

# Reconhecimento de Palavras Isoladas Utilizando Fluxo Óptico e Redes Neurais Artificiais através de Informações Visuais

Thiago Vinícius Dantas Ferraz\*  
thiago.ferraz@aol.com

Dr. Gustavo Fernandes Rodrigues\*  
gfernandes@ufsj.edu.br

**Resumo**— Este trabalho tem como objetivo o desenvolvimento de um sistema de reconhecimento de fala através do rastreamento dos movimentos dos lábios a partir de um vídeo utilizando a análise de fluxo óptico para funcionar em ambientes ruidosos. A maioria dos sistemas de reconhecimento de fala não possui um bom desempenho em ambientes ruidosos. Para a tarefa de reconhecimento foi utilizada uma rede neural com arquitetura feedforward multicamadas com aprendizagem baseada no algoritmo de retropropagação do erro (backpropagation). Foi utilizado exclusivamente informações visuais obtidas através de um banco de dados composto de gravações de vídeos correspondentes aos dígitos de 0 a 9 pronunciados em português por um indivíduo do sexo masculino. A melhor taxa de acertos da rede usando apenas informação visual para o reconhecimento de fala foi de aproximadamente 79%.

**Palavras-Chave**— Fluxo Óptico, Redes Neurais, Componentes Principais.

**Abstract**— This paper proposes a visual speech recognition method using lip movement extracted from a video measured by optical flow to attempt to increase robustness in noise environments. To perform the speech recognition, an artificial neural network is used with multilayer feed forward architecture with back propagation learning algorithm. Experiments were conducted using Portuguese isolated digit speech using visual information only extracted from videos. Recognition accuracy achieved for speech recognition tests using visual information only is about 79%.

**Keywords**— Optical Flow, Neural Network, Principal Component.

## I. INTRODUÇÃO

O avanço da tecnologia do reconhecimento de fala tem permitido alcançar níveis de desempenho satisfatórios em algumas aplicações interessantes, como sistemas de controle e comando, telefonia, transcrição, centrais de atendimento e também na área da robótica. No entanto, a falta de robustez desses sistemas a perturbações que afetam o sinal acústico, continua a ser uma limitação [1]. Esse fato tem conduzido a uma acentuada investigação relacionada ao problema da robustez, tendo como consequência o surgimento de diversos estudos sobre o assunto, como a utilização de novas técnicas ou as combinações das mesmas [2]. Algumas vertentes dessa investigação baseiam-se na extração de características acústicas mais robustas e no aperfeiçoamento dos modelos acústicos. Outra abordagem, que tem-se mostrado

bastante promissora nas pesquisas, consiste na utilização de características de natureza não acústica. Nos últimos 30 anos, extensas pesquisas no campo de estudo envolvendo reconhecimento de fala audiovisual (AVSR, *Audio Visual Speech Recognition*) vêm sendo realizadas. Uma análise abrangente e detalhada apresentando os progressos e desafios do AVSR, são resumidos por Potamianos *et al* [3][4].

Este trabalho aborda o desenvolvimento de um sistema VSR (*Visual Speech Recognition*) utilizando a plataforma MatLab (*Matrix Laboratory*), baseando o processo de reconhecimento em uma combinação adequada das características visuais extraídas através das técnicas Fluxo Óptico (*Optical Flow*) e PCA (*Principal Component Analysis*). Uma rede neural *multi-layer perceptron* com algoritmo *backpropagation* é usada para classificar os dados de fala.

Este artigo está dividido da seguinte forma. As Seções II, III e IV apresentam as técnicas utilizadas para extrair as características importantes do sinal de vídeo. A Seção V mostra o método utilizado para a classificação e reconhecimento dos dígitos de 0 a 9. A Seção VI descreve como foi implementado o sistema de reconhecimento e a Seção VII apresenta os resultados obtidos. Finalmente, algumas considerações deste trabalho encontram-se na Seção VIII.

## II. DETECÇÃO LABIAL

Um dos primeiros passos em um sistema de reconhecimento de fala que utiliza somente informações visuais, é a detecção da Região de Interesse, (ROI - *Region of Interest*). Esta etapa é importante porque elimina informações desnecessárias da imagem.

Paul Viola e Michael Jones [6] propuseram uma abordagem para a detecção de objetos em imagens que se baseia em três conceitos: integração da imagem, treinamento de classificadores e o uso de classificadores em cascata. Embora o algoritmo possa ser treinado para reconhecer qualquer objeto, a motivação principal da abordagem Viola-Jones foi o reconhecimento facial. Muitos algoritmos de detecção facial precisam ser treinados exaustivamente antes de obterem resultados satisfatórios, entretanto, um dos pontos fortes do algoritmo Viola-Jones é a rapidez com que é executado.

A principal função utilizada para a detecção labial neste trabalho é a *CascadeObjectDetector*, disponível no *Toolbox Computer Vision*, que utiliza o algoritmo de Viola-Jones para detectar a ROI, por exemplo, rosto, nariz, olhos ou boca.

\*DETEM - Departamento das Engenharias de Telecomunicações e Mecatrônica, Universidade Federal de São João del-Rei, Ouro Branco, MG.

### III. FLUXO ÓPTICO

Uma importante técnica para estimar movimento a partir de uma sequência de imagens é o chamado Fluxo Óptico (*Optical Flow*). Fluxo Óptico é a distribuição da velocidade aparente do movimento dos padrões de intensidade no plano da imagem [7]. Pode surgir de um movimento relativo de objetos e vista. Conseqüentemente, fluxo óptico pode dar uma informação importante sobre o arranjo dos objetos vistos e a taxa de mudança destes mesmos arranjos. Em outras palavras, o campo de fluxo óptico consiste em um campo denso de velocidade onde cada pixel no plano da imagem está associado a um único vetor de velocidade. Para fins de visualização, o campo é amostrado em uma malha e chamado de mapa de agulhas (*needle map*). Se for conhecido o intervalo de tempo entre duas imagens consecutivas, os vetores da velocidade podem ser convertidos em vetores de deslocamento e vice-versa [8].

Na Seção VII, serão apresentados resultados obtidos a partir da utilização das duas principais técnicas diferenciais para calcular fluxo óptico. A primeira, *Lucas-Kanade*, proposto em [10] e a segunda *Horn-Schunck*, proposto em [7].

#### A. Equação de Restrição do Fluxo Óptico

Os métodos para o cálculo do fluxo óptico podem ser classificados em três grupos principais:

- técnicas diferenciais;
- técnicas de correlação e
- técnicas baseadas em frequência/energia.

Em [11] é possível encontrar uma extensa lista de referências sobre esses três grupos principais.

Nas técnicas diferenciais, a hipótese inicial para a computação do fluxo óptico é a de que a intensidade entre quadros diferentes em uma sequência de imagens seja aproximadamente constante em um intervalo de tempo pequeno, ou seja, em um pequeno intervalo de tempo o deslocamento será mínimo.

A velocidade da imagem é calculada a partir das derivadas espaço-temporal da intensidade na imagem. Considerando uma imagem  $I$  ainda não amostrada (domínio contínuo) e em tons de cinza, pode-se usar a seguinte expressão [7]:

$$I(\vec{x}, t) \approx I(\vec{x} + \delta\vec{x} + \delta t) \quad (1)$$

onde,  $I(\vec{x}, t)$  é a função de intensidade da imagem,  $\vec{x} = (x, y)$  é o vetor posição na imagem e  $\delta\vec{x}$  é o deslocamento de uma região da imagem em  $(\vec{x}, t)$  após o tempo  $\delta t$ , conforme mostra a figura 1. Expandindo o lado direito da equação (1) por Série

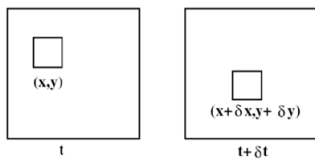


Fig. 1. A imagem na posição  $t$  é a mesma imagem na posição  $t + \delta t$

de Taylor, tem-se:

$$I(\vec{x} + \delta\vec{x}, t + \delta t) = I(\vec{x}; t) + \nabla I \delta\vec{x} + \delta t I_t + O^2 \quad (2)$$

onde,  $\nabla I = (I_x, I_y)$  é o gradiente nas direções  $x$  e  $y$ ;  $I_t$  a derivada parcial de primeira ordem em relação ao tempo de  $I(\vec{x}; t)$ , e  $O^2$  os termos de segunda ordem em diante, que podem ser desprezados. Subtraindo  $I(\vec{x}; t)$  em ambos os lados e dividindo por  $\delta t$ , têm-se:

$$\nabla I \cdot \frac{\delta\vec{x}}{\delta t} + I_t = 0 \Rightarrow \nabla I \cdot \vec{v} + I_t = 0 \quad (3)$$

ou simplificando,

$$I_x u + I_y v + I_t = 0 \quad (4)$$

onde  $\vec{x} = (u, v) = (\frac{\delta x}{\delta t}, \frac{\delta y}{\delta t})$  é a velocidade na imagem. A equação (3) é chamada de Equação de Restrição do Fluxo Óptico e define uma restrição local única sobre o movimento na imagem [7]. As componentes  $u$  e  $v$  são definidas como o fluxo óptico horizontal e vertical, respectivamente.

Para calcular o fluxo óptico entre duas imagens, deve-se resolver a equação (3) ou (4). Pelo fato dessas equações estarem sob restrição, existem diversos métodos para solucionar  $u$  e  $v$ . Nas próximas seções serão citadas duas das principais técnicas.

#### B. Horn & Schunck

*Horn* e *Schunck* usaram uma forma de regularização aplicada à equação (3), chamada de restrição de suavização, significando que o fluxo de vetores varia de uma imagem para outra de forma suave.

$$E = \int \int (I_x u + I_y v + I_t)^2 dx dy + \alpha \int \int \left[ \left( \frac{\delta u}{\delta x} \right)^2 + \left( \frac{\delta u}{\delta y} \right)^2 + \left( \frac{\delta v}{\delta x} \right)^2 + \left( \frac{\delta v}{\delta y} \right)^2 \right] dx dy \quad (5)$$

Nesta equação,  $\frac{\delta u}{\delta x}$  e  $\frac{\delta u}{\delta y}$  são derivadas espaciais das componentes ópticas de velocidade  $u$  e  $v$ , e  $\alpha$  é o fator de suavização. O método de Horn-Schunck minimiza a equação (5) para obter o campo de velocidades,  $[u, v]$ , para cada *pixel* da imagem. A solução desta equação consiste em um cálculo computacional iterativo descrito em [12].

#### C. Lucas & Kanade

Para resolver a equação de restrição do fluxo óptico para  $u$  e  $v$ , o método de *Lucas-Kanade* divide a imagem original em pequenas seções e assume uma velocidade constante em cada seção. Em seguida, ele executa um ajuste ponderado de mínimos quadrados da equação de restrição do fluxo óptico para um modelo constante de  $[uv]^T$  em cada seção,  $\Omega$ , minimizando a seguinte equação:

$$\sum_{x \in \Omega} W^2 [I_x u + I_y v + I_t]^2 \quad (6)$$

Aqui,  $W$  é uma função de janela, que enfatiza as restrições no centro de cada seção. A solução para o problema de minimização é dada pela seguinte equação:

$$\begin{bmatrix} \sum W^2 I_x^2 & \sum W^2 I_x I_y \\ \sum W^2 I_y I_x & \sum W^2 I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum W^2 I_x I_t \\ \sum W^2 I_y I_t \end{bmatrix} \quad (7)$$

$I_t$  é calculado usando por padrão um filtro de diferença. A solução final das componentes  $u$  e  $v$  podem ser consultadas em [12].

#### IV. ANÁLISE DE COMPONENTES PRINCIPAIS

A Análise em Componentes Principais ou *Principal Component Analysis* (PCA) é uma formulação matemática usada na redução da dimensão de dados [13]. Assim, a técnica PCA permite identificar padrões nos dados e expressá-los de uma maneira tal que suas semelhanças e diferenças sejam destacadas. Uma vez encontrados padrões nos dados, é possível comprimi-los, ou seja, reduzir suas dimensões, sem muita distorção.

Na abordagem PCA, a informação contida em um conjunto de dados é armazenada em uma estrutura computacional de dimensão reduzida a partir da projeção integral do conjunto de dados em um subespaço gerado por um sistema de eixos ortogonais [14]. O sistema ótimo de eixos pode ser obtido usando o método de decomposição em valor singular (SVD, *Singular Values Decomposition*) [15]. A estrutura computacional de dimensões reduzidas é selecionada de tal modo que características relevantes dos dados sejam identificadas com poucas perdas de informação [14]. Tal redução é vantajosa em várias aplicações: para compressão de imagens, representação dos dados, redução do cálculo necessário em processamentos subsequentes e reconhecimento de padrões, entre outras.

O uso da técnica PCA na redução da dimensão de dados é justificado pela fácil representação de dados multidimensionais, utilizando a informação contida na matriz de covariância dos dados [14], [16].

#### V. REDES NEURAIS ARTIFICIAIS

As Redes Neurais Artificiais (RNA's) são sistemas de computação adaptativos que foram inspirados nos neurônios biológicos e nas características de suas interconexões [17]. Elas podem ser utilizadas para aproximar funções ou classificar dados em classes semelhantes, por exemplo, sílabas ou palavras isoladas de um vocabulário. A capacidade de aprender através da adaptação é uma propriedade fundamental das RNA's.

#### VI. METODOLOGIA

Nesta seção serão apresentados todos os passos utilizados durante a implementação do sistema, desde as gravações dos vídeos até o processo de reconhecimento. O fluxograma demonstra todos os passos utilizados.

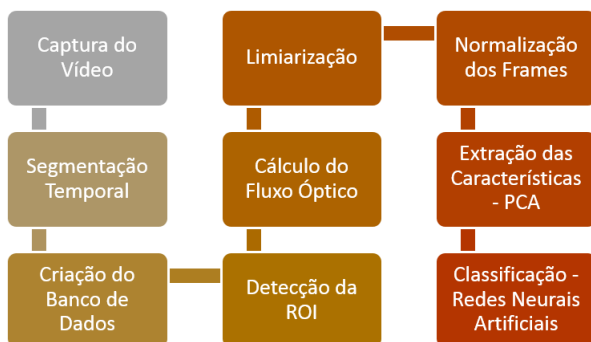


Fig. 2. Fluxograma do Sistema de Reconhecimento proposto

#### A. Captura dos Vídeos

As amostras de vídeo foram capturadas por uma câmera digital de 14 *Megapixels*, no formato MOV (ou formato *Quick-Time*), com resolução de imagem configurada para 640x480 *pixels* e 30 *quadros/segundos* entrelaçados.

#### B. Segmentação Temporal

A segmentação temporal dos vídeos foi realizada utilizando o software de edição de vídeo, *Sony Vegas*. Consiste na identificação do início e fim de cada palavra, deixando-as prontas para serem armazenadas no banco de dados, no formato MPEG-4, resolução 240x320 *pixels*. Os vídeos foram comprimidos para este formato devido à sua alta taxa de compressão e elevada qualidade.

#### C. Banco de Dados

O Banco de Dados consiste em 200 amostras de fala gravadas em vídeo no qual o locutor, do sexo masculino, falava continuamente as palavras correspondentes aos dígitos de “Zero” a “Nove”, em português brasileiro. Para cada palavra foram realizadas 20 gravações.

#### D. Detecção da Região de Interesse

Este é sem dúvida um passo extremamente importante, pois ele detecta informações linguísticas relevantes, pois com a detecção da boca através do algoritmo citado na Seção II, toda informação desnecessária é eliminada, como movimento dos olhos e do rosto em geral, permanecendo somente a região da boca.

A imagem de entrada encontra-se no formato RGB e possui dimensões 240x320 *pixels*. As imagens na saída deste processo são convertidas para escala de cinza, conforme mostra a figura 3. Neste momento somente a região da boca é recortada para cada palavra do banco de dados. Cada palavra é uma matriz de tamanhos variados contendo 10 quadros.

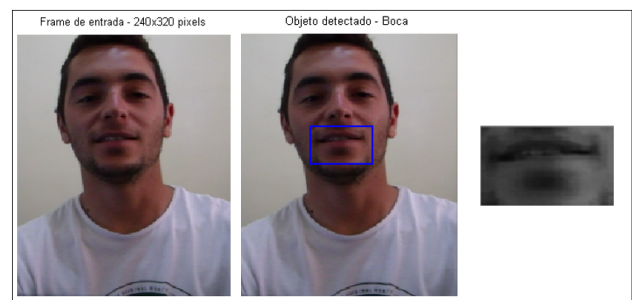


Fig. 3. Detecção da boca utilizando o *Toolbox Computer Vision*

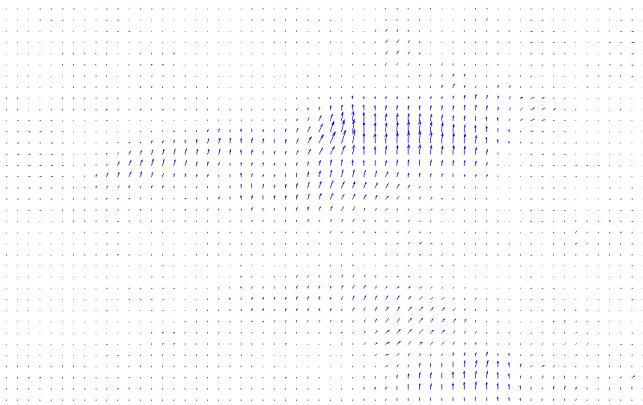
#### E. Cálculo Fluxo Óptico

Esta etapa foi realizada utilizando as duas técnicas descritas na Seção III, *Horn-Schunck* (HS) e *Lucas-Kanade* (LK). Ambas foram aplicadas no banco de dados utilizando o *Toolbox Computer Vision*, e um novo banco foi criado com as matrizes de fluxo óptico referentes aos dígitos gravados. Os

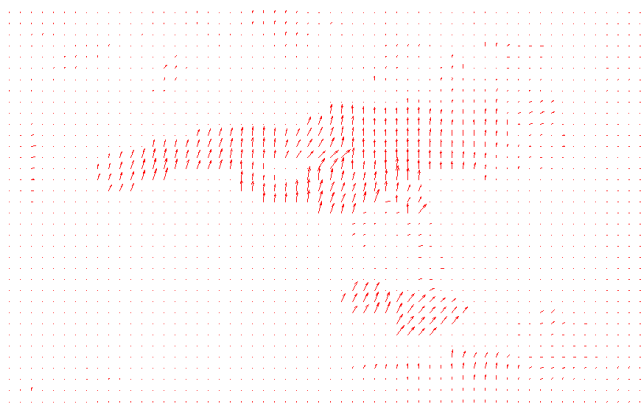
valores resultantes destas matrizes são exatamente o módulo quadrático (*Magnitude-squared*) dos vetores  $u+v$  descritos na Seção III.

Após o cálculo do fluxo óptico, todas as matrizes resultantes foram normalizadas, ou seja, todas possuem mesmas dimensões, sendo neste caso  $37 \times 61$ .

A figura 4 demonstra o resultado do cálculo do fluxo óptico referente a dois quadros da palavra “Seis”. A partir daqui, pode-se afirmar que o método **LK** apresenta uma maior resolução do campo de agulhas, ou seja, os vetores de deslocamento são mais explícitos, o que não o torna necessariamente mais robusto que o método **HS** para a aplicação de reconhecimento, pois ambos podem apresentar vetores de ruídos influenciando nos resultados. Ambos apresentaram o mesmo custo computacional. Uma explicação para o campo de agulhas do método **HS** ser diferente, está diretamente ligado ao fator de suavização que o mesmo possui.



(a) FO Horn-Schunck



(b) FO Lucas-Kanade

Fig. 4. Fluxo Óptico utilizando os métodos Horn-Schunck (a) e Lucas-Kanade (b), respectivamente, entre dois quadros da palavra “Seis”.

Existem deslocamentos que são muito pequenos e desnecessários nesta situação em algumas regiões da imagem, por exemplo, abaixo dos lábios, na região do queixo, os deslocamentos são da ordem de  $10^{-3}$  a  $10^{-8}$  (adimensional). As maiores magnitudes dos vetores de deslocamentos estão na região que contorna os lábios. Portanto, uma função de limiarização foi aplicada em todas as matrizes de fluxo óptico, eliminando valores inferiores a  $10^{-3}$  e substituindo por zero.

#### F. Extração de Características

Como descrito na Seção IV, o PCA consegue extrair informações importantes de um grande conjunto de dados reduzindo suas dimensões sem muita perda de informação.

É importante lembrar que o banco de dados até aqui, é composto por matrizes de fluxo óptico com dimensões  $37 \times 61$  referente a cada palavra gravada. Cada uma dessas palavras contém 10 quadros, ou seja, 10 matrizes  $37 \times 61$ . Após aplicar PCA neste banco de dados, o resultado é uma matriz de características de dimensões  $37 \times 37$  para cada palavra com seus respectivos quadros.

A primeira coluna de cada uma dessas matrizes possuem cerca 85% da variância das componentes principais, sendo esses, os valores mais importantes extraídos. Logo, dos 10 quadros de uma palavra, foi retirada a primeira coluna da matriz de coeficientes de cada um desses quadros, armazenando-os em um novo banco de dados, para então serem inseridos na rede neural. Essa estratégia foi adotada com o intuito de formar uma nova matriz com os dados que tiveram maior variabilidade, ou seja, as informações visuais mais importantes, que é o movimento dos lábios.

O resultado final foi um novo banco de dados com matrizes de componentes principais  $37 \times 10$  para cada palavra. Haviam 10 quadros para cada gravação, por isso formaram-se matrizes com essas dimensões. Essas matrizes foram transformadas em vetores  $1 \times 370$ .

#### G. Classificação e Reconhecimento

Para o reconhecimento dos algarismos, utilizou-se uma Rede Neural *Multi-layer Perceptron* com algoritmo *backpropagation* e aprendizagem supervisionada. Foram considerados os resultados obtidos através da utilização de 10 e 20 neurônios. Isso ocorreu por causa da melhor performance apresentada quando comparada a taxa de acerto utilizando outras quantidades de neurônios. O banco de dados utilizado para as simulações, foi dividido da seguinte forma: utilizou-se 75% dos dados para o treinamento e 25% para testes. Essa configuração foi denominada como *Configuração 1*.

A extração de características utilizando PCA geraram vetores que foram concatenados verticalmente, formando então, os dados para treinamento da rede e teste da mesma. Ainda foi criada uma matriz de valores binários denominada *Target*, referenciando os dígitos de 0 a 9, que é o resultado esperado na saída da rede. Como a rede é do tipo supervisionada, então ela recebe a matriz *Target* no momento do treinamento e dos testes. Durante o treinamento da rede, realizaram-se 20 iterações para cada configuração, e por fim, a média da taxa de acerto de todas as palavras foi calculada.

## VII. RESULTADOS

Nesta seção serão apresentados os resultados das simulações conforme a configuração da rede neural descrita na subseção VI-G.

#### A. Resultados da Configuração 1

Para a configuração de 75% dos dados utilizados para treinamento e 25% para testes, obtiveram-se os valores da

taxa de acertos descritos na Tabela I. Para a taxa de acerto geral do teste da rede, através da utilização dos métodos de *Horn-Schunck* e *Lucas-Kanade* para calcular o fluxo óptico, os valores obtidos com o uso de 10 neurônios apresentaram maior acurácia, sendo 77,0% para HS e 79,0% para LK. Estes resultados são esperados, uma vez que houve uma maior concentração de dados para o treinamento da rede. O sistema proporcionou resultados satisfatórios, embora tenha um vocabulário bastante limitado e, além disso, depende somente de um locutor.

TABELA I  
TAXA DE ACERTOS NA SAÍDA DA REDE (TESTES) - (EM %)

	Horn-Schunck (HS)		Lucas-Kanade (LK)	
	10 neurônios	20 neurônios	10 neurônios	20 neurônios
zero	80,0	68,0	81,0	77,0
um	91,0	92,0	92,0	94,0
dois	58,0	60,0	68,0	60,0
três	60,0	57,0	55,0	44,0
quatro	82,0	83,0	86,0	85,0
cinco	90,0	90,0	92,0	87,0
seis	59,0	57,0	55,0	54,0
sete	80,0	78,0	78,0	78,0
oito	85,0	91,0	89,0	83,0
nove	85,0	82,0	92,0	86,0
<b>Geral</b>	<b>77,0</b>	<b>76,0</b>	<b>79,0</b>	<b>75,0</b>

## VIII. CONCLUSÃO

Neste trabalho foi desenvolvido um sistema de reconhecimento de fala através do rastreamento dos movimentos dos lábios a partir de um vídeo utilizando a análise de fluxo óptico para funcionar em ambientes ruidosos. Os resultados obtidos demonstram experimentalmente que a utilização de características visuais, podem constituir uma abordagem eficaz para aplicações de reconhecimento de fala em ambientes ruidosos. Tal eficácia pôde ser analisada na seção de resultados deste artigo, onde para a configuração da rede neural descrita, obteve-se a taxa de acerto de 79% na execução dos testes utilizando o método de *Lucas-Kanade*. Ainda é possível afirmar que, comparando as técnicas de cálculo do fluxo óptico, as mesmas não apresentaram discrepâncias nos resultados obtidos, sendo possível considerar empate técnico entre os dois métodos utilizados nas simulações.

## AGRADECIMENTOS

Agradeço aos colegas que contribuíram para que este trabalho fosse possível e ao Professor Dr. Gustavo Fernandes pela oportunidade e auxílio, que cooperou de forma imensurável para a realização deste artigo.

## REFERÊNCIAS

- [1] Tomoaki Yoshinaga, Satoshi Tamura, Koji Iwano, and Sadaoki Furui. *Audio-visual speech recognition using lip movement extracted from side-face images*. In PROC. AVSP2003, ST JORIOZ, 2003.
- [2] Kohei Arai, *Voice Recognition Method with Mouth Movement Videos Based on Forward and Backward Optical Flow*, (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 2, No. 2, 2013.
- [3] G. Potamianos, C. Neti, G. Gravier, A. Garg and A. W. Senior, *Recent advances in the automatic recognition of audio-visual speech*, in Proc. of the IEEE 91, pag. 1306-1326, 2003.
- [4] G. Potamianos, C. Neti, J. Luetin and I. Matthews, *Audio-visual automatic speech recognition: An overview*, in *Issues in Visual and Audio-Visual Speech Processing*. MIT Press (Cambridge, MA, 2004).
- [5] MathWorks, *Toolbox Computer Vision*. Disponível em: <http://goo.gl/PuOu0e>. Acessado em Dezembro de 2014.
- [6] Viola, Paul and Michael J. Jones, *Rapid Object Detection using a Boosted Cascade of Simple Features*, Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. Volume: 1, pp.511-518.
- [7] Horn, B. K. P.; Schunck, B. G. *Determining optical flow*. Artificial Intelligence, v. 17, p. 185-203, 1981.
- [8] Shi, Y. Q.; Sun, H. *Image and video compression for multimedia engineering: fundamentals, algorithms and standards*. Boca Raton, CRC, 2000. 480p.
- [9] A. Mitiche and A. reza Mansouri, *On convergence of the Horn and Schunck optical-flow estimation method*, IEEE Transactions on Image Processing, vol. 13, pp. 848-852, 2004.
- [10] B. Lucas and T. Kanade, *An iterative image registration technique with an application to stereo vision (IJCAI)*, in Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI 81), April 1981, pp. 674-679.
- [11] Beauchemin, S. S.; Barron, J. L. *The computation of optical flow*. ACM Computing Surveys, v. 27, n. 3, p. 433-467, 1995.
- [12] Barron, J.L., D.J. Fleet, S.S. Beauchemin, and T.A. Burkitt. *Performance of optical flow techniques*. CVPR, 1992.
- [13] Jolliffe IT. *Principal component analysis*. New York: Springer-Verlag, 1986.
- [14] Ye J, Janardan R, Li Q. *GPCA: an efficient dimension reduction scheme for image compression and retrieval*. In: Conference on Knowledge Discovery in Data Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. Seattle (WA); 2004.
- [15] Golub G.H., Van Loan CF. *Matrix computations*. 3rd ed. Baltimore (MD): The Johns Hopkins University Press, 1996.
- [16] Richardson M., *Principal Component Analysis*. May, 2009.
- [17] Haykin S., *Neural Networks and Learning Machines*. Prentice Hall, 3rd ed., 2008.