Camera Models

Van-Thinh Vo

June 2025

Mục lục

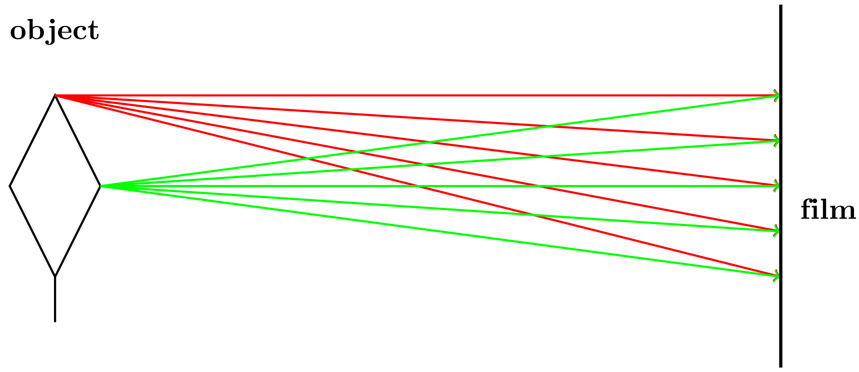
Nếu có bất kỳ sai sót nào,mọi người có thể phản hồi qua email thinh.vovan@hcmut.edu.vn để hỗ trợ mình với nha :v

1 Giới thiệu

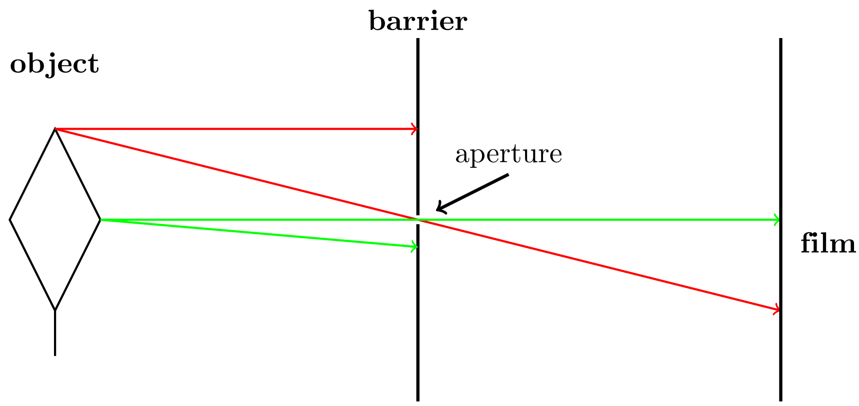
Đã có bao giờ bạn tự hỏi rằng bằng cách nào ta có thể thu được một bức ảnh gồm 3 kênh màu RGB với các điểm ảnh (pixel) từ thế giới thực hay chưa? Trong bài viết này, ta sẽ cùng nhau tìm hiểu về cách xây dựng mô hình camera ghi lại hình ảnh và "số hóa" nó nhé.

2 Máy ảnh lỗ kim (Pinhole Camera)

Ý tưởng đầu tiên và đơn giản nhất về camera là đặt một tấm film phía trước vật thể. Các tia sáng (**rays**) từ vật thể sẽ chiếu vào film và ta sẽ thu được hình của vật trên đó. Tuy nhiên, phương pháp này gặp phải một vấn đề rằng một điểm trên film có thể được chiếu bởi nhiều điểm khác nhau trên vật thể, do đó gây ra hiện tượng ảnh bị mờ.

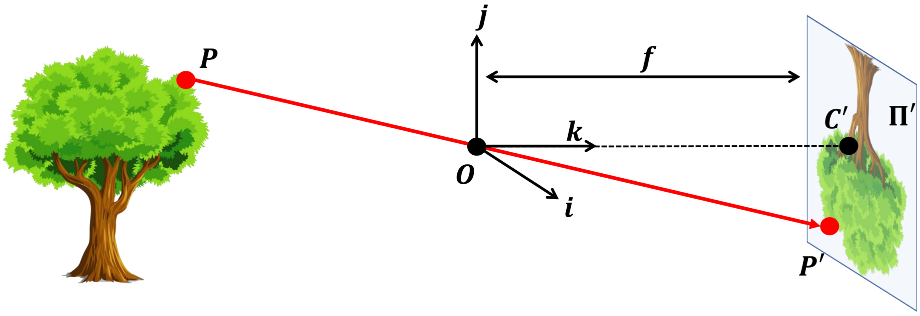


Một giải pháp cho trường hợp này là sử dụng một *màn chắn* (**barrier**) đặt giữa vật và film, ở giữa màn chắn có một khoảng hở đủ nhỏ gọi là *khẩu độ* (**aperture**) để ngăn một tia chiếu nhiều hơn một điểm trên film. Đây chính là *máy ảnh lỗ kim* (**pinhole camera**).



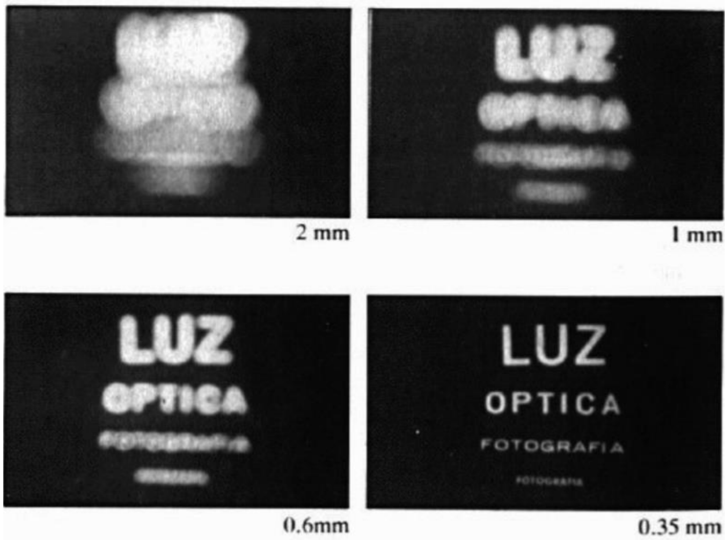
Ta cùng phân tích về mặt hình học của phép chiếu từ ảnh 3D sang màn ảnh. Đặt một hệ trục tọa độ với tâm O là điểm khẩu độ với hệ trục tọa độ $\begin{bmatrix} i & j & k \end{bmatrix}$ như hình [1]. Trục k đi qua tâm O gọi là *trục quang học* (**optical axis**). Khoảng cách từ màn chắn đến màn ảnh (film) f gọi là *tiêu cự* (**focal length**). Gọi $P(x, y, z)$ là một điểm trên vật thể và $P'(x', y', z')$ là hình chiếu của nó lên màn ảnh. Ta có hai tam giác đồng dạng $\Delta OC'P'$ và ΔOCP ($C(0, 0, z)$ là điểm đối xứng với C' qua O) nên $\overrightarrow{OP'} = \frac{f}{z}\overrightarrow{OP}$, vì thế tọa độ của P' (trong hệ trục tọa độ 2 chiều nằm trên màn ảnh có tâm C') sẽ là:

$$P' = \begin{bmatrix} f \times \dfrac{x}{z} \\ \\ f \times \dfrac{y}{z} \end{bmatrix} \tag{1}$$



Hình 1: Cấu trúc của một máy ảnh lỗ kim

Tuy nhiên, vì chỉ cho phép một tia sáng mỗi điểm đi qua khẩu độ nên ảnh sẽ bị tối.



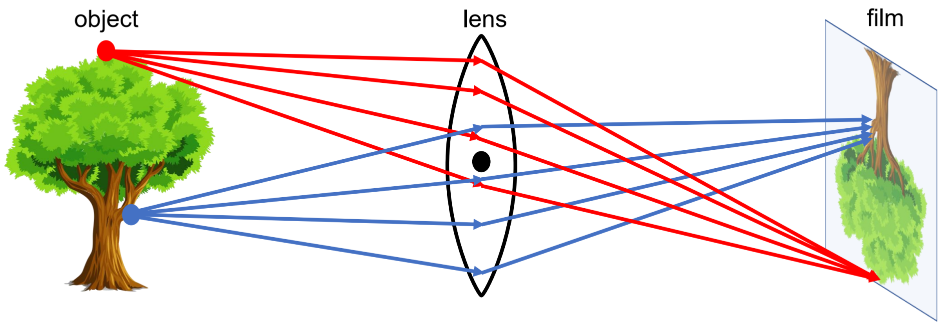
Hình 2: Ảnh với kích thước khẩu độ khác nhau. Khẩu độ lớn thì ảnh sáng nhưng mờ và ngược lại.

3 Camera và thấu kính (lenses)

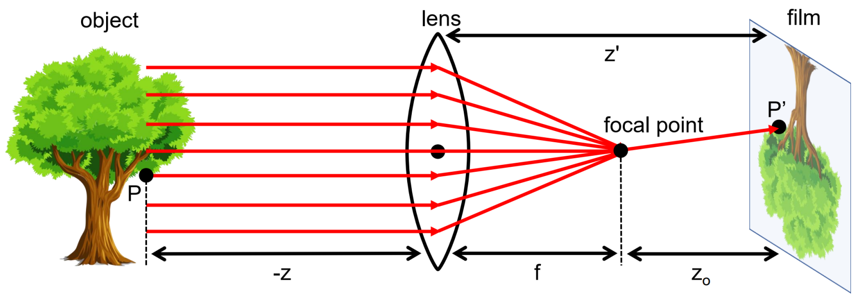
Để khắc phục điểm yếu của mô hình camera lỗ kim ở trên với kích thước khẩu độ, người ta sử dụng một *thấu kính* (**lens**) thay cho màn chắn. Thấu kính cho phép nhiều tia sáng chiếu vào màn ảnh hơn và nhờ đặc tính hội tụ, các tia sáng khác nhau sẽ hội tụ tại một điểm trên màn ảnh tránh hiện tượng mờ nhòe. Mỗi thấu kính sẽ cho ảnh rõ nét trên màn ảnh với một khoảng cách giữa vật và thấu kính cụ thể, phụ thuộc vào tiêu cự của thấu kính.

Các tia sáng đi song song với trục quang học đi qua thấu kính sẽ luôn hội tụ tại một điểm gọi là *tiêu điểm* (**focal point**) của thấu kính và khoảng cách từ tâm O đến tiêu điểm là tiêu cự f [3]. Tương tự như mô hình trước [1], với $z' = f + z_0$:

$$P' = \begin{bmatrix} z' \times \frac{x}{z} \\ z' \times \frac{y}{z} \end{bmatrix} \tag{2}$$

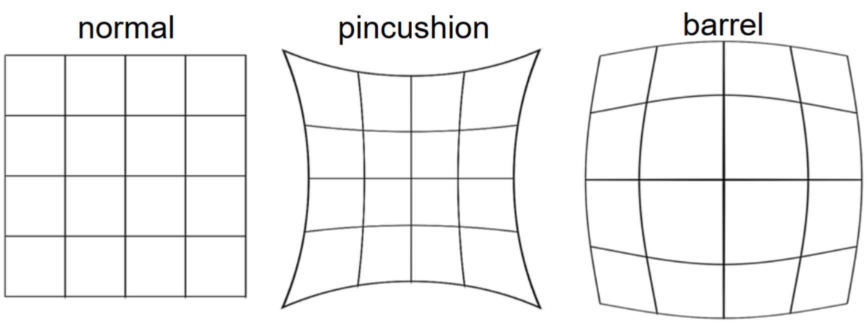


Hình 3: Máy ảnh thấu kính (giả sử khoảng cách từ vật tới thấu kính là ∞ nên $f = z'$)



Hình 4: Các tia sáng vuông góc với thấu kính hội tụ tại tiêu điểm và đi tới màn ảnh

Ảnh từ máy ảnh thấu kính có thể sai lệch do cấu tạo của thấu kính không hoàn hảo, dẫn đến các điểm trên thấu kính có tiêu cự khác nhau dẫn đến hiện tượng *biến dạng xuyên tâm* (**radial distortion**) gồm **pincushion distorsion** (tiêu cự giảm làm tia hội tụ xa trục quang học, làm ảnh trông như bị kéo giãn) và **barrel distortion** (tiêu cự tăng làm tia hội tụ gần trục quang học, làm ảnh như bị dồn vào giữa).



Hình 5: Minh họa các hiện tượng biến dạng xuyên tâm

4 Không gian ảnh số

Ta đã biết cách thu được ảnh trên màn ảnh, nhưng làm sao để biến ảnh đó thành dạng số với các pixel và kênh màu? Quá trình ánh xạ các điểm từ không gian 3D sang 2D $\mathbf{X} : \mathbb{R}^3 \to \mathbb{R}^2$ được gọi là *phép biến đổi xạ ảnh* (**projective transformation**). Ảnh thu được trên màn ảnh vẫn còn một vài điểm chưa thỏa mãn để có ảnh số. Thứ nhất, hệ quy chiếu trong ảnh số có thể khác. Thứ hai, thay vì là các pixel rời rạc thì ảnh trên màn là liên tục. Thứ ba, cần phải xử lý quang sai và biến dạng ảnh do thấu kính gây ra.

4.1 Mô hình ma trận máy ảnh và hệ tọa độ đồng nhất

Mô hình ma trận máy ảnh (**Camera Matrix Model**) là tập hợp các tham số cần cho việc ánh xạ từ không gian 3D sang không gian 2D. Hai tham số đầu tiên là c_x và c_y chỉ offset của hệ quy chiếu của màn ảnh so với ảnh số (hệ quy chiếu tâm của màn ảnh và hệ quy chiếu góc dưới-trái của ảnh). Hai tham số k và l (đơn vị pixel/m) sẽ scale lại trục tọa độ theo pixel, ta tiếp tục suy ra từ công thức 1:

$$P' = \begin{bmatrix} k \times f \times \dfrac{x}{z} + c_x \\ \\ l \times f \times \dfrac{y}{z} + c_y \end{bmatrix} = \begin{bmatrix} \alpha \times \dfrac{x}{z} + c_x \\ \\ \beta \times \dfrac{y}{z} + c_y \end{bmatrix} \tag{3}$$

Dễ thấy đây không phải là một phép biến đổi tuyến tính vì có phép chia cho z. Do đó, ta không thể biến đổi phép chiếu này trực tiếp thành nhân ma trận được. Một phương pháp giải quyết là chuyển hệ trục tọa độ sang *Hệ trục đồng nhất* (**Homogeneous Coordinate**). Một điểm có tọa độ $P(x, y, z)$ chuyển sang hệ tọa độ đồng nhất sẽ có tọa độ là $P(x, y, z, 1)$ (từ đây các phép biến đổi sẽ được hiểu là thực hiện trong hệ tọa đồ đồng nhất). Ta có:

$$\mathbf{M}P_h = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}_h = \begin{bmatrix} \alpha x + c_x z \\ \beta y + c_y z \\ z \end{bmatrix}_h \tag{4}$$

Sau phép biến đổi trong [4], ta có thể chuyển từ hệ tọa độ đồng nhất về hệ tọa độ ban đầu bằng cách chia $d-1$ tọa độ đầu tiên cho tọa độ cuối và chỉ giữ lại $d-1$ tọa độ đầu. Ta thấy đấy chính là P'. Vậy là ta đã có thể thực hiện một phép nhân ma trận cho phép chiếu ảnh. **K** được gọi là *ma trận nội tại* (**instrinsic matrix**) - chứa 5 tham số phụ thuộc vào camera [6].

$$P' = \begin{bmatrix} \alpha x + c_x z \\ \beta y + c_y z \\ z \end{bmatrix}_h \to \begin{bmatrix} \alpha \times \dfrac{x}{z} + c_x \\ \\ \beta \times \dfrac{y}{z} + c_y \end{bmatrix} \tag{5}$$

$$P' = \mathbf{M}P = \mathbf{K} \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} P \tag{6}$$

Nếu hệ trục tọa độ của màn ảnh có hai trục tọa độ tạo một góc θ (gây ra do sai lệch của camera) thay vì vuông góc, ta sẽ có dạng tổng quát của ma trận nội tại như sau:

$$\mathbf{K} = \begin{bmatrix} \alpha & -\alpha \cot \theta & c_x \\ 0 & \frac{\beta}{\sin \theta} & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

Ta đã xét đến phép chiếu từ hệ quy chiếu của camera sang màn ảnh, nhưng làm sao thực hiện ở hệ quy chiếu riêng của vật thể trước khi thực hiện phép chiếu? Ta sẽ thực hiện phép biến đổi và chuyển về hệ quy chiếu camera với **R** là ma trận xoay và **T** là ma trận tịnh tiến:

$$P = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} P_w \tag{8}$$

$$P' = \mathbf{K} \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} P_w = \mathbf{K} \begin{bmatrix} \mathbf{R} & \mathbf{T} \end{bmatrix} P_w \tag{9}$$

Ma trận $\begin{bmatrix} \mathbf{R} & \mathbf{T} \end{bmatrix}$ là *ma trận ngoại tại* (**extrinsic matrix**) đại diện cho những thông số không thuộc camera. Vậy trong công thức [9] có tổng 11 tham số tự do.

5 Hiệu chuẩn camera (Camera Cabliration)

Với một hô hình camera cho trước, làm sao để biết tham số nội tại và ngoại tại của nó? Ta có thể xấp xỉ nó thông qua các tọa độ thực và tọa độ hình chiếu trên màn ảnh. Thật vậy, giả sử ta có :

$$p' = \begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{M}P = \frac{1}{Z}\begin{bmatrix} \mathbf{m_1}^T \\ \mathbf{m_2}^T \\ \mathbf{m_3}^T \end{bmatrix} P = \begin{bmatrix} \frac{\mathbf{m_1}^T P_1}{\mathbf{m_3}^T P_3} \\ \frac{\mathbf{m_2}^T P_1}{\mathbf{m_3}^T P_3} \end{bmatrix} \tag{10}$$

Ta cần tìm $\mathbf{M}$. Dễ thấy với mỗi điểm, ta sẽ có hai phương trình. Vậy cần ít nhất 6 điểm để tính được 11 tham số tự do của $\mathbf{M}$ (thực ra cần nhiều hơn để giảm nhiễu). Ta sẽ mô hình bài toán thành dạng nhân ma trận như sau:

$$\begin{bmatrix} P_1^T & 0^T & -u_1 P_1^T \\ 0^T & P_1^T & -v_1 P_1^T \\ \vdots & \vdots & \vdots \\ P_n^T & 0^T & -u_n P_n^T \\ 0^T & P_n^T & -v_n P_n^T \end{bmatrix} \begin{bmatrix} \mathbf{m}_1^T \\ \mathbf{m}_2^T \\ \mathbf{m}_3 \end{bmatrix} = \mathbf{Pm} = 0 \tag{11}$$

Sẽ có nhiều nghiệm không tầm thường cho phương trình trên, vì thế ta sẽ chuyển bài toán về dạng tối ưu và giải bằng SVD:

$$\begin{aligned} \text{minimize}_{\mathbf{m}} \quad & \|\mathbf{Pm}\|^2 \\ \text{subject to} \quad & \|\mathbf{m}\|^2 = 1 \end{aligned} \tag{12}$$

6 Sự sai lệch của camera

Tuy đã tính toán được các tham số của ma trận camera nhưng ta vẫn chưa tính cho trường hợp sai lệch do thấu kính (radial distortion) dẫn đến thay đổi về khoảng cách từ điểm đến tâm màn ảnh. Ta có thể mô hình hiện tượng này bằng phương trình:

$$p' = \begin{bmatrix} \frac{1}{\lambda} & 0 & 0 \\ 0 & \frac{1}{\lambda} & 0 \\ 0 & 0 & 1 \end{bmatrix} MP \tag{13}$$

với λ là hàm phi tuyến. Do đó ta không thể tính ma trận tham số tự do như [11]. Ta có thể giảm độ phức tạp tính toán bằng cách tính $\mathbf{m}_1$ và $\mathbf{m}_2$ bằng trước, sau đó giải một phương trình phi tuyến.

Tài liệu