



Combinação de Superpixel e Redes Convolucionais para Segmentação de Bovinos

Programa de Pós-Graduação Stricto Sensu - Mestrado em Computação

INOVISAO

Universidade Federal de Mato Grosso do Sul

Março de 2017
Campo Grande - MS

Resumo

Dairy cattle and dairy farming are important segments of the Brazilian economy, especially in the composition of GDP. Despite the high numbers, the production system still presents below-average productive indices compared to other countries. In order to obtain these satisfactory levels of productivity in beef cattle and milk, the objective of this project is the development of a method to segment livestock images. With the automation of processes, the producer can be assisted in real-time decision-making of various management strategies throughout the production system, including growth and nutritional status, establishing the value of the animal's sales and adjust the feed. This method will be based on a combination of superpixel and convolutional neural networks, which are two techniques with good results in segmentation of images and also well known in literature. Thus, the proposed project aims to increase productivity through real-time decision making, increasing the productive potential and income of the state of Mato Grosso do Sul.

1 Antecedentes e Justificativa

A pecuária é um dos segmentos que mais se destaca na economia brasileira. Atualmente, o Brasil é o líder mundial nas exportações de carne bovina e detém o segundo maior rebanho comercial do mundo, atrás apenas dos Estados Unidos. As pecuárias de corte e de leite respondem, respectivamente, por cerca de 11% e 2,8% do Produto Interno Bruto - PIB do agronegócio nacional. Em 2015, o Brasil produziu aproximadamente 10 milhões de toneladas de carne bovina e exportou mais de 1,4 milhão de toneladas para diversos países [Kist, 2016a]. No mesmo período, a pecuária de leite produziu 34.823 milhões de litros de leite, sendo o quarto maior produtor de leite mundial [Kist, 2016b].

Apesar dos números elevados, o sistema nacional ainda apresenta índices produtivos pouco representativos devido à baixa produtividade média por área comparada a outros países. Em 10 anos, o objetivo do Brasil é dobrar a lotação animal por área, passando de 1,3 para 2,6 cabeças por hectare, e saltar de 1,4 mil litros de leite por vaca ao ano para 2 mil litros [Beling, 2014]. Para alcançar esse objetivo, é necessário adotar tecnologias através da pecuária de precisão que permitam incrementos em produtividade e maior rentabilidade ao produtor. A pecuária de precisão consiste na medição de diferentes parâmetros dos animais e do ambiente, a modelagem desses dados e o seu uso para tomada de decisões. Exemplos de medidas que auxiliam na tomada de decisão são a condição corporal e o peso dos animais do rebanho. Em propriedades que utilizam estas medidas, o peso é estimado por meio de avaliação visual e/ou tátil do animal, tornando-se um processo que consome muito tempo, requer treinamento, além da avaliação ser subjetiva e influenciada pelo avaliador ou por animais previamente avaliados [Halachmi et al., 2008].

Para evitar o contato direto com o animal e conseqüentemente diminuir o seu estresse, métodos computacionais com base em imagens estão sendo cada vez mais estudados e propostos na literatura [Frost et al., 1997]. Estes métodos utilizam a visão computacional cujo o objetivo é construir sistemas artificiais para interpretar imagens. Em geral, um sistema de

visão computacional é dividido em cinco etapas: aquisição de imagens, pré-processamento, segmentação, extração de características e reconhecimento de padrões. Em particular, a segmentação de imagens é o processo de dividir uma imagem em regiões com propriedades similares para facilitar a sua análise. Geralmente, a etapa de segmentação afeta diretamente os resultados das etapas posteriores, como a extração de características e classificação. Portanto, a segmentação é considerada uma das etapas mais importantes e complexas de um sistema de visão computacional.

A partir desses direcionamentos, o objetivo deste projeto é o desenvolvimento de um método para segmentar imagens de gados utilizando uma combinação entre superpixels e rede neural convolucionais. Em particular, a segmentação de imagens de gado consiste em dividir a imagem em dois tipos de região: as regiões que contêm gados e as regiões correspondentes ao fundo. Em ambientes não-controlados, como é o caso desse projeto, existem diversos desafios que devem ser resolvidos, tais como diferentes iluminações, escalas e raças de bovinos (e.g., gados com textura malhada, preta e branca). Como a tarefa é considerada complexa, a escolha de técnicas de segmentação se torna muito importante. Neste projeto pretende-se propor uma combinação entre redes neurais convolucionais [Krizhevsky et al., 2012] e superpixel [Achanta et al., 2012], pois ambas as técnicas possuem resultados significativos em segmentação de imagens. Alguns experimentos preliminares foram realizados utilizando o software Pynovisão. Este software utiliza o algoritmo SLIC para dividir a imagem em superpixels e, para cada superpixel, um vetor de características é extraído. Os superpixels foram divididos em duas classes, como mencionado anteriormente: gado e fundo. Nos experimentos foram utilizadas características baseadas em cor e textura, e pode-se observar que as características de textura não apresentam bons resultados para a classe gado, isto porque os superpixels desta classe não apresentam texturas relevantes. Embora a inclusão da característica de cor aumente a medida-f, ainda assim os resultados não foram satisfatórios para a classe gado.

Por meio deste projeto, ambas as áreas, visão computacional e agropecuária, serão beneficiadas. Por um lado, novos desafios são apresentados aos métodos de visão computacional,

tornando possível o aperfeiçoamento e a criação de novos métodos para suprir as necessidades e aumentar a taxa de acerto. Por outro lado, os algoritmos e métodos desenvolvidos para a segmentação de bovinos podem contribuir para novas tecnologias na agropecuária de precisão, aumentando a produtividade e reduzindo os custos da produção.

2 Objetivos

2.1 Objetivo Geral

O objetivo geral deste projeto é propor um método de segmentação de imagens de bovinos por meio da combinação de redes neurais convolucionais e superpixel.

2.2 Objetivos Específicos

1. Criação de um banco de imagens de gados em ambientes não-controlados;
2. Estudar formas de combinações entre rede neural convolucional e superpixel;
3. Desenvolvimento e implementação de um método para segmentar imagens de gados;
4. Validação do método proposto;
5. Divulgação dos resultados obtidos por meio de artigos científicos.

3 Revisão de Literatura

Esta seção apresenta os trabalhos de visão computacional na pecuária de precisão e trabalhos sobre as duas técnicas abordadas neste projeto, redes neurais convolucionais e superpixel.

3.1 Visão computacional na pecuária de precisão

Com os avanços recentes da computação, sistemas assistidos por computador tem tornado possível a automatização de processos em várias atividades, tais como na pecuária de precisão. Vários trabalhos de visão computacional tem sido propostos, como o trabalho proposto por Hertem et. al. [Hertem et al., 2013]. Os autores adaptaram cinco algoritmos de segmentação baseados em subtração de fundo, e testaram em fundos dinâmicos (campo aberto) e estáticos (foi introduzida uma parede atrás dos gados). Eles perceberam que o melhor algoritmo foi baseado em detecção de bordas por subtração de fundo com erro médio absoluto de 6.7 ± 5.7 pixels, e que nenhum algoritmo produziu resultados satisfatórios em fundos dinâmicos.

Viazzi et. al. [Viazzi et al., 2014] compararam sistemas de câmeras 2-D e 3-D para medir a postura traseira em vacas leiteiras. Eles perceberam que sistemas com câmeras 2-D, com vista lateral do gado, não possuem bons resultados principalmente pela diferença de iluminação, podendo ocorrer sombras que dificultam uma boa segmentação. Para contornar este problema, a câmera foi colocada em cima do gado, conseguindo melhorar a precisão na segmentação e obtendo uma acurácia de 91% utilizando câmeras 2-D e 90% utilizando câmeras 3-D.

Alguns trabalhos também utilizam métodos de segmentação para estimar o peso e o escore do gado. Stajanko et. al. [Vindis et al., 2010] propuseram uma segmentação de gado através da imagens térmica para estimar o peso do gado. Uma dificuldade deste método é que o gado precisa estar com uma temperatura mais elevada do que o ambiente a sua volta, então os autores fizeram a captura das imagens com o gado rodeado por paredes resfriadas.

3.2 Segmentação com superpixel

Superpixel é uma abordagem que divide uma imagem em regiões com propriedades similares, como cor, textura e brilho. Estas regiões similares são chamadas de superpixels. Os algoritmos de superpixels podem ser divididos em baseados em grafos e baseados em gradiente.

Em algoritmos baseados em grafos, cada pixel é mapeado em um nó do grafo, e o peso

da conexão entre dois nós é dado pela similaridade entre eles. O algoritmo *Normalized Cuts* [Shi e Malik, 2000] divide a imagem em grafos recursivamente através de similaridades de contornos e texturas. A principal desvantagem desse algoritmo é o seu elevado custo computacional [Levinshtein et al., 2009]. Fezenszwalb e Huttenlocher [Felzenszwalb e Huttenlocher, 2004] também propuseram um algoritmo para obter superpixels baseados em grafos. Este algoritmo agrupa os nós de forma aglomerativa de tal forma que cada superpixel é uma árvore geradora mínima dos pixels.

Por outro lado, os algoritmos baseados em gradiente realizam um agrupamento inicial, e a cada iteração um refinamento baseado em gradiente ocorre para obter uma melhor segmentação. O algoritmo Mean-shift [Comaniciu e Meer, 2002] é um algoritmo baseado em busca onde o centroide de cada agrupamento se move de acordo com o gradiente local. O algoritmo não impõe restrição com relação ao número, tamanho ou compactação dos superpixels. Outro algoritmo baseado em gradiente é o Quick-shift [Vedaldi e Soatto, 2008], que move cada ponto no espaço de característica para o vizinho mais próximo que aumente a densidade de Parzen. Este algoritmo também não restringe o tamanho e a quantidade de superpixels.

Para este trabalho, o algoritmo abordado é o SLIC (*simple linear iterative clustering*) que foi proposto por Achanta et. al. [Achanta et al., 2012]. O SLIC gera agrupamentos com base na similaridade de cor e proximidade no plano da imagem. O algoritmo utiliza o espaço 5-D definido por L, A, B, X e Y, onde L, A e B corresponde ao vetor de cor na espaço CIELAB e os valores X e Y correspondem a posição do pixel no plano da imagem. A escolha deste algoritmo deve-se ao seu desempenho computacional e acurácia [Achanta et al., 2012]. Entretanto, outros algoritmos baseados em superpixel podem ser utilizados na combinação com redes neurais convolucionais, dependendo do desempenho do SLIC.

3.3 Redes Neurais Convolucionais

As redes neurais convolucionais consistem em camadas convolucionais, que utilizam o algoritmo de *backpropagation* para aprende os parâmetros/filtros de cada camada. Desde a sua

criação, a CNN tem sido caracterizada por três propriedades básicas, as conexões locais, o compartilhamento de peso e o *pooling* local. As duas primeiras propriedades permitem que o modelo aprenda os padrões visuais locais importantes com menos parâmetros ajustáveis que um modelo totalmente conectado, e a terceira propriedade prepara a rede para possuir invariância à translação [LeCun et al., 1989].

Uma das primeiras propostas de redes neurais convolucionais é a rede LeNet-5 descrita por LeCun et al. [LeCun et al., 1989] para o reconhecimento óptico de caracteres. Comparada às redes convolucionais profundas atuais, a rede foi relativamente modesta devido aos recursos computacionais limitados da época e aos desafios do treinamento de algoritmos para redes com mais camadas. Embora houvesse muito potencial em redes convolucionais mais profundas, só recentemente elas se tornaram predominantes, devido o aumento do poder computacional atual, da quantidade de dados para treinamento disponíveis na internet e do desenvolvimento de métodos mais eficazes para a formação de tais modelos.

Um exemplo recente e notável do uso de redes convolucionais profundas para classificação de imagens é o desafio Imagenet [Krizhevsky et al., 2012] em que uma CNN obteve um erro consideravelmente menor comparado com o erro das abordagens tradicionais de visão computacional (usando SIFT e Máquinas de Vetores de Suporte). Redes neurais convolucionais também obtiveram recentemente sucesso para diferentes aplicações, incluindo estimação de pose humana [Toshev e Szegedy, 2014], análise de faces [Luo et al., 2012], detecção de pontos-chave facial [Sun et al., 2013], reconhecimento de voz [Graves et al., 2013] e classificação de ação [Karpathy et al., 2014].

Com os resultados promissores em classificação, as redes neurais convolucionais foram recentemente estendidas para o problema de segmentação [Shelhamer et al., 2017]. Nessas extensões, as redes neurais convolucionais possuem apenas camadas de convolução, sendo portanto chamadas de Redes Totalmente Convolucionais (RTC). A saída da RTC é uma imagem do mesmo tamanho da entrada sendo que cada pixel da saída corresponde à probabilidade do pixel pertencer a uma determinada classe. Dessa forma, a imagem de saída é utilizada para segmentação

de imagens.

4 Metodologia

Para alcançar os objetivos propostos, este projeto será desenvolvido obedecendo as seguintes fases:

Meta 1: Criação de um banco de imagens de gados em ambientes não-controlados

Inicialmente acontecerá visitas em fazendas para captura imagens de gados de leite e corte. Pretende-se que estas imagens sejam tanto em ambientes controlados quanto em ambientes não-controlados. Após capturar as imagens, pode-se criar um banco de imagens. Também serão utilizadas imagens de gados das competições Microsoft COCO [Lin et al., 2014] e da Pascal VOC [Everingham et al., 2015]. A grande vantagem é que estas competições disponibilizam as imagens de referência (*ground truth*), como podem ser visualizadas na Figura 1. O banco de imagens será composto com diversas imagens de gados em diferentes ambientes: no curral, solto em um cercado, além de imagens aéreas de rebanhos bovinos. Para tentar melhorar o desempenho do algoritmo, serão realizados pré-processamentos nas imagens.

Meta 2: Aprofundamento da revisão bibliográfica

Nesta etapa a revisão bibliográfica será aprofundada com aplicações da pecuária de precisão em visão computacional com foco em projetos que utilizaram segmentação de imagens de bovinos. Através de consultas aos principais portais de periódicos, como IEEE Xplore, ACM DL, Science Direct e Scopus, serão identificados os trabalhos correlatos relacionados as abordagens usadas neste trabalho: rede neural convolucional e superpixel. Nesta etapa do projeto também serão estudadas as formas de combinações entre estas abordagens propostas na literatura.

Meta 3: Desenvolvimento e implementação de um método para segmentar ima-

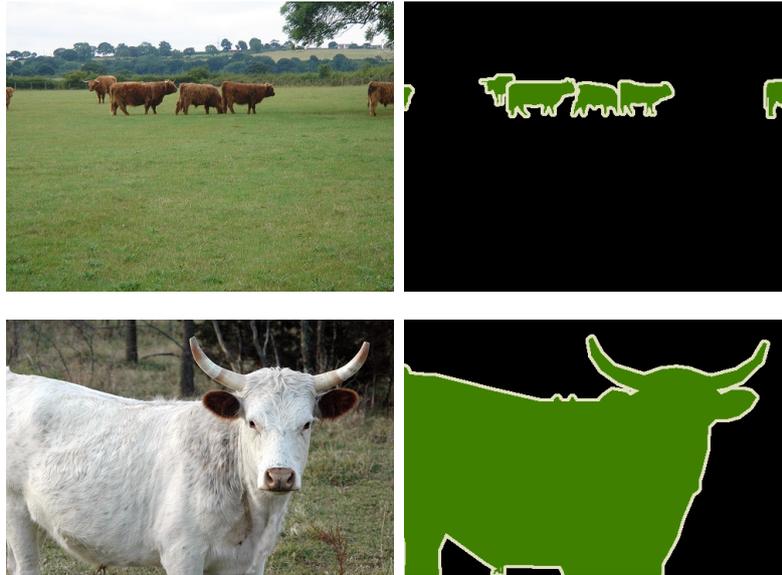


Figura 1: Exemplo de imagens originais e de referência da competição Pascal VOC.

gens de gados e validação do método proposto

Após o estudo de métodos da literatura, alguns destes métodos serão escolhidos para implementação afim de obter um estudo mais aprofundado. Depois será possível o desenvolvimento e implementação de um método que combine superpixel e redes neurais convolucionais. Após a implementação, o método será adicionado ao *software* Pynovisão. O Pynovisão é um *software* desenvolvido pelo grupo INOVISÃO e tem como base o uso do algoritmo superpixel para segmentar imagens. Também será nesta parte do projeto que o método proposto será validado para diferentes tipos de ambientes. Como métricas serão utilizadas a medida-f, porcentagem de classificação correta (PCC), acurácia pixel a pixel e intersecção sobre união entre as regiões preditas e regiões corretas. Para amostragem será adotada a validação cruzada de 10 dobras com 10 repetições (*10-folders cross-validation*). Para comparar o método proposto com métodos da literatura, será utilizada a análise de variância (ANOVA), que identificará se existe diferença estatísticas entre os métodos. Por fim, um pós-teste será realizado para comparar estatisticamente os métodos.

Meta 4: Divulgação dos resultados obtidos por meio de artigos científicos Nesta

etapa serão produzidos os indicadores de desempenho do algoritmo e os resultados obtidos serão analisados. Além disso, esses resultados serão divulgados por meio de artigos científicos e palestras. Revistas, eventos nacionais e internacionais apropriados serão selecionados para submissão dos resultados.

5 Atividades e Cronograma de Execução

As metas descritas na metodologia serão executadas seguindo o cronograma e as atividades da Tabela 1.

Metas	Atividades	Tempo (trimestral)							
		Ano 1				Ano 2			
		1°	2°	3°	4°	1°	2°	3°	4°
Meta 1	1.1 - Definir os critérios/ambientes para a captura de imagens	•							
	1.2 - Capturar as imagens em diferentes ambientes e condições	•	•						
	1.3 - Criação do banco de imagens	•	•						
	1.4 - Pré-processamento das imagens		•						
Meta 2	2.1 - Revisão bibliográfica sobre visão computacional e pecuária de precisão		•	•					
	2.2 - Revisão bibliográfica sobre redes neurais convolucionais		•	•					
	2.3 - Revisão bibliográfica sobre superpixel		•	•					
Meta 3	3.1 - Selecionar métodos estudados na revisão para serem implementados			•	•				
	3.2 - Propor/ajustar métodos para combinar redes neurais convolucionais e superpixel			•	•	•	•		
	3.4 - Testar e avaliar os métodos implementados em diferentes ambientes				•	•	•		
Meta 4	4.1 - Produção dos indicadores de desempenho (porcentagem de classificação e outras medidas)					•	•	•	•
	4.2 - Avaliação e análise estatísticas dos resultados obtidos							•	•
	4.3 - Divulgação dos resultados por meio de artigos científicos e palestras						•	•	•

Tabela 1: Cronograma de execução das metas e atividades.

6 Resultados Preliminares - Software Pynovisão

Alguns experimentos preliminares foram realizados utilizando o software Pynovisão. Este software utiliza o algoritmo SLIC para dividir a imagem em superpixels e, para cada superpixel, um vetor de características é extraído. O software possui três parâmetros para o SLIC, a quantidade de núcleos (k), o σ (s) e a compactação (c). Para os experimentos realizados, os parâmetros utilizados foram $k=500$, $s=5$ e $c=10$. Nos experimentos preliminares foram utilizadas as características baseadas em cor e textura: padrões locais binários (Local Binary Patterns - LBP) [Ojala et al., 2002], matrizes de co-ocorrência em níveis de cinza (Gray-Level Cooccurrence Matrix - GLCM) [Haralick, 1979], histograma de gradientes orientados (Histogram of Oriented Gradients - HOG) [Dalal e Triggs, 2005]. Por fim, as características extraídas são utilizadas para treinar um classificador Máquina de Vetores de Suporte (*Support Vector Machine*) que aprende a classificar um superpixel como pertencente ao gado ou ao fundo. A Figura 2 ilustra as etapas usadas no experimento preliminar.

Nos experimentos foram utilizadas 20 imagens em que os superpixels foram anotados manualmente. Um total de 7.422 superpixels de fundo e 741 superpixels de gado foram obtidos. A Tabela 2 apresenta a medida-f para as classes de fundo e gado usando diferentes extratores de características. Podemos observar que as características de textura não apresentam bons resultados para a classe gado, isto porque os superpixels desta classe não apresentam texturas relevantes. Embora a inclusão da característica de cor aumente a medida-f, ainda assim os resultados não foram satisfatórios para a classe gado. É importante ressaltar que neste experimento preliminar somente uma raça de gado foi utilizada (nelore branco). Com a inclusão de outras raças (cores e texturas diferentes), o desempenho das características de cor devem ser reduzido. A Figura 3 apresenta dois exemplos de segmentação em que os superpixels em vermelho foram classificados como gado e os em azul foram classificados como fundo.

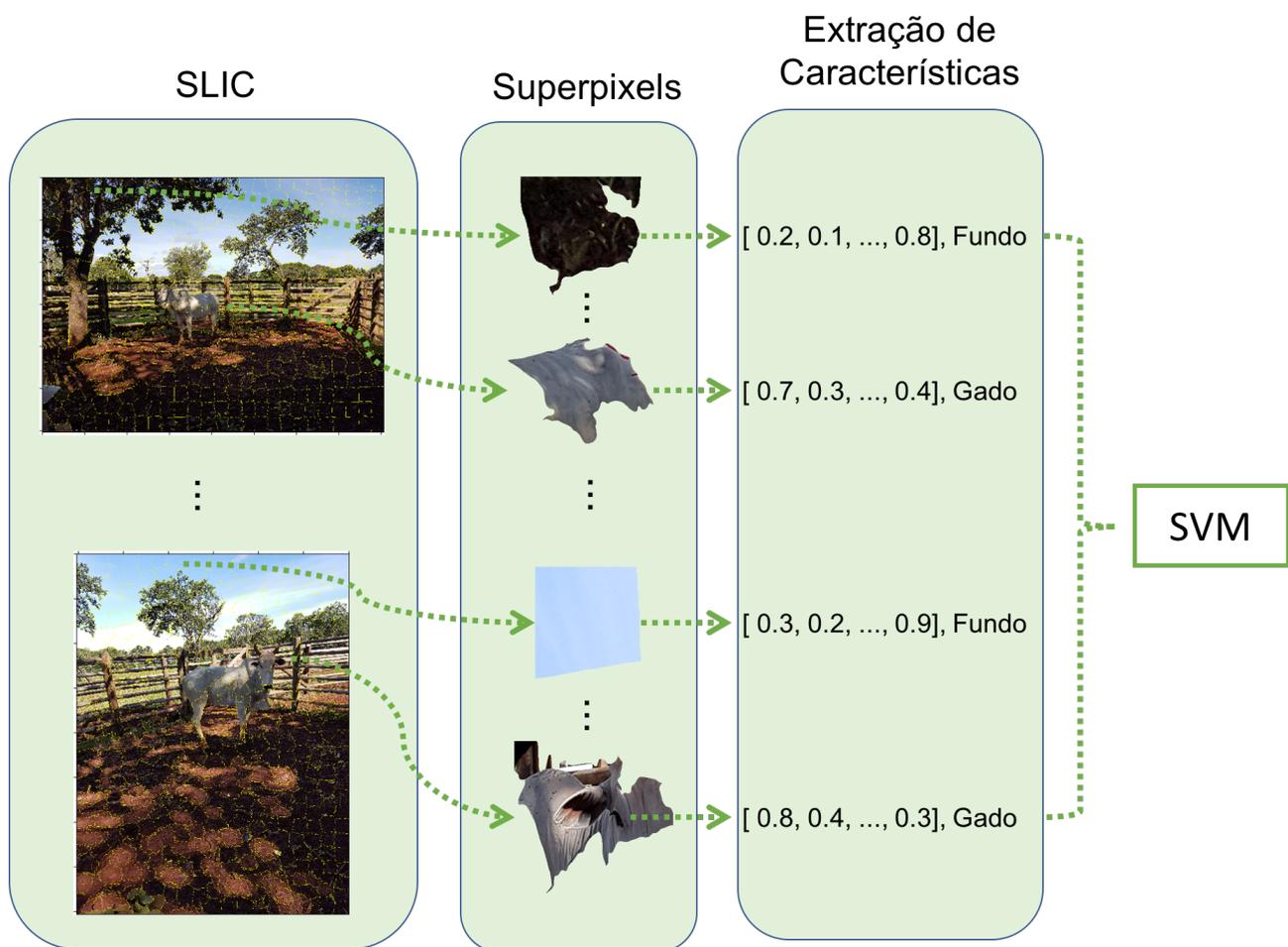


Figura 2: Ilustração da abordagem utilizada nos experimentos preliminares. As imagens de treinamento são processadas pelo SLIC para obtenção dos superpixels. Para cada superpixel, um vetor de características é calculado. Por fim, os vetores de características e as respectivas classes (gado ou fundo) são utilizados para o treinamento do classificador SVM.

Tabela 2: Precisão, revocação e medida-f da classe gado para diferentes extratores de características.

Extrator de Características	Precisão	Revocação	Medida-f
Cor	83.4	59.5	69.4
GLCM	87.3	26.9	41.1
LBP	0.0	0.0	0.0
HOG	0.0	0.0	0.0
Cor+GLCM+LBP+HOG	87.3	72.2	79.0



Figura 3: Duas imagens representando o resultado da segmentação utilizando o software Pynovisão. Os superpixels em vermelho foram classificados como gado e os superpixels em azul foram classificados como fundo.

7 Resultados Esperados, Produtos e Avanços

Este projeto contribuirá para o desenvolvimento teórico e prático da área de visão computacional e segmentação de imagens. A constatação da eficácia do método proposto para a segmentação de imagens de gado permitirá um avanço também na pecuária de precisão, pois a segmentação é considerada como uma parte muito importante em um sistema de visão computacional para caracterização e classificação de imagens. Além disso este método poderá ser utilizado como base de um projeto que visa estimar o peso do gado, que auxiliaria pesquisadores na tomada de decisão para aplicar medidas preventivas para o melhoramento da saúde do animal. Espera-se também auxiliar em diversos outros problemas relacionados a segmentação, como na parte de segmentação de carcaças no frigorífico, o que ajudaria em sua classificação, bem como segmentar rebanhos bovinos para ser possível uma contagem, por exemplo. Estima-se também que o projeto resulte publicações em conferências e revistas nacionais e internacionais.

Referências

[Achanta et al., 2012] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., e Süsstrunk, S. (2012). SLIC Superpixels Compared to State-of-the-art Superpixel Methods. *IEEE Tran-*

- sactions on Pattern Analysis and Machine Intelligence*, 34(11):2274 – 2282.
- [Beling, 2014] Beling, R. R. (2014). Anuário brasileiro da pecuária. *Editora Gazeta Santa Cruz Ltda.*
- [Comaniciu e Meer, 2002] Comaniciu, D. e Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619.
- [Dalal e Triggs, 2005] Dalal, N. e Triggs, B. (2005). Histograms of oriented gradients for human detection. Em *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, páginas 886–893. IEEE.
- [Everingham et al., 2015] Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., e Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136.
- [Felzenszwalb e Huttenlocher, 2004] Felzenszwalb, P. F. e Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59(2):167–181.
- [Frost et al., 1997] Frost, A., Schofield, C., Beaulah, S., Mottram, T., Lines, J., e Wathes, C. (1997). A review of livestock monitoring and the need for integrated systems. *Computers and Electronics in Agriculture*, 17(2):139 – 159.
- [Graves et al., 2013] Graves, A., Mohamed, A., e Hinton, G. (2013). Speech recognition with deep recurrent neural networks. Em *IEEE International Conference on Acoustics, Speech and Signal Processing*, páginas 6645–6649.
- [Halachmi et al., 2008] Halachmi, I., Polak, P., Roberts, D., e Klopčič, M. (2008). Cow body shape and automation of condition scoring. *Journal of Dairy Science*, 91(11):4444–4451.
- [Haralick, 1979] Haralick, R. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804.

- [Hertem et al., 2013] Hertem, T. V., Alchanatis, V., Antler, A., Maltz, E., Halachmi, I., Schlageter-Tello, A., Lokhorst, C., Viazzi, S., Romanini, C., Pluk, A., Bahr, C., e Berckmans, D. (2013). Comparison of segmentation algorithms for cow contour extraction from natural barn background in side view images. *Computers and Electronics in Agriculture*, 91:65 – 74.
- [Karpathy et al., 2014] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., e Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. Em *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, páginas 1725–1732.
- [Kist, 2016a] Kist, B. B. (2016a). Anuário brasileiro do gado de corte. *Editores Gazeta Santa Cruz Ltda.*
- [Kist, 2016b] Kist, B. B. (2016b). Anuário brasileiro do gado de leite. *Editores Gazeta Santa Cruz Ltda.*
- [Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., e Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Em Pereira, F., Burges, C. J. C., Bottou, L., e Weinberger, K. Q., editores, *Advances in Neural Information Processing Systems 25*, páginas 1097–1105. Curran Associates, Inc.
- [LeCun et al., 1989] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., e Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551.
- [Levinshtein et al., 2009] Levinshtein, A., Stere, A., Kutulakos, K. N., Fleet, D. J., Dickinson, S. J., e Siddiqi, K. (2009). Turbopixels: Fast superpixels using geometric flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2290–2297.
- [Lin et al., 2014] Lin, T., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., Perona, P., Ramanan, D., Dollár, P., e Zitnick, C. L. (2014). Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312.

- [Luo et al., 2012] Luo, P., Wang, X., e Tang, X. (2012). Hierarchical face parsing via deep learning. Em *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, páginas 2480–2487.
- [Ojala et al., 2002] Ojala, T., Pietikainen, M., e Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987.
- [Shelhamer et al., 2017] Shelhamer, E., Long, J., e Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651.
- [Shi e Malik, 2000] Shi, J. e Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905.
- [Sun et al., 2013] Sun, Y., Wang, X., e Tang, X. (2013). Deep convolutional network cascade for facial point detection. Em *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, páginas 3476–3483.
- [Toshev e Szegedy, 2014] Toshev, A. e Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. Em *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, páginas 1653–1660.
- [Vedaldi e Soatto, 2008] Vedaldi, A. e Soatto, S. (2008). *Quick Shift and Kernel Methods for Mode Seeking*, páginas 705–718. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Viazzi et al., 2014] Viazzi, S., Bahr, C., Hertem, T. V., Schlageter-Tello, A., Romanini, C., Halachmi, I., Lokhorst, C., e Berckmans, D. (2014). Comparison of a three-dimensional and two-dimensional camera system for automated measurement of back posture in dairy cows. *Computers and Electronics in Agriculture*, 100:139 – 147.

[Vindis et al., 2010] Vindis, P., Brus, M., Stajnko, D., e Janzekovic, M. (2010). Non invasive weighing of live cattle by thermal image analysis. Em *New Trends in Technologies: Control, Management, Computational Intelligence and Network Systems*, capítulo 13, páginas 243 – 257. Meng Joo Er (Ed.), InTech.