Automatic alignment of paintings and photographs depicting a 3D scene
For this work, we will focus on paintings where the artist made an effort to accurately render the 3D scene.

Paintings from Pompeii, Casa di Championnet:

- Scholander
- Blouet (1825)
- Gell (1814-1817)

• Useful for applications where drawings and paintings are the primary record (e.g. archaeology)
• Camera lucida was used to aid the artists
“In this work, we investigate features for aligning paintings and photographs. We primarily build upon recent success in recovering dense 3D points and a triangular mesh from a set of photographs”
Main goal

Photographs → 3D model
Painting → Viewpoint of painting
Why aligning painting is more difficult than aligning photographs?
Main Problem

- work on automatically aligning non-photographic has not done yet.
- difficulties arise when trying to align paintings and photographs:
  - Color
  - Geometry
  - Illumination
  - Shadows
  - Texture
Main Problem

- Shadow:
More Problems

- Time changes
  - Wall murals disappeared.
  - Central columns have changed shape.
More Problems

- Local feature matching using SIFT:
Roman town which founded around 600 BC.
Buried in 79 AD.
Rediscovered in 1599.
First excavated in 1748.
The complex takes its name from General Jean Étienne Championnet who, during the French presence in Naples, was a great supporter of the studies at Pompeii.

• The Championnet complex, an entire residential quarter which opened around 2016

• The complex include house of geometric Mosaics and the municipal building

• The complex is 4100m²

• The complex takes its name from General Jean Étienne Championnet who, during the French presence in Naples, was a great supporter of the studies at Pompeii
Study Contribution

- Combine the gist descriptor with the *view-retrieval algorithm*

- Develop a fine alignment procedure based on the Iterative Closest Point (ICP) algorithm.

*view-synthesis/retrieval of Irschara* - computational model of the recognition of real world scenes that bypasses the segmentation and the processing of individual objects or region
Algorithm overview

1. Recover triangular mesh M from set of photographs J using Bundler, PMVS and Poisson surface reconstruction

2. Coarse alignment by view sensitive retrieval:
   A. Generate virtual cameras that uniformly sample the recovered 3D scene and render the views
   B. Find near by virtual viewpoint \( \theta \) to \( L \) by gist feature matching.

3. Fine alignment by matching view-dependent contours:
   A. Extract contours (ridges, valleys, occluding) for \( \theta \) from M
   B. Use shape context feature to iteratively match contours for \( \theta \) to gPB contours extracted from \( L \) and estimate \( \theta \) from the correspondence.
Algorithm overview

1. Recover triangular mesh $M$ from set of photographs $J$ using Bundler, PMVS and Poisson surface reconstruction

2. Coarse alignment by view sensitive retrieval:
   A. Generate virtual cameras that uniformly sample the recovered 3D scene and render the views
   B. Find near by virtual viewpoint $\theta$ to $L$ by gist feature matching.

3. Fine alignment by matching view-dependent contours:
   A. Extract contours (ridges, valleys, occluding) for $\theta$ from $M$
   B. Use shape context feature to iteratively match contours for $\theta$ to gPB contours extracted from $L$ and estimate $\theta$ from the correspondence.
Bundler matches features extracted at interest points to recover a sparse 3D point set, along with camera parameters for each photograph.

\[
\min_{a_j, b_i} \sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij} d(Q(a_j, b_i), x_{ij})^2
\]

- \(n\) 3D points.
- \(M\) views.
- \(x_{ij}\) - projection of the \(i\)th point on image \(j\).
- \(v_{ij}\) - binary variables equals 1 if point \(l\) is visible in image \(j\).
- \(a_j\) - vector of cameras.
- \(b_i\) - vector of 3D points.
- \(Q(a_j, b_i)\) - Predicated projection of point \(l\) on image \(j\).
- \(D(x, y)\) – Euclidean distance.

*Bundle adjustment minimizes the total reprojection error with respect to all 3D point and camera parameters.
The recovered camera parameters are then used as input for the PMVS algorithm (Patch-based MultiView Stereo).

The algorithm uses photometric matching and region growing to recover a dense 3D point set.
Finally, the dense 3D point set is used as input for the Poisson surface reconstruction algorithm, recover a triangular mesh.
Algorithm overview

1. Recover triangular mesh $M$ from set of photographs $J$ using Bundler, PMVS and Poisson surface reconstruction

2. Coarse alignment by view sensitive retrieval:
   A. Generate virtual cameras that uniformly sample the recovered 3D scene and render the views
   B. Find near by virtual viewpoint $\theta$ to $L$ by gist feature matching.

3. Fine alignment by matching view-dependent contours:
   A. Extract contours (ridges, valleys, occluding) for $\theta$ from $M$
   B. Use shape context feature to iteratively match contours for $\theta$ to gPB contours extracted from $L$ and estimate $\theta$ from the correspondence.
Global image representation.

Algorithm stages:
- Subdivide image in 4*4 sub images.
- Calculate Gabor responses in each of this sub images.
- Create histogram of Gabor responses in each sub-image.
generate virtual camera matrices that uniformly sample viewpoints of the 3D model.

search for the dominant scene plane on the PMVS point set.
Sample camera centers in a grid and use horizontal orientations at each sample point.
- Render each virtual viewpoint.
- Match the appearances of the painting and virtual viewpoints using the gist descriptor.
Coarse alignment by view-sensitive retrieval - 3
Algorithm overview

1. Recover triangular mesh $M$ from set of photographs $J$ using Bundler, PMVS and Poisson surface reconstruction

2. Coarse alignment by view sensitive retrieval:
   A. Generate virtual cameras that uniformly sample the recovered 3D scene and render the views
   B. Find near by virtual viewpoint $\theta$ to $L$ by gist feature matching.

3. Fine alignment by matching view-dependent contours:
   A. Extract contours (ridges, valleys, occluding) for $\theta$ from $M$
   B. Use shape context feature to iteratively match contours for $\theta$ to gPB contours extracted from $L$ and estimate $\theta$ from the correspondence.
Contours for 3D model and painting - 1

- extract contours corresponding to folds, creases, and occlusions from the 3D model.

Red: ridges and valleys, Green: occlusion boundaries
Algorithm overview

1. Recover triangular mesh $M$ from set of photographs $J$ using Bundler, PMVS and Poisson surface reconstruction

2. Coarse alignment by view sensitive retrieval:
   A. Generate virtual cameras that uniformly sample the recovered 3D scene and render the views
   B. Find near by virtual viewpoint $\theta$ to $L$ by gist feature matching.

3. Fine alignment by matching view-dependent contours:
   A. Extract contours (ridges, valleys, occluding) for $\theta$ from $M$
   B. Use shape context feature to iteratively match contours for $\theta$ to gPB contours extracted from $L$ and estimate $\theta$ from the correspondence.

- $M$ – 3D model
- $L$ - painting
we find edges in the painting using the global probability of boundary (gPB) detector.
ICP-Iterative closest point algorithm

\[ S_L(\emptyset) = \{X | B_L(x, \emptyset) = 1\} \]

- \( S_L(\emptyset) \) - Set of edge points for orientation \( \emptyset \) in the binary image for the painting
- \( B_L \) - Binary Image
- \( \emptyset \) - Edge orientation (in painting)
- \( X \) - Point (has a response strength and edge orientation \( \emptyset \)
Cost Function

$$\min_{\theta} \sum_{\emptyset} \sum_{x_i \in S_L(\emptyset)} \min_{x_j \in S_M(\theta,\emptyset)} \min(|X_i - X_j|^2, \gamma)$$

- $S_L$ - Set of Edge point in the painting
- $S_M$ - Set of Edge point in the 3D model
- $\emptyset$ - Edge orientation (in painting)
- $\theta$ - Edge orientation (in 3D model)
- $X$ - Point (has a response strength and edge orientation $\emptyset$
- $\gamma$ – inlier threshold
Shape context is a feature descriptor used in object recognition.

S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts.
1. Sample (randomly) the number of points required to fit the model
2. Solve the model parameters using sample
3. Score by the fraction of inlier within a present threshold ‘t’ of the model.
4. Repeat 1-3 until the best model is found

Point are the shape context features $f_i$, $f_j$ and edge orientations $\phi_i$, $\phi_j$
Fine alignment by matching view-dependent contours

- (a) Extracted edges (painting – red, 3D model - blue).
- (b) Shape context sample points (green lines).
- (c) Dense edge inlier correspondences found by RANSAC
- After each iteration, the viewpoint is updated
Iteration 1

- Painting and 3D model contours
- Shape context sample points and putative correspondences
- Dense inliers
Iteration 2

- Painting and 3D model contours
- Shape context sample points and putative correspondences
- Dense inliers
Iteration 3

- Painting and 3D model contours
- Shape context sample points and putative correspondences
- Dense inliers
Iteration 7

Painting and 3D model contours

Shape context sample points and putative correspondences

Dense inliers
Experimental results

- 9 paintings.
- 563 photographs.
- The final mesh contains 10M vertices and 20M triangles.
Experimental results

Painting

3D model contours overlaid

Rendering from 3D model
Experimental results

Painting

3D model contours overlaid

Rendering from 3D model
Experimental results

3D model contours overlaid

Painting

Rendering from 3D model
Conclusions

- Shown successful alignment of historical painting of an archaeological site to noisy 3D model constructed from modern photographs.
- System handles drastic change in appearance, which is difficult for current system relying on local feature matching.
- 3D model allows alignment from unseen viewpoints.
• Painting-to-3D Model Alignment Via Discriminative Visual Elements, M. Aubry, B. Russell and J. Sivic
Resources

- Shape matching and object recognition using shape contexts, serge Bolongie
- Modeling the shape of the scene: a holistic representation of the spatial envelope, A. Oliva and A. Torralba
- A. Irschara, C. Zach, J.-M. Frahm, and H. Bischof. From structure from-motion point clouds to fast location recognition
- Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis
- M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction.
- http://www.pompeionline.net/pompeii/
- http://www.cs.cornell.edu/~snavely/bundler/
Questions?