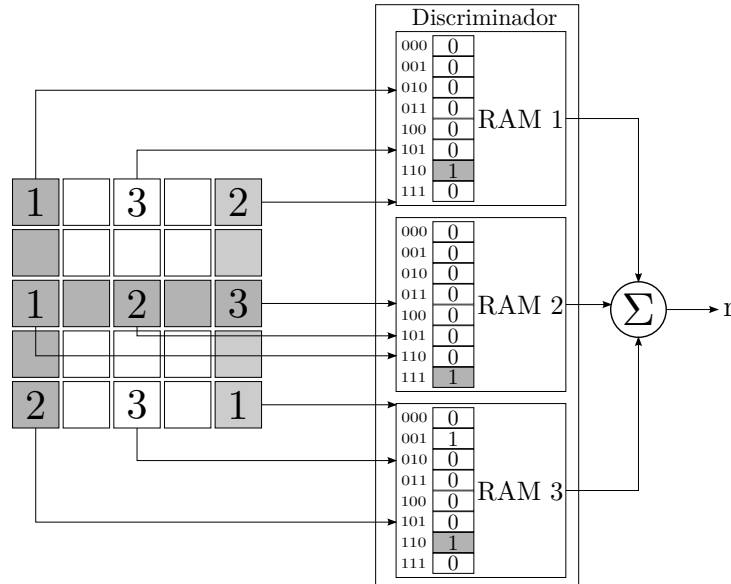


Figura 5: Exemplo de um discriminador WiSARD com três RAMs, adaptado de Bandeira (2010).

distribuição das mesmas no segmento de imagem e a quantidade de *bits* para codificação. O número de RAMs por discriminador foi variável, testando  $\forall x \in \mathbb{N} | 1 \leq x \leq n/3$ , em que  $n$  é o número de *bits* da codificação.

Os *bits* referentes a cada RAM são determinados de modo pseudoaleatório. Isto torna o algoritmo de treinamento não determinístico. Portanto, o processo foi realizado repetidas vezes para cada configuração. A Tabela 1 mostra a quantidade de repetições para cada configuração. As melhores configurações encontradas estão elencadas na Tabela 2.

A próxima seção mostra o processo de validação e os resultados dos experimentos realizados.

## 6 Validação e Resultados

Após o treinamento com cada configuração, a rede foi testada com o conjunto de validação. A qualidade dos resultados foi avaliada a partir dos percentuais de acertos e duplicações; e quantidade de falso-positivos. Entende-se por duplicação os casos em que o maior valor é retornado por discriminadores de classes diferentes.

A Tabela 2 mostra as configurações com maior taxa de acerto encontradas durante o processo

de treinamento. Para o segmento olho direito, os melhores resultados foram encontrados com as configurações OD1 (96% de acerto, 1% de duplicações, 12 discriminadores) e OD2 (95% de acerto, 0% de duplicações, 28 discriminadores).

Para o segmento boca, os melhores resultados foram encontrados com as configurações B3 (98% de acerto, 0% de duplicações, 19 discriminadores e codificação de 800 *bits*) e B4 (95% de acerto, 0% de duplicações, 5 discriminadores e codificação de 800 *bits*). A configuração B2 (88% de acerto, 0% de duplicações, 32 discriminadores e codificação de 200 *bits*), obteve acerto de 100% em todas as classes, exceto tristeza e neutro. Nestes dois casos houve 0% de acerto e 100% de falso-positivos, respectivamente. Todos os padrões da classe tristeza foram classificados como neutro.

Em relação ao segmento nariz, a melhor configuração testada foi a N-C (85% de acerto, 0% de duplicações, 48 discriminadores e codificação de 768 *bits*).

Como pode ser observado na Tabela 3, as emoções com maior taxa de acerto foram felicidade, surpresa, nojo e medo. Com as mesmas, a taxa de acerto foi maior ou igual a 90%, com quase todas as configurações da Tabela 2. A única exceção se deu com a configuração B1 e emoção

Tabela 2: Melhores configurações encontradas após o treinamento da rede neural sem pesos

Configuração	Segmento	Codificação	Nº de RAMs	Acertos	Duplicações
OD1	Olho Direito	OD-A	12	96%	1%
OD2	Olho Direito	OD-A	28	95%	0%
OD3	Olho Direito	OD-B	9	81%	0%
OD4	Olho Direito	OD-C	30	88%	0%
N1	Nariz	N-A	14	66%	1%
N2	Nariz	N-B	32	75%	0%
N3	Nariz	N-C	48	85%	0%
B1	Boca	B-A	8	90%	0%
B2	Boca	B-B	32	88%	0%
B3	Boca	B-C	19	98%	0%
B4	Boca	B-C	5	93%	0%

Tabela 3: Taxa de acerto de cada configuração para cada classe de emoção

	OD1	OD2	OD3	OD4	B1	B2	B3	B4
Raiva	100%	100%	100%	90%	100%	100%	90%	50%
Desprezo	90%	70%	0%	10%	80%	100%	100%	100%
Nojo	100%	100%	100%	100%	80%	100%	100%	100%
Medo	100%	100%	90%	100%	90%	100%	90%	100%
Felicidade	100%	100%	100%	100%	100%	100%	100%	100%
Neutro	90%	100%	90%	100%	70%	100%	100%	90%
Tristeza	100%	90%	70%	100%	100%	0%	100%	100%
Surpresa	90%	100%	100%	100%	100%	100%	100%	100%

Tabela 4: Quantidade de falso-positivos de cada configuração para cada classe de emoção

	OD1	OD2	OD3	OD4	B1	B2	B3	B4
Raiva	0	0	2	0	1	0	0	0
Desprezo	0	0	0	0	0	0	0	0
Nojo	1	0	0	0	2	0	1	0
Medo	0	0	0	0	0	0	0	0
Felicidade	3	2	2	6	0	0	0	0
Neutro	0	2	11	4	0	10	0	2
Tristeza	0	0	0	0	0	0	1	3
Surpresa	0	0	0	0	5	0	0	0

desprezo, em que a taxa de acerto foi de 80%.

Na maioria dos casos, as emoções neutro, raiva, tristeza e desprezo também tiveram taxa de acerto maior ou igual a 90%. Com as seguintes exceções: neutro (B1, com 70%); raiva (B4, com 50%); tristeza (OD3 e B2, com 70% e 0%, respectivamente); desprezo (OD2 e OD3, com 70% e 0%, respectivamente).

Para as emoções desprezo e medo, nota-se na Tabela 4 que não houve casos de falso-positivos. Neste aspecto, os piores casos foram verificados com felicidade e neutro. Com felicidade houve 3, 2, 2 e 6 casos, com as configurações OD1, OD2, OD3 e OD4, respectivamente. Com neutro 2, 11, 4, 10 e 2 casos, com as configurações OD2, OD3, OD4, B2 e B4, respectivamente.

## 7 Considerações Finais

Os experimentos retornaram seus melhores resultados com os segmentos boca e olho direito. Os mesmos mostraram que uma rede neural sem pesos é adequada para a tarefa de identificação de emoções humanas. Entretanto, os resultados não foram tão bons a partir do segmento nariz.

Individualmente, as configurações OD1, OD2

e B3 retornaram os melhores resultados. Entretanto, OD1 é mais interessante por utilizar um número menor de discriminadores. Os padrões das classes felicidade e neutro tiveram muitos falso-positivos. Já o padrão Desprezo foi o de mais difícil detecção. Ainda assim, com a configuração OD1 a taxa de acerto foi de 90%.

Com a configuração B2, a taxa de acerto alcançada foi de 88%. Houve 0% de acerto para a classe tristeza e 10 falso-positivos para a classe neutro. Nos demais casos, 100% de sucesso. Já com a configuração B3, a taxa de acerto para neutro e tristeza foi de 100%. Houve apenas um caso de falso-positivo.

Desta forma, o mais interessante é uma solução híbrida, utilizando a configuração B2. Caso a resposta seja neutro ou tristeza, refaz-se o processo com uma rede de configuração B3. Desta forma, pode-se alcançar um percentual de acerto próximo a 100% para o conjunto de validação.

Como trabalhos futuros, é proposta a implementação de algoritmos para detecção de faces e segmentação de olhos e boca. Desta forma, todo o processo descrito neste trabalho ficaria independente da intervenção de um usuário. Outros aspectos ainda podem ser abordados, como diferen-

tes condições de iluminação e diferentes ângulos para tomada das imagens.

## Referências

- Aifanti, N., Papachristou, C. and Delopoulos, A. (2010). The mug facial expression database, *Image Analysis for Multimedia Interactive Services (WIAMIS), 2010 11th International Workshop on*, pp. 1–4.
- Aleksander, I. (1998). From wisard to magnus: a family of weightless virtual neural machines, *RAM-Based Neural Networks* pp. 18–30.
- Aleksander, I., Gregorio, M. D., França, F. M. G., Lima, P. M. V. and Morton, H. (2009). A brief introduction to Weightless Neural Systems.
- Aleksander, I. and Morton, H. (1990). *An introduction to neural computing*, Vol. 240, Chapman and Hall London.
- Austin, J. (1998). *RAM-based neural networks*, Vol. 9, World Scientific Publishing Company.
- Bandeira, L. C. (2010). *Nc-wisard: Uma interpretação sem pesos do modelo neural neocognitron*, Master's thesis, Universidade Federal do Rio de Janeiro.
- Bartlett, M. S., Littlewort, G., Fasel, I. and Movellan, J. R. (2003). Real time face detection and facial expression recognition: Development and applications to human computer interaction., *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*, Vol. 5, IEEE, pp. 53–53.
- Chen, L., Y. Y.-S. (2007). Taiwanese facial expression image database, *Technical report*, Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan.
- Gu, W., Venkatesh, Y. and Xiang, C. (2010). Composite orthonormal basis for person-independent facial expression recognition, *Industrial Engineering and Engineering Management (IEEM), 2010 IEEE International Conference on*, pp. 1942–1946.
- Gu, W., Xiang, C. and Lin, H. (2009). Modified hmax models for facial expression recognition, *Control and Automation, 2009. ICCA 2009. IEEE International Conference on*, IEEE, pp. 1509–1514.
- Jiang, B., Jia, K.-B. and Wu, Q. (2012). A novel algorithm of facial expression recognition based on discriminative component analysis, *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*, pp. 399–402.
- Kanade, T., Cohn, J. F. and Tian, Y. (2000). Comprehensive database for facial expression analysis, *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, IEEE, pp. 46–53.
- Li, J. and Oussalah, M. (2010). Automatic face emotion recognition system, *Cybernetic Intelligent Systems (CIS), 2010 IEEE 9th International Conference on*, IEEE, pp. 1–6.
- Ludermir, T. B., Carvalho, A., Braga, A. and Souto, M. (1999). Weightless neural models: a review of current and past works, *Neural Computing Surveys* **2**: 41–61.
- Niedenthal, P. M., Halberstadt, J. B., Margolin, J. and Innes-Ker, A. H. (2000). Emotional state and the detection of change in facial expression of emotion, *European Journal of Social Psychology* **30**(2): 211–222.
- Ou, J. (2012). Classification algorithms research on facial expression recognition, *Physics Procedia* **25**: 1241–1244.
- Pantic, M. and Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: The state of the art, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **22**(12): 1424–1445.
- Yu, J. and Bhanu, B. (2006). Evolutionary feature synthesis for facial expression recognition, *Pattern Recognition Letters* **27**(11): 1289–1298.
- Zeng, Z., Fu, Y., Roisman, G. I., Wen, Z., Hu, Y. and Huang, T. S. (2006). Spontaneous emotional facial expression detection, *Journal of Multimedia* **1**(5): 1–8.
- Zhang, Z., Fang, C. and Ding, X. (2011a). Facial expression analysis across databases, *Multimedia Technology (ICMT), 2011 International Conference on*, IEEE, pp. 317–320.
- Zhang, Z., Fang, C. and Ding, X. (2011b). Facial expression analysis across databases, *Multimedia Technology (ICMT), 2011 International Conference on*, IEEE, pp. 317–320.
- Zhao, X. and Zhang, S. (2012). Facial expression recognition based on local binary patterns and least squares support vector machines, in D. Jin and S. Lin (eds), *Advances in Electronic Engineering, Communication and Management Vol.2*, Vol. 140 of *Lecture Notes in Electrical Engineering*, Springer Berlin Heidelberg, pp. 707–712.