Universidade Federal de Pernambuco – UFPE

Centro de Informática - CIn

Detecção de Objetos Usando Características de Histograma Espacial

João Paulo Magalhães

Trabalho de Graduação do Curso de Engenharia da Computação

Orientador: Tsang Ing Ren

Recife, 24 de janeiro de 2008.

João Paulo Magalhães

Detecção de Objetos Usando Características de Histograma Espacial

Trabalho de Graduação apresentado como parte das atividades para obtenção do título de Bacharel, do curso de Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco.

Professor Orientador: Tsang Ing Ren

Recife, 2008.

ERRATA

Folha	Linha	Onde se lê	Leia-se

Dedico este trabalho e, por conseguinte, toda a minha formação a meu pai (in memoriam), pela minha capacidade de sonhar, a minha mãe, pela minha força e capacidade de realizar, a meu irmão, pelo exemplo e incentivo, e a Michele, pela minha capacidade de amar.

Dedico esta conquista principalmente a todos aqueles que nunca tiveram o direito de sonhar e a oportunidade de fazer algo por si próprios.

AGRADECIMENTOS

Agradeço primeiramente a todos os responsáveis pelo que sou hoje: a maravilhosa e grande família, aos meus amigos e pessoas que já fizeram parte da minha vida e a todos os meus professores e mestres.

Agradeço a todos os que me ajudaram na construção deste trabalho: professores e amigos engenheiros, cientistas e do trabalho.

Agradeço sobretudo a Deus, por tornar tudo possível.

A obra interior consiste em que o aluno se converta na matéria-prima de uma criação, de uma realização formal, que termina no domínio da arte escolhida.

Eugen Herrigel

RESUMO

Na comunidade de visão computacional, detecção de objetos tem sido um tópico de pesquisa muito desafiante [1]. O objetivo maior e mais ambicioso do entendimento geral de imagens é reconhecer cada um dos objetos em uma imagem [2]. Além disso, grandes áreas de aplicação, tais como robótica e engenhos de busca, procuram soluções para seus problemas de recuperação de imagem baseada em seu conteúdo (c*ontent-based image retrieval*).

Grandes empresas como o Google©, por exemplo, tentam passar da simples busca baseada em *tags* ou *labels*, que analisa *links* que endereçam as imagens, o texto adjacente as mesmas ou em suas legendas, para uma busca mais avançada baseada no conteúdo das imagens buscadas. O próximo grande avanço em busca de imagens será a habilidade de busca baseada em conceitos visuais.

O problema de detecção de objetos pode ser sumarizado da seguinte forma: dada uma classe de interesse T (tal como pedestres, faces humanas, construções, carros ou textos) e uma imagem P, detecção de objetos é o processo para determinar se existem instâncias de T em P e, se existem, retornar os locais onde as instâncias de T são encontradas na imagem P [1].

Vários métodos têm sido propostos para a detecção de objetos em imagens. Para isso, estes métodos freqüentemente utilizam duas formas de modelagem: de baixo-nível, baseada em características como cor e textura, e semântica, baseada no conteúdo da imagem. Em geral, os métodos baseados na modelagem de baixo-nível podem ser classificados em duas categorias: métodos baseados na aparência global e métodos baseados em componentes.

Zhang et al. [1] propõem características baseadas em histograma espacial para representar objetos, conseguindo assim codificar simultaneamente informações sobre textura e forma de um objeto. Este tipo de características é específico de uma classe de objetos, possibilitando assim discriminar se um objeto pertence ou não a uma determinada classe.

A proposta deste trabalho é estudar métodos de detecção de objetos, em particular o método proposto por Zhang et al. [1] baseado em características de histograma espacial. Além disso, pretende-se apresentar uma visão geral da visão computacional, incluindo uma maior atenção à análise de imagens com ênfase na busca de imagens baseada em seu conteúdo e na detecção de objetos. Fornecendo assim uma visão contextualizada e clara da área juntamente com um eficiente método de detecção de objetos.

Palavras-chave: visão computacional, detecção de objetos, histograma espacial, representação de imagens, aprendizagem de máquina.

SUMÁRIO

INTRODUÇÃO

Uma imagem vale mais que dez mil palavras.
Autor desconhecido

A visão computacional tem sido uma área de grande interesse no meio científico principalmente devido aos desafios e potencialidades desta área. É de grande interesse para o homem prover uma máquina de uma das mais poderosas ferramentas da natureza, a visão. Desta forma, não é mais suficiente capturar e simplesmente processar as imagens em um baixo nível, gerando transformações básicas nas mesmas, é necessário agora construir sistemas completos quanto ao entendimento do conteúdo das imagens. Neste contexto, a análise de imagens e, por conseguinte, a busca em seu conteúdo e a posterior detecção de objetos são peças chaves na construção de aplicações mais próximas a visão humana.

O desenvolvimento de métodos para a detecção automática de objetos em imagens tem sido um tópico de grande interesse nas áreas de pesquisa de visão computacional e análise de padrões [8]. H. Zhang et al. [1] sumariza o problema de detecção de objetos em imagens da seguinte forma: dada uma classe de objetos de interesse T (tais como pedestres, faces humanas, construções, carros ou textos) e uma imagem P, detecção de objetos é o processo de determinar se há instâncias de T em P e, se existem, retornar as localizações onde as instâncias de T se encontram na imagem P.

A principal dificuldade da detecção de objetos é causada pela grande variabilidade na aparência entre objetos da mesma classe [1]. Diferentes objetos pertencentes a mesma categoria têm freqüentemente grandes variações na aparência [8]. Além disso, o mesmo objeto pode aparecer de formas bastante diversas sob diferentes condições, tais como aquelas resultantes de mudanças na iluminação, ângulo, foco e técnicas de captura de imagem [9]. Desta forma, um método de detecção de objetos obterá sucesso caso seja capaz de representar imagens de uma maneira que as modele de forma invariante a variações próprias da classe,

mas que, ao mesmo tempo, seja capaz de distinguir imagens da classe de objetos desejada de todas as outras imagens que não pertencem àquela classe.

Muitos métodos têm sido propostos para a detecção de objetos em imagens. Em muitos deles, o problema é resolvido com um *framework* de conhecimento estático onde amostras de imagens são representadas por um conjunto de características e então métodos de aprendizado são usados para identificar objetos da classe de interesse. Uma grande variabilidade tanto nos tipos de característica usadas quanto nos métodos de aprendizado aplicados têm sido considerados. No entanto, os métodos mais recentes propostos e que merecem maior destaque podem ser classificados em duas categorias: métodos baseados na aparência global e métodos baseados em componentes [1].

Métodos baseados na aparência global consideram um objeto como uma unidade única e então realiza a classificação nas características extraídas do objeto inteiro. Muitos mecanismos de aprendizado estatístico são explorados para caracterizar e identificar padrões de objetos. Rowley et al. [10] e Garcia e Delakis [11] usam redes neurais artificiais como métodos de classificação para a detecção de faces. Osuna et al. [12] e Papageprgiou e Poggio [13] baseiam suas características em *wavelets* adotando *Support Vector Machine* (SVM) para localizar carros e faces humanas. Em outros trabalhos, métodos de aprendizagem supervisionada têm sido aplicados para detectar faces frontais por Viola e Jones [14], e então estendidos para detectar faces sob múltiplas visões por Li et al. [15] e para detecção de textos por Chen e Yuille[16]. Outros métodos de aprendizado usados para detecção de objetos incluem distribuição de probabilidade [17, 18] e análise dos componentes principais (PCA) [19] entre outros.

Métodos baseados em componentes tratam um objeto como uma coleção de partes da imagem. Estes métodos primeiro extraem alguns componentes do objeto e então detectam objetos usando, na maioria das vezes, informações geométricas. Mohan et al. [20] propõem um método de detecção de objetos baseado em componentes. Neste método, uma pessoa é representada por componentes tais como cabeça, braços e pernas e então um classificador SVM é aplicado para detectar estes componentes e decidir quando uma pessoa está presente. Naquest e Ullman [21] usam fragmentos como características e realizam reconhecimento de objetos com informações características e classificação linear.

Dentro deste contexto, destaca-se Agarwal et al. [8]. Em seu trabalho, um vocabulário de partes é extraído automaticamente por meio da aplicação do operador de Forstner [22][23], que busca por pontos de interesse – pontos em uma imagem que têm alto conteúdo informativo em termos de mudanças locais do sinal, tais como pontos de intersecção de linhas e centros de padrões circulares. Esta operação é feita em um conjunto de imagens representativas da classe de objetos estudada, visões laterais de carros em seu trabalho. Este vocabulário é então usado para representar os objetos tanto na etapa de aprendizado supervisionado que gera um classificador discriminante para o conjunto de imagens de treinamento quanto para a fase de detecção da imagem como membro ou não da classe de interesse da detecção com algum grau de segurança.

Uma tarefa visual relacionada à detecção de objetos é o reconhecimento de objetos, cujo objetivo é identificar instâncias específicas de um objeto em imagens. Métodos baseados em descritores locais são atualmente largamente usados na detecção de objetos. Schiele [25] propôs usar derivadas Gaussianas como características locais objetivando criar um histograma multidimensional como representação de um objeto, e então realizar o reconhecimento de muitos objetos em 3 dimensões. Lowe [26] desenvolveu um sistema para reconhecimento de objetos que utiliza descritores SIFT, do inglês *Scale-Invariant Feature Transform*, baseados em histogramas de orientação local. Mesmo assim, estes métodos são projetados para reconhecer um objeto específico, e não generalizar para categorizar classes de objetos.

A extração de características para a representação de objetos é uma das fases mais importantes na automação de sistemas de detecção de objetos. Métodos anteriores usam uma grande variedade de representações para a extração de características dos objetos, tais como intensidade dos pixels [10, 27], margens [28], *wavelets* [29] e LBP [30], do inglês *Local Binnary Pattern*, entre outras.

Motivados pela observação que objetos têm distribuição de textura e configuração de formas, Zhang et al. [1] propõem características baseadas em histogramas espaciais, chamadas características de histograma espacial ou SHF, do inglês *Spatial Histogram Features*, para representar objetos. Como histogramas espaciais consistem de distribuições marginais de uma imagem sobre posições locais, a informação sobre textura e forma dos objetos podem ser armazenadas simultaneamente.

Em contraste com a maioria das características citadas acima, características de histograma espacial são específicas para cada classe de objetos, podendo ser usadas para discriminação entre objetos pertencentes ou não a classe de interesse num processo que requer um baixo custo computacional. Além disso, trabalho anterior também de H. Zhang et al. [31] mostra que características de histograma espacial são efetivas e eficientes para detectar faces humanas em imagens coloridas.

Desta forma, baseado na representação de objetos usando características de histograma espacial, nós apresentaremos o método proposto por Zhang et al. [1], que detecta objetos utilizando uma estratégia de refinamento em duas fases. Um detector hierárquico utiliza-se de combinação de histograma em cascata e SVM para detectar objetos e treinar um classificador durante o processo de aprendizado que consiste na aplicação do critério de Fisher para medir discriminabilidade para cada característica de histograma espacial seguida do cálculo da correlação entre estas características. Assim, um método de treinamento para a combinação de histograma em cascata por meio da seleção automática de características discriminativas é também apresentado.

Além disso, com o objetivo de construir um sistema que é facilmente extensível para diferentes classes de objetos, é de suma importância que a seleção das características tanto na etapa de treinamento como na fase de detecção seja um procedimento automatizado. O método proposto pode ser utilizado para qualquer classe de objetos após passar por um processo de treinamento baseado em bases de dados com imagens pertencentes e não pertencentes à classe, mas nosso trabalho limita-se a detecção de laterais de carros (ver figura 1 abaixo).

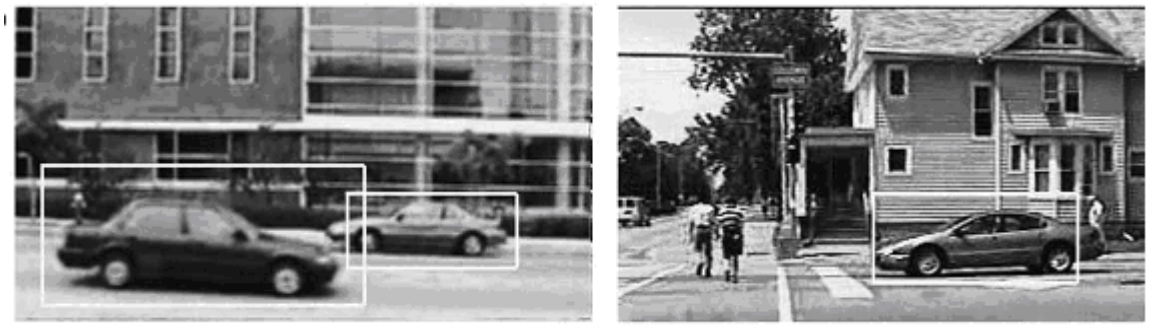

Figura 1: Exemplo de resultados da detecção de carros

Este estudo refere-se então a uma visão geral da área de visão computacional seguida de uma explanação sobre a análise de imagens, focando-se em detecção de objetos e em busca em imagens pelo conteúdo. Especificamente, objetiva-se o estudo e a implementação do método de detecção de objetos usando características de histograma espacial (SHF – *Spatial Histogram Features*) proposto por Zhang et al. [1].

Este trabalho se organiza da seguinte maneira: no primeiro capítulo vemos uma visão geral da área de visão computacional e da análise de objetos, no segundo capítulo vemos o método de detecção de objetos baseado em histogramas espaciais, no terceiro capítulo relatamos nossos experimentos e resultados e, por fim, apresentamos nossas considerações finais a respeito de todo o trabalho desenvolvido, dificuldades encontradas e indicações de próximos trabalhos.

1 VISÃO COMPUTACIONAL

As coisas que vemos não são em sim o que vemos... Continua completamente desconhecido para nós o que objetos podem ser em si mesmos e separados da receptividade dos nossos sentidos. Nada sabemos exceto o nosso modo de percebê-los...

Immanuel Kant

1.1 Visão Geral

Visão computacional é o termo utilizado em ciência e tecnologia para designar a área de pesquisas dedicada aos estudos relacionados a prover uma máquina da capacidade de visão, presente no homem e em muitos outros animais. Como uma disciplina científica, visão computacional dedica-se à teoria da construção de sistemas artificiais que obtenham informações do conteúdo de imagens. Como uma disciplina tecnológica, visão computacional procura aplicar as teorias e modelos da disciplina científica na tentativa de construir sistemas de visão computacional. Exemplos de aplicações de sistemas de visão computacional incluem sistemas para:

- Controle e automação de processos (um robô industrial ou veículo autônomo, por exemplo);
- Detecção de eventos (para vigilância visual ou controle de ambientes);
- Busca, organização e análise de informações (indexação de base de dados, análise de imagens médicas ou modelagem topográfica, por exemplo);
- Interação (como entrada para um dispositivo para interação homem-máquina);
- Próteses (em dispositivos que substituam ou melhorem a visão de humanos);

Visão computacional também pode ser descrita em termos de complementação, e não de contraposição, da visão biológica. Em visão biológica, a percepção visual de humanos e vários animais é estudada, resultando em modelos de como estes sistemas funcionam em termos de processos neuro-fisiológicos. Visão computacional, por outro lado, estuda e descreve sistemas de visão artificial que são implementados por software e/ou hardware com o objetivo de desempenhar a mesma função, ou pelo menos se assemelhar, da visão biológica, a saber, prover informações visuais de forma completa e contextualizada a respeito do conteúdo das imagens. Desta forma, trocas interdisciplinares entre as áreas de estudo de visão biológica e computacional têm sido proveitosas para ambas.

1.1.1 Estado da Arte

O campo de visão computacional pode ser caracterizado como uma área jovem e diversa. Apesar da existência de trabalhos anteriores, somente a partir da década de 70 se iniciaram estudos mais focados nesta área, computadores foram usados para gerenciar o processamento de grandes conjuntos de dados de imagens. De qualquer forma, estes estudos normalmente se originaram de vários outros campos e, conseqüentemente, não havia formulação padrão da problemática da visão computacional.

Esse problema persiste, em menor escala, até os dias atuais. Além disso, e em uma dimensão maior, não há uma formulação padrão geral de como os problemas relacionados a visão computacional devem ser resolvidos. Em vez disso, existe uma abundância de métodos baseados em formulações heurísticas projetados para resolver problemas específicos e bem definidos, onde os métodos freqüentemente são muito específicos das tarefas para as quais foram formulados e raramente podem ser generalizados para uma larga gama de aplicações.

Muitos dos métodos e aplicações estão ainda no estado de pesquisas básicas, mas uma quantidade considerável de métodos têm encontrado seu lugar dentro de produtos comerciais, onde eles normalmente constituem uma parte de um grande sistema, como na área médica, de controle de qualidade ou processos industriais, por exemplo. Na maior parte das aplicações comerciais de visão computacional, os computadores são pré-programados para resolver tarefas particulares, não aproveitando assim as vantagens de métodos baseados em

aprendizagem, que ainda enfrentam a desconfiança e incredibilidade dos profissionais de outras áreas apesar de há algum tempo eles estarem se tornando práticos e confiáveis.

1.1.2 Áreas relacionadas

Várias áreas estão relacionadas a visão computacional e em vários níveis de relacionamento, desde as que servem de base para o estudo da visão computacional até as que utilizam-se de suas técnicas e resultados, passando pelas que são estudadas conjuntamente e/ou deram origem a mesma. A importância de todas elas deve-se ao fato de que vários métodos e pesquisas dessas áreas podem ser utilizados conjuntamente além do que o entendimento delas pode ajudar no progresso e entendimento da visão computacional.

Uma significante área da inteligência artificial lida com planejamento e tomada de decisões autônomas para sistemas que podem realizar ações mecânicas tais como mover um robô em um ambiente real. Este tipo de processamento tipicamente necessita de dados de entrada providos por um sistema de visão computacional, agindo como um dispositivo de visão e provendo informação em alto-nível sobre o ambiente e o próprio robô.

Outras áreas que freqüentemente são descritas como pertencentes a inteligência artificial e que têm técnicas utilizadas conjuntamente com a visão computacional são as áreas de aprendizagem de máquina e reconhecimento de padrões. Como conseqüência, visão computacional é também vista como uma área no campo de inteligência artificial ou da ciência da computação em geral.

Física é outra área que é fortemente relacionada com visão computacional. Uma parte da visão computacional lida com métodos que requerem um completo entendimento dos processos em que a radiação eletromagnética, principalmente nos espectros visível e infra-vermelho, é refletida pelas superfícies de objetos e finalmente é medida por um sensor de iluminação para produzir uma imagem digital. Este processo é baseado em ótica e física do estado sólido. Sensores de imagem mais sofisticados têm sido buscados e requerem ainda mecânica quântica para prover uma compreensão completa do processo de formação da imagem. Além disso, vários problemas da Física podem ser resolvidos usando visão computacional, tais como os problemas de mensuração em movimento dos fluídos, etc.

Conseqüentemente, visão computacional pode também ser visto como uma área de atuação da física.

Um terceiro campo que tem grande importância é o da neuro-biologia, especialmente o estudo dos sistemas de visão biológica. O processamento do estimulo visual em humanos e em várias espécies de animais vem sendo extensivamente estudado através das partes que o compõe: olhos, neurônios e estruturas ou comportamento cerebral. Isto tem gerado uma descrição, apesar de primária, já complexa de como os sistemas reais de visão operam para resolver as tarefas relacionadas à visão. Estes resultados têm levado-nos a um sub-campo dentro da área de visão computacional onde sistemas artificiais são projetados para se assemelhar ao processamento e comportamento de sistemas biológicos, em diferentes níveis de complexidade. Além disso, alguns dos métodos baseados em aprendizado desenvolvidos dentro da visão computacional têm sua inspiração na biologia.

Outro campo relacionado a visão computacional é o de processamento de sinais. Muitos métodos para processamento de sinais de uma variável, tipicamente sinais dependentes do tempo, podem e têm sido expandidos de forma natural para processamento de sinais de duas ou mais variáveis para possibilitar a sua aplicação a imagens. Uma característica destes métodos é que eles são não-lineares, fazendo com que, junto com a multi-dimensionalidade dos sinais, definam um sub-campo em processamento de sinais como uma parte ou de interesse direto da visão computacional.

Além dos campos citados acima, muitos dos tópicos de pesquisa podem também ser estudados de um ponto de vista puramente matemático. Por exemplo, muitos métodos são baseados em análise estatística, métodos de otimização ou análise geométrica. Além disso, uma parte significativa desta área está voltada para o aspecto de aplicabilidade dos métodos: como métodos existentes podem ser implementados em várias combinações de hardware e software; ou como estes métodos podem ser modificados para incrementar velocidade de processamento sem perda de performance.

Finalmente, reconhecimento de padrões é um campo que usa vários métodos para extrair informação de sinais em geral, principalmente baseados em métodos estatísticos. Uma parte significante deste campo é dedicado a aplicar estes métodos a imagens.

Os campos mais próximos à visão computacional são processamento e análise de imagens e visão de robôs e máquinas. Há uma grande sobreposição em termos de técnicas e aplicações destas áreas. Isto significa que as técnicas básicas que são usadas e desenvolvidas são mais ou menos similares. Por outro lado, parece necessário para grupos de pesquisa, jornais científicos, conferências e companhias se apresentar ao mercado como pertencente especificamente a um destes campos e, desta forma, várias caracterizações que distinguem cada um dos campos dos outros têm sido apresentadas.

As seguintes caracterizações são relevantes mas não devem ser consideradas como de aceitação universal: processamento e análise de imagens tem o foco em transformações de imagens de duas dimensões, tais como incremento do contraste, extração de curvas, redução do ruido, rotação, etc (esta definição implica que análise e processamento de imagens não requer suposições e nem produz interpretações sobre o conteúdo das imagens); visão computacional baseia-se fortemente no conteúdo das imagens e no processo cognitivo da visão humana, buscando assim o conhecimento através da percepção e muitas vezes contando com suposições mais ou menos complexas para descrever as cenas em uma imagem; visão de máquina tem seu foco em aplicações principalmente na indústria e seu processamento em tempo real é enfatizado por meio de implementações eficientes em hardware e software.

Abaixo, segue uma pequena ilustração da área de visão computacional. Logicamente, outras inúmeras áreas se relacionam com a visão computacional, sendo esta uma representação ilustrativa, não quer demonstrar assim nenhuma informação hierárquica ou estrutural das áreas envolvidos.

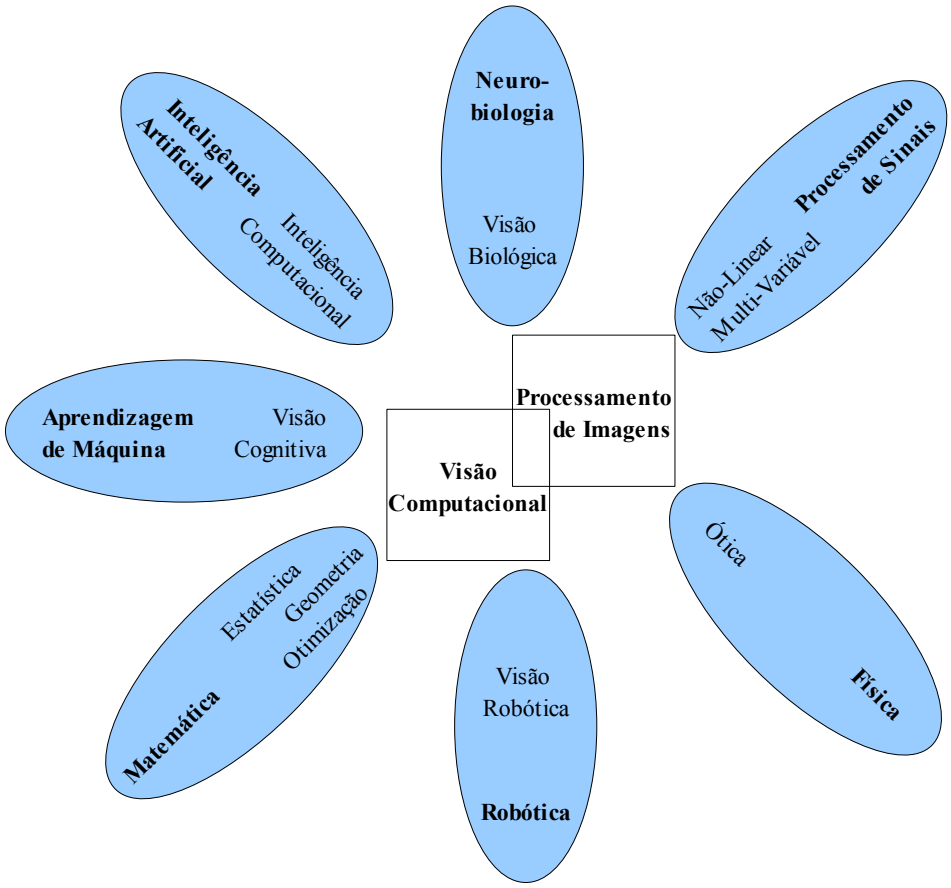

Ilustração 1: Visão geral da área de visão computacional

1.1.3 Aplicações

Um dos mais promissores campos de aplicação da visão computacional é a área médica, tanto com a visão computacional médica quanto com o processamento de imagens médicas. Esta área caracteriza-se principalmente pela extração de informação de imagens do corpo humano com o propósito de diagnosticar patologias. Geralmente, as imagens são extraídas através de microscopia, raios-X, angiografias, ultra-sonografias ou tomografias. Como exemplo de informações que podem ser extraídas de tais imagens pode-se citar a detecção de tumores, arteriosclerose ou outras alterações malignas. Pode-se também obter medidas de dimensões orgânicas, fluxo sanguíneo, etc.

Uma segunda área de aplicação da visão computacional é a industria. Nela, informações são extraídas com o propósito de atender às necessidades dos processos de produção. Um

exemplo é controle de qualidade, no qual detalhes dos componentes ou produtos finais são automaticamente inspecionados para encontrar defeitos de produção. Outra aplicação é a medida de posição e orientação de detalhes com o objetivo de guiar um braço automatizado.

Aplicações militares são provavelmente uma das maiores áreas para a visão computacional. Como exemplos podemos citar a detecção de soldados ou veículos inimigos em campos de batalha, mísseis auto-guiados e processamento de informações para orientar ações estratégicas. Sistemas muito avançados para orientação de mísseis guiam-o possibilitando a seleção de alvos baseada em informações de imagens adquiridas, tratadas e processadas pelo próprio míssil localmente e em tempo-real. Conceitos militares modernos, tais como campo-de-batalha virtual, têm em sua essência a utilização de inúmeros sensores, incluindo os de imagem, provendo assim um rico conjunto de informações sobre o local de combate, podendo assim ser usadas tanto para auxiliar em decisões estratégicas durante o planejamento quanto como vantagem durante as batalhas. Nestes casos, processamento automático de dados é usado para reduzir complexidade e/ou unir informações de múltiplos sensores com o objetivo de aproximar o máximo os campos de batalha real e virtual.

Uma das mais novas áreas é a construção de veículos autônomos, incluindo tanto os submergíveis quanto os terrestre e aéreos, em níveis de automação que vão da automação total (que usam visão computacional principalmente para navegação) a veículos onde os sistemas baseados em visão computacional auxiliam um dispositivo específico ou o piloto em várias situações, tais como para evitar uma colisão ou prevenir um acidente. Várias companhias automobilísticas já demonstraram sistemas com nível variável de autonomia, mas os maiores níveis de automação não está ainda em um nível de mercado, restando para os usuários finais apenas sistemas auxiliares e de pequena abrangência. Já na área militar, vários exemplos de veículos autônomos podem ser encontrados, indo desde os mísseis auto-guiados até aeronaves de reconhecimento. Na exploração espacial, veículos autônomos já há algum tempo utilizam visão computacional em sua navegação, o mais conhecido exemplo nesta área é o Veículo de Exploração de Marte da NASA mostrado na imagem abaixo.

Figura 2: Veículo de Exploração de Marte - NASA

Grandes empresas baseadas na internet, rede mundial de computadores, como Google e Yahoo, entre outras, disputam a ponta em outra área de aplicações, a busca de imagens na internet pelo seu conteúdo. Apesar de não existirem ainda sistemas completos do ponto de vista funcional e mercadológico, vários protótipos têm sido apresentados já contendo funcionalidades em níveis bastante usuais. O grande objetivo desta área de aplicação é obter informações a respeito do conteúdo de imagens em larga escala e de uma forma geral e completa.

1.1.4 Sistemas de visão computacional

A organização de um sistema de visão computacional tem grande dependência da aplicação na qual ele será inserido. Alguns sistemas são pequenos e resolvem apenas problemas específicos, como medição ou reconhecimento, enquanto outros são sub-sistemas de uma grande aplicação envolvendo uma complexidade ainda maior pois geralmente incluem outros sub-sistemas para atuar no ambiente, planejar ações, armazenar informações, receber informações do meio e de humanos, etc. Apesar dessa grande variabilidade, funções tipicamente encontradas em sistemas de visão computacional são:

- Aquisição de Imagem:

  Uma imagem digital é produzida por um ou muitos sensores de imagem, que podem ser geradas por vários tipos de câmeras sensíveis à luz, tomógrafos, radares, câmeras ultra-sônicas, etc. Além disso, as imagens geradas, dependendo do tipo de sensor, podem ser em duas ou três dimensões ou ainda uma seqüência de imagens, como em um vídeo. Outros detalhes ainda podem ser considerados, como o valor unitário

(pixel) gerado que tipicamente corresponde a intensidade luminosa em uma banda espectral como nas imagens em níveis de cinza ou coloridas (caso se use mais de uma banda espectral) mas também pode conter o valor de outras informações físicas como a profundidade ou distância, absorção ou reflectância de ondas sonoras ou eletromagnéticas, entre outras.

- Pré-processamento

  Antes de qualquer outro processamento específico, os sistemas de visão computacional geralmente aplicam algumas técnicas de pré-processamento para garantir que certas pré-condições necessárias adiante sejam satisfeitas. As mais usadas, são: reamostragem para garantir que as coordenadas da imagem estão corretas; tratamento de ruídos evitando que os mesmos afetem o comportamento do sistema com falsas informações; aumento de contraste para que informações relevantes possam ser detectadas; e normalização, garantindo que as informações estejam na mesma escala.

- Extração de características

  A partir das informações unitárias obtidas e pré-processadas, existem vários níveis de complexidade no qual um sistema pode representar as informações a ele necessárias, desde o próprio pixel até estruturas de mais alto nível, como linhas, elementos geométricos, áreas fechadas, pontos de interesse como intersecções, curvas ou cantos, até níveis mais altos como textura, forma, informações espaciais, objetos e até mesmo um conjunto disso tudo.

- Segmentação/Representação

  Em algum ponto do processamento, o sistema deve decidir quais informações da imagem são relevantes ou não para o restante do processamento e assim, representar aquela imagem de forma apropriada. Isto pode incluir a seleção de um específico conjunto de pontos de interesse, ou, ainda, a segmentação de uma ou múltiplas regiões da imagem que contém um objeto de interesse específico antes da representação. A partir disso, o sistema deve descartar ou desconsiderar o que não lhe for útil e passar a

tratar da imagem com a sua representação escolhida, que deve ter tanta informação quanto lhe for necessário nos passos adiantes.

- Busca baseada no conteúdo/Detecção

  Partindo da representação da imagem, os sistemas passam a buscar informações em seu conteúdo. Já não importa tão somente informações físicas ou representativas da imagem, mas sim o que ela significa ou representa em um dado contexto. Isto normalmente utiliza-se de uma etapa de detecção na qual objetos de interesse são buscados na imagem.

- Processamento de alto-nível

  Nesta etapa, que também pode ser considerada já pós-saída de um sistema de visão computacional, as informações geradas anteriormente são verificadas quanto a sua adequação ao modelo. Informações específicas do contexto da aplicação são consideradas e decisões ou informações finais são geradas já em um alto nível, tais como qual direção seguir ou onde está um objeto procurado.

1.2 Análise de imagens

Além da visão geral da área de visão computacional, faz-se necessário uma apresentação mais detalhada a respeito da análise de imagens e, conseqüentemente, da detecção de objetos e da busca baseada no conteúdo das imagens (do inglês, Content-based Image Retrieval – CBIR) visto que é a área mais especifica dentro de visão computacional em que o nosso estudo se insere e devido a importância destas áreas dentro da visão computacional.

A análise de imagens é um processo de descobrimento, de identificação e de entendimento de padrões que sejam relevantes à performance de uma tarefa baseada em imagens. Um dos principais objetivos da análise de imagens por computador é dotar uma máquina com a capacidade de aproximar, em determinado sentido, a capacidade similar dos seres humanos [5] de enxergar. Entenda-se por enxergar não apenas a simples atividade de absorver iluminação em uma determinada faixa espectral, mas sim como um processo de mais alto nível que envolve, além da captura de uma imagem, o seu completo entendimento,

chegando o mais próximo possível à excelente capacidade exibida pelos seres humanos na realização desta tarefa.

Não é tarefa trivial alcançar este objetivo tendo em vista tanto a sua grandeza quanto a sua complexidade e variabilidade de estágios. Por exemplo, o simples ato de identificar pontos negros em uma imagem branca para um determinado fim já pode ser considerado dentro da área de detecção de objetos. Este é um problema de pequena complexidade e um facilmente resolvido pelo atual nível de desenvolvimento da detecção de objetos, assim como a tarefa de reconhecer caracteres datilografados em folhas de papel branco, agrupando-os em palavras e frases mesmo sendo de uma dificuldade maior. Como tarefas de complexidade avançada e de domínio ainda não estabelecido, podemos citar a difícil tarefa de reconhecer os objetos (animais) na Figura 3 ou a detecção de todos os objetos em uma fotografia qualquer, tirada no dia-a-dia.

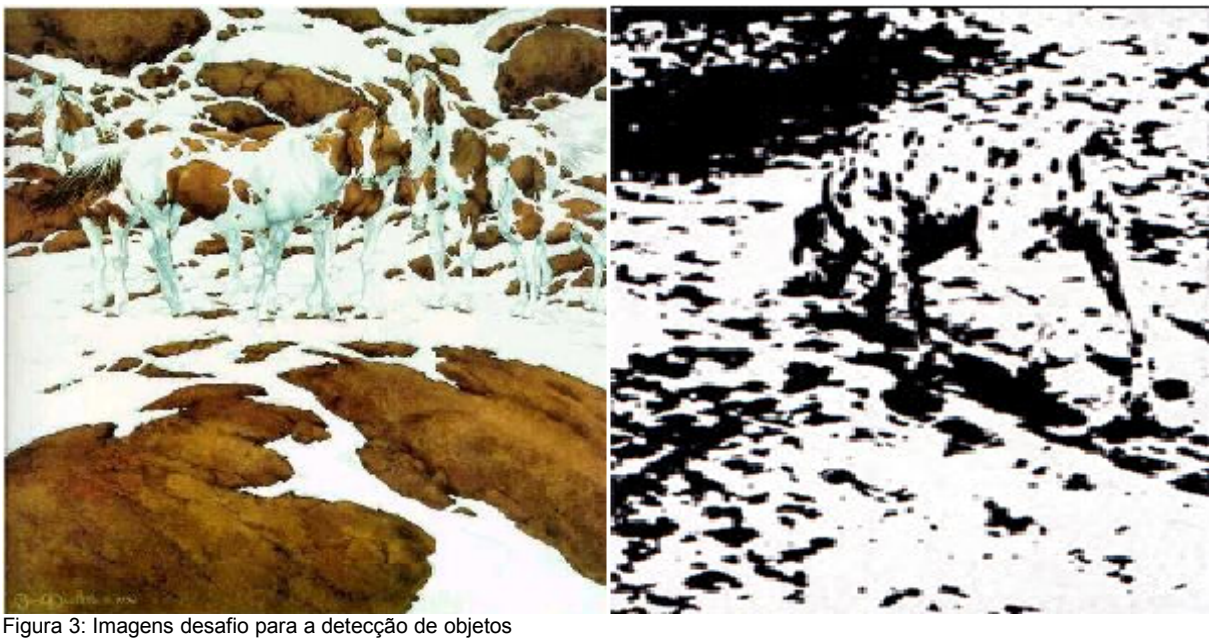

Figura 3: Imagens desafio para a detecção de objetos

Diante disso, um sistema de análise automática de imagens deveria ser capaz de exibir vários graus de inteligência, como: (1) a habilidade de extrair informação pertinente a partir de um fundo de detalhes irrelevantes; (2) a capacidade de aprender a partir de exemplos e de generalizar o conhecimento de forma que ele possa ser aplicado em circunstâncias novas e diferentes; e (3) a habilidade de fazer inferências a partir de informação incompleta.

No contexto da análise de imagens e seus desafios, a busca em imagens baseada em seu conteúdo e a detecção de objetos se encaixam como as ferramentas chaves dos processos descritos acima pois possibilitam a análise do conteúdo das imagens e não somente o seu processamento do ponto de vista de transformações matemáticas.

Content-based Image Retrieval é a área de aplicação da visão computacional que usa conteúdo visual para buscar imagens em grandes bancos de dados de imagens de acordo com os interesses do usuário. Esta busca, ao contrário das primeiras técnicas utilizadas para resolver este problema, não é feita simplesmente baseada em *tags,* modo de atuação atual dos buscadores de imagens baseados na internet, ou *links,* base das ferramentas modernas de busca baseadas na internet. Está área tem ganhado grande atenção desde o começo da década de 1990 com o grane crescimento das ferramentas de busca baseadas em internet.

Pelo fato de ser baseada em conteúdo, isto significa que a busca irá analisar o conteúdo das imagens. Neste contexto, conteúdo pode significar cores, formas, texturas ou qualquer outra informação que possa ser obtida simplesmente da imagem e que possa ser usada como representação daquela imagem. Sem essa capacidade de análise do conteúdo, buscas são feitas baseadas em meta-dados como citado anteriormente. Isto pode ser trabalhoso, custoso do ponto de vista temporal e de utilização de recursos e, ainda assim, não obter bons resultados.

Diante da necessidade de fazer buscas em grandes bases de dados de imagens baseada em seu conteúdo além de inúmeras outras aplicações, tais como as médicas, de segurança, aero-espacial, etc, as ferramentas de detecção de objetos são utilizadas como peça principal na fase de procurar no conteúdo das imagens os objetos de interesse. O propósito geral da detecção de objetos é, dada uma imagem de interesse, identificar se aquela imagem contém uma instância, ou instâncias, de uma determinada classe de interesses.

Com tudo isso, conseguimos contextualizar o nosso trabalho dentro de suas áreas de atuação, tanto em alto nível com a visão computacional e suas áreas e aplicações correlatas, quanto em baixo nível, com a análise de imagens e, por fim, a busca em imagens pelo conteúdo e a detecção de objetos. Muitas outras áreas podem ser relacionadas, assim como muitas outras definições podem ser dadas às áreas de interesse, sem contudo, modificar o grandioso objetivo deste e de muitos outros trabalhos e pesquisas, aproximar o

comportamento de uma máquina do que nós, humanos, definimos como visão, um poderoso, esplêndido e único sentido.

2 DETECÇÃO DE OBJETOS USANDO CARACTERÍSTICAS DE HISTOGRAMA ESPACIAL

Jamais considere seus estudos como uma obrigação, mas como uma oportunidade invejável para aprender a conhecer, para seu próprio prazer pessoal e para proveito da comunidade a qual seu futuro trabalho pertencer.

Albert Einstein

O método objeto principal do nosso estudo é descrito neste capítulo, a detecção de objetos usando características de histograma espacial. Histogramas espaciais consistem da distribuição marginal de uma imagem sobre partes locais, sendo capazes de preservar informações sobre a textura e a forma de um objeto simultaneamente. Características baseadas em histogramas espaciais são utilizadas então para representar as imagens e objetos da classe de interesse por se mostrarem discriminativas. O critério de Fisher é utilizado para medir a discriminabilidade e a correlação entre características de histograma espacial. Além disso, um método de treinamento é proposto para a escolha de um classificador que une a combinação de histograma em cascata, treinado através de características automaticamente selecionadas, com *support vector machine.*

2.1 Visão Geral

O método proposto é projetado para detectar múltiplas instâncias da classe de objetos de diferentes tamanhos em diferentes localizações em uma imagem de entrada [1]. A detecção de carros será usada como exemplo. Uma visão geral da arquitetura proposta, tomando como exemplo a detecção de carros, é ilustrada na figura 3. Nela, tanto componentes quanto sua organização hierárquica são exibidos. Além disso, a seqüência de ações do processo de detecção de objetos também fica clara.

Um componente essencial da arquitetura é o detector de objetos, na figura este componente é mostrado na camada intermediária e é chamado *Object Detector*. O detector de objetos usa características de histograma espacial como representação de um objeto. Este detector é formado por um classificador hierárquico que combina detecção de histograma em cascata com *support vector machine* (SVM) e será chamado de detector de objetos baseado em histogramas espaciais, ou simplesmente detector de objetos baseado em SHF.

No processo de detecção de objetos, foi adotada uma estratégia de busca exaustiva por janelas para encontrar múltiplas instâncias do objeto na imagem de entrada. O processo contém três fases: construção da imagem pirâmide (Passo 1), detecção de objetos em diferentes escalas (Passo 2) e fusão dos resultados de detecção (Passo 3).

No Passo 1, uma pirâmide da imagem é construída a partir da imagem original. Como ilustrado na figura 4, esta pirâmide é construída sub-amostrando a imagem original com fator 1,2. Conseqüentemente, o método proposto pode detectar objetos de diferentes escalas. Além disso, uma pequena janela com tamanho fixo igual ao tamanho das imagens de treinamento (100x40) percorre toda a pirâmide gerando sub-imagens em diferentes escalas. O detector é então aplicado em todas as sub-imagens geradas visando detectar instâncias do objeto em qualquer lugar da mesma.

No Passo 2, todas as sub-imagens passam para o detector de objetos baseado em SHF. Primeiramente, características de histograma espacial são geradas a partir das sub-imagens. Em seguida, aplica-se a combinação de histogramas em cascata, provendo assim uma classificação grosseira que elimina uma grande quantidade de sub-imagens, principalmente pertencentes ao *background*. Finamente, uma classificação baseada em SVM é aplicada em cada janela restante para identificar se existe ou não uma instância do objeto naquela região. Se uma sub-imagem é mapeada como uma instância de um objeto ao fim do detector, ela é mapeada para a imagem de escala correspondente na pirâmide.

O Passo 3 é um estágio para fusão dos resultados de detecção, em que instâncias sobrepostas de diferentes escalas são unidas dentro do resultado de detecção final. Baseado no fato de que o detector de objetos é insensitivo a pequenas mudanças em translação. Todas as detecções em diferentes escalas são inicialmente mapeadas na escala da imagem original, resultando em um mapa de detecção. É utilizado um método de agrupamento para combinar

detecções sobrepostas e para rotular regiões disjuntas. Como resultado, a quantidade de instâncias de objetos, suas posições e tamanhos são destacadas na imagem de saída.

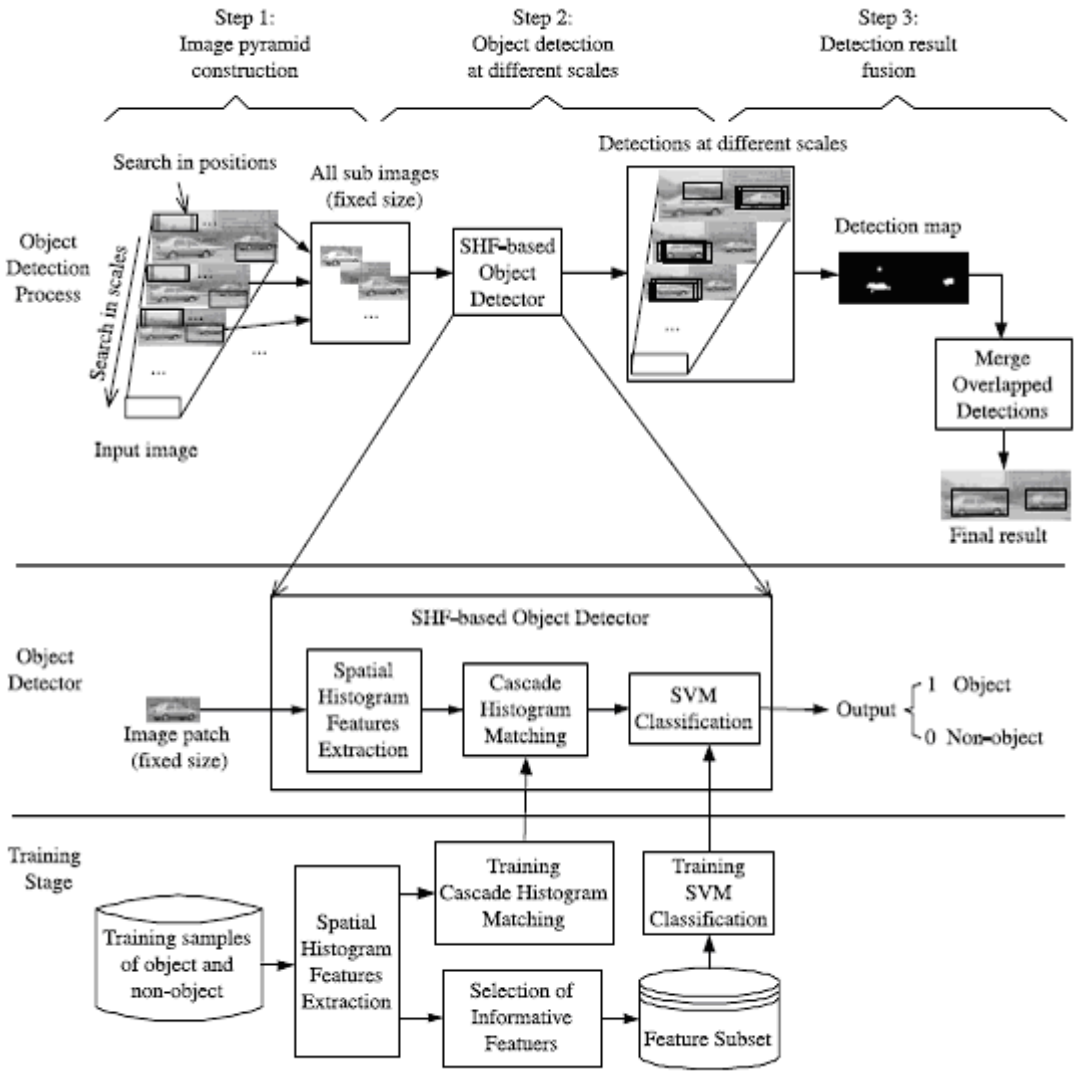

Figura 4: Arquitetura proposta em [1]

O detector de objetos baseado em SHF utiliza uma estratégia de detecção em dois estágios. Uma detecção mais grosseira feita através da combinação de histogramas em cascata, que rapidamente localiza instâncias candidatas. E uma detecção mais refinada, apenas com as instâncias candidatas geradas anteriormente, com SVM, precisamente verificando a pertinência ou não de uma imagem à classe de objetos de interesse. No mais, durante o estágio de treinamento, uma grande quantidade de amostras de objetos e não-

objetos são usados para selecionar as características de histograma espacial mais informativas e para treinar o detector SHF.

2.2 Características de histograma espacial

Representação de objetos e extração de características são essenciais para a detecção de objetos. Nesta seção descrevemos uma nova representação de padrão de objetos combinando textura e estruturas espaciais proposta em [1]. Nela, objetos são modelados por seus histogramas espaciais sobre partes locais e características específicas da classe são extraídas para detecção de objetos.

2.2.1 *Local Binary Pattern* – LBP

Inicialmente o classificador LBP (*Local Binary Pattern*) é utilizado para pré-processar as imagens modelando assim a textura das mesmas. LBP é um classificador de texturas muito simples utilizado para modelagem da textura de uma imagem. LBP tem se mostrado uma característica muito poderosa em classificação de textura [2][3] e é usado para pré-processar amostras de imagens. Entende-se por textura, a disposição ou característica dos elementos constituintes de algum material, especialmente no que se refere a aparência superficial ou táctil. Aplicando-se ao processamento de imagens, textura é uma característica representativa da distribuição espacial dos níveis de cinza dos elementos de uma imagem (*pixels*) em uma região.

A mais importante característica do operador LBP em aplicações do mundo real é sua tolerância contra mudanças de iluminação além de sua simplicidade computacional, que torna possível analisar imagens em configurações de tempo real. LBP é não variante em transformações uniformes de escala de cinza. O operador LBP básico é mostrado na Ilustração 2 e usa a intensidade da vizinhança para calcular o valor do *pixel* central.

g_1	g_2	g_3
g_4	g_0	g_5
g_6	g_7	g_8

Ilustração 2:
Operador LBP

Os *pixels* na vizinhança 3x3 têm seu valor segundo a seguinte fórmula:

$$s(g_0, g_i) = \begin{Bmatrix} 1, g_i \leqslant g_0 \\ 0, g_i < g_0 \end{Bmatrix} 1 \leqslant i \leqslant 8$$

E o valor LBP de cada um dos *pixels* é dado por:

$$LBP(g_0) = \sum s(g_0, g_i) 2^{i-1}, 1 \leqslant i \leqslant 8$$

Desta forma, por sua simplicidade, basta aplicar o operador descrito acima em todos os pontos de uma imagem e obteremos a transformada LBP daquela imagem. A figura 5 abaixo apresenta um exemplo de uma imagem LBP obtida após a aplicação do operador descrito:

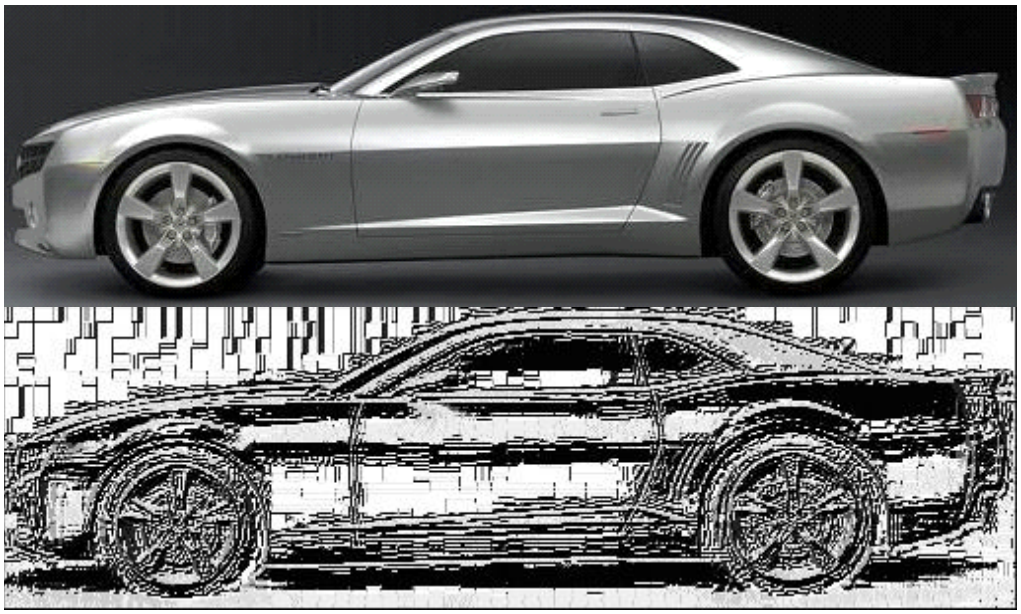

Figura 5: Exemplo de uma imagem real (acima) e de sua transformada LBP

2.3 Histogramas

O histograma de uma imagem digital com níveis de cinza no intervalo [0, L-1] é uma função discreta $h(r_k)=n_k$, onde r_k é o k-ésimo nível de cinza e n_k é o número de *pixels* na imagem que tem esse nível de cinza. É comum realizar a normalização de histograma dividindo cada um dos valores pelo número total de *pixels* na imagem, denotado por n. Dessa forma, o histograma normalizado é dado por $p(r_k)=n_k/n$, para k = 0,..., L-1. Genericamente falando, $p(r_k)$ gera uma estimativa da probabilidade de ocorrência do k-ésimo nível de cinza. Note ainda que a soma de todos os componentes de um histograma normalizado é igual a 1.

Histograma é a base para numerosas técnicas de processamento no domínio espacial. Manipulação de histogramas pode ser usado efetivamente para melhoria de uma imagem. Além de prover estimativas muito úteis a respeito de uma imagem, as informações contidas em histogramas também são muito úteis em algumas aplicações de processamento de imagens tais como compressão e segmentação. Histogramas são simples de calcular por software e implementável também em econômicas implementações de hardware, o que faz de histograma uma ferramenta muito popular para o processamento de imagens em tempo-real.

Um gráfico desta função para todos os valores de k fornece uma descrição global da aparência de uma imagem. Plotando no eixo horizontal os valores de níveis de cinza e no eixo vertical os valores do histograma associado com cada um dos níveis de cinza, imagens escuras são representadas por componentes concentrados do lado esquerdo da escala de cinza, ou seja, onde os valores de nível de cinza são menores, e imagens claras terão uma concentração de seu componentes onde os níveis de cinza são maiores, ou seja, no lado direito da escala. Uma imagem com baixo contraste tem um histograma estreito e centrado na escala de níveis de cinza. Por fim, imagens com alto contraste apresentam distribuição mais ou menos uniforme por toda a escala de níveis de cinza.

A forma do histograma de uma imagem nos dá informação útil para realce do contraste de uma imagem. Vários métodos são conhecidos e têm grande popularidade, tendo abrangência tanto global, tais como a equalização e especificação de histograma, quanto local, como o realce local utilizando as técnicas anteriores. Embora as propriedades e

características associadas a histogramas sejam descrições globais e não informem muito a respeito do conteúdo da imagem, veremos a seguir que elas são capazes de armazenar informações como a textura e a forma de objetos quando associada a informações espaciais, sendo assim propriedades poderosas para descrever e classificar objetos.

O cômputo do histograma de uma imagem LBP como pode ser utilizado como representação desta imagem mas, apesar dessa representação ser não variante à translação e rotação, não é adequada para a detecção de objetos, pois não guarda informações da distribuição espacial dos mesmos.

2.3.1 Histogramas espaciais

Histogramas, uma representação global do padrão de um objeto, é invariante à translação e à rotação. Mesmo assim, histogramas não são adequados para a detecção de objetos porque não armazena informações a respeito da distribuição espacial de objetos. Para algumas imagens pertencentes e não pertencentes à classe de interesse, seus histogramas podem ser muito similares, o que faz de histograma não suficiente para a detecção de objetos.

Para aumentar a habilidade de discriminação de objetos, histogramas espaciais são introduzidos. Neles, *templates* espaciais são usados para armazenar distribuição espacial dos padrões dos objetos. Como ilustrado na figura 6, uma janela de tamanho fixo é usada para amostrar padrões do objeto da imagem pirâmide e então armazenar padrões de distribuição espacial em *templates* espaciais.

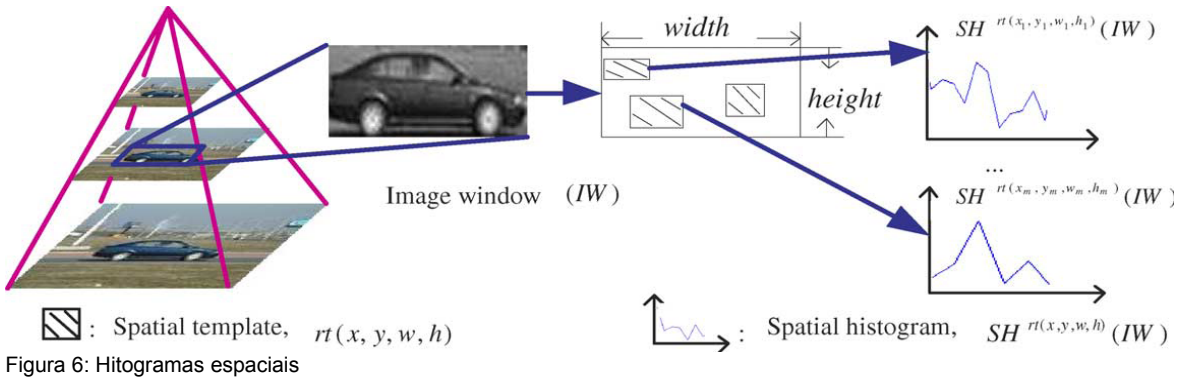

Figura 6: Hitogramas espaciais

Como mostrado na Ilustração 3, cada *template* é um retângulo denotado por rt(x, y, w, h) onde (x,y) é a localização superior esquerda do *template* dentro da imagem, w é a largura da imagem e h sua altura.

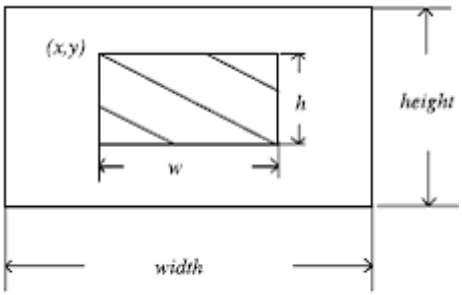

Ilustração 3: Template espacial

Para uma imagem P, seu histograma associado com o *template* rt(x, y, w, h) é denotado $SH^{rt(x,y,w,h)}(P)$. Assim, pode-se representar um modelo de histograma sobre um *template* espacial pela média dos histogramas espaciais de um conjunto de imagens de treinamento com n imagens, como:

$$SH^{rt(x,y,w,h)} = \frac{1}{n}\sum SH^{rt(x,y,w,h)}(P_j), 1 \leqslant j \leqslant n$$

Onde P_j é uma amostra de imagem de treinamento e rt(x, y, w, h) o *template* espacial.

2.3.2 Extração de características a partir de histogramas espaciais

Muitos métodos podem ser usados para medir similaridade entre dois histogramas, como distância quadrática, distância Chi-quadrada e interseção de histogramas [4]. Este trabalho utiliza a intersecção de histogramas por sua estabilidade e pequeno custo computacional. A medida de similaridade por interseção de dois histogramas [5] é calculado por:

$$D(H_{1,}H_2) = \frac{\sum min(H_1^i, H_2^i)}{\sum H_1^i}, 1 \leqslant i \leqslant k$$

Onde H_1 e H2 são dois histogramas e *k* o número de níveis quantizados nos histogramas.

Com esta definição pode-se definir uma característica de histograma espacial para uma amostra P, $f^{rt(x,y,w,h)}(P)$, como sua distância para a média de histograma como:

$$f^{rt(x,y,w,h)}(P) = D(SH^{rt(x,y,w,h)}(P), SH^{rt(x,y,w,h)})$$

Um padrão de um objeto é codificado por um conjunto de *templates* espaciais {rt(1), rt(2), ..., rt(m)}, onde *m* é o número de *templates* espaciais. Desta forma, uma amostra de objeto é representada por um vetor de características de histograma espacial no espaço de características de histograma espacial:

$$F = \{f^{rt}(1), \ldots, f^{rt}(m)\}$$

Como as máscaras podem variar em posição e tamanhos, o conjunto total de características é muito grande. Assim, o espaço de características de histograma espacial armazena completamente textura e distribuição espacial de objetos. Alem disso, características de histograma espacial é um tipo de característica específica a uma classe de objetos devido ao fato do mesmo armazenar similaridade de amostras nos modelos de histograma do objeto.

2.3.3 Habilidade de discriminação de características

Cada tipo de característica de histograma espacial tem habilidade de discriminação entre padrões de objetos e não objetos. Para demonstrar isto, consideremos as características de histograma espacial relacionadas ao padrão de visões laterais de automóveis como um exemplo. Uma base de dados contendo 200 imagens de laterais de carro com tamanho 100x40 é utilizada como base para extração do modelo do objeto a partir do template espacial rt(40, 20, 20, 20), $SH^{rt(40,20,20,20)}$. A característica de histograma espacial $f^{rt(40,20,20,20)}$ é usada como característica de teste.

A Ilustração 4 mostra a distribuição de característica sobre uma base de dados contendo 2000 amostras de carros e 15000 amostras de imagens não pertencentes ao padrão. No eixo horizontal, o valor da característica indica o valor da característica testada. No eixo vertical, freqüência significa o valor da distribuição da característica das amostras pertencentes ou não à classe estudada. Um *threshold* é usado para classificar a característica testada. Testando

com limiar igual a 0,7 é obtida uma taxa de detecção de 99,1% com taxa de erro de 45,1%, um limiar igual a 0,8 produz 93,7% de acertos e 12,1% de falsas detecções.

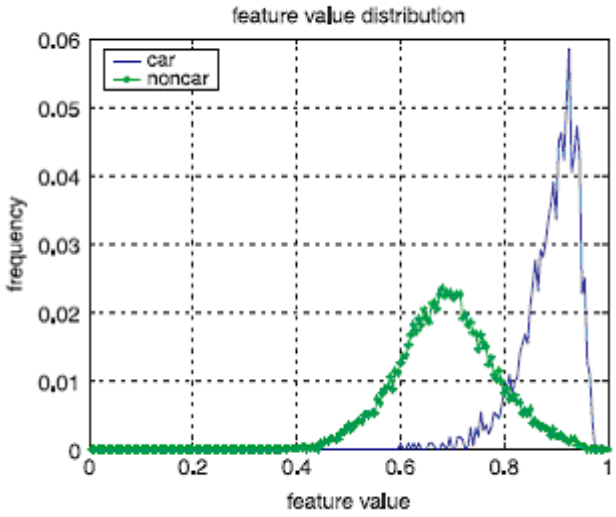

Ilustração 4: Distribuição de Características

2.3.3.1 Critério de Fisher

Para medir a capacidade de discriminação de cada característica de histograma espacial, foi utilizado o critério de Fisher. Para uma determinada característica de histograma espacial f_j (0 < j < m), supõe-se ter um conjunto de N amostras $x_1,..., x_n$, onde cada x_i é um valor escalar da característica de histograma espacial. No conjunto de dados, N_1 amostras são do conjunto w_1 e N_2 do conjunto w_2. A dispersão entre classes, $S_b=(m_1-m_2)^2$, é a distância média entre duas classes, sendo m_i a média de x_i em um conjunto de dados, e a dispersão total dentro da classe, $S_w=S_1+S_2$, é a soma das variâncias (S_i) dentro de cada uma das classes.

Desta forma, o critério de Fisher de uma característica de histograma espacial f_j é então a taxa de dispersão entre classes relacionada à dispersão total dentro das classes, dada por:

$$J(f_j)=\frac{S_b}{S_w}$$

A grande vantagem do Critério de Fisher é que ele é tão mais discriminante quanto mais for discriminante a característica de histograma espacial. Sendo assim um método direto e simples.

2.4 Aprendizado para detecção de objetos

Para classificação de objetos é usada uma combinação de histograma em cascata e SVM em uma estrutura hierárquica. No entanto, como o espaço de características de histograma espacial é muito grande, é crucial selecionar um subconjunto compacto de características para uma classificação eficiente. Nesta seção, são vistas as características básicas dos métodos utilizados, com um enfoque na combinação de histograma em cascata.

2.4.1 Combinação de histograma em cascata

Combinação de histograma é um método direto para reconhecimento de objetos. Nele, o modelo de histograma de um padrão de objeto é gerado a partir de um template espacial. Se o histograma de uma amostra é similar ao histograma modelo atingindo um limiar pré-definido, a amostra é classificada como pertencente ao padrão. Assim, seja P uma amostra e $f^{rt(x,y,w,h)}$ (P) a sua característica de histograma espacial associado ao *template* rt(x, y, w, h). P é classificado como um objeto pertencente ao padrão se $f^{rt(x,y,w,h)}$(P) é maior ou igual a θ (*threshold* para a classificação).

Aplicar a combinação de histograma com apenas um *template* espacial não é aceitável para um sistema de detecção de objetos. Desta forma, é selecionado um conjunto dessas características e elas são combinadas em cascata para realizar a combinação de histograma, dando origem então a combinação de histogramas em cascata, que funciona como se segue: selecionam-se *n* características de histograma espacial $f_1, f_2, ..., f_n$ e respectivos *thresholds* de classificação $\theta_1, \theta_2, ..., \theta_n$. A regra de decisão da combinação de histograma em cascata é a seguinte:

$$H(P)=\begin{cases} 1 \rightarrow objeto, se(f_1(P) \geqslant \theta_1) \wedge ... \wedge (f_n(P) \geqslant \theta_n), \\ 0 \rightarrow não-objeto, caso\ contrário \end{cases}$$

A contribuição de cada característica é medida por seu Critério de Fisher e pela taxa de detecção, quantidade de detecções sobre a quantidade total de imagens em uma base positiva – na qual todas as imagens são pertencentes ao padrão – de imagens. Para a seleção de um subconjunto de características, é proposto um método de treinamento que seleciona um conjunto F_{select} de características juntamente com um conjunto de limiares ThreSet de classificação para construir um classificador para a combinação de histograma em cascata.

Supondo que já se tem (1) o espaço de características de histograma espacial F = {f_1, f_2, ..., f_m}; (2) conjuntos de treinamento, um com imagens pertencentes ao padrão e um com imagens não-pertencentes ao padrão: SP e SN; (3) Conjuntos de validação, também pertencentes e não pertencentes ao padrão: VP e VN; (4) uma taxa de detecção aceitável, D. São usados ainda dois parâmetros baseados na precisão da classificação, Acc(Pre) (precisão anterior) e Acc(Cur) (precisão corrente). O método de treinamento é o seguinte:

1. Inicialização: F_{select} = {}, ThreSet = {}, t = 0, accPre = 0, AccCur = 0;

2. Compute o Critério de Fisher J(*f*) usando os conjuntos SP e SN para cada característica *f* pertencente a F;

3. Encontre a característica f_t que maximiza o critério de Fisher;

4. Execute a combinação de histograma com f_t no conjunto de validação V = VP e encontre um limiar θ_t tal que a taxa de detecção *d* em VP seja maior ou igual a D;

5. Compute a precisão de classificação corrente em VN:

   $$Acc(cur) = 1 - Fp$$

   Onde Fp é o número de falsas detecções ou falsos positivos;

6. Se Acc(cur) satisfaz: $Acc(Cur) - Acc(Pre) \leqslant \varepsilon$, onde ε é um pequena constante positiva, o procedimento finaliza e retorna F_{select} e ThreSet, se não, segue o procedimento seguinte:

1. Acc(Pre) = Acc(Cur), SN = {}, $F_{select} = F_{select}$ U { f_t }, F = F\{f_t}, ThreSet = ThreSet U { θ_t }, t = t + 1;

2. Bootstrap: realize a combinação de histograma em cascata para F_{select} e ThreSet em um conjunto de imagens contendo nenhum dos objetos alvo, adicione as falsas detecções a SN;

3. Volte ao passo 2.

2.4.2 Support Vector Machine para o reconhecimento de objetos

Combinação de histograma em cascata é o estágio de detecção de objetos mais grosseiro e, apesar de obter alta taxa de detecção para objetos verdadeiros, a taxa de detecção positiva para falsos objetos ainda é alta. Para aumentar a performance de detecção, é utilizada a classificação SVM como um detector de objetos mais refinado [1]. SVM [6] realiza reconhecimento de padrões de um problema com duas classes determinando o hiper plano de separação que maximiza a distância para os mais próximos pontos de um conjunto de treinamento.

Não utilizaremos o SVM em nossa implementação, pois foge ao escopo deste trabalho, e por isso não nos aprofundaremos no seu estudo, mas trabalhos futuros podem adicioná-lo a esta implementação. Mais informações sobre a aplicação do SVM na detecção de objetos usando características de histograma espacial são encontradas em [1] e um estudo mais completo é encontrado em [6].

3 EXPERIMENTOS E RESULTADOS

É lenta a experiência de todas as fontes profundas: elas precisam esperar muito para saber o que caiu na sua profundidade.

Friedrich Nietzsche

Com o intuito de validar os estudos, foi feita uma implementação de uma instância da detecção de objetos utilizando características de histograma espacial proposta em [1]. A instância implementada se propõe a avaliar o comportamento do método sem a utilização de SVM, que possibilita uma classificação mais refinada e assim aumenta substancialmente a precisão do método. Outros pontos não foram implementados por se apresentar de modo confuso ou não serem devidamente detalhados, como a formação da pirâmide da imagem, o que possibilita a detecção de objetos de tamanhos variáveis.

Algumas medidas de performance foram usadas para avaliar nosso sistema de detecção de objetos:

- Taxa de detecção – definida como o número de detecções corretas sobre o número total de objetos da classe de interesse no conjunto de dados de teste;

- Taxa de falsos positivos – definida como o número de detecções falsas positivas, ou seja, em que o detector afirmou haver um objeto da classe e não havia, sobre o número total de possibilidades negativas na base de dados;

- Precisão – definida como o número de detecções corretas sobre a soma de detecções corretas e de detecções falsas positivas.

Os experimentos foram feitos testando-se a detecção de visões laterais de carros, uma classe de objetos que apresentar uma certa variabilidade de formas e tamanhos, mas conta com uma configuração de componentes espaciais relativamente constante. Visões laterais de

carros consistem do agrupamento espacial de partes bem distintas como pneus, portas e janelas além de outros componentes de menor tamanho ou maior variabilidade como faróis e visões laterais da traseira e frente do carro. Estas partes são arranjadas em uma configuração espacial relativamente fixa. Visões laterais de carros têm enormes mudanças na configuração devido aos vários estilos e *design* dos carros.

Para o treinamento, geração de modelos e testes em geral, foi utilizada a base de dados da UIUC disponível em [7]. Nela, é encontrada uma boa quantidade de imagens tanto de treinamento quanto de testes, contando com:

- Base de treinamento positiva – 550 imagens de tamanho 100X40 contendo apenas uma lateral de carro centrada e de tamanho equivalente que foi usada como base positiva de treinamento;

- Base de treinamento negativa – 500 imagens de tamanho 100X40 não contendo lateral de carro que foi usada como base negativa de treinamento;

- Base de testes – 170 imagens de tamanhos e em situações diversas apresentando ao menos uma lateral de carro por foto, com um total de 200 carros, sendo todos eles de tamanho aproximadamente 100X40, que foi usada como base de testes para todos os nossos experimentos;

- Base de testes com escala – 108 imagens contando com 139 carros de diferentes tamanhos que não foi utilizada mas que poderia ser a base para testes envolvendo objetos em diferentes escalas.

As bases de teste são consideradas difíceis para detecção visto que contêm visões de laterais de carros parcialmente ocultas, em baixo contraste com o fundo da imagem, em fundos com alto contraste, de carros em movimento e em diversas localizações e tamanhos além de terem sido tiradas em diferentes resoluções. As imagens foram todas capturadas na área urbana de Champaign-França e foram adquiridas parte tiradas normalmente com uma câmera e parte retiradas de seqüências de imagens de vídeo de carros em movimento.

A busca exaustiva de *templates* para cobrir todo o espaço de características dentro de imagens de tamanho 100X40 é impraticável, gerando em torno de 3.600.000 *templates*. No

entanto, além de impraticável do ponto de vista computacional, este conjunto de *templates* é excessivamente completo, e a maioria dos *templates* espaciais têm tamanhos pequenos e insignificantes na representação de características além de serem sobrepostos por *templates* maiores. Para sanar esses problemas e reduzir o espaço de *templates,* criamos um *template* de tamanho mínimo, 10X10, que é movido em passos de 5 *pixels* nas direções horizontal e vertical e então aumentado em tamanhos múltiplos de 10. Isto nos deu um espaço de 289 *templates* usado como entrada do processo de treinamento.

A implementação foi realizada em linguagem JAVA 1.6 e todos os testes foram executados em computador com processador AMD Athlon 2300+, sistema operacional Windows XP Service Pack 2 e memória volátil de 1 MB.

Como resultados dos testes, pode-se observar visualmente que a detecção de objetos em cascata cumpre o seu papel, fazer uma filtragem inicial de baixo custo nas possibilidades de detecção de objetos por toda a imagem para alguns pontos específicos. Um exemplo dos *templates* e *thresholds* associados são exibidos abaixo:

- Template1(15, 20, 10, 10) – θ = 0,52;
- Template2(15, 15, 10, 10) – θ = 0,53;
- Template3(15, 15, 10, 30) – θ = 0,73;
- Template4(20, 15, 10, 20) – θ = 0,64;
- Template5(20, 15, 10, 309 – θ = 0,76;

Em média foram geradas 811 possíveis localizações do objeto por imagem de teste antes da aplicação do detector, que gerou como saída 6,4 posições de possíveis objetos em uma base tem uma média de 1,18 carros por imagem. Diante da imensa base de possíveis objetos recebida como entrada merece destaque a taxa total de falsos positivos de 0,7% em toda a base de treinamento. Abaixo segue uma comparação com o método proposto:

	Detecções corretas	Detecções falsas	Taxa de detecção	Precisão
Zhang et al. [1]	164/200	187	82%	46,70%
Nossa implementaçã	142/200	952	71%	13%
Zhang et al. [1]	196/200	441	98%	30,70%
Nossa implementaçã	153/200	1284	76,50%	10,60%

Tabela 1: Comparação entre o proposto e o implementado.

Abaixo seguem algumas boas detecções alcançadas pelo método, Figura 7. Apesar de ser apenas um estágio de filtragem para uma posterior detecção mais acurada, algumas imagens já apesentam resultados similares aos finais desejados.

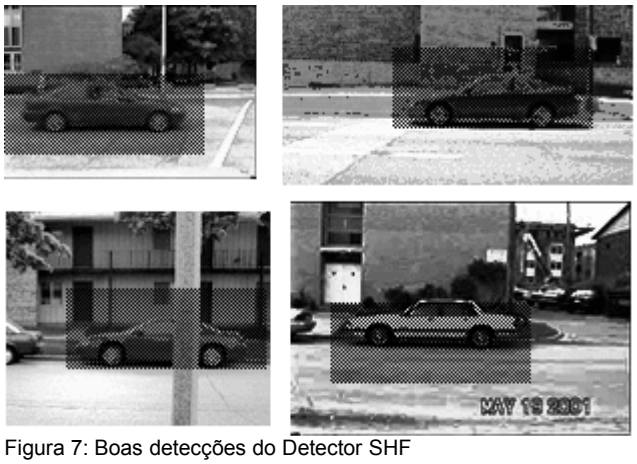

Figura 7: Boas detecções do Detector SHF

No grupo de imagens abaixo, são mostradas como a maioria das imagens se comporta. Percebe-se que apesar de não indicar o local exato, o algoritmo seleciona poucas áreas de maior probabilidade da existência de um objeto procurado.

Figura 8: Média das detecções nas imagens

Por fim, observa-se o grupo das más detecções. Não indicando uma área como uma possível área de localização de um objeto o método fica incapacitado de aplicar uma outra detecção mais adiante, originando uma falha na detecção já no primeiro estágio do classificador.

Figura 9: Erros de detecção

4 CONSIDERAÇÕES FINAIS

Quantas coisas terá de vencer e deixar para trás até que, por fim, encontre a verdade... Quantas vezes será acometido, durante sua caminhada, da sensação de estar aspirando o impossível... E, não obstante, chegará o dia em que o impossível se tornará no possível e, mais ainda, no natural.

Eugen Herrigel

Ao fim deste trabalho, concluímos que a visão computacional se mostra como uma área de grandes desafios, contando atualmente com uma elevada quantidade de pesquisas e grupos de trabalhos nas maiores e melhores faculdades e empresas tecnológicas, além de já ter diversos periódicos e grandes eventos no mundo inteiro. Por outro lado, o mercado ainda apresenta carência tanto de mão-de-obra qualificada quanto de produtos completos tecnologicamente, o que aumenta ainda mais a demanda por trabalhos e meios de difundir o seu aprendizado.

Em um nível mais abaixo, a área de análise de imagens apresenta-se muito interessante por ainda ter grandes problemas em abertos, como o próprio problema da detecção de objetos, objetivo maior deste trabalho. Áreas correlatas como segmentação, processamento de imagens, matemática, estatística, inteligência artificial, aprendizagem de máquina, redes neurais e até mesmo a biologia e a psicologia podem trazer grandes avanços e seus estudos são de grande importância. Além disso, a busca de imagens pelo conteúdo e a detecção de objetos têm causado grandes avanços nesta área.

A busca de imagens pelo conteúdo e a detecção de objetos é um tópico de grande disputa principalmente no mercado tecnológico de buscas baseadas na internet – com grandes empresas como o Google© sendo muito ativas nesta área – e outras áreas como a robótica, a segurança e a automação veicular e de processos, entre outras. Por isso, essas duas áreas têm

recebido muito destaque dentro de visão computacional e causado grandes avanços tanto teóricos quanto tecnológicos.

Todo este interesse e boas disputas nestas áreas têm trazido mais estudantes e pesquisadores para a visão computacional assim como atraído o mercado para as suas aplicações. Acreditamos assim estar passando por uma uma fase chave na evolução destas áreas, principalmente Conten-Based Image Retrieval e detecção de imagens onde as suas aplicações vão passar a fazer parte em larga escala de nossas vidas.

Quanto ao método de detecção de objetos baseado em características de histograma espacial proposto em [1], tanto os resultados exibidos no próprio trabalho quanto os obtidos através da sua implementação parcial mostram que a representação usando características de histograma espacial é geral para diferentes tipos de classe e, juntamente com o método de seleção de características apresentado, mostra-se eficiente para extrair características específicas de uma determinada classe para a detecção de objetos.

A implementação do método proposto em [1], mostrou-se bastante eficiente para a detecção de objetos mesmo não contando com o SVM, que incrementaria o grau de refinamento da detecção. Diante dos problemas enfrentados, obtivemos bons resultados, apresentando grande quantidade de detecções positivas mas também alta taxa de falsos positivos. Um resultado logicamente já esperado.

A detecção de objetos utilizando características de histograma espacial se mostrou bastante satisfatória para um primeiro estágio rápido de seleção de áreas com maior probabilidade da presença de objetos da classe de interesse para posterior utilização de algum outro método mais custoso como o próprio SVM. Por outro lado, devido as suas características como discriminante, maiores estudos poderiam incrementar a sua precisão.

Como trabalhos futuros deixamos um maior estudo a respeito de outros métodos de representação tais como os de pontos de interesse mesclados com algumas boas ferramentas utilizadas no método proposto, como as informações espaciais, o critério de Fisher, o método de aprendizado, entre outros.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Zhang, H.[Hongming], Gao, W.[Wen], Chen, X.[Xilin], Zhao, D.[Debin], Object detection using spatial histogram features, IVC(24), No. 4, 1 April 2006, pp. 327-341.

[2] M. Pietikainen, T. Ojala, Z. Xu, Rotation-invariant texture classification using feature distributions, Pattern Recognition 33 (2000), pp. 43–52.

[3] T. Ojala, M. Pietikainen, T. Maenpa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7) (2002), pp. 971–987.

[4] B.Schiele, Object recognition using multidimensional receptive field histograms, PhD Thesis, I.N.P. Grenoble. English translation, 1997.

[5] GONZALEZ, R.C., WOODS, R.E. Processamento de Imagens Digitais. Editora Edgard Blucher LTDA, 2000.

[6] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

[7] UIUC Image Database for Car Detection, http://l2r.cs.uiuc.edu/~cogcomp/Data/Car/, 2004.

[8] Agarwal, S. [Shivani], Awan, A. [Aatif], Roth, D. [Dan], Learning to detect objects in images via a sparse, part-based representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (11) (2004), pp. 1475-1490.

[9] S. Ullman, High-Level Vision: Object Recognition and Visual Cognition. MIT Press, 1996.

[10] H.A. Rowley, S. Baluja, T. Kanade, Neural network-based face detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (1) (1998) 29–38.

[11] C. Garcia, M. Delakis, Convolutional face finder: a neural architecture for fast and robust face detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (11) (2004) 1408–1423.

[12] E. Osuna, R. Freund, F. Girosi, Training support vector machines: an application to face detection, Proceedings of the 1997 IEEE Conference on Computer Vision and Pattern Recognition, 1997, pp. 130–136.

[13] C.P. Papageprgiou,T. Poggio, A training object system: car detection in static images, MIT AI Memo No. 180, 1999.

[14] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition 1 (2001) 511–518.

[15] S.Z. Li, L. Zhu, Z.Q. Zhang, A. Blake, H.J. Zhang, H. Shum, Statistical learning of multi-view face detection, Proceedings of the Seventh European Conference on Computer Vision 4 (2002) 67–81.

[16] X.R. Chen, A. Yuille, Detecting and reading text in natural scenes, Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (2004) 366–373.

[17] K.K. Sung, T. Poggio, Example-based learning for view-based human face detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (1) (1998) 39–50.

[18] C.J. Liu, A Bayesian discriminating features method for face detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2003) 725–740.

[19] B. Menser, F. Muller, Face detection in color images using principal component analysis, Proceedings of the Seventh International Congress on Image Processing and its Applications, 1999, pp. 13–15.

[20] A. Mohan, C. Papageorgiou, T. Poggio, Example-based object detection in images by components, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (4) (2001) 349–361.

[21] M.V. Naquest, S. Ullman, Object recognition with informative features and linear classification, Proceedings of the Ninth International Conference on Computer Vision, 2003, pp. 281–288.

[22] M. Weber, M. Welling, and P. Perona, Unsupervised Learning of Models for Recognition *Proc. Sixth European Conf. Computer Vision,* pp. 18-32, 2000.

[23] R.M. Haralick and L.G. Shapiro, *Computer and Robot Vision II.* Addison-Wesley, 1993.

[24] H. Schneiderman, T. Kanade, A statistical method for 3D object detection applied to faces and cars, Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition 1 (2000) 746–751.

[25] B.Schiele, Object recognition using multidimensional receptive field histograms, PhD Thesis, I.N.P. Grenoble. English translation, 1997.

[26] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.

[27] A.J. Colmenarez and T.S. Huang, “Face Detection with Information-Based Maximum Discrimination,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 782-787, 1997.

[28] F. Bernhard, K. Christian, Real-time face detection using edge-orientation matching, Proceedings of the Third International Conference Audio- and Video-Based Biometric Person Authentication, 2001, pp. 78–83.

[29] C. Garcia, G. Tziritas, Face detection using quantized skin color regions merging and wavelet packet analysis, IEEE Transactions on Multimedia 1 (3) (1999) 264–277.

[30] A. Hadid, M. Pietikainen, T. Ahonen, A discriminative feature space for detecting and recognizing faces, Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, pp. 797–804.

[31] H.M. Zhang, D.B. Zhao, Spatial histogram features for face detection in color images, 5th Pacific Rim Conference on Multimedia, Lecture Notes in Computer Science 3331 (2004) 377–384.