

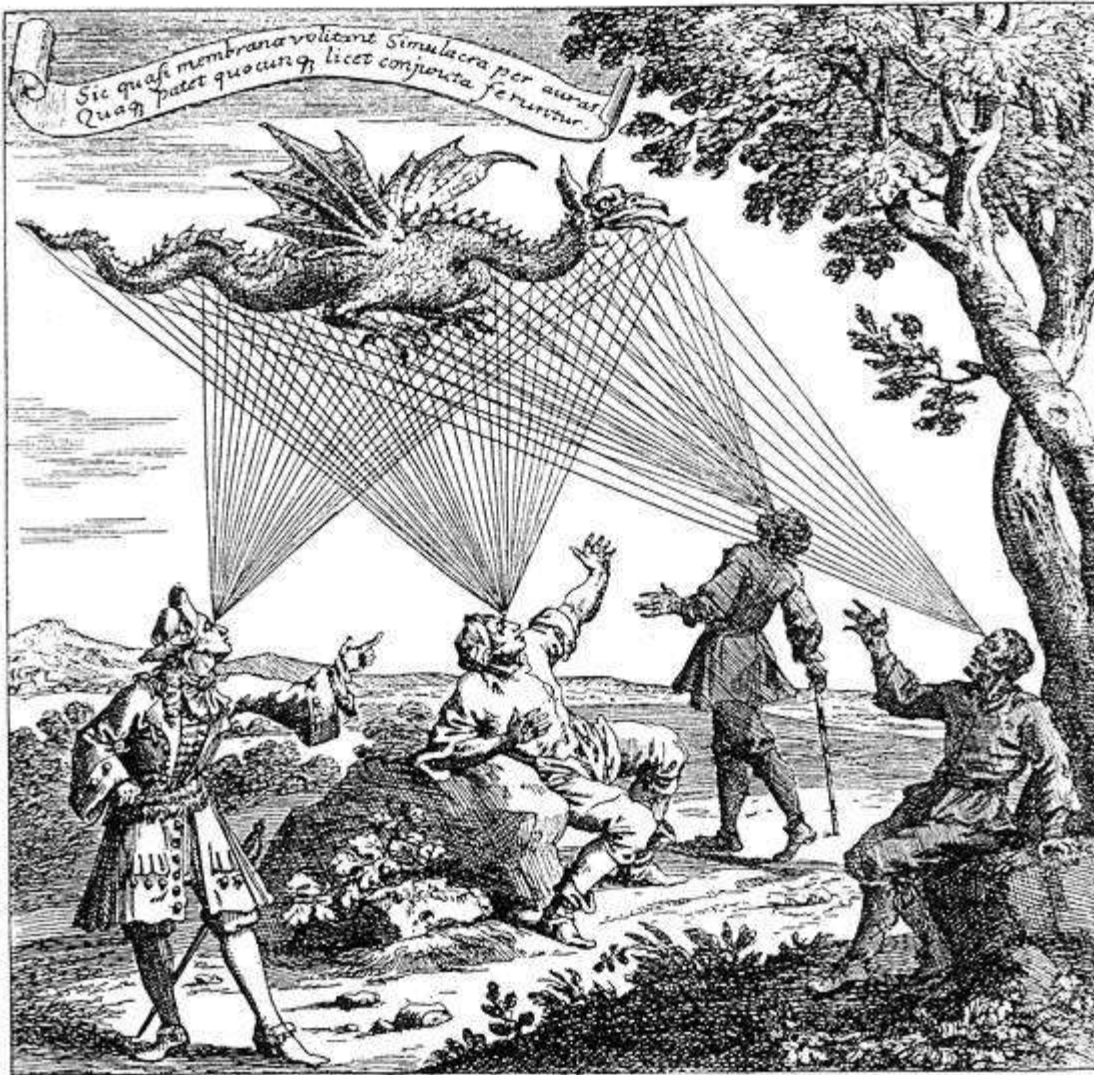
Курс «Трёхмерное компьютерное зрение»

Тема №4А

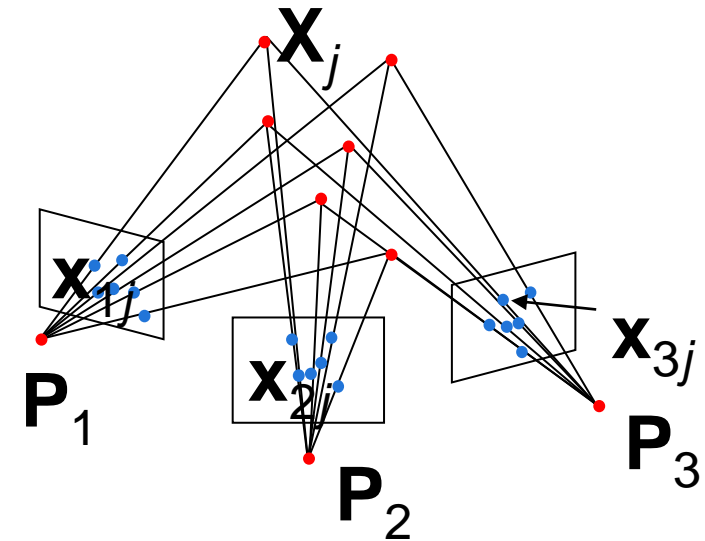
«Разреженная 3D реконструкция»

Антон Конушин

Структура из движения



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



Движение проекций 3D точек, вызванное движением камеры, позволяет определить координаты 3D точек (*структура, structure*) и само движение камеры (*движение, motion*).

Это задача «структура из движения» (Structure from Motion, SfM)

Пример работы SfM

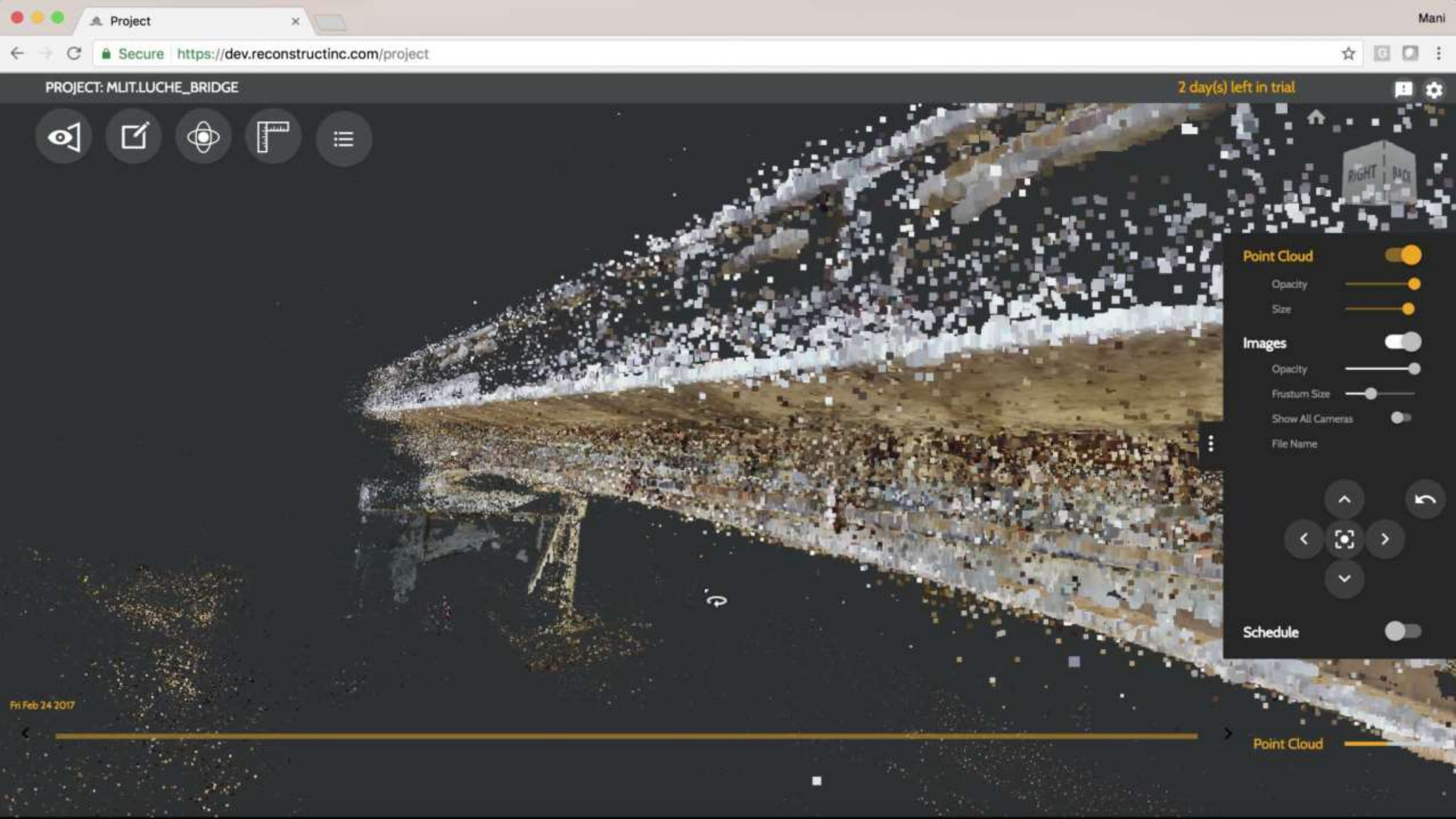


Пример приложения

Инспекция в
труднодоступных местах по
съёмке с дронов

- Строим 3D модель по изображениям
- Даём инструменты для визуального изучения изображения и отображения взаимодействия в 3D





Point Cloud

Opacity

Size

Images

Opacity

Frustum Size

Show All Cameras

File Name

Schedule



- Разреженная реконструкция по изображениям
- BigSFM
- COLMAP
- Задача SLAM
- Классический подход к SLAM
- Нейросетевой подход к SLAM



Схема разреженной 3D реконструкции по изображениям

3D реконструкция по изображениям



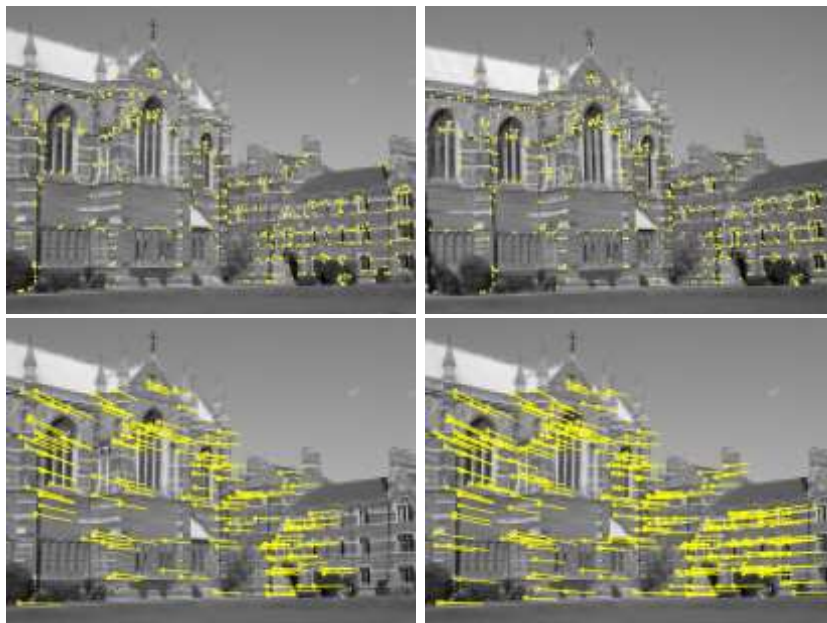
Набор изображений



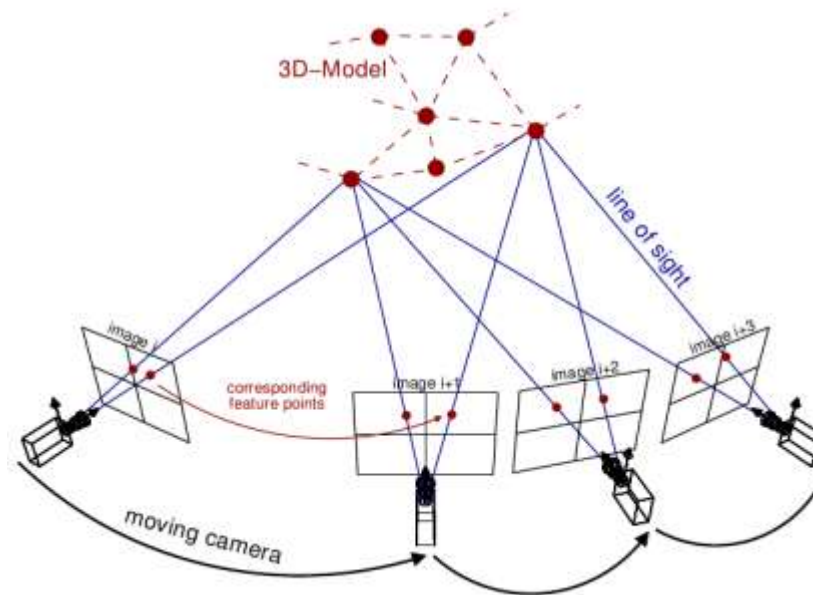
Трёхмерная (3D) модель

Пример задачи: построение 3D модели реального объекта по набору фотографий из интернета

Разреженная реконструкция



Point features and matching



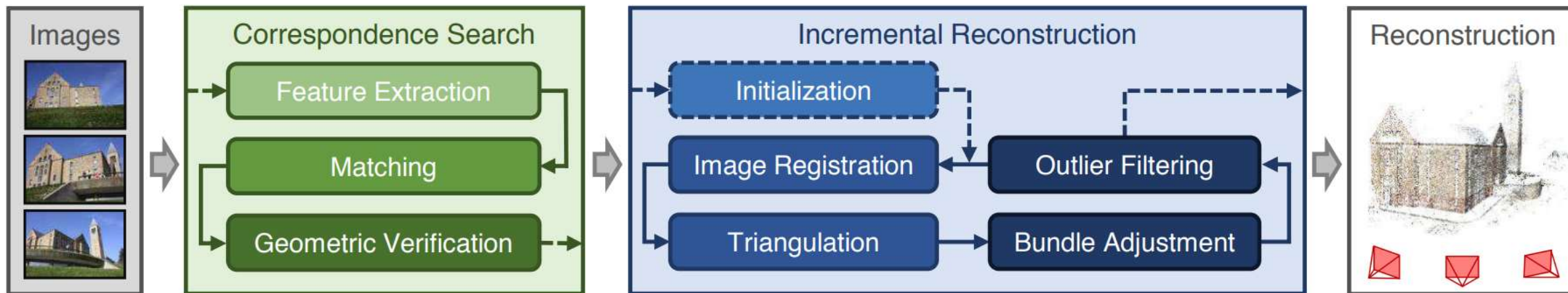
Camera localization and structure estimation

- Опираемся на сопоставление ключевых точек между изображениями
- 3D реконструкция в форме облака точек, соответствующих найденным ключевым точкам на изображениях
- Решаем совместно определение 3Д точек, соответствующих ключевым, и положение камер в пространстве относительно них
- 3D реконструкция – обратная задача к задаче компьютерной графики

Инкрементальный подход



Классический инкрементальный подход, в котором изображения по одному добавляются в модель до полного перебора всех изображений



Source: [Structure-From-Motion Revisited](#). CVPR2016

Коллекции изображений



Несколько миллионов изображений по тегу
«Rome»

Обычные фотографии



- Получены разными пользователями с разных фотоаппаратов



iPhone 3G



Nikon D3

Инициализация внутренней калибровка



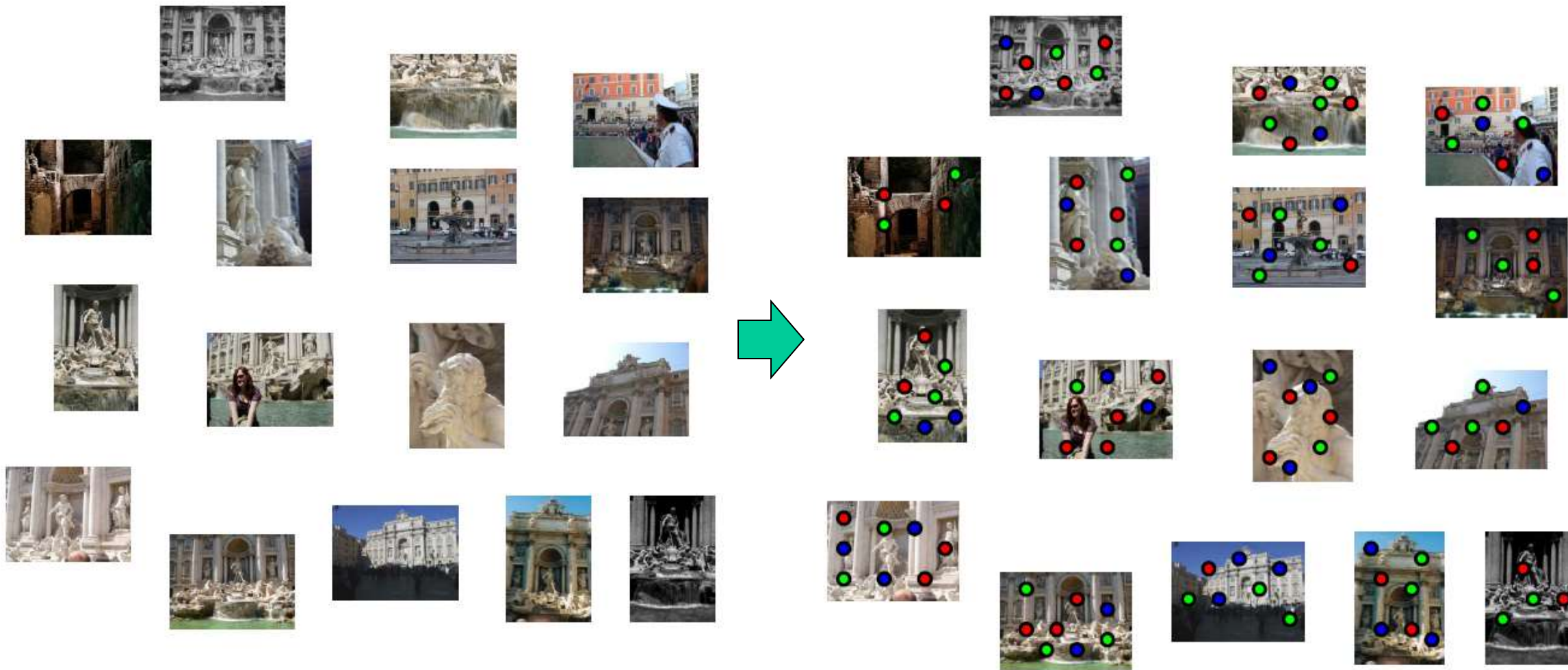
File size : 85111 bytes
File date : 2005:12:16 04:17:12
Camera make : **Panasonic**
Camera model : **DMC-FZ20**
Date/Time : 2005:03:19 12:52:33
Resolution : 450 x 600
Flash used : No
Focal length : **6.0mm**
Exposure time: 0.0012 s (1/800)
Aperture : f/5.6
ISO equiv. : 80
Whitebalance : Auto
Metering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

Структура EXIF-данных

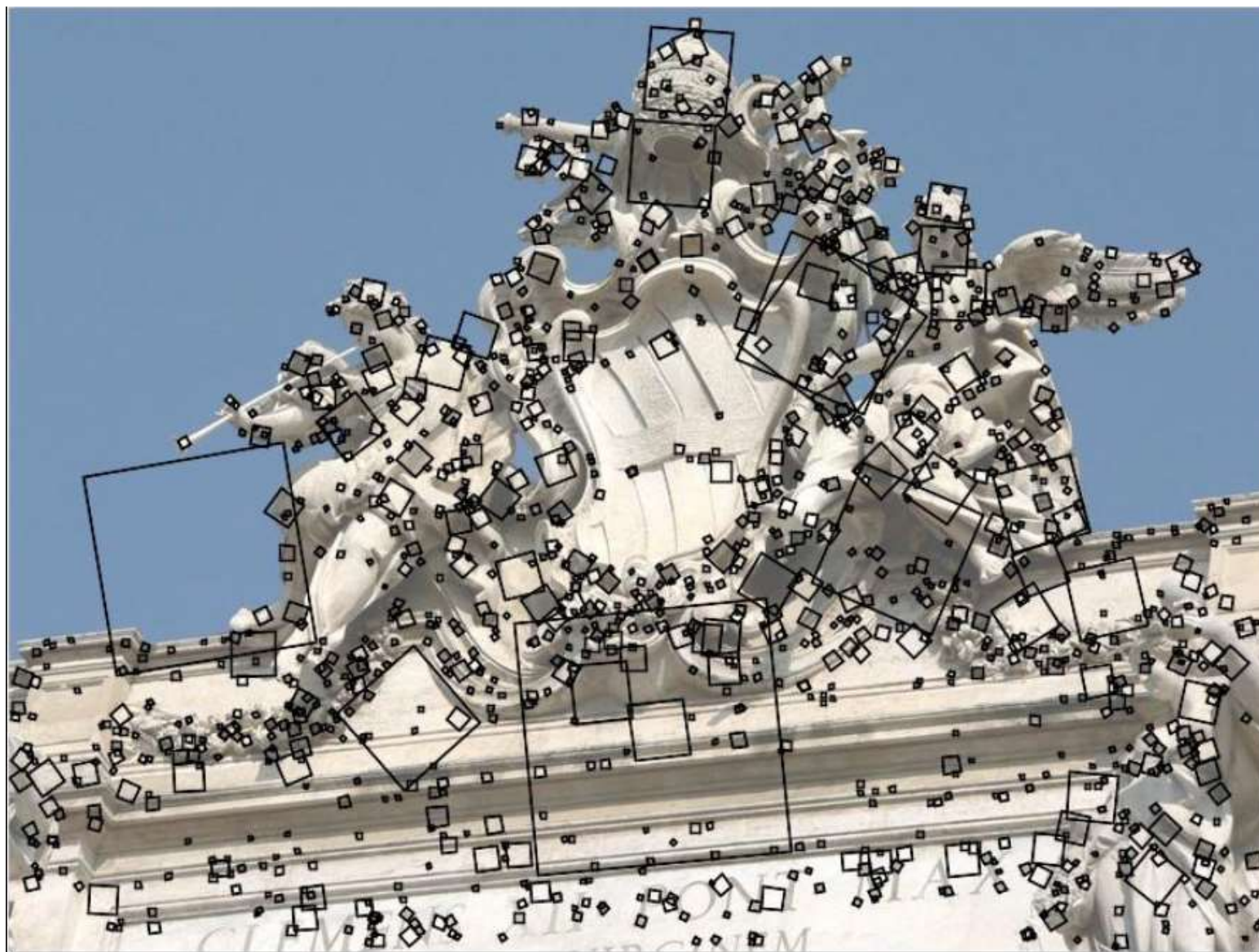
Panasonic Lumix DMC-FZ20 digital camera specifications	
	
Image	
More information	DSLR system Sample gallery Announced 21 Jul 04 All Panasonic products All Panasonic cameras
Discussion	Panasonic Talk Forum Find related discussion
Owners opinions	Read owners reviews (301) Post / edit your opinion
Format	Powerlite, 8.8µm
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, 960 x 720
Image ratio w/h	4:3, 16:9
Effective pixels	5.0 million
Sensor photo detectors	5.0 million
Sensor size	1/2.5" (5.75 x 4.21 mm, 0.24 cm²)
Pixel density	80 Megapixels
Sensor type	CCD

Матрица из спецификации камер

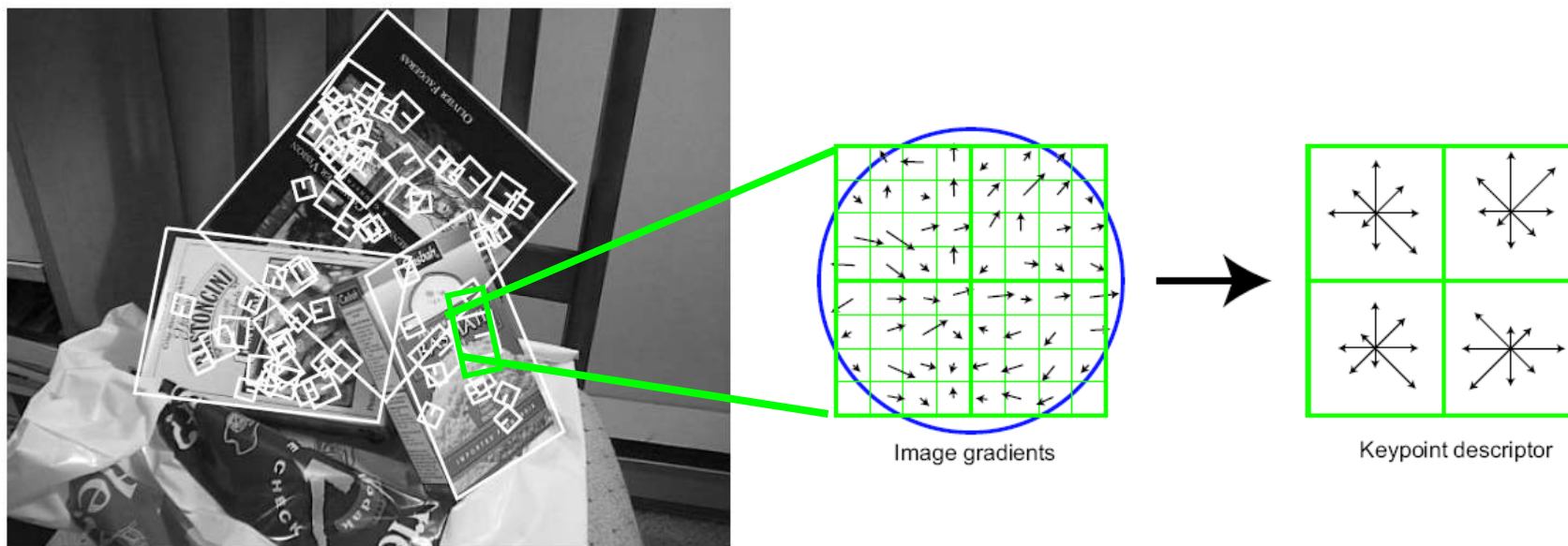
Найдем соответствующие точки



Найденные особые точки



Локальные дескрипторы: SIFT Descriptor



Histogram of oriented gradients (HoG)

- Описываем текстуру в окрестности
- Устойчивы к небольшим смещениям и аффинным преобразованиям

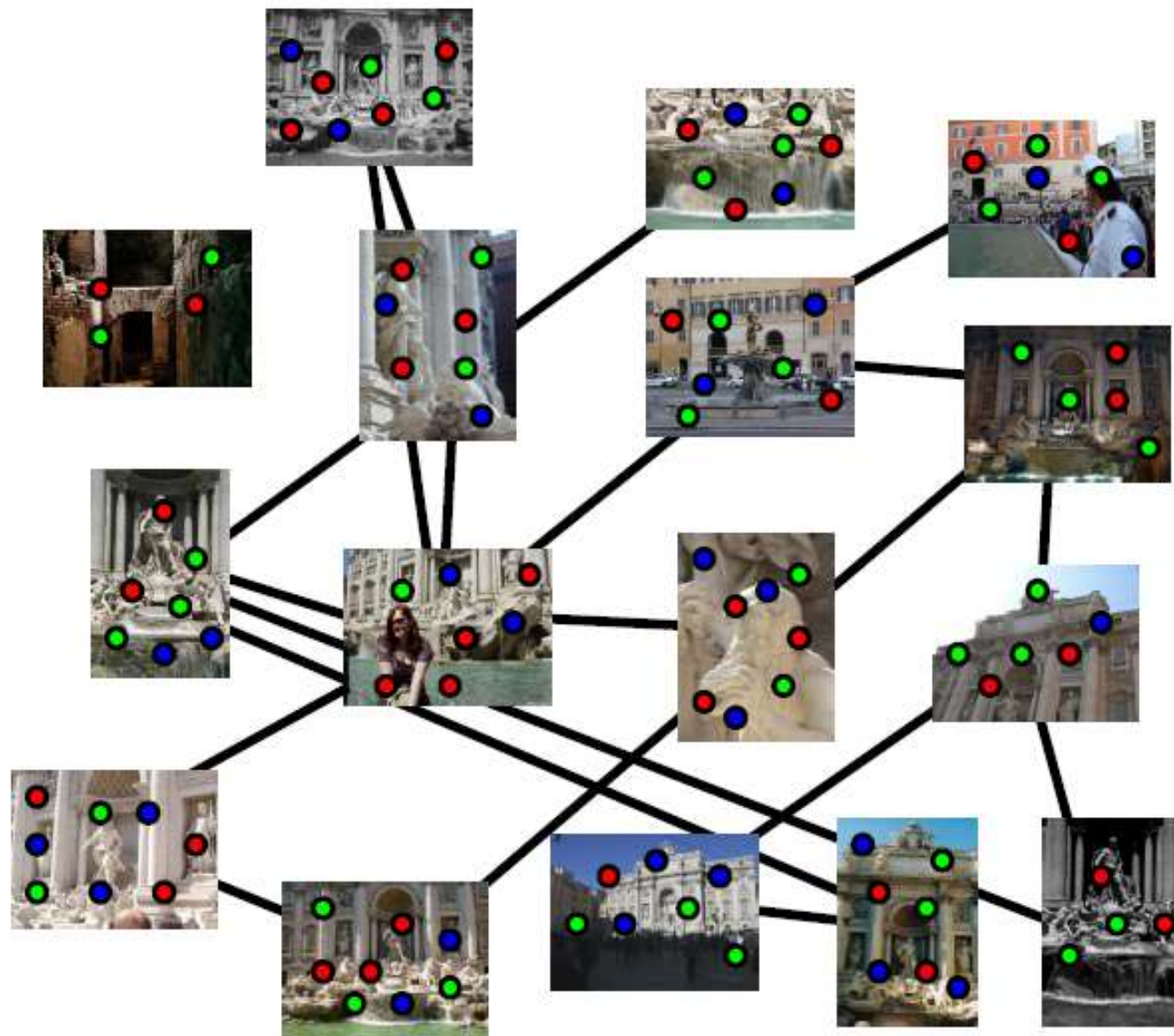
[Lowe, ICCV 1999]



Сопоставление изображений

- Сопоставляем точки по дескрипторам
- Фильтруем ложные соответствия по эпиполярной геометрии
- Для этого будем с помощью RANSAC вычислять фундаментальную матрицу
- Пары, между которыми много соответствий, будем считать «связанными»

Граф связности изображений



Объем задачи

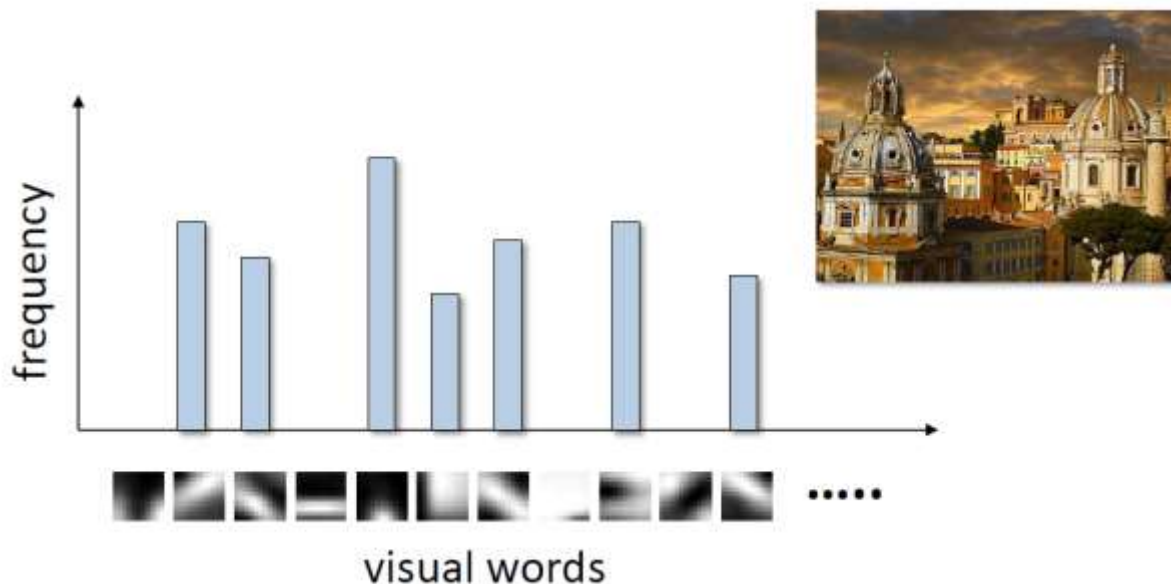


- 250K изображений -> 31M пар изображений
- 2 пары изображений в секунду
 - 1 год на 500 машинах
- 1M изображений -> 500 000M пар изображений
 - 15 лет на 500 машинах

Поиск похожих изображений



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



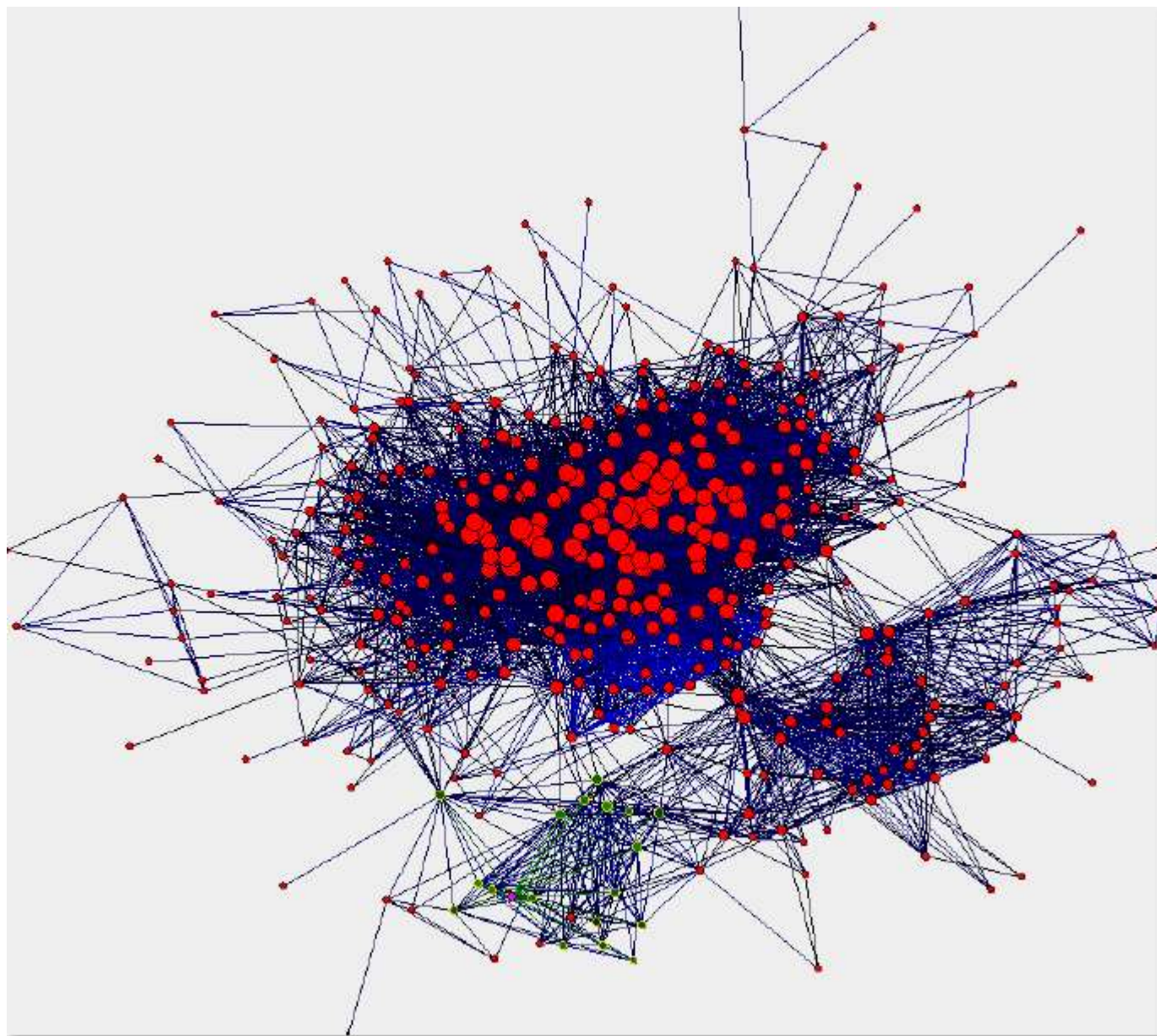
- Для каждого изображения ищем 40 наиболее похожих
- Сопоставляем точки между этими парами изображений
- Выбираем хорошие пары

Результат сопоставлений

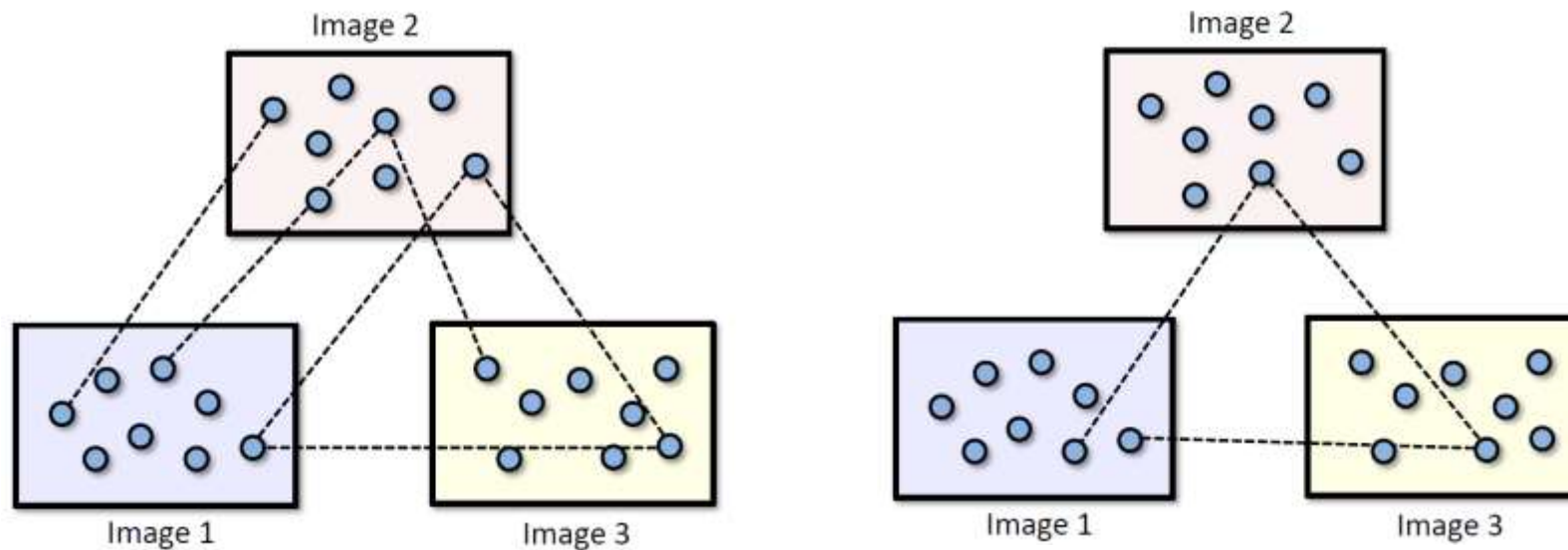


Мощность дескрипторов SIFT!

Граф связности изображений

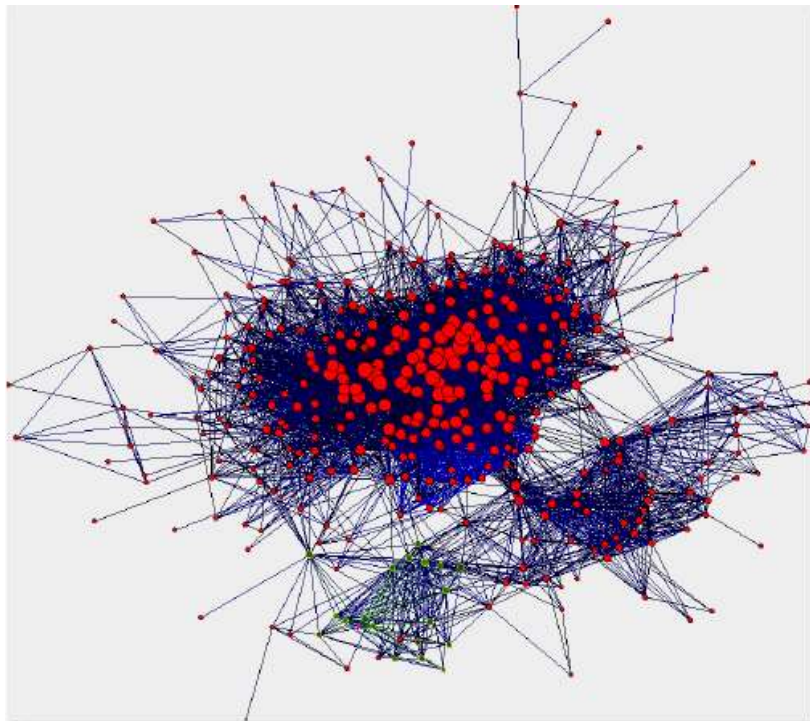


Построение «следов»

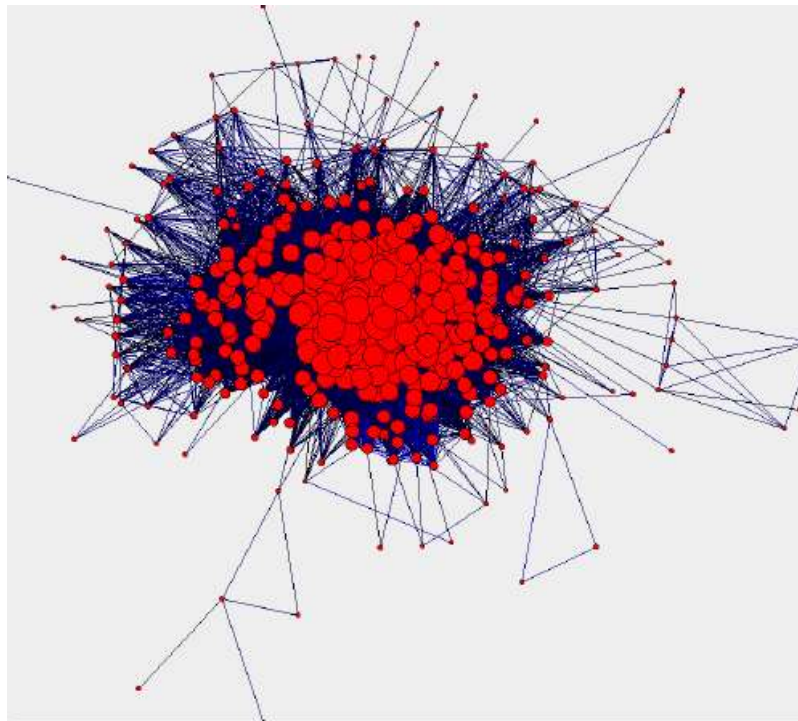


- Объединяем соответствия в «следы»
 - Отбрасываем незамкнутые следы, как потенциальные ошибки
- Пример (3000 изображений), всего 1.5M+ следов
 - 79% длины 2
 - 90% длины ≤ 3
 - 98% длины ≤ 10
 - Самый длинный след 385 точек

Граф связности изображений



Исходные соответствия

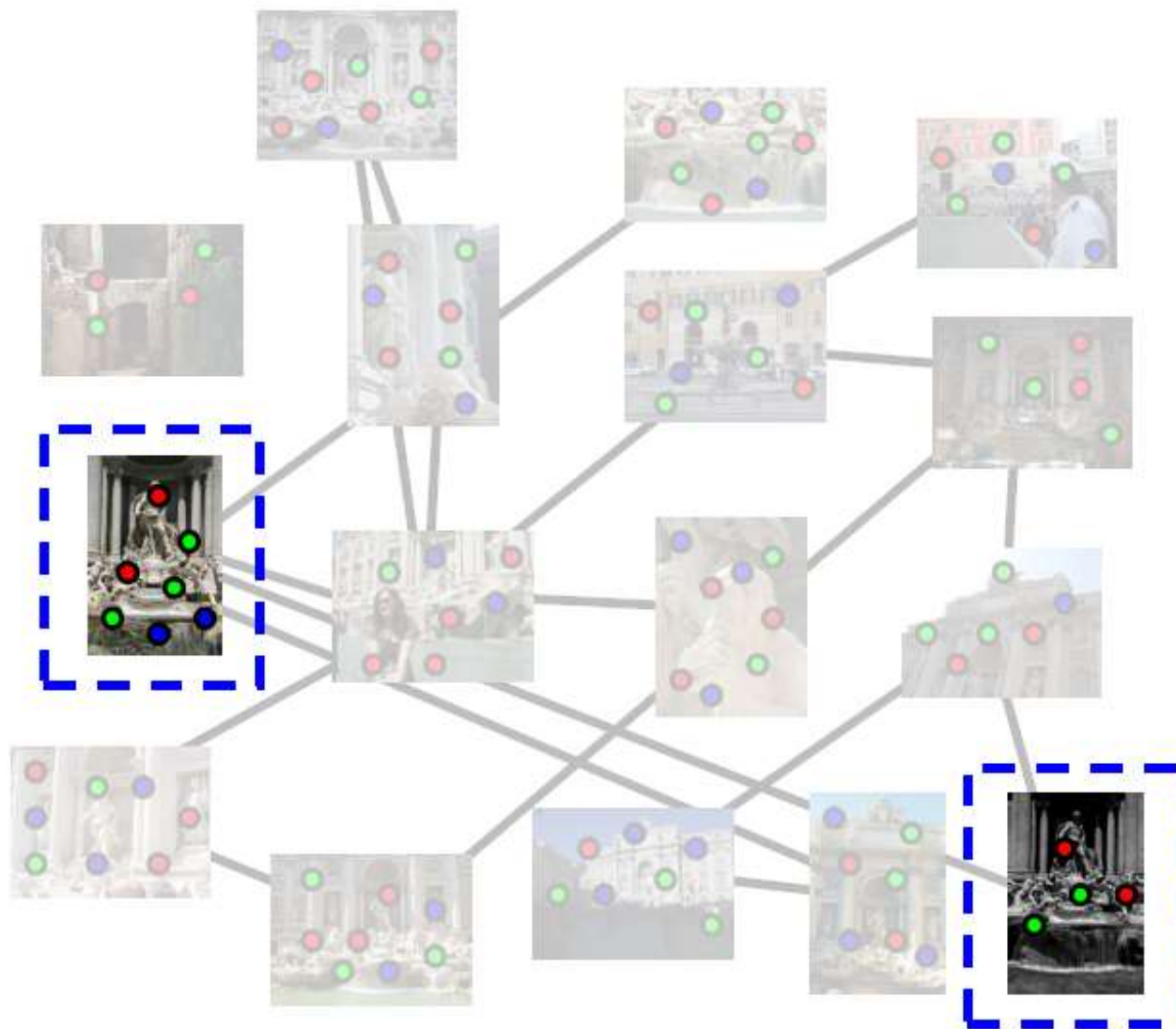


После построения следов



- 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:

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

Примеры реконструкций



Dubrovnik, Croatia.

images downloaded: 57,845

Size of largest connected component: 4,619

points: ~2M

observations: ~11.3M

Time: 18 hours



Дубровник



Примеры реконструкций





BigSFM



BigSFM: Reconstructing the World from Internet Photos

Our group is working on building a 3D model of the world from online photo collections, and our research spans several areas, from image features, to large-scale image matching, to structure-from-motion optimization, to applications such as location recognition. This page summarizes our work, has links to [code](#) and [datasets](#) we have made available, and has a description of each [project](#).

Cornell Participants (and alumni): [Noah Snavely](#) (Cornell), [David Crandall](#) (Indiana University), [Daniel Hauagge](#) (Cornell), [Kyle Wilson](#) (Cornell), [Song Cao](#) (Cornell), [Yin Lou](#) (Cornell), [Yunpeng Li](#) (EPFL), [Andrew Owens](#) (MIT), [Johannes Gehrke](#) (Cornell) [Dan Huttenlocher](#) (Cornell)

Other Collaborators: [Sameer Agarwal](#) (University of Washington and Google), [Brian Curless](#) (UW), [Yasutaka Furukawa](#) (Google), [Steve Seitz](#) (UW), [Ian Simon](#) (UW), [Rick Szeliski](#) (Microsoft Research)

We gratefully acknowledge the support of the following sponsors:



[National Science Foundation](#)



[Amazon Web Services in Education](#)



[MIT Lincoln Labs](#)

Microsoft

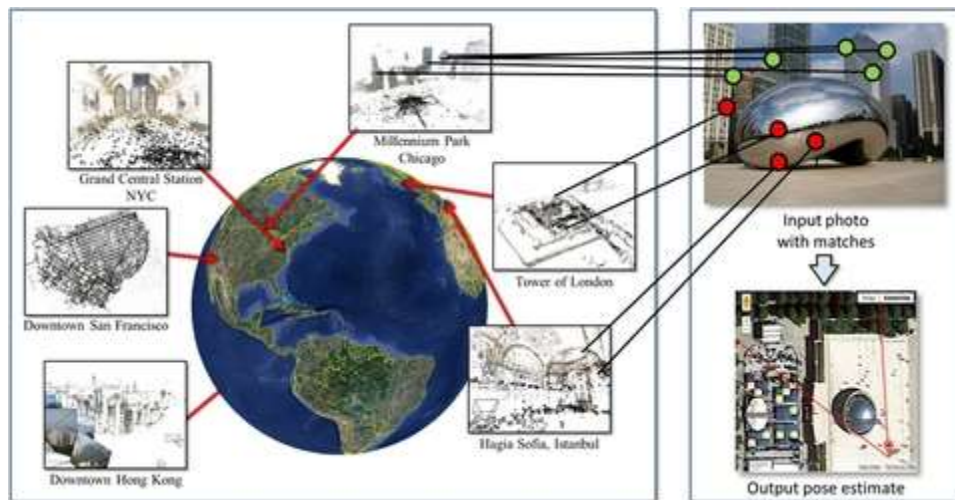
[Microsoft](#)



[IU Data to Insight Center](#)

<http://www.cs.cornell.edu/projects/bigsfm/>

Глобальное позиционирование



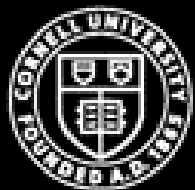
Ищем похожие фотографии +
сопоставляет по ключевым
особенностям с 3д точками





Scene Chronology

Kevin Matzen and Noah Snavely
Cornell University




ECCV 2014



COLMAP



 COLMAP
3.6

Search docs

[Installation](#)
[Tutorial](#)
[Database Format](#)
[Camera Models](#)
[Output Format](#)
[Datasets](#)
[Graphical User Interface](#)
[Command-line Interface](#)
[Frequently Asked Questions](#)
[Changelog](#)
[Contribution](#)
[License](#)
[Bibliography](#)

Docs » COLMAP

[View page source](#)

COLMAP



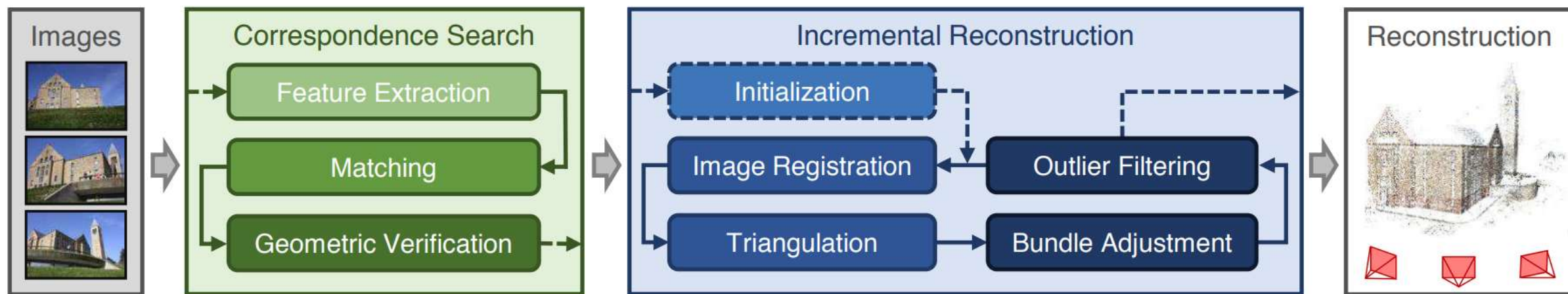
Sparse model of central Rome using 21K photos produced by COLMAP's SfM pipeline.



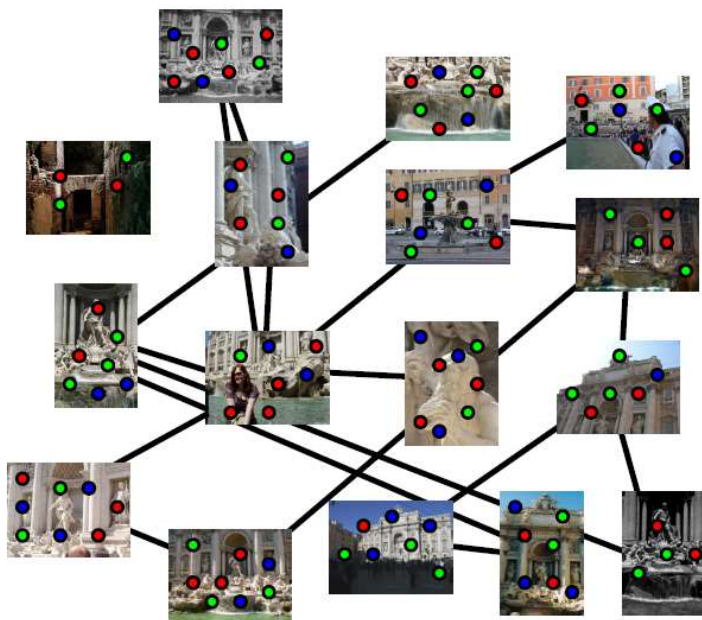
Dense models of several landmarks produced by COLMAP's MVS pipeline.

<https://colmap.github.io/>

Улучшение графа сцены

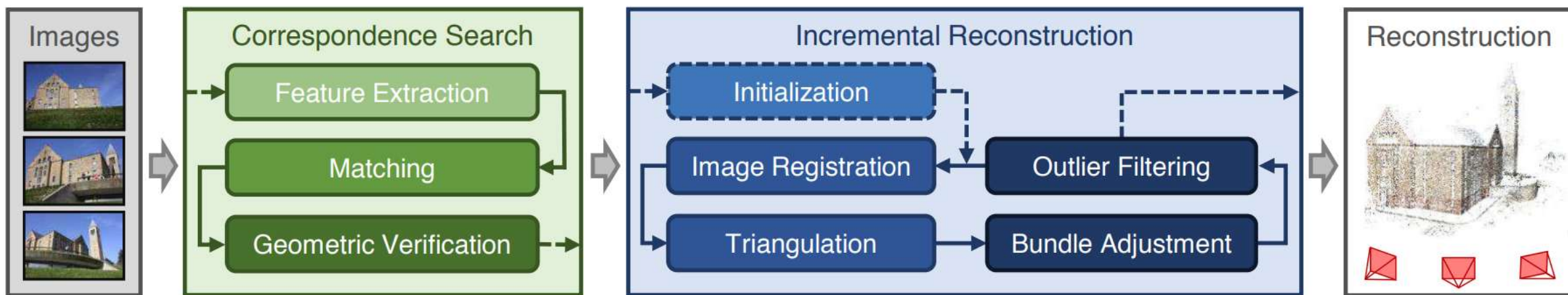


Source: [Structure-From-Motion Revisited](#). CVPR2016



- Для калиброванных пар изображений считаем также существенную матрицу, движение камеры, триангулируем точки и a_m - медианный угол
- Если пара кадров хорошо описывается гомографией, то это позволит нам различать чистый поворот (панораму) и плоскую сцену
- Для точек на границе считаем преобразование подобия, и если доля таких точек, удовлетворяющих F или E высока, то вероятно это точки водяных знаков, поэтому такие кадры мы выбрасываем

Выбор следующего лучшего кадра



Source: [Structure-From-Motion Revisited](#). CVPR2016

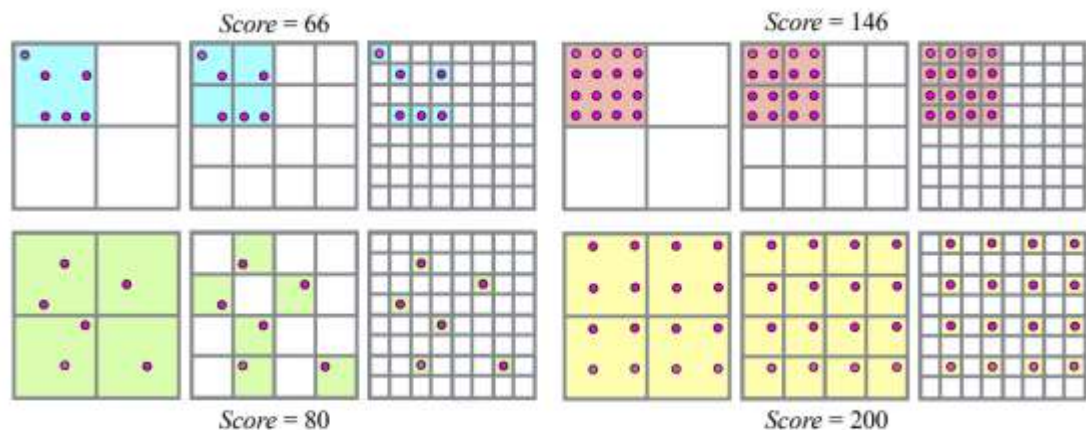
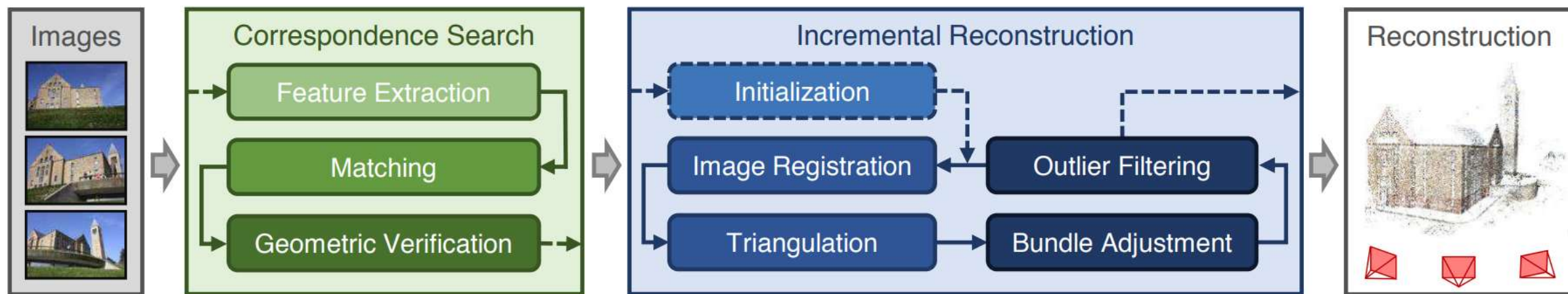


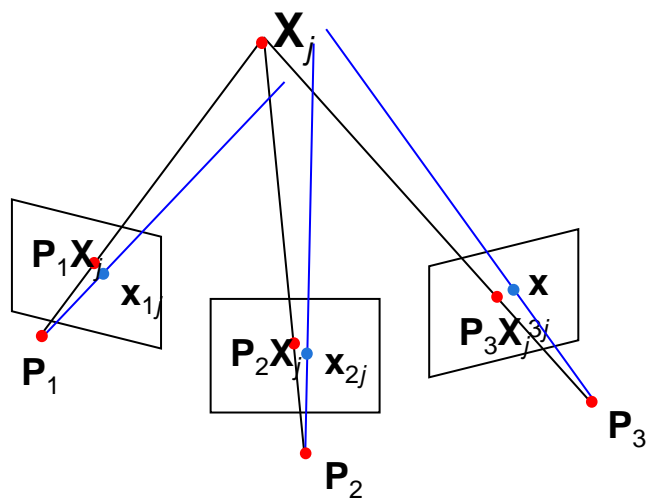
Figure 3. Scores for different number of points (left and right) with different distributions (top and bottom) in the image for $L = 3$.

- Выбор следующего кадра для итеративного метода особенно важен, т.к. ошибки могут накапливаться
- Рассматриваем кадры, у которых $> N_T$ триангулированных точек
- Предложена процедура оценки качества распределения точек
- Разбиваем изображение сеткой, считаем число заполненных клеток, каждой назначая вес пропорционально площади

Робастная триангуляция

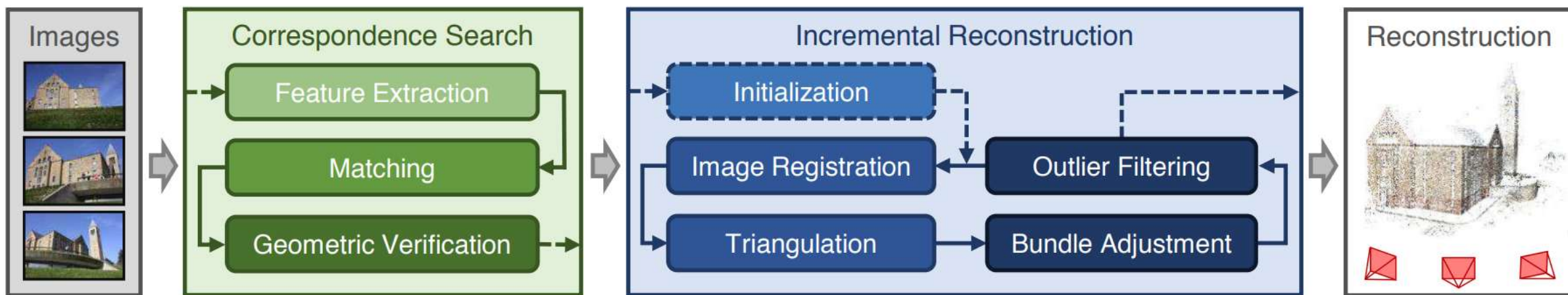


Source: [Structure-From-Motion Revisited](#). CVPR2016

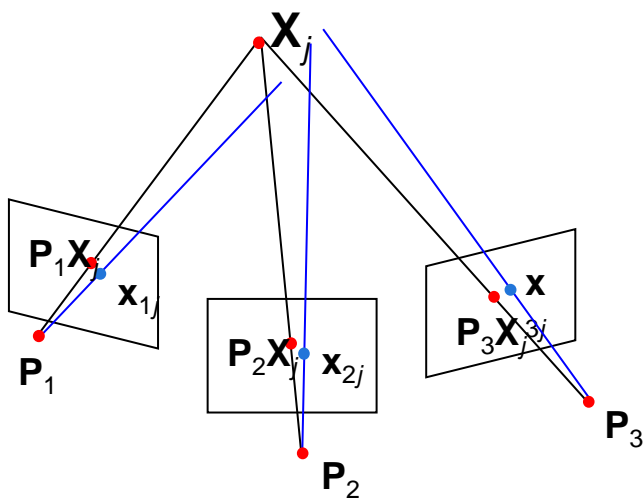


- Предполагаем, что в следах могут быть ошибки, поэтому применяем метод RANSAC для многокадровой триангуляции
- Сэмплируем только уникальные пары точек
- Изначально предполагаем малую долю истинных соответствий, пересчитываем оценку в процессе RANSAC

Bundle Adjustment

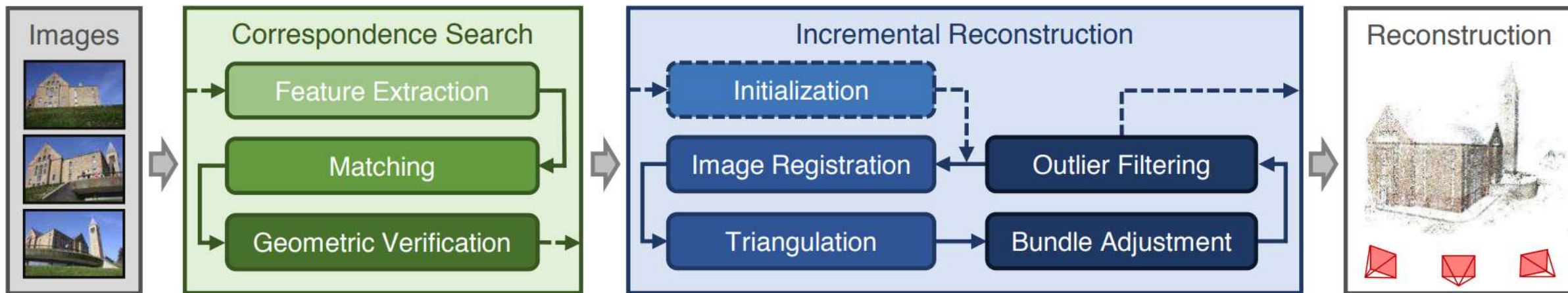


Source: [Structure-From-Motion Revisited](#). CVPR2016

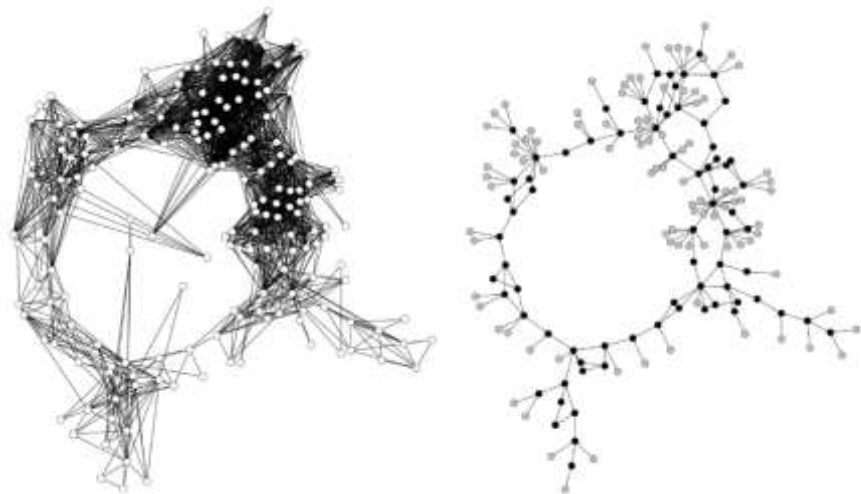


- Применяем Local BA на маленьких группах тесно связанных кадров после добавления каждого кадра
- Применяем Global BA только при увеличении 3D модели на определённую величину
- Итеративная схема с фильтрацией:
 - Перед BA повторно триангулируем точки и фильтруем ложные (чтобы не портили BA)
 - После BA снова триангулируем, чтобы дополнить модель точками, которые могли плохо оцениваться из-за ошибок в позах и т.д., и отфильтровать ложные.
 - Повторяем BA, поскольку в прошлую итерацию могли использоваться ложные и исказить всё

Redundant View Mining



Source: [Structure-From-Motion Revisited](#). CVPR2016



- Можем «схлопывать» очень близкие ракурсы в один
- Это существенно ускоряет работу, в первую очередь ВА без потери точности

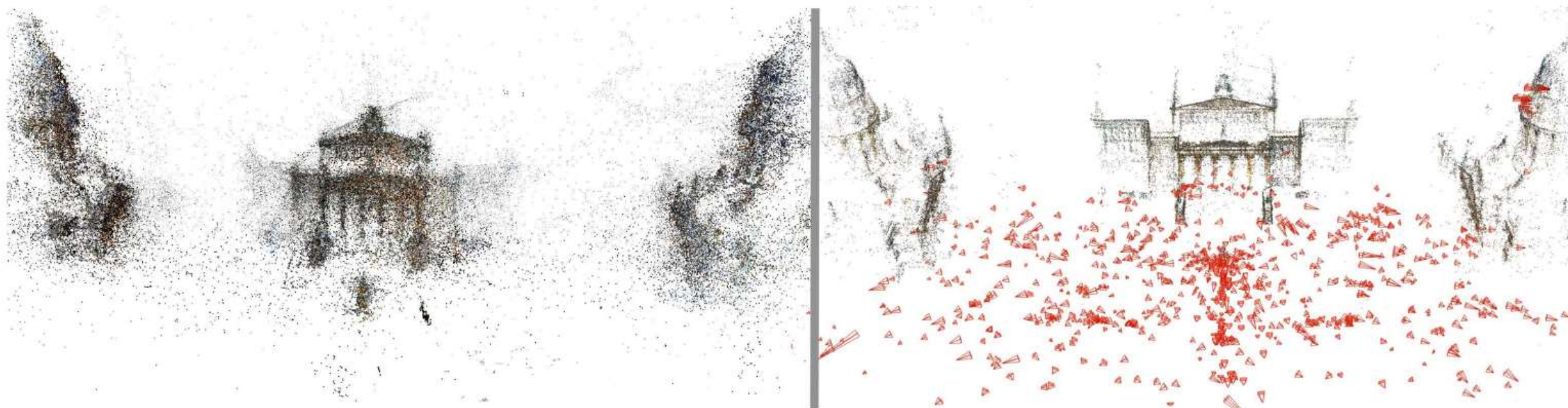


Figure 9. Reconstruction of Gendarmenmarkt [61] for *Bundler* (left) and our method (right).

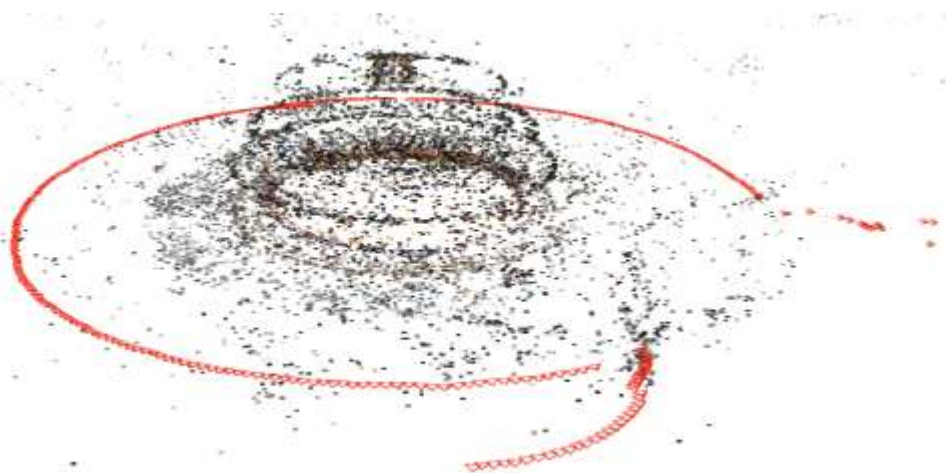
Сравнение с аналогами



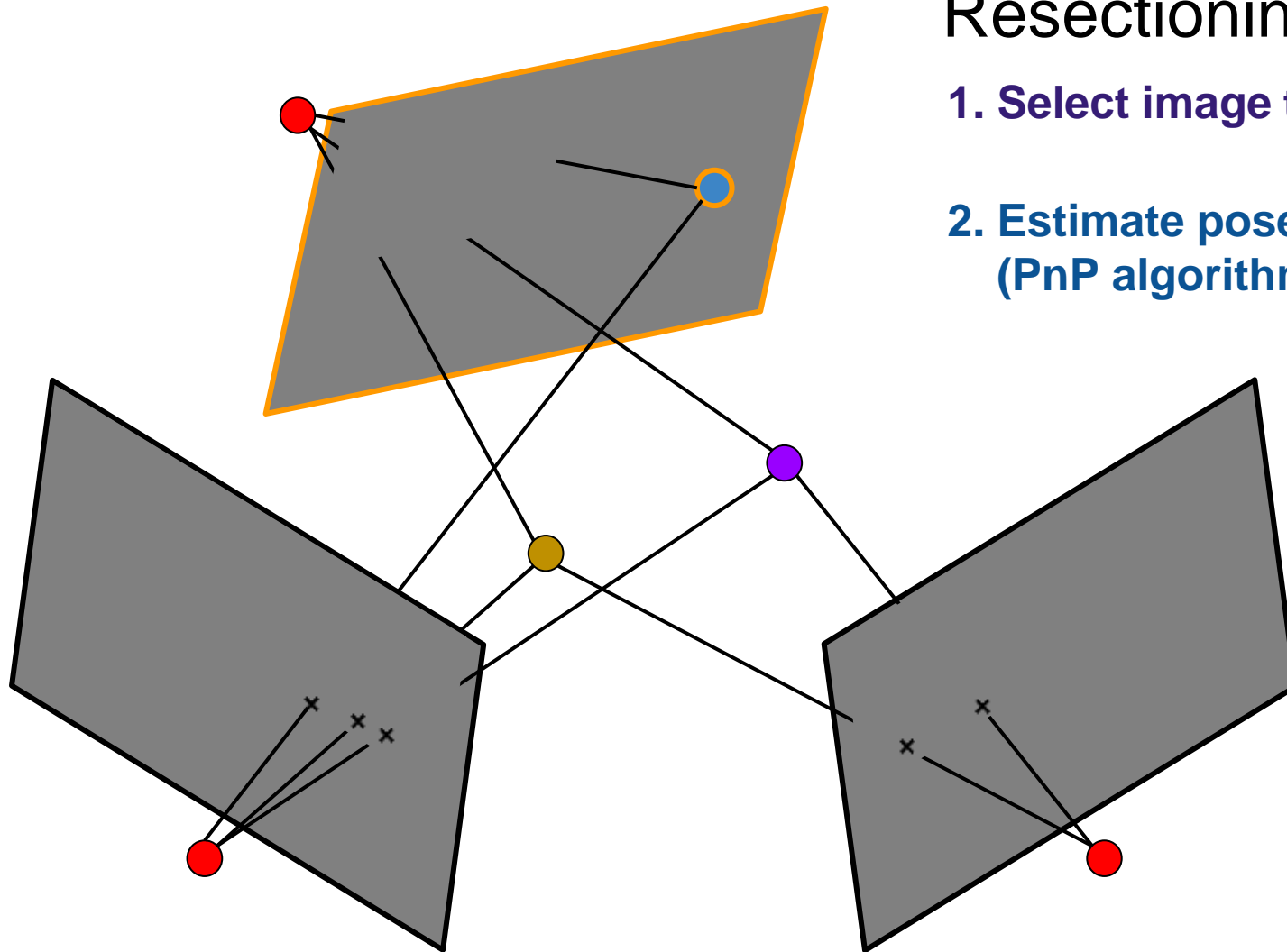
	# Images	# Registered				# Points (Avg. Track Length)				Time [s]				Avg. Reproj. Error [px]			
		<i>Theia</i>	<i>Bundler</i>	<i>VSFM</i>	<i>Ours</i>	<i>Theia</i>	<i>Bundler</i>	<i>VSFM</i>	<i>Ours</i>	<i>Theia</i>	<i>Bundler</i>	<i>VSFM</i>	<i>Ours</i>	<i>Theia</i>	<i>Bundler</i>	<i>VSFM</i>	<i>Ours</i>
Rome [14]	74,394	–	13,455	14,797	20,918	–	5.4M	12.9M	5.3M	–	295,200	6,012	10,912	–	–	–	–
Quad [14]	6,514	–	5,028	5,624	5,860	–	10.5M	0.8M	1.2M	–	223,200	2,124	3,791	–	–	–	–
Dubrovnik [36]	6,044	–	–	–	5,913	–	–	–	1.35M	–	–	–	3,821	–	–	–	–
Alamo [61]	2,915	582	647	609	666	146K (6.0)	127K (4.5)	124K (8.9)	94K (11.6)	874	22,025	495	882	1.47	2.29	0.70	0.68
Ellis Island [61]	2,587	231	286	297	315	29K (4.9)	39K (4.1)	61K (5.5)	64K (6.8)	94	12,798	240	332	2.41	2.24	0.71	0.70
Gendarmenmarkt [61]	1,463	703	302	807	861	87K (3.8)	93K (3.7)	138K (4.9)	123K (6.1)	202	465,213	412	627	2.19	1.59	0.71	0.68
Madrid Metropolis [61]	1,344	351	330	309	368	47K (5.0)	27K (3.2)	48K (5.2)	43K (6.6)	95	21,633	203	251	1.48	1.62	0.59	0.60
Montreal Notre Dame [61]	2,298	464	501	491	506	154K (5.4)	135K (4.6)	110K (7.1)	98K (8.7)	207	112,171	418	723	2.01	1.92	0.88	0.81
NYC Library [61]	2,550	339	400	411	453	66K (4.1)	71K (3.7)	95K (5.5)	77K (7.1)	194	36,462	327	420	1.89	1.84	0.67	0.69
Piazza del Popolo [61]	2,251	335	376	403	437	36K (5.2)	34K (3.7)	50K (7.2)	47K (8.8)	89	33,805	275	380	2.11	1.76	0.76	0.72
Piccadilly [61]	7,351	2,270	1,087	2,161	2,336	197K (4.9)	197K (3.9)	245K (6.9)	260K (7.9)	1,427	478,956	1,236	1,961	2.33	1.79	0.79	0.75
Roman Forum [61]	2,364	1,074	885	1,320	1,409	261K (4.9)	281K (4.4)	278K (5.7)	222K (7.8)	1,302	587,451	748	1,041	2.07	1.66	0.69	0.70
Tower of London [61]	1,576	468	569	547	578	140K (5.2)	151K (4.8)	143K (5.7)	109K (7.4)	201	184,905	497	678	1.86	1.54	0.59	0.61
Trafalgar [61]	15,685	5,067	1,257	5,087	5,211	381K (4.8)	196K (3.7)	497K (8.7)	450K (10.1)	1,494	612,452	3,921	5,122	2.09	2.07	0.79	0.74
Union Square [61]	5,961	720	649	658	763	35K (5.3)	48K (3.7)	43K (7.1)	53K (8.2)	131	56,317	556	693	2.36	3.22	0.68	0.67
Vienna Cathedral [61]	6,288	858	853	890	933	259K (4.9)	276K (4.6)	231K (7.6)	190K (9.8)	764	567,213	899	1,244	2.45	1.69	0.80	0.74
Yorkminster [61]	3,368	429	379	427	456	143K (4.5)	71K (3.9)	130K (5.2)	105K (6.8)	164	34,641	661	997	2.38	2.61	0.72	0.70

Table 1. Reconstruction results for state-of-the-art SfM systems on large-scale unordered Internet photo collections.

Проблема повторяющихся структур



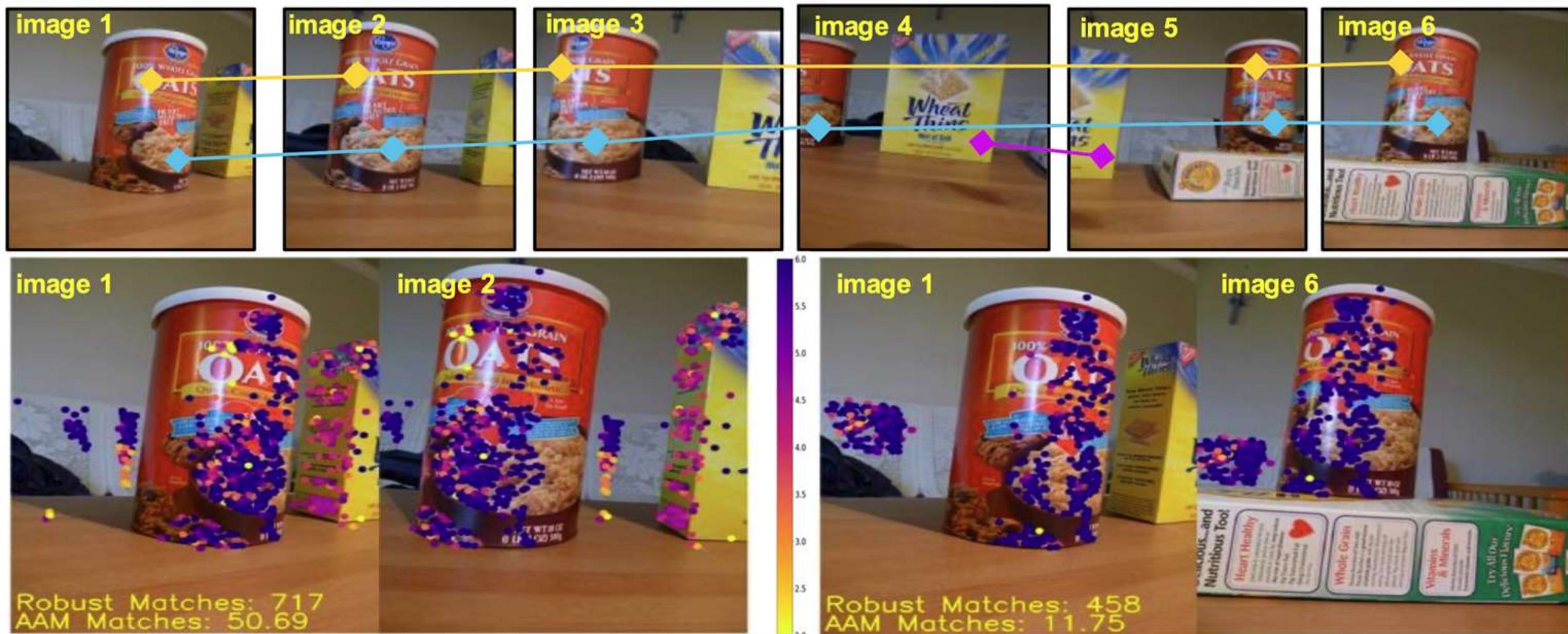
Resectioning – добавление нового кадра



Resectioning is a critical step

1. Select image that views the most triangulated points
2. Estimate pose of image using all the triangulated points (PnP algorithm using RANSAC)

Пенализация длинных следов

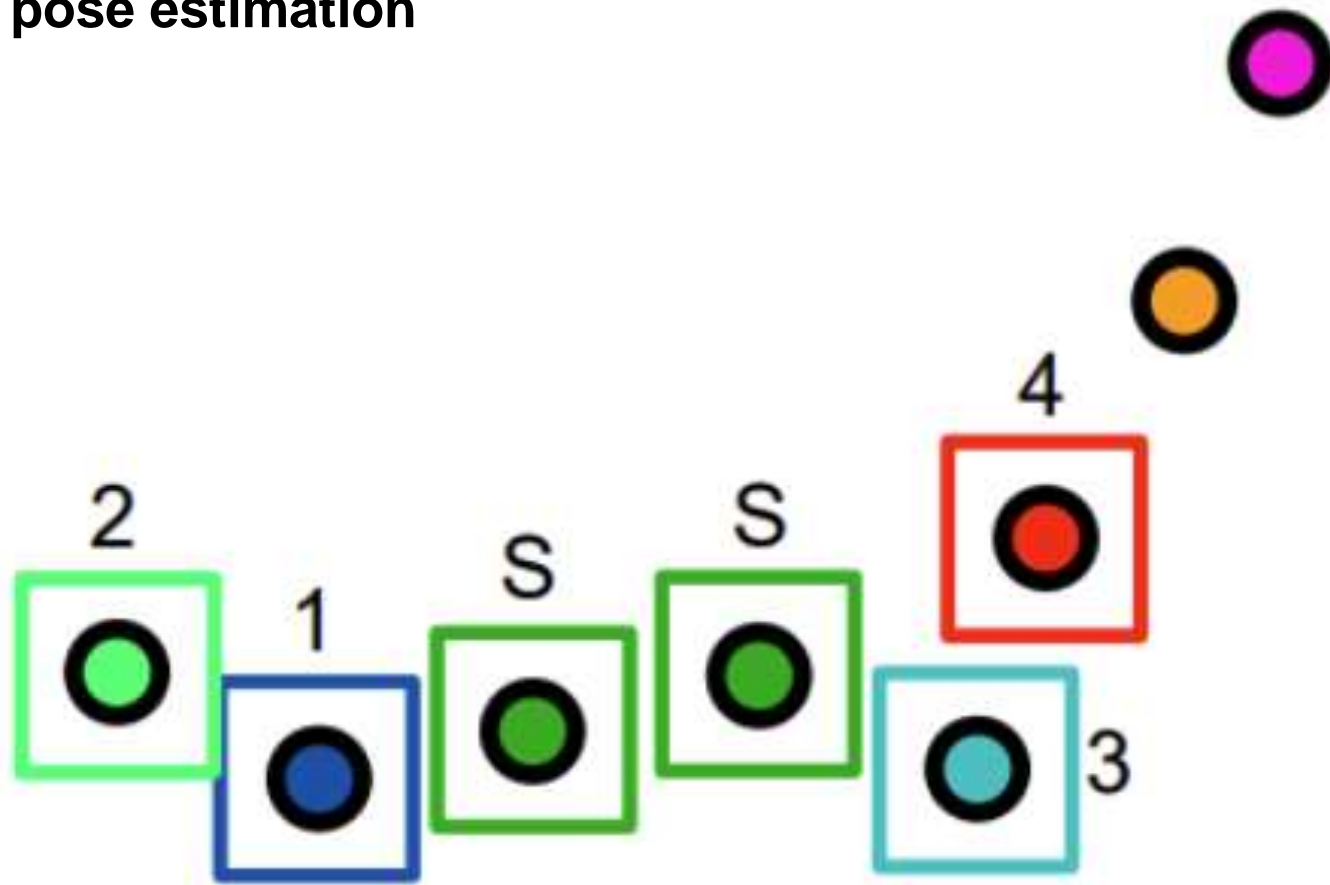


Ambiguity-adjusted match score (AAM): Discount longer tracks that are more likely to correspond to duplicate structures

Сортировка кадров по AAM score



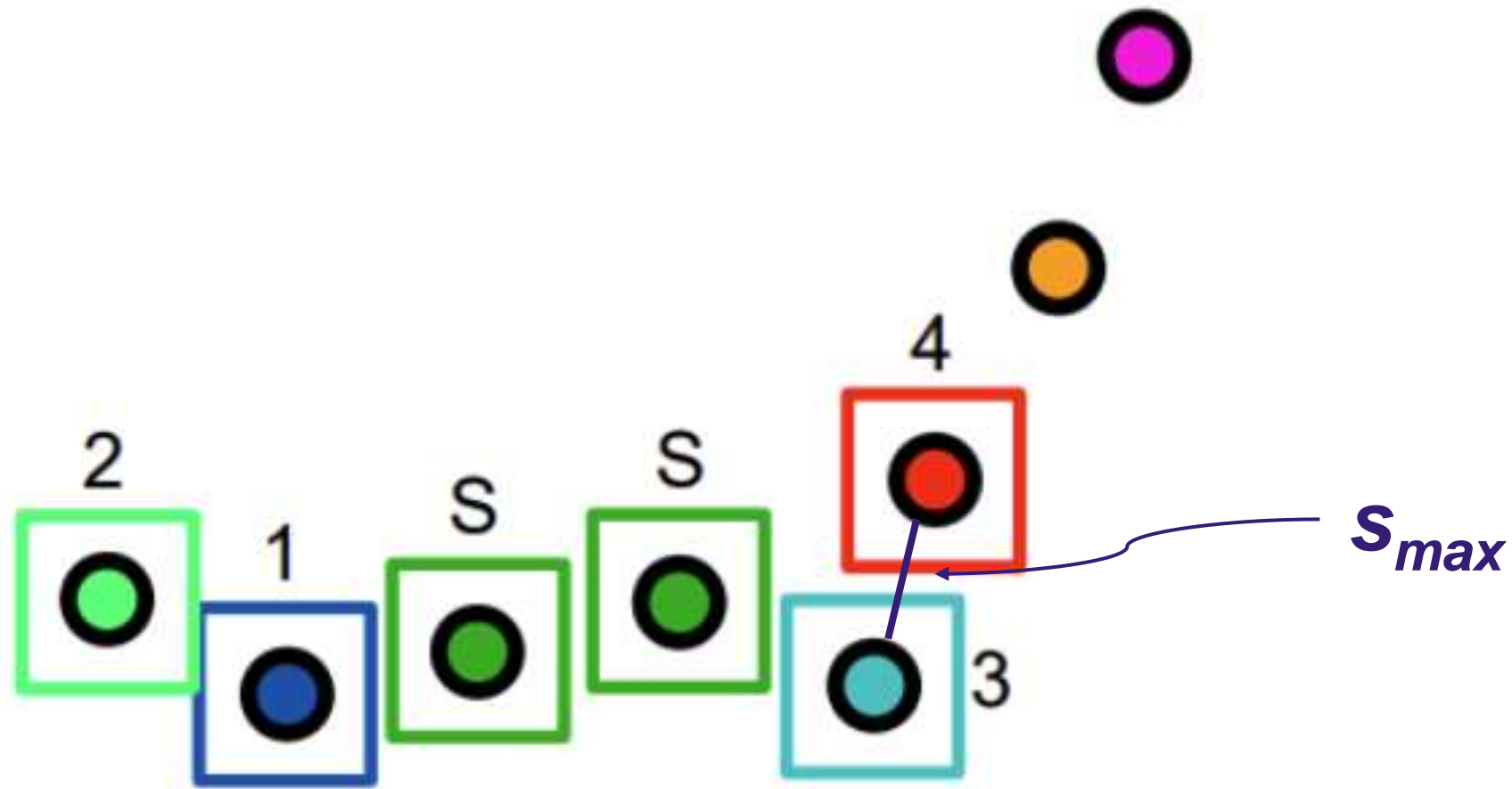
- Local resectioning order uses most similar image
- We use points from a smaller set of reliable images to determine **resectioning order** and **pose estimation**



Подход к ресекции



- Local pose estimation uses reliable images only
- Use points from a smaller set of reliable images to determine **resectioning order** and **pose estimation**



-



Улучшение стандартных подходов



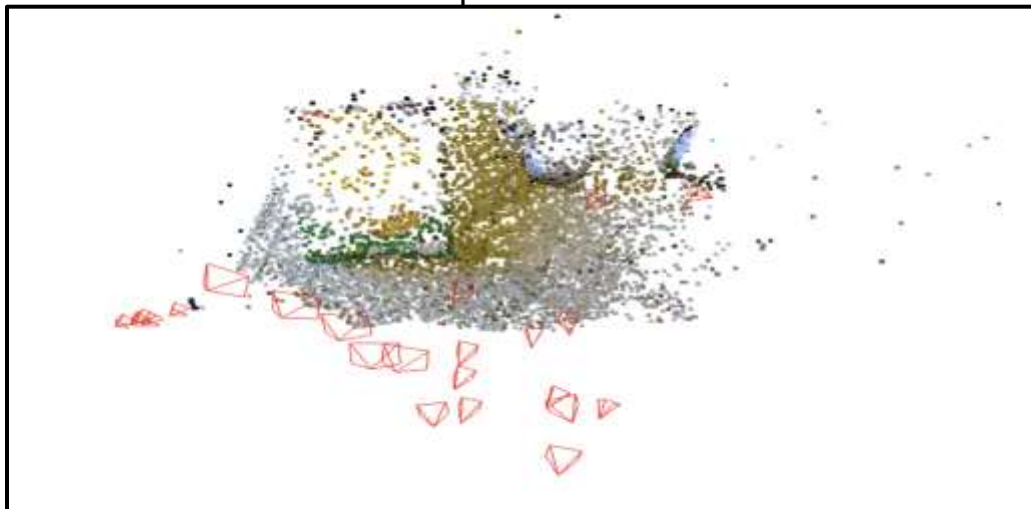
Local resectioning using ambiguity-adjusted matches compared against baselines (standard OpenSfM and COLMAP pipelines)

	OpenSfM	OpenSfM w/ Our Resectioning	COLMAP	COLMAP w/ Our Resectioning
Duplicate Structures Dataset	6 Failures	6 Successes	6 Failures	2 Failures 4 Successes
UIUCTag Dataset	9 Failures 3 Partial Successes 4 Successes	3 Failures 13 Successes	7 Failures 2 Partial Successes 7 Successes	5 Failures 3 Partial Successes 8 Successes
TanksAndTemples Dataset	4 Failures 3 Successes	1 Failure 6 Successes	1 Failure 6 Successes	7 Successes

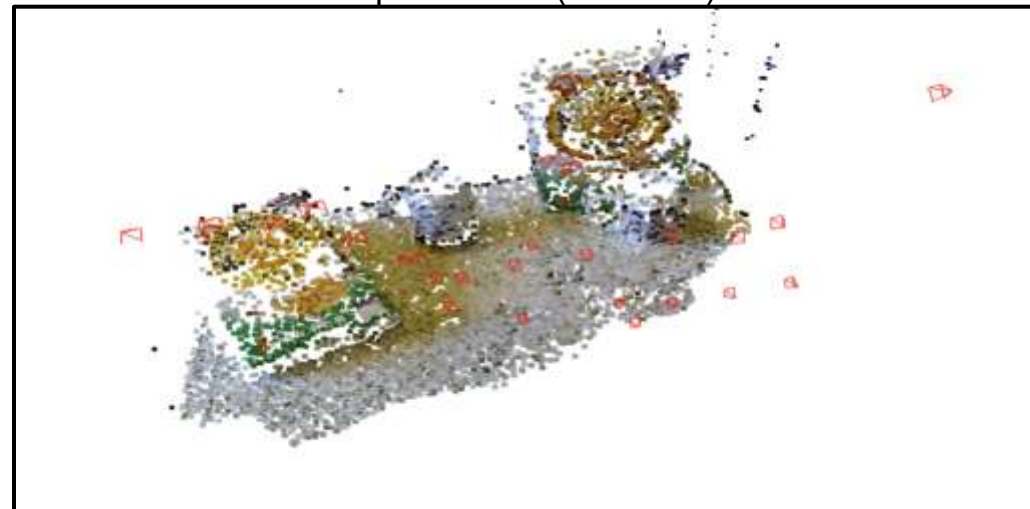
Пример - Cereal (DuplicateStructures)



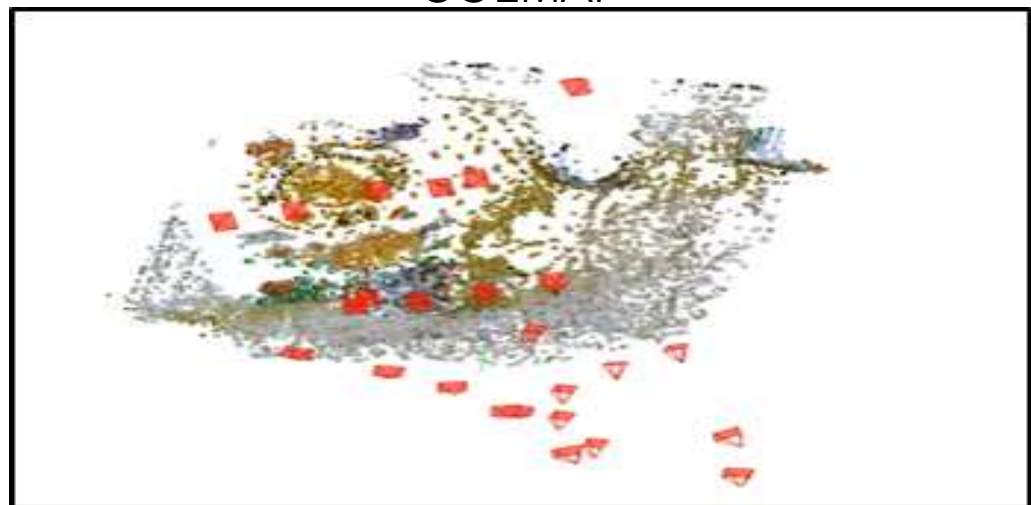
OpenSfM



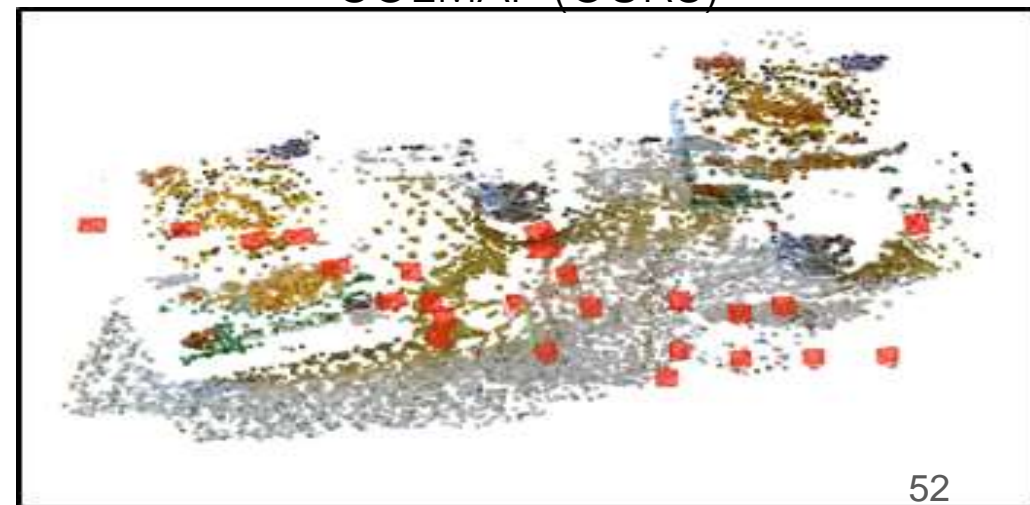
OpenSfM (OURS)



COLMAP



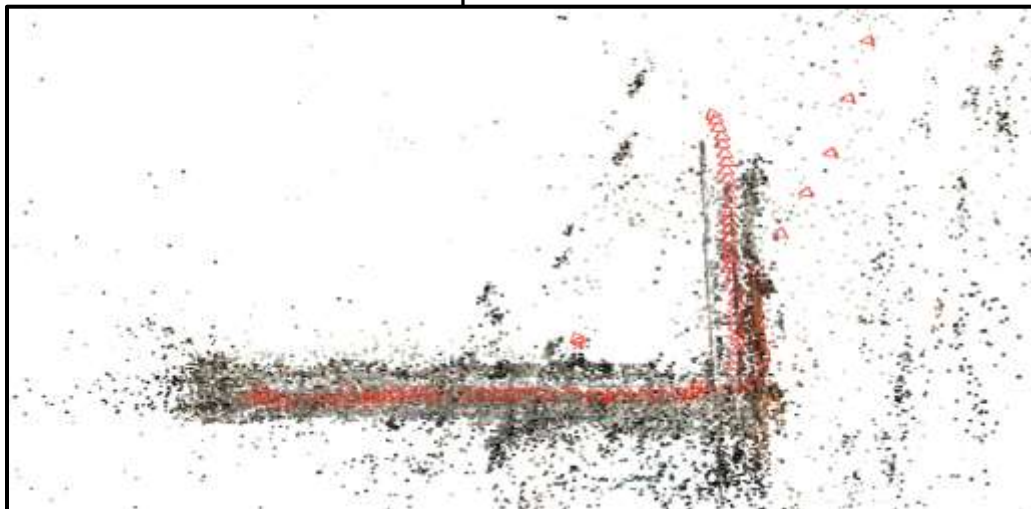
COLMAP (OURS)



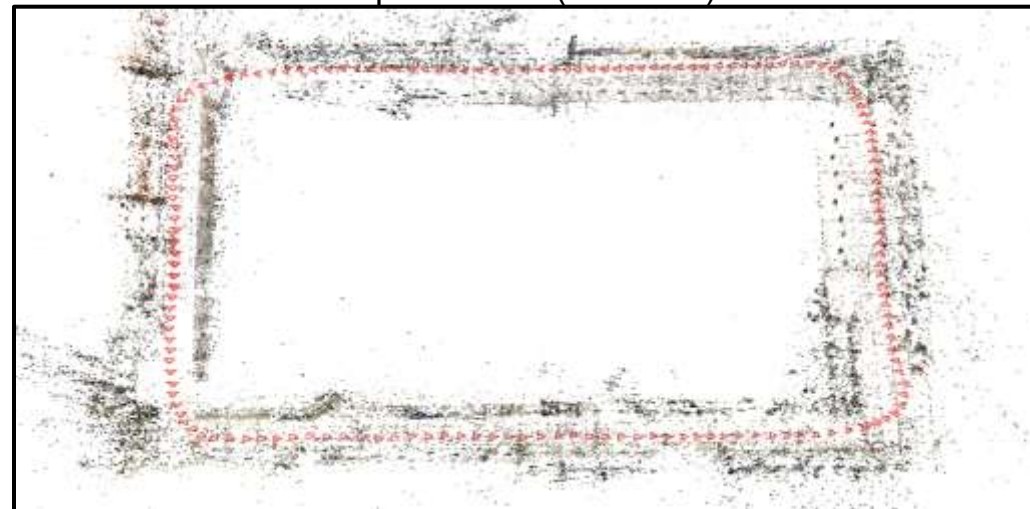
Пример - ece_floor3_loop_cw (UIUCTag)



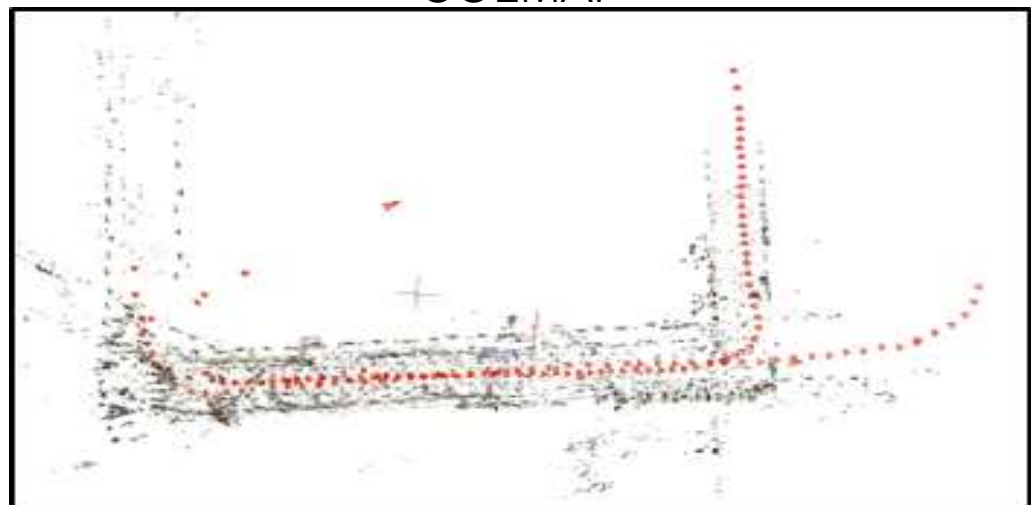
OpenSfM



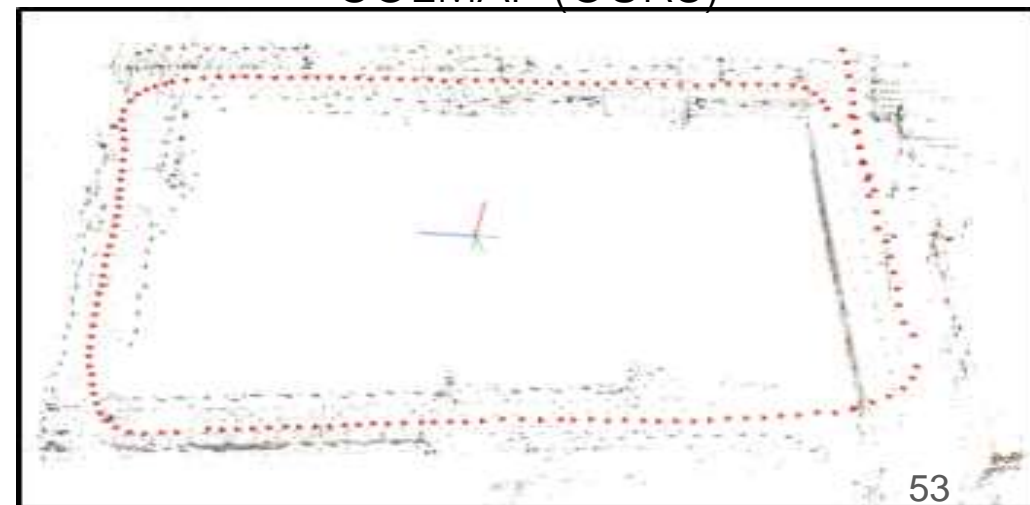
OpenSfM (OURS)



COLMAP



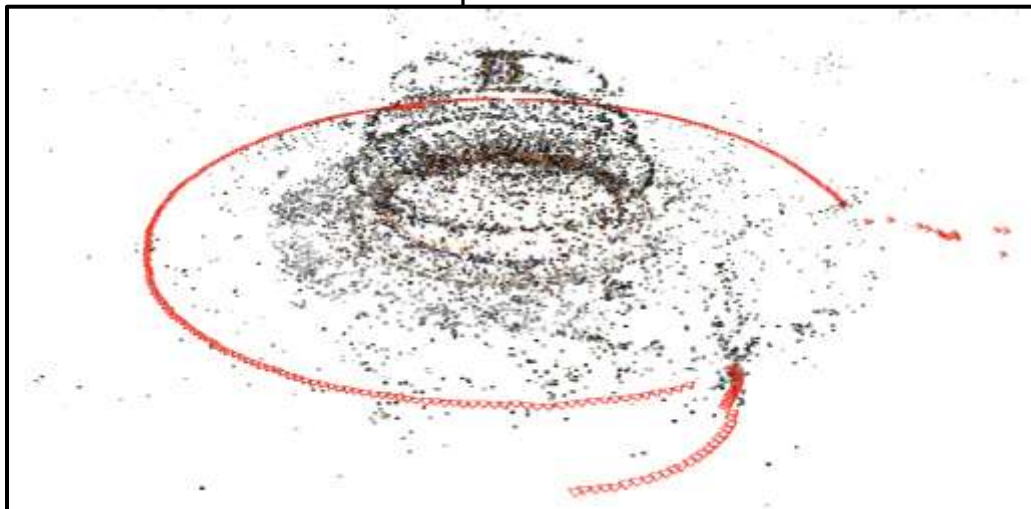
COLMAP (OURS)



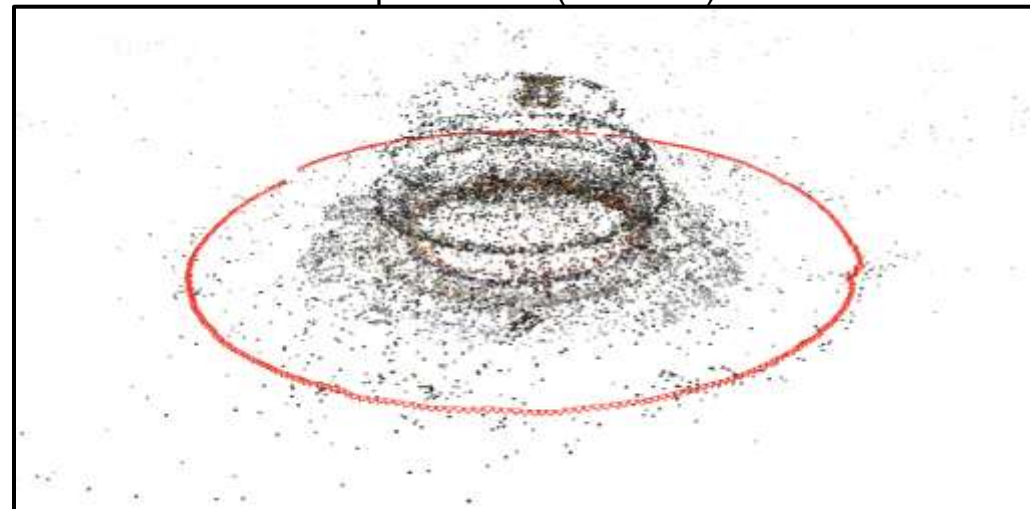
Пример - TempleOfHeaven (Internet)



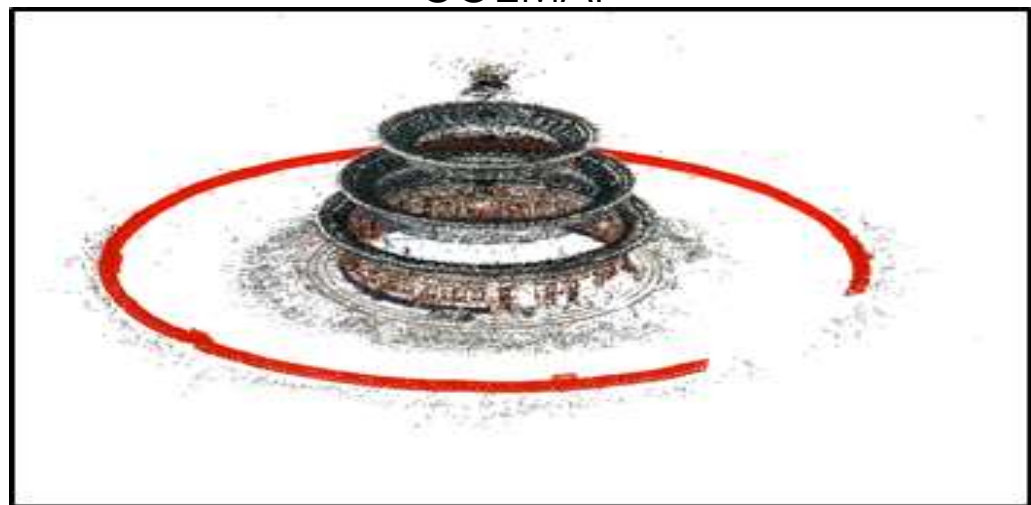
OpenSfM



OpenSfM (OURS)



COLMAP



COLMAP (OURS)

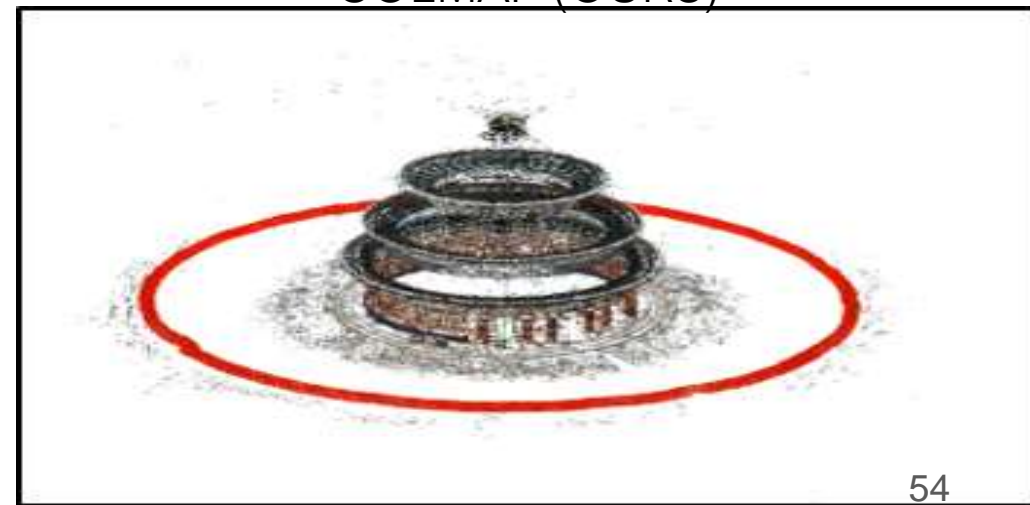




Image Matching Across Wide Baselines: From Paper to Practice

Yuhe Jin · Dmytro Mishkin · Anastasiia Mishchuk · Jiri Matas · Pascal Fua ·
Kwang Moo Yi · Eduard Trulls

Local featured type	Number of Images				
	100	200	400	800	all
SIFT [54]	0.06°	0.09°	0.06°	0.07°	0.09°
SIFT (Upright) [54]	0.07°	0.07°	0.04°	0.06°	0.09°
HardNet (Upright) [62]	0.06°	0.06°	0.06°	0.04°	0.05°
SuperPoint [34]	0.31°	0.25°	0.33°	0.19°	0.32°
R2D2 [80]	0.12°	0.08°	0.07°	0.08°	0.05°

Table 2 Standard deviation of the pose difference of three COLMAP runs with different number of images. Most of them are below 0.1° , except for SuperPoint.

Local feature type	Number of images			
	100 vs. all	200 vs. all	400 vs. all	800 vs. all
SIFT [54]	0.58° / 0.22°	0.31° / 0.08°	0.23° / 0.05°	0.18° / 0.04°
SIFT (Upright) [54]	0.52° / 0.16°	0.29° / 0.08°	0.22° / 0.05°	0.16° / 0.03°
HardNet (Upright) [62]	0.35° / 0.10°	0.33° / 0.08°	0.23° / 0.06°	0.14° / 0.04°
SuperPoint [34]	1.22° / 0.71°	1.11° / 0.67°	1.08° / 0.48°	0.74° / 0.38°
R2D2 [80]	0.49° / 0.14°	0.32° / 0.10°	0.25° / 0.08°	0.18° / 0.05°

Table 3 Pose convergence in SfM. We report the mean/median of the difference (in degrees) between the poses extracted with the full set of 1179 images for “Sacre Coeur”, and different subsets of it, for four local feature methods – to keep the results comparable we only look at the 100 images in common across all subsets. We report the maximum among the angular difference between rotation matrices and translation vectors. The estimated poses are stable, with as little as 100 images.

Можно ли улучшить за счёт DL-точек?



Reference	Compared to			
	SIFT (Upright)	HardNet (Upright)	SuperPoint	R2D2
SIFT [54]	0.20° / 0.05°	0.26° / 0.05°	1.01° / 0.62°	0.26° / 0.09°

Table 4 Difference between poses obtained with different local features. We report the mean/median of the difference (in degrees) between the poses extracted with SIFT (Upright), HardNet (Upright), SuperPoint, or R2D2, and those extracted with SIFT. We use the maximum of the angular difference between rotation matrices and translation vectors. SIFT (Upright), HardNet (Upright), and R2D2 give near-identical results to SIFT.

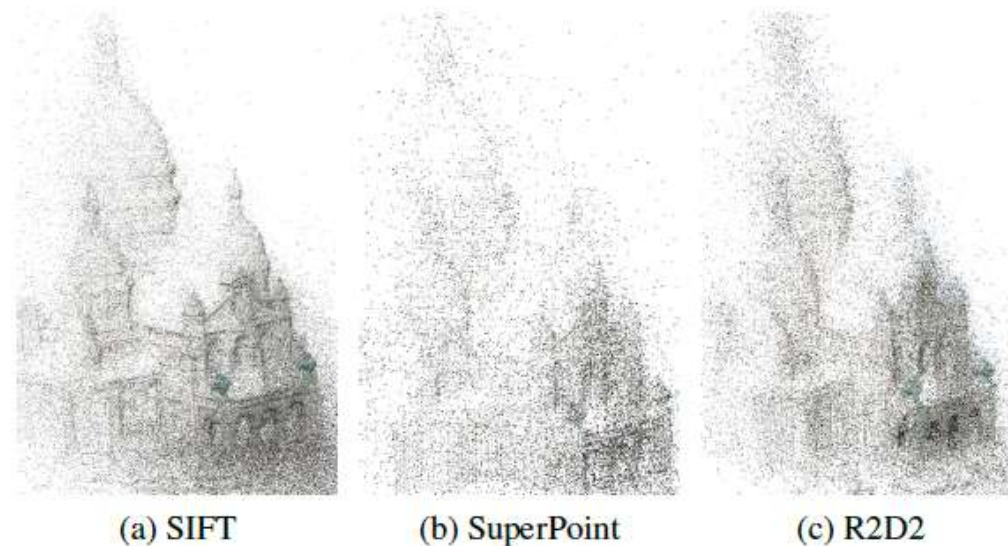


Fig. 6 COLMAP with different local features. We show the reconstructed point cloud for the scene “Sacre Coeur” using three different local features: SIFT, SuperPoint, and R2D2, using all images available (1179). The reconstructions with SIFT and R2D2 are both dense, albeit somewhat different. The reconstruction with SuperPoint is quite dense, considering it can only extract a much smaller number of features effectively, but its poses appear less accurate.