Reconhecimento de objectos 3D a partir de imagens 2D usando protótipos

Raquel César, Nº 46020
raquelcesar@netcabo.pt

Instituto Superior Técnico
Engenharia Informática e de Computadores
Engenharia Biomédica 2002

RESUMO

O reconhecimento de objectos encontra-se no topo de uma hierarquia de tarefas visuais. Na sua forma geral, este é um problema computacional muito difícil, que desempenhará, provavelmente, um papel significativo na eventual construção de máquinas inteligentes.
Um número cada vez maior de resultados de estudos comportamentais e neurofisiológicos vem dar suporte à ídeia de que os seres humanos representam internamente os objectos tridimensionais na forma de um conjunto reduzido de imagens bidimensionais. Neste trabalho apresentamos um esquema para reconhecimento de objectos 3D a partir de imagens 2D. O esquema proposto começa por identificar a classe do objecto observado e só depois procura determinar a sua identidade individual. Desta forma, diminuem-se os custos computacionais de uma comparação exaustiva com todos os objectos conhecidos. Por outro lado, parte do processamento efectuado na fase de categorização pode ser reutilizado na fase de identificação. O sistema desenvolvido não possui qualquer conhecimento prévio e constrói a base de objectos enquanto vai funcionando.

INTRODUÇÃO

Cada objecto tridimensional pode produzir padrões de excitação na retina consideravelmente diferentes, dependendo da posição do objecto relativamente ao observador. Apesar disto, somos capazes de perceber que estes sinais diferentes são produzidos pelo mesmo objecto. Esta capacidade de reconhecimento constante a partir de tais sinais de entrada inconstantes é-nos conferida pela capacidade que o nosso cérebro possui de estabelecer representações internas dos objectos. A natureza de tais representações invariáveis ao ponto de vista e a forma como elas podem ser adquiridas é ainda um dos maiores problemas por resolver em neurociência e em visão por computador.

Existe um número incontável de estudos comportamentais com primatas que suportam o modelo de uma representação dos objectos tridimensionais baseada em vistas pelo nosso sistema de visão. Se apresentarmos a um humano um conjunto de vistas de objectos desconhecidos, o seu tempo de resposta e as taxas de erro durante o reconhecimento crescem com o aumento da distância angular entre o objecto aprendido e a vista desconhecida [11]. Este efeito diminui se forem consideradas vistas intermédias. O desempenho não depende linearmente da menor distância angular em três dimensões à vista melhor reconhecida mas correlaciona-se de forma significativa com a distância entre a vista apresentada e a "melhor" vista (menor tempo de reconhecimento e menor taxa de erro) em termos da deformação, no plano bidimensional da imagem, de um conjunto de características identificativas do objecto [2].

Desta forma, a medição da semelhança entre planos de imagem e alguns padrões de características parece ser um modelo apropriado para o processo de reconhecimento humano de objectos tridimensionais. Experiências com macacos mostram que a familiarização com um número limitado de vistas de um novo objecto pode dar origem a reconhecimento independente do ponto de vista. Vários estudos fisiológicos também fornecem evidência de um processamento baseado em vistas pelo cérebro durante o reconhecimento de objectos. Resultados de medições em neurónios no cortex temporal inferior dos macacos, que se sabe estar relacionado com o reconhecimento de objectos, suportam os resultados dos estudos comportamentais. Foram encontradas populações de neurónios no cortex inferior temporal que respondem selectivamente a apenas algumas vistas de um objecto e cuja resposta diminui à medida que o objecto é rodado, afastando-se de um ponto de vista preferencial [7].

Em suma, podemos dizer que a representação de objectos na forma de vistas únicas ligadas entre si parece ser suficiente para uma vasta variedade de situações e tarefas de percepção.

O trabalho aqui apresentado descreve uma tentativa de incorporação de reconhecimento de objectos tridimensionais a partir de imagens bidimensionais partindo de trabalho apresentado em [14]. O esquema considerado baseia-se na projecção ortográfica de objectos 3D em imagens 2D e é composto por duas fases. Na primeira fase, a fase de categorização, a imagem é comparada a objectos protótipo. Para cada protótipo determina-se a vista que mais se aproxima da imagem e, se essa vista for semelhante à imagem, classifica-se o objecto na classe representada pelo protótipo. Na segunda fase, a fase de identificação, o objecto observado é comparado com os modelos individuais da sua classe. Cada classe agrupa objectos com formas relativamente próximas. Para cada modelo procura-se uma vista que coincida com a imagem. No caso de se encontrar uma vista nestas condições, a identidade específica do objecto é determinada. O processo de categorização do objecto (antes da identificação) oferece duas vantagens essenciais:

em primeiro lugar, a imagem é comparada com um número menor de modelos, já que apenas é necessário considerar modelos que pertencem à mesma classe que o objecto; em segundo lugar, o custo de comparar uma imagem com cada modelo de uma classe é muito reduzido porque as correspondências são computadas uma única vez para toda a classe. Mais concretamente, as correspondência e pose do objecto computadas no processo de categorização para alinhar a imagem com o protótipo são reutilizadas no estágio de categorização para alinhar os modelos individuais com a imagem. Desta forma, a identificação reduz-se a uma série de comparações simples.

Este processo de reconhecimento segue de perto o esquema proposto por Basri [8]. No entanto, diferencia-se do trabalho aí apresentado porque tentámos desenvolver um processo para reconhecimento em que a base de conhecimento fosse construída de forma incremental, sem a prévia construção/categorização de uma base de imagens.

O sistema não possui qualquer conhecimento prévio e as classes e modelos de objectos vão sendo construídas à medida que novos objectos vão sendo "observados". A ídeia fundamental na base deste procedimento é a seguinte: quando é observada uma nova imagem do objecto, se ela não difere significativamente das vistas já observadas do mesmo objecto, então ela será reconhecida. Se a nova vista reconhecida for suficientemente diferente das vistas armazenadas, podemos guardá-la, juntamente com as restantes vistas. Desta forma, poderemos cobrir todo o espaço das vistas de cada objecto com um número reduzido de imagens, de uma forma incremental. Evidentemente, pode acontecer (e é mesmo provável que aconteça) que duas vistas distintas do mesmo objecto sejam identificadas como pertencendo a objectos diferentes, por se tratar de vistas com poucos pontos em comum. A ídeia, então, é que, em determinado momento, irá surgir alguma nova vista do objecto que se assemelhará a ambos os objectos. Nessa altura, podemos reconhecer que estamos perante o mesmo objecto e unificar as duas representações.

REPRESENTAÇÃO DOS OBJECTOS

Um objecto é modelado por uma matriz *M*, de dimensão *n x k*, onde *n* é o número de pontos característicos e *k,* o número de colunas em *M*, está relacionado com o número de graus de liberdade do objecto.

Este esquema de representação resulta do modelo de combinação linear para objectos 3D proposto por [12]. Neste trabalho é demonstrado que o conjunto de imagens possíveis de um objecto 3D que sofre transformações rígidas e escalamento entre imagens seguidos de projecção ortográfica pertence a um espaço linear gerado por um número restrito de imagens 2D do mesmo objecto.

Seja *O* um objecto 3D que contem *n* pontos característicos $(X_i, Y_i, Z_i), 1 \le i \le n$. Sob projecção perspectiva fraca, a posição do objecto, após uma rotação *R*, translação $\vec{t}$ e escalamento *s*, é dada por

$$\begin{aligned} x_i &= sr_{11}X_i + sr_{12}Y_i + sr_{13}Z_i + st_x, \\ y_i &= sr_{21}X_i + sr_{22}Y_i + sr_{23}Z_i + st_y, \end{aligned} \quad (1)$$

onde r_{ij} são os componentes da matriz de rotação *R*, t_x, t_y são os componentes horizontal e vertical, respectivamente, do vector de translação $\vec{t}$ e *s* é o factor de escalamento.

Denotemos por $\vec{X}, \vec{Y}, \vec{Z}, \vec{x}, \vec{y} \in \Re^n$ os vectores dos valores X_i, Y_i, Z_i, x_i e y_i, respectivamente, e denotemos $\vec{1} = (1,...,1) \in \Re^n$. Então, podemos escrever, sob a forma vectorial

$$\begin{aligned} \vec{x} &= a_1\vec{X} + a_2\vec{Y} + a_3\vec{Z} + a_4\vec{1}, \\ \vec{y} &= b_1\vec{X} + b_2\vec{Y} + b_3\vec{Z} + b_4\vec{1}, \end{aligned} \quad (2)$$

onde

$a_1 = sr_{11}$	$b_1 = sr_{21}$
$a_2 = sr_{12}$	$b_2 = sr_{22}$
$a_3 = sr_{13}$	$b_3 = sr_{23}$
$a_4 = st_x$	$b_4 = st_y$

Ou seja,

$$\vec{x}, \vec{y} \in span\left\{\vec{X}, \vec{Y}, \vec{Z}, \vec{1}\right\}$$

Note-se que a componente de translação pode ser ignorada se os centróides dos pontos (X_i, Y_i, Z_i) e (x_i, y_i) forem deslocado para a origem, isto é, se transladarmos os pontos do objecto e da imagem de forma que

$$\sum_{i=1}^{n}(X_i, Y_i, Z_i) = (0,0,0),$$

$$\sum_{i=1}^{n}(x_i, y_i) = (0,0),$$

Logo, todas as vistas do objecto rígido *O* estão contidas num espaço linear 4D (ou 3D, se ignorarmos a translação). A ídeia, agora, é usar imagens do objecto para construir uma base para este espaço. Mostra-se que, em geral, duas vistas são suficientes [12].

Seja $p_1 = (\vec{x}_1, \vec{y}_1)$ uma imagem 2D de *O* e seja $p_2 = (\vec{x}_2, \vec{y}_2)$ a imagem de *O* que se obtém após uma rotação por *R* (uma matriz *3 x 3*). Considere-se, então, uma nova vista de *O*, $p_3 = (\vec{x}_3, \vec{y}_3)$, obtida por aplicação de uma nova rotação a *O*. Ter-se-á:

$$\begin{aligned} \vec{x}_3 &= a_1\vec{x}_1 + a_2\vec{y}_1 + a_3\vec{x}_2, \\ \vec{y}_3 &= b_1\vec{x}_1 + b_2\vec{y}_1 + b_3\vec{x}_2, \end{aligned} \quad (3)$$

desde que as duas imagens p_1 e p_2 não difiram apenas por uma rotação pura em torno da linha de vista [12].

A utilização da combinação linear de duas vistas descrita é aplicável a transformações lineares gerais do objecto e, sem mais restrições, é impossível distinguir entre transformações rígidas e transformações lineares não rígidas. Para impôr rigidez (com possível escalamento), os coeficientes $(a_1, a_2, a_3, b_1, b_2, b_3)$ devem obedecer a duas restrições simples

$$\begin{aligned} &a_1b_1 + a_2b_2 + a_3b_3 + (a_1b_3 + a_3b_1)r_{11} + \\ &+(a_2b_3 + a_3b_2)r_{12} = 0, \\ &a_1^2 + a_2^2 + a_3^2 - b_1^2 - b_2^2 - b_3^2 = \\ &= 2(b_1b_3 - a_1a_3)r_{11} + 2(b_2b_3 - a_2a_3)r_{12} \end{aligned} \quad (4)$$

em que r_{11} e r_{12} são componentes da matriz de rotação R que podem ser determinados, a menos de um factor de escala, a partir das duas primeiras vistas. Quando estas duas restrições não são satisfeitas são geradas imagens do objecto distorcidas.

Este esquema de combinação linear de imagens assume que os mesmos pontos do objecto estão visíveis em vistas diferentes. Quando as vistas são suficientemente difrerentes esta abordagem deixa de ser válida, devido a auto oclusão. Para representar um objecto a partir de todas as direcções possíveis (por exemplo, visto de frente e de trás), são necessários vários modelos diferentes deste tipo.

Para resumir, seguindo o esquema exposto, um objecto é representado por uma matriz *M* cujas colunas são construídas a partir de vistas do objecto, transladadas por forma a ter o centróide na origem, que formam uma base do espaço 3D.

Vistas do objecto podem ser construídas como se segue

$$\begin{aligned} \vec{x} &= M\vec{a}, \\ \vec{y} &= M\vec{b}, \end{aligned} \quad (5)$$

onde $\vec{a}, \vec{b} \in \Re^k$ são os vectores dos coeficientes na equação (3). Note-se que os dois sistemas lineares podem ser reunidos num só através da construção de uma matriz modelo modificada, da forma seguinte

$$\begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix} = \begin{bmatrix} M & 0 \\ 0 & M \end{bmatrix} \begin{bmatrix} \vec{a} \\ \vec{b} \end{bmatrix} \quad (6)$$

Para objectos rígidos, nem todos os pares de vectores $\vec{a}, \vec{b}$ são válidos, é necessário que os seus componentes satisfaçam as duas restrições quadráticas (4). O reconhecimento envolve a obtenção dos vectores de transformação $\vec{a}, \vec{b}$ e a verificação de que as suas componentes satisfazem as duas restrições. No que se segue as restições são largamente ignoradas, mas elas podem ser verificadas tanto na fase de categorização como na fase de identificação dos objectos.

CATEGORIZAÇÃO

O reconhecimento consiste, antes de mais, na determinação da categoria do objecto através da sua comparação com objectos protótipo que constituem exemplares típicos das suas classes. Para um dado protótipo, obtém-se a vista que apresenta mais semelhanças com a imagem. Essa vista é comparada com a imagem actual e o resultado desta comparação determina a identidade da classe do objecto.

Uma classe de objectos é um par $C = (P, \{M_1, M_2, ..., M_l\})$, onde *P* é um objecto protótipo para a classe e $M_1, M_2, ... M_l$ são objectos modelo. Tanto o protótipo como os modelos são representados por martizes *n x k*, de acordo com a descrição acima.

Uma classe contém objectos com forma idêntica. Estes objectos partilham, grosso modo, a mesma topologia e existe uma correspondência "natural" entre eles. Esta correspondência é explicitada pela ordem dos vectores linha nos modelos. Especificamente, dado um protótipo *P* e modelos $M_1, M_2, ... M_l$, ordenamos as linhas destes modelos de tal forma que o primeiro ponto característico de *P* corresponde ao primeiro ponto característo de cada um dos modelos $M_1, M_2, ... M_l$, o segundo ponto característico de *P* corresponde ao segundo ponto característo de cada um dos modelos $M_1, M_2, ... M_l$, e assim por diante. A importância desta ordenação tornar-se-á evidente adiante.

Para proceder à categorização do objecto observado na imagem, é necessário, antes de mais, alinhar os objectos protótipo com a imagem e compará-los com ela. Para cada protótipo, resolve-se, em primeiro lugar, a correspondência entre o protótipo e a imagem. Em seguida, usando a correspondência determinada, calcula-se a vista do protótipo mais próxima.

Dados um protótipo *P* e uma imagem *I*, geramos um vector $\vec{v}$ a partir da imagem que contem a localização dos pontos característicos da imagem ordenados em correspondência com os pontos do protótipo: o primeiro ponto $\vec{v}$ corresponde ao primeiro ponto em *P* e assim por diante. O vector de transformação $\vec{a}$ que mais aproxima os pontos do protótipo dos pontos da imagem é o vector que minimiza a distância euclidiana entre os pontos do protótipo e os pontos da imagem

$$\min_{\vec{a}'} \left\| P\vec{a}' - \vec{v} \right\|$$

Se *P* é uma matriz sobredeterminada, isto é, se *P* tem dimensão *n x k*, com *n* > *k* e verifica *rank(P) = k*, então a solução da equação acima é dada por

$$\vec{a} = P^+\vec{v} \quad (7)$$

onde $P^+ = \left(P^T P\right)^{-1} P^T$ denota a matriz pseudo-inversa de *P*, e a vista do protótipo mais próxima, $\vec{p}$, é obtida por aplicação de *P* a $\vec{a}$, isto é

$$\vec{p} = P\vec{a} = PP^+\vec{v}$$

A vista $\vec{p}$ é então comparada com a imagem e a sua semelhança determina a classificação do objecto. A qualidade do emparelhamento entre protótipo e imagem é dada por

$$D(P, \vec{v}) = \frac{\left\| \vec{p} - \vec{v} \right\|}{\left\| \vec{v} \right\|} = \frac{\left\| \left(PP^+ - I\right)\vec{v} \right\|}{\left\| \vec{v} \right\|} \quad (8)$$

onde *I* representa a matriz identidade. A divisão pela norma de $\vec{v}$ normaliza a medida (8) permitindo eliminar efeitos devidos ao escalamento do objecto.

Se o objecto pertence à classe representada por *P*, então a função definida por (8) atinge o seu valor mínimo quando $\vec{v}$ está ordenado em correspondência com *P*. Qualquer outra ordenação dos pontos aumentará o valor de *D*. Portanto, a função *D* pode ser utilizada como função objectivo para o problema da determinação da correspondência entre o protótipo e a imagem. Formalmente, denotando por π uma matriz permutação, definimos:

$$\hat{D}\left(P,\vec{v}\right)=\min_{\pi} D\left(P,\pi\vec{v}\right) \quad (9)$$

Se definirmos o custo de emparelhar um ponto p_i na imagem $\vec{p}$ com o ponto q_j na imagem $\vec{v}$ como

$$Cij=\left(p_i-q_j\right)^2$$

então a minimização de (9) é equivalente à minimização da função

$$H\left(\pi\right)=\sum_{i=1}^{n} C_{ij}\left(p_i,q_{\pi(i)}\right) \quad (10)$$

sujeita à restrição de que o emparelhamento seja um para um, isto é, π seja uma permutação. Este problema é uma instância do problema de atribuição quadrado (ou emparelhamento bipartido pesado), que pode ser resolvido em tempo $O(n^3)$ usando o método Hungarian. Na nossa implementação usámos o método mais eficiente de [9]. A entrada para o problema de atribuição é uma matriz quadrada de custos C_{ij} e a saída é uma permutação π tal que (10) é minimizada.

De forma a ter-se um tratamento robusto de pontos sem correspondência, adicionamos pontos "dummy" a cada conjunto de pontos com um custo de emparelhamento constante ε_d. Assim, um ponto é emparelhado com um "dummy" sempre que não existe um emparelhamento real disponível com custo inferior a ε_d. Desta forma, ε_d pode ser encarado como um parâmetro de "threshold" para a detecção de "outliers". De forma análoga, quando o número de pontos nos dois conjuntos não é igual, a matriz de custos pode tornar-se quadrada através da adição de pontos "dummy" ao conjunto de pontos menor.

Um objecto observado numa vista $\vec{v}$ pertence à classe representada pelo protótipo *P* se

$$\hat{D}\left(P,\vec{v}\right)<\varepsilon$$

para uma certa constante $\varepsilon > 0$.

Resumindo, dados um protótipo *P* e uma imagem *I*, a correspondência entre *P* e *I* é resolvida minimizando a medida (9) sobre todas as permutações possíveis de $\vec{v}$ e, se o mínimo obtido estiver abaixo do threshold ε, então a classe do objecto é determinada.

A medida $\hat{D}$ aqui definida determina a semelhança entre o protótipo *P* e a vista $\vec{v}$ usando apenas distâncias entre pontos característicos. Em geral, como é difícil de estabelecer uma correspondência perfeita, esta medida não é robusta. Uma forma de tornar este esquema mais robusto será incorporar na medida de semelhança informação adicional sobre os pontos característicos.

Embora o esquema geral de classificação aqui definido não dependa da escolha específica da métrica de distância, a medida escolhida afecta a divisão dos modelos em classes e a selecção dos protótipos óptimos para essas classes. Mais adiante mostraremos como é possível escolher os protótipos óptimos utilizando a medida especificada por (8).

Como veremos na secção seguinte, o esquema de categorização aqui definido mostra-se útil mesmo quando a categorização do objecto não é possível e é necessário comparar a imagem com todos os modelos existentes. Aí mostra-se como o vector de transformação do protótipo pode ser reutilizado para alinhar a imagem com os modelos específicos. Assim, após a categorização, o custo de comparar a imagem com cada um dos modelos específicos é substancialmente reduzido, pois a parte complicada de recuperar a transformação que relaciona os modelos com a imagem é aplicada apenas aos objectos protótipos.

IDENTIFICAÇÃO

Após a categorização do objecto, procura-se determinar a sua identidade individual. Nesta fase, a imagem é comparada com todos os modelos pertencentes à classe identificada no processo de categorização, ou, se não foi possível identificar a classe do objecto, com todos os modelos existentes. Para cada modelo, determina-se a transformação que alinha o modelo com a imagem, se existir, usando a informação obtida na categorização.

Seja $\vec{v}$ uma vista de um objecto modelo M_i, verificando

$$\vec{v}=M_i\vec{b} \quad (11)$$

para um certo vector de transformação $\vec{b}$. Então, pode-se mostrar sem dificuldade que

$$\vec{b}=A_i\vec{a} \quad (12)$$

onde $\vec{a}$ é o vector transformação do protótipo dado por (7) e $A_i=\left(P^{+}M_i\right)^{-1}$, supondo que $\det\left(P^{+}M_i\right)\neq 0$.

Este resultado é válido porque os pontos característicos no protótipo e nos modelos estão alinhados.

A transformação linear definida pela matriz A_i é independente da vista $\vec{v}$ considerada, ou seja, para qualquer vista do objecto, a mesma transformação mapeia a transformação do protótipo que corresponde a essa vista na transformação do modelo correcta. Isto significa que a transformação A_i pode ser computada à partida e guardada juntamente com o modelo. Mais, a transformação A_i permite recuperar a transformação do modelo independentemente da qualidade do emparelhamento entre o protótipo e a imagem. Isto é, mesmo quando o protótipo alinha mal com a imagem, a transformação que

alinha o modelo com a imagem é determinada correctamente.

Como vimos, A_i existe se P^+M_i é invertível. Esta condição é equivalente a exigir que os dois espaços coluna de P e M_i não sejam ortogonais em nenhuma direcção. Esta condição verifica-se, em geral, desde que os dois objectos sejam relativamente semelhantes.

Denotemos $M_i^{'} = M_i A_i$ o modelo M_i alinhado com o protótipo P. $M_i^{'}$ modela o mesmo objecto que M_i, já que os vectores coluna de ambas as matrizes geram o mesmo espaço. Para além disso, o modelo alinhado $M_i^{'}$ é posto pela transformação do protótipo $\vec{a}$ em alinhamento perfeito com a imagem. De facto, podemos rescrever (11) sob a forma

$$\vec{v} = M_i^{'}\vec{a} \qquad (13)$$

Assim, se os modelos estiverem alinhados com o protótipo, a transformação calculada na fase de categorização pode ser usada para identificação sem mais manipulações. Este resultado permite simplificar o esquema de identificação. Os modelos $M_1, ..., M_l$ são alinhados com o protótipo P aplicando as transformações correspondentes $A_1, ..., A_l$. No reconhecimento, a transformação do protótipo $\vec{a} = P^+\vec{v}$ é aplicada aos modelos alinhados $M_1^{'},...,M_l^{'}$.

Na descrição acima suposemos que existe uma correspondência total entre o protótipo e a imagem. Esta suposição não é, no entanto, mandatória. Se a correspondência não é total, os resultados anteriores continuam válidos desde que se elimine, nas matrizes P e M, as linhas que correspondem a pontos que não têm correspondência na imagem.

CONSTRUÇÃO DE PROTÓTIPOS ÓPTIMOS

Nesta secção mostraremos como é possível determinar os protótipos óptimos para uma dada classe sob a métrica (8).

Dada uma classe de objectos, o protótipo óptimo para esta classe é o objecto que mais se assemelha aos objectos da classe. Na formulação utilizada, um tal objecto deverá partilhar o máximo número possível de pontos característicos com os objectos da sua classe, as posições destes pontos no protótipo deverão estar tão próximas quanto possível das suas posições nos objectos e as transformações protótipo para modelo destes objectos deverão set tão estáveis quanto possível. O protótipo pode ser calculado, então, usando uma análise de componentes principais, isto é, calculando os vectores próprios que correspondem aos valores próprios dominantes de uma certa matriz determinada pelos modelos da classe.

O protótipo óptimo para uma dada classe é definido como o objecto que minimiza a seguinte função de custo

$$E(P) = \sum_{i=1}^{n} \int_{\|\vec{v}_i\|=1} \left\| \left(PP^T - I\right)\vec{v}_i \right\| d\vec{v}_i \qquad (14)$$

que corresponde ao somatório, para todos os modelos da classe, da distância $D\left(P, \vec{v}_i\right)$ a todas as possíveis vistas, de norma unitária, de cada modelo.

Em [8] prova-se que o protótipo que minimiza a equação (14) pode ser obtido usando o seguinte algoritmo:

1 – Verificar que os vectores coluna de cada uma das matrizes dos modelos M_i $(1 \le i \le l)$ são ortonormados. Em caso negativo, aplicar o método de ortonormalização de Gram-Schmidt.

2 – Construir a matriz simétrica $n \times n$:

$$F = \sum_{i=1}^{l} M_i M_i^T \qquad (15)$$

3 – Encontrar os k vectores próprios de F que correspondem aos valores próprios dominantes. A matriz P óptima é construida a partir destes vectores.

O protótipo determinado por este processo é independente da escolha da base para os modelos. Isto implica que, para construir o protótipo, não é necessário que os objectos modelo $M_1, ..., M_l$ estejam alinhados.

IMPLEMENTAÇÃO

A implementação do processamento descrito acima é trivial. O algoritmo implementado consiste nos seguintes passos:

1 – Dada uma imagem I, aplicar o processamento desenvolvido em [14] para identificar os objectos presentes na imagem. Para cada objecto encontrado, obter um vector $\vec{v}$ com a localização dos pontos característicos da imagem e proceder como se segue.

2 – Obter o vector $\vec{v}\,'$ que tem o centróide na origem e resulta de uma translação de $\vec{v}$, dado por

$$\vec{v}\,' = \vec{v} - \sum_{i=1}^{n} \frac{v_i}{n}$$

onde $v_i = (x_i, y_i)$ é um ponto de $\vec{v}$ e n é o número de pontos em $\vec{v}$. Normalizar $\vec{v}\,'$.

3 – Seja P o conjunto de todos os protótipos e seja Cl o conjunto de todas as classes. Se $P = \varnothing$, prosseguir para 7.1.

4 – Para cada protótipo $P_j \in P$ determinar a distância $\hat{D}\left(P_j, \vec{v}\,'\right)$, dada por (9).

5 – Seja $\sigma_j = \arg\min_{\sigma} D\left(P_j, \sigma\vec{v}\,'\right)$. Determinar $d = \min_{j \in \{j : P_j \in P\}} \hat{D}\left(P_j, \sigma_j \vec{v}\,'\right)$.

6 – Se $d < \varepsilon$, determinar

$$\wp = \left\{ j : P_j \in P \wedge \hat{D}\left(P_j, \sigma_j \vec{v}\,'\right) = d \right\}$$

Figura 1 – Primeira e décima vistas do objecto casa.

6.2 – Determinar $d' = \min_{j\in\wp} \min_{i\in\Im_j} \left\| M_i P_j^+ \sigma_j \vec{v}' - \vec{v}' \right\|$,

onde $\Im_j = \left\{ i : C\left(P_j, M\right) \in Cl \wedge M_i \in M \right\}$.

6.3 – Se $d' < \varepsilon'$, determinar

$M = \left\{ \left(j,i\right) : P_j \in \wp \wedge i \in \Im_j \wedge \left\| M_i P_j^+ \sigma_j \vec{v}' - \vec{v}' \right\| = d' \right\}$

6.3.1 – Para cada $P_j \in \left\{ P_j : j \in M \right\}$,

tomar $A = \left\{ M_i : i \in \Im_j \right\} - \left\{ M_i : i \in A_j \right\} \cup \left[M_i \right]_{i\in A_j}$,

onde $A_j = \left\{ i : \left(j,i\right) \in M \right\}$ e $\left[M_i \right]_{i\in A_j}$ representa a matriz formada por todas as colunas das matrizes com índices em A_j.

6.3.2 – Se $d' \geq \varepsilon''$, fazer $A := A \cup \left\{ \sigma_j \vec{v}' \right\}$.

6.3.3 – Se $d' \geq \varepsilon'$, fazer $A = \left\{ M_i : i \in \Im_j \right\} \cup \left\{ \sigma_j \vec{v}' \right\}$.

6.3.4 – Fazer $Cl := Cl - \left(P_j, M\right) \cup \left(P_j', M'\right)$, em que P_j' resulta da aplicação do algoritmo para obtenção do protótipo óptimo ao conjunto A e

$M' = \left\{ M_i \left(\left(P_j' \right)^+ M_i \right)^{-1} : M_i \in A \right\}$.

7 – Se $d \geq \varepsilon$

7.1 – Fazer $Cl := Cl \cup \left(\vec{v}', \left\{ \vec{v}' \right\} \right)$.

O threshold ε', na linha 6.3, é o equivalente, para os modelos, ao threshold ε usado na categorização. O threshold ε'', na linha 6.3.2, destina-se a restringir a inclusão de novas vistas nos modelos. A nova vista não é incluída no modelo a não ser que difira do modelo por um valor superior a ε''. Em todas as simulações aqui reportadas tomou-se $\varepsilon = 0.25$, $\varepsilon' = 0.15$ e $\varepsilon'' = 0.01$ (obviamente, deverá ter-se sempre $\varepsilon'' \leq \varepsilon' \leq \varepsilon$).

RESULTADOS

Para testar a capacidade do sistema reconhecer o mesmo objectos de vários pontos de vista, apresentámos-lhe um conjunto de dez imagens de um modelo 3D de uma casa (Figura 1) obtidas por rotações sucessivas em torno do eixo vertical de 3.6º. Desta forma, a primeira e a última imagem apresentadas diferem entre si por uma rotação no plano horizontal de 36º. Todas as imagens foram reconhecidas como correspondendo a um único objecto. Os resultados da comparação da décima vista com o protótipo e com o único modelo são mostrados nas figuras 2 e 3, respectivamente.

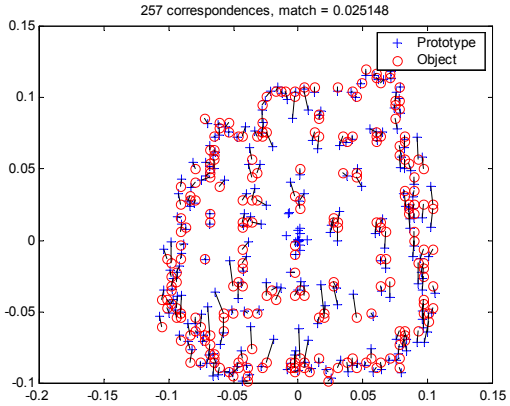

Figura 2 – Resultados da comparação do objecto casa com o protótipo (match = 0.025148).

Seguidamente, apresentámos uma imagem de um outro objecto diferente para verificar se o sistema seria capaz de distinguir entre as dois objectos. Desta vez escolhemos um modelo 3D de um cão (Figura 4). Os resultados obtidos são mostrados na figura 5. Como se pode comprovar, o sistema foi capaz de reconhecer estar na presença de um novo objecto.

DISCUSSÃO

Os resultados obtidos, se bem que em número reduzidos, são encorajadores. Para testar a validade do modelo proposto ter-se-á que efectuar uma bateria de testes mais exigentes. O modelo é muito simples e atractivo do ponto de vista matemático e computacional. Os parâmetros threshold utilizados foram escolhidos sem grande critério e a investigação do modelo exigirá uma pesquisa dos melhores valores a utilizar.

REFERÊNCIAS

[1] C. M. Cyr, B. B. Kimia "3d object recognition using shape similarity-based aspect graph" In ICCV, A aparecer, 2001.

[2] F. Cutzu, S. Edelman "Canonical Views in Object Representation and Recognition". Vision Research, 34:3037-3056, 1994.

[3] G. Peters "Theories of Three-Dimensional Object Perception - A Survey", *Recent Research Developments in Pattern Recognition*, Vol. 1, pp. 179-197, (Part-I), *Transworld Research Network*, rivandrum, Kerala, India, 2000.

[4] G. Peters, Christoph von der Malsburg "View Reconstruction by Linear Combination of Sample Views", *Proceedings of the 12th British Machine Vision Conference* (BMVC:2001), edição de Tim Cootes e Chris Taylor, University of Manchester, Vol. 1, pp. 223-232, Manchester, UK, September 10-13, 2001.

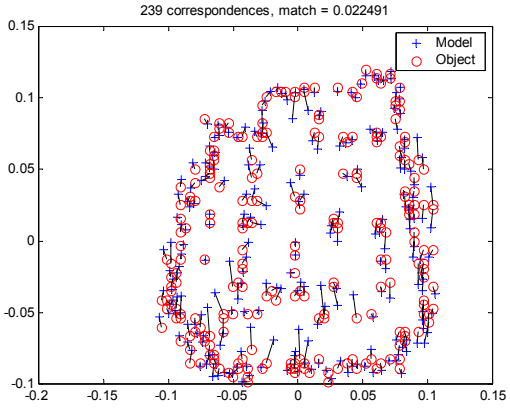

Figura 3 – Resultados da comparação do objecto casa com o modelo casa (match = 0.022491).

[5] G. Peters, C. von der Malsburg "Learning Sparse Representations of Three-Dimensional Objects", *Proceedings of the 10th European Symposium on Artificial Neural Networks* (ESANN 2002), edited by Michel Verleysen, d-side, pp. 245-250, Bruges, Belgium, April 24-26, 2002.

[6] H. M. Gomes, R. B. Fisher "Structural Learning from Iconic Representations." IBERAMIA-SBIA 2000, pp.399-408, 2000.

[7] N. K. Logothetis, J. Pauls, H.H. Bulthof, Poggio T. "Shape Representation in the Inferior Temporal Cortex of Monkeys", Current Biology, 5(5): 552-563, 1995.

[8] R. Basri "Recognition by Prototypes", *International Journal of Computer Vision*, 19(2): 147-168, 1996.

[9] R. Jonker, A. Volgenant "A Shortest Augmenting Path Algorithm for Dense and Sparse Linear Assignment Problems", Computing, 38:325-340, 1987.

[10] S. Belongie, J. Malik, J. Puzicha "Shape Matching and Object Recognition Using Shape Contexts", Vol. 24, No. 4, 2002.

[11] S. Edelman, H.H. Bülthoff "Orientation dependence in the recognition of familiar and novel views of Three-Dimensional Objects". Vision Research 32(12):2385-2400, 1992.

[12] S. Ullman, R Basri "Recognition by linear combinations of models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):pp. 992–1006, 1991.

[13] S. Z. Li, J. Yan, X.W. Hou, Z.Y. Li, and H.J. Zhang "Learning Low Dimensional Invariant Signature of 3-D Object under Varying View and Illumination from 2-D Appearances". In *Proceedings of 8th IEEE International Conference on Computer Vision*. Vancouver, Canada. July 9-12, 2001.

[14] T. Silva "Reconhecimento Visual de Objectos por Coerência Estrutural de Características", Tese de Doutoramento, IST, 2001.

[15] Z.Q. Zhang, L. Zhu, S.Z. Li, H.J. Zhang "Real-Time Multi-view Face Detection". *Proceedings de 5th International Conference on Automatic Face and Gesture Recognition*. Washington, DC, USA. 20-21 May, 2002.

Figura 4 – Imagem do modelo 3D do objecto cão.

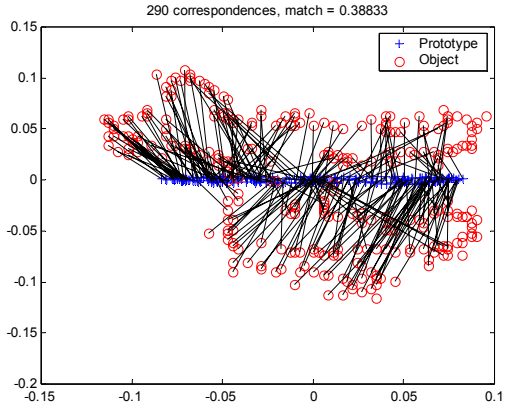

Figura 5 – Resultados da comparação do objecto cão com o protótipo casa (match = 0.38833).