

Um olhar pela visão computacional

HELDER DE JESUS ARAÚJO *

Uma das características mais marcantes do Homem é, sem dúvida, alguma, a sua extraordinária capacidade de processamento da informação. É essa capacidade que estabelece diferenças intransponíveis entre os outros seres vivos e o Homem, nomeadamente no que se refere ao tipo de relação com a Natureza.

Entre os vários tipos de informação que o Homem é capaz de adquirir e processar, a codificada em imagem ocupa um lugar muito especial e particularmente importante. O desenvolvimento dos computadores (enquanto máquinas de processamento de informação) permitiu ao Homem dispor de instrumentos com possibilidades de simularem, ou, eventualmente, assumirem comportamentos inteligentes. Utilizando máquinas capazes de tratar a informação codificada em imagens, a Visão Computacional (V.C.) tem como objectivo fornecer descrições simbólicas dum universo físico a partir de imagens desse universo.

A aquisição das imagens é feita por dispositivos sensores que convertem a radiação luminosa em sinais eléctricos. Basicamente utilizam-se dois tipos de sensores: as câmaras usuais de TV e as câmaras de estado sólido.

Adquirida a imagem, é necessário dar-lhe uma

forma que possibilite o seu tratamento por computador. Uma imagem a preto e branco pode ser considerada como uma matriz de pontos cujos valores são proporcionais à intensidade luminosa. Cada elemento da matriz é um número que codifica a intensidade luminosa do ponto correspondente através de um valor situado por exemplo, entre 0 (para completamente branco) e 255 (para completamente negro). Neste caso utilizar-se-ia um octeto¹ por cada elemento da matriz. Se se tratar de uma imagem a cores pode ter-se uma matriz para cada uma das cores fundamentais (vermelho, verde, azul). Claro que, para posterior tratamento, a imagem a cores pode ser «representada» por outras características.

Como foi anteriormente referido é objectivo da V.C. fornecer uma descrição da cena que corresponde à imagem. Vejamos algumas definições que permitem esclarecer melhor o âmbito e objectivo da V.C. Habitualmente designa-se por *processamento de imagem* o conjunto de métodos e de técnicas que trabalham sobre as imagens, procurando melhorar algumas das suas características, como por exemplo, a relação sinal-ruído, realçar os contrastes para facilitar a sua inspecção visual, «comprimi-las» para facilitar a sua transmissão ou ar-

* Grupo de Instrumentação e Controle — Departamento de Engenharia Electrotécnica da Universidade de Coimbra

¹ Octeto — grupo de oito dígitos binários.

mazenamento, etc., sem atender ao seu conteúdo. Outra área que, de certo modo, se sobrepõe à da V.C. é a do *reconhecimento de formas*. O objectivo desta área é o de classificar as imagens em categorias predeterminadas. Para isso um conjunto de características previamente definidas é extraído da imagem e, com base nesses valores e em regras de decisão (na grande maioria dos casos de natureza estatística), determina-se qual a categoria ou categorias em que a imagem (ou parte dela) se insere. No caso da V.C. pretende descrever-se não a imagem propriamente dita mas sim a cena que a imagem representa. Mais do que classificar a V.C. pretende descrever a cena. Claro que, em certa medida, a descrição comporta sempre uma classificação.

Em todo o processo de tratamento de informação em V.C. podem distinguir-se vários níveis de processamento. Ao nível mais baixo recorre-se a técnicas de processamento de imagem para se efectuar a extracção das características a serem tratadas nos níveis seguintes. A este nível pretende-se, por exemplo, determinar as zonas onde se verificam as variações bruscas de intensidade (que constituem as arestas) e a orientação espacial dessas mesmas arestas.

No nível de processamento imediatamente superior procura-se extrair informações tais como linhas, regiões, orientação de superfícies, etc. Trata-se de fazer o que é designado por segmentação da imagem, ou seja, a decomposição da imagem nos seus elementos constituintes. Um dos tipos mais usados é o da segmentação em linhas, em que os algoritmos mais eficientes são aqueles em que se introduz um certo «conhecimento» do domínio onde se pretende determinar as linhas.

A segmentação em regiões ou zonas é outra técnica complementar da anterior. Enquanto os algoritmos de detecção de arestas são baseados em princípios de descontinuidade, os que efectuem a determinação das regiões baseiam-se nas similaridades. O objectivo desta análise é agrupar os pixels² de uma imagem que tenham valores idênticos de determinada característica. Por exemplo, uma imagem pode ser segmentada em regiões com idênticos níveis de cinzento desde que se considere que essas regiões correspondem a superfícies dis-

tintas na cena que produziu a imagem.

Estas técnicas de segmentação são extremamente vantajosas para os estágios posteriores do processamento da imagem, pois para além de constituírem uma forma de compressão de dados (há muito menos zonas homogêneas que pixels), as regiões são, em princípio grupos de pixels com a mesma interpretação semântica. Esta última vantagem pode constituir simultaneamente o grande inconveniente desta técnica, se à uniformidade das regiões não corresponderem também elementos da cena com prioridades análogas.

O passo seguinte do tratamento da imagem situa-se já no nível mais elevado. Trata-se de construir descrições de cenas. É então necessário incorporar nos algoritmos uma forte dose de conhecimento do domínio específico da imagem. Um dos primeiros programas desenvolvidos com este objectivo foi o de L. G. Roberts. O programa era capaz de identificar os objectivos na cena, e de determinar as posições e orientações respectivas. O seu domínio de aplicação era bastante restrito: um universo constituído por prismas hexagonais, paralelepípedos, cubos, e prismas triangulares. Para efectuar a identificação dos objectos o programa dispunha de modelos tridimensionais de cada um dos objectos básicos (um cubo, um paralelepípedo, um prisma hexagonal e um prisma triangular) representados pelas coordenadas tridimensionais de cada um dos seus vértices. Os modelos podiam ser rolados, projectados e escalados para tornar possível a identificação, pois os objectos podiam estar colocados na cena numa posição arbitrária.

Outro modo de identificar objectos é por intermédio da incorporação de heurísticas nos programas. Por exemplo, o programa SEE desenvolvido por A. Guzman não tinha modelos de objectos armazenados, efectuando a identificação a partir de regras puramente heurísticas. Este programa, contudo, não fornecia descrições tridimensionais dos objectos representados na imagem e, ao contrário do programa de Roberts, utilizava métodos não numéricos (simbólicos).

Os programas utilizando este tipo de metodologia partilhavam um defeito fundamental: nenhum deles fazia um tratamento explícito das características tridimensionais da cena. Toda a heurística se baseava nas características bidimensionais da imagem, sem a preocupação de as relacionar explicitamente com as características da cena representada.

² Pixel — cada um dos pontos que constituem a imagem na sua representação susceptível de ser tratada por computador.

Ao contrário das aproximações ao problema até então dominantes, Huffman e Clowes em 1971 estabeleceram a diferença entre o domínio da cena e o domínio da imagem. Esta aproximação resultou da necessidade de tornar explícito o processo de formação da imagem. Enquanto o domínio da cena envolve aspectos tridimensionais e físicos, o domínio da imagem tem que ver com a projecção dessas características num plano. Nesta perspectiva o objectivo da interpretação da imagem, deve ser o de entender os respectivos elementos da como propriedades no domínio da cena. Quer Huffman quer Clowes, limitaram o seu tratamento a um universo poliédrico.

A consideração das características físicas do processo de formação das imagens veio permitir integrar nos algoritmos de interpretação as restrições impostas pelo mundo físico. Uma das consequências mais importantes deste tipo de abordagem foi a utilização de informação sobre a forma dos objectos. Antes da consideração desse tipo de características era possível caracterizar a forma apenas de um modo qualitativo, além de que os objectos deviam pertencer a um universo restrito, o universo triédrico. Utilizando essas características físicas é possível determinar a forma não só de um modo qualitativo, mas também quantitativo. É o que faz Takeo Kanade, que, para além de determinar a forma de um modo quantitativo, consegue determinar a imagem correspondente à mesma cena vista sob diferentes ângulos. Além disso, com a introdução do que se chama «universo de Origami», tornou possível tratar imagens de cenas que admitem mais objectos que o universo triédrico.

Como é óbvio a cada imagem pode corresponder mais do que uma cena tridimensional. Contudo, quando vemos uma imagem, geralmente, temos ideias definidas quanto à cena tridimensional que a originou. Isto significa que usamos algumas suposições, conhecimento dos objectos ou do processo de formação das imagens. Kanade conseguiu identificar algumas dessas suposições (a maior parte das quais é do tipo geométrico) e introduziu-as também no seu programa de modo a que este rejeitasse as interpretações menos comuns.

Existem contudo outras técnicas para determinar de um modo único a cena que deu origem à imagem, uma das quais é a designada por estereo fotométrico. A técnica usada neste tipo de estereo

é a de variar a direcção da iluminação incidente entre imagens sucessivas, conservando constante a direcção sob a qual se obtêm as imagens. Assim os valores das intensidades nas várias imagens são suficientes para determinar as várias orientações. O estereo fotométrico pode ser usado quer para determinar a orientação de superfície em cada ponto da imagem quer para determinar pontos do objecto que tenham uma orientação de superfície particular. Este tipo de aproximação remete-nos para uma outra, o estereo tradicional, que lhe é análoga, pois é um tipo de processamento que também utiliza mais do que uma imagem da mesma cena. Neste tipo de técnica usam-se duas imagens da mesma cena, obtidas em duas posições diferentes. Determinados os pontos correspondentes nas duas imagens, e conhecidas as posições relativas sob as quais se obtiveram as imagens é então possível calcular a forma tridimensional dos objectos. A maior dificuldade neste tipo de estereo é a da identificação dos pontos correspondentes nas duas imagens. Na visão em robótica a técnica que geralmente se usa para se efectuar o reconhecimento dos objectos, é a da utilização (conjuntamente com imagens) de informação quanto à distância de cada um dos pontos visíveis da câmara.

Até agora abordámos o tratamento de imagens nas quais o problema do movimento não se colocava. A complexidade das questões acima referidas aumenta significativamente quando se trata de processar sequências de imagens obtidas ao longo do tempo. Geralmente a sequência de imagens representa a cena em instantes de tempo muito próximos, e pode ter várias origens: ou porque os objectos se movem na cena, sendo o sensor fixo, ou porque o sensor se move relativamente à cena, ou, finalmente porque objecto e sensor se movem. O objectivo último do processamento destas sequências de imagens é a assimilação de informação da sequência como um todo, informação essa que não podia ser obtida de cada uma das imagens isoladamente.

Os sistemas de visão, processando ou não imagens em movimento, têm tido múltiplas aplicações e têm-se revelado extremamente poderosos nas contribuições prestadas ao desenvolvimento das áreas onde são utilizados. As aplicações do processamento de sequências de imagens são inúmeras e vão desde a análise de imagens de satélites à robótica, passando pela medicina e biologia (angiocardiografia, circulação de sangue, fenómenos

de crescimento, etc.) e pelas aplicações militares (seguimento de alvos em tempo real, de forma automática).

Na indústria a utilização deste tipo de sistemas permite aumentar a produtividade, melhorar a qualidade, eliminar tarefas repetitivas, e automatizar processos que se desenvolvem em meios ambientes perigosos ou desagradáveis. As tarefas que os sistemas de visão industriais normalmente desempenham vão desde a montagem e manipulação de peças ao controlo de qualidade. A operação em tempo real é, sem dúvida, uma das características determinantes do projecto de um sistema de visão industrial, sendo a sua viabilidade extremamente dependente da rapidez com que desempenha a tarefa e também o custo. A rapidez vai influenciar significativamente a escolha da resolução e tamanho da imagem, do número de níveis de cinzento, e ainda a escolha entre as implementações por «hardware» ou «software» dos algoritmos. Vários sistemas de visão industriais estão já em funcionamento em todo o mundo com grande sucesso. Um deles desenvolvido pela Hitachi, faz as ligações internas nas pastilhas de transístores. O sistema localiza visualmente a pastilha e liga automaticamente os fios de ouro entre os terminais externos e os eléctrodos internos do transístor. Este sistema permite a montagem de 2000 «chips» por hora com uma precisão superior a 99 %, podendo montar-se vários tipos de transístores.

Outro sistema industrial famoso é o CONSIGHT-I desenvolvido pela General Motors. Trata-se de um robô que apanha peças colocadas ao acaso num tapete rolante. O subsistema da visão, que pode operar em ambientes com iluminação não controlada, como é, tipicamente, o de uma fábrica, determina a posição e orientação das peças no tapete rolante. Após a identificação um braço robótico apanha as peças e transfere-as para locais predeterminados. O CONSIGHT-I tem dois modos de funcionamento: o modo de inicialização e o modo operacional. Durante a inicialização o sistema é calibrado e programado para manipular determinado conjunto de peças. Ensina-se o sistema a reconhecer determinada peça pela identificação de descritores como, por exemplo, a área, pontos extremos, perímetro, etc. No modo operacional o sistema executa as transferências de peças automaticamente. As peças são colocadas em posições arbitrárias (sem contudo se poderem tocar) no tapete rolante. O tapete transporta as

peças para o subsistema de visão que continuamente o analisa. O subsistema de visão identifica o tipo de peça, a sua posição e orientação e envia esta informação para o sistema de controlo do braço de robô. As peças continuam a mover-se e, na posição apropriada, o braço apanha-as e coloca-as no local predeterminado. A reprogramação do CONSIGHT-I é extremamente simples, bastando colocar a nova peça no tapete, deixá-la passar pelo subsistema de visão, parar o tapete e manualmente mover o braço para a posição em que se deseja que ele apanhe a peça. A calibração do CONSIGHT-I faz-se sempre que, ou o tapete, ou a câmara, ou o braço são deslocados das suas posições iniciais. Cada operação de calibração dura cerca de 15 minutos.

O campo das aplicações da Visão Computacional tem vindo a registar um crescimento paralelo ao das tecnologias que lhe servem de suporte. Por um lado o desenvolvimento da electrónica digital tem permitido cada vez mais tratar rapidamente quantidades de informação crescentes e de forma cada vez mais barata. Por outro o desenvolvimento da Inteligência Artificial permitiu o aparecimento de técnicas para manipular informações a níveis relativamente elevados de abstracção e generalização. A própria natureza das informações tratadas nos sistemas de V.C. torna-os extremamente úteis em variados domínios porque a informação fornecida pelas imagens é, de certa forma, qualitativamente superior à fornecida por outros sensores. É portanto previsível que daqui a alguns anos o sistema computadorizado, dispondo de visão, penetre no nosso quotidiano mais imediato.

RESUMO

Aborda-se um dos temas da Ia, o da codificação da informação em imagens, vulgo visão computacional, discutindo o seu âmbito e objectivo. E, situa-se o tema nas aplicações da robótica industrial.

SUMMARY

We present the scope and goal of computational vision, explaining its content.

REFERÊNCIAS

- ANDERSON, J. R. e G. BOWER (1973) – *Human Associative Memory*, Erlbaum, N. Y.
- BELL, C. G. e A. NEWELL (1971) – *Computer Structures*, McGraw-Hill, N. Y.
- BROADBENT, D. M. (1958) – *Perception and Communication*, Pergamon Press, London.
- BRUNER, J.; J. GOODNOW; G. AUSTIN (1956) – *A Study in Thinking*, Wiley, N. Y.
- CHOMSKY, N. (1957) – *Syntactic Structures*, Mouton, The Hague.
- NEWELL, A.; J. C. SHAW; H. A. SIMON (1957) – *Preliminary Description of General Problem Solving Program-I*, Tech. Rep., Carnegie Institute of Technology.
- QUILLIAN, M. R. (1968) – «Semantic Memory», in M. Minsky (Ed.), *Semantic Information Processing*, M.I.T. Press, Mass.
- TANNER, J. P. E. J. A. SWETS (1954) – «A decision-making theory of visual detection», *Psychological Review*, 61:401-409.

PSICOLOGIA

PRÓXIMOS NÚMEROS

GRUPOS E ORGANIZAÇÕES – HOMENAGEM A MANUEL TAVARES DA SILVA

Organização de Mário Ceitil

A SEXUALIDADE: PERSPECTIVAS PSICANALÍTICAS

Organização de Pedro Luzes

A EDUCAÇÃO PRÉ-ESCOLAR

Organização de Joaquim Bairrão