

UMA ARQUITETURA DE VISÃO ARTIFICIAL: APLICAÇÃO EM ROBÓTICA MÓVEL BASEADA EM FUSÃO DE SENSORES

LUCIANO OLIVEIRA*, AUGUSTO COSTA*, LEIZER SCHNITMAN*

* *Universidade Federal da Bahia*

Emails: lrebouca@ufba.br, augusto.loureiro@ufba.br, leizer@ufba.br

Abstract— Mobile robots are usually used in order to liberate the human being of dangerous or boring tasks. Each part of a mobile robot possesses proper aspects, that must be integrated in order to enable it to the accomplishment of a particular task. Amongst these parts, the sensing makes possible the construction of a representation of landmarks of the environment with the objective to supply relevant informations for robot navigation. This article describes an architecture of a sensing system based on fusing of data of a CMOS camera and distance sensors. The objective of the architecture proposal is the determination and localization of objects in a soccer field. An SVM is used in such a way for the recognition how much for the localization of objects. The fusing process is carried through a Fuzzy system using a TSK model.

Keywords— Artificial Vision, Support Vector Machine, Fuzzy Logic, Mobile Robot.

Resumo— Robôs móveis são usualmente empregados a fim de liberar o ser humano de tarefas perigosas ou enfadonhas. Cada parte de um robô móvel possui aspectos próprios, que devem estar integrados a fim de capacitá-lo para a realização de uma tarefa particular. Dentre estas partes, o sensoramento possibilita a construção de uma representação dos marcos do ambiente com o objetivo de fornecer informações relevantes para a navegação do robô. Este artigo descreve uma arquitetura de um sistema de sensoramento baseado em fusão de dados de uma câmera CMOS e sensores de distância. O objetivo da arquitetura proposta é a determinação e localização dos objetos num campo de futebol. Uma SVM é utilizada tanto para o reconhecimento quanto para a localização dos objetos. O processo de fusão é realizado através de um sistema difuso utilizando um modelo TSK.

Palavras-chave— Visão Artificial, Máquina de Vetor de Suporte, Lógica Difusa, Robôs móveis.

1 Introdução

O campo da robótica móvel tem demonstrado vasta possibilidade de pesquisas e experimentos reais. Robôs móveis que apresentam comportamentos inteligentes são constituídos de diversas partes que isoladamente revelam potenciais de estudo (Kortenkamp et al., 1998).

Particularmente, o sensoramento é responsável por prover o robô de informações necessárias para a construção de uma representação do ambiente, onde o robô está inserido, permitindo, assim, uma descrição dinâmica dos obstáculos e marcos úteis para a orientação do robô. Tal tarefa deve ser realizada através de métodos computacionais cujo objetivo é reduzir a natureza imprecisa dos sensores. Para isso, técnicas de fusão de dados de sensores vêm sendo aplicadas com sucesso, fornecendo através de redundância e complementaridade dos dados, uma descrição mais precisa e completa do ambiente.

Para se avaliar um sistema de visão artificial e sua arquitetura, deve-se aplicá-los a alguma tarefa específica. Para o sistema proposto, o futebol de robôs é escolhido e as regras da Robocup são utilizadas para as restrições do sistema de visão. Para um time de robôs participar de uma partida de futebol, várias tecnologias devem estar presentes: princípios de projetos de agentes autônomos, colaboração multi-agente, robótica, fusão de sensores, entre outros. A aplicação da robótica móvel no ambiente de futebol de robôs tem como objetivo principal, portanto, a análise de múltiplas áreas do

conhecimento que servem de suporte a problemas socialmente significativos e na área industrial.

Para o sistema de visão artificial proposto neste trabalho, uma câmera CMOS e sensores de distância infravermelhos são utilizados. O sistema possui duas etapas: reconhecimento de objetos e fusão de dados dos sensores, obtendo, ao final, a completa localização dos objetos reconhecidos. Para a primeira etapa, uma SVM (*Support Vector Machine*) é aplicada a fim de classificar os objetos caracterizados por uma cor única. Outra SVM é utilizada para a obtenção das coordenadas polares dos objetos, através de uma regressão, a partir da imagem; na segunda etapa, um sistema difuso utilizando o modelo TSK (*Takagi-Sugeno-Kang*) de ordem zero integra as informações dos ângulos dos objetos com os dados dos sensores de distância.

Este artigo possui a seguinte estrutura: na Seção 2, alguns trabalhos correlatos são apresentados. Na Seção 3, aspectos sobre reconhecimento de padrões em imagem e o classificador utilizado são abordados. A Seção 4 descreve algumas arquiteturas de fusão de sensores. A arquitetura proposta, bem como seus módulos são apresentados na Seção 5 e a Seção 6 apresenta os resultados gerais obtidos em cada etapa. Por fim, conclusões são propostas na Seção 7.

2 Trabalhos Correlatos

Wallraven et al. (2003) relatam um estudo sobre aplicações de SVM em imagens através de dois tipos de representações distintas: pixels e histo-

grama, demonstrando o alto desempenho da técnica. M. Quinlan and Middleton (2003) aplicam uma SVM para a classificação de cor em um robô Sony Aibo. Uma SVM é empregada para cada cor a ser reconhecida e uma tabela de busca é criada para ser utilizada em tempo de execução.

Na área de fusão de sensores em robótica móvel, Bastos et al. (1999) e Bonci et al. (2002) apresentam um trabalho de integração de dados de sonares e uma câmera CCD para detecção e localização de objetos. Neste último, o sistema proposto é aplicado para detecção de segmentos de linhas do ambiente e pode ser visto como uma perspectiva futura ao trabalho aqui apresentado.

O objetivo da robótica móvel aplicada ao ambiente de futebol de robôs é o estudo de técnicas que podem ser aplicadas em problemas significativos. Nesse contexto, pode-se citar o trabalho de Nashman et al. (1998). Os autores propõem um sistema para inspeção dimensional de máquinas utilizadas para medição de coordenadas de peças mecânicas, através da fusão de dados de uma câmera e sensores de toque.

3 Visão Artificial

A visão é considerada um dos sentidos mais completos do ser humano, pois provê uma vasta quantidade de informações indispensáveis para a interação do homem com o ambiente. Atualmente, muitas aplicações utilizam sistemas de visão artificial. Sensores de visão juntamente com softwares com funcionalidades específicas, são capazes de extrair informações do ambiente a fim de capacitar máquinas para a tomada de decisão.

Na área de robótica móvel, sistemas utilizando visão artificial são essenciais para garantir uma navegação segura ao robô móvel, através do reconhecimento e localização dos objetos. Tais sistemas são muito eficientes devido a alta densidade de informação espaço-temporal fornecida, ou seja, maior quantidade de bits em uma única aquisição. Estes sistemas são também mais complexos devido à necessidade de extração de dados a partir das imagens obtidas. Sob esse ponto de vista, um método adequado deve ser selecionado para a classificação dos padrões de imagens.

No sistema de visão proposto, pelo fato do ambiente de futebol de robôs ter representações de cores bem definidas para cada objeto, a escolha do classificador utilizado é ponderada entre dois métodos determinísticos: RNA (Redes Neurais Artificiais) e SVM.

Uma SVM possui as mesmas vantagens de uma RNA. No entanto, as SVMs vêm demonstrando, nos últimos anos, maior desempenho para algumas aplicações (Wallraven et al., 2003). As principais desvantagens das RNAs são a susceptibilidade a *overfitting*, ou seja, a possibilidade de uma rede não generalizar (“decorar”), e pou-

cas definições sobre a estrutura desta para cada problema. Devido a natureza convexa e quadrática das soluções para os hiperplanos de separação, as chances de ocorrência de *overfitting* são reduzidas em SVMs e o grau de generalização também é maior que uma RNA.

3.1 Support Vector Machine

SVM é uma técnica híbrida entre uma abordagem estatística e determinística. Isto significa que, para encontrar o melhor espaço de hipóteses de classificação, uma distribuição de probabilidade é determinada a partir do espaço de entrada. A técnica tem origem no trabalho de Vapnik (1995) sobre o Princípio da Minimização de Risco, na área de Aprendizagem Estatística.

A técnica utiliza os seguintes estágios: determinação de linearidade da separação do espaço; caso espaço linear, determina-se os hiperplanos de separação através de um problema de otimização (Eq. (1)); caso espaço não linear, uma função de *kernel* é aplicada e o novo espaço, de dimensão maior que o original, é denominado de espaço de característica.

$$\begin{aligned} \text{minimizar}_{w,b} \quad & < w \cdot w >, \\ \text{sujeito a} \quad & y_i(< w \cdot x_i > + b) \geq 1 \quad (1) \\ & i = 1, \dots, \ell \end{aligned}$$

A Fig. 1 ilustra a aplicação de um *kernel* no espaço de entrada. No espaço de característica, os hiperplanos determinados para a separação do espaço são lineares.

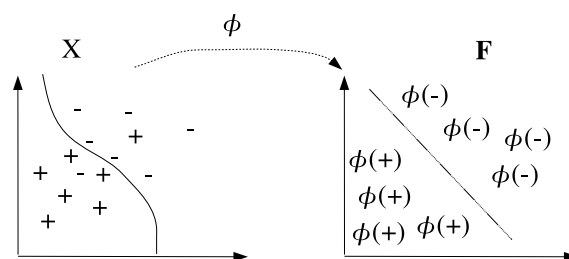


Figura 1: Mapeamento de um espaço de entrada não-linearmente separável para um espaço de características

Outro aspecto importante é a relação de dimensão VC (*Vapnik-Chervonenkis*), número de vetores de suporte e generalização da classificação. Para se determinar a dimensão VC de um espaço é necessário determinar também o número de *Vetores de Suporte*; neste caso, encontrar os hiperplanos orientados, dentro do menor espaço de hipótese definido pela função de estimação de risco, que determinam a fronteira dos dados classificados.

A probabilidade de se classificar erroneamente os dados de teste, depois de encontrados os vetores de suporte, pode ser dada pela Eq. (2):

$$E_{\ell}[P(error)] \leq \frac{E[nVS]}{\ell} \quad (2)$$

onde E_{ℓ} denota a esperança matemática sobre o conjunto de dados de entrada de tamanho ℓ e nVS é o número de vetores de suporte.

Pela Eq. (2), quanto menor o número de vetores de suporte da SVM, maior será o seu grau de generalização, obviamente respeitando um limite mínimo, que será determinado empiricamente.

4 Fusão de Sensores

Fusão de dados de sensores, ou simplesmente fusão de sensores, representa o processo de combinação de dados ou informações de múltiplos sensores para estimativa ou predição de estados de entidades (objetos de medição) (Hall and Llinas, 2001).

Alguns objetivos para diversos cenários de aplicações de fusão de sensores podem ser citados: detecção da presença de um objeto ou condição ambiental, reconhecimento de um objeto, rastreamento de objetos ou monitoramento de um evento continuado, combinação de informações para tomada inteligente de decisão, entre outros

A combinação de dados de múltiplos sensores podem apresentar três vantagens imediatas:

- **Redundância** – dados provenientes de vários sensores semelhantes provêm N observações independentes e a combinação destas resultará em uma estimativa melhorada das medidas, assumindo que estes dados serão combinados de maneira eficiente.
- **Cooperatividade** – a disposição relativa e a movimentação de múltiplos sensores auxilia no processo de inferência na observação. Por exemplo, dois sensores que medem direção angular de um objeto pode determinar a posição espacial deste por triangulação.
- **Complementaridade** – dados provenientes de dois tipos de sensores diferentes podem ser combinados a fim de se obter uma estimativa mais completa do estado de uma entidade. Esta estimativa não poderia ser obtida por qualquer um deles isoladamente.

Na área de fusão de sensores, o principal órgão que regulamenta a terminologia é o JDL (*Joint Directors of Laboratories Data Fusion Workgroup*). Este órgão foi estabelecido em 1986 e define um modelo de processos para fusão de dados, através de níveis de informação, além de um dicionário de termos relativos à área (Hall and Llinas, 2001). A JDL determina três níveis para fusão de sensores: dados, características e decisão.

Um interessante trabalho na área de fusão de sensores é apresentado por Bruder et al. (1991). O trabalho é dedicado a questão temporal da fusão de sensores e quatro tipos são sugeridos: centralizada, descentralizada, sequencial e estática. Os dois últimos tipos são também descentralizados. Caso o sistema de sensoramento seja composto somente por sensores passivos, ou um sensor ativo e vários sensores passivos, a fusão centralizada é a mais adequada, ou seja, a linha de tempo de um sensor determina o processo de fusão. Quando somente sensores ativos são utilizados, a fusão descentralizada (seja esta pura, sequencial ou estática) é a mais comumente utilizada. Na fusão descentralizada, a integração dos dados dos sensores é mais complexa, devido à necessidade de alinhamento temporal.

Outro importante trabalho sobre uma arquitetura flexível para fusão de sensores é proposto por Dasarathy (1997). O autor propõe seis categorias baseadas nos três níveis da JDL, porém utilizadas como modos de processamento de entrada/saída.

5 Arquitetura Proposta

O sistema proposto é implementado utilizando a linguagem C++ em ambiente Linux, explorando as características de modularização e reutilização de código. A arquitetura do sistema, ilustrada na Fig. 2, consiste em quatro módulos principais: aquisição de dados, extração de característica/classificação de padrões, fusão de dados dos sensores e representação interna de conhecimento dos objetos.

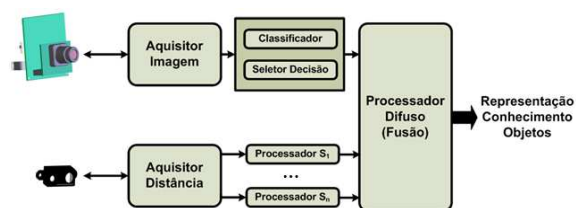


Figura 2: Arquitetura do sistema de visão

Os módulos *Aquisitores* são responsáveis pela extração dos dados da câmera e sensores de distância. Uma vez obtido um quadro de imagem, este é classificado através de uma SVM. O centróide, de cada objeto extraído, é calculado e enviado para o *Seletor de Decisão* a fim de se obter informações como ângulo (θ) relativo ao robô e distância estimada (d) de cada objeto. Ao mesmo tempo, os dados adquiridos de todos os sensores de distância são separados e enviados para cada *Processador* S_i , onde $i \in [1, 5]$, e uma função de interpolação é aplicada com o objetivo de encontrar a distância a partir do valor decimalizado de

cada um destes. Por fim, todas as informações extraídas são enviadas para o *Processador de Fusão* a fim de gerar uma representação de conhecimento de cada objeto da cena. A localização dos objetos reconhecidos na imagem é dada através de suas coordenadas polares relativas ao centro do robô (θ, d).

Os processadores S_i , bem como o Classificador e Seletor de Decisão, são implementados em duas *treads* distintas e sincronizados no Processador de Fusão, uma vez que a fusão é realizada a partir de sensores ativos.

5.1 Aquisição

A câmera utilizada é a CMUCam2, projetada pela Universidade Carnegie-Mellon e consiste de um microprocessador SX52 fabricado pela Ubicom e uma câmera CMOS OV6620. Os dados deste sensor podem ser obtidos tanto por uma interface RS232 quanto por uma porta serial TTL (a primeira opção é utilizada neste projeto). A comunicação com a CMUCam2 ocorre através de comandos no formato ASCII.

A fim de reduzir ao máximo, sem perda de qualidade no reconhecimento, o tempo de aquisição de um quadro da câmera, a imagem obtida é redimensionada para 44x36 pixels. O tempo de aquisição para este quadro é de 245 ms em média e é determinado pelo hardware da câmera.

Os sensores de distância utilizados são os GP2D02 fabricados pela Sharp e utilizam a tecnologia de infravermelho. O protocolo para aquisição dos dados destes sensores é ilustrado pelo gráfico da Fig. 3. O tempo para aquisição de um valor de distância é de 75 ms.

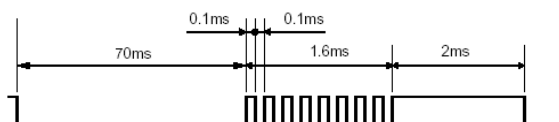


Figura 3: Protocolo de comunicação do sensor Sharp GP2D02

5.2 Reconhecimento e localização de objetos

O módulo de Classificação de padrões de imagem e Seleção de Decisão é responsável pelo processamento dos dados provenientes do módulo Aquisitor da imagem. A classificação de padrões na imagem tem como objetivo a identificação dos objetos móveis no campo de futebol. Esses objetos podem ser: robôs do time, robôs adversários e bola.

A CMUCam2 oferece uma funcionalidade de rastreamento de cor. A técnica utilizada é a de limiar direto e consiste em rastrear determinados pixels cujas cores pertencem ao intervalo R_{min} ,

G_{min} , B_{min} e R_{max} , G_{max} , B_{max} , no espaço RGB. A desvantagem dessa funcionalidade oferecida pela câmera é a alta sensibilidade à variação de luminosidade do ambiente. Em vista disso, uma SVM é utilizada a partir da biblioteca proposta em (Chang and Lin, 2001) com o objetivo de implementar um classificador mais robusto às variações de luminosidade.

5.2.1 Implementação do Classificador

Para a aplicação de uma SVM no espaço de cor da imagem obtida, quatro hipóteses são analisadas, a fim de avaliar o melhor espaço para classificação: Imagens no espaço de cor RGB, imagens no espaço de cor YCrCb, imagens no espaço de cor RGB com segmentação e imagens no espaço de cor YCrCb com segmentação.

Inicialmente, imagens nos espaços RGB e YCrCb sem nenhum tipo de segmentação são avaliadas quanto ao *kernel* de menor número de vetores de suporte, na fase de treinamento. Neste estágio, os objetos são rotulados por cores distintas (azul para bola, roxo para corpo do robô, marrom para robô do mesmo time e vermelho para robô adversário) através da interface gráfica de usuário, utilizando o *mouse*.

Em seguida, a partir do conjunto de *pixels* rotulados é gerado um arquivo de treinamento. As cores neste arquivo são normalizadas no intervalo de $[-1, 1]$ e submetidas a um algoritmo de seleção de parâmetros do *kernel* da SVM, denominado *K-fold cross validation*.

Esta fase é responsável pela determinação do tipo de *kernel* mais adequado. A Tabela 1 demonstra o resultado de número de vetores de suporte obtidos para cada amostra nos espaços RGB e YCrCb:

Tabela 1: Determinação do *kernel* adequado

RGB		YCrCb	
Kernel	nVS	Kernel	nVS
RBF	1311	RBF	656
linear	1295	linear	1220
polinomial	1421	polinomial	1313
sigmóide	1347	sigmóide	1553

Pode-se observar, portanto, que o *kernel* RBF utiliza a menor quantidade de vetores de suporte, 656. Portanto, de acordo com a Eq. (2), possui a melhor generalização. Devido a quantidade de pixels da bola, da marca do robô do mesmo time e da marca do robô do time adversário ser menor que a quantidade de pixels do corpo dos robôs, são aplicados diferentes pesos para os diversos objetos: 10 para a bola, 25 para a marca do robô do mesmo time e para o do time adversário e 1

para o corpo do robô. Estes pesos aplicados durante a fase de treinamento, não altera o número de vetores de suporte totais, porém auxilia na obtenção de vetores de suporte em áreas limitrófes, que melhoram a classificação.

Uma vez definido o melhor *kernel*, duas imagens no espaço RGB e YCrCb são segmentadas utilizando os algoritmos de normalização compreensiva proposto por Finlayson et al. (1998) e a segmentação por deslocamento de média, proposto por Comaniciu and Meer (2002). Em seguida, cada imagem é treinada com uma SVM, utilizando *kernel* RBF, e parâmetros definidos pelo algoritmo de *cross validation*. A Tabela 2 é construída para avaliação do número de vetores de suporte e, consequentemente, da qualidade da classificação.

Tabela 2: Resultados em diferentes espaços de imagem

RGB		YCrCb	
Imagem	nVS	Imagem	nVS
Pura	694	Pura	656
Compreensão normalizada	965	Compreensão normalizada	752
Deslocamento de média	1705	Deslocamento de média	1111

Apesar de serem algoritmos reconhecidamente eficientes para segmentação de imagens, a sua utilização não auxilia o trabalho do classificador. Além disso, o próprio espaço YCrCb já normaliza, de uma certa forma, o espaço RGB, diminuindo os efeitos de variação de luminosidade, o que explica, portanto, a melhor classificação utilizando imagens em YCrCb puro.

Por fim, uma etapa de pós classificação dos pixels classificados é realizada. Esta etapa é responsável por diminuir o número de pixels classificados fora do contorno dos objetos e, portanto, definidos como ruído de classificação. O objetivo principal dessa etapa é aumentar a precisão da determinação dos atributos de centróide e altura dos objetos. O algoritmo de pós classificação é ilustrado a seguir (TH é um limiar definido pelo especialista). Ao final do processo, todos os objetos de um quadro são reconhecidos e localizados num tempo de aproximadamente 55 ms.

```

Dividir a imagem em N subregiões ( $N = 6$ )
Para cada região, calcular a frequência  $F_i$  de
pixels de cada classe  $c_i$ 
if  $F_i \leq TH$  ( $TH = 3$ ) then
    Retirar os pixels da classe
end if

```

5.3 Localização dos objetos

A partir das características extraídas da imagem (área do objeto, altura, largura, centróide), a localização dos objetos é determinada através de suas coordenadas polares (θ, d), onde θ representa o ângulo e d a distância relativos ao centro do robô. Para isso, duas SVM responsáveis por efetuar uma regressão são utilizadas: uma para a determinação do ângulo do objeto e realiza um mapeamento de cada pixel a seu ângulo correspondente; outra para a distância e realiza um mapeamento da altura de cada objeto na imagem e sua distância ao robô.

Para a obtenção da função pixel-ângulo, raios e linhas paralelas são definidos em uma folha branca de papel. A bola é colocada na intersecção entre cada raio e linha, onde a origem dos raios é o centro do robô. A Fig. 4 ilustra esta situação. Um conjunto de pares (pixel, ângulo) é então obtido a partir do centróide de cada objeto e a respectiva localização do ângulo relativo ao centro do robô é determinada.

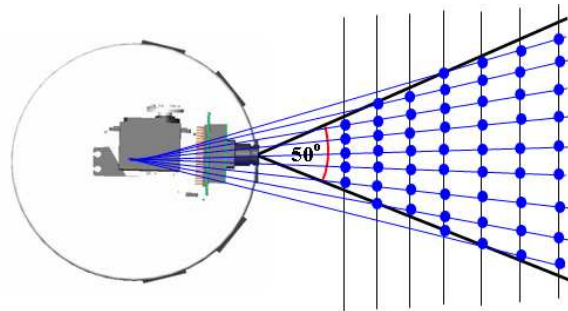


Figura 4: Determinação da relação entre pixel e ângulo

A cabeça de visão artificial proposta é controlada através de dois servo-motores e, portanto, possui dois graus de liberdade de movimentação (movimento panorâmico através da rotação pelo eixo y do sistema de coordenadas da câmera e movimento vertical através de rotação pelo eixo z). A fim de reduzir o efeito da distorção angular da imagem, ocorrido quando a cabeça de visão artificial se encontra em diferentes ângulos em relação ao eixo x, um valor de ângulo vertical (da cabeça) é adicionado ao par (pixel, ângulo) obtido anteriormente. A Fig. 5 ilustra os ângulos utilizados. Para cada ângulo (20° , 35° , 50° e 65°) um conjunto de pares (pixel, ângulo) é determinado e a tupla (pixel, ângulo, ângulo da cabeça) é submetida ao classificador.

Para o treinamento, uma SVM de *kernel* linear e um modelo com 30 vetores de suporte são utilizados, escolhida através da mesma avaliação feita para a classificação em relação ao número de vetores de suporte por *kernel* treinado.

Para a determinação da distância, uma SVM de *kernel* linear é novamente utilizada, agora sem

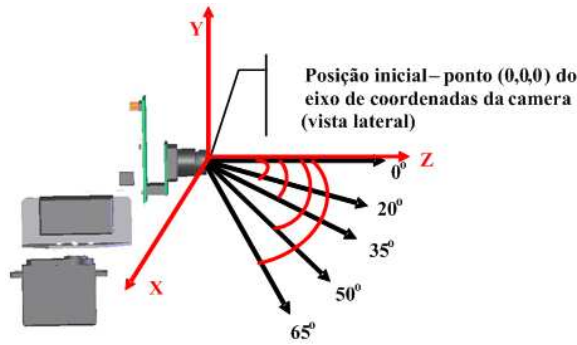


Figura 5: Ângulos da cabeça móvel utilizados para a determinação da função de mapeamento pixel-ângulo. O ângulo zero representa a posição inicial da câmera quando o robô é ativado

a necessidade de avaliação de precisão, visto que o próprio método de determinação da distância é de baixa confiabilidade e só se aplica em circunstâncias especiais.

A informação de distância extraída da imagem é utilizada apenas como estimativa bruta, caso algum objeto se encontre em áreas de sombra dos sensores de distância. De fato, a estimativa de distância funciona apenas como um valor de referência para a localização do robô em relação aos obstáculos móveis. À medida que o robô se locomove, este pode refinar as informações de distância rotacionando seu corpo, por exemplo, a fim de obter uma medida mais confiável.

5.4 Processador dos sensores de distância

O processador dos sensores de distância é responsável em separar os dados obtidos pela *thread* de aquisição dos dados dos sensores e determinação da distância correspondente aos valores decimalizados de cada sensor.

Uma vez obtidos os dados dos sensores, estes são submetidos a uma função polinomial de interpolação. A faixa útil de operação destes sensores é de 10 a 80 cm. Para uma interpolação com maior precisão, dois polinômios são definidos em duas faixas: um para a faixa de 10 a 35 cm e outro para a faixa de >35 até 80 cm.

5.5 Processador de fusão

Uma vez selecionados os valores de decisão de cada sensor (câmera e sensores de distância), estes são submetidos ao Processador de Fusão. A fusão dos dados de sensores é realizada através duas etapas distintas: seleção do ângulo e distância a partir da imagem, e fusão entre estes valores de decisão e os valores obtidos dos sensores de distância. A primeira etapa é fundamental a fim de sincronizar as linhas de tempo dos sensores. Visto que o sensor de visão é o mais lento dos sensores utili-

zados e possui maior densidade de dados espaço-temporais por aquisição, este é utilizado como sincronizador do processado de fusão.

A linha de tempo das duas classes de sensores utilizados é ilustrada na Fig. 6. O tempo de aquisição de um valor de distância e conversão para centímetros é de aproximadamente 75 ms. Isto corresponde a aproximadamente quatro leituras de distância em um processamento de um quadro de imagem e seleção de variáveis de decisão na imagem (300 ms). Em vista disso, as três primeiras leituras dos sensores de distância são descartadas e somente a última representa as distâncias dos objetos reconhecidos. Isto é ilustrado através de linhas pontilhadas na Fig. 6.

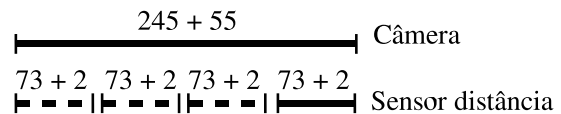


Figura 6: Linha de tempo dos sensores envolvidos no processo de fusão (aquisição + processamento)

O alinhamento temporal das informações dos sensores é necessário, pois os sensores utilizados são do tipo ativo. Após realizado este alinhamento, os valores de decisão são utilizados para a fusão de sensores propriamente dita. Para isso, os conjuntos difusos são determinados a partir da disposição física dos sensores de distância no protótipo, conforme Fig. 7.

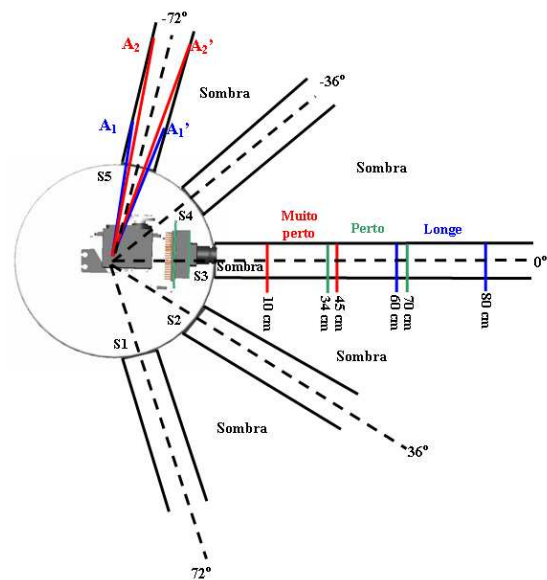


Figura 7: Determinação dos conjuntos difusos a partir da disposição dos sensores de distância

Na Fig. 7, o sensor três (S3) é utilizado como referência ilustrativa dos conjuntos difusos utilizados para o processo de fusão através da variável

linguística distância em todos os sensores. A distância e o ângulo finais são determinados através de um sistema difuso utilizando o modelo de TSK de ordem zero. A razão da escolha é determinada pelo desempenho deste sistema para processamento de amostras de dados (Jang et al., 1997).

Uma vez determinado o ângulo através da câmera, seu valor pode ser refinado através dos sensores de distância. Uma vez que a disposição dos sensores de distância é conhecida, por questões de simplicidade, os ângulos são fornecidos a partir da linha imaginária (linha hachurada na Fig. 7) que parte da origem do robô e passa por cada ponto central dos sensores de distância. Pode-se observar na Fig. 7, que quanto mais longe o objeto se encontra do sensor de distância, mais preciso seu ângulo é determinado. Isto se verifica da seguinte forma: já que o espalhamento de cada sensor de distância é realizado através de um cilindro de 3 cm de largura (espaço entre o receptor e emissor), um objeto entre A_1 e A'_1 possuirá um ângulo menos preciso do que um objeto entre A_2 e A'_2 .

Portanto, a partir da geometria dos sensores, obtém-se os conjuntos difusos ilustrados na Fig. 8.

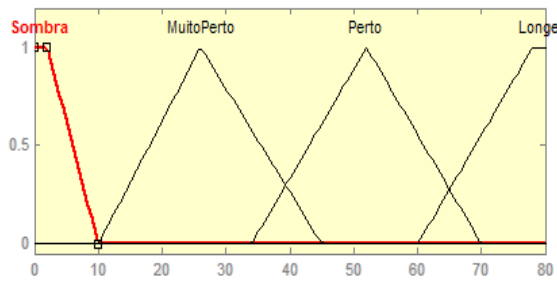


Figura 8: Conjuntos difusos para a variável linguística Distância

A partir dos conjuntos difusos, oito regras são propostas:

R_1 : SE distância = SOMBRA ENTÃO distância = $d(\text{camera})$

R_2 : SE distância = SOMBRA ENTÃO ângulo = $a(\text{camera})$

R_3 : SE distância = MUITOPERTO ENTÃO distância = $d(S_i)$

R_4 : SE distância = MUITOPERTO ENTÃO ângulo = $a(\text{camera})$

R_5 : SE distância = PERTO ENTÃO distância = $d(S_i)$

R_6 : SE distância = PERTO ENTÃO ângulo = $a(\text{camera}) * 0,5 + a(S_n) * 0,5$

R_7 : SE distância = LONGE ENTÃO distância = $d(S_i)$

R_8 : SE distância = LONGE ENTÃO ângulo = $a(S_i)$

As funções $a(\cdot)$ e $d(\cdot)$ representam, respectivamente, ângulo e distância obtidos pela câmera

e pelos sensores de distância. Os valores reais das distâncias e ângulos, após a avaliação das regras, são determinados através da Eq. (3).

$$S = \frac{\sum \psi_i z_i}{\sum \psi_i} \quad (3)$$

onde ψ_i é a T-norma ou T-conorma de cada antecedente e z_i é o resultado da função $f(x, y)$, responsável por descrever a relação entre os conjuntos difusos do antecedente.

6 Análise dos Resultados

São analisados três parâmetros com o objetivo de avaliar o desempenho geral do sistema: para o reconhecimento dos objetos, a aplicação do classificador em diferentes faixas de luz; para a localização dos objetos reconhecidos, o erro quadrático e o desvio padrão médio dos ângulos encontrados.

Para a avaliação do classificador, a Tabela 3 ilustra a taxa de acerto para diferentes valores de luminosidade. A faixa de luminosidade testada se encontra entre 570 and 980 luxs e está de acordo com as regras da Robocup que define uma iluminação com variação entre 700 e 1000 luxs, para cada torneio.

Tabela 3: Resultados em diferentes faixas de luz

Iluminação (lux)	Precisão
570	87.75%
660	84.08%
780	84.87%
800	86.36%
920	87.72%
980	90.00%

As Figs. 9(a) e 9(b) ilustram os resultados da classificação.

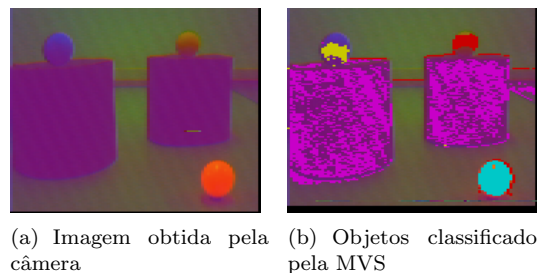


Figura 9: Exemplos de padrões

Para a avaliação do método de determinação das medidas de ângulos do objeto em relação ao robô, duas medidas são determinadas: o erro quadrático entre o valor real e o determinado, e o desvio padrão entre as medidas de ângulo medido e real para cada posição escolhida da cabeça de

visão. O primeiro valor é de 5,16%. Os valores para os desvios padrões de 6 medidas aleatórias (determinadas e reais) são ilustrados na Tabela 4.

Tabela 4: Desvios padrões

Ângulo da cabeça	Desvio médio
20	2,34
50	2,00
65	2,18

Os parâmetros analisados anteriormente apontam, portanto, a uma robustez de reconhecimento de um objeto e eficiência em sua localização. Através do processo de fusão, os valores obtidos podem ser refinados em cada instante de atuação do robô.

7 Conclusões

Um sistema de visão artificial aplicado em robótica móvel, baseado em fusão de sensores, é apresentado neste artigo. A técnica SVM é aplicada tanto para o reconhecimento de padrões quanto para a extração de atributos de decisão dos objetos na imagem. Para a fusão de sensores, a lógica difusa é utilizada.

Um ponto crítico do sistema é o tempo de aquisição de um quadro de imagem e deve ser avaliado e alterado através de circuitos dedicados e um processador de imagem mais rápido.

As contribuições deste trabalho são: um sistema com arquitetura modular, facilitando a reutilização e inclusão de módulos no sistema; reconhecimento de objetos robusto a variação de luminosidade, com apenas uma amostra de treinamento e um tempo de classificação eficiente através de uma SVM aplicada ao espaço de cor YCrCb; um sistema adaptativo de informações dos sensores utilizando lógica difusa aplicada ao Processador de Fusão.

Por fim, pode-se afirmar que os resultados gerais apontam para um bom desempenho do sistema.

Referências

- Bastos, T., Sarcinelli, M. and Freitas, R. (1999). A multi-sensorial integration scheme to help mobile robot navigation through obstacle recognition, *International Conference on Emergin Technology and Factory Automation*, Vol. 1, IEEE, EUA, pp. 549–558.
- Bonci, A., Leo, T. and Longhi, S. (2002). Ultra-sonic and video data fusion for mobile robot navigation, *Proceedings of the 10th Mediterranean Conference on Control and Automation*.
- Bruder, S., Faronq, M. and Bayoumi, M. (1991). Robotic heterogenous multi-sensor fusion with spatial and temporal alignment, *Proceedings of the 30th IEEE Conference on Decision and Controle*, IEEE, pp. 506–511.
- Chang, C.-C. and Lin, C.-J. (2001). Libsvm: A library for support vector machines. Biblioteca disponível em: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis, in IEEE (ed.), *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 24, pp. 603–619.
- Dasarathy, B. (1997). Sensor fusion potential exploitation – innovative and illustrative applications, *Proceedings of the IEEE Special Issue on Sensor Fusion*, Vol. 85, IEEE, pp. 24–38.
- Finlayson, G., Schiele, B. and Crowley, J. (1998). Comprenhesive colour normalization, *ECCV'98 Fifth European Conference on Computer Visio*.
- Hall, D. and Llinas, J. (2001). *Handbook of multi-sensor data fusion*.
- Jang, J., Sun, C. and Mizutani, E. (1997). *Neuro-fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice Hall.
- Kortenkamp, D., Bonasso, R. and Murphy, R. (1998). *Artificial Intelligence and Mobile Robots*, Menlo Park.
- M. Quinlan, S. C. and Middleton, R. (2003). Techniques for improving vision and locomotion on the sony aibo robot, *Australasian Conference on Robotics and Automation (ACRA)*.
- Nashman, M., Yoshimi, B., Hong, T., Rippey, W. and Herman, M. (1998). A unique sensor fusion system for coordinate measuring machine tasks, *Technical Report 172*, International Society for Optical Engineering.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*, Springer Verlag.
- Wallraven, C., Caputo, B. and Graf, A. (2003). Recognition with local features: the kernel recipe, *IEEE International Ninth Conference on Computer Vision*, Vol. 1, pp. 257–264.