

2

Visão Computacional e Obtenção de Formas

Neste capítulo são abordadas as principais características e áreas de atuação da Visão Computacional e seu caráter multidisciplinar. Depois são apresentados os conceitos teóricos da captura de formas a partir de estéreo ativo, focando nos diversos métodos de codificação de luz estruturada, seus benefícios e limitações.

2.1.

Visão Computacional

Visão Computacional foi definida em [1] como sendo o conjunto de técnicas computacionais para estimar ou explicitar as propriedades geométricas e dinâmicas do mundo tridimensional a partir de imagens.

A palavra visão está relacionada à capacidade que um ser tem de ver e entender o mundo em que habita. A imagem é formada na mente através das organizações física, química e biológica dos olhos. Em cima da visão, o cérebro realiza diversas funções. Por exemplo, em um movimento para se pegar um copo, a visão tem papel fundamental junto com a coordenação motora. Através da visão guiamos a nossa mão até o copo, encaixando os dedos no seu corpo cilíndrico. Gestos simples como esses realizamos aos milhares por dia sem nos darmos conta, porém que se tornam extremamente difíceis sem a visão, como andar. Diariamente utilizamos nossa visão para reconhecer objetos, pessoas, calcular distâncias, verificar se há buracos na rua, nuvens cinzas no céu, encontrar uma vaga livre no estacionamento etc.

O mesmo se sucede na Visão Computacional. Ao invés dos olhos, câmeras são utilizadas para se obter imagens digitais. Em cima destas imagens são aplicadas técnicas computacionais para extrair informações desejadas do mundo tridimensional. Estas informações variam muito de natureza e podem ser empregadas para variadas aplicações em diferentes áreas. Na medicina, imagens de órgãos e células são utilizadas no diagnóstico e tratamento de doenças. Na robótica a contribuição é ampla com identificação de objetos, localização,

inspeção e locomoção. Há também aplicações militares para guiar mísseis e veículos através de trajetórias pré-definidas. A astronomia utiliza imagens de satélites e telescópios para identificar a composição química de planetas e criar um modelo da sua superfície enquanto que a indústria automobilística utiliza imagens digitais em sistemas de inspeção industrial. Estes são alguns exemplos de áreas e aplicações onde são empregadas técnicas de Visão Computacional.

Computação Gráfica e Processamento de Imagens são duas áreas correlacionadas à Visão Computacional. De uma maneira geral, a Computação Gráfica utiliza informações, como grafos de cena, tipos de materiais, geometria dos objetos, tipo de projeção utilizada etc. para gerar imagens. A Visão Computacional faz o caminho contrário, procurando obter informações de cenas a partir de imagens digitais previamente capturadas. Para isso quase sempre faz uso de métodos de Processamento de Imagens. O interesse nestes métodos surgiu da necessidade de melhorar a qualidade da informação pictórica para a interpretação humana. Hoje em dia estes métodos são também utilizados em aplicações de Visão Computacional, de maneira que as duas áreas se confundem.

Por fim, apresentamos uma outra definição, segundo Vishvjit Nalwa [2]: a Visão Computacional visa inferir e descrever automaticamente estruturas e propriedades de um mundo 3D, possivelmente dinâmico, a partir de uma ou múltiplas imagens 2D do mundo. As imagens podem ser coloridas ou monocromáticas, captadas por um ou múltiplos sensores que, por sua vez, podem ser estacionários ou móveis.

2.2.

Forma a partir de Estéreo Ativo

Uma das áreas de pesquisa da Visão Computacional é a aquisição de modelos geométricos de objetos reais a partir de imagens. O processo de se obter a forma e aparência de objetos reais, utilizando-se câmeras e luzes, é chamado de fotografia 3D. A rápida evolução de equipamentos eletrônicos, como câmeras e projetores, permite que se monte sistemas de aquisição de baixo custo e cada vez mais eficazes. Modelos geométricos de objetos reais são utilizados por diferentes áreas, tipicamente design industrial, jogos, educação, arte, arqueologia, comércio

eletrônico etc. Com o advento de métodos e equipamentos, mais áreas poderão fazer uso destes modelos como, por exemplo, inspeção industrial.

Uma classificação dos métodos para aquisição de formas é dada na figura 1. Existem diferentes métodos ópticos para aquisição de formas, cada um com suas qualidades e limitações, tais como resolução e precisão. A escolha deve ser feita tendo em vista o contexto de utilização. Por exemplo, modelar um navio é diferente do que modelar um carro. As principais limitações dos métodos ópticos é que estes podem obter apenas porções visíveis da superfície e são sensíveis a suas propriedades como transparência, brilho, textura, falta de cor etc. Esta dissertação está focada no estudo dos métodos ópticos, mais precisamente em métodos estéreos ativos, em destaque na figura.

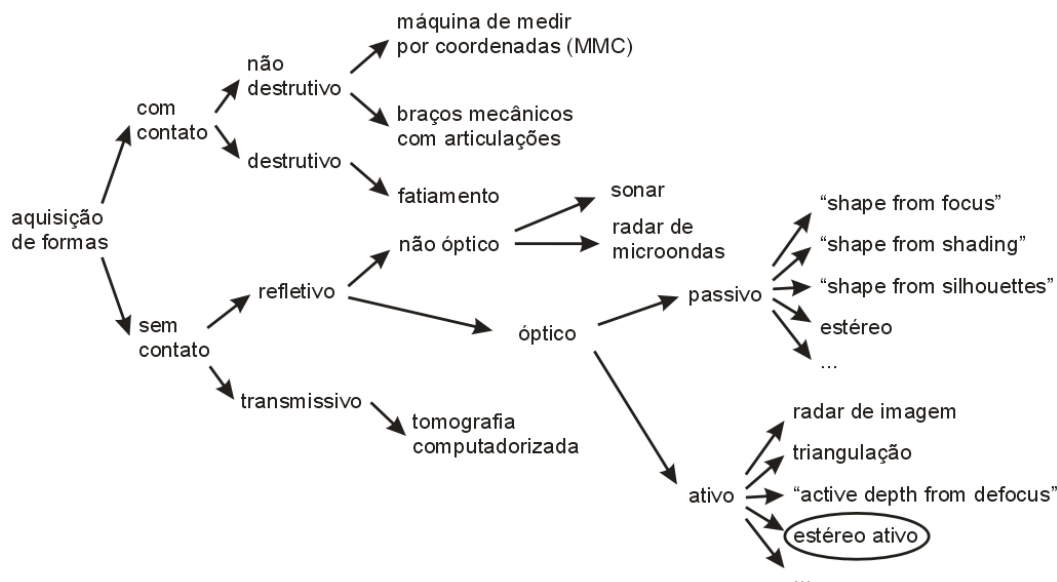


Figura 1 – Classificação dos métodos de aquisição de formas.

Os princípios básicos utilizados pelos métodos ópticos são os mesmos da visão estéreo. Se duas câmeras observam um mesmo ponto na cena, então a sua posição pode ser obtida através da interseção dos raios correspondentes à projeção na imagem de cada câmera. Esse processo é chamado de triangulação.

A grande dificuldade de se utilizar visão estéreo para recuperar formas 3D é realizar a correspondência entre os pontos nas duas imagens. Os algoritmos que realizam esta correspondência têm dificuldades quando a cena é muito complexa ou simples demais. Por exemplo, objetos lisos e monocromáticos geram uma falta

de textura nas imagens capturadas, impossibilitando o processamento da correspondência.

Uma maneira de contornar este problema é empregar o método estéreo ativo. Utilizando uma fonte de luz, a cena é marcada por padrões conhecidos. Esta fonte de luz é chamada de luz estruturada. Cada tipo de luz estruturada possui um conjunto de padrões próprios que são projetados sobre a cena para marcá-la. Por exemplo, podemos projetar um padrão de apenas um ponto com um dispositivo laser. Este ponto é projetado sobre o objeto, cuja forma gostaríamos de obter. Basta então encontrar este mesmo ponto nas duas imagens obtidas e fazer a triangulação posteriormente para encontrar a sua localização no espaço. Poderíamos repetir este processo deslocando o ponto de lugar para obtermos mais pontos do modelo. Porém este procedimento seria extremamente demorado. Daí a necessidade de utilizarmos luz estruturada para acelerar a velocidade de captura.

Podemos, por exemplo, projetar um padrão com vários pontos sobre a cena. Para distinguir entre os diversos pontos e realizar a triangulação, devemos codificá-los apropriadamente de maneira a identificá-los posteriormente para que o problema de correspondência possa ser feito sem ambigüidades. Assim estaremos acelerando a velocidade de captura ao aumentar o número de pontos projetados simultaneamente.

A figura 2 mostra um exemplo de sistema estéreo ativo utilizado em [3]. O projetor é usado para projetar padrões de luz estruturada com listras codificadas sobre a cena que é capturada por uma câmera digital.

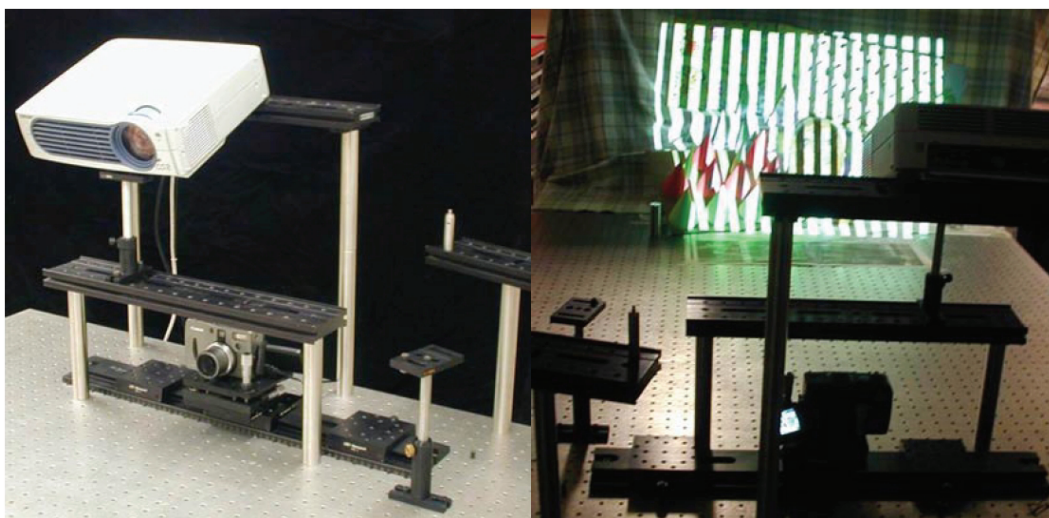


Figura 2 – Exemplo de sistema estéreo ativo capturando uma cena.

Podemos utilizar duas configurações distintas. A primeira seria manter as duas câmeras, enquanto que a luz estruturada serviria para marcar o objeto com vários pontos distintos que são utilizados posteriormente na triangulação. A segunda seria utilizar uma câmera e uma fonte de luz calibrada, onde sua posição e direção de projeção são conhecidas. A triangulação é feita entre a câmera e o projetor. Para cada ponto codificado no padrão projetado podemos traçar um raio, que sai do centro da fonte de luz, de maneira semelhante com o que fazemos com uma câmera calibrada.

Em resumo, os três passos importantes para recuperar formas com um sistema estéreo ativo são:

- Calibrar as câmeras.
- Estabelecer a correspondência entre os pontos nas imagens e os elementos dos padrões projetados.
- Reconstrução das coordenadas 3D de pontos da cena.

A calibração de câmera é tratada no próximo capítulo, assim como a reconstrução das coordenadas 3D de pontos da cena. Esta seção se concentra no segundo passo. Dois estudos foram feitos a respeito de métodos que utilizam luzes estruturadas codificadas para resolver o problema de correspondência em [4] e [5]. Esta seção foi baseada neste último e não tem por objetivo reproduzir todas as descrições de métodos de codificação de luz estruturada.

As diferentes abordagens para codificação da luz estruturada são apresentadas a seguir.

2.2.1. Codificação Temporal

A codificação temporal é o método de codificação mais utilizado pelos trabalhos atuais. A idéia principal se baseia em projetar uma seqüência de padrões onde cada elemento do padrão codifica os dígitos binários de um código. Como mostra a figura 3, padrões com listras são projetados sobre a cena seqüencialmente. O código binário é convertido para listras brancas e pretas. As brancas correspondem ao dígito binário um, enquanto que as listras pretas representam o dígito binário zero.

Para distinguir entre as diversas listras, deve-se codificá-las apropriadamente de maneira que o problema de correspondência possa ser resolvido sem ambigüidades. Isso é feito projetando-se uma seqüência de padrões com listras, um após o outro, em uma codificação temporal.

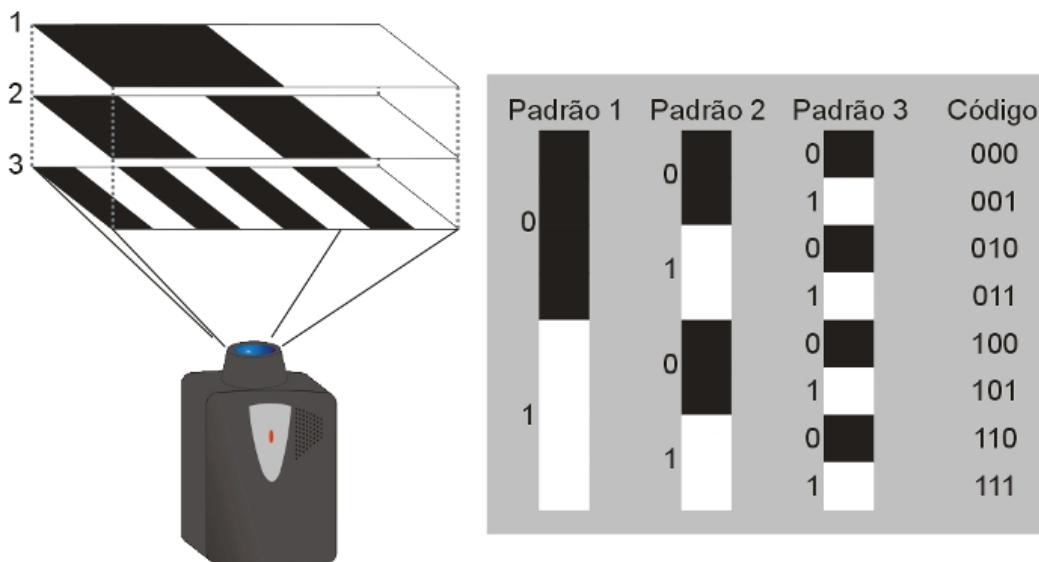


Figura 3 – Codificação temporal binária.

Os padrões são projetados em diferentes momentos, em uma seqüência definida. Cada listra em um padrão projetado corresponde a um dígito binário, de maneira que a seqüência de padrões fornece os dígitos de um código binário. Cada listra é univocamente determinada pelo seu código. A figura 3 mostra que a seqüência de n padrões produz 2^n listras codificadas e a resolução cresce exponencialmente com o número de padrões utilizados. Repetimos a mesma seqüência de padrões, porém com listras horizontais. No final, teremos iluminado a cena estática com $2^n \times 2^n$ regiões, cada uma com seu código horizontal e vertical. A figura 4 ilustra o aspecto de um objeto iluminado por um dos padrões da luz estruturada codificada.

Em [6] foi proposta a codificação temporal binária apresentada acima. Já em [7] foi proposta a mesma codificação binária, porém para cada padrão projetado também se projetava o seu inverso correspondente, onde as linhas pretas eram projetadas brancas e vice-versa. Isso permite uma melhor identificação de cada listra, como iremos ver na seção Captura e Processamento de Imagens do terceiro capítulo.

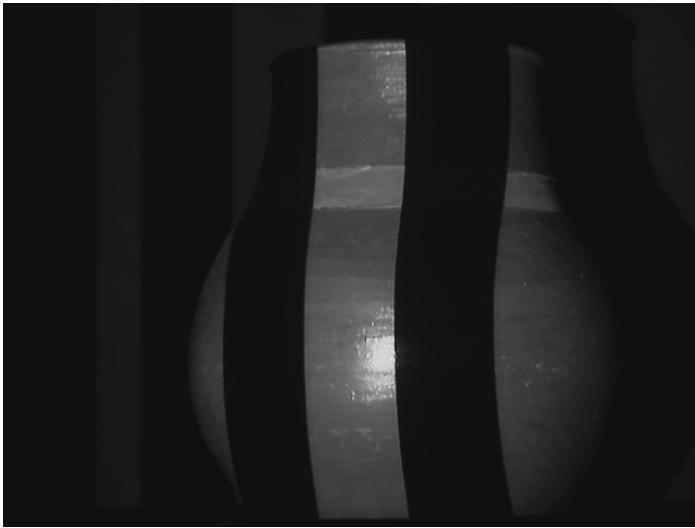


Figura 4 – Objeto iluminado por luz estruturada.

A codificação binária foi substituída pela codificação de Gray em [8]. Esta segunda codificação mais robusta, utilizada neste trabalho, é apresentada também no terceiro capítulo.

2.2.2. Codificação Espacial

A codificação espacial requer apenas um único slide a ser projetado, permitindo a aquisição de cenas dinâmicas. Porque se utiliza apenas um slide, este possui um maior número de elementos padronizados de maneira a garantir a resolução desejada. Na codificação espacial a vizinhança de um pixel é utilizada para codificar a sua posição.

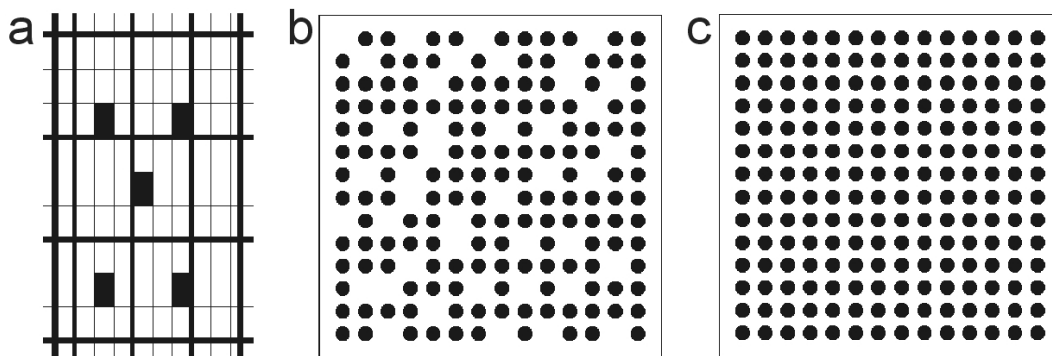


Figura 5 – Padrões de codificação espacial: (a) padrão de grid; (b) padrão com matriz codificada; (c) padrão inicial para matriz codificada.

A maior dificuldade deste método é recuperar o padrão projetado sobre o objeto modelado, limitando o tipo de superfícies capturadas. Os dois códigos seguintes são apresentados representando as várias abordagens existentes de codificação espacial.

Padrões de grid

Esta codificação consiste em utilizar um conjunto de listras como guia. Muitos padrões diferentes baseados neste princípio podem ser criados. Um exemplo pode ser visto na figura 5a onde a grade de listras é parcialmente codificada com quadrados usados como referência [9]. Ambos os eixos são codificados simultaneamente. A limitação desta abordagem é que descontinuidades na superfície do objeto produzem também uma descontinuidade nos identificadores, podendo fazer com que o decodificador se perca.

Matriz Codificada

Esta codificação, proposta em [10], consiste em projetar uma matriz de pontos. Cada linha horizontal da matriz codifica uma palavra. Inicialmente é projetada uma matriz contendo todos os pontos para extrair a posição de cada um deles. Posteriormente projeta-se a matriz codificada. As figuras 5b e 5c ilustram as matrizes utilizadas. A dificuldade deste método está em determinar cada ponto, que pode ter sua forma modificada pela superfície do objeto capturado. O tamanho de cada ponto deve manter uma relação com a resolução de captura, limitando o tamanho da matriz.

2.2.3. Codificação Modulada

Outra possibilidade é modular a intensidade da luz projetada, como foi proposta em [11]. A posição de um pixel pode ser codificada, sem utilizar informação de vizinhança, em função da intensidade da luz modulada. Projeta-se posteriormente um padrão de iluminação constante, o qual é utilizado também para a decodificação, limitando a sua aplicação a cenas estáticas. Utiliza-se a diferença de intensidade para recuperar a posição dos pixels.



Figura 6 – Padrão de luz modulada.

Esta abordagem é sensível a ruídos da câmera e a certas superfícies. Texturas podem interferir na decodificação, tornando o método não robusto.

2.2.4. Codificação com Cores

Os avanços tecnológicos na captura de imagens permitiram a utilização de cores nos padrões projetados. A vantagem está em se utilizar três canais simultaneamente ao invés de apenas um. O grande problema da sua utilização está na recuperação. Luz colorida em superfícies coloridas não se comporta de maneira desejada, o que limitava a sua utilização a cenas com cores neutras nos primeiros trabalhos.

Em [12] são utilizadas listras coloridas com fendas entre elas. A figura 7a ilustra este padrão. A codificação é feita através da sequência utilizada. A mesma idéia de se modular a intensidade da luz foi feita em [13] utilizando-se cores, gerando o padrão arco-íris da figura 7b. Já em [14] utilizou-se um padrão periódico, ilustrado na figura 7c, feito a partir da união dos padrões de intensidade senoidal das três cores fora de fase entre si, representadas na figura 7d.

Um estudo teórico para determinar o maior tamanho possível de uma matriz formada por pontos e codificada com um certo número de cores foi feito em [15]. Quando maior a matriz, maior a resolução obtida. A posição de cada ponto é codificada pela sua cor e as cores dos seus quatro vizinhos: superior, inferior, esquerdo e direito. Ambos os eixos são codificados simultaneamente. A figura 8a ilustra uma matriz gerada. Dependendo da cena, algumas vezes são difíceis de se determinar os vizinhos, como é ilustrado na figura 8b.

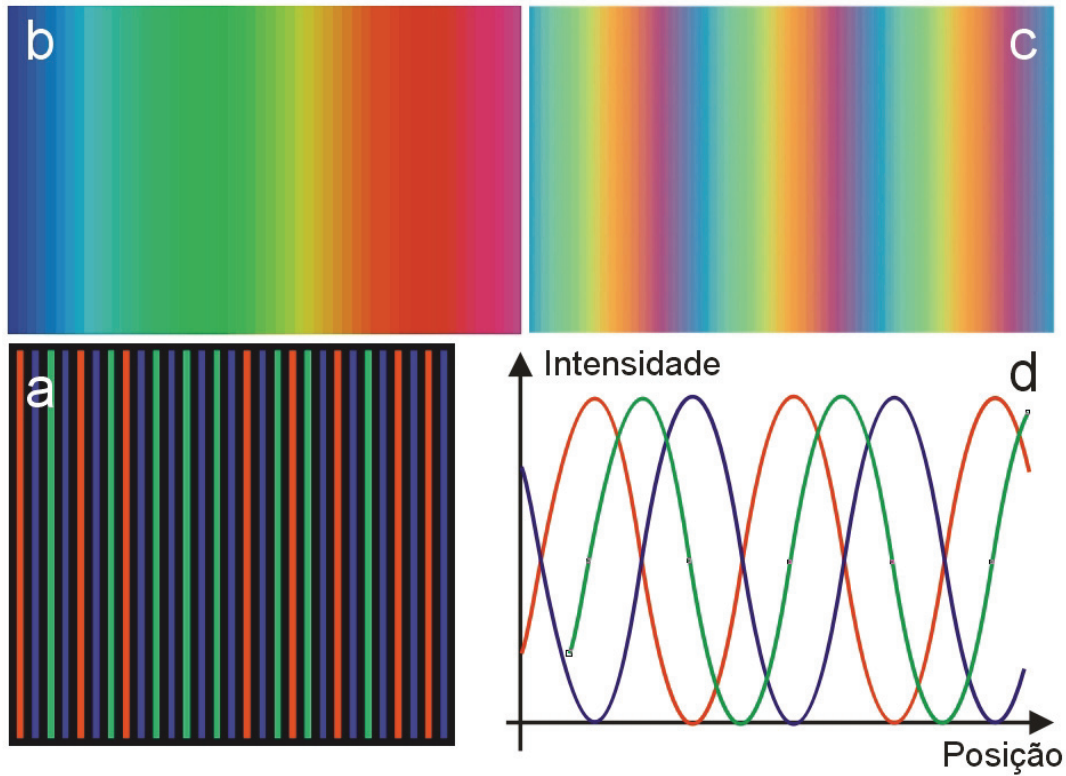


Figura 7 – Padrões com cores: (a) padrão com listras; (b) padrão arco-íris; (c) padrão periódico; (d) gráfico da intensidade das cores no padrão periódico.

Mais recentemente foi proposto em [5] a utilização de um código de Gray colorido, onde cada um dos três canais de cores contém um slide que corresponde a um slide da sequência de slides de Gray. Essa abordagem divide o número de slides necessários por três do padrão de Gray original para uma mesma resolução. Na prática esta redução permite capturar cenas dinâmicas utilizando um método de codificação temporal como mostram os resultados apresentados.

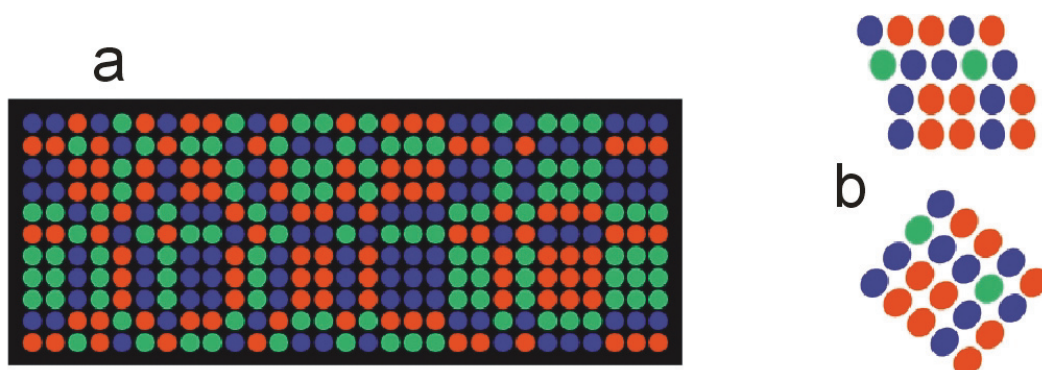


Figura 8 – Padrão com matriz de pontos coloridos: (a) exemplo de matriz gerada; (b) problemas ocorridos na projeção do padrão.

2.2.5. Classificação dos Métodos

Como vimos ao longo desta seção, há três possíveis maneiras de se codificar a luz projetada: codificação espacial, codificação temporal e modulação de cores através da intensidade projetada. Uma taxonomia foi proposta em [16] baseada na diferenças entre os padrões e as restrições que estes impõem à cena a ser capturada.

Modulação cromática impõe restrições às cores permitidas na cena. Quando codificação espacial é utilizada, continuidade local da cena é necessária para se recuperar o código transmitido. Já a codificação temporal restringe o movimento da cena. A tabela abaixo apresenta o resumo destas idéias.

Hipóteses	Métodos	Restrições
Coerência espacial	Métodos baseados na vizinhança	Descontinuidades na superfície do objeto não são permitidas
Coerência temporal	Muitos slides Versus Um slide	Restrito a cenas estáticas Versus Permite cenas dinâmicas
Refletividade	Binário / tons de cinza Versus Codificação com cores	Nenhuma restrição de cor Versus Objetos com cores neutras

Tabela 1 – Taxonomia para métodos de codificação de luz estruturada baseada nas suas características e restrições.

Uma outra taxonomia foi proposta em [5], na qual é feita uma analogia entre luz estruturada codificada e um sistema de comunicação digital. As coordenadas do projetor são codificadas pelos padrões e transmitidas para a cena. Para cada ponto na imagem da câmera uma transmissão com ruído é recebida e precisa ser decodificada. Portanto cada sensor da câmera age como um receptor e a imagem capturada como um conjunto de símbolos recebidos. O meio de transmissão é a superfície do objeto capturado e a mensagem é a posição codificada pelo pixel do projetor. Tendo em vista esta analogia, duas questões devem ser estudadas: as limitações do canal de transmissão, relacionado com as propriedades da superfície, e qual codificação utilizar, o que irá restringir os tipos de objetos capturáveis.

Ao contrário da taxonomia anterior, nesta outra é o código utilizado que impõe restrições à cena (meio de transmissão). Por exemplo, se utilizarmos codificação espacial, a cena deve preservar a estrutura espacial, caso contrário haverá perda de informação. As principais características de um código são o número de símbolos distintos (sinais básicos), o tamanho da palavra composta por símbolos e, no caso de codificação espacial, a geometria utilizada.

Método	Número de slides (tamanho da palavra)	Intensidade de modelação / canais (número de caracteres)	Vizinhança (região do caractere)	Resolução (tamanho do alfabeto)
Código de Gray	n	Binário (2) / monocromático	Pixel único	2^n linhas
Cód. de Gray colorido	n	Binário (2) / RGB	Pixel único	2^{3n} linhas
Padrão arco-íris	2	2^8 / RGB	Pixel único	$(2^8)^3$ linhas
Matriz de pontos	1	3 (R, G, B)	Quatro vizinhos	3^5

Tabela 2 – Taxonomia para métodos de codificação de luz estruturada baseada em transmissão de dados digitais.

Deste ponto de vista fica claro de perceber que o código impõe restrições aos objetos capturados. A estrutura espacial da codificação é perdida quando há descontinuidades na superfície do objeto capturado. Também é necessário recuperar a intensidade de modulação corretamente, o que requer que a superfície não distorça intensidades ou cores demasiadamente. Finalmente, o tamanho da palavra codificada impede que haja movimento enquanto a palavra ainda não esteja completa.

2.2.6. Sistema Estudado

O sistema estudado neste trabalho para aquisição de formas geométricas pode ser classificado como um sistema óptico estéreo ativo. Ele é composto por duas câmeras de vídeo e um projetor digitais ligados ao computador. O par de câmeras é calibrado simultaneamente com o mesmo padrão de calibração. Como foi visto, a correspondência entre pontos nas imagens das duas câmeras é feita projetando-se luz estruturada codificada na cena.

A codificação temporal foi escolhida para este trabalho, pois permite capturar uma maior variedade de cenas e objetos sem restrições de continuidade nas superfícies capturadas e de cores neutras. A princípio não estamos interessados em capturar cenas dinâmicas, mas na robustez da decodificação e na precisão do sistema. Utilizou-se o código binário de Gray monocromático que permite capturar cenas com diferentes cores e exige menos do hardware utilizado.

O modelo 3D final é constituído por pontos no espaço, obtidos triangulando pontos correspondentes nas duas câmeras. A figura 9 mostra a disposição física do sistema. Cada câmera é posicionada de um lado do projetor, de maneira a melhor capturar o par de imagens.

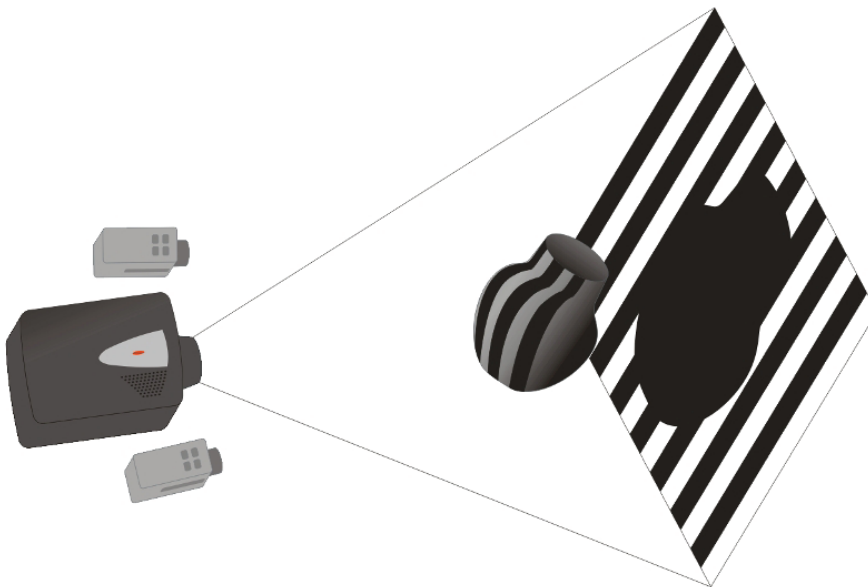


Figura 9 – Disposição física do sistema estudado.

O próximo capítulo apresenta os conceitos teóricos envolvidos no trabalho que não foram apresentados neste capítulo.