



RECONHECIMENTO DE EDIFÍCIOS EM AMBIENTES URBANOS COM A UTILIZAÇÃO DE QUANTIZAÇÃO VETORIAL

Eduardo Silva Vasconcelos*¹, Débora Vasconcelos Melo², Gilberto Carrijo¹

¹FEELT – Universidade Federal de Uberlândia

²FENG – Universidade Federal de Catalão

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 10 imagens por classe, na escala de cores cinza, vermelha, verde e azul, 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, treinos e cores. Os resultados alcançados obtiveram uma alta taxa de acerto, de 98,75%, na cor cinza, para a imagem quantizada com 64 e 256 centroides, no treino de 60%.

Palavras-Chave – Algoritmo Linde-Buzo-Gray, Compressão de Imagens, Processamento de Imagens, Quantização Vetorial, Reconhecimento de Edifícios.

RECOGNITION OF BUILDINGS IN URBAN ENVIRONMENTS USING VECTOR QUANTIZATION

Abstract – Building recognition is essential for a variety of applications. Due to its importance, this work aims to analyze the image recognition rates of buildings, for 40 classes and 10 images per class, in gray, red, green and blue color scale, using a recognition system based on the technique of Vector Quantization for image compression using the LBG algorithm, varying the number of centroids, trainings and colors. The results achieved had a high accuracy rate of 98.75%, in gray, for the quantized image with 64 and 256 centroids, in 60% training.

Keywords – Building Recognition, Image Compression, Image Processing, Linde-Buzo-Gray Algorithm, Vector Quantization.

I. INTRODUÇÃO

Técnicas de processamento de imagens são aplicadas a praticamente todas as ciências, segundo Albuquerque [1] 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 [2] 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 agrupamento; 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 [3].

A QV traz como objetivo representar distribuições de dados utilizando um número de referência de padrões significativamente menor que o número de dados [4]. Dessa forma, a compressão de um conjunto de dados busca diminuir a quantidade de informação, tanto para transmissão quanto para armazenamento [5], pois apenas os centroides precisam ser armazenados, ao invés da base de dados inteira.

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. 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 a 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 e chamado de partição do espaço.

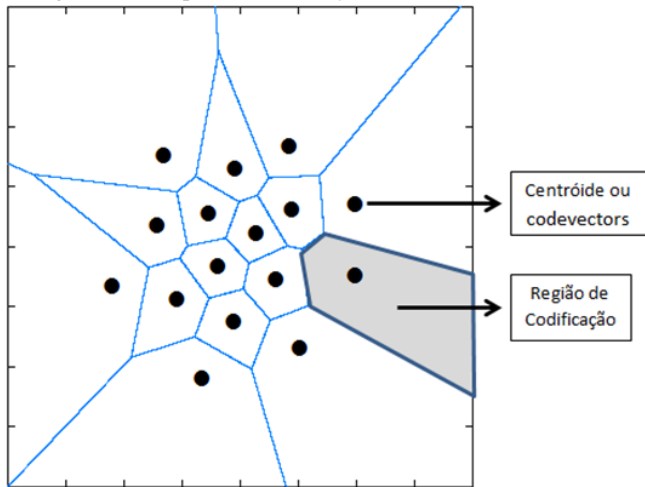
Ressalta-se que, além da compressão de imagens, há um amplo espectro de aplicações para a QV, como, por exemplo, esteganografia [7], marca d'água digital [3], identificação vocal [8] e classificação de sinais de voz com patologias [9].

Contudo na QV a complexidade computacional existente na fase de codificação dos vetores a serem quantizados,

*educelos1@gmail.com

constitui alguns desafios, tais como o projeto de criação de codebooks e a sensibilidade da técnica aos erros de transmissão [3]. Sendo assim, alguns algoritmos podem ser empregados.

Figura 1: Exemplo de Quantização Vetorial em duas dimensões.



Fonte: Adaptado de Madeiro e Lopes [6].

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 centróides, que servem como elemento representativo dos agrupamentos obtidos, criando um novo *codebook* que é iterativamente atualizado [10]. A quantidade de centróides (C) é escolhida com base na Equação 1.

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

Onde:

m é um valor de entrada pré-estabelecido [11].

Essas técnicas são essenciais para muitos serviços de localização, aplicações geoespaciais, exército, engenharia e arquitetura [12], como, a área imobiliária, planejamento urbano, turismo virtual de uma cidade, navegação de carro [13], 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.

II. 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 [14]. Esse reconhecimento consiste em três partes: (i) representação das características, (ii) correspondência das características, (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 [14].

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.

A. O banco de dados e sua complexidade: SBID

Conforme Li et al. [14], 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.

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.

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.

B. O método proposto

O método proposto compreende o processamento das imagens nas escalas de cores vermelha, verde, azul e cinza, separadamente. 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 [10].

Os resultados foram analisados a partir das variáveis: cor, centróides e porcentual de treino. Foi fixado para esse trabalho a análise de 10 imagens em 40 classes. Estipulou-se as primeiras 10 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 400 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, 160 imagens foi para o treino e 240 para o reconhecimento, para 50% de treino, 200 imagens foram para o treino e outras 200 para o reconhecimento e para 60% de treino, 240 imagens foram para o treino e 160 para o reconhecimento. Todo o procedimento foi analisado para 16, 32, 64, 128 e 256 centroides.

As análises considerando as cores, 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.

Por fim, tem-se a execução do algoritmo iniciando o processo de reconhecimento e classificação da imagem a partir dos parâmetros selecionados. O resultado final consiste na obtenção do percentual de acertos do reconhecimento (Equação 2).

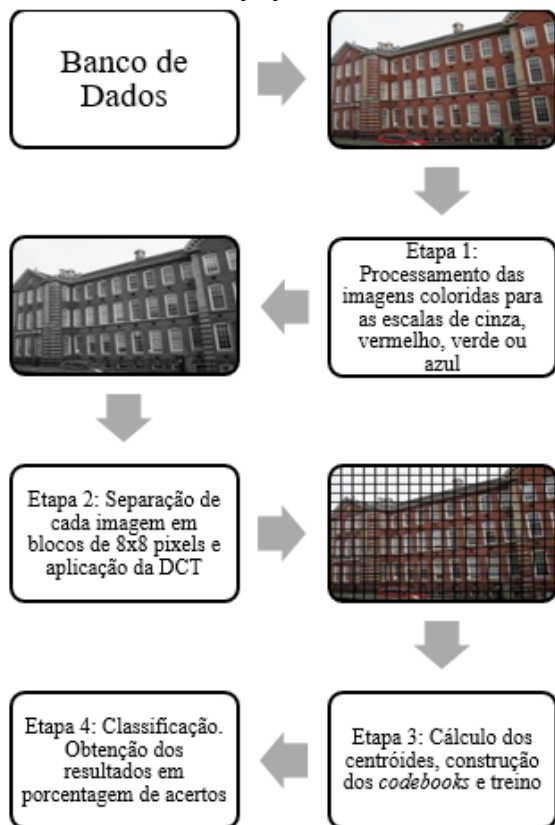
$$Porcentagem = \frac{Q_a}{Q_c \cdot (Q_{ic} - Q_{it})} \cdot 100 \quad (2)$$

Onde:

- Q_a - Quantidade de acerto
- Q_c - Quantidade de classes
- Q_{ic} - Quantidade de imagens em cada classe
- Q_{it} - Quantidade de imagens para treino

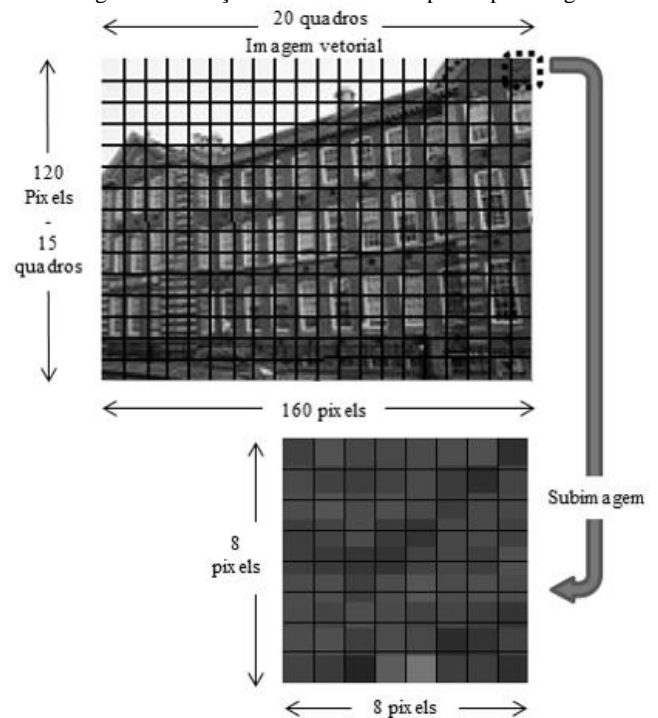
As etapas propostas no método empregado são apresentadas na Figura 2.

Figura 2: Diagrama em blocos das etapas da metodologia proposta.



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 e converte os dados de amplitude espacial em coeficientes de frequência espacial.

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



Com isso, o algoritmo realiza os treinamentos nesse conjunto de imagem e gera 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.

$$x_k = \{x_m; m = 1, 2, \dots, 64\}; k = 1, 2, \dots, 300 \quad (3)$$

Onde:

x_m - Ordenada da escala da cor utilizada

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.

III. RESULTADOS E DISCUSSÃO

Os dados apresentados foram obtidos pelo método de QV juntamente com o algoritmo LBG, pelo Matlab, processadas nas cores cinza, vermelha, verde e azul. Os resultados foram analisados e são apresentados conforme as cores, centroe (16, 32, 64, 128 e 256) e treino (40%, 50% e 60%), como observado na Tabela I.

Tabela I: Taxa percentual de reconhecimento do banco de dados analisado para 10 imagens em 40 classes.

Cor	Centroide	Treino		
		40%	50%	60%
Cinza	16	76,67%	89,00%	96,87%
	32	80,00%	93,00%	96,87%
	64	85,83%	95,50%	98,75%
	128	87,50%	95,00%	98,12%
	256	87,92%	95,50%	98,75%
Vermelha	16	63,75%	78,50%	86,25%
	32	76,25%	89,00%	95,63%
	64	76,67%	90,00%	95,00%
	128	79,17%	91,50%	96,88%
	256	80,83%	92,00%	96,88%
Verde	16	78,75%	88,50%	94,38%
	32	83,33%	94,00%	96,25%
	64	83,75%	95,50%	97,50%
	128	85,00%	96,00%	96,88%
	256	87,08%	95,00%	98,13%
Azul	16	79,58%	90,00%	94,38%
	32	86,25%	91,00%	93,75%
	64	87,92%	93,00%	93,75%
	128	88,75%	94,00%	95,63%
	256	88,75%	96,50%	96,25%

No geral, observa-se que para todas as cores analisadas o melhor percentual descritivo encontra-se com 256 centroides, para o treino de 60%, exceto para a cor azul, a qual apresentou-se maior taxa de reconhecimento no treino de 50%. Destaca-se também que na cor cinza o percentual de reconhecimento quando utilizado 64 centroides apresentou o mesmo que quando utiliza 256 centroides, para o treino de 60%, esse resultado foi a maior taxa descritiva de reconhecimento encontrada em toda a análise.

Nota-se também que para a cor vermelha o percentual de reconhecimento quando utilizado 128 centroides apresentou o mesmo quando utilizado 256 centroides. Para a estatística inferencial, comparou-se, separadamente os grupos cor e centroides com os treinos, a fim de verificar se existe diferença significativa entre eles. Para o centroide foi observado que há diferença significativa para os treinos de 40% e 50% (Tabela II).

Tabela II: Análise entre as taxas de reconhecimento dos centroides com os treinos.

Treino	Centroide				
	16	32	64	128	256
40%	74,69a	81,46ab	83,54ab	85,1ab	86,14b
50%	86,5a	91,75ab	93,5b	94,12b	94,75b
60%	92,97a	95,62a	96,25a	96,88a	97,5a

Letras diferentes (“a” e “b”) na linha da tabela representa diferença estatística significativa, pela Anova com *post-hoc* de Tukey, com p-valor $\leq 0,05$.

Observou-se que no treino de 40% existe uma diferença significativa ao utilizar a taxa média de reconhecimento entre 16 e 256 centroides, esse resultado implica que as imagens processadas no treino de 40% têm uma maior taxa de reconhecimento quando utilizado 256 centroides. Para o treino de 50%, existe diferença significativa entre 16 centroides com 64, 128 e 256 centroides.

Para o treino de 60%, não foi encontrado nenhuma diferença significativa entre os centroides, implicando que se

tem uma boa taxa de reconhecimento utilizando qualquer quantidade de centroide analisado. Entre as taxas de reconhecimento em relação as cores, encontrou-se diferença significativa apenas no treino de 40% (Tabela III).

Tabela III: Análise entre as taxas de reconhecimento das cores com os treinos.

Treino	Cores			
	Cinza	Vermelha	Verde	Azul
40%	83,58ab	75,33a	83,58ab	86,25b
50%	93,6a	88,2a	93,8a	92,9a
60%	97,87a	94,13a	96,63a	94,75a

Letras diferentes (“a” e “b”) na linha da tabela representa diferença estatística, pela Anova com *post-hoc* de Tukey, com p-valor $\leq 0,05$.

Não foi encontrado diferença significativa entre as cores nos treinos de 50% e 60%. Para o treino de 40%, observa-se que existe uma diferença significativa na taxa média de reconhecimento, entre as cores vermelha e azul, podendo inferir que as imagens quando processadas na cor azul tem uma maior taxa de reconhecimento, nesse treino.

IV. 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 uma taxa de reconhecimento de 98,75%, na cor cinza, para a imagem quantizada com 64 e 256 centroides, no treino de 60%, para quando trabalhado com 10 imagens em 40 classes. Foi encontrado diferença significativa, em relação aos centroides, apenas no treino de 40% e 50%, onde as taxas de reconhecimento são mais significativas quando utilizado 256 centroides.

Em relação a cor, foi encontrado uma diferença estatística significativa apenas no treino de 40%, na cor azul em relação a vermelha. As maiores taxas de reconhecimento foram encontradas no treino de 60%, independente dos centroides e das cores.

REFERÊNCIAS

- [1] M. P. Albuquerque. “Processamento de imagens: Métodos e análise”, 2000. Acedido em 03 de setembro de 2021, em <http://www.cbpf.br/~mpa/curso2.htm>.
- [2] R. Gonzales, R. Woods. “Processamento de Imagens Digitais”, 3 ed., São Paulo: Edgard Blucher Ltda., 2000.
- [3] F. Madeiro, R. Galvão, F. Ferreira, D. Cunha. “Uma alternativa de aceleração do algoritmo *fuzzy k-means* aplicado à Quantização Vetorial”, *TEMA*, 13, n. 2, 193-206, 2012.
- [4] K. L. Wiggers, L. H. L. Veloso, E. A. Caldas, A. M. Guimarães. “Análise de abordagens para extração de características de sementes usando processamento digital de imagens”. In: IX Congresso Brasileiro de Agro Informática, 2013.
- [5] A. Gersho, R. M. Gray. “*Vector Quantization and signal compression*”. Kluwer Academic Publishers, Boston, MA, 1992.

- [6] 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.
- [7] C. Chang, C. Lin, Y. Hsieh. “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.
- [8] A. Srinivasan. “Speaker identification and verification using Vector Quantization and mel frequency cepstral coefficients”, *Engineering and Technology*, 4:1, 33-40, 2012.
- [9] R. Vieira, N. Brunet, S. Costa, S. Correia, B. Neto, J. Fechine. “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.
- [10] Y. Linde, A. Buzo, R. M. Gray. “An algorithm for Vector Quantizer design”, *IEEE Transactions on Communications*, New York, v. 28, n. 1, 84-95, 1980.
- [11] D. Scolari. “Estudo e desenvolvimento de algoritmos para o reconhecimento de arritmias cardíacas”, dissertação de mestrado em Engenharia Elétrica: PUCRS, Porto Alegre, 96p., 2007.
- [12] L. A. Silva, E. L. Flôres, M. B. P. Carneiro, S. Nomura, C. F. Santos. “Utilização de filtros orientados como método de seleção de características de imagens usadas em reconhecimento de edifícios”, in: XII CEEL, 2014.
- [13] C. Rodrigues, P. Redweik. “Extração de fachadas de edifícios em imagens obtidas com UVA”, in: VIII CNCG, 2015.
- [14] J. Li, W. Huang, L. Shao, N. Allison. “Building recognition in urban environments: A survey of state-of-the-art and future challenges”, *Information Sciences*, 277, 406-420, 2014.