



# Структура из движения

---



Many slides adopted from Svetlana Lazebnik, Steve Seitz, Richard Hartley and Noah Snavely



# Что умеем

---

- Калибровка камеры
  - Оценка  $P$  ( $K, R, T$ ) по набору соответствий 2D и 3D точек  $\{(x_i, X_i)\}$ ,  $i=1, k$
- Триангуляция
  - Оценка 3D точки  $X$  по проекциям  $x_i$  на калиброванные камеры  $P_i$
- Вычисление существенной матрицы  $E$ 
  - По набору соответствий 2D точек  $\{(x_i, x'_i)\}$ ,  $i=1, k$  на 2х изображениях и внутренней калибровке  $K, K'$
  - Извлечение  $R, T$  (внешней калибровки) из  $E$
- Вычисление фундаментальной матрицы  $F$ 
  - По набору соответствий 2D точек  $\{(x_i, x'_i)\}$ ,  $i=1, k$  на 2х изображениях



# Структура из движения



Драконъ, видимый подъ различными углами зрѣнія  
По гравюру изъ книги наз. „Oculus artificialis teleiopicus“ Нема. 1702 года.

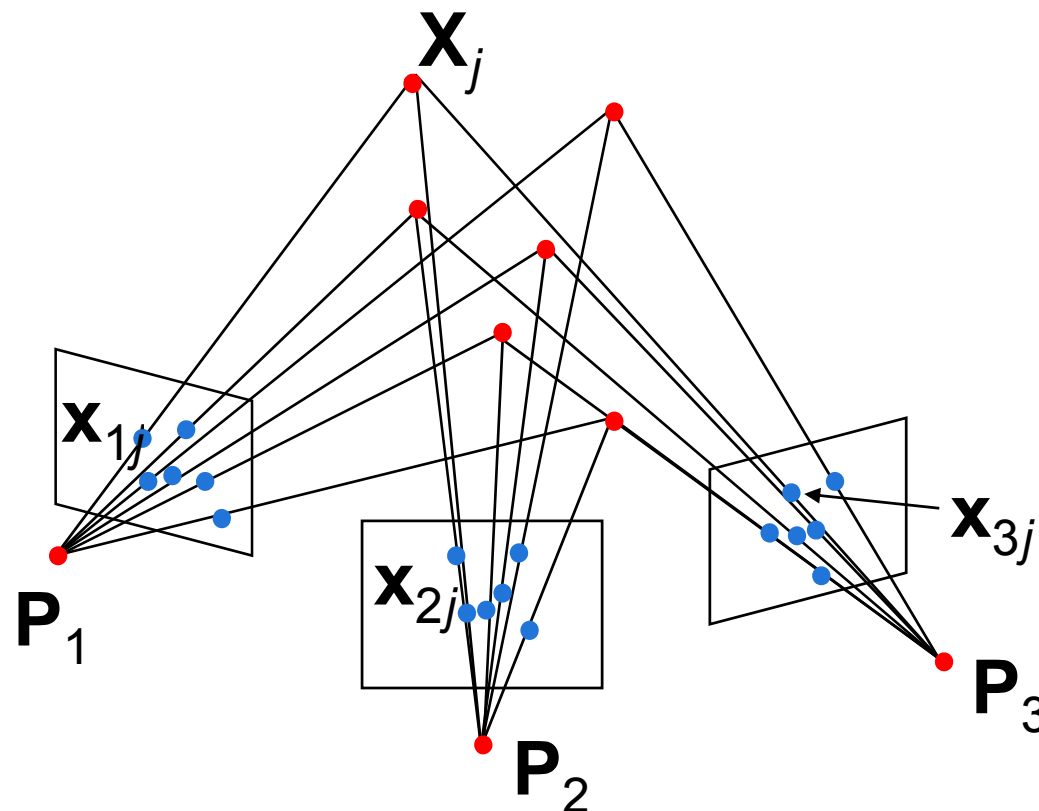


# Структура из движения

• Дано:  $m$  изображений  $n$  фиксированных 3D точек

$$\bullet \mathbf{x}_{ij} = \mathbf{P}_i \mathbf{X}_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

• Задача: оценить  $m$  матриц проекции  $\mathbf{P}_i$  и  $n$  3D точек  $\mathbf{X}_j$  из  $mn$  соответствий  $\mathbf{x}_{ij}$





# Неоднозначность решения

---

- Если мы умножим всю сцену на некоторый коэффициент  $k$  и в то же время умножим матрицы камер на  $1/k$ , проекции точек сцены на изображения не изменятся:

$$\mathbf{x} = \mathbf{P}\mathbf{X} = \left( \frac{1}{k} \mathbf{P} \right) (k \mathbf{X})$$

Вывод: оценить абсолютный масштаб (размеры) сцены по изображениям невозможно!



## Неоднозначность решения

---

- Обобщение: если мы преобразуем сцену с помощью преобразования  $\mathbf{Q}$  и применим обратное преобразование к матрицам камер, изображения не изменятся:

$$\mathbf{x} = \mathbf{P}\mathbf{X} = (\mathbf{P}\mathbf{Q}^{-1})(\mathbf{Q}\mathbf{X})$$



# Иерархия 3D преобразования

---

Проективное  
15dof

$$\begin{bmatrix} A & t \\ v^T & v \end{bmatrix}$$



Сохраняет пересечения  
и касание

Аффинное  
12dof

$$\begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix}$$



Сохраняет  
параллельность,  
отношения объемов

Подобие  
7dof

$$\begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix}$$



Сохраняет углы,  
отношения длин

Евклидово  
6dof

$$\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$$

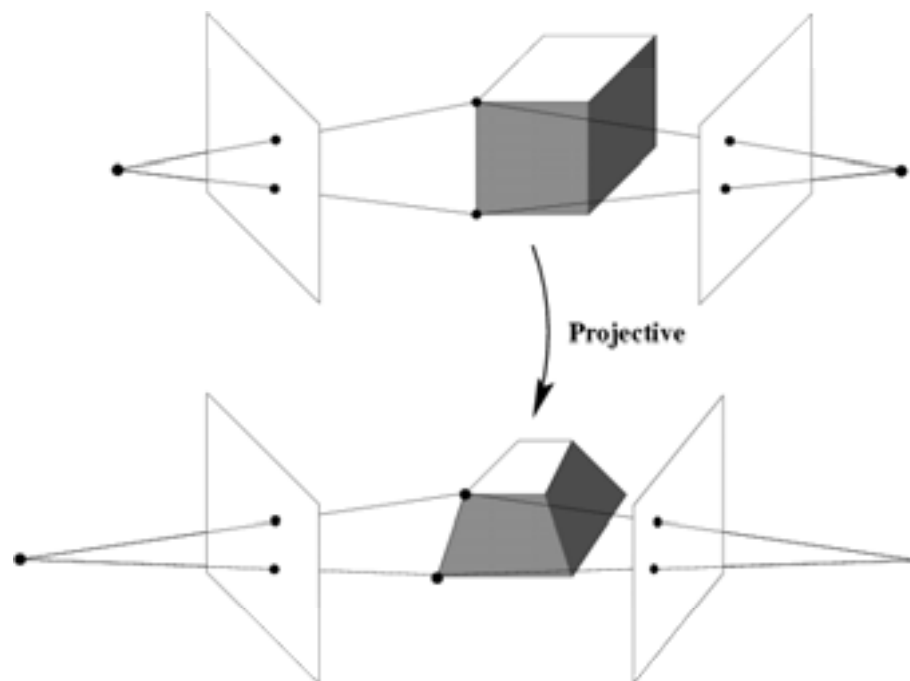


Сохраняет углы, длины



# Проективная неоднозначность

---



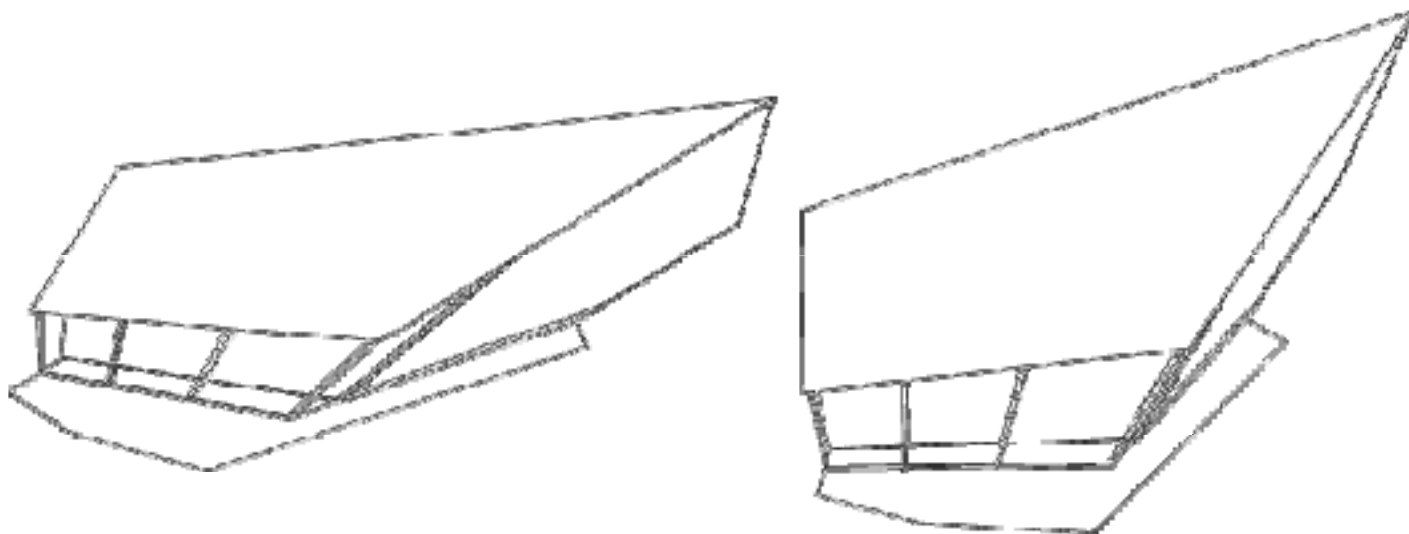
$$\mathbf{x} = \mathbf{P}\mathbf{X} = (\mathbf{P}\mathbf{Q}_P^{-1})(\mathbf{Q}_P \mathbf{X})$$





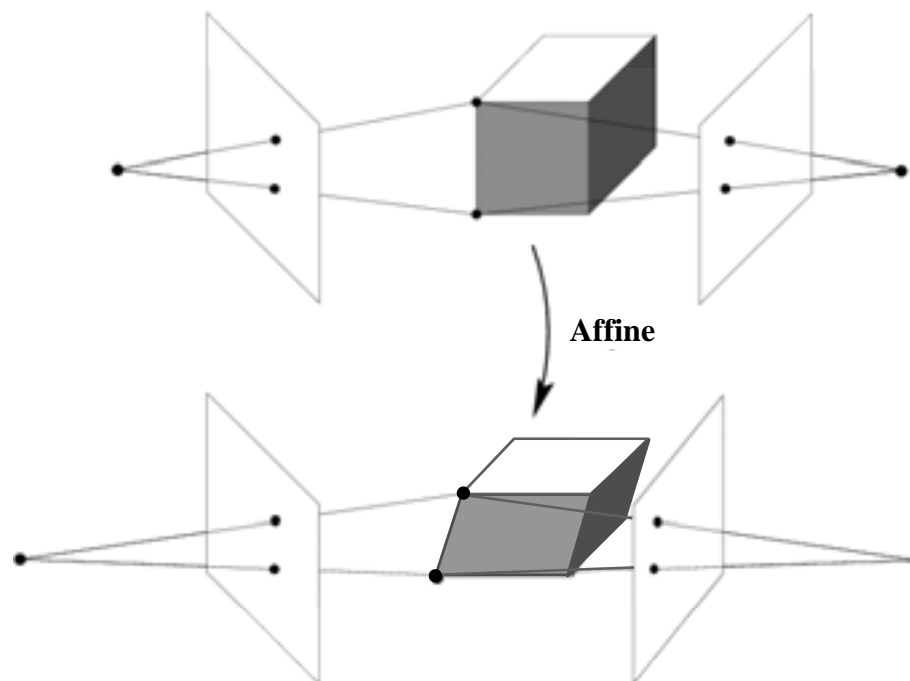
# Проективная неоднозначность

---





# Аффинная неоднозначность

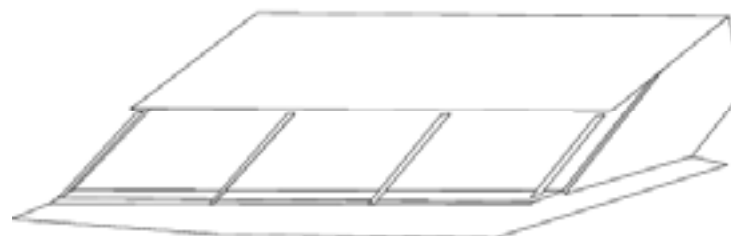
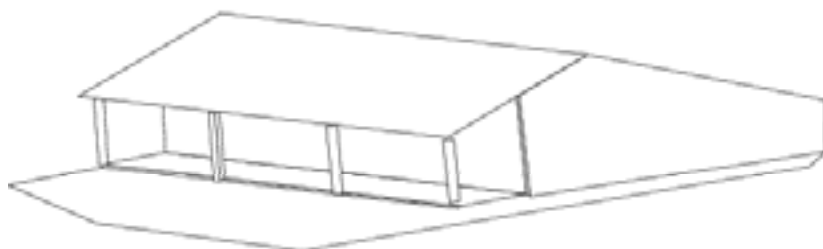


$$\mathbf{x} = \mathbf{P}\mathbf{X} = (\mathbf{P}\mathbf{Q}_A^{-1})(\mathbf{Q}_A \mathbf{X})$$



# Аффинная неоднозначность

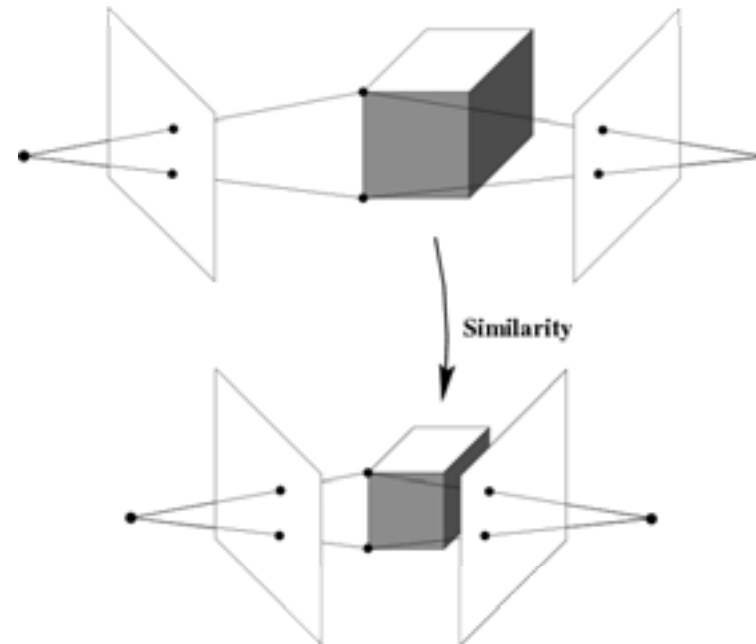
---





# Неоднозначность подобия

---

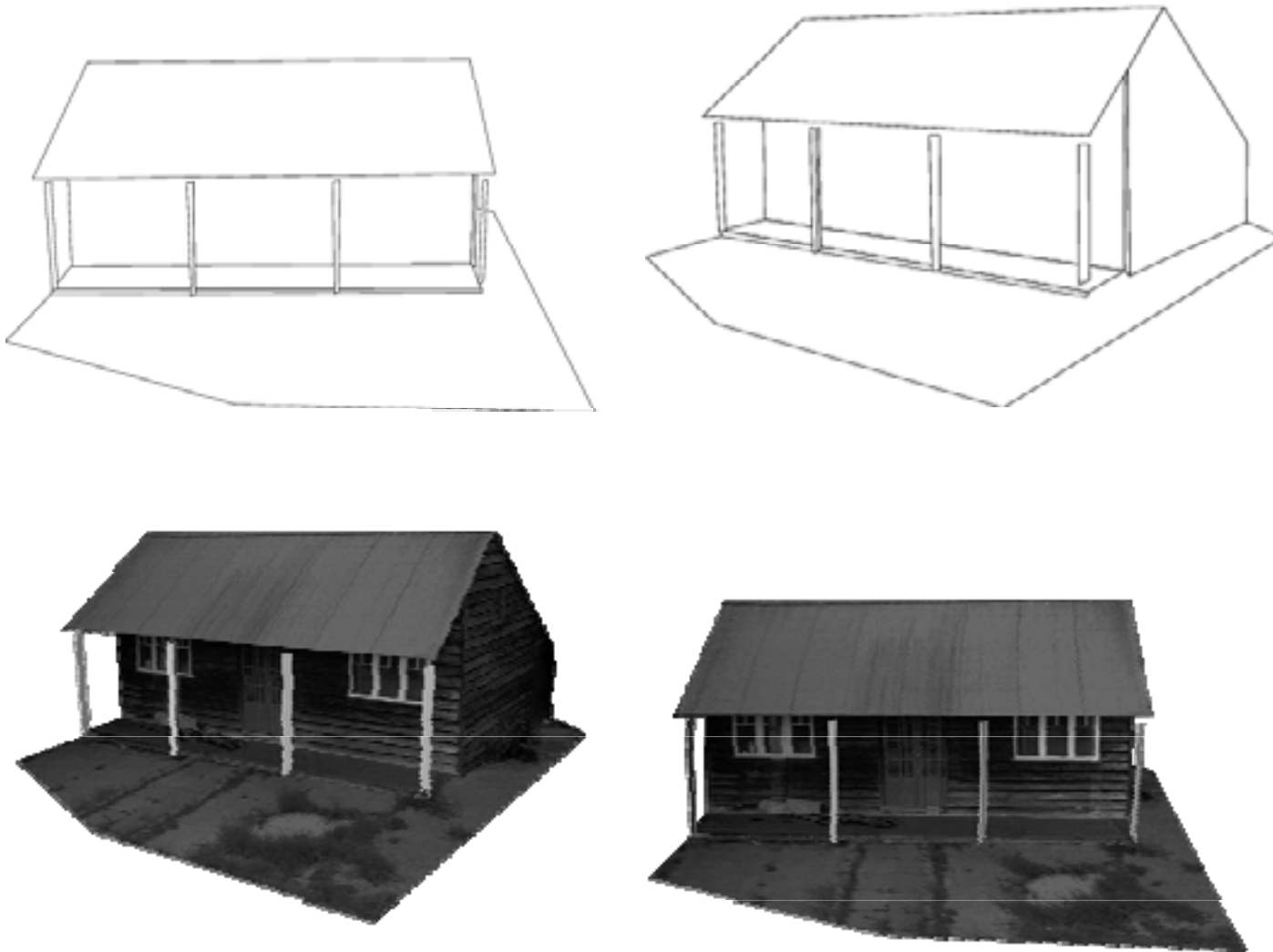


$$\mathbf{x} = \mathbf{P}\mathbf{X} = (\mathbf{P}\mathbf{Q}_s^{-1})(\mathbf{Q}_s\mathbf{X})$$



# Неоднозначность подобия

---





# Иерархия реконструкций

---

- Если нет ограничений на калибровку камеры или сцену, мы получим проективную реконструкцию

Проективное  
15dof

$$\begin{bmatrix} A & t \\ v^T & v \end{bmatrix}$$

- С добавлением дополнительных ограничений мы можем усовершенствовать (upgrade) реконструкцию до аффинной, подобия или евклидовой

Аффинное  
12dof

$$\begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix}$$

Подобие  
7dof

$$\begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix}$$

- Если известна матрица внутренней калибровки, тогда получаем реконструкцию с точностью до подобия

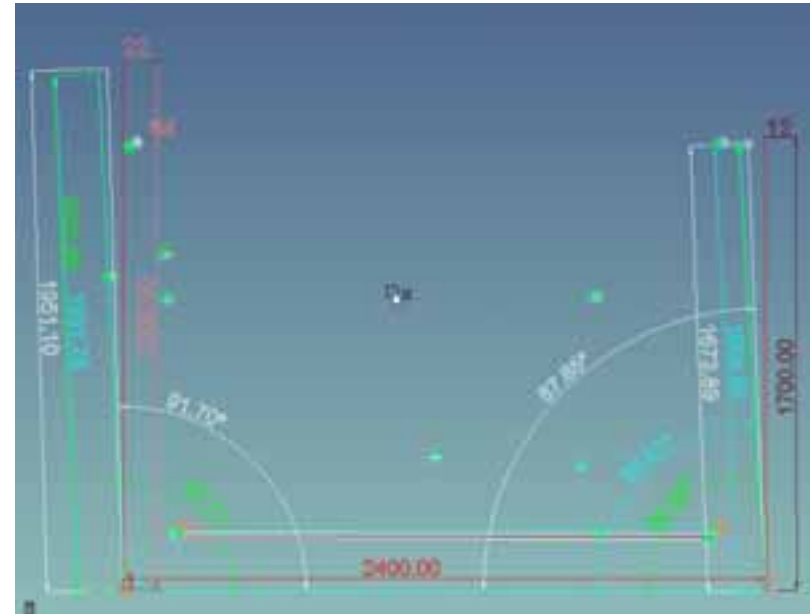
Евклидово  
6dof

$$\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$$

- Если есть и расстояние – тогда с точностью до евклидова преобразования



# Фотограмметрия

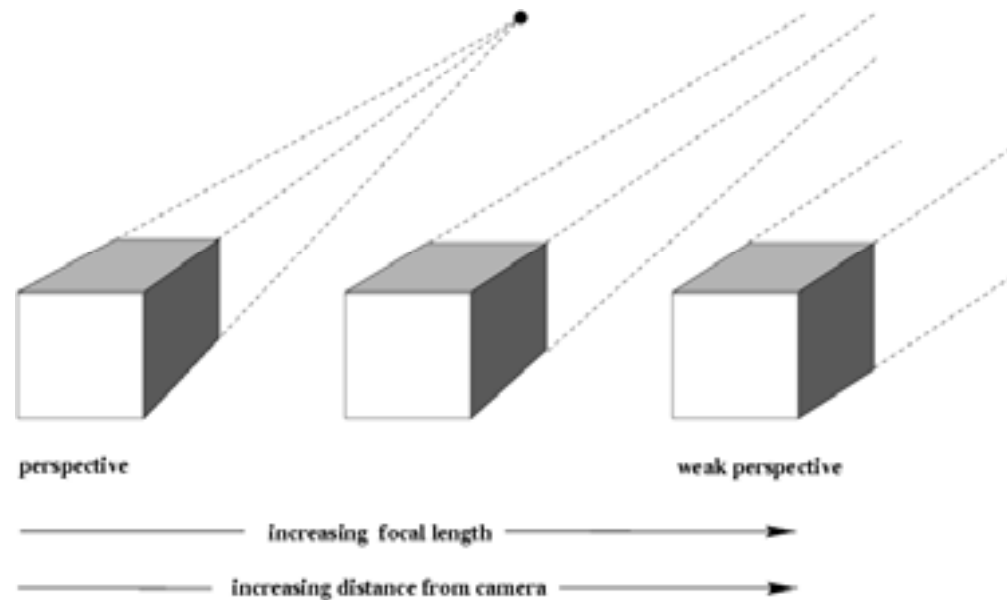


Чтобы провести измерения по изображениям (задача фотограмметрии), необходимо знать расстояние между минимум 2мя точками и внутреннюю калибровку камеры



# Структура из движения

- Рассмотрим случай аффинных камер



Центр в  
бесконечности

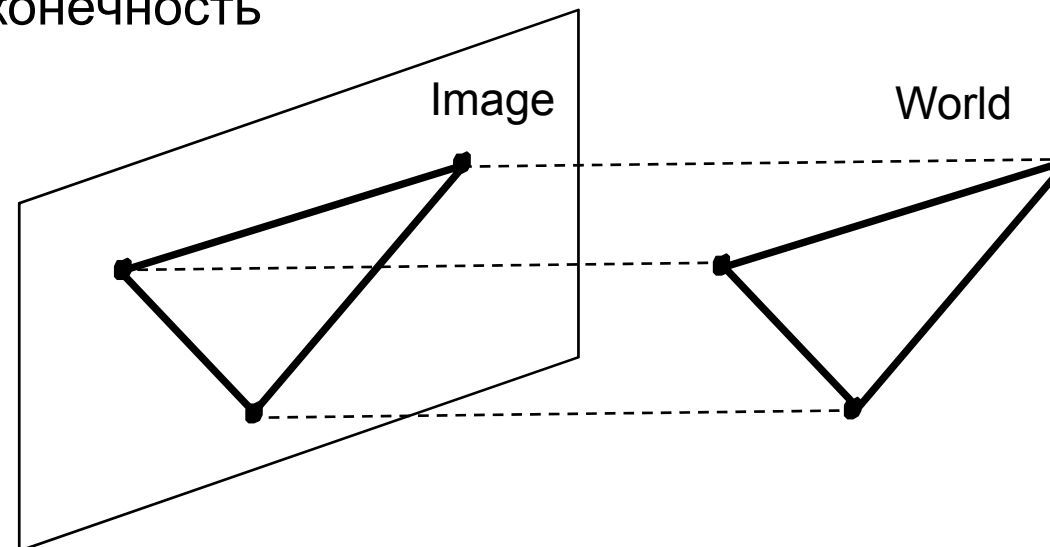






# Ортографическая проекция

- Вырожденный случай перспективной
  - Расстояние от центра проекции до картинной плоскости - бесконечность



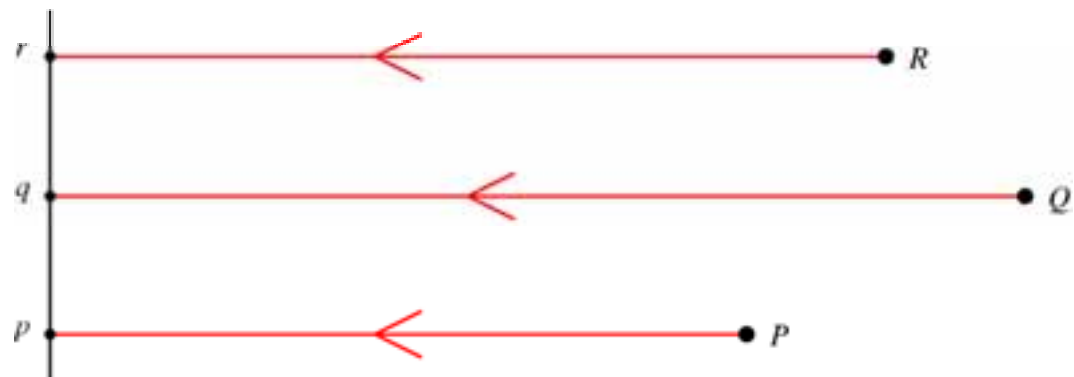
- Матрица ортографической проекции:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \Rightarrow (x, y)$$

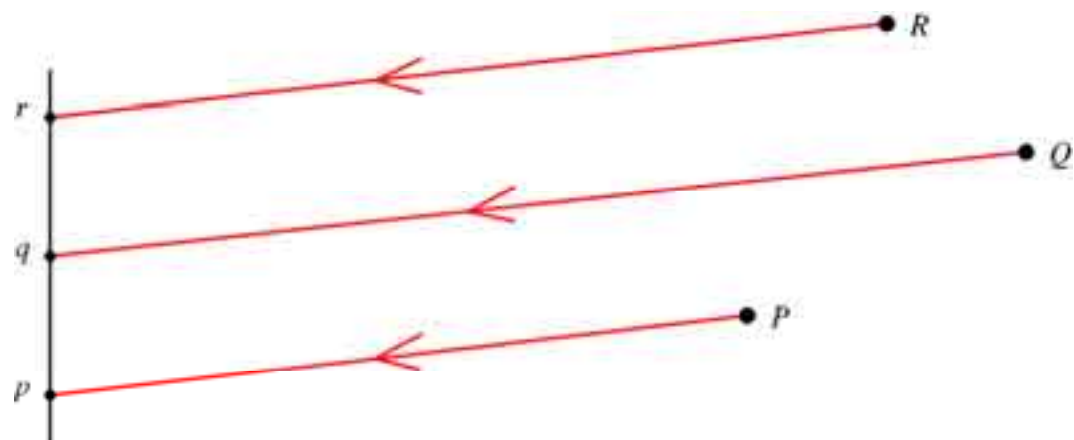


# Аффинные камеры

Ортографическая  
проекция



Параллельная  
проекция



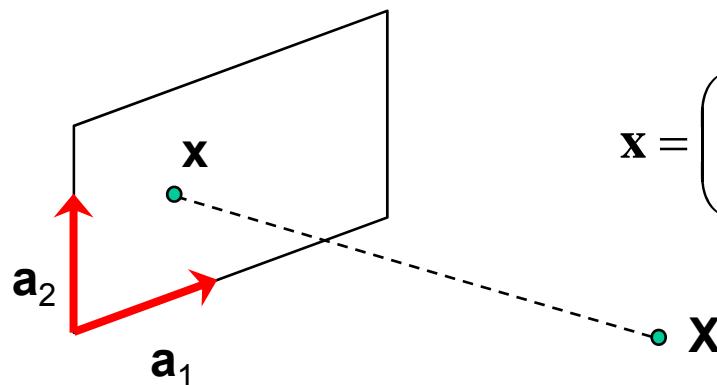


# Аффинная камера

- Общая аффинная камера объединяет аффинное преобразование в 3D пространстве, ортографическую проекцию, и аффинное преобразование изображения

$$\mathbf{P} = [3 \times 3 \text{ affine}] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} [4 \times 4 \text{ affine}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$$

- Аффинная проекция это линейное отображение плюс сдвиг в неоднородных координатах



$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

Проекция центра  
Мировых координат



# Аффинная реконструкция

---

- Дано:  $m$  изображений  $n$  фиксированных 3D точек:

$$\mathbf{x}_{ij} = \mathbf{A}_i \mathbf{X}_j + \mathbf{b}_i, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

- Задача: по  $mn$  соответствиям  $\mathbf{x}_{ij}$  оценить  $m$  матриц проекции  $\mathbf{A}_i$  и векторов сдвига  $\mathbf{b}_i$  и  $n$  точек  $\mathbf{X}_j$

- Реконструкция определена с точностью до произвольного аффинного преобразования  $\mathbf{Q}$  (12 степеней свободы):

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \rightarrow \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \mathbf{Q}^{-1}, \quad \begin{pmatrix} \mathbf{X} \\ \mathbf{1} \end{pmatrix} \rightarrow \mathbf{Q} \begin{pmatrix} \mathbf{X} \\ \mathbf{1} \end{pmatrix}$$

- Даны  $2mn$  известных и  $8m + 3n$  неизвестных (минус 12 dof из-за аффинной неоднозначности)
- Значит,  $2mn \geq 8m + 3n - 12$  (должно быть)
- Для 2х видов нужно 4 соответствия



# Аффинная реконструкция

---

- Центровка: вычитаем центроид точек изображения

$$\begin{aligned}\hat{\mathbf{x}}_{ij} &= \mathbf{x}_{ij} - \frac{1}{n} \sum_{k=1}^n \mathbf{x}_{ik} = \mathbf{A}_i \mathbf{X}_j + \mathbf{b}_i - \frac{1}{n} \sum_{k=1}^n (\mathbf{A}_i \mathbf{X}_k + \mathbf{b}_i) \\ &= \mathbf{A}_i \left( \mathbf{X}_j - \frac{1}{n} \sum_{k=1}^n \mathbf{X}_k \right) = \mathbf{A}_i \hat{\mathbf{X}}_j\end{aligned}$$

- Для простоты, пусть центр мировых координат – центроид 3D точек
- После центровки, каждая нормализованная точка  $\mathbf{x}_{ij}$  относится к  $\mathbf{X}_j$  следующим образом:

$$\hat{\mathbf{x}}_{ij} = \mathbf{A}_i \mathbf{X}_j$$



# Аффинная реконструкция

- Запишем данные  $2m \times n$  в виде матрицы измерений

$$\mathbf{D} = \begin{bmatrix} \hat{\mathbf{x}}_{11} & \hat{\mathbf{x}}_{12} & \cdots & \hat{\mathbf{x}}_{1n} \\ \hat{\mathbf{x}}_{21} & \hat{\mathbf{x}}_{22} & \cdots & \hat{\mathbf{x}}_{2n} \\ & & \ddots & \\ \hat{\mathbf{x}}_{m1} & \hat{\mathbf{x}}_{m2} & \cdots & \hat{\mathbf{x}}_{mn} \end{bmatrix}$$

↓ камеры ( $2m$ )

→ точки ( $n$ )



# Аффинная реконструкция

- Запишем данные  $2m \times n$  в виде матрицы измерений

$$\mathbf{D} = \begin{bmatrix} \hat{\mathbf{x}}_{11} & \hat{\mathbf{x}}_{12} & \cdots & \hat{\mathbf{x}}_{1n} \\ \hat{\mathbf{x}}_{21} & \hat{\mathbf{x}}_{22} & \cdots & \hat{\mathbf{x}}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{x}}_{m1} & \hat{\mathbf{x}}_{m2} & \cdots & \hat{\mathbf{x}}_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_m \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_n \end{bmatrix}$$

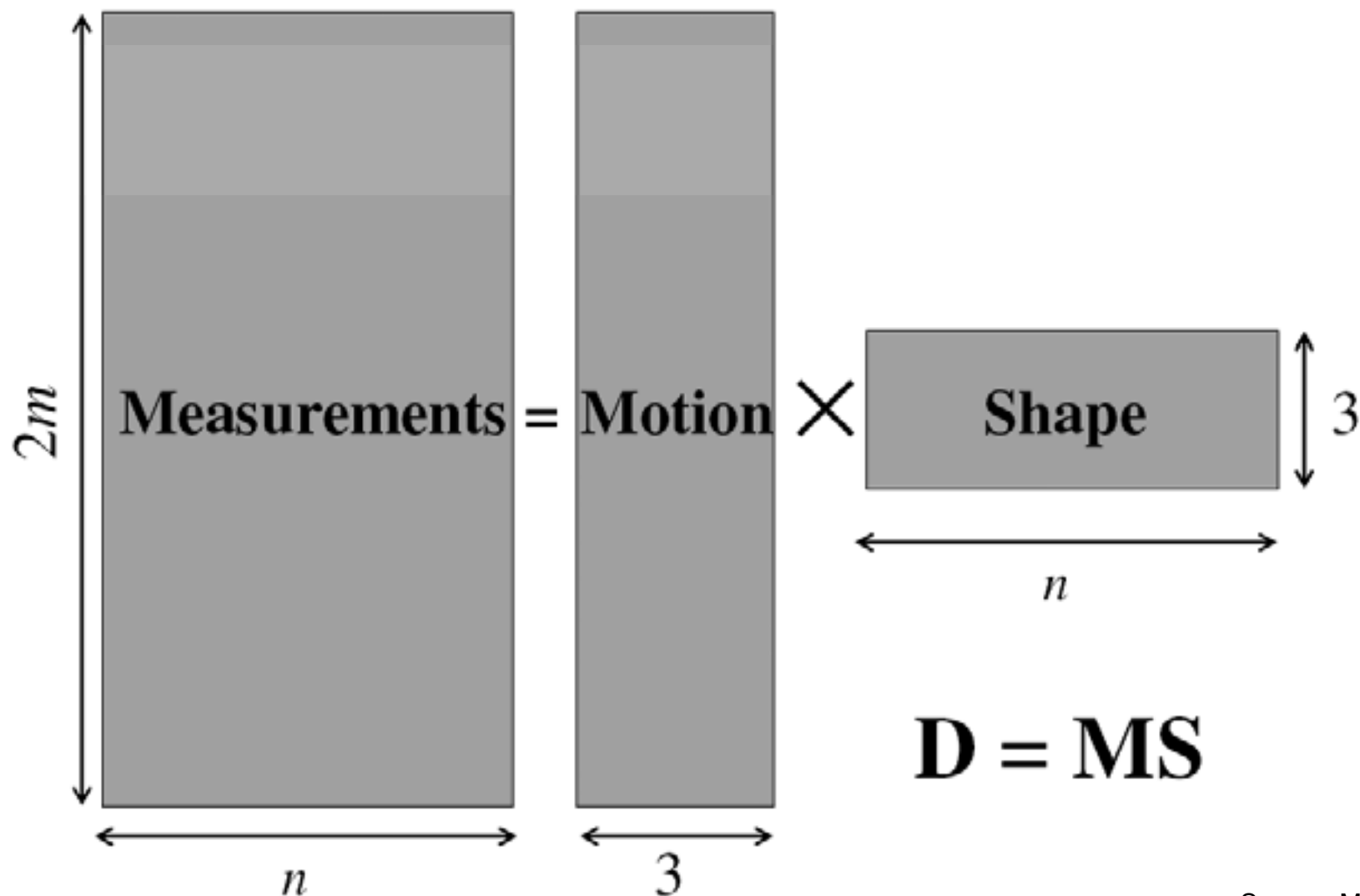
points ( $3 \times n$ )

cameras  
( $2m \times 3$ )

У матрицы измерений  $\mathbf{D} = \mathbf{MS}$  должен быть ранг 3!



# Факторизация матрицы измерений

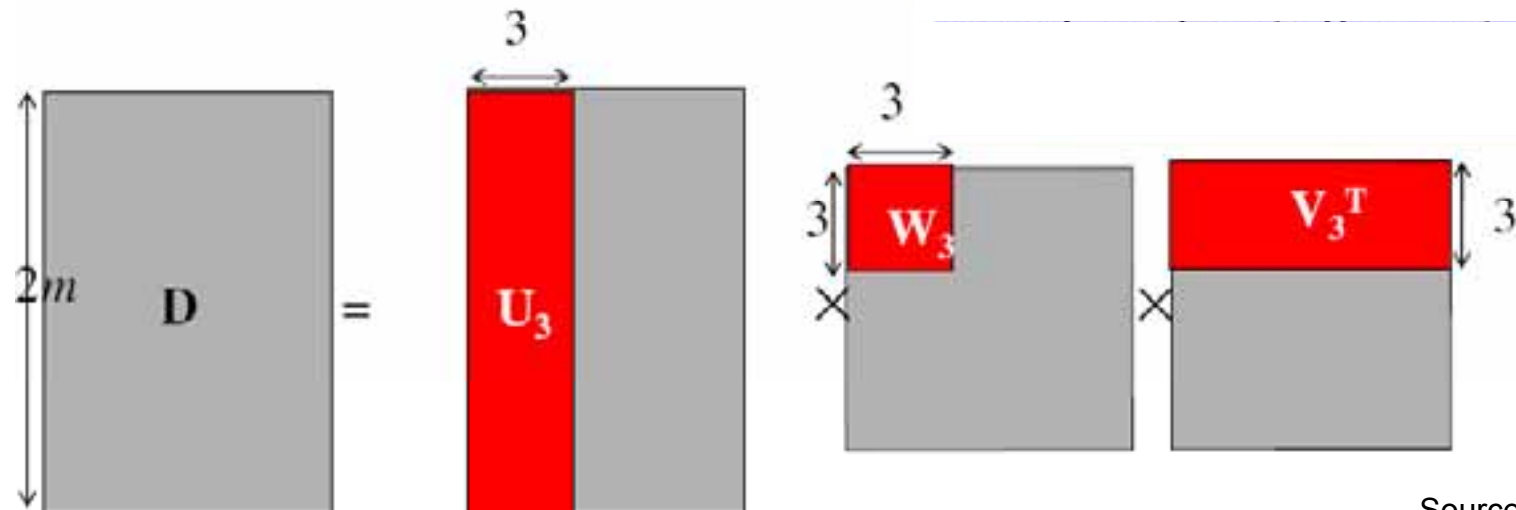
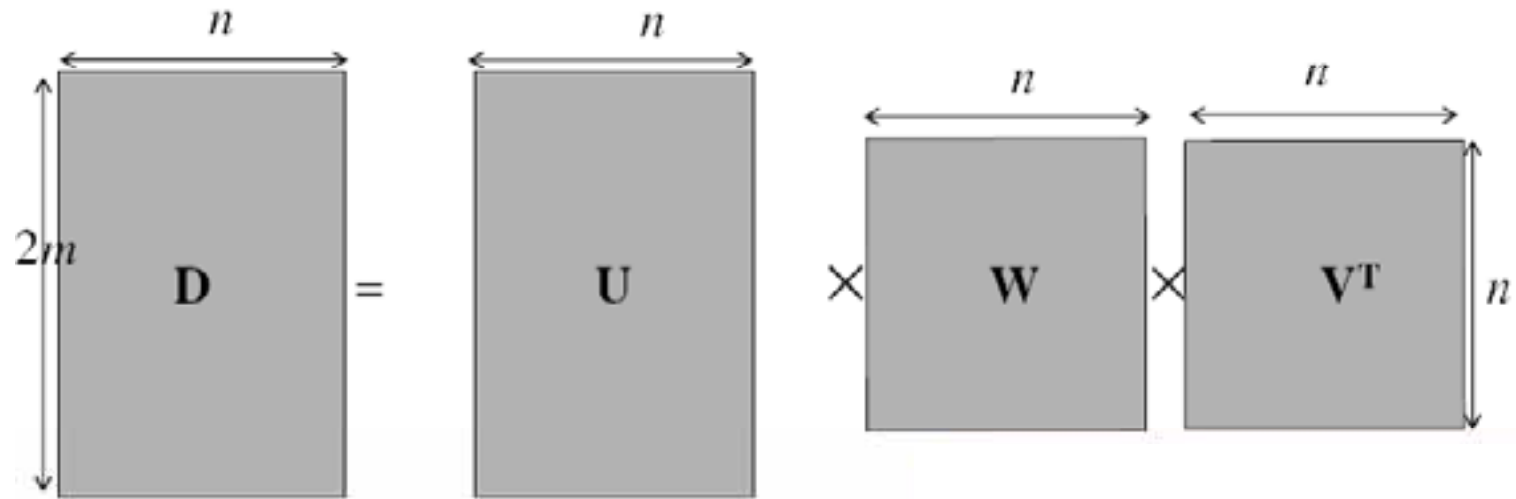






# Факторизация матрицы измерений

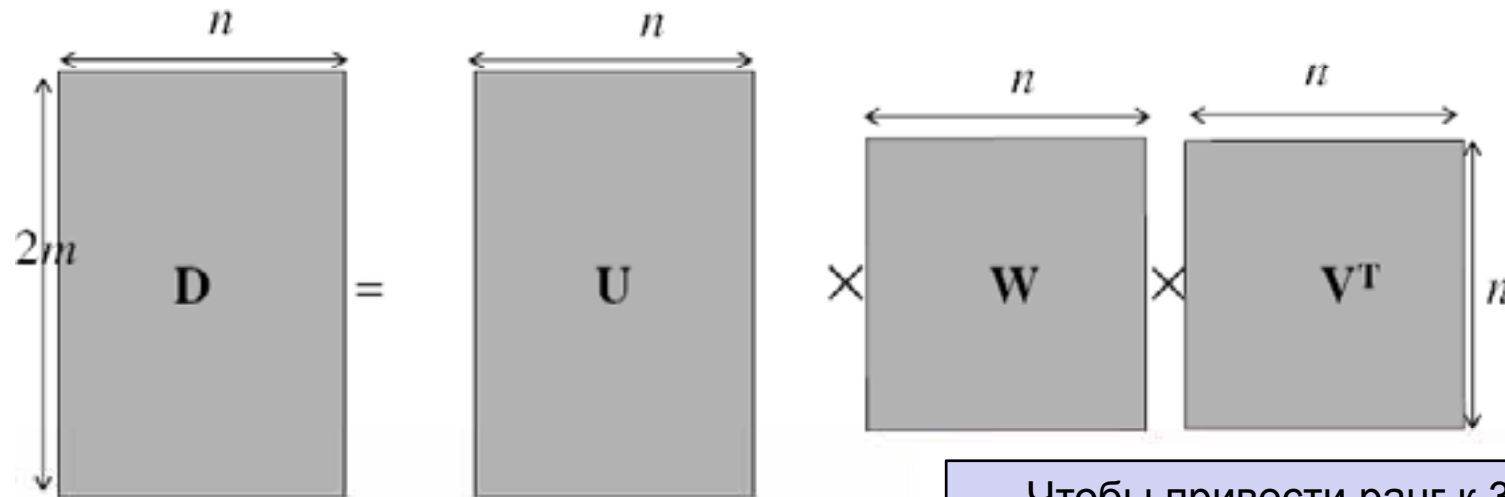
- SVD разложение матрицы D:



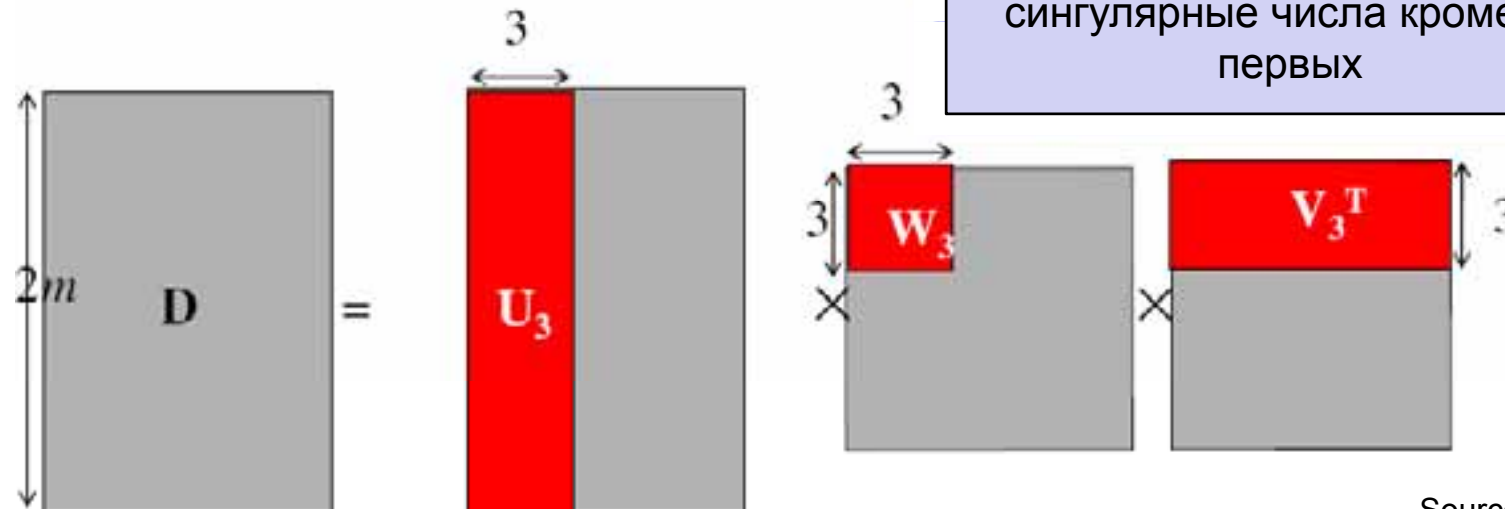


# Факторизация матрицы измерений

- SVD разложение матрицы  $D$ :



Чтобы привести ранг к  $3m$ ,  
просто обнулим все  
сингулярные числа кроме  $3 \times$   
первых





# Факторизация матрицы измерений

- Вычисляем разложение через SVD:

The diagram illustrates the SVD decomposition of a matrix  $D$ . Matrix  $D$  is a gray rectangle with a vertical dimension of  $2m$  and a horizontal dimension of  $n$ . It is equal to the product of three matrices:  $U_3$ ,  $W_3$ , and  $V_3^T$ . Matrix  $U_3$  is a red vertical rectangle with a vertical dimension of  $2m$  and a horizontal dimension of  $3$ . Matrix  $W_3$  is a red square with a horizontal dimension of  $3$  and a vertical dimension of  $3$ . Matrix  $V_3^T$  is a red horizontal rectangle with a horizontal dimension of  $n$  and a vertical dimension of  $3$ . The dimensions are indicated by arrows and labels:  $2m$  for the vertical axis of  $D$  and  $U_3$ ,  $3$  for the horizontal axis of  $U_3$  and the vertical axis of  $W_3$  and  $V_3^T$ , and  $n$  for the horizontal axis of  $D$  and  $V_3^T$ .

$$\begin{matrix} 2m \\ \downarrow \\ \text{D} \\ \uparrow \\ 2m \end{matrix} = \begin{matrix} \downarrow \\ U_3 \\ \uparrow \\ 3 \end{matrix} \times \begin{matrix} 3 \\ \leftarrow \\ W_3 \\ \rightarrow \\ 3 \end{matrix} \times \begin{matrix} \leftarrow \\ V_3^T \\ \rightarrow \\ n \end{matrix}$$



# Факторизация матрицы измерений

- Вычисляем разложение через SVD :

$$\begin{matrix} 2m \\ \downarrow \\ \text{D} \\ \uparrow \\ 2m \end{matrix} = \begin{matrix} \downarrow \\ \text{U}_3 \\ \uparrow \\ 3 \end{matrix} \times \begin{matrix} 3 \\ \leftarrow \\ \text{W}_3 \\ \rightarrow \\ 3 \end{matrix} \times \begin{matrix} \leftarrow \\ \text{V}_3^T \\ \rightarrow \\ n \end{matrix} \begin{matrix} \downarrow \\ 3 \\ \uparrow \end{matrix}$$

Возможное разложение:  
 $M = U_3 W_3^{1/2}$     $S = W_3^{1/2} V_3^T$

$$\text{D} = \text{M} \times \text{S}$$

Это минимизирует  
 $|D - MS|^2$



# Аффинная неоднозначность

---

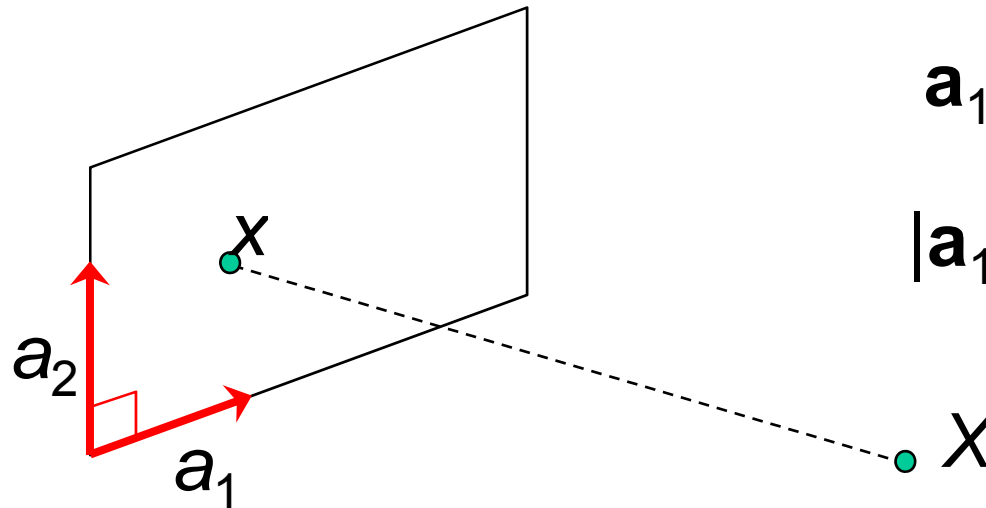
$$\mathbf{D} = \mathbf{M} \times \mathbf{S}$$

- Разложение неоднозначно. Мы получим ту же  $\mathbf{D}$  используя любую  $3 \times 3$  матрицу  $\mathbf{C}$  и применив преобразование  $\mathbf{M} \rightarrow \mathbf{MC}$ ,  $\mathbf{S} \rightarrow \mathbf{C}^{-1}\mathbf{S}$
- Это происходит потому, что у нас есть только аффинные преобразования и мы не требуем соблюдение Евклидовых ограничений (например, перпендикулярности осей изображения)



# Удаление неоднозначности

- Ортография: оси изображения перпендикулярны и масштабированы к 1



$$\mathbf{a}_1 \cdot \mathbf{a}_2 = 0$$

$$|\mathbf{a}_1|^2 = |\mathbf{a}_2|^2 = 1$$

- Это даёт  $3m$  уравнений на  $\mathbf{L} = \mathbf{C}\mathbf{C}^T$ :

$$\mathbf{A}_i \mathbf{L} \mathbf{A}_i^T = \mathbf{Id}, \quad i = 1, \dots, m$$

- Находим  $\mathbf{L}$
- Вычисляем  $\mathbf{C}$  с помощью разложения Cholesky  $\mathbf{L} = \mathbf{C}\mathbf{C}^T$
- Обновляем  $\mathbf{M}$  и  $\mathbf{S}$ :  $\mathbf{M} = \mathbf{M}\mathbf{C}$ ,  $\mathbf{S} = \mathbf{C}^{-1}\mathbf{S}$



# Алгоритм

---

- Дано:  $m$  изображений и  $n$  точек  $\mathbf{x}_{ij}$
- Для каждого изображения  $i$ , центрируем координаты точек
- Строим матрицу измерений  $2m \times n$   $\mathbf{D}$ :
  - Столбец  $j$  содержит проекции точки  $j$  на все виды
  - Строка  $i$  содержит координаты проекций всех  $n$  точек на изображение  $i$
- Разложение  $\mathbf{D}$ :
  - Вычисляем SVD:  $\mathbf{D} = \mathbf{U} \mathbf{W} \mathbf{V}^T$
  - Строим  $\mathbf{U}_3$  из 3 первых столбцов  $\mathbf{U}$
  - Строим  $\mathbf{V}_3$  из 3 первых столбцов  $\mathbf{V}$
  - Строим  $\mathbf{W}_3$  из верхнего левого  $3 \times 3$  блока  $\mathbf{W}$
- Строим матрицы движения и структуры:
  - $\mathbf{M} = \mathbf{U}_3 \mathbf{W}_3^{1/2}$  и  $\mathbf{S} = \mathbf{W}_3^{1/2} \mathbf{V}_3^T$  (или  $\mathbf{M} = \mathbf{U}_3$  и  $\mathbf{S} = \mathbf{W}_3 \mathbf{V}_3^T$ )
- Устраняем аффинную неоднозначность



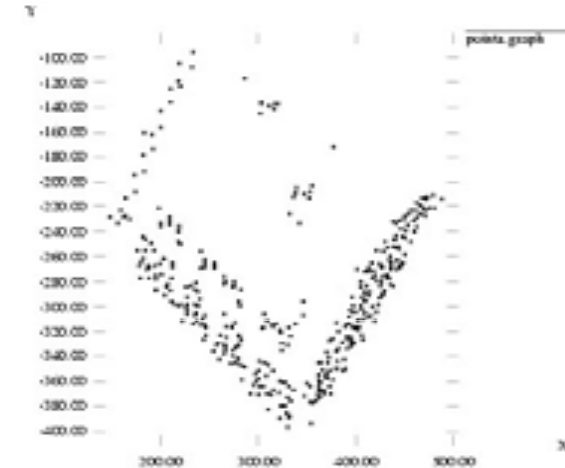
# Результат реконструкции



1



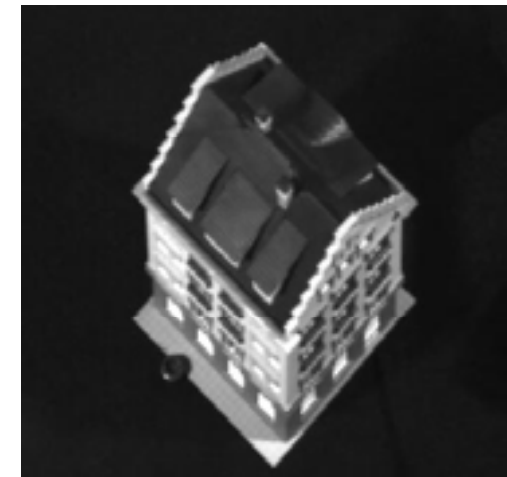
60



120



150



C. Tomasi and T. Kanade. [Shape and motion from image streams under orthography: A factorization method.](#) *IJCV*, 9(2):137-154, November 1992.





# Пропущенные данные

- В методе факторизации используется предположение, что все точки видны на всех изображениях
- В действительности, матрица измерений выглядит обычно так:



камеры

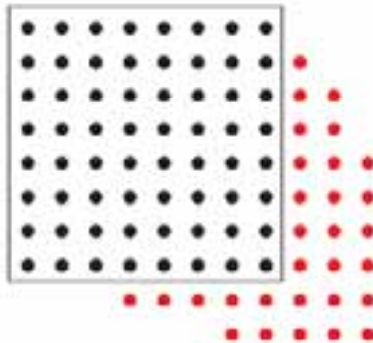
ТОЧКИ



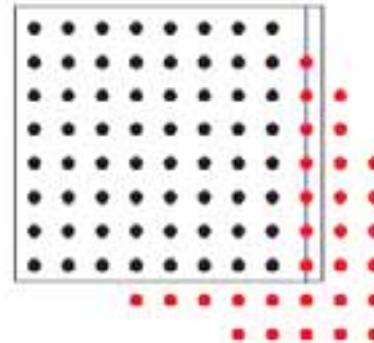
# Пропущенные данные

- Решение: разбиваем матрицу на плотные блоки, факторизуем каждый блок, объединяем результаты

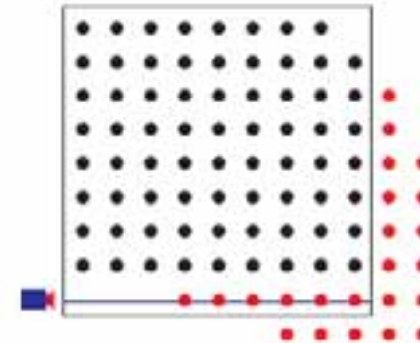
- Поиск максимальных блоков в матрице NP-полная задача (эквивалентная поиску максимальных клик в графе) maximal



(1) Факторизация блока



(2) Триангулируем 3D точки, видимые хотя бы на 2х камерах



(3) Вычисляем калибровку камер, которые видят хотя бы 3 известные 3D точки

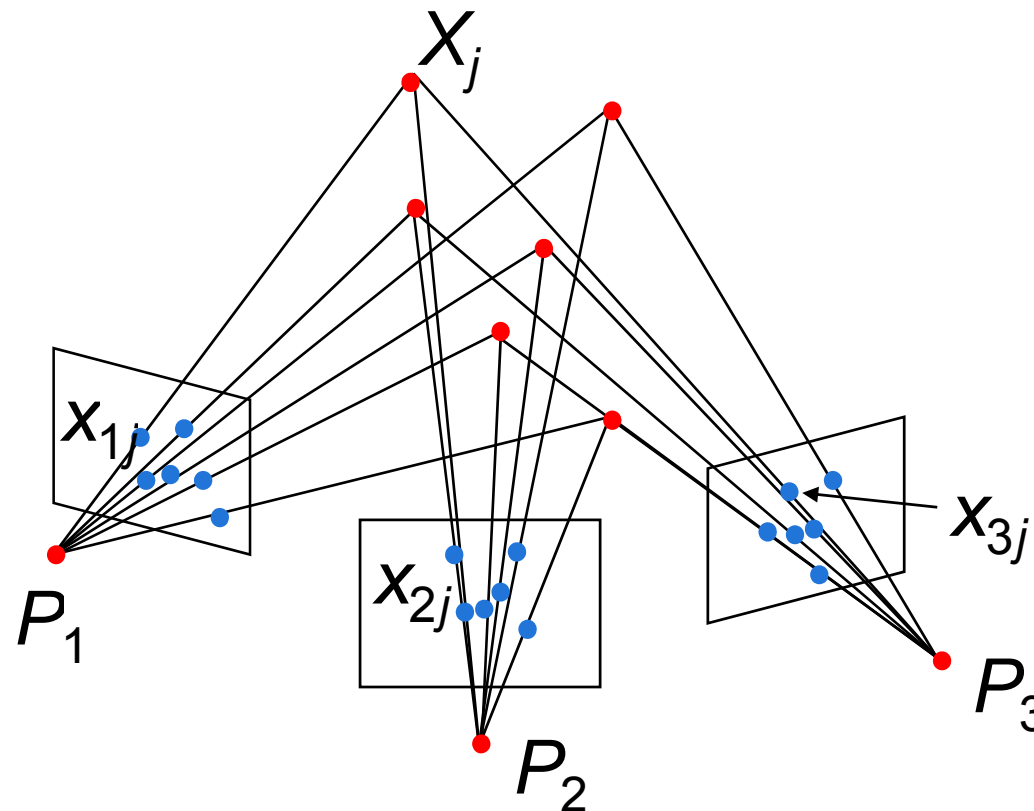


# Проективная реконструкция

- Дано:  $m$  изображений  $n$  фиксированных 3D точек

$$\bullet z_{ij} \mathbf{x}_{ij} = \mathbf{P}_i \mathbf{X}_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

- Задача: оценить  $m$  матриц проекции  $\mathbf{P}_i$  и  $n$  3D точек  $\mathbf{X}_j$  из  $mn$  соответствий  $\mathbf{x}_{ij}$





# Проективная реконструкция

---

- Дано:  $m$  изображений  $n$  фиксированных 3D точек

$$\bullet z_{ij} \mathbf{x}_{ij} = \mathbf{P}_i \mathbf{X}_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

- Задача: оценить  $m$  матриц проекции  $\mathbf{P}_i$  и  $n$  3D точек  $\mathbf{X}_j$  из  $mn$  соответствий  $\mathbf{x}_{ij}$
- В отсутствие информации о калибровке, камеры и точки можно восстановить с точностью до 4x4 проективного преобразования  $\mathbf{Q}$ :

$$\bullet \mathbf{X} \rightarrow \mathbf{QX}, \mathbf{P} \rightarrow \mathbf{PQ}^{-1}$$

- Решить задачу можно в следующем случае

$$\bullet 2mn \geq 11m + 3n - 15$$

- Для 2х камер, нужно не менее 7 точек



## SFM: Случай 2х камер

---

- Вычислим фундаментальную матрицу  $\mathbf{F}$
- Первая матрица камеры  $[\mathbf{I}|\mathbf{0}]$
- Вторая матрица камеры  $[\mathbf{A}|\mathbf{b}]$
- Тогда  $z\mathbf{x} = [\mathbf{I} | \mathbf{0}]\mathbf{X}$ ,  $z'\mathbf{x}' = [\mathbf{A} | \mathbf{b}]\mathbf{X}$

$$z'\mathbf{x}' = \mathbf{A}[\mathbf{I} | \mathbf{0}]\mathbf{X} + \mathbf{b} = z\mathbf{A}\mathbf{x} + \mathbf{b}$$

$$z'\mathbf{x}' \times \mathbf{b} = z\mathbf{A}\mathbf{x} \times \mathbf{b}$$

$$(z'\mathbf{x}' \times \mathbf{b}) \cdot \mathbf{x}' = (z\mathbf{A}\mathbf{x} \times \mathbf{b}) \cdot \mathbf{x}'$$

$$\mathbf{x}'^T [\mathbf{b}_\times] \mathbf{A} \mathbf{x} = 0$$

$$\mathbf{F} = [\mathbf{b}_\times] \mathbf{A} \quad \mathbf{b}: \text{эпиполь } (\mathbf{F}^T \mathbf{b} = 0), \quad \mathbf{A} = -[\mathbf{b}_\times] \mathbf{F}$$



# Проективная факторизация

$$\mathbf{D} = \begin{bmatrix} z_{11}\mathbf{X}_{11} & z_{12}\mathbf{X}_{12} & \cdots & z_{1n}\mathbf{X}_{1n} \\ z_{21}\mathbf{X}_{21} & z_{22}\mathbf{X}_{22} & \cdots & z_{2n}\mathbf{X}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{m1}\mathbf{X}_{m1} & z_{m2}\mathbf{X}_{m2} & \cdots & z_{mn}\mathbf{X}_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \\ \mathbf{P}_m \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_n \end{bmatrix}$$

ТОЧКИ ( $4 \times n$ )

камеры  
( $3m \times 4$ )

$$\mathbf{D} = \mathbf{MS} \text{ ранг } 4$$

- Если мы знаем глубину  $z$ , мы можем разложить  $\mathbf{D}$  и оценить  $\mathbf{M}$  и  $\mathbf{S}$
- Если мы знаем  $\mathbf{M}$  и  $\mathbf{S}$ , мы можем найти  $z$
- Решение: итеративный метод (чередуют оба вышеуказанных шага)



# Недостатки факторизации

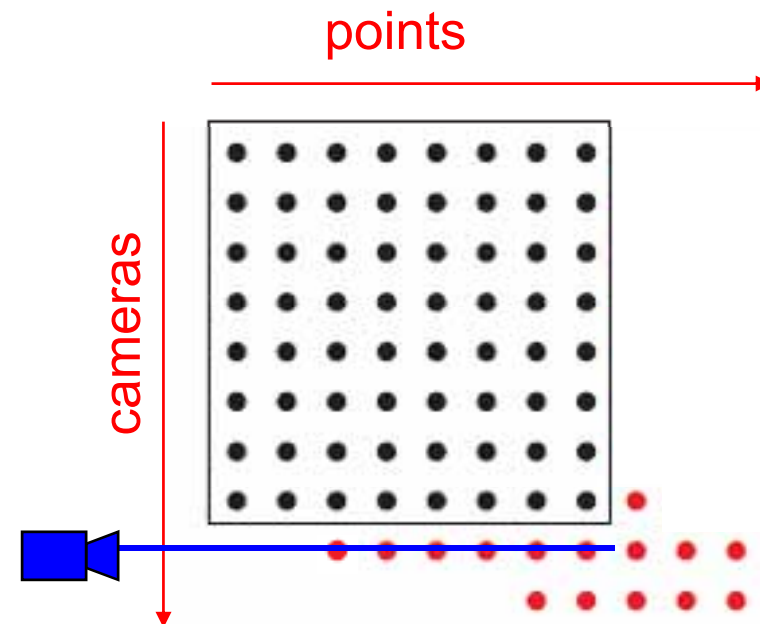
---

- Предполагается, что все точки видны на всех кадрах
  - Приходится разбивать на отдельные задачи и решать их
- Предполагается, что все точки правильные, нет ложных соответствий
  - В факторизации нет простого способа фильтровать ложные соответствия
- Не оптимизируется оптимальный («gold standard») критерий
- Поэтому обычно используется последовательный метод «структуры и движения»



# Последовательный SFM

- Инициализируем движение через существенную матрицу двух камер
- Инициализируем структуру
- Для каждого вида:
  - Оцениваем матрицу проекции новой камеры по всем известным 3D точкам, видимым на этом изображении – калибровка

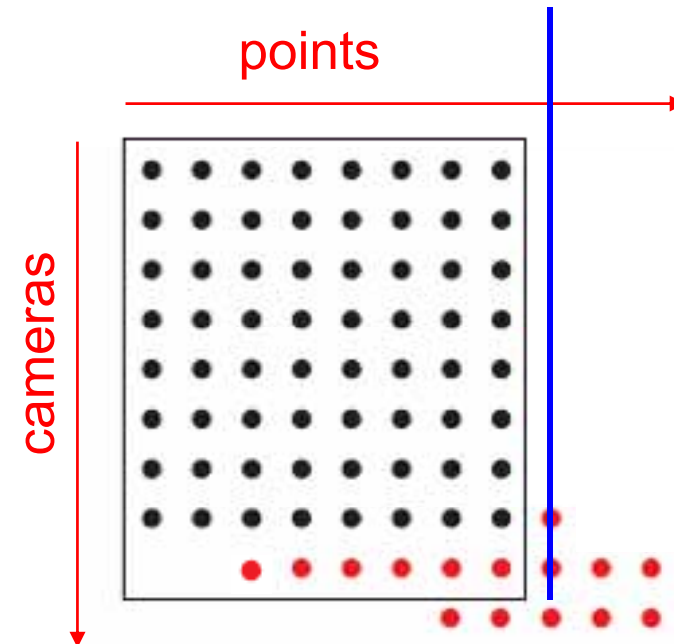






# Последовательный SFM

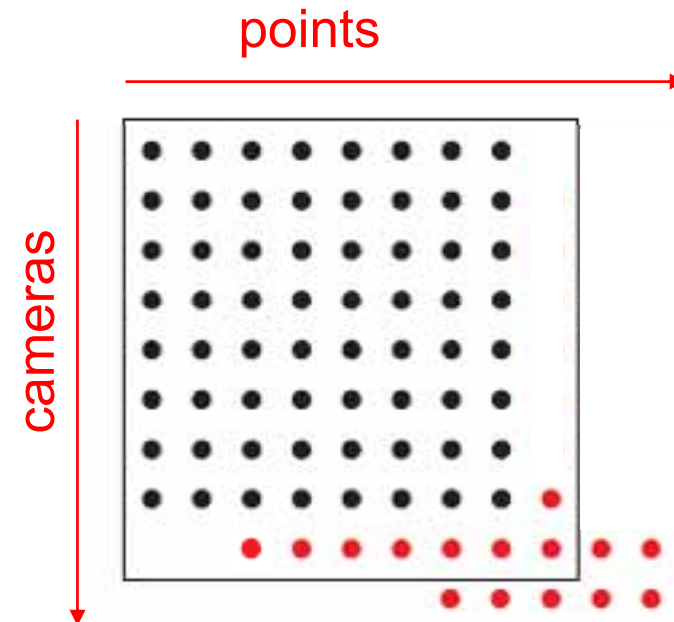
- Инициализируем движение через существенную матрицу двух камер
- Инициализируем структуру
- Для каждого вида:
  - Оцениваем матрицу проекции новой камеры по всем известным 3D точкам, видимым на этом изображении – калибровка
  - Уточняем и дополняем структуру, вычисляем новые 3D точки, уточняем существующие точки, видимые на камере - триангуляция





# Последовательный SFM

- Инициализируем движение через существенную матрицу двух камер
- Инициализируем структуру
- Для каждого вида:
  - Оцениваем матрицу проекции новой камеры по всем известным 3D точкам, видимым на этом изображении – калибровка
  - Уточняем и дополняем структуру, вычисляем новые 3D точки, уточняем существующие точки, видимые на камере - триангуляция
  - Уточняем структуру и в движение методом связок

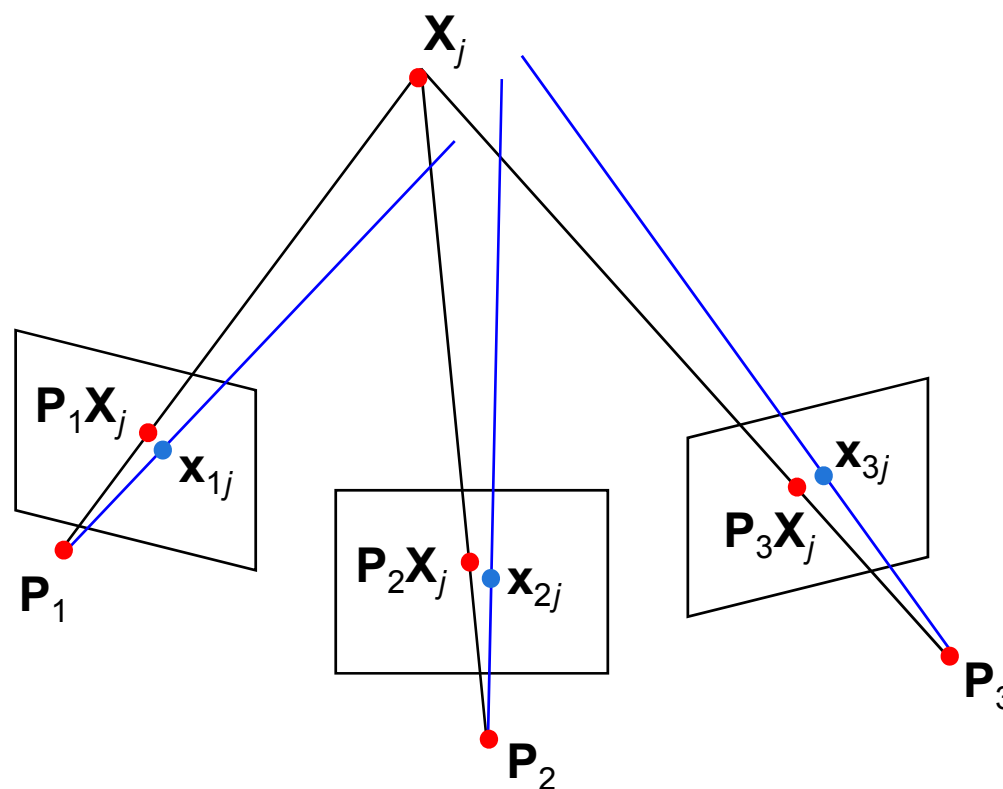




# Метод связок (Bundle adjustment)

- Нелинейный метод для уточнения структуры и движения
- Минимизируем сумму ошибок проекций всех точек на все камеры:

$$E(\mathbf{P}, \mathbf{X}) = \sum_{i=1}^m \sum_{j=1}^n D(\mathbf{x}_{ij}, \mathbf{P}_i \mathbf{X}_j)^2$$





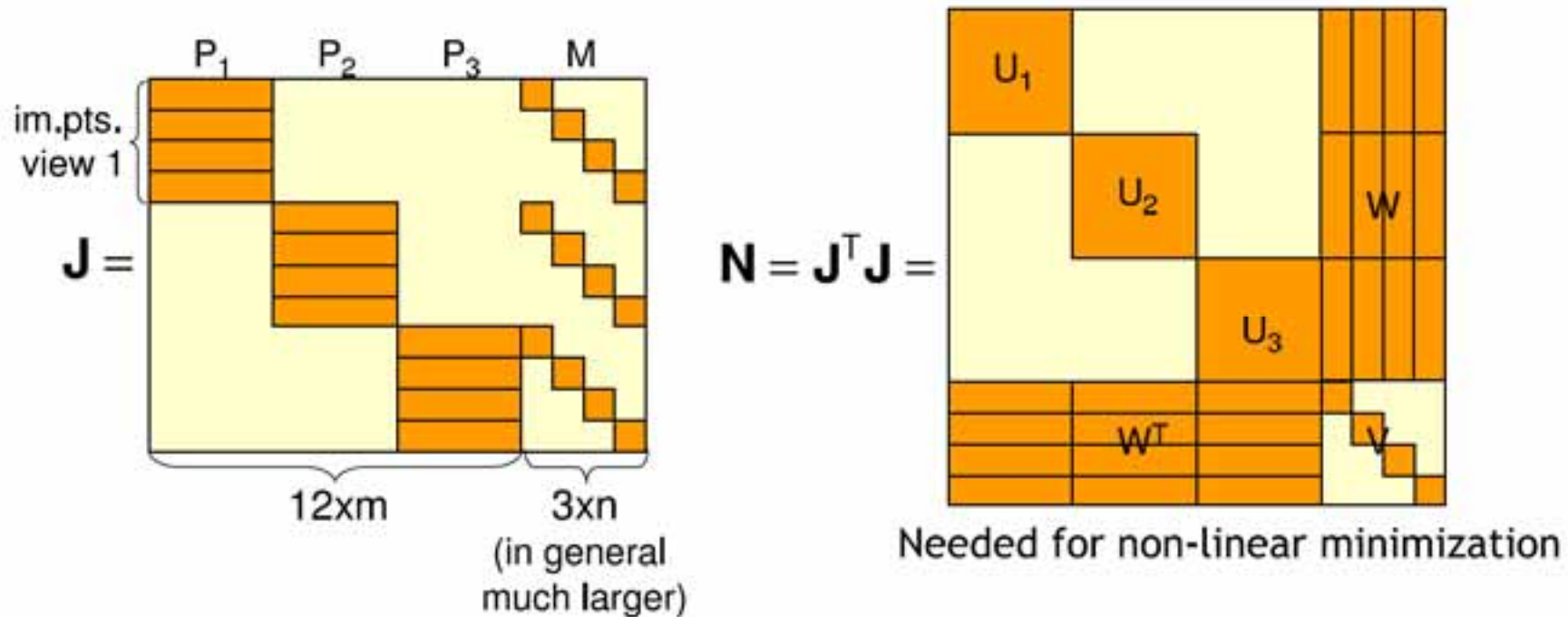
# Итеративная оптимизация

---

- Имеем задачу  $X = f(P)$ , где  $X$  – вектор измерений,  $P$  – параметры,  $f()$  – функция
- Есть начальное приближение  $e_0 = f(P_0) - X$
- Предположим, что можно приблизить  $f(P_0 + \Delta) \approx f(P_0) + J\Delta$
- Ищем  $P_1$ , минимизирующее
$$f(P_1) - X = f(P_0) + J\Delta - X = e_0 + J\Delta$$
- Задача, найти минимум  $\|e_0 + J\Delta\|$  по  $\Delta$
- Или  $(e_0 + J\Delta)^T (e_0 + J\Delta)$
- Нужно решить нормальные уравнения  $J^T J\Delta = -J^T e_0$



# Блочная структура матриц



- Огромный размер матриц, но блочная структура
- Специальные реализация методов оптимизации для матриц такой формы
- Методы улучшения структуры матриц для повышения производительности



# Проективная реконструкция

---

- Вместо существенной матрицы можем оценивать фундаментальную и использовать специальную триангуляцию
- Затем наложим дополнительные ограничения для устранения неоднозначности реконструкции



# Автокалибровка (Self-calibration)

---

- Автокалибровка – это оценка внутренних параметров камеры непосредственно по некалиброванным изображениям
- Например, если изображения получены с одной движущейся камеры, мы можем использовать ограничение, что матрица калибровки фиксирована для всех изображений
  - Вычислим исходную проективную реконструкцию и найдем такое преобразование  $\mathbf{Q}$ , что матрицы проекций всех камер можно представить в виде  $\mathbf{P}_i = \mathbf{K} [\mathbf{R}_i | \mathbf{t}_i]$
- Можем использовать ограничения на форму матрицы калибровки, пр. нулевой скос (прямоугольную форму пикселей)



# Большая задача

---



Коллекция изображений  
из интернета

Трёхмерная модель

В 24 часа...





## Несколько миллионов изображений по тегу «Rome»



## Обычные фотографии

---

- Получены разными пользователями с разных фотоаппаратов



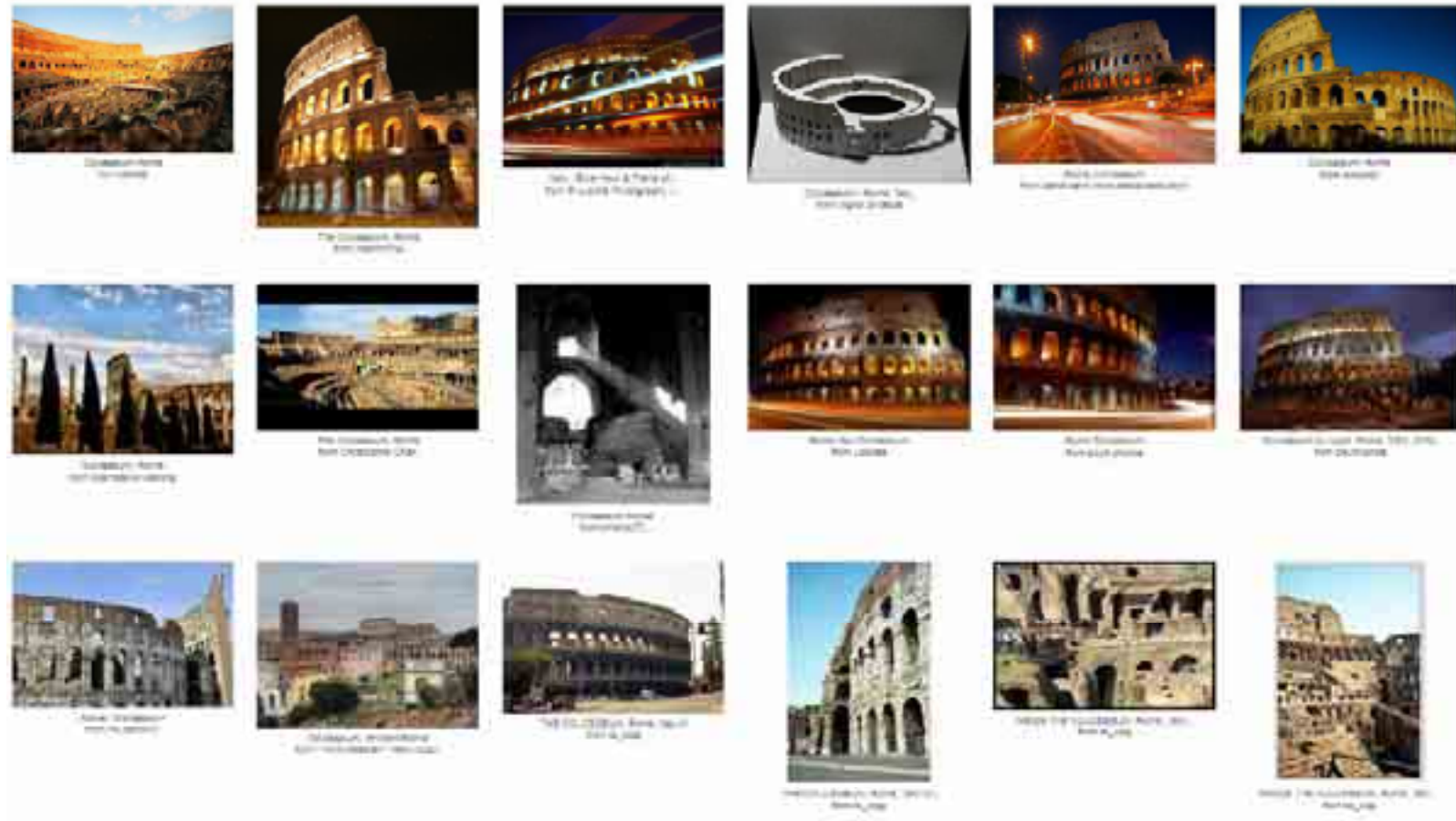
iPhone 3G



Nikon D3



# Обычные фотографии



- Разное время года, разное время суток
- Случайный порядок изображений



# Загрузим из интернета



[http://graphics.cs.cmu.edu/projects/im2gps/flickr\\_code.html](http://graphics.cs.cmu.edu/projects/im2gps/flickr_code.html)



# Найдем соответствующие точки

---







# Поиск особых точек (SIFT)

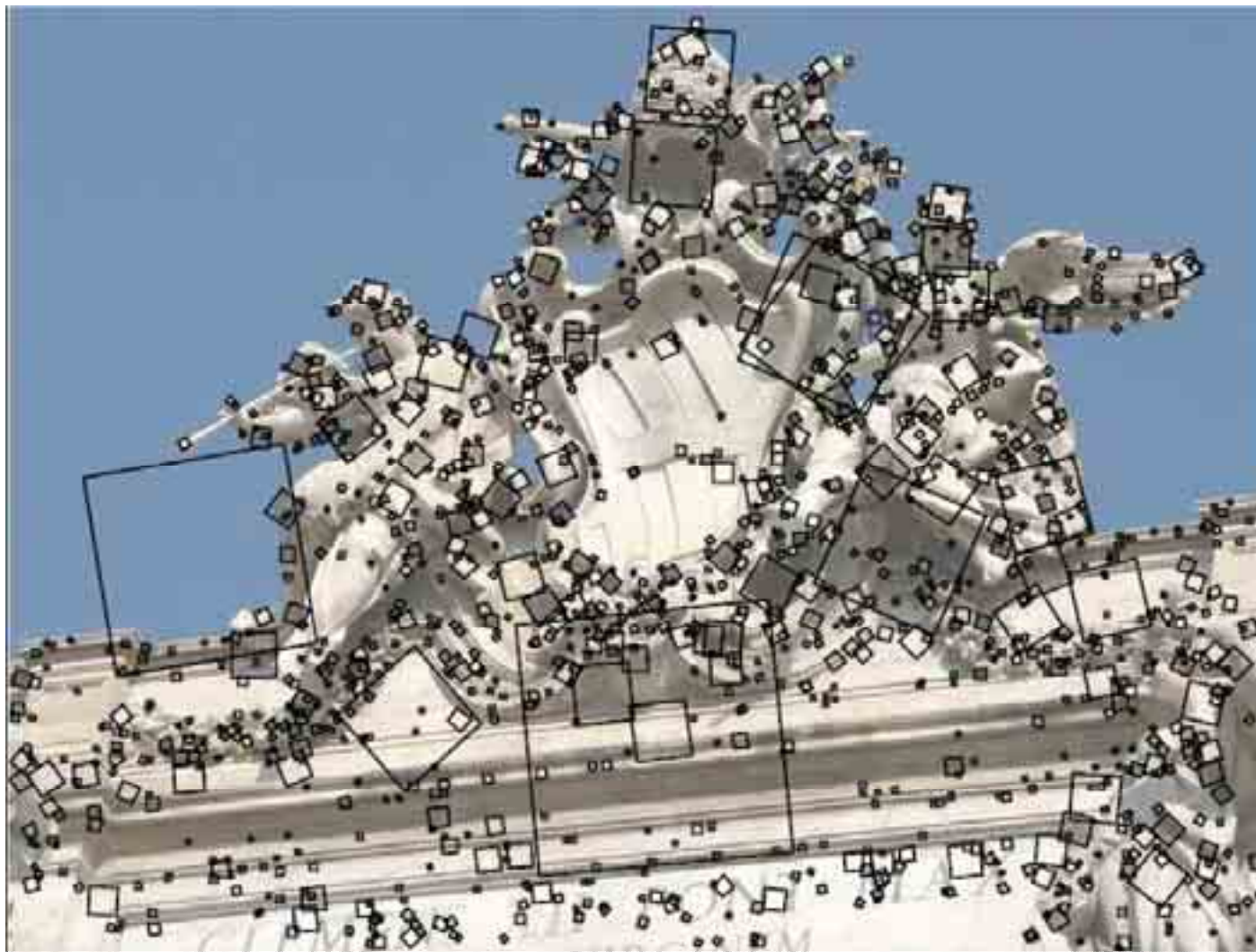
---





# Найденные особые точки

---





# Сопоставление изображений

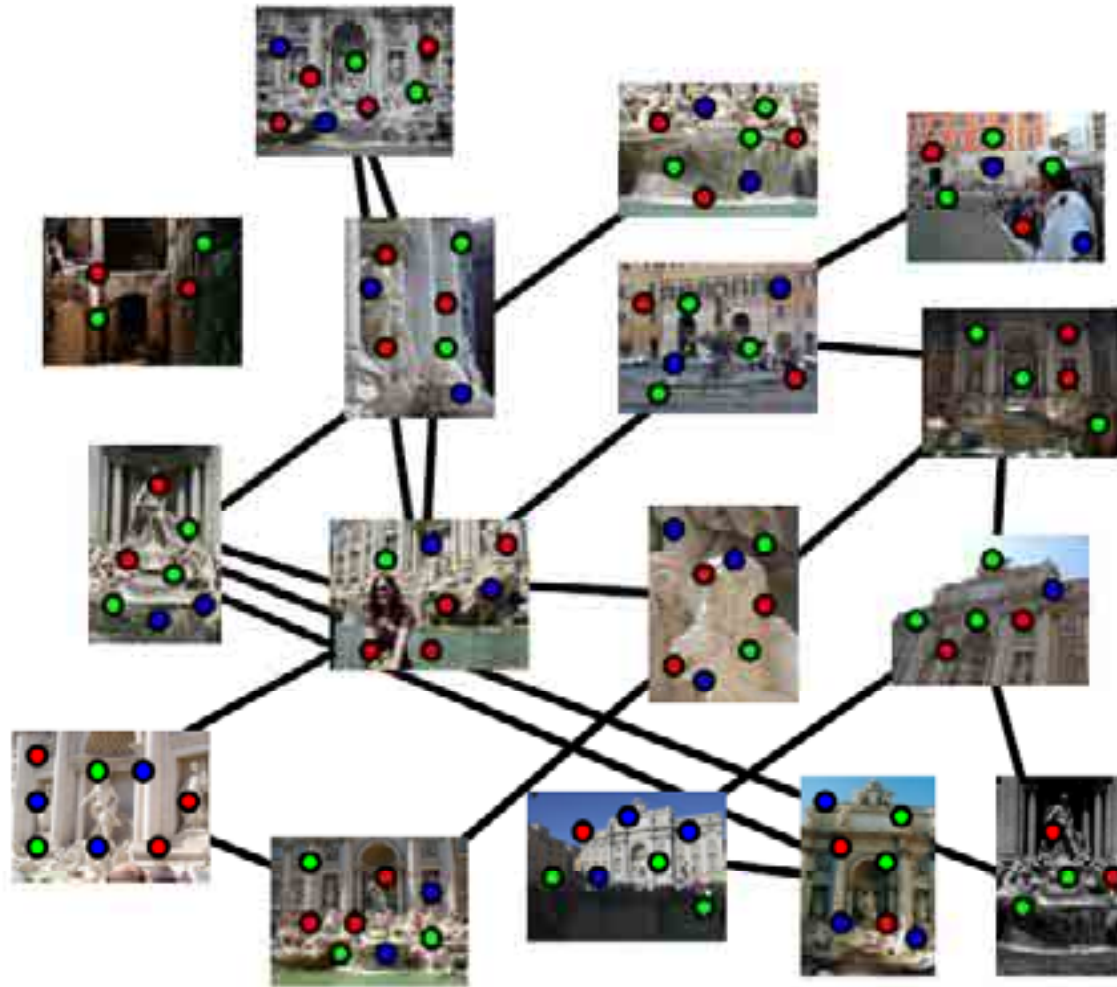
---

- Сопоставляем точки по дескрипторам SIFT
- Фильтруем ложные соответствия по эпиполярной геометрии
- Для этого будем с помощью RANSAC вычислять фундаментальную матрицу





# Граф связности изображений





# Объем задачи

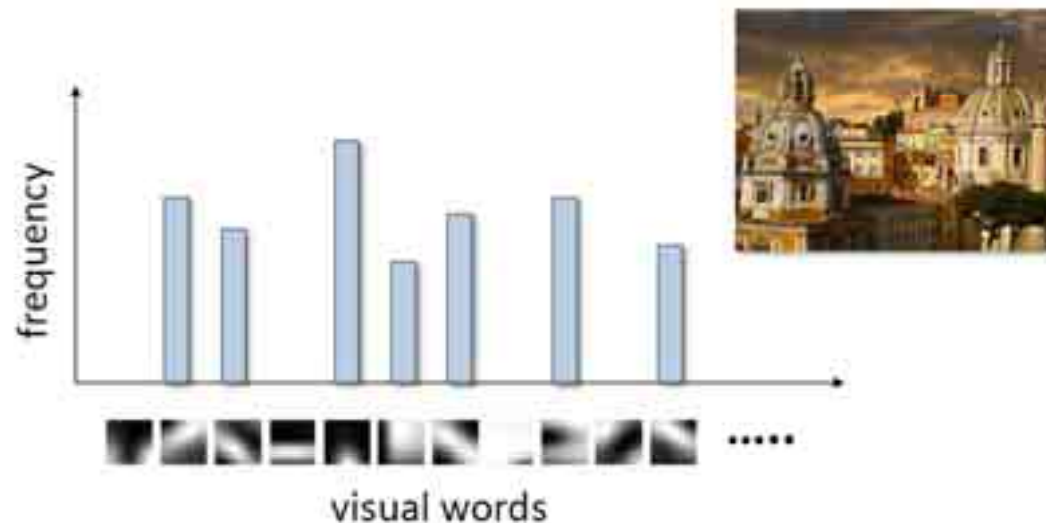
---

- 250K изображений -> 31M пар изображений
- 2 пары изображений в секунду
  - 1 год на 500 машинах
- 1M изображений -> 500 000M пар изображений
  - 15 лет на 500 машинах



# Поиск похожих изображений

- Метод «Мешок слов»
  - Обучаем словарь визуальных слов размером 100К
  - Строим гистограмму частот для каждого изображения:

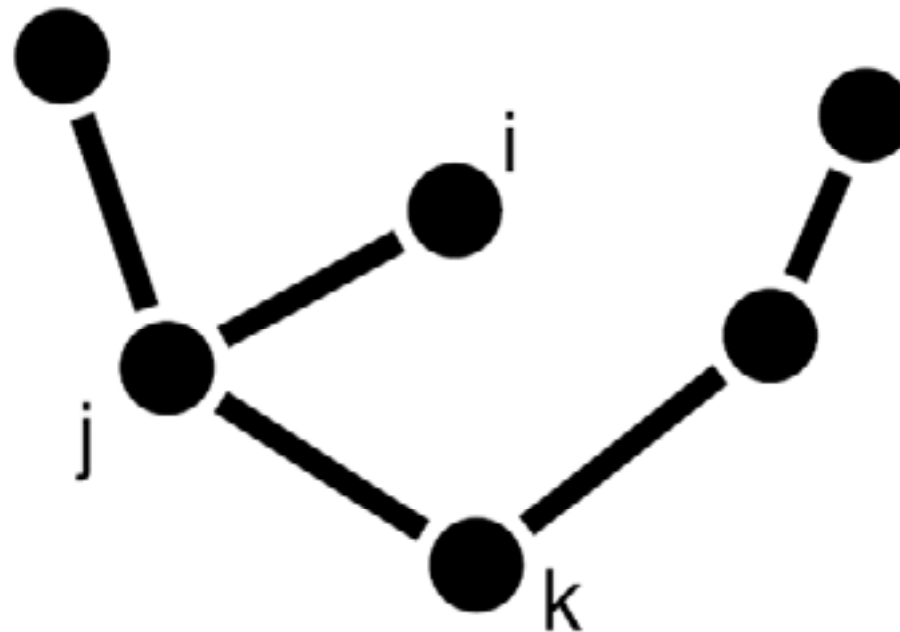


- Для каждого изображения ищем 40 наиболее похожих
- Сопоставляем точки между этими парами изображений
- Раскрываем запросы (query expansion)



# Раскрытие запросов

---

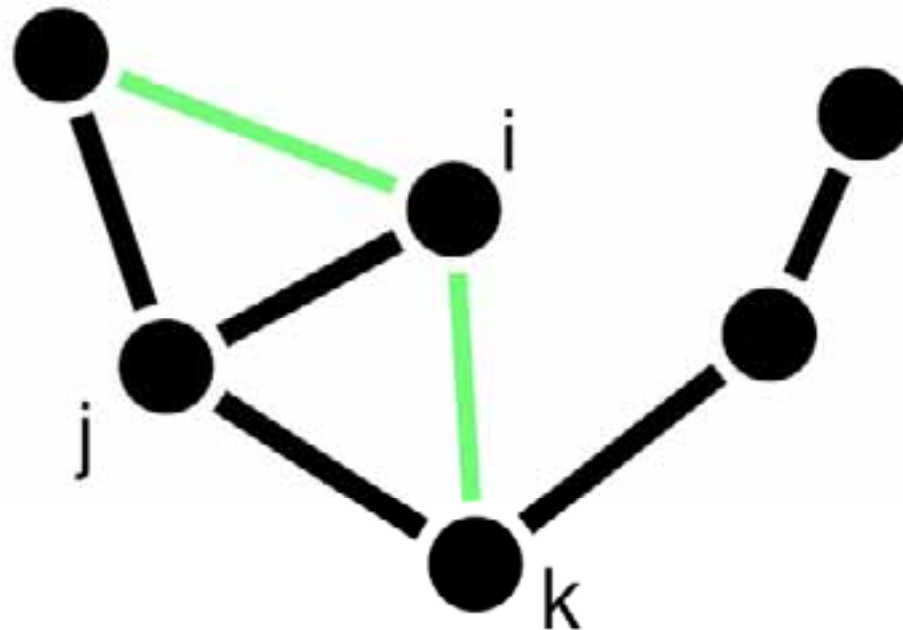


O. Chum *et al.* *Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval*. ICCV '07.



# Раскрытие запросов

---

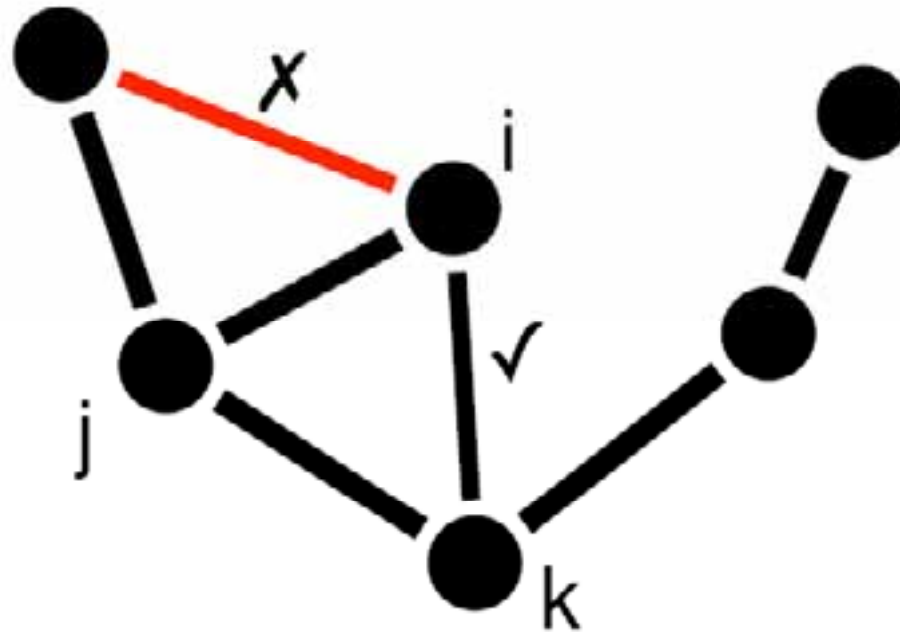


O. Chum *et al.* *Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval*. ICCV '07.



# Раскрытие запросов

---

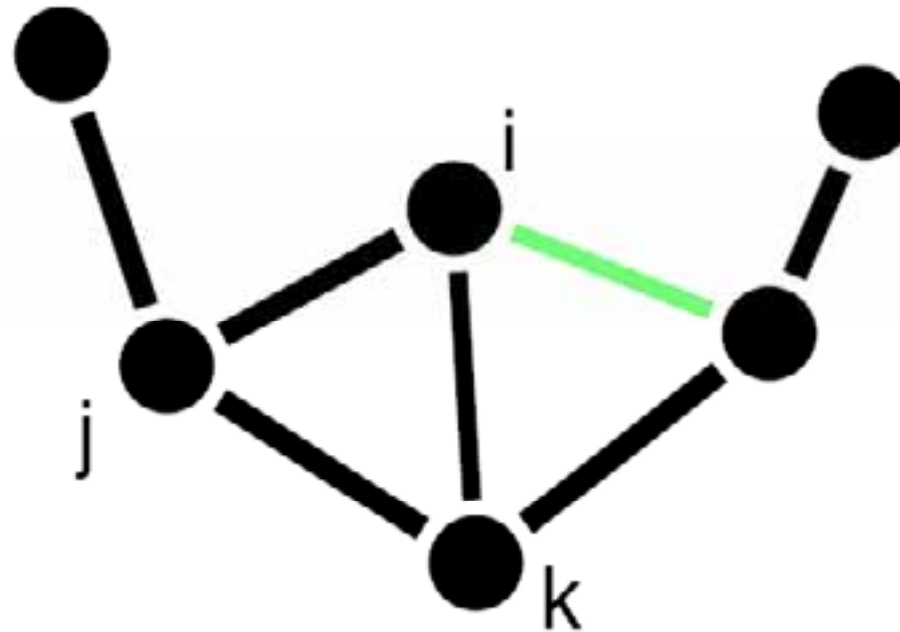


O. Chum *et al.* *Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval*. ICCV '07.



# Раскрытие запросов

---



O. Chum *et al.* *Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval*. ICCV '07.



# Результат сопоставлений

---



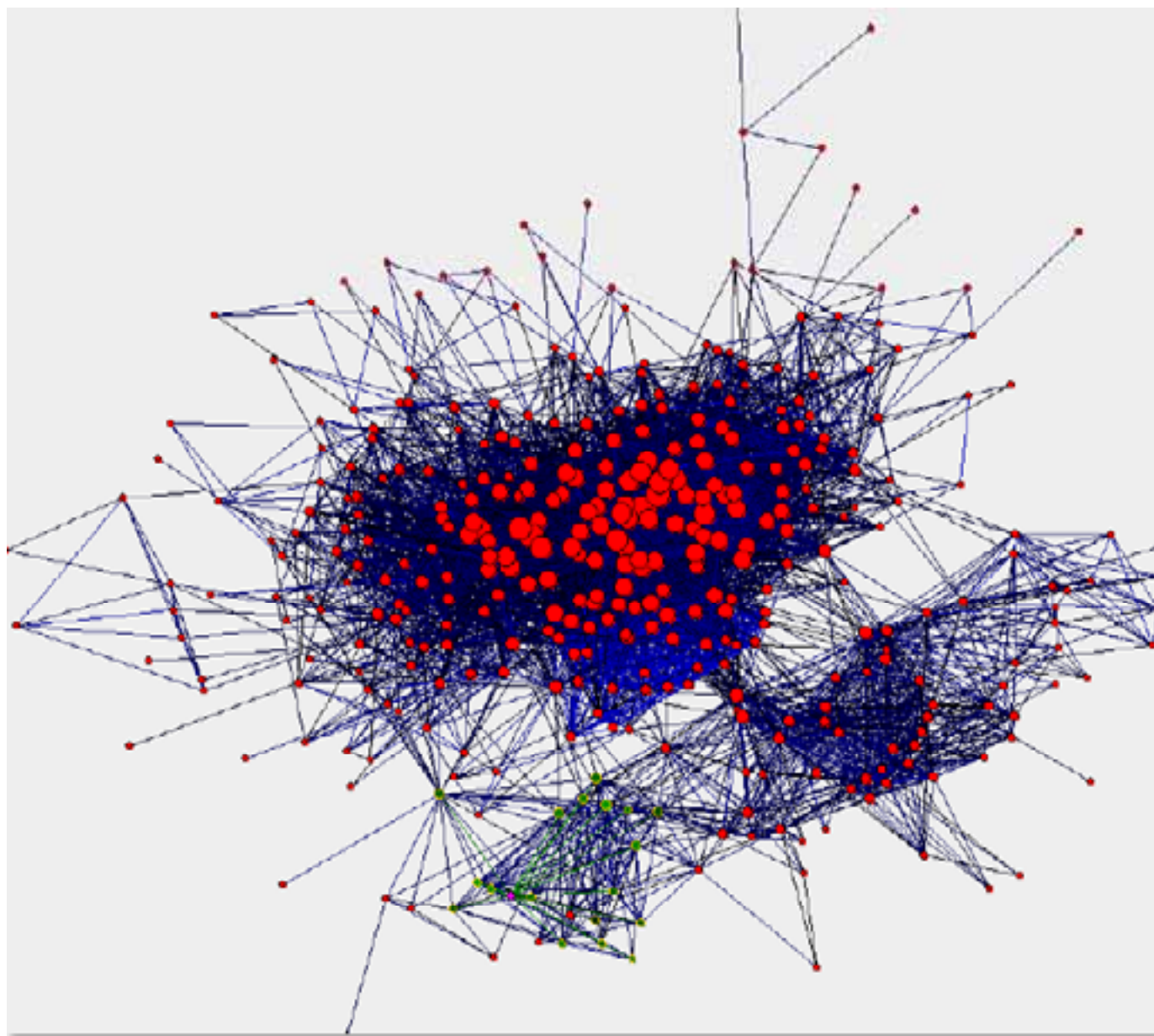
Мощность SIFT!





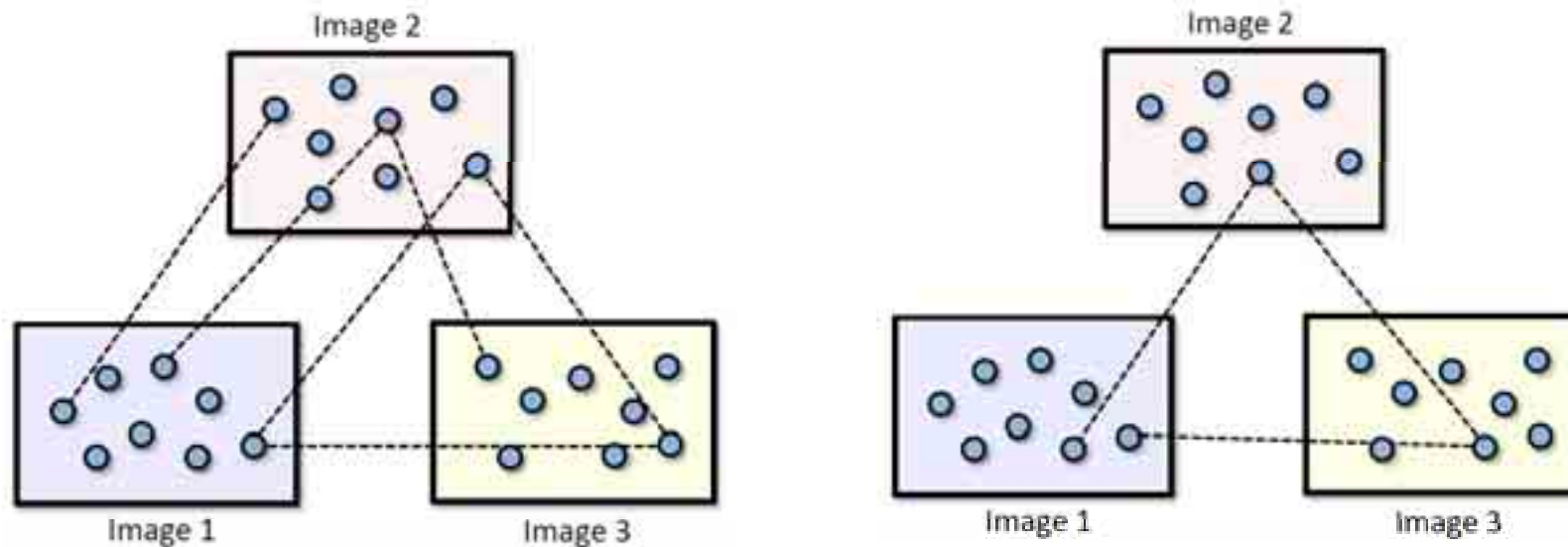
# Граф связности изображений

---





# Построение «следов»

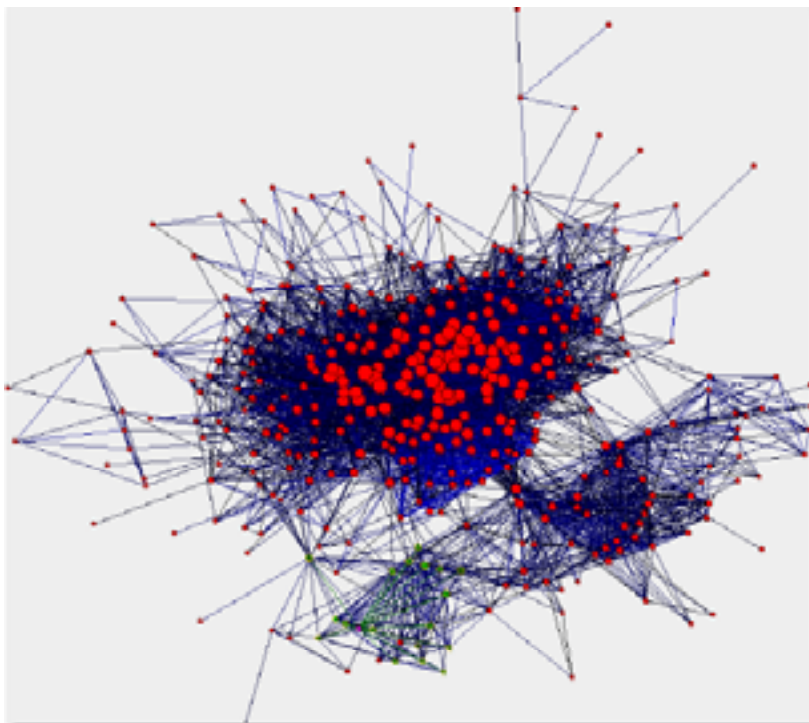


- Объединяем соответствия в «следы»
  - Отбрасываем незамкнутые следы, как потенциальные ошибки
- Пример (3000 изображений), всего 1.5M+ следов
  - 79% длины 2
  - 90% длины  $\leq 3$
  - 98% длины  $\leq 10$
  - Самый длинный след 385 точек

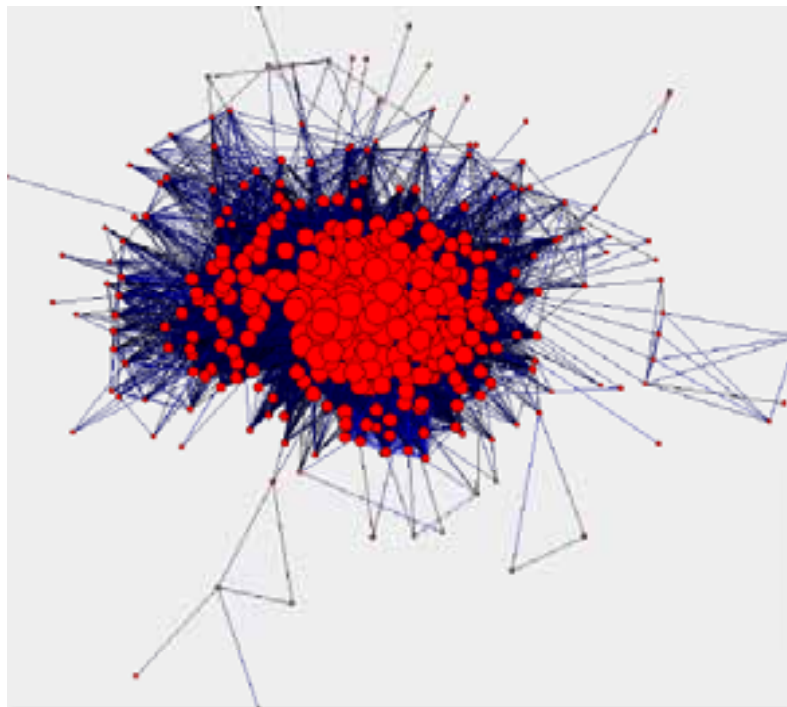


# Граф связности изображений

---



Исходные  
соответствия

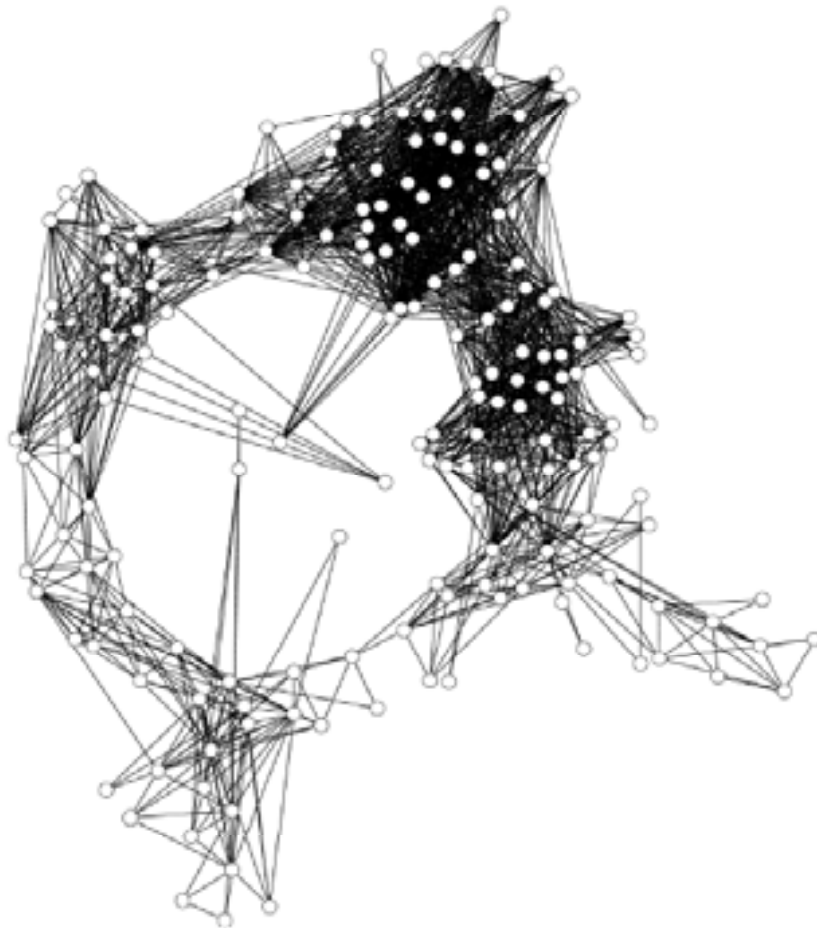


После построения следов



# «Скелет» графа

---



[Snavely, *et al.*, "Skeletal sets for efficient structure from motion," CVPR 2008]



# Стандартная схема

---

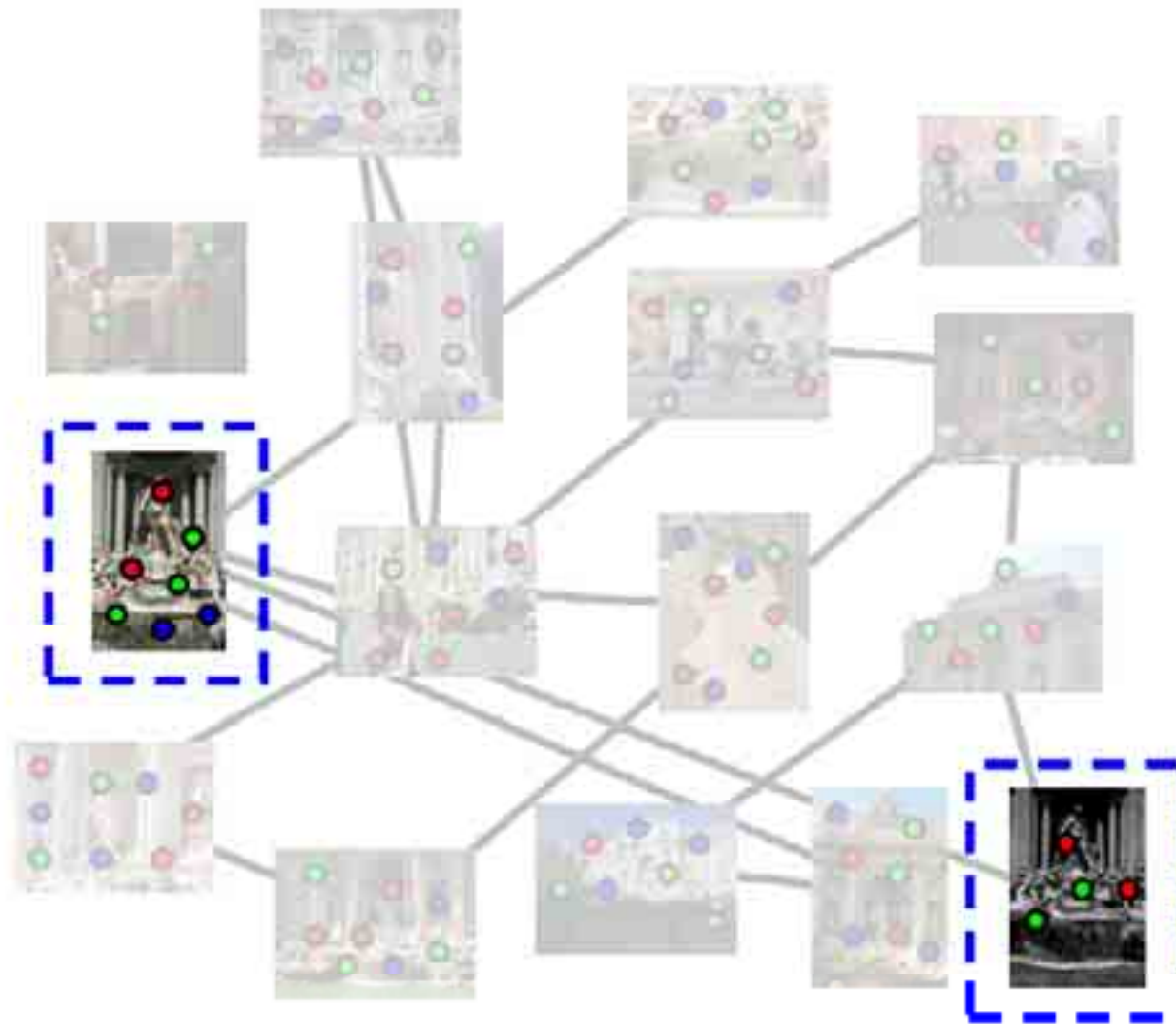
- Выбор опорных пар изображений
- Вычисление структуры по паре изображений
  - Построение матрицы внутренней калибровки
  - Вычисление фундаментальной и существенной матрицы
  - Триангуляция точек
  - Уточнение методом связок
- Добавление новых изображений
  - Калибровка камеры по известным 2D/3D соответствиям
  - Триангуляция новых точек
  - Уточнение методом связок





# Выбор опорной пары

---





# Выбор опорной пары

---



(+) много точек  
(-) малая база



(+) большая база  
(-) мало точек



(+) много точек  
(+) большая база





# Выбор опорной пары

---

- Возможны разные эвристики
- Гомография плохо описывает сцену, когда сцена не плоская, и существенный параллакс (большая база)
- Пример:
  - Возьмем пары, в которых  $> 100$  соответствий
  - Вычислим гомографию и фундаментальную матрицу
  - Выберем пару изображений, минимизирующее отношение:

$$\frac{\text{numInliers}_{\text{homography}}}{\text{numInliers}_{\text{Fmatrix}}}$$





# Внутренняя калибровка

---



File size : 85111 bytes  
File date : 2005:12:16 04:17:12  
**Camera make : Panasonic**  
**Camera model : DMC-FZ20**  
Date/Time : 2005:03:19 12:52:33  
Resolution : 450 x 600  
Flash used : No  
**Focal length : 6.0mm**  
Exposure time: 0.0012 s (1/800)  
Aperture : f/5.6  
ISO equiv. : 80  
Whitebalance : Auto  
Metering Mode: matrix  
Exposure : program (auto)

Структура EXIF-данных



# Параметры камеры

**Panasonic Lumix DMC-FZ20 digital camera specifications**

**Panasonic Lumix DMC-FZ20**



**Image**

**More information**

- [In-depth review](#)
- [Sample gallery](#)
- [Announced 21-Jul-04](#)
- [All Panasonic products](#)
- [Panasonic website](#)

**Discussion**

- [Panasonic Talk Forum](#)
- [Find related discussion](#)

**Owners opinions**

★★★★★

[Read owners comments \(101\)](#)  
[Post / Edit your opinion](#)

**Format** Compact, 1/2.5"

Price (street)	
Also known as	
Release Status	Discontinued
Max resolution	2560 x 1920
Low resolution	2048 x 1536, 1920 x 1080, 1600 x 1200, 1280 x 960, 640 x 480
Image ratio w/h	4:3, 16:9
Effective pixels	5.0 million
Sensor photo detectors	5.8 million
Sensor size	1/2.5" (5.75 x 4.31 mm, 0.24 cm <sup>2</sup> )
Pixel density	40.8 MV/cm <sup>2</sup>
Sensor type	CCD

[http://www.dpreview.com/reviews/specs/Panasonic/panasonic\\_dmc-fz20.asp](http://www.dpreview.com/reviews/specs/Panasonic/panasonic_dmc-fz20.asp)







## Dubrovnik, Croatia.

# images downloaded: 57,845

Size of largest connected component: 4,619

# points: ~2M

# observations: ~11.3M

Time: 18 hours





# Дубровник

---

Microsoft  
Research





# Photosynth

Microsoft  
Research





# Коды

---

- Pedagogical Bundle
  - <http://www.maths.lth.se/matematiklth/personal/nicolas/bundle.html>
- SBA
  - <http://www.ics.forth.gr/~lourakis/sba/>
- Bundler
  - <http://phototour.cs.washington.edu/bundler>



## На следующих лекциях

---

- Стереореконструкция
- Реконструкция по 1 изображению