

INTRODUÇÃO ÀS TÉCNICAS PARA LEVANTAMENTO DE FORMA TRIDIMENSIONAL SEM CONTACTO

Cândida Pinto Coelho¹, João Manuel R. S. Tavares^{1, 2}

¹Laboratório de Óptica e Mecânica Experimental, Instituto de Engenharia Mecânica e Gestão Industrial, Rua Dr. Roberto Frias, s/n 4200-465 PORTO PORTUGAL, {candide@fe.up.pt, tavares@fe.up.pt}

²Faculdade de Engenharia da Universidade do Porto, Departamento de Engenharia Mecânica e Gestão Industrial, Rua Dr. Roberto Frias, s/n 4200-465 PORTO PORTUGAL, {<http://www.fe.up.pt/~tavares>}

RESUMO

As técnicas de levantamento de forma tridimensional (3D) sem contacto são hoje uma ferramenta muito usada para a obtenção da forma 3D de objectos ou cenas. Este artigo descreve, resumidamente, algumas dessas técnicas que têm sido utilizadas na aquisição de informação 3D no domínio de Visão Tridimensional, bem como as vantagens e dificuldades que lhes são apontadas.

1 INTRODUÇÃO

Ao longo dos últimos anos, tem-se assistido ao desenvolvimento de várias técnicas de aquisição de informação tridimensional sem contacto que, directa ou indirectamente, determinam as coordenadas tridimensionais de pontos de uma cena ou de um objecto, a distância desses pontos a um determinado referencial ou a orientação local das superfícies que os constituem.

Neste artigo, optou-se por fazer a divisão das técnicas de aquisição de informação 3D a apresentar em duas classes: passivas e activas; estando cada uma delas dividida em várias subclasses, como se observa no esquema da Fig. 1. Esta divisão, habitual neste domínio, teve em atenção os princípios básicos utilizados por cada uma dessas técnicas.

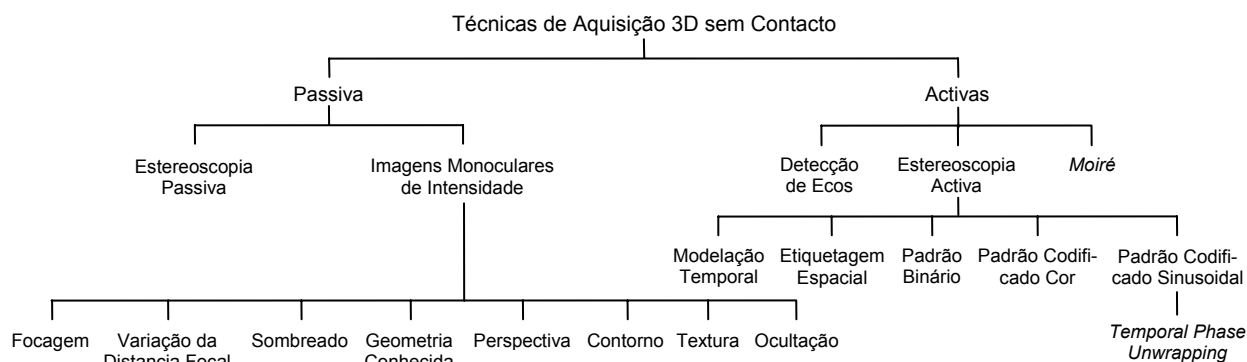


Figura 1 – Divisão das técnicas sem contacto para obtenção de informação 3D de uma cena ou objecto.

Os termos cena e objecto irão ser utilizados, ao longo deste artigo, de uma forma não distinta, assim, a aquisição da informação tridimensional tanto pode ser relativa a uma cena como a um objecto.

2 TÉCNICAS PASSIVAS

As técnicas passivas caracterizam-se por utilizarem apenas a luz ambiente para iluminar a cena. A informação de distância é extraída a partir de uma ou mais imagens de intensidade da cena. As técnicas passivas que têm a designação genérica de “forma com base em x ”, procuram extrair a informação de distância a partir de uma ou mais imagens monoculares de intensidade, obtidas com uma câmara fixa. De um modo geral, estas técnicas apenas permitem obter informação acerca da orientação das superfícies ou da distância relativa entre os pontos da cena, não permitindo medir distâncias absolutas. No entanto, existem técnicas de medida de distância absoluta usando imagens monoculares: o exemplo mais conhecido é o das técnicas de focagem que recorrem à equação das lentes para determinar a distância às regiões bem focadas de uma ou mais imagens de intensidade. Das técnicas passivas a mais conhecida é a estereoscopia que permite obter informação de distância a partir de duas ou mais imagens de intensidade da cena, tomadas com uma ou mais câmaras colocadas em posições diferentes, recorrendo ao princípio da *Triangulação*.

2.1 Técnicas Baseadas em Imagens Monoculares de Intensidade

As técnicas descritas nesta subsecção procuram extrair informação de distância com base em imagens monoculares de intensidade, obtidas com uma câmara fixa. Estas técnicas têm a designação genérica de “forma com base em x ”, em que “ x ” é a informação usada para extrair informação tridimensional. Normalmente, estas técnicas são divididas em dois grupos: aquelas que permitem obter medidas de distância absoluta com base na focagem, na variação

da distância focal ou no conhecimento prévio da geometria da cena, como o tamanho dos objectos ou de outros elementos da cena; e aquelas que permitem obter informação acerca da orientação das superfícies ou da proximidade relativa dos objectos com base no sombreado, na perspectiva, na textura, nos contornos, na oclusão ou noutras características que de forma indirecta traduzem a geometria da cena. Nestes casos, a obtenção da informação de orientação das superfícies só é possível quando se dispõe de várias imagens da cena obtidas da mesma posição, mas em condições ambientais diferentes.

2.1.1 Focagem

A técnica de focagem [Awcock, 1996; Jain, 1995; Silva, 1994] consiste no cálculo da distância aos pontos da imagem que estejam bem focados (r), tendo o conhecimento prévio da distância focal da lente da câmara (f) e da distância entre o centro óptico da lente e o plano imagem (s), a partir da equação das lentes finas (Fig. 2):

$$\frac{1}{f} = \frac{1}{r} + \frac{1}{s}$$

Esta equação é aplicável também em lentes mais espessas, desde que as distâncias r , s e f sejam ajustadas aos pontos principais que resultam da intersecção dos planos principais da lente com o eixo óptico. O sucesso da obtenção de distância dos pontos da imagem depende essencialmente da qualidade da focagem. A precisão das medidas é inversamente proporcional à profundidade do campo, por essa razão a abertura do campo de visão e a distância focal devem ser elevadas.

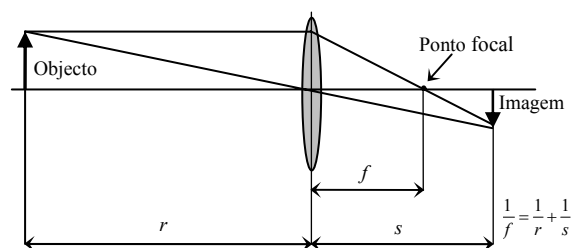


Figura 2 – Técnica baseada na focagem: formação de imagens a partir da equação das lentes finas.

A principal vantagem desta técnica é permitir obter informação de distância absoluta usando uma só câmara, sem conhecimento prévio da cena. Contudo, esta técnica está limitada a cenas que não sejam visualmente homogêneas e à rápida diminuição da precisão e resolução com o aumento da distância.

2.1.2 Variação da distância focal

A técnica de variação da distância focal [Awcock, 1996; Jain, 1995; Silva, 1994] consiste no cálculo da distância absoluta a partir da comparação de um determinado ponto característico em duas imagens de intensidade adquiridas com variação da distância focal (Δf) do sistema de aquisição (Fig. 3).

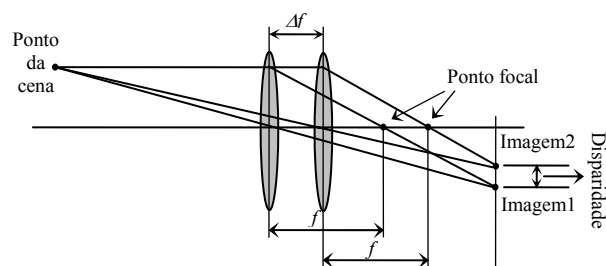


Figura 3 – Técnica baseada na variação da distância focal: disparidade entre pontos da imagem.

A principal desvantagem desta técnica é obrigar à detecção e emparelhamento dos pontos característicos nas duas imagens, tornando-a pesada computacionalmente. Outra dificuldade ocorre quando se trata de estudar objectos distantes, pois a distância absoluta tem de ser determinada com elevada precisão.

2.1.3 Sombreado

A técnica baseada no sombreado [Awcock, 1996; Jain, 1995; Silva, 1994] pretende extrair informação sobre a forma de um objecto a partir do sombreado das superfícies, resultante da variação da sua orientação local relativamente à fonte de luz. Uma câmara mede a intensidade de um elemento de superfície, a qualidade desta medição depende da iluminação incidente, das características reflectoras da superfície, da sua orientação relativamente à fonte de

luz e à câmara, e das características da câmara (Fig. 4). Esta técnica baseia-se na aquisição de várias imagens da cena, captadas da mesma posição, mas iluminadas por uma fonte de luz diferente. A aquisição é conseguida se a superfície for perfeitamente difusora, a fonte de luz e a câmara estiverem em posições fixas e suficientemente afastadas da superfície de modo que a intensidade da iluminação incidente sobre cada ponto desta se possa considerar constante, e a direcção de visão se possa admitir a mesma para todos os pontos da cena. Considerando estes pressupostos, a intensidade medida no plano imagem, $I(i, j)$, é dada pela função:

$$I(i, j) = R(p, q),$$

em que $I(i, j)$ são os dois parâmetros que determinam a orientação da superfície e $R(p, q)$ especifica a intensidade de um elemento de superfície em função da sua orientação.

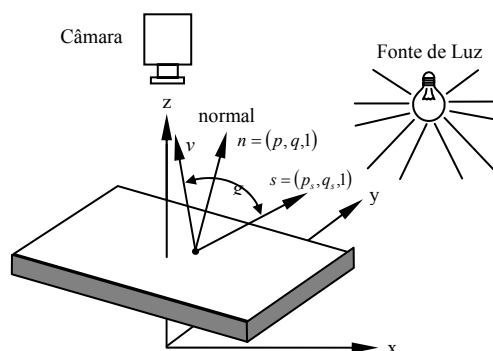


Figura 4 – Técnica baseada no sombreado: informação da forma de um objecto extraída a partir da variação da sua orientação local relativamente a uma fonte de luz, imagem adaptada de [Silva, 1994].

A principal desvantagem desta técnica é não ser aplicável a cenas que apresentem descontinuidades, pois a distância relativa entre dois pontos é obtida por integração. A determinação de distâncias absolutas só é possível se for conhecida previamente a distância a um ponto da cena, visível nas imagens.

2.1.4 Geometria conhecida

Na técnica baseada na geometria conhecida dos objectos [Awcock, 1996; Jain, 1995; Silva, 1994] é possível

determinar a distância absoluta através de uma relação simples, conhecendo as dimensões dos objectos da cena (R) e os parâmetros geométricos do sensor de imagem (eixo óptico, centro de projecção, R' , e distância focal, f). Esta técnica é particularmente simples para objectos esféricos se o ângulo de visão for pequeno (Fig. 5).

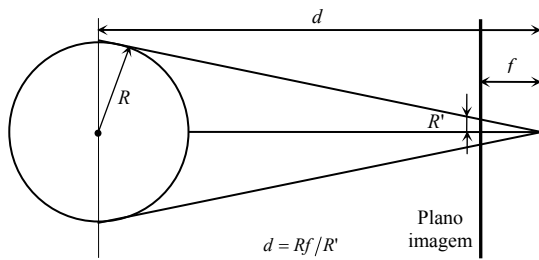


Figura 5 – Técnica baseada em geometria conhecida: distância absoluta de um objecto esférico.

No entanto, esta técnica tem a desvantagem de ser de difícil aplicação em cenas com objectos polidédricos ou em que o campo de visão seja elevado.

2.1.5 Perspectiva

Esta técnica [Jain, 1995; Silva, 1994] pressupõe que existam linhas paralelas na cena, por exemplo arestas de objectos. Sabendo-se que duas linhas de uma imagem correspondem a linhas paralelas na cena é possível impor “restrições” à orientação do plano que contém essas linhas. Quando as imagens das linhas se intersectam, o ponto de intersecção define, com o centro óptico da lente do sistema formador de imagem, uma recta no espaço *coplanar* às referidas linhas, restringindo assim as posições possíveis do plano (Fig. 6). Estes pontos

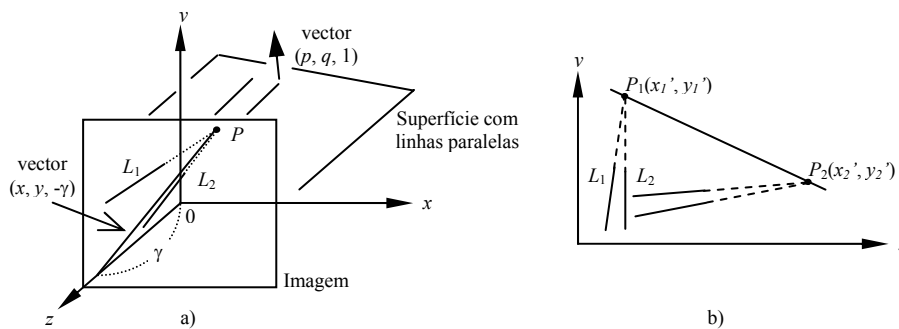


Figura 6 – Técnica baseada na perspectiva: a) as linhas (paralelas) da superfície projectadas na imagem intersectam-se no ponto de fuga (P); b) os pontos de fuga P_1 e P_2 de dois conjuntos de linhas.

têm a designação de pontos de fuga. Dispondo-se de um outro conjunto de linhas paralelas, no mesmo plano e com direcção diferente, é possível obter outra recta no espaço que também pertence ao plano. O plano em questão é pois a superfície definida por essas duas rectas.

A principal desvantagem desta técnica é a necessidade de existência, na cena, de superfícies planas que contenham pelo menos dois conjuntos de linhas paralelas. Outra dificuldade é a existência de outras linhas não paralelas.

2.1.6 Contorno

A técnica baseada no contorno [Jain, 1995; Silva, 1994] consiste na determinação da orientação da superfície a que pertence um contorno de uma imagem dada, fazendo a sua projecção inversa sobre planos com diversas orientações e escolhendo a orientação do plano a partir da maximização do factor de forma do contorno projectado (F), dado por:

$$F = \frac{\text{area}}{\text{perimetro}^2}.$$

Deste modo, uma elipse, um trapézio e um triângulo são interpretados, respectivamente, como uma circunferência, um rectângulo e um triângulo equilátero inclinados.

A principal dificuldade deste método é a possibilidade de um contorno bidimensional poder ser a projecção de uma infinidade de contornos tridimensionais, o que o torna ambíguo (Fig. 7).

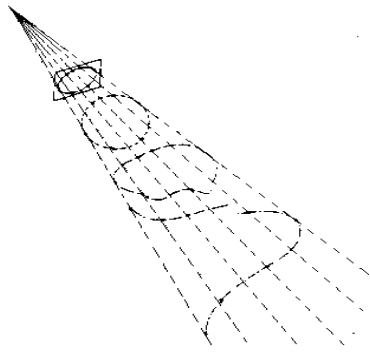


Figura 7 – Técnica baseada no contorno: um contorno bidimensional poder ser a projecção de uma infinidade de contornos tridimensionais, imagem de [Silva, 1994].

2.1.7 Textura

A textura de uma superfície [Awcock, 1996; Jain, 1995; Silva, 1994] pode ser usada para obter informação de orientação e de distância relativa. As variações de tamanho, densidade ou orientação, que os elementos de uma textura uniforme têm numa imagem, dependem da distância e da inclinação da superfície relativamente ao plano de imagem. É possível também obter informação de distância absoluta, se o tamanho dos elementos da textura forem conhecidos.

Existem variantes desta técnica, no entanto, irá ser apresentada neste artigo como exemplo apenas uma. Assim, esta variante implica que os elementos da textura da cena tenham uma densidade uniforme, ou seja, o número de elementos por unidade de área seja constante. Este gradiente traduz uma inclinação da superfície relativamente ao plano de imagem, sendo possível assim obter a sua orientação (Fig. 8). Esta variante tem como dificuldade a identificação e a contagem exacta dos elementos da textura.

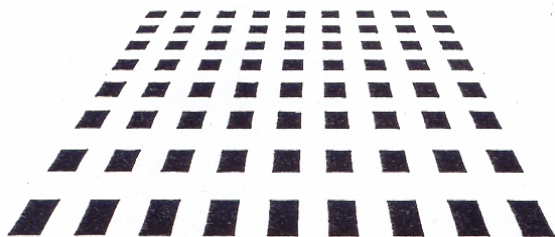


Figura 8 – Técnica baseada na textura: o gradiente de textura pode ser usado para determinar a orientação das superfícies, imagem de [Silva, 1994].

As principais desvantagens da técnica de determinação da forma com base na textura são: a necessidade de existência de superfícies texturadas na cena, o conhecimento prévio da textura de modo a poder fazer a distinção entre as distorções devidas à projecção e as características da textura, e o elevado custo computacional.

2.2 Estereoscopia Passiva

A técnica que recorre a duas imagens de uma cena, adquiridas de posições diferentes, para obter informação de distância, designa-se estereoscopia passiva [Awcock, 1996; Jain, 1995; Silva, 1994; Tavares, 1995]. Esta técnica baseia-se na forma como se processa o sistema visual humano para obtenção da informação tridimensional. A percepção de distância deste sistema resulta principalmente de um processo que opera sobre duas imagens, chamado “fusão”. Apesar de cada imagem por si só não conter informação de distância, quando as duas imagens são “fundidas” pelo sistema de visão humano, é obtida uma estrutura tridimensional resultante da disparidade, ou diferença de posição, entre os pontos das duas imagens (Fig. 9).

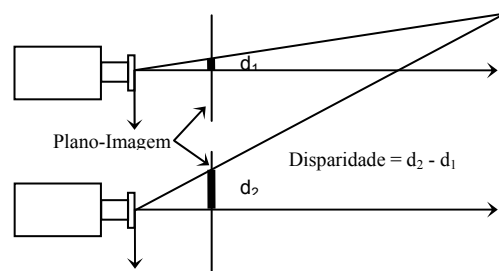


Figura 9 – Ilustração do conceito de disparidade.

O processo de medida de distância por estereoscopia passiva pode ser dividido nos três passos seguintes:

- i) aquisição de duas imagens da cena, captadas por câmaras afastadas uma da outra (ou por uma só câmara em posições/instantes distintas) o suficiente de modo a que exista alguma sobreposição entre essas imagens;
- ii) estabelecer as correspondências entre pontos característicos presentes nas duas

imagens, que sejam a projecção do mesmo ponto na cena;

- iii) para cada par de pontos correspondentes determinar o ponto de intersecção das rectas por eles determinadas no espaço, obtendo assim as coordenadas do ponto da cena. Este é o principio usado nas técnicas estereoscópicas para obtenção da informação tridimensional, o qual tem a designação de *Triangulação*.

Existem duas configurações básicas do sistema de aquisição para obtenção das duas imagens da cena: uma em que as câmaras estão colocadas no mesmo plano, com eixos ópticos paralelos, e orientadas de modo a que as linhas de varrimento das imagens sejam paralelas à linha que une os centros ópticos das duas câmaras, facilitando assim o estabelecimento de correspondências entre os pontos das duas imagens (Fig. 10a); outra em que os eixos ópticos das duas câmaras são convergentes e não necessariamente *coplanares* (Fig. 10b).

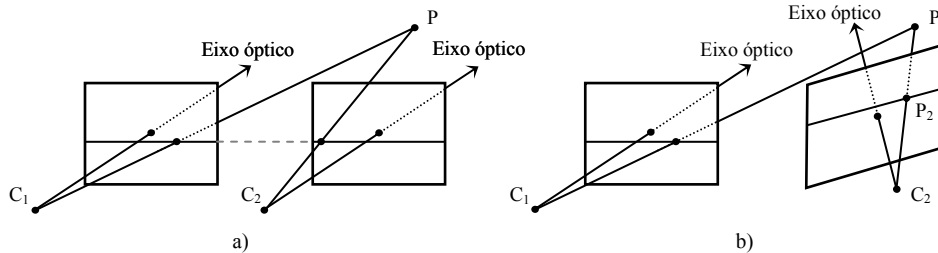


Figura 10 – Estereoscopia passiva: câmaras com os eixos ópticos a) paralelos; b) convergentes.

No entanto, o principal desafio das técnicas de estereoscopia passiva reside no processo de estabelecimento de correspondência entre os pontos das duas imagens, que pode ser dividido em três fases: a primeira consiste no processamento das imagens no sentido de detectar pontos ou outras características a usar no emparelhamento; a segunda consiste em efectuar um emparelhamento local entre as duas imagens; e a terceira consiste em resolver as ambiguidades e erros no emparelhamento. Os procedimentos que podem ser usados em cada uma dessas fases são descritos em seguida:

- Na primeira fase, é feito um processamento das duas imagens, no sentido de detectar pontos ou outras

características a usar no emparelhamento. Para evitar ambiguidades no emparelhamento são escolhidos pontos, em torno dos quais se define uma janela de correlação, e que se distinguem bem dos restantes pontos que se encontram nessa janela. Estes pontos são designados por pontos de interesse, e resultam normalmente de características locais da cena, como arestas, vértices, orlas dos objectos, ou outras resultantes da textura, variação de tonalidade ou de sombras. Os pontos de interesse podem ser detectados usando um operador sensível às variações locais de intensidade, como um operador de gradiente [Tavares, 1995].

- Na segunda fase, para efectuar o emparelhamento local têm sido usadas duas técnicas, designadas por emparelhamento baseado em áreas e emparelhamento baseado em elementos característicos. A primeira técnica consiste em seleccionar uma região em

torno de um ponto de interesse de uma das imagens e procurar, na outra imagem, uma região com semelhança máxima. Essa semelhança é medida a partir do coeficiente de correlação das intensidades de cada região. Esta técnica é condicionada por diversos factores como: a escolha do tamanho da janela de correlação, a sensibilidade às diferenças de intensidade entre regiões correspondentes das duas imagens e a existência de discontinuidades na cena. A segunda técnica é a mais utilizada e consiste em estabelecer correspondências entre características discretas extraídas, geralmente, dos segmentos das orlas da cena. Essas características são do tipo geométrico ou “físico”, como por exemplo, o

comprimento, as coordenadas dos pontos extremos, a orientação ou o contraste médio ao longo do segmento. Nesta técnica coloca-se um problema, que também surge na técnica anterior, que é o da selecção do tamanho da máscara de convolução usada para extrair as características. Esta técnica é considerada vantajosa relativamente à anterior por diversos motivos, entre eles, a maior rapidez no emparelhamento devido ao número de candidatos ser, em princípio, inferior; e a menor sensibilidade às diferenças de intensidade entre as duas imagens. No entanto, tem a desvantagem de não permitir obter uma imagem de distância mais densa.

- Na terceira fase, com o objectivo de resolver certas ambiguidades e impor consistência global nos emparelhamentos têm sido usadas técnicas de relaxação. Estas técnicas podem ser utilizadas no emparelhamento baseado em áreas ou em características, e consistem no seguinte procedimento: para cada ponto de interesse ou segmento característico de uma imagem é determinado um conjunto de candidatos ao emparelhamento na outra imagem; a cada par é atribuída uma probabilidade baseada num critério de semelhança que depende do método de emparelhamento e é actualizada através de um processo iterativo (relaxação). Este processo consiste em reforçar ou enfraquecer essa probabilidade de emparelhamento, consoante a disparidade associada ao candidato seja próxima ou afastada à associada a pontos ou características vizinhos que tenham probabilidades de emparelhamento elevado.

Estabelecidas as correspondências, pretende-se determinar o ponto de intersecção das rectas determinadas por cada par de pontos correspondentes no espaço de forma a obter as coordenadas tridimensionais do ponto da cena. Para tal, é usado o princípio da *Triangulação* que também é aplicável quando uma das

imagens é substituída pela projecção de um padrão de luz sobre a cena (estas técnicas são descritas na secção seguinte). O padrão a projectar pode ser visto como uma imagem do padrão que a projecção gera na cena.

O princípio de *Triangulação* explora o facto de, no espaço tridimensional, um ponto da cena em análise poder ser definido pela intersecção de duas rectas (ou de um plano com uma recta) no espaço [Jähne, 2001; Jain, 1995; Paiva, 1997; Silva, 1994; Tavares, 1995]. Cada recta é a linha de visão que une esse ponto da cena (P) com o centro óptico da lente da câmara utilizada na aquisição (C_i) e que passa pelo ponto da imagem correspondente (P_i) (Fig. 11) (ou o plano é um plano de luz projectado que passa por esse ponto da cena (P)). A equação destas rectas (ou do plano e da recta) é calculada a partir das coordenadas do ponto na imagem e dos parâmetros geométricos da câmara.

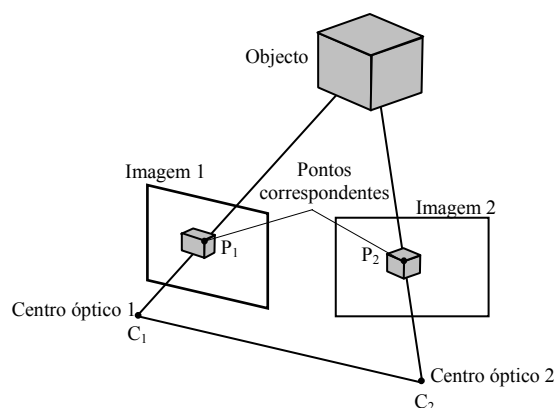


Figura 11 - Triangulação usando duas câmaras.

3 TÉCNICAS ACTIVAS

As técnicas activas de aquisição de informação tridimensional são caracterizadas pela projecção directa e controlada de energia sobre a cena, recorrendo a um ou mais projectores e um ou mais sensores. A energia projectada, geralmente sob a forma de luz ou ultrasons, é reflectida pelos objectos da cena, sendo detectada por sensores que, de forma mais ou menos directa, fornecem a informação de distância. Estas técnicas podem ser divididas em: técnicas de *Moiré*, técnicas de detecção de ecos e técnicas de

luz estruturada. Nas técnicas de *Moiré*, a informação de distância é obtida a partir da interacção entre um padrão de luz projectado sobre uma cena e um padrão de referência. As técnicas de detecção de ecos baseiam-se na projecção de energia (luz ou ultra-sons) sobre a cena e na medida do tempo que decorre entre a emissão e a recepção dessa energia, após reflexão na cena. A estereoscopia activa é a designação comum para um conjunto de técnicas que utilizam a projecção de um padrão de luz sobre a cena, da qual se adquire uma imagem de intensidade, com uma câmara (sensor) afastada do projector.

3.1 Técnica de *Moiré*

A técnica de *Moiré* [Benoit, 1975; Brooks, 1969; Chen, 2000; Gonzalez, 2002; Silva, 1994] consiste na projecção de um conjunto de linhas paralelas sobre a cena e a aquisição de uma imagem da mesma, recorrendo a uma câmara deslocada em relação ao projector, em frente da qual está colocada uma grelha idêntica à usada para projectar as linhas. O padrão de *Moiré* é um padrão de interferência de baixa frequência formado pela sobreposição de duas grelhas com padrões regularmente espaçados de frequência espacial mais elevada (Fig. 12a). O resultado é um padrão constituído por franjas alternadamente brilhantes e escuras, chamadas franjas de *Moiré*. Os pontos da cena que são iluminados pelo projector e captados na imagem aparecem brilhantes, os restantes aparecem escuros.

Existe uma variante desta técnica, designada por *Moiré* de sombra, que usa uma única grelha colocada em frente de toda a extensão da superfície em estudo, tão próxima desta quanto possível (Fig. 12b). A grelha utilizada define, com os centros de projecção e de aquisição de imagem do sistema, dois conjuntos de planos no espaço. As intersecções destes planos dão origem a superfícies de contorno com forma cilíndrica. A intersecção destas superfícies com a superfície do objecto dá origem a um conjunto de franjas de *Moiré*. Conhecido o espaçamento das linhas da grelha

projectada, a distância entre o projector e a câmara, a distância do projector/câmara à grelha e o número de ordem da superfície de contorno é possível determinar a variação de distância entre os pontos de duas franjas consecutivas.

Ambas as variantes usam o mesmo princípio na aquisição de informação tridimensional, sendo apenas o modo de geração das franjas diferente.

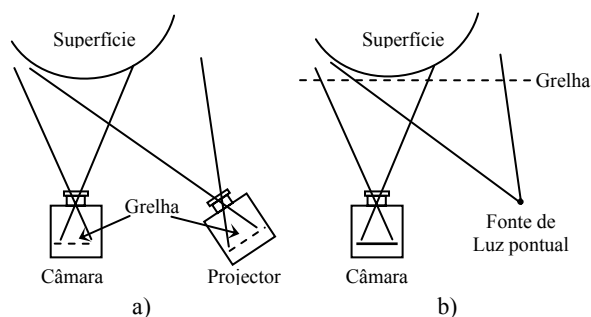


Figura 12 – Técnicas de *Moiré*: a) *Moiré* de projecção; b) *Moiré* de sombra.

A variante de sombra tem a vantagem de ser mais simples de implementar, mas não pode ser aplicada em cenas de grandes dimensões devido à necessidade de uma grelha do tamanho da cena. A principal vantagem da técnica de *Moiré* é a sua elevada resolução, sendo a da técnica de sombra superior à de projecção. A principal dificuldade é de só permitir obter informação de distância relativa entre as franjas. Outra desvantagem é a difícil determinação da linha central de cada franja.

3.2 Técnicas de Detecção de Ecos

A técnica de detecção de ecos [Chen, 2000; Jähne, 2001; Jain, 1995; Silva, 1994] consiste na determinação de distância a partir do tempo de percurso desde que uma fonte de energia (como por exemplo um feixe de luz colimado, ultra-sons, etc.) é emitida até ser detectada. Esta técnica implica a utilização de um emissor de energia, um sistema de varrimento da cena e um receptor que detecta uma parte da energia reflectida pela cena. Existem duas técnicas para determinar o tempo de percurso que se distinguem pelo tipo de modulação utilizada. Na primeira, mede-se

o tempo que decorre entre a emissão de um impulso de energia e o seu retorno (Fig. 13a). A principal desvantagem desta técnica ocorre quando os objectos da cena estão muito próximos do sensor, sendo necessário uma maior precisão na medição dos intervalos de tempo, visto se tratar de intervalos muito curtos. Na segunda, mede-se o tempo com base na diferença de fase entre um feixe de energia emitido, modulado em amplitude por uma função sinusoidal, e o sinal reflectido (Fig. 13b). Esta técnica falha quando a diferença de fase é superior a 2π . As formas de contornar esta situação seriam a utilização de uma frequência de modulação mais baixa ou, caso se disponha de informação acerca da gama de distância dos pontos da cena, a imposição de limites aos valores de distância medidos.

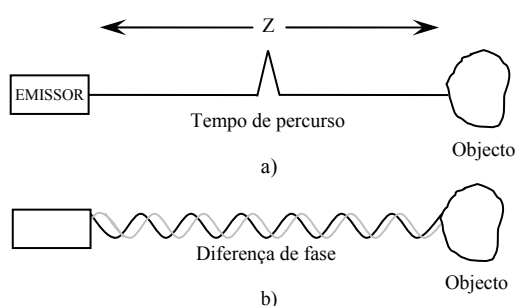


Figura 13 – Técnicas de detecção de ecos: medição de distância usando: a) tempo de percurso e b) diferença de fase.

As principais vantagens das técnicas de detecção de ecos são permitirem obter informação em cenas visualmente uniformes e controlarem a densidades de pontos das imagens obtidas. No entanto, têm a desvantagem de não poderem ser aplicadas a superfícies com muito elevada ou muito baixa reflectividade, transparentes ou translúcidas.

3.3 Estereoscopia Activa

As técnicas de luz estruturada [Awcock, 1996; Chen, 2000; Jain, 1995; Paiva, 1997; Silva, 1993, 1994], também designadas por estereoscopia activa, baseiam-se na projecção controlada de luz (um feixe, um plano ou outro padrão de luz) sobre a cena e na captação da luz reflectida pela cena por um sensor (câmara ou sensor linear),

afastado do projector. Em ambientes em que a iluminação pode ser controlada, estas técnicas são preferíveis relativamente às técnicas passivas, dado que a luz projectada dá origem a pontos luminosos que podem ser detectados na imagem com maior facilidade. No entanto, a dificuldade da obtenção desses pontos característicos varia consoante o tipo de padrão projectado.

Das técnicas de luz estruturada, a mais simples consiste na projecção de um feixe de luz que determina a iluminação de um ponto da cena de cada vez. As coordenadas tridimensionais do ponto em questão podem ser determinadas por triangulação depois de conhecida a posição e orientação do projector relativamente à câmara. A obtenção da informação tridimensional de toda uma cena, usando esta técnica, implica que se faça um varrimento da cena com um feixe de luz (Fig. 14). Assim, para uma imagem de $n \times m$ pontos são necessárias $n \times m$ aquisições de imagens, tornando-se portanto, num processo lento.

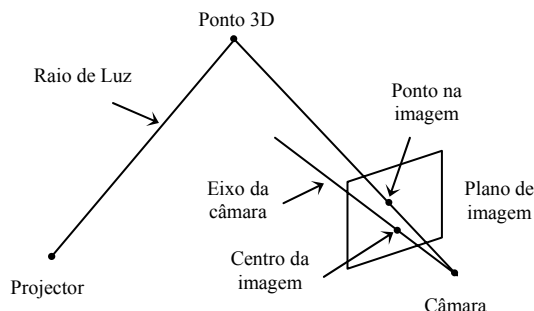


Figura 14 – Projecção de um raio de luz.

A projecção de um plano de luz, em vez de um feixe, tem a vantagem de permitir um varrimento mais rápido da cena (Fig. 15). Assim, para uma imagem de $n \times m$ pontos são necessários n ou m aquisições de imagens conforme os planos sejam projectados horizontal ou verticalmente. As técnicas que utilizam este tipo de projecção tem sido muito usadas para a inspecção, identificação e posicionamento de objectos tridimensionais.

Para eliminar a necessidade de varrimento da cena pode ser usada a projecção de vários planos de luz paralelos (ou em grelha), contudo, o estabelecimento das correspondências entre os pontos

projectados e os pontos visíveis na imagem captada pela câmara torna-se num problema mais complexo, dado que os pontos projectados não se diferenciam uns dos outros (Fig. 16).

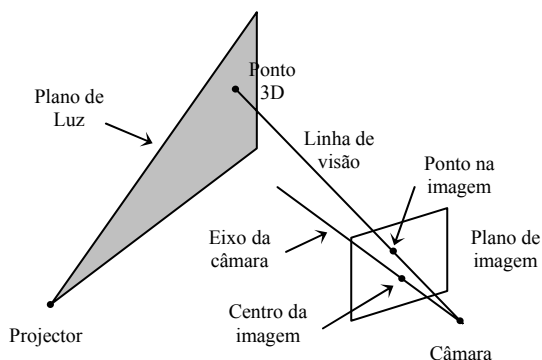


Figura 15 – Projeção de um plano de luz.

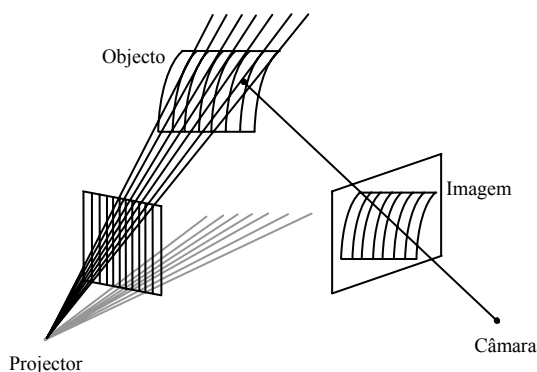


Figura 16 – Projeção simultânea de vários planos de luz.

Ao longo dos anos foram propostas várias soluções para resolver, ou pelo menos atenuar, o problema do estabelecimento de correspondências dos sistemas que recorrem à projecção controlada de luz. Em seguida são apresentadas algumas dessas propostas.

3.3.1 Técnica de modulação temporal

A técnica de modulação temporal foi proposta por *Altschuler* e *Posdamer* [Altschuler, 1982; Paiva, 1997; Rechsteiner, 1996; Silva, 1994; Wahl, 1984] e utiliza uma matriz de feixes laser controlados por um obturador. Esta matriz corresponde aos planos de *bits* de um código binário, sendo atribuído a cada feixe um código binário único. Começa-se por criar uma imagem de referência projectando todos os feixes simultaneamente, em seguida são projectados vários padrões,

sendo cada um definido pelos planos de *bits* correspondentes aos códigos dos feixes. O feixe de referência é alterado quando os *bits* do código valem 0 (zero), mantendo-se inalterado quando valem 1 (um). O sistema codifica os feixes em grupos, segundo as colunas (Fig. 17).

Se o sistema gerar na cena $N - 1$ planos de luz, serão necessárias $\log_2 N$ projecções. No caso de se pretender obter uma imagem de distância densa, esta técnica torna-se bastante lenta. No entanto, o número de projecções necessárias para obter os pontos da cena é bastante inferior aquele necessário quando se faz um varrimento da cena com um plano de luz.

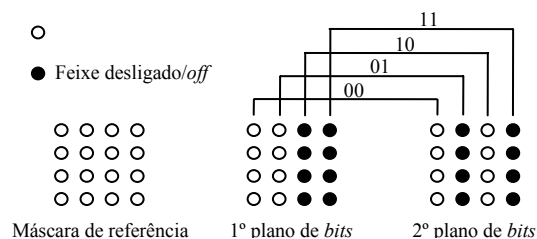


Figura 17 – Técnica de modulação temporal: codificação dos feixes usando um plano de *bits*.

A principal vantagem desta técnica é permitir a identificação dos pontos ou linhas projectados sem ambiguidades, bem como a identificação daqueles que não são detectados pela câmara.

3.3.2 Técnica de projecção de um padrão codificado através da cor

Boyer e *Kak* propuseram um sistema que ilumina a cena através da projecção de um único padrão com linhas vermelhas, verdes, azuis e brancas formando subconjuntos de quatro linhas [Boyer, 1987; Gonzalez, 2002; Jain, 1995; Paiva, 1997; Silva, 1994]. Cada um desses subconjuntos é identificado pela sequência de cores presentes no padrão projectado, evitando assim ambiguidades (Fig. 18). O trabalho proposto por *Boyer* e *Kak* teve como objectivo demonstrar que era possível extrair as linhas projectadas e determinar o seu índice de posição.

A principal vantagem desta técnica é a obtenção de uma imagem de distância com base na aquisição de uma única imagem de

intensidade. No entanto, esta técnica está limitada a cenas com cores neutras. A principal desvantagem reside no facto de poder existir iluminação mútua entre

objectos da cena, alterando assim a cor das linhas captadas pela câmara, conduzindo à determinação de índices de posição errados.

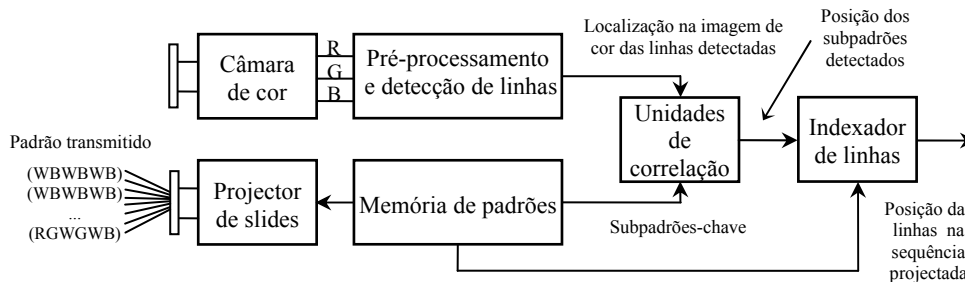


Figura 18 – Técnica de projecção de um padrão codificado através da cor: esquema básico do sistema, imagem adaptada de [Boyer, 1987].

3.3.3 Técnica de etiquetagem espacial

A técnica de etiquetagem espacial foi proposta por *Le Moigne* e *Waxman* [Le Moigne, 1988; Paiva, 1997; Silva, 1993, 1994] e consiste em projectar uma grelha de luz, à qual são acrescentados alguns quadrados iluminados (Fig. 19). Estes quadrados são usados como referência para se proceder à identificação dos pontos de cruzamento da grelha. O projector e a câmara são posicionados de forma a que as linhas verticais da grelha projectada apareçam igualmente na vertical na imagem captada. Após a extracção das linhas verticais e horizontais e dos pontos de cruzamento, estes pontos são etiquetados partindo de um dos pontos de referência, seguindo as linhas detectadas e contando os pontos de cruzamento já considerados.

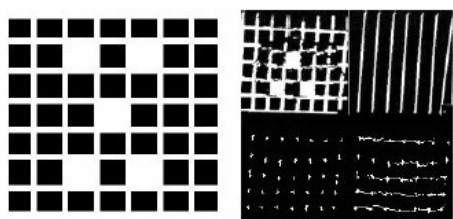


Figura 19 – Técnica de etiquetagem espacial: grelha com pontos de referência e passos de processamento, imagem de [Silva, 1994].

Uma das principais desvantagens desta técnica é não poder ser aplicada em situações em que ocorrem descontinuidades nas linhas, quer por não ser possível detectar todos os pontos, quer devido a descontinuidades da cena, ou quando os pontos de cruzamento não são detectados. Este problema pode ser resolvido

recorrendo a várias etiquetagens, seguindo caminhos diferentes, e usando uma técnica de relaxação para resolver eventuais ambiguidades na etiquetagem.

3.3.4 Técnica de projecção de um padrão binário

Esta técnica foi proposta por *Vuylsteke* e *Oosterlinck* [Paiva, 1997; Silva, 1993, 1994; Vuylsteke, 1990] e baseia-se na projecção de um padrão binário constituído por quadrados opacos e transparentes, em forma de xadrez, cujos vértices de junção são sobrepostos por outros mais pequenos, também opacos ou transparentes conforme representam o código binário 0 (zero) ou 1 (um) (Fig. 20). Estes pequenos quadrados são os pontos característicos que se pretende identificar. Assim, o padrão é construído de tal forma que o plano vertical ao qual um ponto característico pertence pode ser identificado por análise do padrão binário de uma janela definida em torno do ponto.

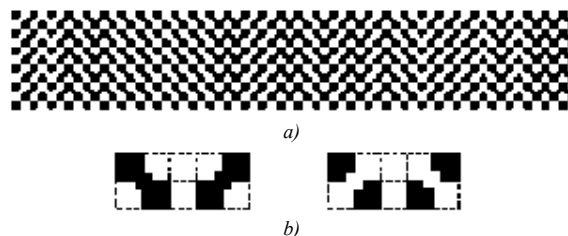


Figura 20 – Técnica de projecção de um padrão binário (imagens de [Paiva, 1997]): a) Padrão binário usado por *Vuylsteke* e *Oosterlinck*; b) Constituintes básicos desse padrão, ampliados.

A principal vantagem desta técnica, relativamente às técnicas de modelação temporal apresentadas anteriormente, é a de

permitir identificar os pontos característicos usando uma única imagem da cena. No entanto, tem a desvantagem de os pontos característicos não poderem ser identificados quando ocorrem descontinuidades na janela definida em torno desses pontos, ou a binarização ser inadequada. Outra desvantagem desta técnica é o elevado tempo computacional e a limitada densidade espacial da informação tridimensional obtida.

3.3.5 Técnica de projecção de um padrão codificado com perfil sinusoidal

A técnica de projecção de um padrão codificado com perfil sinusoidal foi proposto por Gomes [Coggrave, 2000; Gomes, 1992; Paiva, 1997; Silva, 1994] e baseia-se na codificação do espaço onde se situa um objecto em estudo, usando iluminação com uma configuração de intensidade de perfil periódico sinusoidal (Fig. 21). Primeiro, esse padrão codificado com período e geometria conhecidos é projectado sobre um plano de referência e todos os pontos desse plano são caracterizados por um valor de fase relativo a um ponto de referência. Em seguida, são determinadas as coordenadas tridimensionais de cada ponto do objecto a partir da diferença entre os registos de fase da intensidade das imagens codificadas e do plano de referência.

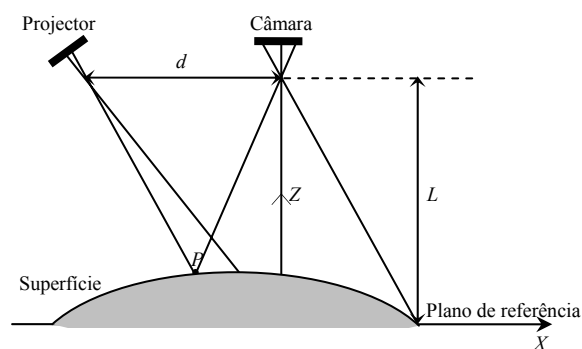


Figura 21 – Técnica de projecção de um padrão codificado com perfil sinusoidal: a altura h é medida a partir da diferença entre registos de fase da intensidade das imagens da cena e do plano de referência.

Esta técnica tem a vantagem de permitir obter imagens de distância densas com erros pequenos. A principal dificuldade é a

que resulta da necessidade de ter pontos vizinhos com valores de fase contínuos, cuja as variações sejam inferiores a $\pm \pi$. Esta necessidade implica o ajuste do padrão sinusoidal à inclinação da superfície em estudo, dificultando a determinação de distância em superfícies com descontinuidades.

Na subsecção seguinte é apresentada uma variante desta técnica de aquisição de informação tridimensional, aplicável a superfícies com descontinuidades.

3.3.6 Técnica de *Temporal Phase Unwrapping*

A técnica de *temporal phase unwrapping* [Coggrave, 1999; Coggrave, 2000; Huntley, 1993; Saldner, 1997] é uma das técnicas activas de projecção de um padrão de luz/sombra com perfil sinusoidal apresentado na subsecção anterior (3.3.5). Esta técnica é aplicável ao problema da existência de descontinuidades no perfil dos objectos e consiste na projecção de franjas para medição desse perfil. O padrão de franjas projectado é formado por linhas paralelas com perfil de intensidade aproximadamente sinusoidal. A ideia base desta técnica é variar a largura das franjas ao longo do tempo, de forma a obter uma sequência de mapas de fase. Cada mapa de fase é obtido a partir da projecção de um padrão de franjas (com a mesma largura) cujo valor de fase varia num intervalo de $-\pi$ a $+\pi$ como ilustra a Fig. 22. A sequência de mapas de fase forma a distribuição de fase tridimensional do objecto que se pretende reconstruir.

As franjas projectadas são geradas de forma a que a intensidade da luz que passa pela coordenada do *pixel* (i, j) do *SLM*, $(i, j = 0, 1, 2, \dots, N - 1)$, seja dada por:

$$I_{SLM}(i, j, k, t) = I_0 + V \cos \left\{ 2\pi \left(t \left(\frac{i}{N} - \frac{1}{2} \right) + \frac{k-1}{4} \right) \right\}$$

onde I_0 é a intensidade média, V é a visibilidade da franja, k é o índice do salto de fase ($k = 1, 2, 3, 4$), e t é o índice da largura da franja definindo o número de

franjas no vector imagem ($t = 1, 2, \dots, s$). A sequência de padrões de franjas é enviada para o *SLM*, começando na imagem $(t, k) = (1, 1)$ e abrangendo todas as combinações de índices de largura de franja e de salto de fase.

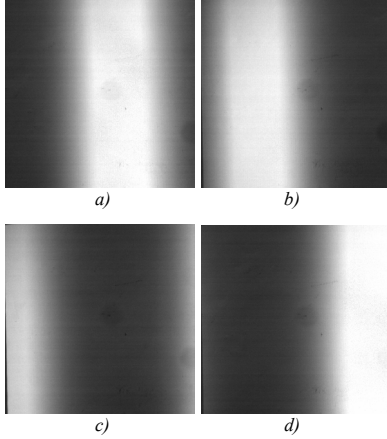


Figura 22 – Padrão de franjas com perfil de intensidade aproximadamente sinusoidal para construção de um mapa de fase. Imagem (t, k) : a) (1,1), b) (1,2), c) (1,3), d) (1,4).

A distribuição de fase das franjas medidas pela câmara pode ser calculada usando, por exemplo, a técnica *quatro-janelas*, que consiste em determinar cada mapa de fase $\Phi(m, n, t)$ a partir de quatro janelas (imagens) de intensidade, cuja fase em cada uma é, respectivamente, $-\pi, -\pi/2, 0, \pi/2$. Para cada novo mapa de fase a determinar, t é incrementado provocando o aumento de uma franja no padrão projectado; a franja central mantém-se fixa, com as restantes franjas a moverem-se nos limites do *SLM*. Para o método das *quatro-janelas*, o t -ésimo mapa de fase, $\Phi(m, n, t)$, pode ser calculado como:

$$\Delta\Phi(m, n, t) = \tan^{-1} \left[\frac{\Delta I_{42}(m, n, t) \Delta I_{13}(m, n, t-1) - \Delta I_{13}(m, n, t) \Delta I_{42}(m, n, t-1)}{\Delta I_{13}(m, n, t) \Delta I_{13}(m, n, t-1) - \Delta I_{42}(m, n, t) \Delta I_{42}(m, n, t-1)} \right]$$

onde,

$$\Delta I_{ij}(m, n, t) = I(m, n, i, t) - I(m, n, j, t).$$

Quando determinado o s -ésimo mapa de fase, a variação de fase “desembrulhada” total $\Psi(m, n)$ pode ser calculada somando as diferenças de fase, ou seja,

$$\Psi(m, n) = \Phi(m, n, s) - \Phi(m, n, 0) = \sum_{t=1}^s \Delta\Phi(m, n, t).$$

O que distingue a técnica de *temporal phase unwrapping* das restantes técnicas espaciais de “desembrulhar” da fase é o facto deste processo ocorrer ao longo do índice t , em vez dos índices m ou n . Os cálculos, em cada *pixel*, são realizados independentemente dos restantes *pixels*, evitando as dificuldades causadas pelas descontinuidades da cena, normalmente, encontradas nas outras técnicas espaciais de “desembrulhar” da fase.

4 SUMÁRIO

O presente artigo descreve algumas técnicas que têm sido utilizadas em sistemas de aquisição de informação tridimensional no domínio de Visão 3D. Assim, foram descritas técnicas que utilizam apenas a luz ambiente para iluminar a cena, designadas técnicas passivas, e técnicas que utilizam projecção directa e controlada de energia sobre a cena, designadas técnicas activas. As técnicas passivas foram divididas em duas categorias: a primeira, onde se inserem as técnicas baseadas em imagens monoculares de intensidade que extraem a informação de distância a partir de uma ou mais imagens monoculares de intensidade captadas com uma câmara fixa; e a segunda, onde se inserem as técnicas baseadas em estereoscopia passiva que permitem obter informação de distância a partir de duas ou mais imagens de intensidade, captadas com uma ou mais câmaras colocadas em posições distintas. Por sua vez, as técnicas activas foram divididas em três categorias: a primeira, designada por técnica de *Moiré*, consiste na projecção de um conjunto de linhas paralelas e na aquisição de uma imagem da mesma, utilizando uma câmara em frente da qual está colocada um grelha idêntica à usada para projectar as linhas; a segunda, designada por técnica de detecção de ecos que consiste na determinação de distância a partir do tempo de percurso desde que uma fonte de energia é emitida até ser detectada; e a terceira, onde se

inserir as técnicas baseadas em estereoscopia activa que se caracterizam pela projecção controlada de luz sobre a cena e na captação da luz reflectida por esta a partir de um sensor, afastado do projector.

Todas as técnicas descritas neste artigo apresentam dificuldades e limitações, por essa razão não existe uma técnica que possa ser considerada “a melhor em geral”. Contudo, em cenas de interior, com ambientes controlados, as técnicas activas são mais utilizadas e em cenas de exterior ou ambientes não controlados são, geralmente, utilizadas as técnicas passivas.

5 REFERÊNCIAS

- Altschuler, M.D., Bae, K., Robot Vision by Encoded Light Beams in Three-Dimensional Robot Vision, Ed. T. Kanade, pp. 97-149, 1982
- Awcock, G.W., Thomas, R., Applied Image Processing, McGraw-Hill International Editions – Electrical & Electronic Series, 1996
- Benoit, P., Mathieu, E., Hormière, J., Thomas, A., Characterization and Control of Threedimensional Objects Using Fringe Projection Techniques, *Nouv. Rev. Optique*, Vol. 6, No. 2, pp.67-86, 1975
- Boyer, K.L., Kak, A.C., Color-Encoded Structured Light for Rapid Active Ranging, *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 9, No. 1, pp. 14-28, Jan. 1987
- Brooks, R.E., Heflinger, L.O., Moiré Gauging Using Optical Interference Patterns, *Applied Optics*, Vol. 8, No. 5, pp.935-939, May 1969
- Chen, F., Brown, G.M., Song, M., Overview of Three-Dimensional Shape Measurement Using Optical Methods, *Opt. Eng.*, Vol. 39, No. 1, pp. 10-22, Jan. 2000
- Coggrave, C.R., Huntley, J.M., High-Speed Surface Profilometer Based on a Spatial Light Modulator and Pipeline Image Processor, *Opt. Eng.*, Vol. 38, No. 9, pp. 1573-1581, Sept. 1999
- Coggrave, C.R., Huntley, J.M., Optimization of a Shape Measurement System Based on Spatial Light Modulators, *Opt. Eng.*, Vol. 39, No. 1, pp. 91-98, Jan. 2000
- Gomes, A.S., Techniques Utilisant La Lumière Structurée, *Innov. Tech. Biol. Med.*, Vol. 13, No. spécial 1, 1992
- Gonzalez, R.C., Woods, R.E., Digital Image Processing, Prentice Hall, Second Edition, 2002
- Huntley, J.M., Saldner, H.O., Temporal Phase-Unwrapping Algorithm for Automated Interferogram Analysis, *Applied Optics*, Vol. 32, No. 17, pp. 3047-3052, Jun. 1993
- Jähne, B., Haubecker, H., Computer Vision and Applications, Academic Press, 2000
- Jain, R., Kasturi, R., Schunck, B.G., Machine Vision, McGraw-Hill International Editions – Computer Science Series, 1995
- Le Moigne, J.J., Waxman A.M., Structured Light Patterns for Robot Mobility, *IEEE Journal of Robotics and Automation*, Vol. 4, No. 5, pp. 541-548, Oct. 1988
- Paiva, A.R., Dissertação de Mestrado: Aquisição e Processamento de Informação Tridimensional com Luz Estruturada Codificada, Faculdade de Engenharia da Universidade do Porto, 1997
- Rechsteiner, M., Schneuwly, B., Guggenbühl, W., Fast and Precise 3D-Sensor Insensitive to Ambient Light, Electronics Laboratory ETH Zurich, 1996
- Saldner, H.O., Huntley, J.M., Profilometry Using Temporal Phase Unwrapping and a Spatial Light Modulator-Based Fringe, *Optical Engineering*, Vol. 36, No. 2, pp. 610-615, Feb. 1997
- Silva, J.A., Campilho, A.C., Marques, J. C., Structured-light Techniques: Basic Principle and Comparative Analysis, Faculdade de Engenharia da Universidade do Porto, 1993
- Silva, J.A., Tese de Doutoramento: Aquisição e Processamento de Informação

Tridimensional, Faculdade de Engenharia da Universidade do Porto, 1994

Tavares, J.M., Dissertação de Mestrado: Obtenção de Estrutura Tridimensional a Partir de Movimento de Câmara, Faculdade de Engenharia da Universidade do Porto, 1995

(<http://www.fe.up.pt/~tavares/downloads/publications/teses/tese%20msc.zip>)

Vuylsteke, P., Oosterlinck, A., Range Image Acquisition with a Single Binary-Encoded Light Pattern, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 12, No. 12, pp. 148-164, Feb. 1990

Wahl, F.M., A Coded-Light Approach for 3-Dimensional (3D) Vision, IBM Zurich Research Laboratory, Switzerland, September 1984