

6 Resultados

Neste capítulo apresentamos resultados das Galerias Inteligentes e de sua aplicação ao problema de posicionamento de câmera em cenas tridimensionais usando os descritores apresentados no capítulo 5. Analisamos sua capacidade de aprender e reproduzir as preferências do usuário em diferentes instâncias do problema.

Usamos como instâncias do problema modelos estáticos usuais de Computação Gráfica (ver figuras 6.1, 6.5 e 6.6); modelos de difícil compreensão, como nós tridimensionais (ver figuras 6.8 e 6.9); e cenas com grande potencial de oclusão (ver figuras 6.10 e 6.11).

Usamos como espaço de parâmetros Ω o espaço de câmera composto pela posição da câmera w^p , direção de visão w^d e orientação da câmera w^o . Em nossos experimentos, este espaço foi restrito de acordo com cada aplicação. Analisamos a seguir os resultados.

6.1 Capacidade de aprendizagem e reprodução de preferências

Reprodutibilidade em modelos similares Nosso experimento inicial tem como objetivo verificar se nosso método de aprendizagem combinado com os descritores desenvolvidos para o problema de posicionamento de câmera são capazes de reproduzir definições simples de boas visões. Usamos o modelo de uma vaca como instância de treino, e restringimos a posição da câmera a uma esfera centrada no centro do objeto, e a direção de visão de modo que o observador estivesse sempre visualizando o centro do objeto.

Navegamos pela galeria preferindo sempre visões laterais da vaca (ver figura 6.1). Em seguida, aplicamos o procedimento de seleção automática do melhor parâmetro em modelos de animais similares à vaca, na esperança de obter visões laterais, respeitando o treino.

Os resultados apresentaram-se satisfatórios em todos os casos, inclusive para o cavalo, que tem uma pose um pouco diferente dos demais modelos



Figura 6.1: Posicionamento automático de câmera usando um treino voltado para as visões laterais de uma vaca (esquerda): a máquina é capaz de reproduzir com sucesso visões laterais em outros animais.

(a cabeça não está alinhada com o corpo). Isso mostra que nosso método de aprendizagem é capaz de aprender as preferências subjetivas do usuário e imitar seu comportamento não somente na instância (modelo) treinada, mas também em outras instâncias similares. A escolha dos descritores, fundamental para o sucesso do método, também é validada neste caso: aplicando a análise dos descritores apresentada no capítulo 4, concluímos que os descritores mais relevantes neste caso foram a área projetada, que valoriza visões laterais, e as altas saliências, que evitam visões oblíquas (figura 6.2).

Respeito à subjetividade Gostaríamos também de testar se a máquina é capaz de escolher os parâmetros (visões) levando em conta preferências de diferentes usuários. Nesta análise, pedimos para dois profissionais usarem a galeria para escolher visões em um quadro de uma simulação de fluidos 3D. O primeiro profissional é um *designer* gráfico, que classificou visões do ponto de vista artístico. O segundo é um especialista em dinâmica de fluidos, que posicionou a câmera para visualizar as feições do fluido, gerando visões diferentes das visões do primeiro, como ilustra a figura 6.3.

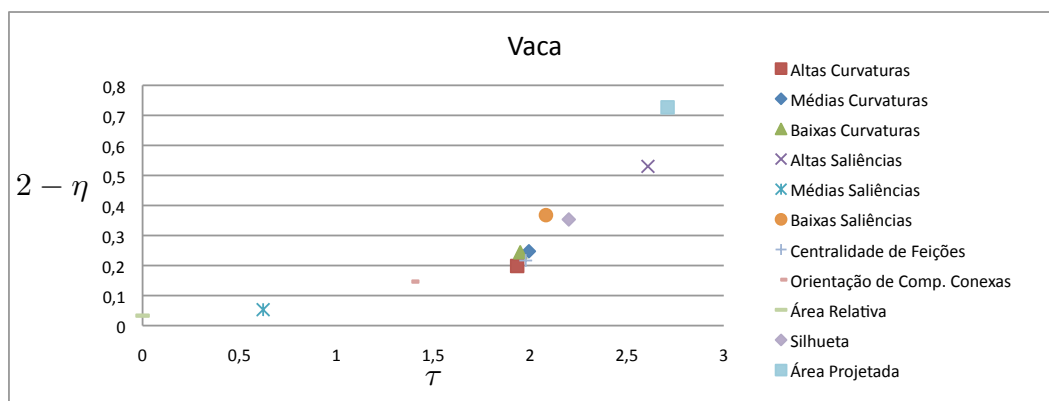


Figura 6.2: Análise da influência dos descritores no treino lateral da vaca (figura 6.1): a região superior direita representa os descritores mais influentes, onde destacam-se a área projetada e as altas saliências.

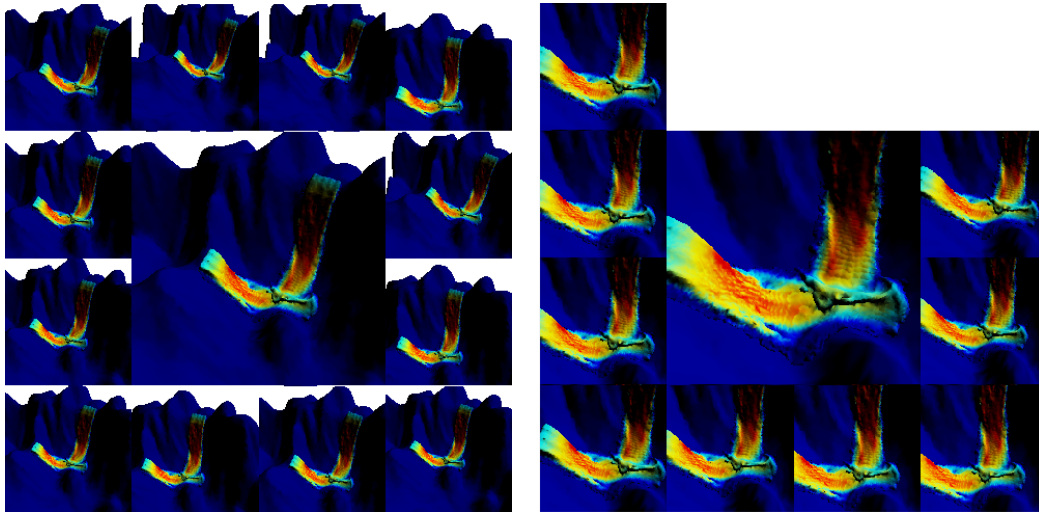


Figura 6.3: Resultado das seleções de visões de um *designer* gráfico (esquerda) e de um especialista em dinâmica de fluidos (direita).

Os bancos de dados Γ^1 e Γ^2 referentes aos treinos dos respectivos profissionais são usados para gerar duas funções classificadoras \hat{f}^1 e \hat{f}^2 . Em seguida, aplicamos o procedimento de seleção automática do melhor parâmetro em outro quadro da simulação usando cada uma das funções, e obtemos resultados similares às preferências de cada usuário, e conseqüentemente diferentes entre si, como exibido na figura 6.4.

Este resultado valida a capacidade da máquina de respeitar regras subjetivas, gerando resultados diferentes de acordo com as preferências individuais de cada usuário. Além disso, a precisão do método para posicionar a câmera em um quadro diferente da simulação, respeitando o treino em um quadro anterior, nos motiva a realizar experimentos para posicionar câmera em seqüências de quadros de video 3D.

Neste experimento, a posição da câmera w^p esteve restrita a um paralelepípedo acima do terreno, e a direção de visão w^d a um cone cujo eixo é perpendicular ao terreno.

Comparação A realização de comparações com outros métodos de posicionamento automático de câmera é delicada, visto que nesses casos são levadas em conta regras objetivas, enquanto nosso método é capaz de fornecer diferentes resultados de acordo com o treino do usuário.

Porém, dependendo do contexto, é possível encontrar boas visões canônicas de modelos, ou seja, visões que são indiscutivelmente ótimas, como a visão frontal da face em um modelo de uma cabeça. Seguindo esta linha, realizamos um treino duplo: primeiro navegamos na galeria selecionando visões

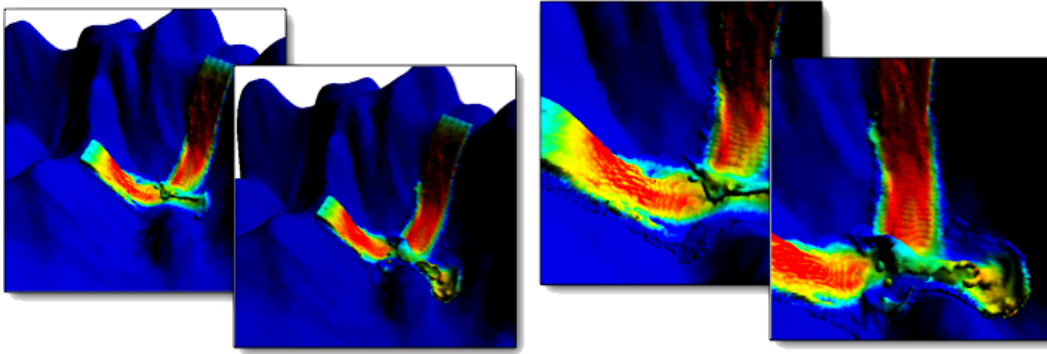


Figura 6.4: Comparação entre as melhores visões escolhidas pelos profissionais em um quadro da simulação de fluidos e as visões resultantes do procedimento de seleção automática em um outro quadro da simulação: a máquina imita com sucesso as preferências do *designer* gráfico (imagens à esquerda) e do especialista em dinâmica de fluidos (imagens à direita), selecionando visões semelhantes às selecionadas por cada profissional.

frontais do rosto do modelo da cabeça do David; e aumentamos a precisão do treino selecionando visões frontais do rosto do Max Planck, como mostra a figura 6.5. Neste caso, o banco de dados resultante é a união das seleções dos dois modelos.

Usando este treino, comparamos nosso método com três regras objetivas descritas em trabalhos recentes: maximização da área projetada (34), maximização do comprimento da silhueta (34) e maximização das saliências (23). Selecionamos alguns modelos com visões canônicas para testar a robustez dos métodos, e destacamos três resultados, exibidos na figura 6.6, onde as regras objetivas falharam em achar a melhor visão, enquanto nosso método teve sucesso, usando os dados do treino duplo.

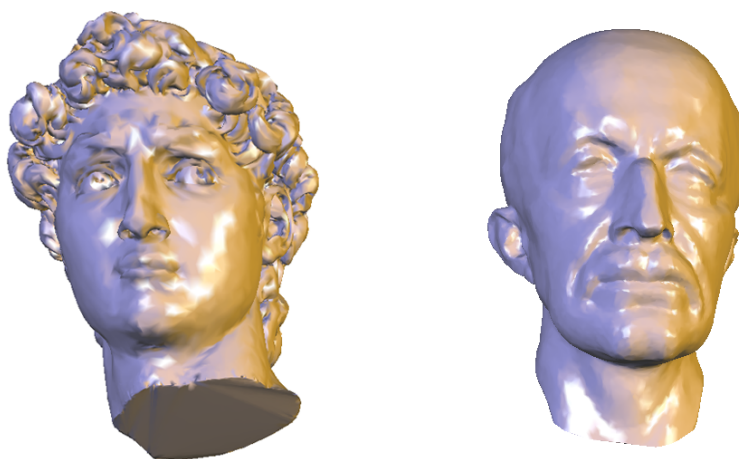


Figura 6.5: Visões preferidas em um treino iniciado na cabeça do David (esquerda) e completada com a cabeça do Max Planck (direita).

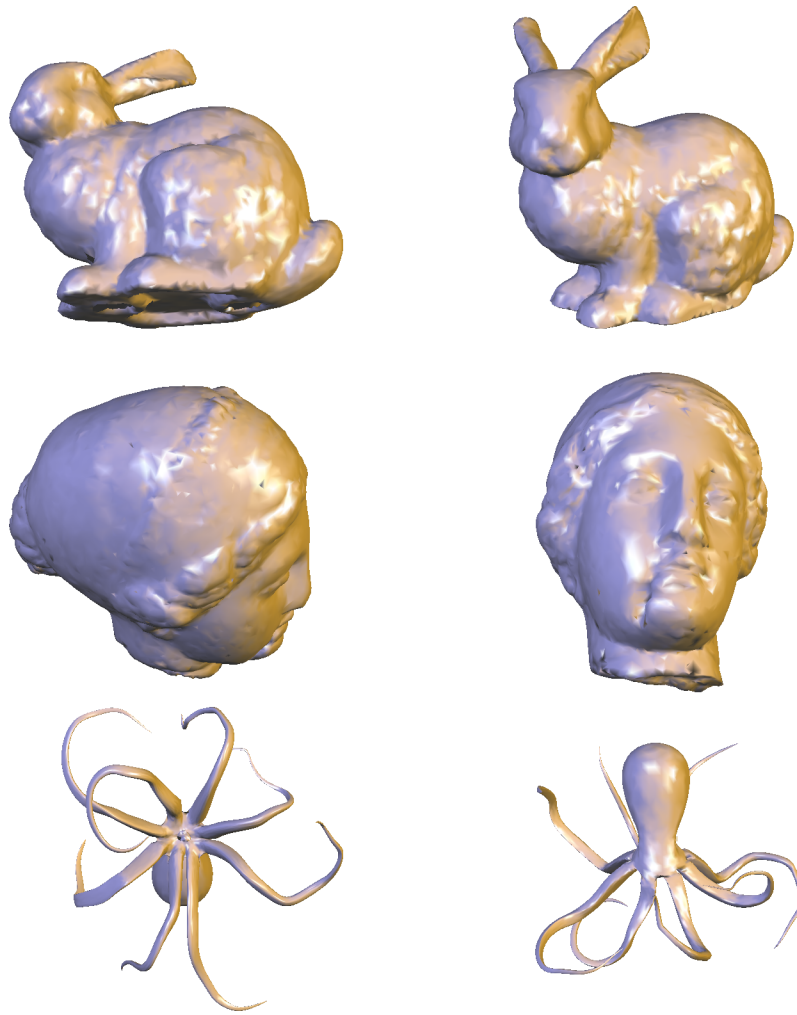


Figura 6.6: Comparação de nosso método de posicionamento automática de câmera (coluna da direita) usando os dados de treino da figura 6.5 com outros trabalhos recentes: maximização das saliências (23) (superior esquerda), maximização da área projetada (34) (meio esquerda) e maximização do comprimento da silhueta (34) (inferior esquerda).

Esses resultados ilustram bem as dificuldades do problema de posicionamento de câmera, onde é difícil definir um conjunto de regras objetivas que funcione bem em todos os casos. Além disso, a natureza subjetiva do problema impossibilita a definição de uma visão canônica em todos os modelos e aplicações. Portanto, trata-se de um problema onde as Galerias Inteligentes se apresentam como uma boa alternativa, aprendendo regras subjetivas através da combinação de diversos descritores, e mostrando-se robusta em uma maior coleção de modelos.

O resultado da análise de influência dos descritores neste treino duplo é exibida na figura 6.7, onde é possível identificar 6 descritores com boa influência, com destaque para curvaturas médias e baixas, e saliências médias. Este resultado nos leva a concluir que a combinação de vários descritores

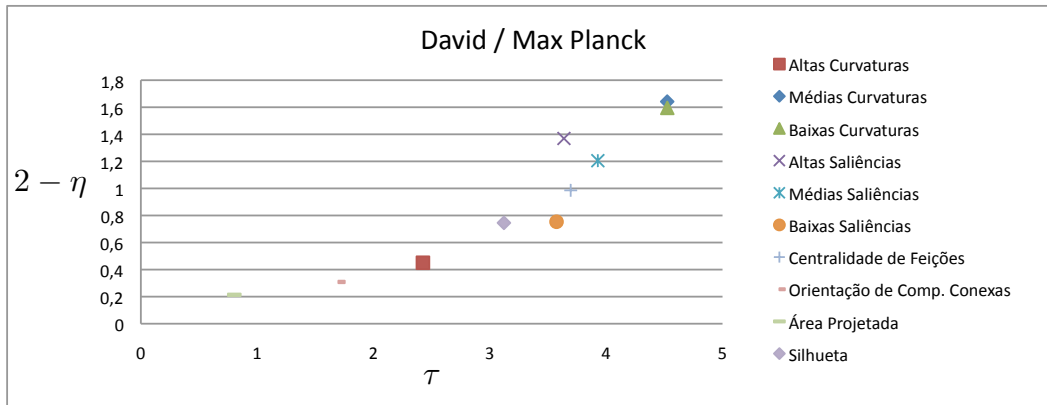


Figura 6.7: Análise da influência dos descritores no treino frontal duplo do David e do Max Planck (figura 6.5): seis descritores se destacam na região superior direita, com destaque para curvaturas médias e baixas, e saliências médias.

é fundamental para gerar uma função classificadora robusta para todos os modelos testados, como predito na conclusão do trabalho.

Neste experimento, o espaço de câmera foi o mesmo usado no experimento que avaliou a reprodutibilidade em modelos similares.

Estabilidade da função classificadora Examinamos também a estabilidade da aprendizagem realizando testes de precisão *out-of-sample*. Removemos aleatoriamente 20% das amostras de Γ dos dados de treinamento, geramos uma função classificadora sem essas amostras e testamos se a função classificadora é capaz de avaliar corretamente as amostras removidas. Realizamos 10 experimentos usando os dados de treino da vaca (figura 6.1), e 10 experimentos usando os dados do treino duplo (figura 6.5). As médias de classificações erradas são exibidas na tabela 6.1.

O baixo índice de erros nos leva a crer que nossa máquina de aprendizagem em combinação com o conjunto de descritores é estável do ponto de vista de não depender de um subconjunto pequeno de amostras de seu banco de dados Γ para gerar funções classificadoras robustas.

Instâncias difíceis Realizamos experimentos em cenas tridimensionais complexas, incluindo nós tridimensionais e cenas com grande potencial de oclusão.

Treino	Falsos positivos	Falsos negativos
vaca	5.7%	0.2%
duplo	2.7%	3.0%

Tabela 6.1: Estabilidade da função classificadora: baixo índice de erros.

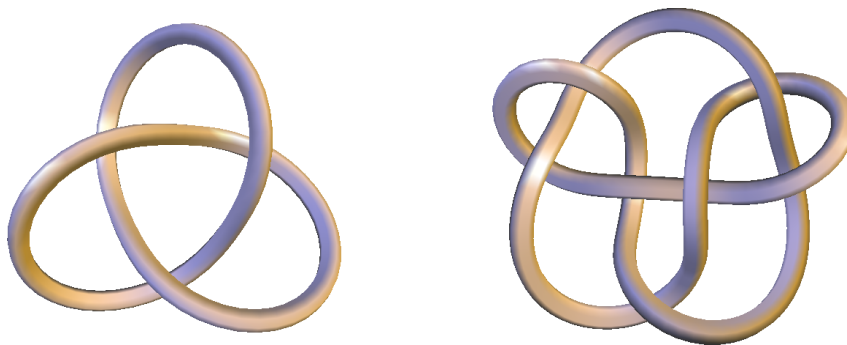


Figura 6.8: Treino do nó trifólio (esquerda) e posicionamento automático da câmera no nó 7.7c (direita): neste treino, foram selecionadas visões descritivas do nó, convergindo para a visão da esquerda. A melhor visão do nó 7.7c segundo a máquina reproduz a preferência do usuário por visões descritivas (direita).

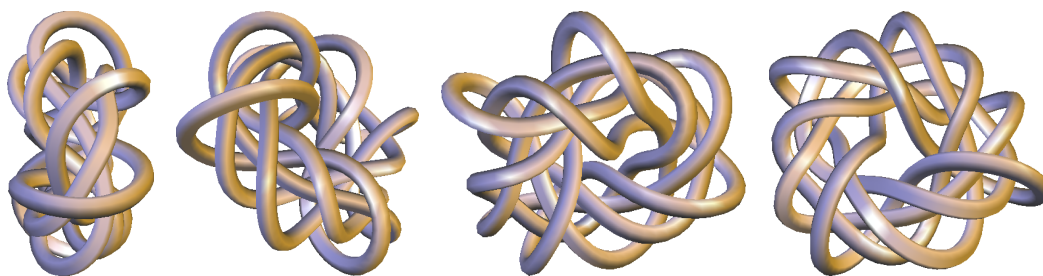


Figura 6.9: Classificação coerente de visões do nó 7.7c (ordem crescente de qualidade de visão da esquerda para direita).

A compreensão de nós tridimensionais por meio de imagens em geral é difícil, sendo necessário encontrar pontos de vista específicos que gerem imagens com boa descrição do nó. Realizamos um treino em um nó simples (trifólio), classificando como boas as visões mais descritivas do nó (figura 6.8), e pedimos para a máquina usar este treino para avaliar visões em nós mais complexos. Vale notar que o mergulho dos nós já resulta de uma otimização de forma (36) independente do uso das Galerias Inteligentes.

A figura 6.8 também exibe o melhor ponto de vista do nó 7.7c de acordo com a avaliação da máquina. Na figura 6.9 exibimos o resultado de uma ordenação de diferentes pontos de vista em um nó mais complexo, de acordo com a avaliação da máquina treinada no nó mais simples. É possível ver claramente que a máquina tem sucesso em classificar coerentemente as visões de acordo com as preferências do usuário, ou seja, visões mais descritivas do nó receberam melhor avaliação.

Neste experimento, o espaço de câmera foi o mesmo usado no experimento que avaliou a reprodutibilidade em modelos similares.

Para testar a capacidade do método em cenas com grande potencial de oclusão, criamos diversos modelos de florestas com árvores posicionadas

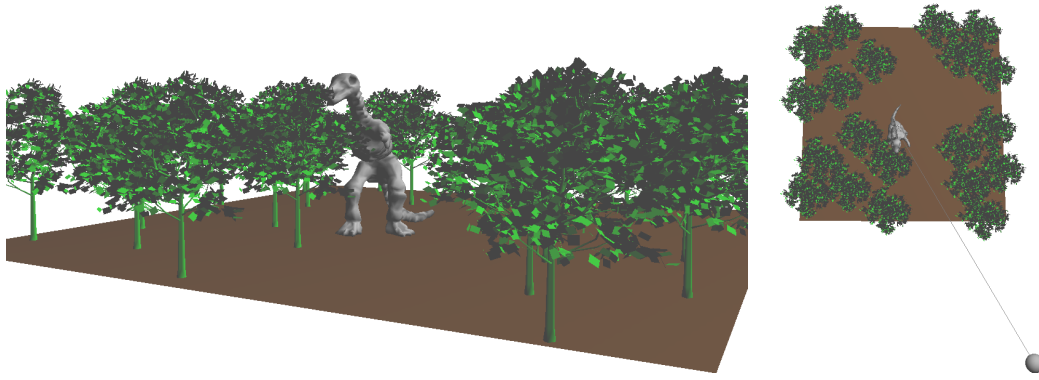


Figura 6.10: Treino em cena de floresta com grande potencial de oclusão, na tentativa de visualizar o dinossauro. Melhor visão do usuário (esquerda) e visão de cima da floresta, com a respectiva posição e direção da câmera (direita).

em posições aleatórias, e com um dinossauro posicionado em algum lugar aleatório da floresta. A posição da câmera w^p ficou restrita a um anel fora da floresta, contido no plano da floresta. A direção de visão w^d foi restrita a uma circunferência paralela ao plano da floresta.

Realizamos um treino em uma floresta com pouca densidade de árvores dando preferência aos pontos de vista onde seria possível visualizar a frente do dinossauro sem a oclusão das árvores, como ilustrado na figura 6.10. Usamos este treino para selecionar automaticamente pontos de vista em cenas com maior densidade de árvores. Os resultados podem ser observados na figura 6.11.

A máquina claramente tem sucesso em encontrar pontos de vista onde o dinossauro é visível, inclusive quando os pontos de vista que geram visões satisfatórias são escassos, como no exemplo inferior da figura 6.11. A análise de influência dos descritores revela grande influência do descritor de área relativa. A figura 6.12 exibe o gráfico deste descritor, revelando visualmente alta inclinação da reta e baixa variância.

6.2

Eficiência da interface das Galerias Inteligentes

Para analisar a eficiência da interface, realizamos uma pré-avaliação simples com usuários inexperientes, comparando a interface das Galerias Inteligentes com uma interface clássica, no caso o *Blender* com *trackball*. Apesar de não ser uma interface intuitiva, ela contém as ferramentas usuais de interação 3D. Uma avaliação mais adequada está ainda por ser feita mas saíria do escopo deste trabalho.

Inicialmente, os usuários escolhem um grupo de modelos para avaliar com ambas as interfaces, usando apenas uma das interfaces em cada modelo

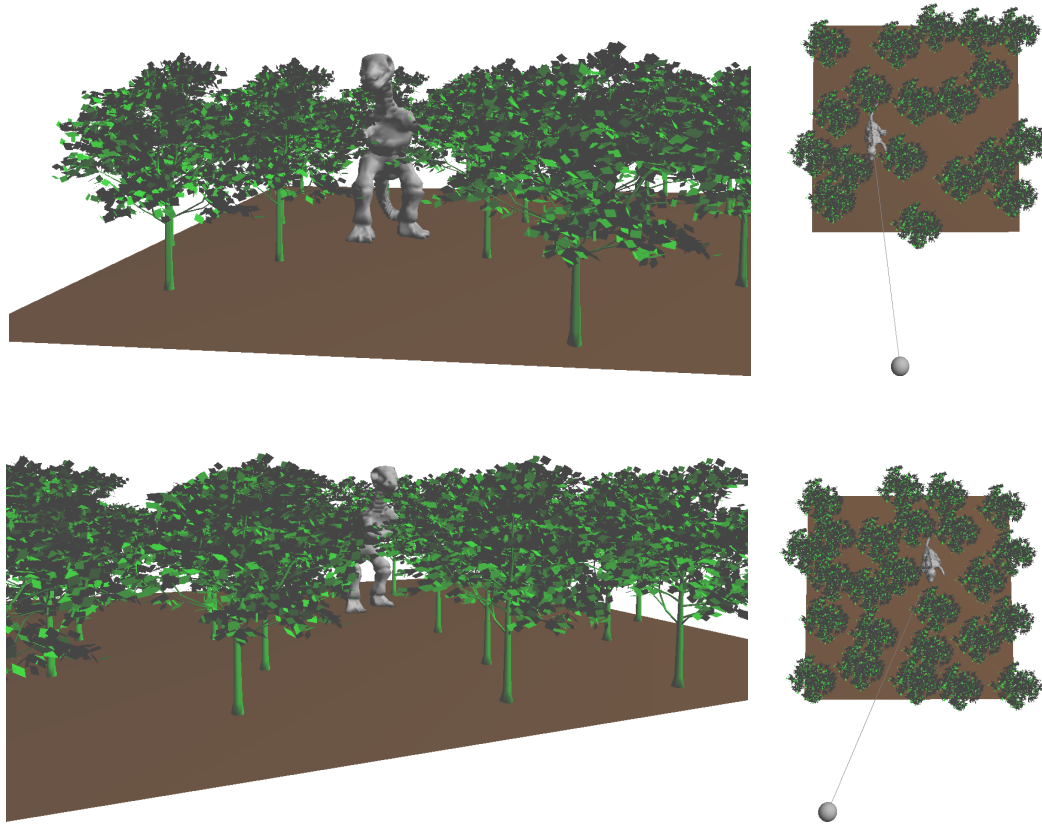


Figura 6.11: Resultados da seleção automática da melhor visão na cena da floresta: a máquina teve sucesso no posicionamento da câmera em outras florestas, gerando visões onde era possível visualizar o dinossauro.

do grupo. Usando as Galerias Inteligentes, o usuário deve navegar pelas galerias classificando visões como boas ou ruins e gerando novas galerias até chegar a uma visão final satisfatória. Ele pode também solicitar ajuda da máquina para selecionar automaticamente boas visões, e ordenar as galerias de acordo com a avaliação da máquina.

Nessa pré-avaliação simples, analisamos as curvas de aprendizagem e tempos dos usuários. As curvas de aprendizagem exibidas na figura 6.13 mostram que, após treinar dois ou três modelos, o tempo de interação do usuário com as Galerias Inteligentes diminuiu consideravelmente, enfatizando a capacidade de aprendizagem das Galerias Inteligentes.

Os tempos de interação estão diretamente relacionados com o acúmulo dos dados de treinamento, que possibilita a assistência das Galerias Inteligentes ao usuário, realizando seleções automáticas e ordenações nas galerias. Consequentemente, isto reduz o tempo que o usuário gasta para avaliar as visões de cada galeria. Por outro lado, os tempos de interação do usuário com o *Blender* permanecem praticamente constantes.

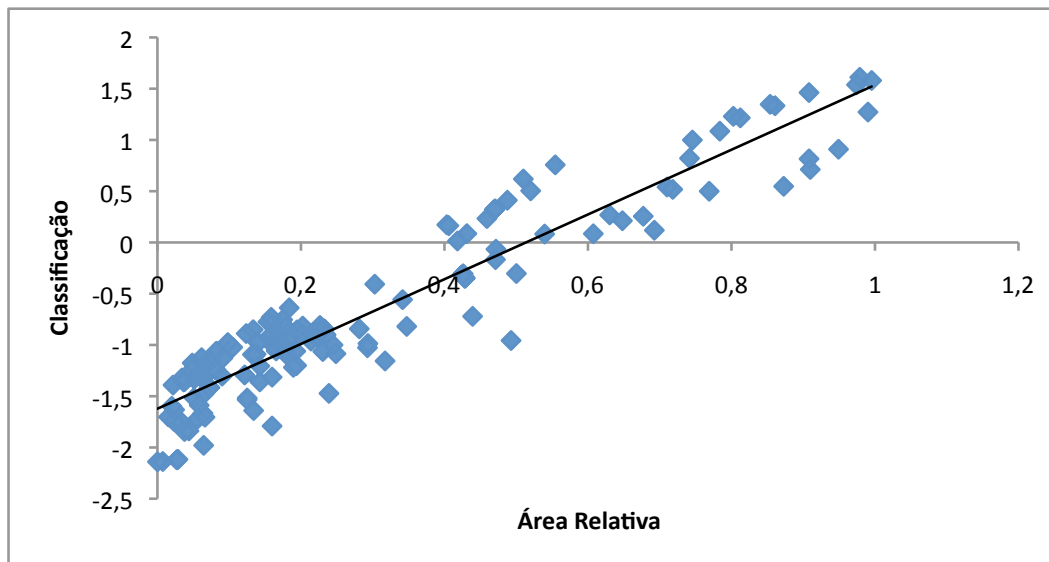


Figura 6.12: Regressão linear do gráfico do descritor de área relativa referente ao treino do dinossauro (figura 6.10): alta correlação linear.

6.3

Tempos de execução

A tabela 6.2 exibe dados de algumas das interações realizadas por usuários inexperientes com a interface das Galerias Inteligentes, incluindo o total de visões e galerias geradas, tempos de pré-processamento, processamento e de interação do usuário. Os tempos obtidos nas fases de processamento nos mostram que a interação com as Galerias Inteligentes foi realizada em tempo *quasi-real*.

Como boa parte do processamento intrínseco dos modelos é realizada em fase de pré-processamento, sua complexidade não afeta consideravelmente os tempos de processamento durante a interação do usuário. O cálculo da função de classificação do SVM e a avaliação de parâmetros também têm custo computacional desprezível, visto que os bancos de dados compostos por parâmetros avaliados por usuários são da ordem de algumas dezenas ou centenas de parâmetros. Conseqüentemente, o maior gargalo das Galerias Inteligentes está na fase de extração de descritores, mais especificamente na avaliação de cada imagem, por cada descritor. Neste caso, a resolução da imagem influencia fortemente o tempo de execução, visto que boa parte dos descritores deve varrer todos os pixels da imagem.

Além disso, descritores mais elaborados, como o que mede a complexidade da silhueta, podem ser adaptados para processamento em GPU. Em particular, podemos avaliar a complexidade da silhueta no espaço do objeto, extraíndo sua silhueta 3D com tempo de execução menor.

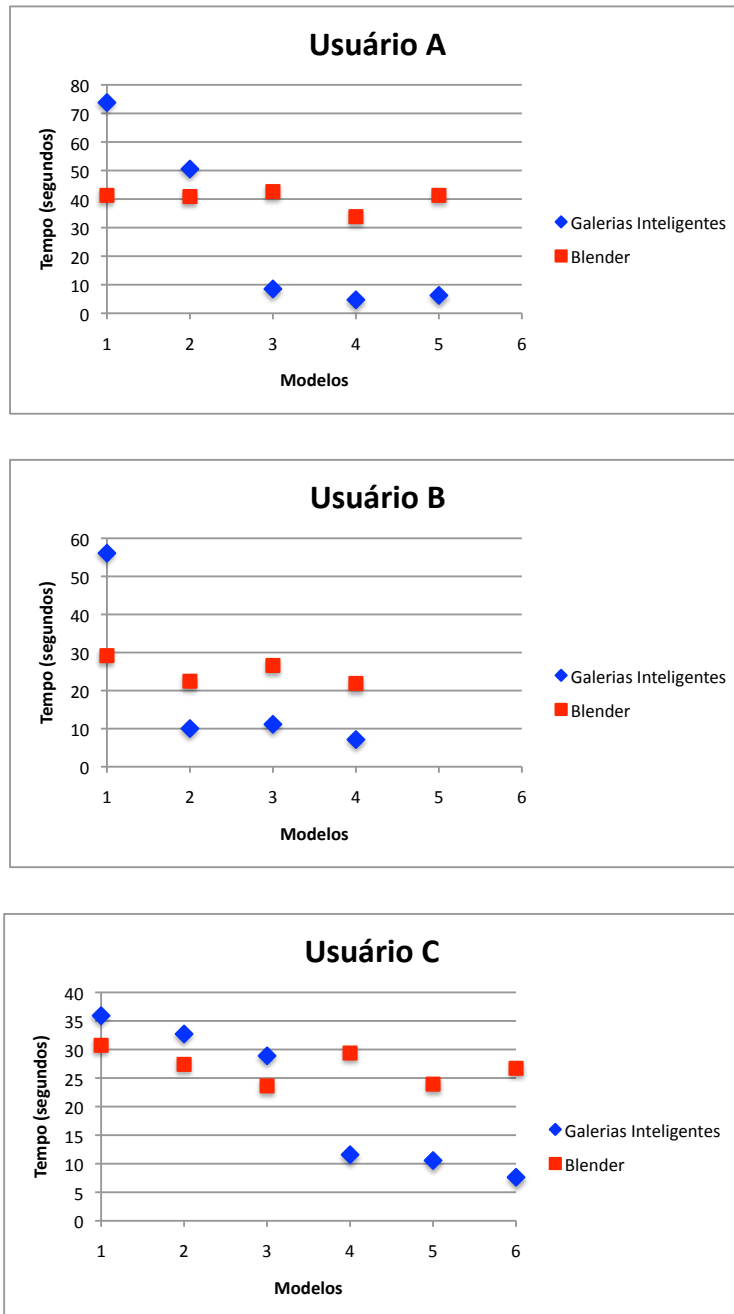


Figura 6.13: Curvas de aprendizagem de três usuários: o tempo de interação com a interface das Galerias Inteligentes diminui com o tempo, em contraste com interfaces convencionais, onde o tempo de interação permanece constante.

Finalmente, considerando que os tempos obtidos foram consequência de um processamento sequencial de parâmetros e descritores, estimamos que para atingirmos tempos interativos, pode ser suficiente paralelizar a avaliação dos descritores e dos parâmetros.

Modelo	# de visões	# de galerias	Pré-Proc. seg.	Proc. seg.	Interação seg.
vaca (0)	60	4	2.140	8.280	73.8
cão (1)	44	2	4.218	6.595	50.
gato (2)	20	0	8.875	3.281	8.5
dinossauro (3)	20	0	17.172	3.469	4.7
pantera (4)	20	0	4.703	3.234	6.3
nó 3.1c (0)	36	2	6.890	4.671	56.1
nó 7.7c (1)	20	0	6.672	3.235	10.0
nó 6.13c (2)	20	0	3.375	3.250	11.1
nó 138k (3)	20	0	3.735	3.250	7.1
triceratops (0)	28	1	2.313	4.906	35.9
dragão (1)	24	1	21.328	3.250	32.7
polvo (2)	24	1	9.531	4.375	28.9
mosquito (3)	20	0	5.141	3.234	11.6
mulher (4)	20	0	4.359	3.297	10.6
esquilo (5)	20	0	5.281	3.328	7.6

Tabela 6.2: Eficiência da interface das Galerias Inteligentes. As linhas horizontais dividem treinos de diferentes usuários.