

## TÉCNICAS DE SEGMENTAÇÃO BASEADAS EM SUBTRAÇÃO DE FUNDO E MODELOS DE CORES: UM ESTUDO COMPARATIVO

Wesley Nunes Gonçalves, Vinícius Assis Saueia, Bruno Brandoli Machado,  
Jonathan de Andrade Silva, Kleber Padovani de Souza\* e Hemerson Pistori

Grupo de Pesquisa em Engenharia e Computação  
Universidade Católica Dom Bosco  
Av. Tamandaré, 6000, Jardim Seminário, 79117-900, Campo Grande, MS  
e-mail: {wnunes, vsaueia, bmachado, jsilva, kleber}@acad.ucdb.br, pistori@ucdb.br  
web: <http://www.gpec.ucdb.br>

**Palavras chave:** Segmentação, Subtração de Fundo, Modelos de Cores

**Resumo.** Neste trabalho foram comparados dois grupos de técnicas de segmentação de imagens, o primeiro baseado em subtração de fundo e o segundo em aprendizagem supervisionada de modelos de cores. Entre as técnicas de subtração de fundo, foram testadas tanto técnicas adaptativas, quanto não-adaptativas. Na abordagem por aprendizagem supervisionada, foram realizados experimentos com modelos Gaussianos, árvores de decisão inferidas pelo algoritmo C4.5, redes neurais artificiais e máquinas de vetores de suporte. As comparações foram realizadas com base em imagens reais, segmentadas manualmente, referentes a dois importantes problemas que vêm sendo estudados por diversos grupos de pesquisa em visão computacional: a interpretação de língua gestual de sinais e a análise de comportamentos de camundongos em experimentos na área da saúde. No caso da interpretação de língua de sinais, foram utilizadas imagens com diferentes sinalizadores, com diferentes tons de pele, em ambientes de fundo estático, em laboratório, com iluminação artificial, e em ambientes externos, com iluminação natural e fundo complexo. Para esse problema, testou-se a capacidade dos segmentadores de extrair da imagem regiões contendo a face e as mãos de um ser humano. O experimento com camundongos baseou-se no teste do campo aberto, um teste bastante utilizado na área da saúde, em que os animais são colocados em uma arena e seu comportamento é analisado para que se possa identificar mudanças significativas após a utilização de alguma droga ou modificação genética. Nesse caso, o problema para os segmentadores consistia em extrair da imagem as regiões contendo camundongos. Para cada técnica foi escolhida a melhor configuração de parâmetros seguindo as mesmas métricas utilizadas na comparação entre as técnicas. As métricas utilizadas foram baseadas no trabalho de M. Sezgin e B. SanKur, e que consideram, além da taxa de acerto na classificação de cada pixel, informações relacionadas com as bordas e a forma geral do objeto segmentado em comparação com a imagem segmentada manualmente (ground truth).

## 1 INTRODUÇÃO

A segmentação é uma importante etapa para a análise de imagens em sistemas de visão computacional. O objetivo da segmentação é o particionamento de imagens em objetos de interesse e objetos irrelevantes. Este processo simplifica a representação da imagem de forma mais significativa para análise do problema. Em segmentação de imagens não existe uma técnica geral e robusta que obtenha um resultado aceitável para todas as aplicações [6]. Por isso, a escolha da técnica é influenciada pela aplicação e requerimentos do problema, sendo importante avaliar diversas técnicas propostas. Entretanto, a dificuldade está em avaliar a qualidade dos resultados obtidos pelos algoritmos de segmentação. Geralmente, a acurácia da segmentação é realizada pela comparação dos resultados dos algoritmos com a segmentação manual, que consiste em contornar os limites do objeto de interesse por um ser humano.

Um importante problema que vêm sendo estudado é o reconhecimento automático de gestos de línguas de sinais [7]. A segmentação da pele humana é um fator essencial em situações que a imagem é a fonte de informação e o sinalizador não utiliza recursos de auxílio, como luvas coloridas ou marcadores. A identificação de determinadas partes do corpo na imagem, como as mãos e a face, possibilita a coleta de informações relevantes para esse processo de reconhecimento, dado que o sinalizador é o objeto de interesse nessa aplicação. Através da utilização de outras técnicas na imagem resultante do processo de segmentação, é possível obter características que podem contribuir na classificação correta dos gestos, como informações relacionadas ao posicionamento espacial de membros do corpo, configurações das mãos e expressões faciais.

Um outro problema estudado é a análise de comportamentos de camundongos em experimentos do campo aberto<sup>1</sup> [11]. O estudo do comportamento de camundongos é de suma importância para responder questões etológicas básicas [13]. O uso de animais de laboratório em pesquisas, em particular os camundongos, serve de modelos simplificados do comportamento humano [8, 3], permitindo que os novos fármacos sejam avaliados antes de serem testados em seres humanos. A segmentação do camundongo contribui para a automatização dos sistemas no monitoramento do animal. Essa observação automática é particularmente apropriada para registrar os comportamentos que ocorrem momentaneamente após períodos longos, em que o observador humano é incapaz de estimar com exatidão as informações espaciais, como por exemplo, distância percorrida e velocidade. Essas características provavelmente não seriam alcançadas sem a distinção de alvo e fundo obtida na segmentação da imagem original.

O objetivo deste trabalho é comparar técnicas de segmentação aplicada aos dois importantes problemas que vêm sendo estudados por diversos grupos de pesquisa em visão computacional: a interpretação de língua gestual de sinais e a análise de comportamentos

---

<sup>1</sup>O campo aberto é um teste onde são realizados experimentos com roedores, introduzidos em uma arena, geralmente de forma circular com raio de 20cm. O piso da arena é revestido de fórmica de cor branca ou preta para que haja um contraste entre o camundongo e o piso, facilitando a observação do pesquisador e os algoritmos de segmentação.

de camundongos em experimentos na área da saúde. Os experimentos foram realizados sobre um conjunto de 40 imagens para camundongos e 240 para língua gestual de sinais. Nas imagens, foi realizado um processo de segmentação manual e segmentação automática através dos algoritmos. Para cada técnica de segmentação foi escolhida a melhor configuração de parâmetros seguindo as mesmas métricas utilizadas na comparação entre as técnicas. As métricas utilizadas foram baseadas no trabalho de M. Sezgin e B. SanKur [16].

Este artigo está estruturado em sete seções. A Seção 2 apresenta alguns trabalhos correlatos que utilizam técnicas de segmentação. Na Seção 3, são descritas as técnicas de segmentação avaliadas neste presente trabalho. Na próxima seção, são apresentados os métodos de avaliação dos resultados dos algoritmos de segmentação. Na Seção 5, os experimentos realizados são mostrados com maiores detalhes. A Seção 6 apresenta os resultados obtidos seguido de uma análise. Finalmente, na Seção 7 são mostradas a conclusão e os trabalhos futuros.

## 2 TRABALHOS CORRELATOS

As técnicas de segmentação podem ser classificadas através das características utilizadas no processo de análise de imagens [5]. As principais características descritas por Cucchiara são os movimentos da câmera (e.g. fixa ou em movimento), sistemas de aquisição de imagens (e.g. única câmera - 2D, duas câmeras - 3D), características de segmentação (e.g. movimento, visuais e características híbridas) e utilização em tempo real.

Em [5] é apresentado um método de segmentação baseado em cores seguido de uma junção de movimentos através dos campos aleatórios de Markov (MRF). O sistema foi desenvolvido para ser executado em tempo real, processando 10 *frames* por segundo. O método foi avaliado em imagens sintéticas e reais comparadas as imagens segmentadas manualmente. As métricas utilizadas foram o tempo de processamento, falsos positivos, falsos negativos, erro médio da área e do centro do objeto. O método apresentou resultados satisfatórios, entretanto, é limitado com relação à mudança de iluminação do ambiente.

A segmentação implementada por Haritaoglu e Harwood [9] combina a análise do fundo da cena e um simples pós-processamento na imagem binária resultante. O fundo da cena é modelado estatisticamente pelos valores de máximo e mínimo de cada *pixel* e a maior diferença entre quadros consecutivos. Os *pixels* são marcados como objeto se a diferença entre o *pixel* e o valor máximo ou mínimo for maior que a diferença obtida na fase de treinamento. Esse método de segmentação é utilizado em um sistema de rastreamento em tempo real de pessoas utilizando imagens em níveis de cinza.

Um novo algoritmo para segmentação de objetos em movimento foi desenvolvido em [4]. Este novo algoritmo combina a subtração de fundo adaptativo com o algoritmo de diferenciação de quadros. A diferenciação de quadros é executada para determinar quais regiões estão em movimento e a subtração de fundo adaptativo auxilia a identificação da região total em movimento.

Uma característica muito utilizada na segmentação é a cor. A cor é a característica

mais saliente a ser explorada [5, 17], pois oferece diversas vantagens sobre outras características, como mudanças de escalas e resoluções, possibilidade de movimento da câmera e robustez em oclusões [1]. Mas uma das maiores dificuldades das técnicas que utilizam a cor é a seleção do espaço de cores. Os espaços de cores que separam a crominância dos componentes de luminância são geralmente preferíveis [1]. Com isso, a mudança da iluminação do ambiente não irá interferir drasticamente na segmentação. Terrillon e Fukamachi [17] compararam nove espaços de cores no processo de segmentação. Neste trabalho, são apresentados dois modelos de segmentação que utilizam a cor, o modelo Gaussiano baseado na distância de Mahalanobis e o modelo de mistura de Gaussianas, com o objetivo de segmentar faces em imagens coloridas.

Em [12] é realizada uma comparação de diversas técnicas de subtração de fundo descritas na literatura. Entre as oito técnicas, está a mistura de Gaussianas adaptativa. Para esta técnica em particular, para cada *pixel* é construída uma mistura de  $K$  Gaussianas, representando o modelo de fundo. Após a detecção dos objetos, os parâmetros das Gaussianas são atualizados com base em uma constante de atualização. As diferentes técnicas foram avaliadas através de três principais parâmetros: distinção dos *pixels* de frente em relação aos *pixels* do fundo, armazenamento e atualização do fundo sobre o tempo e pós-processamento das áreas dos objetos com intuito de eliminar falsos positivos da imagem resultante.

### 3 SEGMENTAÇÃO

Nesta seção é descrita uma revisão das técnicas de segmentação utilizadas neste presente estudo. Entre as técnicas de subtração, foram descritas a subtração de fundo e a subtração de fundo adaptativo. Na abordagem por aprendizagem supervisionada, foram descritos modelos Gaussianos, árvores de decisão, redes neurais artificiais e máquinas de vetores de suporte.

#### 3.1 Subtração de Fundo

A subtração de fundo [12] é uma das técnicas mais utilizadas devido a sua fácil implementação e seu baixo custo de processamento. O princípio básico é subtrair cada nova imagem de uma imagem de referência, obtida a partir de um fundo estático e sem o objeto de interesse. A grande dificuldade apresentada pela subtração de fundo é não se adaptar a modificações no fundo da cena, como mudança de iluminação ou objetos que entram em cena e em seguida ficam estáticos e sem relevância. Para contornar esse problema, existem técnicas de subtração de fundo mais sofisticadas, como a subtração de fundo adaptativo.

A imagem segmentada através da subtração de fundo é obtida pela Equação 1, onde  $B_n(x)$  é a intensidade da imagem de referência para o *pixel*  $x$ ,  $T_n(x)$  é o limiar atual atribuído ao *pixel*  $x$ ,  $I_n(x)$  é a intensidade do quadro que se deseja segmentar e  $n$  é a quantidade de quadros capturados até o momento. Para simplificação, em diversos sistemas as imagens são processadas em níveis de cinza e o limiar  $T_n(x)$  é o mesmo para

todos os *pixels* da imagem.

$$|I_n(x) - B_n(x)| > T_n(x) \quad (1)$$

### 3.2 Subtração de Fundo Adaptativo

A subtração de fundo adaptativo [12] possui a vantagem de se adaptar a pequenas modificações na imagem de referência, sendo útil em longos períodos de tempo e ambientes com iluminação natural. Um exemplo de longos períodos de tempo se aplica em experimentos com camundongos em que possui a necessidade de avaliar um novo fármaco em diversas horas de duração.

A imagem segmentada é obtida pela mesma equação da subtração de fundo, a Equação 1. A principal diferença está na modificação da imagem de referência e dos limiares. A imagem de referência e o limiar são determinados estatisticamente a partir das seqüências de imagens  $I_k$ , para  $k < n$ . O fundo e o limiar são atualizados a cada nova imagem através das Equações 2 e 3.

$$B_{n+1}(x) = \begin{cases} \alpha B_n(x) + (1 - \alpha)I_n(x), & \text{se } x \text{ não está em movimento} \\ B_n(x), & \text{se } x \text{ está em movimento} \end{cases} \quad (2)$$

$$T_{n+1}(x) = \begin{cases} \alpha T_n(x) + (1 - \alpha)(y * |I_n(x) - B_n(x)|), & \text{se } x \text{ não está em movimento} \\ T_n(x), & \text{se } x \text{ está em movimento} \end{cases} \quad (3)$$

onde  $\alpha$  é um valor constante que especifica o peso das novas informações. O fundo é atualizado somente para os *pixels* que não apresentam movimento, considerando assim, apenas os *pixels* que realmente correspondem ao fundo.

### 3.3 Modelo Gaussiano

Esta técnica assume que as intensidades dos *pixels* do objeto de interesse pode ser modelado por uma distribuição Gaussiana, definida como  $N(\mu, \Sigma)$  [17]. Os parâmetros da Gaussiana são estimados através de um conjunto de treinamento, calculado na Equação 4.

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad \Sigma = \frac{1}{n} \sum_{i=1}^n [x_i - \mu][x_i - \mu]^T \quad (4)$$

onde  $\mu$  representa a média,  $\Sigma$  representa a matriz de covariância,  $x$  representa um vetor do conjunto de treinamento e  $n$  a quantidade de vetores de treinamento. Para classificar um *pixel*  $p$  é pertencente ao objeto de interesse modelado por uma Gaussiana  $N(\mu, \Sigma)$ , é necessário medir a similaridade do *pixel*  $p$  à distribuição. Qualquer medida de similaridade pode ser utilizada, como distância euclidiana ou distância de Mahalanobis. A medida de similaridade adotada neste artigo é a distância de Mahalanobis descrita pela Equação 5.

$$D_m = [x_i - \mu]^T \Sigma^{-1} [x_i - \mu] \quad (5)$$

### 3.4 Aprendizagem Supervisionada

A aprendizagem supervisionada é um método de aprendizagem de máquina baseado em experiências acumuladas contidas no conjunto de treinamento. O conjunto de treinamento é composto por atributos, geralmente vetores, e suas respectivas classes etiquetadas por um especialista. Os algoritmos de aprendizagem de máquina utilizados neste trabalho foram as árvores de decisão, máquinas de vetores de suporte e redes neurais do tipo *perceptron* de várias camadas (*Multi-Layer Perceptron*).

As árvores de decisão [18] é um dos algoritmos mais utilizados na área de aprendizagem de máquina. A idéia utilizada nesse método é recursivamente dividir o conjunto de treinamento mais complexo em um subconjunto mais simples. Os algoritmos mais utilizados na construção das árvores são os C4.5, ID3 e *Assistant*. Geralmente, estes algoritmos utilizam os conceitos de ganho e entropia para a escolha dos melhores atributos que irão constituir os primeiros atributos da árvore.

As máquinas de vetores de suporte (*Support Vector Machine - SVM*) [2] são aplicáveis à problemas lineares e não linearmente separáveis. O princípio desse método é minimizar o risco estrutural criando hiperplanos de separação ótimos. Um hiperplano ótimo é um subespaço que separa o espaço de características de forma robusta e computacionalmente eficiente. Essa eficiência é uma importante característica quando se trabalha com grandes quantidades de dados, como é o caso da visão computacional e o processamento de imagens. Dado um conjunto de treinamento  $X$  e assumindo que o problema é linearmente separável, a classificação através de uma SVM é baseada em um hiperplano constituído por alguns pontos do conjunto de treinamento (vetores de suporte). Em casos onde os dados não são separáveis linearmente, a criação do hiperplano é baseada em uma transformação dos dados iniciais através de algoritmos de separabilidade de padrões. A transformação nos dados iniciais são realizadas através de funções de núcleos, como função polinomial, função de base radial e *spline* linear [10].

As redes neurais são sistemas computacionais representados por modelos matemáticos para o processamento de informações [14]. As redes adquirem conhecimento através de experiências que definem o processo de aprendizagem. Uma rede é composta por diversas unidades de processamento interconectadas por canais de comunicações com um determinado peso. O processo de aprendizagem consiste em um ajuste iterativo aplicado aos pesos das conexões de acordo com o fluxo de informações recebidas. Cada rede neural possui sua classe com um determinado algoritmo para o treinamento. Para este trabalho foi utilizada a classe de rede *perceptron* com várias camadas (*Multi-Layer Perceptron*), geralmente interconectadas em uma forma de realimentação. Para o treinamento da rede, foi utilizado o algoritmo de propagação reversa (*Backpropagation*).

## 4 MÉTODOS DE AVALIAÇÃO

Os métodos de avaliação geralmente utilizam *pixels* individuais ou regiões segmentadas para avaliar o desempenho dos algoritmos de segmentação através de uma imagem segmentada manualmente. Esses métodos de avaliação são de extrema importância pois são resultados confiáveis do desempenho demonstrado pelos algoritmos, visto que, a avaliação visual não exprime o verdadeiro desempenho dos algoritmos.

### 4.1 Avaliação por Pixel

A avaliação por *pixels* consiste na análise individual de cada *pixel* da imagem segmentada manualmente com o mesmo *pixel* da imagem segmentada pelos algoritmos. Estas métricas de avaliação são baseadas no número de ocorrências de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN). O verdadeiro positivo consiste no número de *pixels* do objeto de interesse classificados corretamente pelo algoritmo. O verdadeiro negativo é o número de *pixels* do fundo corretamente detectado. O falso positivo corresponde ao número dos *pixels* do fundo inferidos incorretamente como objeto. O falso negativo é o número de *pixels* do objeto classificados incorretamente como *pixels* do fundo.

Em seguida, pode-se aplicar os valores obtidos nas métricas de percentual de classificação correta, coeficiente de Jaccard e coeficiente de Yule, descritos nas Equações 6, 7 e 8, [15]. Essas métricas são delimitadas pelo intervalo  $[0, 1]$ , onde o valor 1 representa acerto máximo.

$$PCC = \frac{VP + VN}{VN + VP + FN + FP} \quad (6)$$

$$CJ = \frac{VP}{VP + FN + FP} \quad (7)$$

$$CY = \left| \frac{VP}{VP+FP} + \frac{VN}{VN+FN} - 1 \right| \quad (8)$$

### 4.2 Avaliação por Regiões

A análise por regiões baseia-se na distribuição dos *pixels* da imagem. Essas métricas avaliam as informações relacionadas com as bordas e a forma geral do objeto segmentado em comparação com a imagem segmentada manualmente [16]. A métrica RAE ou erro relativo de área, mostrado na Equação 9, relaciona a área da imagem segmentada de forma manual  $A_o$  e automática  $A_t$ . Essa métrica é normalizada na escala  $[0, 1]$ , onde RAE é 0 para um casamento máximo das regiões da imagem segmentada, enquanto que 1 indica a sobreposição mínima das regiões.

$$RAE = \begin{cases} \frac{A_o - A_t}{A_o}, & se A_t < A_o \\ \frac{A_t - A_o}{A_t}, & se A_t \geq A_o \end{cases} \quad (9)$$

A métrica ME (Misclassification Error) calcula a porcentagem de *pixels* incorretamente classificados como fundo ou objeto de interesse. A Equação 10 mostra o cálculo dessa métrica para o problema de segmentação, onde são classificadas somente duas classes.

$$ME = 1 - \frac{|B_o \cap B_t| + |F_o \cap F_t|}{|B_o| + |F_o|} \quad (10)$$

onde  $B_o$  e  $F_o$  denotam o fundo e os objetos de interesse da imagem segmentada manualmente,  $B_t$  e  $F_t$  denotam o fundo e os objetos de interesse da imagem segmentada pelo algoritmo e  $|\cdot|$  é a cardinalidade do conjunto. A métrica ME possui uma classificação perfeita quando assume o valor 0 e uma classificação totalmente incorreta quando assume o valor 1.

## 5 EXPERIMENTOS

Nesta seção, é descrito a forma com que os experimentos foram realizados. As imagens digitais utilizadas possuem dimensão de 640x480 e 800x600 *pixels*, capturadas por uma câmera Sony Cyber-shot® com resolução de 8 *MPixels*. O ambiente computacional utilizado nos experimentos foi um computador com processador Intel® Pentium 4™ com frequência de *clock* de 3.00 GHz, 512 *MBytes* de memória e sistema operacional Linux Fedora Core 5.

Os experimentos foram realizados com dois grupos de técnicas de segmentação de imagens. O primeiro grupo baseado em subtração de fundo e o segundo em aprendizagem supervisionada de modelos de cores, descritos na Seção 3. As comparações foram realizadas com base em conjunto de 240 imagens para a interpretação de língua gestual de sinais e um outro conjunto de 40 imagens para a análise de comportamentos de camundongos em experimentos na área da saúde.

O conjunto de imagens para a interpretação de língua gestual de sinais possui 10 posturas que exploram situações comuns nos gestos, apresentados na Figura 1. As imagens das posturas foram capturadas com diferentes sinalizadores e diferentes tons de pele, em ambientes de fundo estático, em laboratório, com iluminação artificial, e em ambientes externos, com iluminação natural e fundo complexo.

O conjunto de imagens para a análise de experimentos com camundongos possui dois comportamentos, o comportamento locomotor e o de exploração dos animais, mostrados na Figura 2. O movimento locomotor é o deslocamento entre um ponto a outro da arena realizado pelo animal. O movimento de exploração ou não locomotor é aquele que o animal pode realizar sem a necessidade de deslocamento, como por exemplo exploração vertical. A exploração vertical consiste na elevação das patas posteriores do camundongo.

Nos experimentos, foram utilizadas um total de 280 imagens. As imagens passaram por um processo de segmentação manual, que consiste em delimitar os objetos de interesse na imagem por um ser humano. Este método de segmentação foi realizado por cinco pessoas durante 70 horas, consumindo um tempo considerável para a realização dos experimentos.



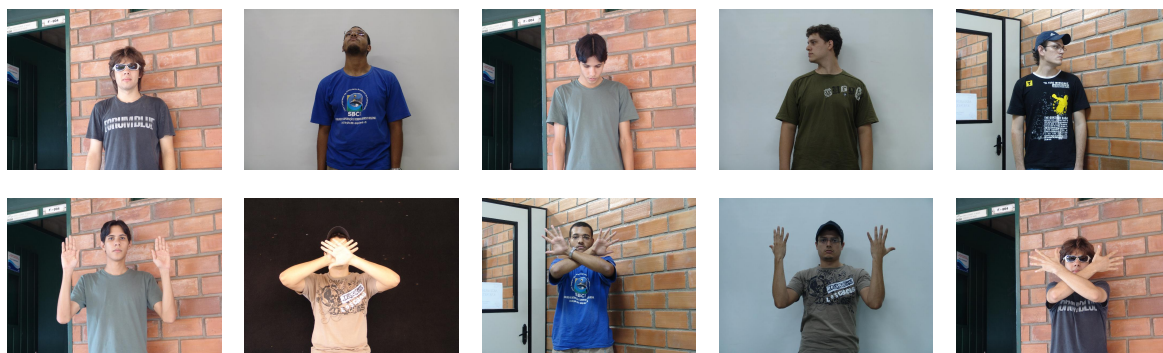


Figura 1: Posturas exploradas pelos sinalizadores para interpretação de língua gestual de sinais.

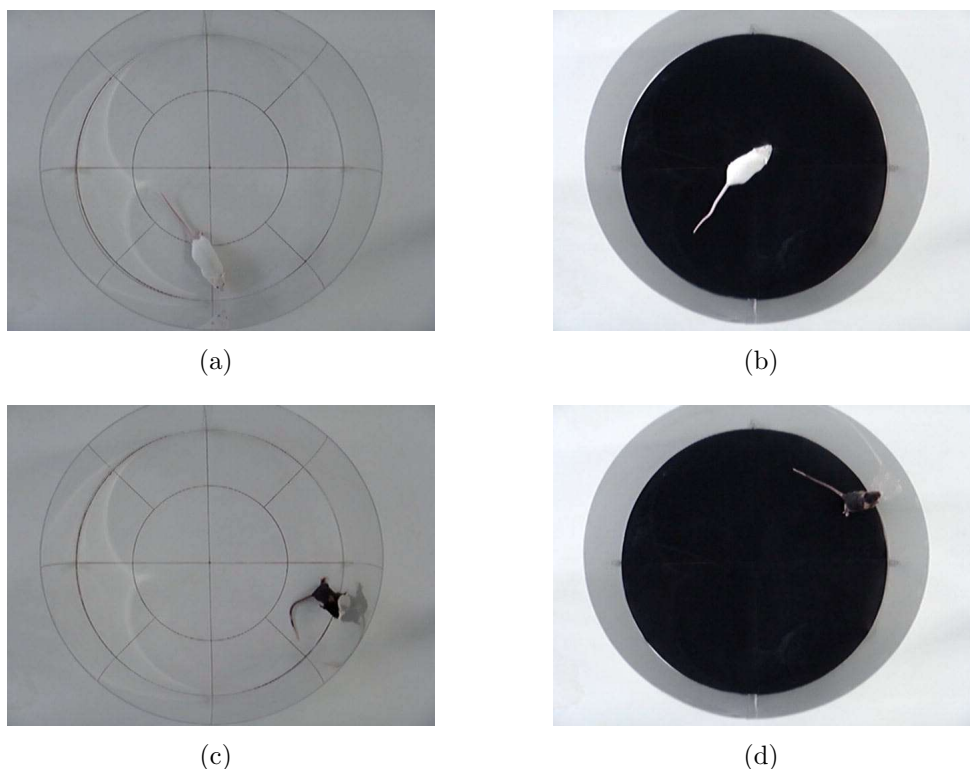


Figura 2: (a) (b) Exemplos de movimento locomotor e (c) (d) exploração vertical aplicados à análise de comportamentos de camundongos em experimentos para o teste do campo aberto.

Para o treinamento das técnicas de aprendizagem supervisionada foram coletas amostras de todas as imagens. Para as imagens do conjunto de interpretação de língua de sinais foram coletas cinco amostras da pele de diferentes partes do corpo e cinco imagens de fundo com cores que se destacavam, como roupa, cabelo entre outras. Para as imagens do conjunto de análise de experimentos com camundongos foram coletas cinco amostras

do corpo do animal e cinco amostras do fundo. Para as técnicas de subtração de fundo foi utilizado um quadro de referência com fundo estático.

Para cada técnica de segmentação foi escolhida a melhor configuração de parâmetros. A variação dos parâmetros das técnicas de segmentação geraram cerca de 70000 imagens segmentadas. A avaliação da segmentação automática utilizou as mesmas métricas na comparação entre as técnicas implementadas. As métricas utilizadas estão descritas na Seção 4. O valor final da avaliação é obtido pela soma das métricas de comparações dividido pela quantidade de métricas.

## 6 RESULTADO E ANÁLISE

Nesta seção, são apresentados os resultados obtidos nos experimentos. Esses resultados foram avaliados em cada grupo de imagens. Para as imagens da interpretação de língua gestual de sinais, os grupos estão divididos de acordo com o fundo, a iluminação e a tonalidade da pele. Para as imagens da análise de experimentos com camundongos, foram divididos em relação à cor do camundongo e do fundo da arena do experimento.

A porcentagem de acerto para cada grupo é dada nas Tabelas 1 e 2. Cada tabela corresponde à um conjunto de imagens e o melhor resultado obtido pelas técnicas de segmentação devido aos seus parâmetros. A porcentagem de acerto é calculada através da média das métricas para o conjunto.

Técnica	Iluminação Artificial			
	Fundo Preto		Fundo Branco	
	*Escuro	Claro	Escuro	Claro
Sub. de Fundo	73,81	91,03	86,37	68,24
Sub. de Fundo Adaptativo	71,20	87,14	72,68	65,74
Molelo Gaussiano	65,68	44,34	81,11	39,98
Árvores de Decisão	40,03	39,21	79,43	35,10
SVM	45,28	43,54	84,91	40,22
Redes Neurais	42,32	38,76	81,12	35,61

Tabela 1: Resultados obtidos pelos segmentadores para o conjunto de imagens aplicados à análise de comportamentos de camundongos. \*Indica que é um camundongo com o tom de sua pelagem, neste caso Escuro ou Claro.

Os resultados mostraram que para as imagens da análise de comportamentos de camundongos o melhor resultado foi obtido pela técnica de subtração de fundo. Esse bom resultado é obtido porque o fundo é estático e a iluminação é controlada. Com isso, o fundo se mantém praticamente o mesmo para todo o experimento. Ainda é possível notar, que essa técnica com o limiar bem ajustado para cada situação pode obter bons resultados, como foi mostrado na segmentação de camundongos pretos com fundo preto.

Para as imagens que representam a interpretação de língua gestual de sinais, o melhor resultado foi alcançado por diversas técnicas. Com fundo complexo, as técnicas baseadas

Técnica	Iluminação Artificial				Iluminação Natural			
	F. Complexo		F. Simples		F. Complexo		F. Simples	
	*Escura	Clara	Escura	Clara	Escura	Clara	Escura	Clara
Sub. de Fundo	59,92	53,36	60,62	47,34	55,20	50,90	65,76	73,29
Sub. de Fundo Adap.	59,14	56,12	61,206	58,083	56,005	53,686	69,531	74,229
Modelo Gaussiano	54,50	57,70	80,30	71,75	42,80	59,95	68,99	72,38
Árvores de Decisão	45,67	41,42	90,09	87,13	47,14	48,78	88,88	80,11
SVM	49,21	43,91	92,18	90,87	51,01	52,84	89,23	86,61
Redes Neurais	44,45	44,53	89,65	88,10	50,71	48,16	89,12	84,83

Tabela 2: Resultados obtidos pelos segmentadores para o conjunto de imagens da interpretação de língua gestual de sinais. \*Indica a tonalidade da pele do sinalizador, neste caso Escura ou Clara.

em subtração de fundo obtiveram melhores resultados para ambas as tonalidades de pele e iluminação. Em fundo simples, a segmentação pelas máquinas de vetores de suporte obtiveram um bom resultado. Com esses resultados, é possível mostrar que as técnicas de segmentação são bastante sensíveis ao fundo em questão.

## 7 CONCLUSÕES E TRABALHOS FUTUROS

Neste estudo, foram realizadas comparações de diferentes técnicas de segmentação. Para isso, essas técnicas foram aplicadas em imagens de dois importantes problemas, reconhecimento automático de gestos para línguas de sinais e análise de comportamentos de camundongos. A avaliação dos resultados foi baseada em métricas que comparam uma imagem segmentada manualmente com uma imagem segmentada pelo algoritmo.

Os experimentos realizados mostraram que as técnicas baseadas em subtração de fundo possuem o melhor resultado em ambientes com fundo complexo, enquanto imagens com fundo simples, o melhor resultado para segmentação automática das imagens foi alcançado pela segmentação das máquinas de vetores de suporte, incluindo grande parte do conjunto de imagens da interpretação de língua gestual de sinais. Para a análise de comportamentos de camundongos, a técnica que obteve melhor resultado foi a subtração de fundo. Os experimentos realizados são úteis na avaliação das técnicas. Porém, é importante que os experimentos sejam expandidos para diferentes configurações de fundo, como ambientes com árvores produzindo sombras. Para isso, seria interessante incluir um pré-processamento nas imagens. Contudo, vale ressaltar que este estudo foi aplicado a somente dois problemas, sendo interessante ampliar o conjunto de técnicas de segmentação e realizar novos experimentos com outros problemas.

## AGRADECIMENTOS

Este trabalho recebeu apoio financeiro da Universidade Católica Dom Bosco, UCDB, da Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul, FUNDECT e da Agência Financiadora de Estudos e Projetos, FINEP. Alguns dos acadêmicos que participaram do projeto SIGUS receberam bolsas

PIBIC/CNPQ e da própria UCDB.

## REFERÊNCIAS

- [1] Antonis A. Argyros and Manolis I. A. Lourakis, *Real time tracking of multiple skin-colored objects with a possibly moving camera*, Proceedings of the European Conference on Computer Vision (ECCV 2004), vol. 3, Springer, May 11-14 2004, pp. 368–379.
- [2] Christopher J. C. Burges, *A tutorial on support vector machines for pattern recognition*, Data Mining and Knowledge Discovery **2** (1998), no. 2, 121–167.
- [3] Tais H. F. Carvalho and Oswaldo U. Lopes, *O emprego de camundongo geneticamente modificado como modelo de estudo para doenças cardiovasculares*, X Simpósio Brasileiro de Fisiologia Cardiovascular (Ribeirão Preto, Brasil), vol. 39, 2006, pp. 110–116.
- [4] Robert Collins, Alan Lipton, Takeo Kanade, Hironobu Fujiyoshi, David Duggins, Yanghai Tsin, David Tolliver, Nobuyoshi Enomoto, and Osamu Hasegawa, *A system for video surveillance and monitoring*, Tech. Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, May 2000.
- [5] Rita Cucchiara, Andrea Prati, and Roberto Vezzani, *Real-time motion segmentation from moving cameras.*, Real-Time Imaging **10** (2004), no. 3, 127–143.
- [6] Marcos Carneiro de Andrade, *An interactive algorithm for image smoothing and segmentation*, ELCVIA **4** (2004), no. 1, 32–48.
- [7] J. B. Dias, K. P. Souza, and H. Pistori, *Conjunto de treinamento para algoritmos de reconhecimento de libras*, II Workshop de Visão Computacional, São Carlos (2006).
- [8] Djalma José Fagundes and Murched Omar Taha, *Modelo animal de doença: critérios de escolha e espécies de animais de uso corrente*, Acta Cirúrgica Brasileira **19** (2004), 59–65.
- [9] Ismail Haritaoglu, David Harwood, and Larry S. Davis, *W4: Who? when? where? what? a real time system for detecting and tracking people*, FG, 1998, pp. 222–227.
- [10] Allan Reffson Granja Lima, *Máquinas de vetores suporte na classificação de impressões digitais*, 2002.
- [11] Bruno Brandoli Machado, Jonathan de Andrade Silva, Wesley Nunes Gonçalves, Hemerson Pistori, and Albert Schiaveto de Souza, *Topolino: Software livre para automatização do experimento do campo aberto*, XV Seminário de Computação (2006), 19–28.

- [12] Alan McIvor, *Background subtraction techniques*, Proceedings of Image and Vision Computing, 2000.
- [13] J. Morrow-Tesch, J. W. Dailey, and H. JIang, *A video data base system for studying animal behavior*, Journal of Animal Science **76** (1998), no. 10, 2605–2608.
- [14] Richard A. Peters and Colin J. Davis, *Discriminating signal from noise: recognition of a movement-based animal display by artificial neural networks*, Behavioural Processes **1** (2006), 52–64.
- [15] Paul L. Rosin and Efstathios Ioannidis, *Evaluation of global image thresholding for change detection*, Pattern Recogn. Lett. **24** (2003), no. 14, 2345–2356.
- [16] Mehmet Sezgin and Bulent Sankur, *Survey over image thresholding techniques and quantitative performance evaluation*, Journal of Electronic Imaging **13** (2004), no. 1, 146–168.
- [17] Jean-Christophe Terrillon, Hideo Fukamachi, Shigeru Akamatsu, and Mahdad Nouri Shirazi, *Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images*, IEEE International Conference on Automatic Face and Gesture Recognition (Washington, DC, USA), IEEE Computer Society, 2000, pp. 54–63.
- [18] Dirk Wilking and Thomas Röfer, *Realtime object recognition using decision tree learning.*, RobuCup, 2004, pp. 556–563.