

Utilização de Quantização Vetorial para Reconhecimento de Edifícios em Ambientes Urbanos

Using Vector Quantization for Building Recognition in Urban Environments

Eduardo Silva Vasconcelos ^{1*}, Débora Vasconcelos Melo ², Leandro Aureliano da Silva ³,
Gilberto Arantes Carrijo ¹

RESUMO

O reconhecimento de edifícios é essencial para uma variedade de aplicações. Devido à sua importância, este trabalho tem por objetivo analisar as taxas de reconhecimento de imagens de edifícios, para 40 classes e 30 imagens por classe, na escala de cor RGB, utilizando um sistema de reconhecimento baseado na técnica de Quantização Vetorial para compressão de imagens por meio do algoritmo LBG, variando a quantidade de centroides e treinos. Para a discussão dos resultados fez-se uma análise estatística descritiva e para a estatística inferencial utilizando o teste paramétrico da ANOVA, com post-hoc de Tukey, com o nível de significância de $p \leq 0,05$. Os resultados alcançados obtiveram uma alta taxa de acerto, de 100%, quando trabalhado com 128 e 256 centroides. Não foi encontrado diferença significativa entre os treinos. Encontrou-se diferença significativa somente entre as taxas de reconhecimento processadas com 16 centroides e as demais quantidades de centroides.

Palavras-chave: Reconhecimento de Imagens, Quantização Vetorial, Reconhecimento de Edifícios.

ABSTRACT

Building recognition is essential for a variety of applications. Due to its importance, this work aims to analyze the recognition rates of building images, for 40 classes and 30 images per class, in RGB color scale, using a recognition system based on the Vector Quantization technique for image compression through the LBG algorithm, varying the number of centroids and training. For the discussion of the results a descriptive statistical analysis was performed, and for the inferential statistics using the parametric ANOVA test, with Tukey post-hoc, with the significance level of $p \leq 0.05$. The results achieved a high hit rate of 100% when working with 128 and 256 centroids. No significant difference was found between the training. A significant difference was found only between recognition rates processed with 16 centroids and the other quantities of centroids.

Keywords: Image Recognition, Vector Quantization, Building Recognition.

¹ Universidade Federal de Uberlândia

*E-mail: educelos1@gmail.com

² Universidade Federal de Catalão

³ Universidade de Uberaba

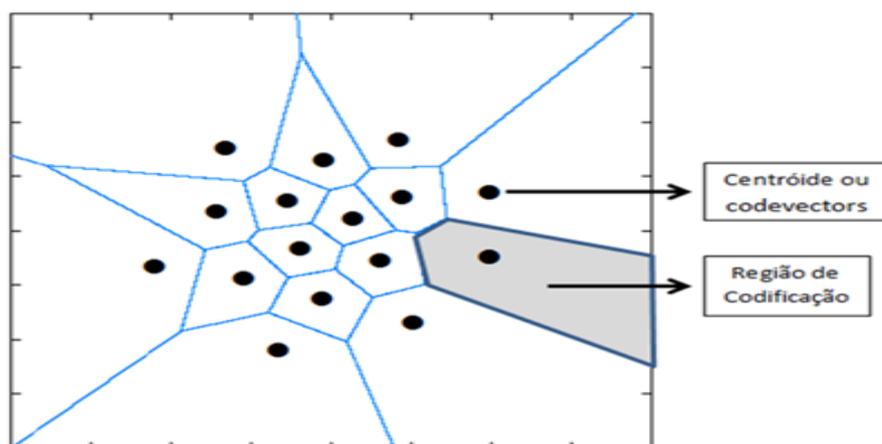
INTRODUÇÃO

Técnicas de processamento de imagens são aplicadas a praticamente todas as ciências, segundo Albuquerque (2000) processar uma imagem tem o objetivo de viabilizar a extração das informações nela presentes. Além disso, técnicas de agrupamento de dados tem se tornado um tópico muito importante dentro de uma área mais ampla conhecida como processamento de dados e imagens digitais.

Gonzales e Woods (2000) destacam que o alvo do processamento de imagens é a digitalização de valores de amostras de uma imagem por meio da compressão de seus dados e agrupamentos. A compressão dos dados e dos agrupamentos é conhecido como clusterização, que é o processo de agrupar um conjunto de dados em classes ou grupos, também chamados de clusters, de forma que amostras do mesmo grupo apresentem alta similaridade entre si. Para tanto, existe a possibilidade de melhoria do desempenho dessa compressão utilizando técnicas de Quantização Vetorial (QV), uma técnica não atual, mas relevante em sistemas de mapeamento e imagens (MADEIRO, 2012).

A QV traz como objetivo representar distribuições de dados utilizando um número de referência de padrões significativamente menor que o número de dados (WIGGERS, 2013). Dessa forma, a compressão de um conjunto de dados busca diminuir a quantidade de informação, tanto para transmissão quanto para armazenamento (GERSHO, 1992), pois apenas os centroides precisam ser armazenados, ao invés da base de dados inteira.

Figura 1 – Ilustração de Quantização Vetorial em duas dimensões.



Fonte: VASCONCELOS, E. S.; MELO, D. V.; CARRIJO. (2022)

A Figura 1 apresenta os pontos conhecidos como centroides ou codevectors, ela mostra um exemplo de Quantização Vetorial (QV) em duas dimensões. Considera-se que a Quantização Vetorial é uma forma de compressão de dados, pois apenas os centroides precisam ser armazenados, ao invés da base de dados inteira (MADEIRO, 2011).

As regiões definidas pelas linhas de fronteiras são intituladas regiões de codificação. A região destacada exemplifica uma região de codificação, sendo que essa possui um centroide correspondente à ela, de forma que para mapear qualquer dado dessa região é suficiente o centroide. O conjunto de todos os centroides é conhecido por codebook ou livro de códigos, que é uma lista previamente computada, e o conjunto de todas as regiões de codificação é chamado de partição do espaço.

Ressalta-se que, além da compressão de imagens, há um amplo espectro de aplicações para a QV, por exemplo, esteganografia (CHANG, 2012), marca d'água digital (MADEIRO, 2012), identificação vocal (SRINIVASAN, 2012) e classificação de sinais de voz com patologias (VIEIRA, 2012).

Contudo, na QV a complexidade computacional existente na fase de codificação dos vetores a serem quantizados, constitui alguns desafios, tais como o projeto de criação de codebooks e a sensibilidade da técnica aos erros de transmissão (MADEIRO, 2012). Sendo assim, alguns algoritmos podem ser empregados.

Quando incorporados leva a uma redução do tempo de execução no processamento de imagens além de encontrar um codebook ótimo para a representação vetorial do conjunto de dados processado, de forma a minimizar o erro inerente quando utilizado para comprimir imagens, entre eles está o algoritmo Linde-Buzo-Gray (LBG).

O LBG é uma técnica que inferi a maneira como os dados estão organizados e relacionados em cada grupo, ele divide a imagem em vetores-referências ou centroides, que servem como elemento representativo dos agrupamentos obtidos, criando um novo codebook que é iterativamente atualizado (LINDE, 1980). A quantidade de centroides (C) é escolhida com base na Equação 1, onde, m é um valor de entrada pré-estabelecido (SCOLARI, 2007).

$$C = 2^m \quad (1)$$

A similaridade da imagem processada com a imagem original depende diretamente do codebook. A distância entre os vetores e os centroides pode ser medida de diversas formas, sendo a mais usada, a Distância Euclidiana ao Quadrado (Equação 2).

$$d(x, \hat{x}) = \sum_{i=0}^{k-1} |x_i - \hat{x}_i|^2 \quad (2)$$

Essas técnicas são essenciais para muitos serviços de localização, aplicações geoespaciais, exército, engenharia e arquitetura (SILVA, 2014), área imobiliária, planejamento urbano, turismo virtual de uma cidade, navegação de carro (RODRIGUES, 2016), detecção automática de alvos e reconhecimento de edifícios. O reconhecimento de edifícios é uma ação significativa e apresenta uma grande escala de desafios, pois para cada edifício contém uma enorme variabilidade de parâmetros, como pontos de vista, condições de iluminação, bloqueio por árvores, veículos e de outras construções.

Diante disso, o trabalho tem como objetivo analisar os resultados no reconhecimento de imagens de edifícios do banco de dados SBID (<http://eeepro.shef.ac.uk/building/dataset.rar>), para 40 classes (pastas) e 10 imagens por classe, nas escalas de cores cinza, vermelha, verde e azul, utilizando a técnica QV com o algoritmo LBG.

METODOLOGIA

Uma técnica de reconhecimento de edifícios ideal deve-se investigar os seguintes aspectos: (1) modelos visuais que podem descrever com precisão os edifícios e ser sensíveis a pequenas alterações na imagem e (2) rápida compressão para melhorar a eficiência de um algoritmo, enquanto diminui o espaço de armazenamento de dados e complexidade computacional (LI, 2014). Esse reconhecimento consiste em três partes, (i) representação das características, (ii) correspondência das características e (iii) classificação das características.

Após essa representação a correspondência de características é conduzida para encontrar semelhanças entre as imagens, isto é, a imagem de consulta e uma imagem de referência no banco de dados. Finalmente, a classificação é conduzida para determinar a melhor correspondência, onde os classificadores em relação aos modelos estatísticos combinam as saídas dos vetores de características globais ou locais de mesma aparência para maximizar a qualidade da saída em um conjunto de treinamento (LI, 2014).

Para processamento dos dados e obtenção dos resultados foi utilizado o programa Matlab versão R2015a para todos os processamentos. E, o computador usado foi um ProBook da HP, com processador Intel Inside, core i7 vPro.

O banco de dados e sua complexidade: SBID

Conforme Li (2014), o banco de dados SBID torna a tarefa de reconhecimento de edifícios desafiadora, pois combina diferentes condições de iluminação e pontos de vista. As imagens do SBID possuem rotação, escalonamento, diferentes condições de iluminação, alterações de pontos de vista, oclusões e vibração. O banco de dados é composto por 3.192 imagens, no formato JPEG, de 40 edifícios que incluem igrejas e uma variedade de edifícios modernos, como salas de exposições e prédios de escritórios (Figura 2).

O tamanho das imagens é fixado em 160 x 120 pixels, a fim de garantir eficiência computacional. Contudo, para cada classe o número de imagens é diferente.

Figura 2 – Exemplo de imagens do conjunto de dados SBID.



Fonte: Database da pesquisa

Essas imagens foram tiradas de prédios da Universidade de Sheffield e do centro da cidade de Sheffield em diferentes dias e horários ano de 2008.

O método proposto

O método proposto compreende o processamento das imagens nas escalas de cores RGB. A tarefa de reconhecimento, apresentada neste trabalho, utiliza a Quantização Vetorial (QV) para construção dos padrões representativos. Sendo que para o projeto dos dicionários do quantizador vetorial é realizado o algoritmo tradicional LBG (LINDE, 1980).

Os resultados foram analisados a partir das variáveis: centroides e porcentual de treino. Foi fixado para esse trabalho a análise de 30 imagens em 40 classes. Estipulou-se as primeiras 30 imagens de cada classe, tendo em vista que a distribuição, dessas imagens nas classes, é aleatória. Assim, como trabalhou-se com 40 classes, teve-se no total 1.200 imagens analisadas.

Por padrão, cada treino foi executado, separadamente, com 40%, 50% e 60% das imagens, por classe, e o restante das imagens foram usadas para o reconhecimento. Dessa maneira, para 40% de treino, 480 imagens foi para o treino e 720 para o reconhecimento, para 50% de treino, 600 imagens foram para o treino e outras 600 para o reconhecimento e para 60% de treino, 720 imagens foram para o treino e 480 para o reconhecimento (Tabela I). Todo o procedimento foi analisado para 16, 32, 64, 128 e 256 centroides.

Tabela I – Quantidade de imagens utilizadas para o treino e para o reconhecimento.

Quantidade de	Treino	Quantidade	Quantidade total de	Total de
30	40%	480	720	1.200
	50%	600	600	
	60%	720	480	

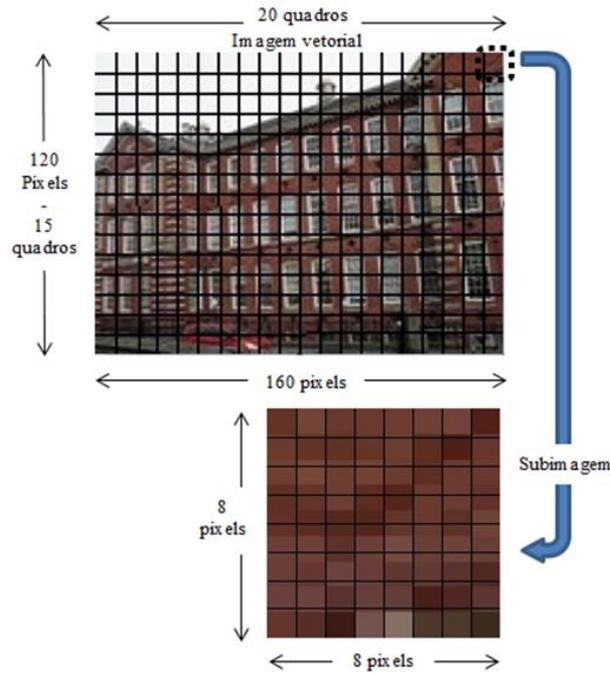
Como o algoritmo utilizado é iterativo, os codebooks gerados são obtidos pelo método de divisão das imagens, onde divide cada imagem em blocos 8 x 8 pixels, totalizando assim 300 blocos por imagem. Conforme ilustrado pela Figura 3 Com essa imagem dividida a técnica aplica a Transformada Discreta do Cosseno (DCT) e converte os dados de amplitude espacial em coeficientes de frequência espacial.

Com isso, o algoritmo realiza os treinamentos nesse conjunto de imagem e constroe um novo codebook, assim por diante. Esse codebook é utilizado no processo de reconhecimento das imagens em suas respectivas classes. Cada vetor gerado é um bloco de dimensão 64 (Equação 3), pois considerou-se somente uma escala de cor, por vez.

$$X_K = \{ \{ X \}_m, Y_m, Z_m; m=1,2,\dots,64 \}; K=1,2,\dots,300 \quad (3)$$

Em que, X_m : ordenadas da escala da cor vermelha, Y_m : ordenadas da escala da cor verde e Z_m : ordenadas da escala da cor azul.

Figura 3 – Extração de blocos de 8x8 pixels por imagem.



Fonte: Dados da pesquisa

Executa-se o algoritmo LBG dando início ao processo de reconhecimento e a classificação da imagem por classe. O resultado final consiste na obtenção do percentual de acertos do reconhecimento (Equação 4).

$$Percentage = \frac{Qa}{Qc \times (Qic - Qit)} \times 100 \quad (4)$$

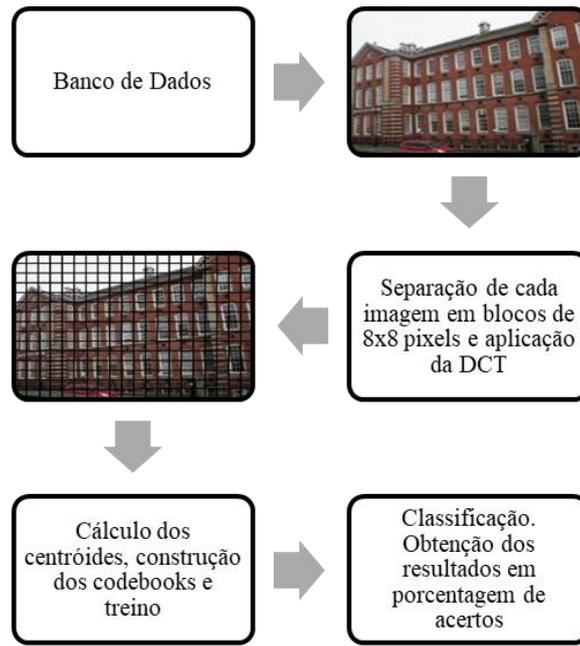
Em que, Qa : quantidade de acerto, Qc : quantidade de classes, Qic : quantidade de imagens em cada classe e Qit : quantidade de imagens para treino.

As análises considerando porcentagem de treinos e quantidade de centroides foram para verificação se ocorreriam diferenças significativas nos resultados. Considera-se, nesse trabalho, diferença significativa pela Anova com post-hoc de Tukey os p-valor $> 0,05$.

Os dados foram inseridos e processados por meio de banco de dados do programa SPSS® 20.0 e do programa Excel®, analisados por meio de estatística descritiva e inferencial. Para a estatística descritiva, os resultados foram analisados por meio de frequência percentual e média.

As etapas propostas no método empregado são apresentadas na Figura 4.

Figura 4 – Diagrama em blocos das etapas da metodologia proposta.



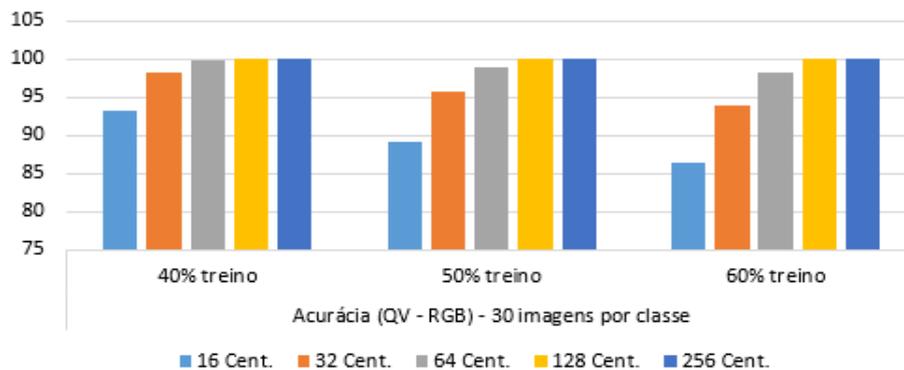
Fonte: Dados da pesquisa

METODOLOGIA

Os dados apresentados foram obtidos pelo método de QV juntamente com o algoritmo LBG. Utilizou-se o software Matlab. Foram utilizadas as 30 primeiras imagens de cada uma das 40 classes, totalizando 1.200 imagens.

Os resultados foram analisados e são apresentados conforme as quantidades de centróides (16, 32, 64, 128 e 256) e treino (40%, 50% e 60%), como observado nas Figuras 5 e 6.

Figura 5 – Porcentual de Reconhecimento – Treino em Relação à quantidade de Centróides



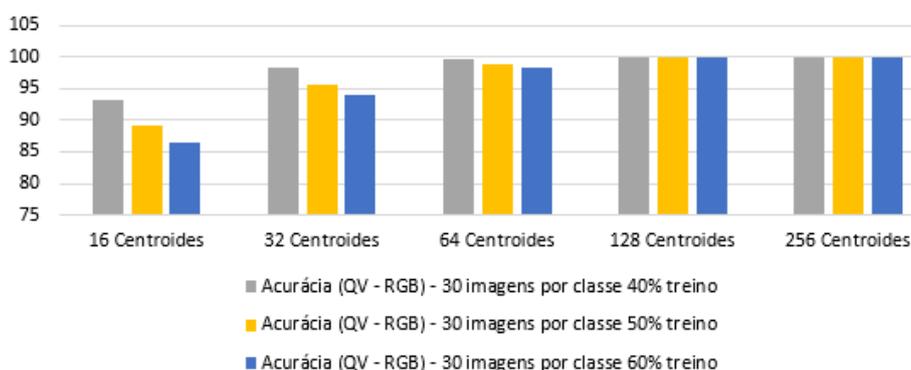
Fonte: Dados da pesquisa

Observa-se na Figura 5 que foi encontrado taxa de reconhecimento de 100% com 128 e 256 centroides, independente do treino, logo para 128 e 256 centroides tem-se variância zero.

Para 16, 32 e 64 centroides tem-se uma redução no percentual da taxa de reconhecimento conforme aumenta-se o treino.

Nota-se pela Figura 6 que ocorre aumento percentual na taxa de reconhecimento conforme aumenta-se a quantidade de centroides, sendo que fixa na taxa de 100% de reconhecimento para 128 e 256 centroides independentemente do percentual de treino utilizado.

Figura 6 – Percentual de Reconhecimento – Quantidade de Centroides em Relação ao Treino



Fonte: Dados da pesquisa

Diante dos dados apresentados das Figuras 5 e 6 nota-se uma variabilidade das taxas de reconhecimento de cada treino em relação aos centroides juntando as diferentes quantidades de imagens processadas por classe. Destaca-se que o reconhecimento para 16 centroides é o que possui a maior distribuição dos dados e para 128 e 256 centroides não houve variação do reconhecimento.

Devido às variabilidades encontradas nas taxas apresentadas e com o objetivo de verificar se existe diferenças significativas e determinar o efeito de duas variáveis independentes em uma variável dependente, para a inferência dos dados foi feito o cálculo da ANOVA de duas vias, pelo post-hoc de Tukey, com o nível de significância de $p \leq 0,05$, entre as variáveis treino e centroide, para a quantidade de 30 imagens por classe (Tabela II).

Para a estatística inferencial, comparou-se os grupos Porcentagem de Treino e Quantidade de Centroides com o objetivo de verificar a existência ou não de diferença

significativa entre eles. Para o centroide foi observado que há diferença significativa para os treinos de 40% e 50% (Tabela II).

Tabela II – Relação entre as taxas de reconhecimento dos centroides com os treinos

	Porcentagem do Treino		
	40%	50%	60%
16 cent.	93,3333b	89,1667b	86,4583b
32 cent.	98,1944a	95,6667a	93,9583a
64 cent.	99,7222a	98,8333a	98,3333a
128 cent.	100a	100a	100a
256 cent.	100a	100a	100a

Para a discussão dos dados foi caracterizado que letras iguais nas colunas das tabelas representam que não existe diferença estatística significativa entre eles, e por outro lado letras diferentes representam que existe diferença significativa. Isto é, os dados nas colunas que possuem a mesma letra, seja “a” ou “b”, significa que não existe significância entre eles, e por sua vez, quando apresenta letras diferentes, “a” e “b”, significa que foi encontrado um nível de significância de $p \leq 0,05$.

Observa-se na Tabela II que não foi encontrado diferença significativa entre os treinos. No entanto, encontrou-se diferença estatística significativa somente entre as taxas de reconhecimento processadas com 16 centroides e as demais quantidades de centroides. Observa-se que imagens processadas a partir de 32 centroides não apresentam taxas de reconhecimento significativa.

A média de reconhecimento obtida entre as imagens por classe e todos os treinos, da técnica da técnica de QV - 32 centroides é de 95,94% e para QV - 64 centroides é de 98,96%. No entanto, quando trabalhado com 128 centroides e 256 centroides obteve-se média de 100% de reconhecimento.

Comparação com outros trabalhos

Tabela III – Resultados encontrados na literatura

Trabalhos	Base de dados	Taxa de reconhecimento
Li e	SBID	94,6%
Li e	SBID	85,25%

Verifica-se que o método proposto nesse trabalho apresentou taxas de reconhecimento melhores, a partir de 32 centroides, do que as apresentadas por Li, et al. (2014). Deve-se levar em consideração, também, que o presente trabalho não pretende

fazer a análise de reconhecimento para todo o conjunto SBID como propõem os trabalhos apresentados na Tabela III.

CONCLUSÕES

Destaca-se que a técnica de Quantização Vetorial utilizando o algoritmo LBG tem um excelente desempenho para reconhecimento de edifícios. Tendo em vista que se obteve taxas de reconhecimento de 100%.

Foi encontrado diferença significativa, em relação somente quando utilizado 16 centroides.

Quando mantido a quantidade de centroides e variado os treinos os resultados ao obtiveram diferenças significativas.

Percentualmente encontrou-se taxas de reconhecimento de 100% quando trabalhado com 128 e 256 centroides. No entanto, a partir de 32 centroides, ao comparar os resultados obtidos pela análise de todos os valores encontrados, não se teve diferença significativa quando analisado as taxas utilizando o teste paramétrico da ANOVA, com post-hoc de Tukey, com o nível de significância de $p \leq 0,05$.

REFERÊNCIAS

- ALBUQUERQUE, M. P. **Processamento de imagens: Métodos e análise**, 2000. Acedido em 03 de setembro de 2021, em <http://www.cbpf.br/~mpa/curso2.htm>.
- CHANG, C.; LIN, C.; HSIEH, Y. **Data Hiding dor Vector Quantization images using mixed-base notation and dissimilar patterns without loss of fidelity, information sciences**. Information Sciences, 201, 70-79, 2012.
- GERSHO, A.; GRAY, R. M. **Vector Quantization and signal compression**. Kluwer Academic Publishers, Boston, MA, 1992.
- GONZALES, R.; WOODS, R. **Processamento de Imagens Digitais**, 3ª ed., São Paulo: Edgard Blucher Ltda., 2000.
- LI, J.; ALLINSON, N. M. **Building Recognition Using Local Oriented Features**, IEEE Transactions on Industrial Informatics, vol. 9, no. 3, pp. 1697-1704, Aug. 2013
- LI, J.; ALLINSON, N. M. **Subspace learning-based dimensionality reduction in building recognition**, Neuro computing, vol. 73, no. 1-3, pp. 324-330, 2009.
- LI, J.; HUANG, W.; SHAO, L.; ALLISON, N. **Building recognition in urban environments: A survey of state-of-the-art and future challenges**. Information Sciences, 277, 406-420, 2014
- LINDE, Y.; BUZO, A.; GRAY, R. M. **An algorithm for Vector Quantizer desing**. IEEE Transactions on Communications, New York, v. 28, n. 1, 84-95, 1980.
- MADEIRO, F.; GALVÃO, R.; FERREIRA, F.; CUNHA, D. **Uma alternativa de aceleração do algoritmo fuzzy k-means aplicado à Quantização Vetorial**, TEMA, 13, n. 2, 193-206, 2012.
- MADEIRO, F.; LOPES, W. T. A. **Introdução à Compressão de Sinais**. Revista de Tecnologia da Informação e Comunicação, vol. 1, nº. 1, pp. 33-40, outubro 2011.
- RODRIGUES, C.; REDWEIK, P. **Extração de fachadas de edifícios em imagens obtidas com UVA**. Dissertação de mestrado em Mestrado em Engenharia Geográfica: Faculdade de Ciências - Departamento de Engenharia Geográfica, Geofísica e Energia, Lisboa, Portugal, 54p., 2016.
- SCOLARI, D. **Estudo e desenvolvimento de algoritmos para o reconhecimento de arritmias cardíacas**. Dissertação de mestrado em Engenharia Elétrica: PUCRS, Porto Alegre, 96p., 2007.
- SILVA, L. A.; FLÔRES, E. L.; CARNEIRO, M. B. P.; NOMURA, S.; SANTOS, C. F. **Utilização de filtros orientados como método de seleção de características de imagens usadas em reconhecimento de edifícios**. XII Conferência de Estudos em Engenharia Elétrica, 2014.
- SRINIVASAN, A. **Speaker identification and verification using Vector Quantization and mel frequency cepstral coefficients**. Engineering and Technology, 4:1, 33-40, 2012.

VASCONCELOS, E. S.; MELO, D. V.; CARRIJO, G. A. **Reconhecimento de edifícios em ambientes urbanos com a utilização de quantização vetorial.** Brazilian Journal of Development, Curitiba, v.8, n.4, p.23332-23342, apr.,2022.

VIEIRA, R.; BRUNET, N.; COSTA, S.; CORREIA, S.; NETO, B.; FECHINE, J. **Combining entropy measurements and cepstral analysis for pathological voice assessment.** Journal of Medical and Biological Engineering, 2012. Acedida em 08 de setembro de 2021 em https://www.researchgate.net/publication/234059530_Combining_entropy_measurements_and_cepstral_analysis_for_pathological_voice_assessment.

WIGGERS, K. L.; VELOSO, L. H. L.; CALDAS, E. A.; GUIMARÃES, A. M. **Análise de abordagens para extração de características de sementes usando processamento digital de imagens.** In: IX Congresso Brasileiro de Agro Informática, 2013.

Recebido em: 01/09/2022

Aprovado em: 30/09/2022

Publicado em: 05/10/2022