

RECONHECIMENTO DE MARCOS VISUAIS COMO CENAS PARA LOCALIZAÇÃO DE ROBÔS MÓVEIS

Valguima Odakura Martínez

Anna Helena Reali Costa

Laboratório de Técnicas Inteligentes

Departamento de Engenharia de Computação e Sistemas Digitais

Escola Politécnica da Universidade de São Paulo

Av. Prof. Luciano Gualberto, 158, tv. 3 – 05508-900 São Paulo, SP

{valguima.odakura,anna.reali}@poli.usp.br

Resumo

Este trabalho descreve uma técnica para reconhecimento de marcos visuais em ambientes internos para localizar robôs móveis. Cada marco é representado por uma cena, e um grupo de marcos similares define um local do ambiente. As cenas são representadas através de PCA – Análise de Componentes Principais e o reconhecimento é realizado por um classificador estatístico supervisionado, não paramétrico, que estima as funções densidade de probabilidade condicionais utilizando o estimador de Parzen. Resultados preliminares da utilização desta técnica são apresentados.

Abstract

This work describes a technique to recognize visual landmarks for indoor places to localize mobile robots. Each landmark is represented by a scene, and a group of similar landmarks defines a place. The scenes are represented using PCA – Principal Components Analysis and the recognize is realized by a supervised statistic classifier, non parametric, that estimates the probabilities densities functions using Parzen estimator. The results obtained with the application of this techniques are described.

1 INTRODUÇÃO

O propósito do presente trabalho é o reconhecimento de cenas em ambientes internos para localizar robôs móveis. Cenas do ambiente interno serão utilizadas como marcos visuais, sendo que cada cena será um marco. Esta abordagem difere das técnicas em que marcos são baseados em objetos individuais da cena.

A representação da cena para o reconhecimento de ambientes internos baseada na aparência, ao invés de utilizar características de objetos individuais, foi utilizada em [Torralba and Sinha, 2001], [Kröse and Bunschoten, 1999], [Kröner et al., 2001] e [Pourraz and Crowley, 1998]. Ao utilizar marcos de objetos individuais encontra-se a desvantagem de que os marcos nem sempre estão disponíveis em todas as imagens. Com a representação das cenas, cada cena preserva a organização dos maiores elementos, como mesas, janelas, etc., e não é tão afetada por elementos menores como suplementos de escritório, livros, etc.

[Torralba and Sinha, 2001] acreditam que as propriedades globais da cena são relevantes para discriminar cenas diferentes. Segundo eles, ambientes distintos possuem diferentes elementos estruturais que os compõem e a organização espacial desses elementos possui uma combinação espacial particular. Por exemplo, uma sala é composta por paredes, mesas e cadeiras e uma rua é composta de uma estrada cercada de construções nos dois lados.

Para reconhecer cenas será utilizado um classificador estatístico supervisionado, não paramétrico, que estima a função densidade de probabilidade condicional utilizando o estimador de Parzen. Tal técnica foi empregada para o reconhecimento de ambientes internos por [Torralba and Sinha, 2001] e [Kröner et al., 2001]. Como o reconhecimento é supervisionado, conjuntos de cenas serão fornecidos para o sistema, em que cada conjunto representa um local do ambiente.

No entanto, antes de reconhecer as cenas é necessário representá-las. Neste trabalho, as cenas serão representadas como vetores de atributos (chamados marcos). Um problema com essa representação é que imagens possuem alta dimensão. Torna-se, então, necessária a extração dos atributos mais discriminantes das cenas a fim de reduzir a dimensão do vetor de atributos. O método de extração de atributos utilizado será a PCA – Análise de Componentes Principais. Nos trabalhos de [Torralba and Sinha, 2001], [Kröner et al., 2001], [Kröse and Bunschoten, 1999], e [Pourraz and Crowley, 1998] as imagens das cenas foram representadas utilizando PCA.

Este trabalho está relacionado com o projeto de doutorado intitulado VisLoc – Identificação de Marcos Visuais para Localização de Robôs Móveis, e dá o primeiro passo em direção a uma solução para o problema. Uma breve descrição do projeto VisLoc será fornecida na Seção 2. As seções seguintes tratam das técnicas utilizadas para o reconhecimento

de cenas de ambientes internos. Na Seção 3 será desenvolvido o assunto de reconhecimento de padrões estatístico, em especial o estimador de Parzen. A representação das cenas utilizando PCA será tratada na Seção 4. Os resultados obtidos serão apresentados na Seção 5. Por fim, as considerações finais serão apresentadas na Seção 6.

2 VISLOC – IDENTIFICAÇÃO DE MARCOS VISUAIS PARA LOCALIZAÇÃO DE ROBÔS MÓVEIS

Robôs móveis, segundo [Ribeiro et al., 2001], são agentes artificiais, ativos, com capacidade de locomoção, imersos no mundo físico real. Para operar de forma autônoma, um robô móvel deve ser capaz de reconhecer sua localização em seu ambiente, ou seja, a definição de sua posição no espaço e de sua orientação.

Existem vários métodos de localização de robôs móveis na literatura. Quase todos os métodos extraem um pequeno conjunto de atributos das medidas do sensor do robô, sendo que há uma variedade desses atributos. [Borenstein et al., 1996] descrevem um apanhado geral das técnicas existentes.

Quando há a necessidade de efetuar um deslocamento adequado dentro de um ambiente desconhecido o robô deve estimar, com razoável precisão, seu deslocamento. Segundo [Deans and Hebert, 2000] esta tarefa torna-se relativamente fácil se um conhecimento prévio do ambiente é obtido, por exemplo utilizando um mapa ou recursos externos como o GPS – Sistema de Posicionamento Global (*Global Position System*), para estimar seu posicionamento.

No entanto, caso o robô não possua previamente um mapa do ambiente ou informações sobre sua posição global, a busca por pontos de referência, ou marcos, tem se mostrado aplicável e vem sendo adotada em vários projetos de pesquisa. O problema de busca por marcos normalmente é conhecido por CLM – Mapeamento e Localização Concorrentes (*Concurrent Localization and Mapping*), ou por SLAM – Mapeamento e Localização Simultâneos (*Simultaneous Localization and Mapping*). Trabalhos relacionados com esse problema podem ser encontrados em [Choset and Nagatani, 2001, Marco et al., 2001].

Marcos podem ser extraídos por diversos sensores. Um dos mais flexíveis e versáteis, com possibilidade de fornecimento de várias informações relevantes, por exemplo textura, formas, etc., é a visão computacional [Draper et al., 1996]. Assim, o uso de marcos visuais para tarefas de localização é altamente relevante.

Entretanto, reconhecer marcos visuais em uma sequência de imagens é um problema desafiador, pois a aparência de qualquer dado marco varia de uma observação para outra. Além disso, variações devido a aspectos diferentes como mudanças na iluminação, oclusões parciais, imprecisão de sen-

sos na formação de imagens, ruídos na captura, entre outros, são fatores que afetam os marcos observados.

Por essas razões, nem sempre é possível utilizar técnicas de reconhecimento de objetos fortemente baseadas em modelos geométricos. Uma alternativa é usar técnicas baseadas em imagens em que marcos são representados por coleções de imagens, sendo que em tais imagens encontra-se a aparência típica dos objetos. As informações mais relevantes para o reconhecimento são extraídas da coleção de imagens e utilizadas como modelos de identificação. Este processo é conhecido como aprendizado visual.

Assim, considerando que marcos visuais são utilizados para localização de robôs, duas fases são necessárias. Na fase de aprendizado o robô desloca-se em uma área por ele desconhecida, capturando imagens de objetos e classificando-as em grupos conforme a similaridade entre seus atributos. Cada grupo descreve um marco encontrado. Na fase de localização, numa passagem futura pela mesma área, o robô captura novas imagens e o algoritmo classifica essas imagens como integrantes de um dos grupos que foram previamente aprendidos. Assim sendo, o robô tem uma estimativa do seu posicionamento, uma vez que passou a conhecer o local.

O projeto de doutorado tem por objetivo propor um método de aprendizado automático de marcos visuais para tarefa de localização de robôs móveis em ambientes internos como escritórios, hospitais, museus, etc., explorando representações e seleções eficientes de marcos e rápidas indexações no conjunto de marcos selecionados.

O projeto propõe uma arquitetura de software constituída de duas fases, sendo uma de aprendizado de marcos e a outra de localização, conforme descrita na Figura 1. Na fase de aprendizado de marcos são realizadas as etapas de aquisição de imagens, seleção e representação de marcos, associação de um posicionamento para cada marco, e por fim, uma base de conhecimento estruturada é construída. A fase de localização consiste da aquisição e representação das imagens e do reconhecimento dos marcos em relação aos marcos armazenados na base de conhecimentos. Por fim, realiza-se um aprimoramento da estimativa de localização resultando na localização do robô.

Nas seções seguintes serão abordadas as fases de aprendizagem e localização respectivamente.

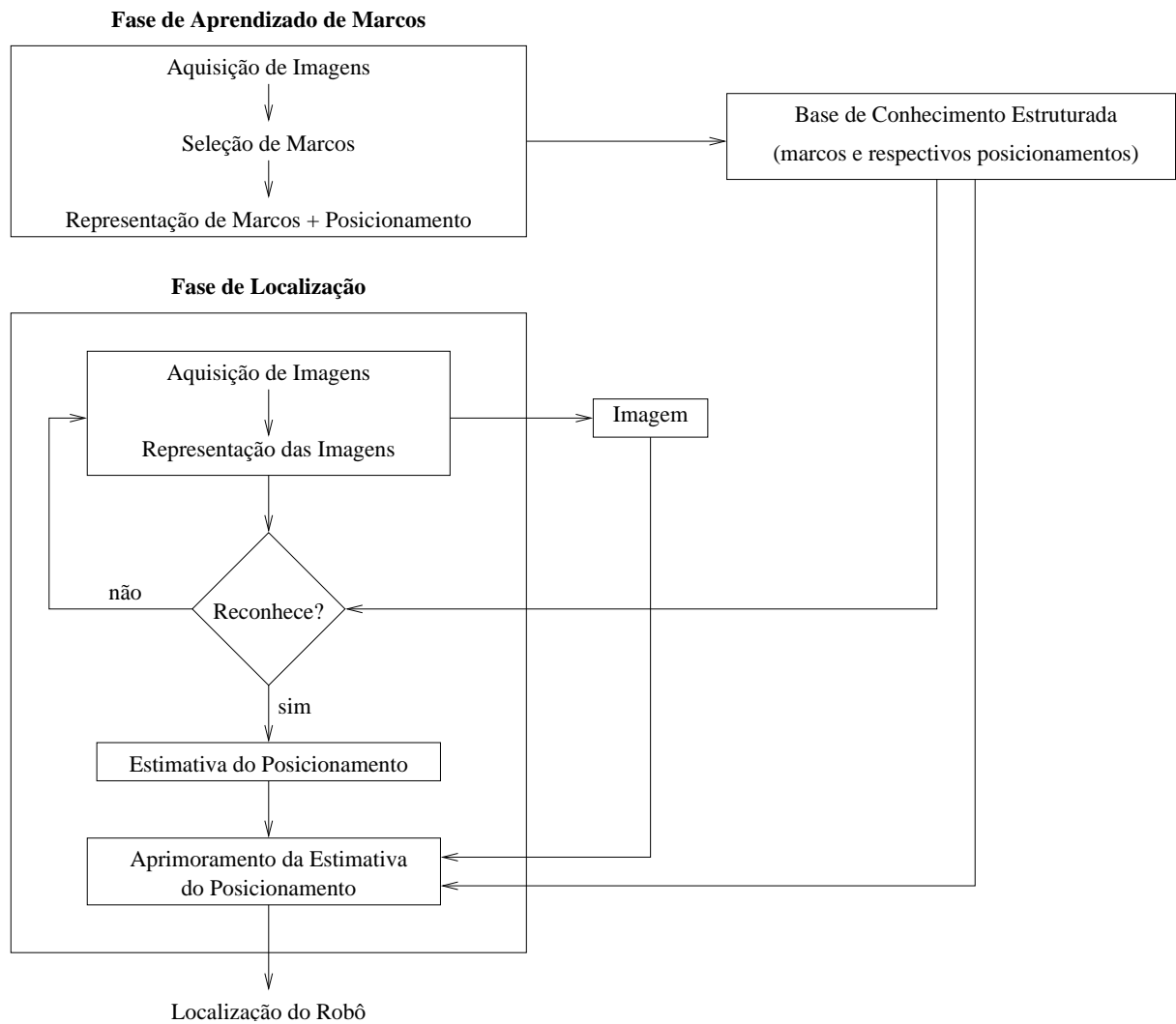


Figura 1: Sistema de Localização de Robôs Móveis Utilizando Marcos Visuais.

2.1 Aprendizado dos Marcos Visuais

Na fase de aprendizado são selecionadas características adequadas do ambiente para serem utilizadas como marcos. A escolha dos marcos deve considerar que a aparência de um marco é sensível a mudanças nas condições da iluminação do ambiente, que podem causar reflexões e sombras nas imagens, e é sensível a mudanças nas distâncias e ângulos dos marcos, uma vez que aspectos diferentes dos objetos podem se tornar visíveis, e oclusões podem ser geradas.

Nesta fase, o sistema recebe um conjunto de imagens em seqüência tomada de uma câmera em movimento e seleciona marcos para representar as imagens. Em seguida, as imagens são classificadas em grupos baseados na similaridade dos marcos obtidos. A escolha de representações adequadas para armazenar informações relevantes dos marcos é um tópico de grande interesse e que deve ser definido com cuidado.

A aquisição de marcos compreende a escolha de atributos característicos que distinguem um marco, a

representação desses atributos de forma apropriada e, por fim, o treinamento do sistema.

A escolha de atributos deve ser realizada de forma que as informações relevantes contidas na imagem da cena sejam preservadas. Esse cuidado é necessário já que as imagens são extraídas de ambientes que apresentam variações significativas nas condições de iluminação e nos ângulos de observação, por exemplo.

Neste trabalho, a escolha dos atributos será efetuada levando-se em consideração a aplicação do sistema na área de robótica móvel em ambientes internos que, entre outras particularidades, tem como fonte de imagens uma câmera acoplada a um robô móvel, sendo suscetível às vibrações decorrentes de irregularidades no solo e possíveis mudanças significativas nos ângulos de observações a cada novo deslocamento efetuado pelo robô no ambiente. Ambientes internos se caracterizam por possuir formas geométricas mais definidas, repletas de ângulos retos, superfícies planas, etc.

A escolha dos atributos levanta a questão de quais atributos devem ser extraídos das imagens,

considerando que tais atributos devem produzir os melhores resultados de localização. Os principais atributos considerados nesta proposta de trabalho são cores, níveis de cinza, tipos e número de segmentos, segmentos paralelos, agrupamento de segmentos e bordas. Não necessariamente todas serão utilizadas. Uma idéia inicial é empregar testes estatísticos sobre os atributos que suportem a escolha daqueles que melhor se aplicam ao problema.

Uma vez definidas quais atributos irão representar cada marco, uma estrutura vetorial, normalmente chamada de vetor de atributos, será utilizada para armazená-los. O vetor de atributos pode ser expresso por:

$$X(M_j) = (x_1, x_2, \dots, x_n), \text{ para } j = 1, \dots, m, \quad (1)$$

onde M_j é um marco, m é o número de marcos, x_i são os atributos para $i = 1, \dots, n$ e n é o número de atributos do marco.

2.2 Estimação da Localização do Robô Móvel

A localização de robôs móveis é muito importante em robótica móvel, uma vez que o robô nem sempre consegue executar precisamente seus comandos. Ao executar um comando de ação, a posição e orientação atuais do robô sempre diferem da posição e orientação para que foi efetivamente programado. Esse erro deve-se principalmente ao deslizamento das rodas em superfícies lisas e irregularidades no terreno, erros de odometria, etc.

Localização é o problema de encontrar as coordenadas do robô relativas ao seu ambiente, assumindo que um mapa do ambiente é conhecido. Para lidar com a incerteza da informação do sensor, muitas técnicas utilizam uma abordagem probabilística para estimar a posição do robô como em [Fox, 1998], [Thrun, 1998] e [Kröner et al., 2001], tais técnicas apresentam bons resultados. No caso de não ser fornecido um mapa do ambiente, marcos visuais são utilizados para estimar a localização.

Uma vez que os marcos aprendidos na fase anterior foram armazenados, o sistema pode estimar a posição do robô assim que estes marcos forem identificados e reconhecidos no ambiente. Quando é necessário estimar a posição do robô, imagens de uma câmera são requeridas. São extraídos marcos dessas imagens utilizando o mesmo procedimento de seleção de marcos da fase de aprendizado e então uma técnica de reconhecimento é empregada para encontrar o grupo mais similar ao marco extraído.

2.2.1 Reconhecimento de Marcos

O reconhecimento de um marco em tempo real envolve uma comparação entre uma representação de uma nova imagem e todas dos modelos. Uma vez que os vetores contendo os atributos de cada marco estejam devidamente armazenados em uma base de

conhecimentos, quando uma nova imagem for apresentada ao sistema, um vetor com seus atributos será gerado e então comparado com os aprendidos.

A decisão de uma dada representação na imagem pertencer ou não a um modelo pode ser tomada utilizando algumas técnicas para a comparação dos vetores armazenados com aqueles extraídos das novas imagens. Desta forma, espera-se que o sistema possa classificar marcos. [Takeuchi and Herbert, 1998] descreve abordagens por distância euclidiana ponderada, por redes de uma única camada e redes neurais de várias camadas. Os melhores resultados no trabalho de Takeuchi foram obtidos com a utilização de redes neurais com 6 camadas escondidas. [Livatino and Madsen, 1999] realizam o reconhecimento do marco utilizando o método de correlação cruzada normalizada e [Lee, 1997] aplica o método de verossimilhança máxima para reconhecer o marco. Em outro trabalho, [Torralba and Sinha, 2001] e [Kröner et al., 2001] utilizam estimador de Parzen para o reconhecimento.

Uma vez feita a identificação do marco, a coordenada (x, y) a ele associada fornece ao robô a possibilidade de estimar o seu posicionamento relativo aos demais marcos e ao ambiente. Observe que esse posicionamento é apenas uma estimativa, não havendo portanto uma maior preocupação com um alto nível de precisão. Para que uma maior precisão seja obtida é necessário um ajuste fino do posicionamento do robô com o objetivo de coincidir as imagens aprendidas.

Uma vez que se conhece uma estimativa inicial do posicionamento do robô, obtida com o reconhecimento, pode-se buscar um aprimoramento da estimativa da localização do robô para obter maior precisão na informação de localização.

3 RECONHECIMENTO ESTATÍSTICO DE PADRÕES

O reconhecimento estatístico de padrões pode ser realizado em duas fases: treinamento (ou aprendizado) e classificação. O classificador toma decisões baseado no conjunto de amostras de padrões obtidas na fase de treinamento.

Um padrão x , em reconhecimento estatístico de padrões, é um vetor de atributos d -dimensional, modelado como um vetor aleatório em que cada $x(i)$ para $i = 1, \dots, d$ é um atributo do espaço de atributos d -dimensional. Uma classe é um conjunto de padrões que possuem similaridade entre si.

Um classificador pode ser visto como uma função $F: X \rightarrow C$, tal que $F(x) = C_i$. Cada C_i é a i -ésima classe do conjunto de classes C , e x é um padrão de X .

O objetivo do reconhecimento de padrões é escolher vetores de atributos que permitam que padrões pertencentes a diferentes classes ocupem regiões compactas e disjuntas do espaço d -dimensional de atributos.

A tarefa da fase de treinamento é estabelecer fronteiras de decisão no espaço de atributos que separem padrões pertencentes a classes distintas. Em reconhecimento estatístico de padrões as fronteiras de decisão são determinadas pelas distribuições de probabilidade pertencentes a cada classe.

Quando o conjunto de treinamento possui as classes pré-definidas, o reconhecimento de padrões é dito supervisionado. Se as classes devem ser aprendidas, baseando-se nas similaridades dos padrões, então, o reconhecimento é não supervisionado.

A escolha do classificador depende de quais informações sobre o problema são conhecidas. Se todas as densidades condicionais de classe são conhecidas, então a regra de decisão de Bayes pode ser empregada para projetar o classificador. No entanto, quase nunca as densidades condicionais das classes são conhecidas; deve-se então, aprendê-las, a partir dos padrões de treinamento disponíveis.

Se a forma da densidade condicional é conhecida, mas alguns de seus parâmetros são desconhecidos, tem-se um problema de decisão paramétrico. A solução comum para esse problema é substituir os parâmetros desconhecidos nas funções densidade por seus valores estimados. Por exemplo, sabe-se que a forma da densidade condicional é gaussiana, mas a sua média e covariância são desconhecidas, então deve-se estimar esses parâmetros.

Por outro lado, se a forma da densidade condicional não é conhecida, tem-se um problema de decisão não paramétrico. Deve-se, então, estimar a função densidade de probabilidade ou construir diretamente uma fronteira de decisão a partir do conjunto de treinamento. Um exemplo de estimador de densidade é o estimador de Parzen e um exemplo de construção de fronteiras de decisão é o método de vizinhos mais próximos.

A Seção 3.1 trata do classificador de Bayes, introduzindo os conceitos necessários para possibilitar a formalização do estimador de Parzen, que será tratado na Seção 3.2. Técnicas de reconhecimento estatístico de padrões podem ser encontradas no livro de [Duda and Hart, 1973] e, mais recentemente, em uma revisão das técnicas por [Jain et al., 2000].

3.3 Classificador Bayesiano

A teoria de decisão de Bayes é fundamental para o problema de classificação estatística de padrões. As decisões são tomadas utilizando estimativas de distribuição de probabilidade. Essa técnica é baseada na suposição que todos os valores de probabilidade são conhecidos.

A probabilidade a priori da classe C_i , denotada por $P(C_i)$, representa o conhecimento de que um dado vetor x pertença a uma classe C_i , $i = 1, \dots, c$, onde $\sum_{i=1}^c P(C_i) = 1$.

A função densidade de probabilidade do padrão x , quando se sabe que x pertence à classe C_i , é $p(x|C_i)$, e a função densidade de probabilidade de

x é definida por $p(x) = \sum_{i=1}^c p(x|C_i)P(C_i)$.

Dado um padrão x com classificação desconhecida, a probabilidade de x pertencer à classe C_i , para $i = 1, \dots, c$, denotada por $p(C_i|x)$, é a probabilidade a posteriori da classe C_i e pela regra de Bayes pode ser calculada como:

$$P(C_i|x) = \frac{p(x|C_i)P(C_i)}{p(x)}, \quad (2)$$

onde,

$$p(x) = \sum_{j=1}^c p(x|C_j)P(C_j). \quad (3)$$

Sempre que um padrão é atribuído a uma classe, pode ocorrer um erro de classificação. Para medir o erro cometido pode-se utilizar uma função erro definida por:

$$e_i(x) = 1 - P(C_i|x), \text{ para } i = 1, \dots, c. \quad (4)$$

A regra de decisão de Bayes para minimizar o erro procura selecionar a ação que minimize a função erro. Assim, deve-se selecionar a classe C_i que maximiza a probabilidade a posteriori $P(C_i|x)$, ou seja, decidir a classe C_i se $P(C_i|x) > P(C_j|x)$ para todo $j \neq i$.

3.4 Estimador de Parzen

O estimador de Parzen é uma técnica supervisionada, não paramétrica para estimar, a partir do conjunto de treinamento, a função densidade de probabilidade quando sua forma não é conhecida. A função densidade de probabilidade que se quer estimar é $p(x|C_i)$, ou seja, a função densidade de probabilidade do padrão x , quando se sabe que x pertence à classe C_i .

Na estimação de parzen, a função densidade de probabilidade $p(x)$ pode ser estimada a partir de uma função janela $\varphi(u)$. A função $\varphi(u)$ deve ser uma função densidade, ou seja, deve atender os seguintes requerimentos:

$$\varphi(u) \geq 0 \quad (5)$$

$$\int \varphi(u)du = 1. \quad (6)$$

Assim, a estimação da função densidade de probabilidade $p(x)$ em relação à função janela $\varphi(u)$, pode ser obtida como segue:

$$p(x) = \frac{1}{t} \sum_{k=1}^t \frac{1}{h^t} \varphi\left(\frac{(x - x_i)}{h}\right), \quad (7)$$

onde h é a largura da janela e t é o número de amostras. A altura da janela h deve satisfazer $\lim_{n \rightarrow \infty} h^t = 0$ e $\lim_{n \rightarrow \infty} th^t = \infty$ e para que a estimação de $p(x)$ possa convergir para $p(x)$.

A equação acima expressa a estimação de $p(x)$ como uma média de funções de x e amostras x_i . A função janela está sendo usada para interpolação,

em que cada amostra contribui para a estimação de acordo com sua distância de x .

Uma função janela amplamente utilizada é a gaussiana multivariada, com a largura da janela determinada por h ,

$$\varphi(x - x_i) = \frac{1}{(2\pi)^{d/2} h^d |\Sigma|^{1/2}} \times \exp\left(-\frac{1}{2h^2}(x - x_i)^T \Sigma^{-1}(x - x_i)\right), \quad (8)$$

onde Σ é a matriz de covariância, $|\Sigma|$ é o determinante de Σ e $(x - x_i)^T$ é a matriz transposta de $(x - x_i)$.

Assim, a estimação da função densidade de probabilidade $p(x)$ é uma média das densidades normais centradas nas amostras,

$$p(x) = \frac{1}{t} \sum_{i=1}^t \varphi(x - x_i), \quad (9)$$

onde t é o número de amostras, ou tamanho do conjunto de treinamento.

Uma dificuldade com o método de janela é a escolha da matriz de covariância [Kröner et al., 2001], [Duda and Hart, 1973]. Uma das estimativas é assumir independência estatísticas entre as variáveis, mudando todo elemento fora da diagonal principal para zero. Embora a suposição de independência seja quase sempre incorreta, essa estimativa apresenta melhor resultado em comparação com o estimador de verossimilhança máxima.

4 PCA – ANÁLISE DE COMPONENTES PRINCIPAIS

Antes de realizar o reconhecimento de padrões é necessário representar os dados, de forma que cada padrão seja um vetor de atributos. No caso de reconhecimento de ambiente, os dados são as cenas que serão representadas como vetores de atributos, chamados marcos. A necessidade de redução de dimensionalidade em visão computacional é muito grande, uma vez que imagens são padrões de alta dimensão. Uma imagem pode ser considerada como um padrão, que possui dimensão $d = m \times n$, onde m é a altura e n é a largura da matriz de imagem, em pixels. Para o reconhecimento de padrões, nem todos os atributos da imagem precisam ser utilizadas. Assim, é possível reduzir a dimensão do padrão de imagem a fim de obter apenas os atributos com maior poder de discriminação.

Para solucionar o problema de dimensionalidade foram propostas muitas técnicas. Um procedimento estatístico clássico é a PCA – Análise de Componentes Principais (*Principal Component Analysis*), também conhecido como transformação de Karhunen-Löve [Duda and Hart, 1973], [Gonzalez and Woods, 1992]. O objetivo da PCA é encontrar uma representação de menor dimensão que considere a variância dos atributos.

A análise de componentes principais é em geral utilizada para reduzir dimensionalidade e/ou extrair atributos. No entanto, ao utilizar o modelo baseado em aparência, PCA é utilizada também como parte do reconhecimento, ou localização. PCA foi utilizada para representar cenas de ambientes internos em [Kröse and Bunschoten, 1999], [Kröner et al., 2001], [Pourraz and Crowley, 1998] e [Torralba and Sinha, 2001] e para representar imagens de faces humanas em [Moghaddam and Pentland, 1997] e [de Campos, 2001].

Seja uma imagem I em níveis de cinza, representada por uma matriz de dimensão $m \times n$, na qual cada elemento representa a intensidade da imagem naquele ponto. A imagem pode ser representada na forma de um vetor através da leitura coluna a coluna da matriz de imagem e armazenando cada pixel em um vetor coluna. Assim,

$$x(l) = I(i, j), \quad (10)$$

para $i = 1, \dots, n, j = 1, \dots, m$ e $l = i + (j-1) \times m$. A dimensão do vetor de atributos é dada por $d = m \times n$.

Considere que são conhecidos t padrões de treinamento x_1, x_2, \dots, x_t . O conjunto de treinamento pode ser visto como uma matriz X , em que cada coluna possui um padrão de treinamento,

$$X = \begin{bmatrix} x_1(1) & x_2(1) & \dots & x_t(1) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(d) & x_2(d) & \dots & x_t(d) \end{bmatrix}. \quad (11)$$

A matriz Σ_X de covariância da matriz de treinamento X , pode ser definida como:

$$\Sigma_X = (X - \mu)(X - \mu)^T, \quad (12)$$

onde μ é a matriz média de dimensão $d \times t$, sendo que cada coluna da matriz contém o valor esperado dos padrões de X .

$$\mu(i, j) = \frac{1}{t} \sum_{j=1}^t (X(i, j)), \quad (13)$$

para $i = 1, \dots, d$ e $j = 1, \dots, t$.

Dada a matriz de covariância Σ_X , calcula-se seus auto-vetores e auto-valores de maneira que:

$$\lambda_i e_i = \Sigma_X e_i, \quad (14)$$

onde λ_i e e_i , para $i = 1, \dots, d$ são os auto-valores e auto-vetores de Σ_X , respectivamente.

Os auto-valores indicam a relevância dos auto-vetores. No caso de PCA, se o auto-vetor possui grande auto-valor, significa que o auto-vetor está em uma direção com grande variância nos padrões. Dessa forma, os auto-vetores são organizados em ordem decrescente de auto-valores.

O conjunto ordenado de auto-vetores compõe a matriz de transformação H da seguinte maneira,

$$H = [e_1, e_2, \dots, e_d]. \quad (15)$$

A transformação do vetor de atributos é realizada através de uma mudança de base, onde a matriz H é a matriz de mudança de base,

$$Y = H^T X, \quad (16)$$

e H^T é a matriz transposta de H .

A matriz de novos atributos Y obtida não apresenta redução da dimensionalidade, sendo apenas uma mudança de base do espaço de atributos. No entanto, a matriz H pode ser construída apenas com os auto-vetores que correspondem aos maiores auto-valores. Assim, se forem escolhidos k auto-valores, $k < d$, a dimensão do vetor de atributos é reduzida para uma representação k -dimensional.

No trabalho de [Kröse and Bunschoten, 1999], a escolha de k foi realizada ajustando a fração de variância preservada para 80%, ou seja,

$$v = \sum_{i=1}^k (\lambda_i) / \sum_{i=1}^d (\lambda_i). \quad (17)$$

Em reconhecimento de padrões busca-se por características discriminantes, ou seja, busca-se por vetores de atributos que não possuam covariância entre si. Através da matriz de covariância Σ_X é possível verificar se há covariância entre os atributos existentes. Ao realizar a transformação dos atributos X para os novos atributos Y utilizando PCA, observa-se que a matriz Y possui média zero e matriz de covariância Σ_Y , diagonal, em que os elementos da diagonal principal são os auto-valores de Σ_X , e os elementos fora da diagonal são zero. Isso significa que os elementos da matriz Y são descorrelacionados.

Para reconhecimento de ambientes internos pode-se explorar a extensão da representação utilizando PCA em uma descrição da cena baseada em bordas. Utiliza-se um detector de bordas que resulta em um mapa de bordas. Tal mapa é uma representação alternativa e tem a vantagem de ser menos suscetível a mudanças na iluminação. [Moghaddam and Pentland, 1997] utilizaram, com sucesso, o detector de Canny combinado com PCA para representar faces humanas.

5 RECONHECIMENTO DE PADRÕES USANDO PCA

O reconhecimento de ambientes internos foi realizado utilizando marcos visuais, onde cada marco é representado por uma cena, e um grupo de marcos similares define um local do ambiente. As cenas são representadas através de PCA e o reconhecimento é realizado por um classificador estatístico supervisionado, não paramétrico, que estima as funções densidade de probabilidade condicionais utilizando o estimador de Parzen.

Para testar a técnica descrita foram utilizadas imagens de três salas do LTI – Laboratório de Técnicas Inteligentes da EPUSP. Uma amostra das imagens

utilizadas pode ser vista na Figura 2.

O sistema foi treinado para diferenciar os três locais. O treinamento foi realizado utilizando uma sequência de vídeo obtida posicionando a câmera na entrada das salas e realizando uma rotação de aproximadamente 180 graus na câmera. As imagens da sequência foram identificadas como pertencentes a cada um dos locais antes de iniciar o treinamento.

Para o reconhecimento, foram avaliadas as probabilidades $P(C_i|x)$, onde x é o vetor de atributos de uma imagem que se quer identificar, obtido através de PCA, e C_i são as três classes possíveis, ou seja,

$$P(C_i|x) = \frac{p(x|C_i)P(C_i)}{\sum_{j=1}^c p(x|C_j)P(C_j)}. \quad (18)$$

O estimador de Parzen foi utilizado para obter as densidades condicionais $p(x|C_i)$. A probabilidade a priori de cada classe utilizada foi $P(C_i) = 1/3$, considerando que todas as classes são equiprováveis.

A classificação de uma imagem desconhecida é realizada atribuindo um rótulo para uma classe que possui a maior probabilidade $P(C_i|x)$, se essa probabilidade está acima de um limiar pré-definido.

As Figuras 3 e 4 ilustram imagens classificadas corretamente.

A sequência de vídeo obtida para o treinamento foi possui, 174 imagens para a classe 1, 119 para a classe 2 e 130 para a classe 3. Todas as imagens foram usadas para treinamento. Para testar o sistema foi utilizada uma nova sequência de vídeo.

Nos testes realizados as imagens foram reduzidas para 32×24 pixels, a escolha desse tamanho de imagem foi necessário, pois com imagens maiores, o classificador tornou-se muito lento. Foram utilizados 15 atributos por imagem, o que corresponde 80% da variância preservada. Variamos o parâmetro h do estimador de Parzen, entre 0.1 e 1. Os melhores resultados encontrados foram com $h = 0.25$. O resultado da classificação foi dividido em classificação correta, incorreta ou sem classificação. Os resultados encontram-se na tabela abaixo.

classes	correta	incorreta	sem classificação
1	78%	19%	3%
2	55%	24%	21%
3	20%	71%	9%

Ao observar os dados da tabela é possível notar que a classe 3 não foi classificada corretamente em 71% das imagens de teste. Esse resultado foi atribuído ao fato que a aparência da classe 1 é muito semelhante a da classe 3, o que tornou incorreta a classificação. No entanto, um bom classificador depende da relação entre o número de atributos utilizados e o número de amostras de treinamento, e no caso do estimador de Parzen, da escolha de um valor para h . Dessa forma, acreditamos ser possível melhorar o classificador ajustando os parâmetros envolvidos.

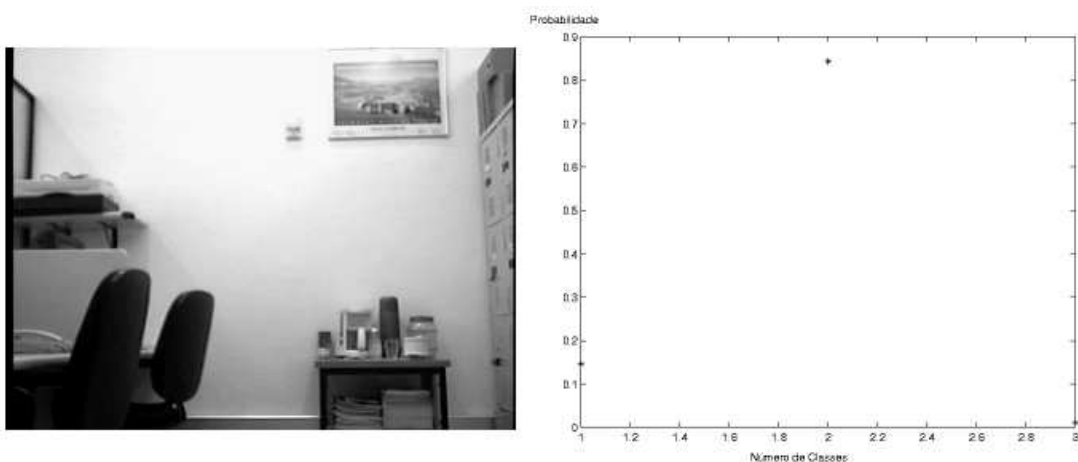


Figura 3: Resultado da classificação: uma imagem desconhecida, classificada como pertencente a classe 2.

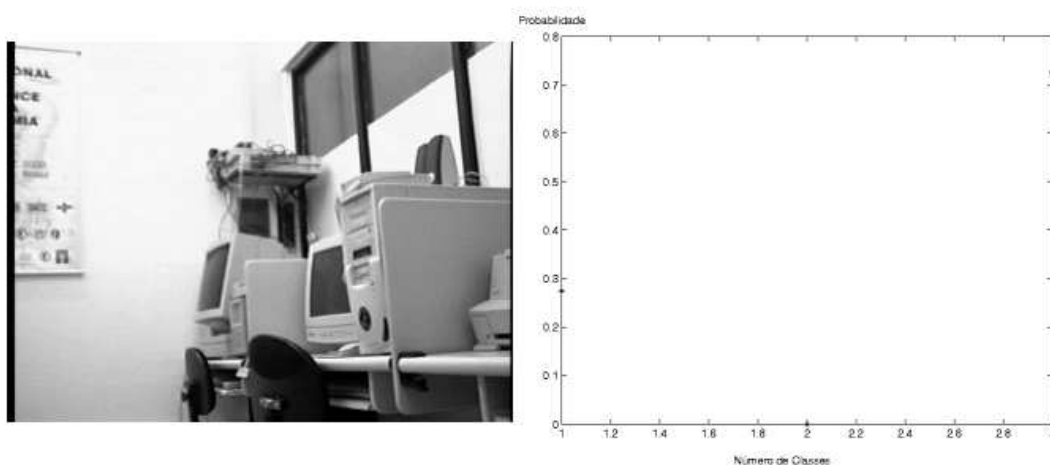


Figura 4: Resultado da classificação: uma imagem desconhecida, classificada como pertencente a classe 3.

6 DISCUSSÃO E TRABALHOS FUTUROS

O presente artigo apresentou uma técnica para reconhecimento de ambientes internos baseada em marcos visuais, em que cada cena é representada como um marco. Para reconhecer os marcos, um classificador estatístico supervisionado foi utilizado.

O fato da técnica de reconhecimento empregada ser supervisionada, não atende os objetivos do projeto VisLoc, que pretende realizar uma identificação autônoma dos marcos para posterior localização. Assim, uma extensão natural deste trabalho, na parte de reconhecimento, é a utilização de técnicas não supervisionadas.

Na busca por um conjunto de atributos compacto e discriminante, pode-se investigar outras técnicas de extração de atributos, além de explorar outros atributos, como tipos e número de segmentos, segmentos paralelos, agrupamento de segmentos, bordas, cores e níveis de cinza.

AGRADECIMENTOS

Este trabalho faz parte do projeto AACROM – Aprendizagem Autônoma da Coordenação para

Robôs Móveis, FAPESP N^o.01/14588-2 e do projeto MAPPEL - *Multi-Agent Collaborative and Adversarial Perception, Planning, Execution, and Learning*, financiado pelo NSF/CNPq-PROTeM CC, Proc. N^o 68.0033/99-8.

7 *

Referências

- [Borenstein et al., 1996] Borenstein, J., Everett, B., and Feng, L. (1996). *Navigating Mobile Robots: Systems and Techniques*. Wellesley.
- [Choset and Nagatani, 2001] Choset, H. and Nagatani, K. (2001). Topological simultaneous localization and mapping (SLAM): toward exact localization without explicit localization. *IEEE Transactions on robotics and automation*, 17(2):125–137.
- [de Campos, 2001] de Campos, T. E. (2001). Técnicas de seleção de características com aplicações em reconhecimento de faces. Master's thesis, Universidade de São Paulo.

- [Deans and Hebert, 2000] Deans, M. and Hebert, M. (2000). Invariant filtering for simultaneous localization mapping. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.
- [Draper et al., 1996] Draper, B., Hanson, A. R., and Riseman, E. M. (1996). Knowledge-direct vision: control, learning and integration. *Proceedings of the IEEE*, 84(11):1625–1637.
- [Duda and Hart, 1973] Duda, R. O. and Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. John Wiley & Sons.
- [Fox, 1998] Fox, D. (1998). *Markov Localization: A Probabilistic Framework for Mobile Robot Localization and Navigation*. PhD thesis, University of Bonn.
- [Gonzalez and Woods, 1992] Gonzalez, R. C. and Woods, R. E. (1992). *Digital image processing*. Addison-Wesley Publishing Company.
- [Jain et al., 2000] Jain, A. K., Duin, R. P. W., and Moa, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(37):4–37.
- [Kröner et al., 2001] Kröner, B. J. A., Vlassis, N., Bunschoten, R., and Motomura, Y. (2001). A probabilistic model for appearance-based robot localization. *Image and Vision Computing*, 18:381–391.
- [Kröse and Bunschoten, 1999] Kröse, B. J. A. and Bunschoten, R. (1999). Learning and evaluating visual features for pose estimation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 2255–2260.
- [Lee, 1997] Lee, J. (1997). Indoor robot navigation by landmark tracking. *Mathl. Comput. Modelling*, 26(4):79–89.
- [Livatino and Madsen, 1999] Livatino, S. and Madsen, C. B. (1999). Autonomous robot navigation with automatic learning of visual landmarks. In *SIRS'99 7th International Symposium on Intelligent Robotic System*.
- [Marco et al., 2001] Marco, M. D., Garulli, A., Lacroix, S., and Vicino, A. (2001). Set membership localization and mapping for autonomous navigation. *International Journal of robust and nonlinear control*, 11(7):709–732.
- [Moghaddam and Pentland, 1997] Moghaddam, B. and Pentland, A. (1997). Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7).
- [Pourraz and Crowley, 1998] Pourraz, F. and Crowley, J. L. (1998). Continuity properties of the appearance manifold for mobile robot position estimation. In *Symposium for Intelligent Robotics Systems*.
- [Ribeiro et al., 2001] Ribeiro, C. H. C., Costa, A. H. R., and Romero, R. A. F. (2001). Robôs móveis inteligentes: princípios e técnicas. In *Congresso da Sociedade Brasileira de Computação*, volume 3, pages 257–306, Fortaleza. Jornada de Atualização em Inteligência Artificial.
- [Takeuchi and Herbert, 1998] Takeuchi, Y. and Herbert, M. (1998). Evaluation of image-based landmark recognition techniques. Technical Report CMU-RI-TR-98-20, Carnegie Mellon University.
- [Thrun, 1998] Thrun, S. (1998). Bayesian landmark learning for mobile robot localization. *Machine Learning*, 31(1):29–53.
- [Torralba and Sinha, 2001] Torralba, A. and Sinha, P. (2001). Recognizing indoor scenes. Technical Report AI Memo 2001-015 CBCL Memo 202, Massachusetts Institute of Technology.