INTERPRETAÇÃO DE OBJETOS EM CONTEXTO

MARIANA C. SPERANDIO[1], FELLIPE A. S. SILVA[1], PAULO E. SANTOS[2].

1. *Centro Universitário da FEI*
Av. Humberto de Alencar Castelo Branco 3972, 09850-908 São Bernardo do Campo, SP, Brasil
E-mails: mariana.sperandio@gmail.com, fellipeaugusto.ssilva@gmail.com

2. *Departamento de Engenharia Elétrica, Centro Universitário da FEI*
Av. Humberto de Alencar Castelo Branco 3972, 09850-908 São Bernardo do Campo, SP, Brasil
*E-mail:*psantos@fei.edu.br

Resumo— Um tema ainda em questão para sistemas artificiais de visão é a interpretação de cenas, apesar de esta ser uma tarefa que nos pareça natural. Um dos principais motivos para o lento avanço nessa área é a ausência de estruturas que representem conhecimento de alto nível. Tais descrições poderiam incluir, por exemplo, proposições sobre reconhecimento de imagens em cenas quaisquer a partir de uma imagem inicial. O objetivo deste trabalho é o estudo e verificação de melhorias em um sistema de interpretação de imagens que utiliza o contexto dos objetos na imagem para realizar sua identificação. Isto é feito através do método de interpretação de imagens conhecido como algoritmo SIFT, capaz de detectar pontos específicos em uma cena qualquer, juntamente com comparações probabilísticas entre objetos em contexto utilizando redes Bayesianas.

Palavras-chave— processamento inteligente de imagens, redes Bayesianas, SIFT e visão computacional.

1 Introdução

No campo de visão computacional, até o início da década de 70, existia uma enorme barreira ao se tentar analisar e segmentar imagens digitalmente, já que não havia algoritmos eficientes que tratassem e representassem incertezas. Desta forma, não era possível realizar o desenvolvimento de um sistema de visão de alto nível que pudesse fazer com que um agente inteligente interpretasse o mundo de forma autônoma, e agisse de acordo com essa interpretação.

Após a década de 80, quando a evolução dos computadores já permitia o processamento de grandes conjuntos de dados, iniciaram-se estudos mais aprofundados de processamento de imagens.

Um métodos capaz de processar imagens é o SIFT (do inglês *Scale Invariant Feature Transform*) (Lowe, 2004). SIFT é um algoritmo de visão computacional que tem a habilidade de detectar e descrever características locais de uma imagem, podendo assim identificar objetos contidos nesta imagem, mesmo que eles estejam em desordem, rotacionados ou sob oclusão parcial (Lowe, 2004). Isso ocorre pelo fato de o algoritmo SIFT ser invariante em relação à escala, rotação, ponto de vista, distorção e parcialmente invariante à mudança de iluminação (Lowe, 2004).

Este artigo tem como objetivo relatar o desenvolvimento de um sistema de reconhecimento de objetos em imagens utilizando contexto modelado por redes Bayesianas.

O algoritmo SIFT foi utilizado neste sistema de interpretação de imagens para reconhecer objetos em imagens quaisquer. Para a realização do sistema de reconhecimento contextual, fez-se uso de redes Bayesianas (Charniak, 1991) através do toolbox para Matlab BNT (do inglês *Bayes Net Toolbox for Matlab*) (BNT, 2011). Por contexto entende-se a relação física espacial natural entre objetos partindo-se de uma premissa conhecida pelo ser humano, como por exemplo, quando se observa um livro em uma biblioteca e não o vê em um jardim.

Foram realizadas diversas experiências simuladas através da plataforma Matlab para se analisar a aplicação deste sistema, e, a partir destas simulações pôde-se verificar a melhora de confiabilidade em um sistema de identificação de objetos que considera o contexto em que estes se encontram na imagem.

2 Métodos

Nesta seção encontram-se as descrições dos métodos utilizados no desenvolvimento deste artigo, sendo eles, respectivamente, um método de reconhecimento de objetos e um método de cálculo de probabilidades.

O método para a geração das características das imagens, *Scale Invariant Feature Transform* (SIFT) (Lowe, 2004), foi utilizado neste artigo para a realização do reconhecimento de objetos. Este método possui num total seis procedimentos principais, descritos a seguir.

O primeiro procedimento é a preparação inicial do algoritmo, a construção de um espaço escala, onde são criadas representações internas da imagem original para garantir a invariância da escala.

Em seguida a aproximação LoG (do inglês *Laplacian of Gaussian*) é utilizada para encontrar pon-

tos de interesse, ou pontos-chave, em uma imagem. Como esta aproximação é computacionalmente cara, utiliza-se o espaço escala criado anteriormente para este fim.

A partir da aproximação anterior, é possível encontrar pontos-chave, estes pontos são os máximos e mínimos valores da diferença de Gaussiana da imagem.

Posteriormente, deve-se excluir as bordas e as regiões de baixo contraste, estas são classificadas como pontos-chave ruins, estes pontos são eliminados para tornar o algoritmo mais eficiente.

Uma orientação é criada para cada ponto chave, os próximos cálculos são feitos a partir desta orientação, desta forma, as características da imagem tornam-se invariantes à rotação.

Por fim, gera-se uma identificação para cada ponto-chave, assim estes podem ser reconhecidos quando comparados com objetos similares.

O entendimento do algoritmo SIFT pode ser mais aprofundado analisando-se (Lowe, 2004).

O método de cálculo de incertezas e probabilidades foi feito através de redes Bayesianas (Russell, 2009). Redes Bayesianas são representações gráficas que, quando utilizadas em sistemas podem simplificar as relações de causalidade entre suas variáveis (Russell, 2009).

Uma rede Bayesiana possui nós que representam as variáveis, contínuas ou discretas, de um domínio, e também arcos que interligam diferentes nós. Para representar a dependência dos nós conectados entre si, utilizamos probabilidades (Jain, 2009).

Matematicamente, uma rede Bayesiana é a associação de um gráfico acíclico a um conjunto de distribuição de probabilidade Bayesiana (Korb, 2003). Cada nó da rede está relacionado à uma variável (Xi), e seus arcos indicam a relação de probabilidade presente entre o nó atual e os anteriores, estes nós que antecedem e derivam o nó atual estão relacionados à variáveis denominadas de pais $(pa(Xi))$. A probabilidade condicional de uma rede Bayesiana é a probabilidade de a variável (Xi) depender de seus pais $(pa(Xi))$, ela pode ser denotada por $p(Xi \mid pa(Xi))$. Esta distribuição de probabilidade é dada por (1).

$$p(Xi...Xn) = \prod_{i=1}^{n} p(Xi \mid pa(Xi)) \qquad (1)$$

A utilização de redes Bayesianas pode ser considerada uma boa estratégia para lidar com sistemas que tratam incertezas, sendo que nestes sistemas não se pode construir conclusões derivadas apenas do conhecimento precedente do problema (Charniak, 1991). Isso se deve pelo fato de o raciocínio probabilístico poder tomar decisões racionais mesmo que haja pouca informação que prove alguma ação (Charniak, 1991).

Neste artigo, utilizam-se redes Bayesianas para levar em consideração o contexto da imagem (Russell, 2009) em que o objeto se encontra para calcular a probabilidade da interpretação do algoritmo SIFT estar correta.

Concluindo, este sistema verifica a localização de um objeto correto em um local apropriado, como por exemplo, a identificação da imagem de um forno microondas como sendo um forno microondas, e não um televisor ou um monitor de computador, levando em consideração outros objetos referentes à imagem de uma cozinha. Para que isso ocorra é necessário que haja um encadeamento de idéias que se baseiem em outros objetos da imagem para pressupor a veracidade do objeto identificado.

3 Resultados

Para o desenvolvimento de um sistema probabilístico que melhore a confiabilidade de reconhecimento do algoritmo SIFT, foram criadas cinco redes Bayesianas distintas através das quais se pôde calcular a probabilidade de a imagem reconhecida ser real de acordo com o contexto em que se encontra.

Dentre estas cinco redes criadas serão exemplificadas apenas duas, como se pode observar nas figuras abaixo (figuras 1 e 2), sendo as outras três redes análogas a estas.

Na figura 1 pode-se notar a rede Bayesiana conhecida como 'Imagem_Caneca' e também as relações de probabilidade presente em todas as redes de forma equivalente a esta. Enquanto que na figura 2 somente apresenta a estrutura da rede Bayesiana 'Imagem_Secador'.

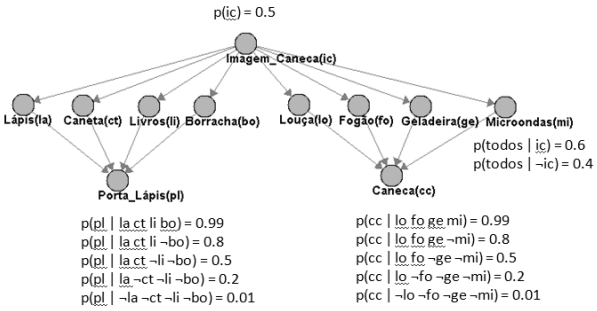

Figura 1. Rede Bayesiana 'Imagem_Caneca'

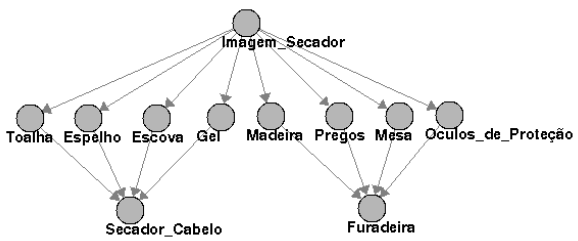

Figura 2. Rede Bayesiana 'Imagem_Secador'

Pode-se descrever as redes ilustradas anteriormente da seguinte forma, a rede ‘Imagem_Caneca’ associa a imagem de uma caneca às imagens ‘lápis’, ‘caneta’, ‘livros’ e ‘borracha’ para identificar esta caneca como ‘porta-lápis’, enquanto que esta mesma associação é feita com as imagens ‘louça’, ‘fogão’, ‘geladeira’ e ‘micro-ondas’ para identifica-la como ‘caneca’. Da mesma forma a rede ‘Imagem_Secador’ associa a imagem de um secador de cabelos às imagens ‘toalha’, ‘espelho’, ‘escova’ e ‘gel’ para identifica-la como ‘secador’ enquanto que esta mesma associação é feita com as imagens ‘madeira’, ‘pregos’, ‘mesa’ e ‘óculos de proteção’ para identificá-la como ‘furadeira’.

Neste sistema de interpretação de imagens, o algoritmo SIFT identifica em uma cena todos os objetos pertencentes às redes, descritos anteriormente, e a partir de reconhecimentos falsos e verdadeiros, calcula-se através da inferência Bayesiana, a probabilidade deste resultado ser confiável.

Isso significa que, se o algoritmo SIFT identificar um objeto como sendo um ‘secador’ e identifica mais três objetos na cena, entre eles a ‘madeira’, os ‘pregos’ e os ‘óculos de proteção’, então, através do cálculo de probabilidades da rede Bayesiana, verifica-se que é mais provável que o objeto identificado seja a ‘furadeira’ e menos provável que seja o ‘secador’.

Após a criação das cinco redes Bayesianas, pôde-se desenvolver um sistema de interpretação de imagens baseado em probabilidade, com o propósito de maximizar a confiabilidade de reconhecimento do algoritmo SIFT.

Pode-se, através deste sistema, identificar diversas imagens e a partir do contexto dos objetos localizados, analisar se o objeto que foi encontrado na imagem, verdadeiramente é o objeto que deveria ter sido encontrado.

Para realizar o cálculo de probabilidade em redes Bayesianas fez-se uso do toolbox para Matlab BNT (do inglês *Bayes Net Toolbox for Matlab*) (BNT, 2011), este toolbox nos possibilita calcular facilmente as probabilidades de qualquer nó em uma rede bayesiana descrita utilizando-se o Matlab.

Para isso, o software Matlab recebe os nomes das variáveis da rede Bayesiana em questão e realiza o reconhecimento de cada objeto, cada nó da rede, utilizando o algoritmo SIFT, em seguida inicia os cálculos de probabilidade para mostrar qual dos nós da rede é o mais provável de ser verdadeiro.

Pode-se ilustrar a operação deste sistema de interpretação de imagens contextuais através da rede Bayesiana ‘Imagem_Caneca’, apresentada na figura 1. Para cada nó de cada rede foram utilizadas trinta imagens distintas, totalizando 300 imagens para cada rede do sistema (1500 imagens totais no sistema). Com o intuito de demonstrar o funcionamento do sistema, só serão ilustradas uma imagem para cada nó. Para a rede ‘Imagem_Caneca’, estas imagens podem ser vistas nas figuras 3, 4, 5, 6, 7, 8, 9, 10, 11 e 12.

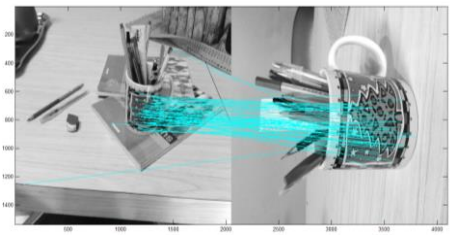

Figura 3. ‘Imagem_Caneca’ e ‘Porta_Lápis24’

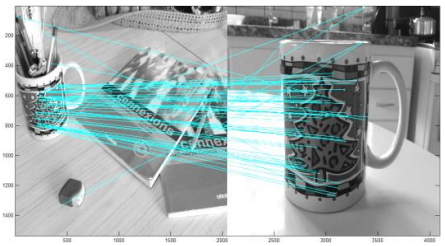

Figura 4. ‘Imagem_Caneca’ e ‘ Caneca24’

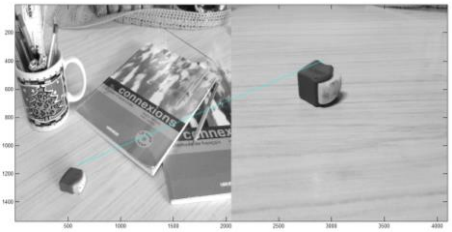

Figura 5. ‘Imagem_Caneca’ e ‘Borracha24’

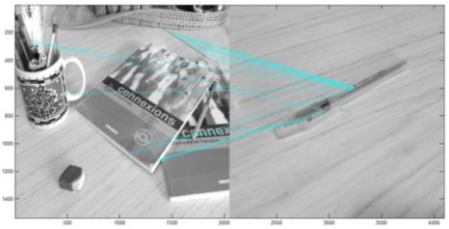

Figura 6. ‘Imagem_Caneca’ e ‘Caneta24’

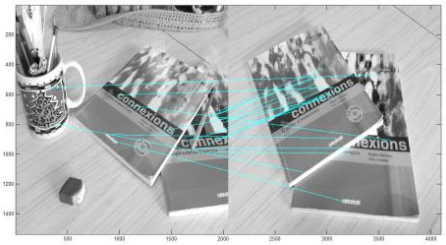

Figura 7. ‘Imagem_Caneca’ e ‘Livros24’

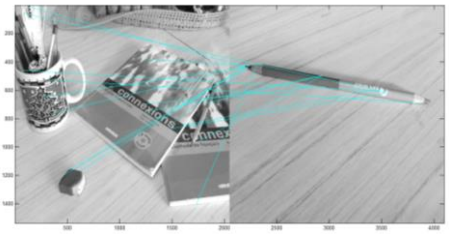

Figura 8. ‘Imagem_Caneca’ e ‘Lápis24’

Figura 9. ‘Imagem_Caneca’ e ‘Fogão24’

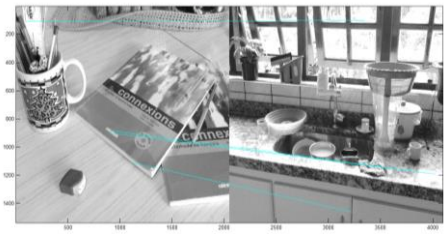

Figura 10. ‘Imagem_Caneca’ e ‘Louça24’

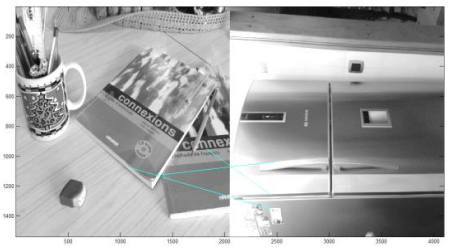

Figura 11 ‘Imagem_Caneca’ e ‘Geladeira24’

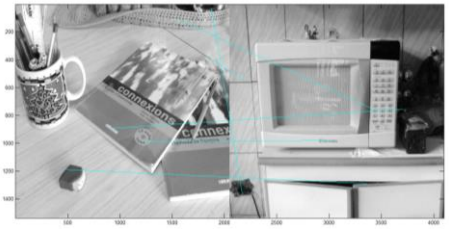

Figura 12. ‘Imagem_Caneca’ e ‘Microondas24’

Através das imagens anteriores pode-se confirmar a invariância de rotação do algoritmo SIFT, ou seja, mesmo que os objetos possuam rotações diferentes entre as imagens, o algoritmo SIFT as identifica de forma positiva.

O sistema classifica uma imagem como similar quando o algoritmo SIFT encontra números de pontos-chave iguais ou superiores a 10. Na tabela abaixo pode-se entender a lógica utilizada para realizar o cálculo de probabilidades do sistema.

Tabela 1. Tabela de funcionamento ‘Imagem_Caneca’

Imagem_Caneca.pgm	Encontrado?	Pontos-chave
Porta_Lapis24.pgm	Sim	129
Caneca24.pgm	Sim	61
Borracha24.pgm	Não	1
Caneta24.pgm	Sim	18
Livros24.pgm	Sim	20
Lapis24.pgm	Sim	18
Fogao24.pgm	Não	4
Louca24.pgm	Não	5
Geladeira24.pgm	Não	3
Microondas24.pgm	Não	6
Probabilidade de ser:	Resultado	Probabilidade
Porta_Lapis	Verdadeiro	0.9140
Caneca	Falso	0.6086

A tabela 1 fornece as quantidades de pontos-chave totais e sua interpretação de similaridade, ou seja, se foram identificados 10 ou mais pontos-chave, o sistema classifica o objeto como encontrado, se não, o classifica como não encontrado.

Novamente, para a verificação do funcionamento do sistema desenvolvido por este projeto foram realizados 30 testes para cada rede Bayesiana criada, ou seja, o processo descrito anteriormente foi feito trinta vezes com trinta imagens diferentes, para as cinco redes criadas, uma a uma. Os resultados de probabilidade obtidos em cada teste foram utilizados para calcular a veracidade do sistema, estes cálculos serão descritos em seguida.

4 Discussões

Para o sistema de reconhecimento de imagens baseado em contexto pode-se obter quatro diferentes resultados: Verdadeiro Positivo (VP), Verdadeiro Negativo (VN), Falso Positivo (FP) e Falso Negativo (FN).

Um objeto que foi identificado de forma correta possui um resultado Verdadeiro Positivo (VP). Já se o objeto foi encontrado, porém não está presente na imagem, este resultado é classificado como Falso Positivo (FP). Se o objeto não foi identificado, e realmente não deveria ter sido, pois não está contido na imagem, o resultado é dito Verdadeiro Negativo (VN). Da mesma forma, se o objeto não foi identificado, porém deveria ter sido, pois está presente na imagem, o resultado é considerado Falso Negativo (FN). A tabela 2 nos ajuda a entender melhor estas definições.

Tabela 2. Tabela de definição de VP, VN, FP e FN

	Identificação da Imagem	
Resultado do Sistema	Presente	Ausente
Positivo	VP	FP
Negativo	FN	VN

Analisando os resultados obtidos através dos testes descritos anteriormente, podem-se verificar as quantidades de resultados dos tipos VP, VN, FP e FN que foram obtidos no sistema. Considerando o reconhecimento Positivo como aquele que obteve probabilidade maior ou igual a 75% de certeza. Essas quantidades podem ser observadas na tabela 3.

Tabela 3. Resultados do sistema

	VP	VN	FP	FN
Imagem_Boné	26	30	0	3
Imagem_Caneca	30	25	5	0
Imagem_Celular	26	93	3	6
Imagem_Secador	19	28	2	11
Imagem_TV	22	102	4	10
Sistema	123	278	14	30

A fim de avaliar o sistema desenvolvido neste projeto foram calculados quatro parâmetros de verificação possíveis. São eles a acurácia, a sensibilidade, a precisão e a especificidade do sistema (Souza, 2011).

A acurácia é a medida de confiabilidade do sistema, ou seja, é o numero de previsões corretas sabendo-se o numero total de previsões (Souza, 2011). A acurácia nos informa o quanto o valor estimado do sistema se aproxima do valor real. É dada pela equação (2).

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (2)$$

Outro parâmetro utilizado é a sensibilidade do sistema. A sensibilidade é a fração de previsões verdadeiras existentes que foram detectadas como tal (Souza, 2011). A sensibilidade é dada pela equação (3).

$$Sensibilidade = \frac{VP}{VP + FN} \quad (3)$$

A precisão do sistema é dada pela equação (4) e indica a detecção de predições verdadeiras e positivas (Souza, 2011).

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (4)$$

Por fim, calcula-se a especificidade do sistema, dada pela equação (5), que nada mais é que a medida de frequência na qual o sistema caracteriza um resultado falso como um resultado verdadeiro e negativo (Souza, 2011).

$$Especifidade = \frac{VN}{VN + FP} \quad (5)$$

A tabela 4 mostra o cálculo dos parâmetros para a avaliação deste sistema de reconhecimento de imagens considerando o contexto entre cenas.

Tabela 4. Cálculo de avaliação do sistema

	Acurácia (%)	Sensibilidade (%)	Precisão (%)	Especificidade (%)
Imagem_Boné	94,92	89,66	100,00	100,00
Imagem_Caneca	91,67	100,00	85,71	83,33
Imagem_Celular	92,97	81,25	89,66	96,88
Imagem_Secador	78,33	63,33	90,48	93,33
Imagem_TV	89,86	68,75	84,62	96,23
Sistema	90,11	80,39	89,78	95,21

Sabendo-se que quanto maior a eficácia de um sistema, maior são seus valores de acurácia, sensibilidade, precisão e especificidade, podemos afirmar que nosso sistema possui alto valor de eficiência (Souza, 2011).

Nosso sistema obteve valores elevados para todos os parâmetros descritos, sendo que obteve 80% de sensibilidade, aproximadamente 90% de acurácia e de precisão e 95% para especificidade.

5 Conclusões

O SIFT pode ser considerado um algoritmo de alta confiabilidade, pois apresenta ótimos resultados experimentais, mesmo tendo identificado alguns pontos falsos em reconhecimentos verdadeiros. Em grande parte dos testes realizados, o objeto requerido foi identificado com sucesso dentro da imagem, mesmo quando ele se encontrava parcialmente oculto ou em outro ângulo de visão.

O algoritmo se mostrou invariante à escala, rotação, ponto de vista e à distorção da imagem, como já havia sido mencionado anteriormente, porém mostrou-se sensível as mudanças na intensidade de iluminação e também quanto à variedade de coloração das imagens. Foi possível perceber estas considerações através das redes 'Imagem_Secador' e 'Imagem_TV', estas obtiveram valores reduzidos do pa-

râmetro sensibilidade. Isto se deu justamente pela variação precária de coloração e também pela alta intensidade de iluminação das imagens utilizadas.

Como o algoritmo SIFT não possui a capacidade de levar em consideração o contexto em que o objeto se encontra na cena determinada, para poder identificá-lo com veracidade e precisão, foi preciso criar um sistema baseado em redes Bayesianas que fornecesse a probabilidade de uma imagem ser identificada com confiança.

Através dos resultados adquiridos e demonstrados, pode-se afirmar que este sistema atingiu alto grau de credibilidade e pode ser utilizado com diversas finalidades de interpretação de imagens. Possibilitando uma melhoria em um sistema de reconhecimento de imagens, pois agora se pode levar em conta toda a análise de contexto presente em uma cena.

Este sistema pode ser implementado, por exemplo, em um robô móvel real, que interpreta uma cena e identifica seus objetos com eficácia enquanto caminha pelo ambiente, possibilitando sua localização em um mapa conhecido ou até mesmo na elaboração deste em um ambiente desconhecido. Como evolução deste sistema de interpretação de imagens pode-se automatizar este processo de criação de redes bayesianas utilizando aprendizado de classificadores Naïve Bayes (Neapolitan, 2004).

O desenvolvimento do presente projeto possui extrema importância para o mundo científico e acadêmico, já que contribui com áreas de pesquisas que puderam ser integradas, formando um sistema autônomo de reconhecimento de objetos em imagens, que leva em consideração o contexto destas para melhor atribuir um resultado satisfatório e eficiente.

Referências Bibliográficas

BNT. Bayes Net Toolbox para MATLAB. (2011). Disponível em: <http://code.google.com/p/bnt/>.

Charniak, Eugene. (1991). Bayesian networks without tears. AI Magazine, pp.50-63.

Jain, Dominik; Mosenlechner, Lorenz; Beetz, Michael. (2009). Equipping Robot Control Programs with First-Order Probabilistic Reasoning Capabilities. International Conference on Robotics and Automation.

Korb, Kevin B.; Nicholson, Ann E. (2003). Bayesian Artificial Intelligence. Technometrics, pp101-102.

Lowe, David G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, Vol. 2, pp. 91-110.

Neapolitan, Richard E. (2004). Learning Bayesian networks. Prentice Hall.

Russell, Stuart J.; Norving, Peter. (2009). Artificial Intelligence: A Modern Approach. AI Book, 5ª Edição.

Souza, Carlos R. C.; Santos, Paulo E. (2011). Probabilistc Logic Reasoning About Traffic. Annual conference on Towards autonomous robotic systems. Berlin.