

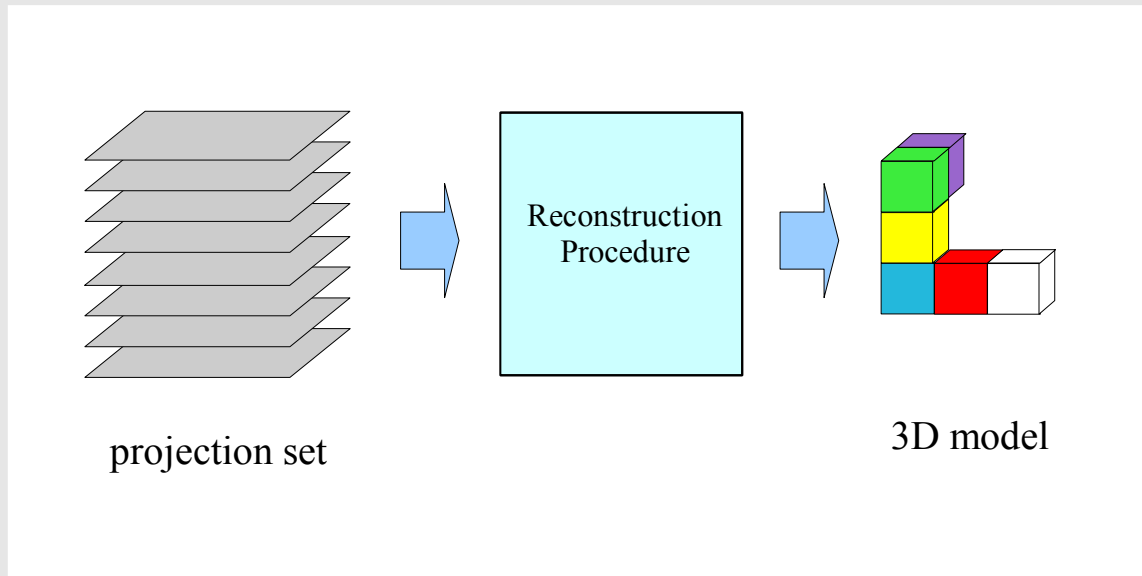
**COMBINATORIAL CLASSIFICATION TO
SEPARATE HOMOGENEOUS
SUBSETS OF HETEROGENEOUS PROJECTION
SETS**

Gabor T. Herman and
Miroslaw Kalinowski

Introduction/Background

Introduction/Background

3D reconstruction procedure

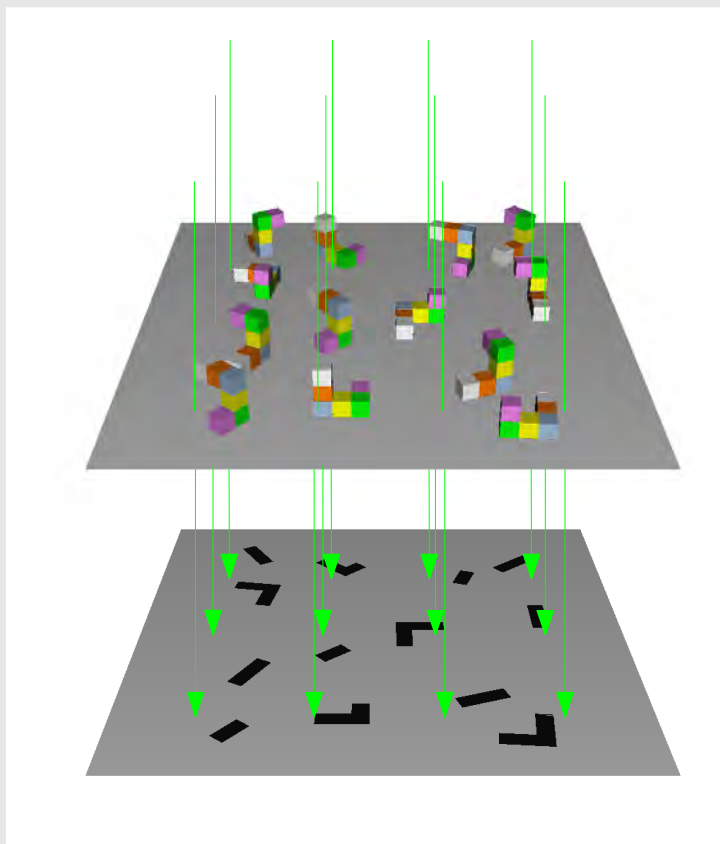


The 3D reconstruction procedure.

A set of 2D projection images is used to produce a 3D model of the object from which these projections images were obtained

Introduction/Background

Single Particle Reconstruction

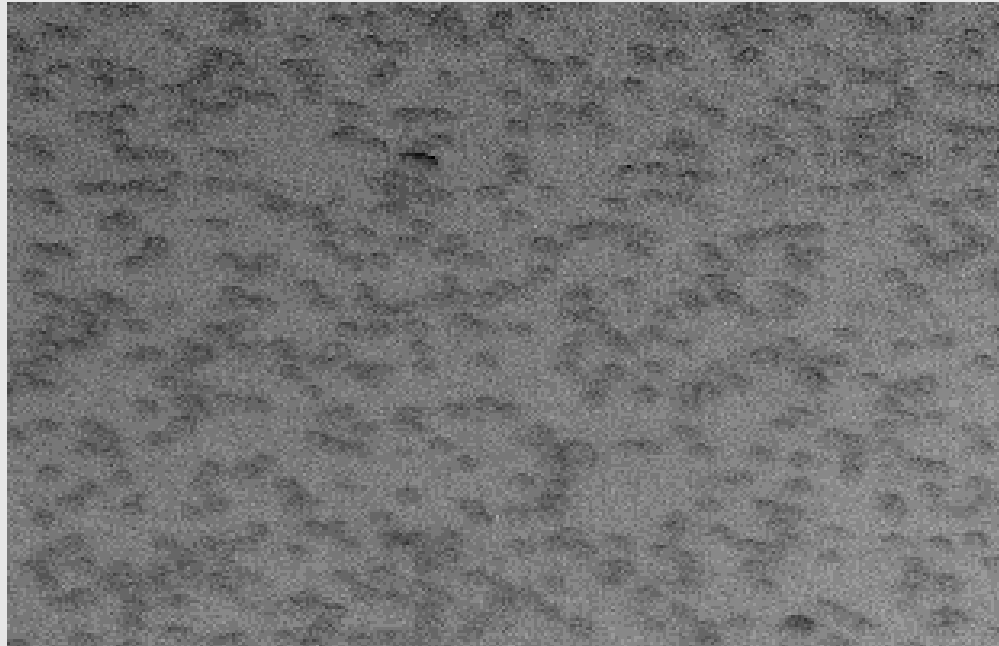


Process of obtaining projections.

Many identical randomly oriented molecules are simultaneously projected

Introduction/Background

Single Particle Reconstruction



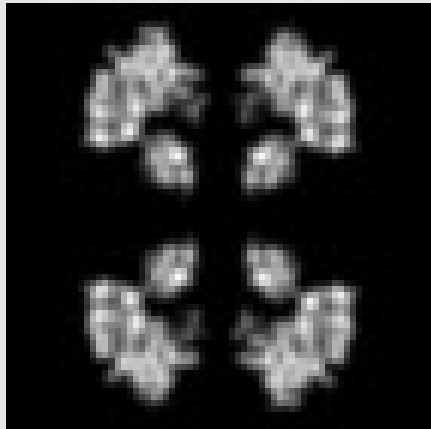
EM micrograph.

Each dark spot is a 2D projection of the molecule

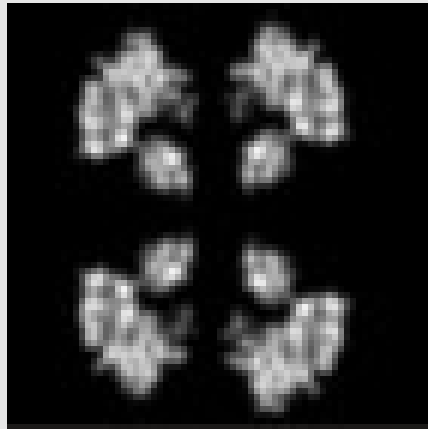
Introduction/Background

Heterogeneity

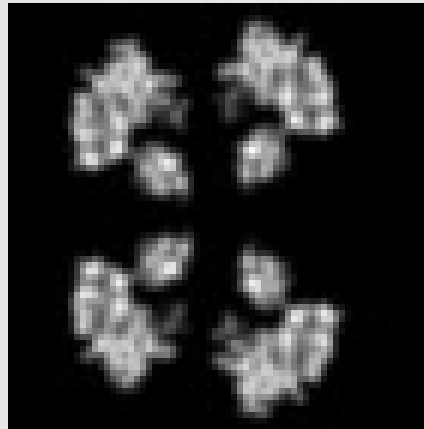
Conformation A



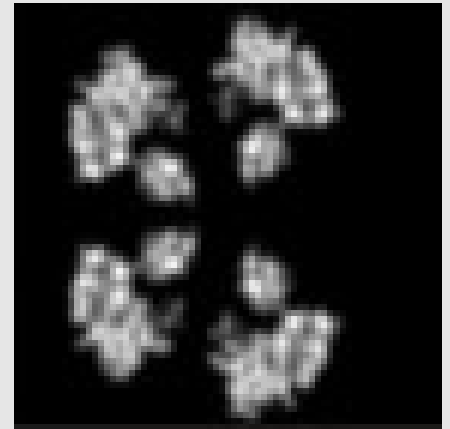
Conformation B



Conformation C



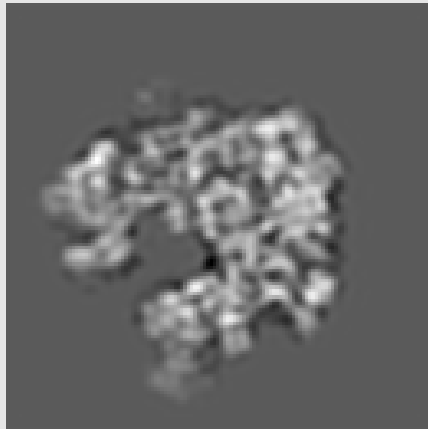
Conformation D



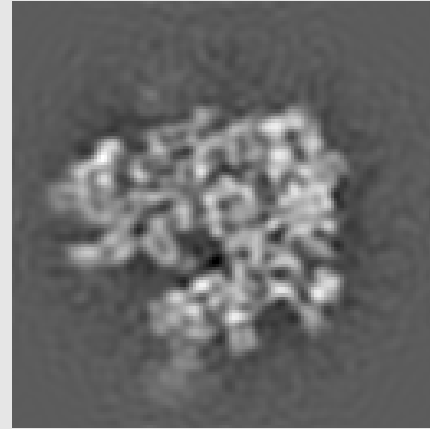
Heterogeneity - deformations of 3D structure.

Introduction/Background

Heterogeneity



Conformation A

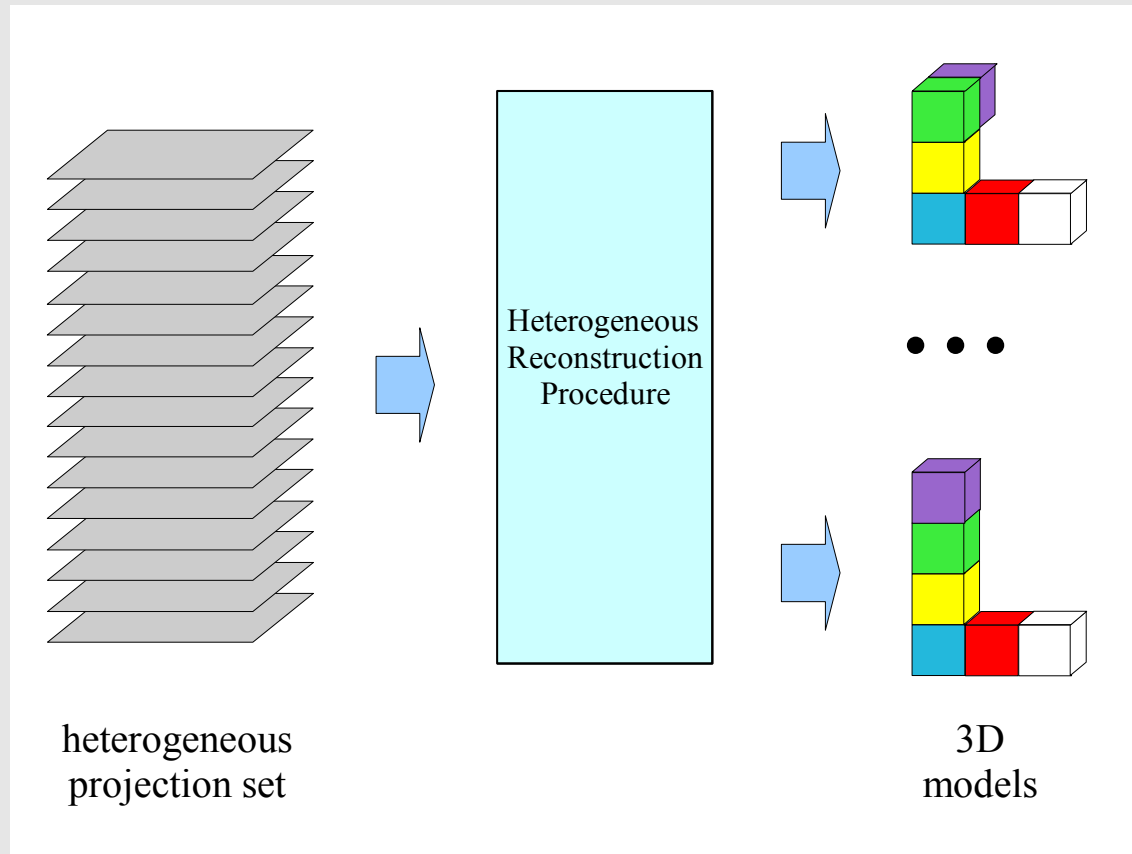


Conformation B

Heterogeneity - bound and unbound molecule

Introduction/Background

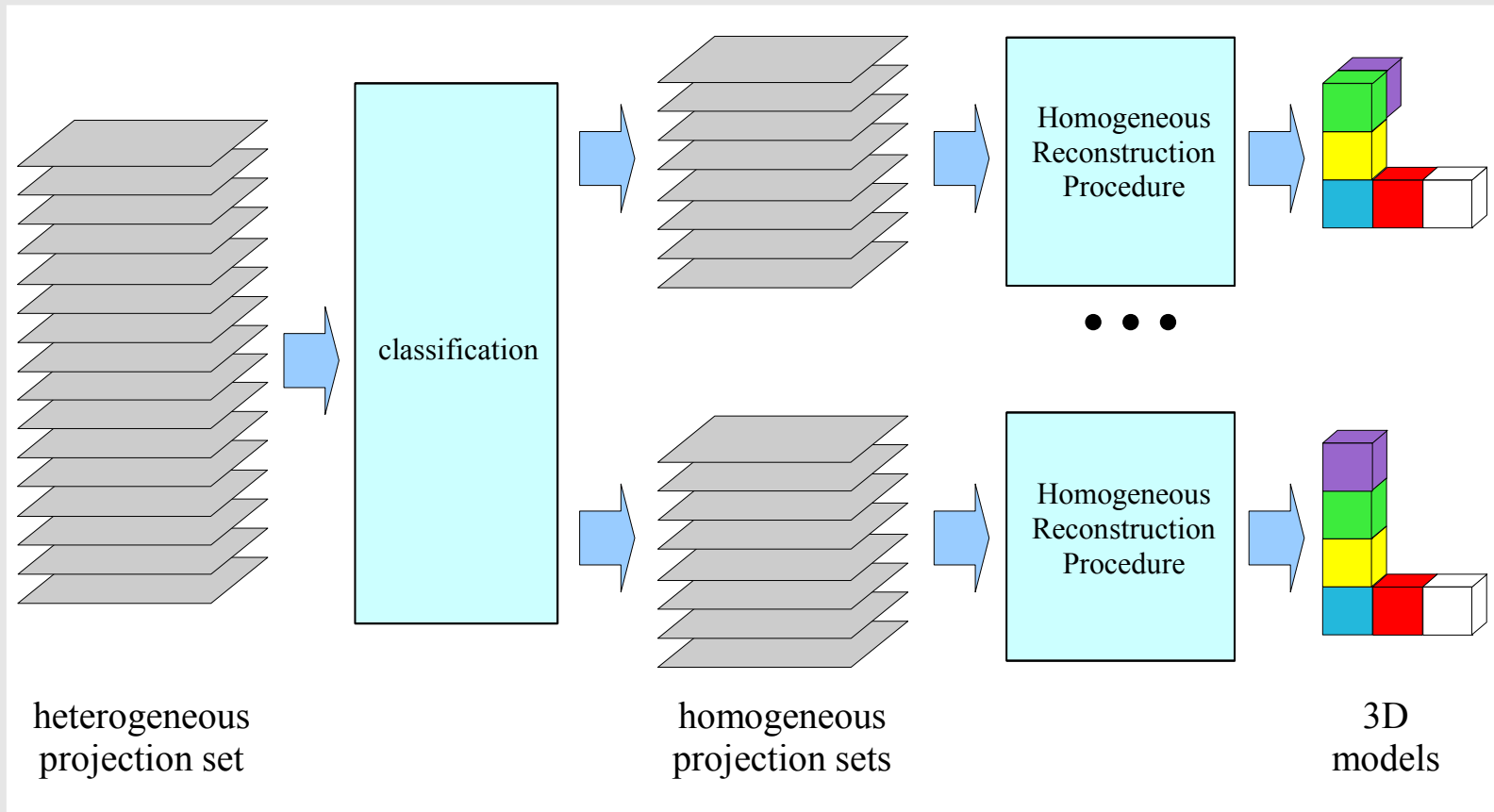
Reconstruction from Heterogeneous Sets



Reconstruction procedure returns 3D models of all conformations represented in the projection set.

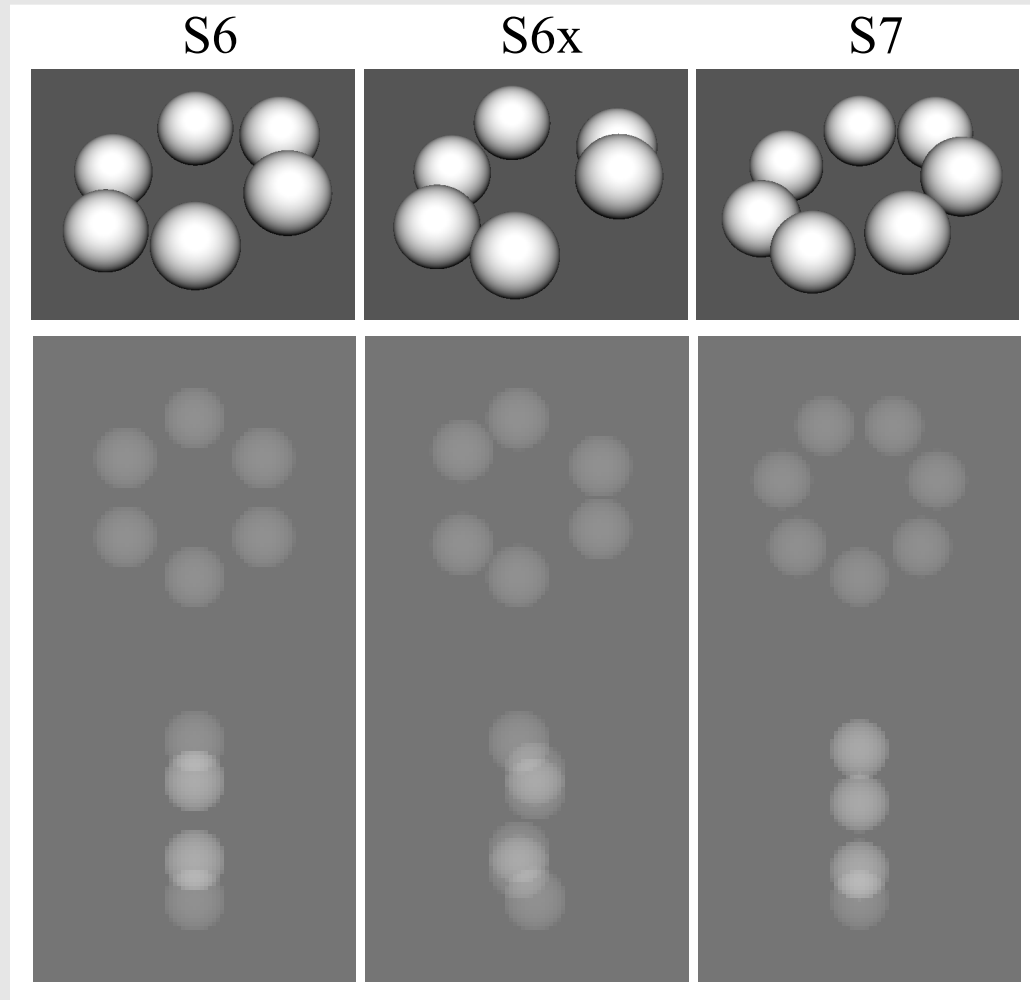
Introduction/Background

Classification Based Approach



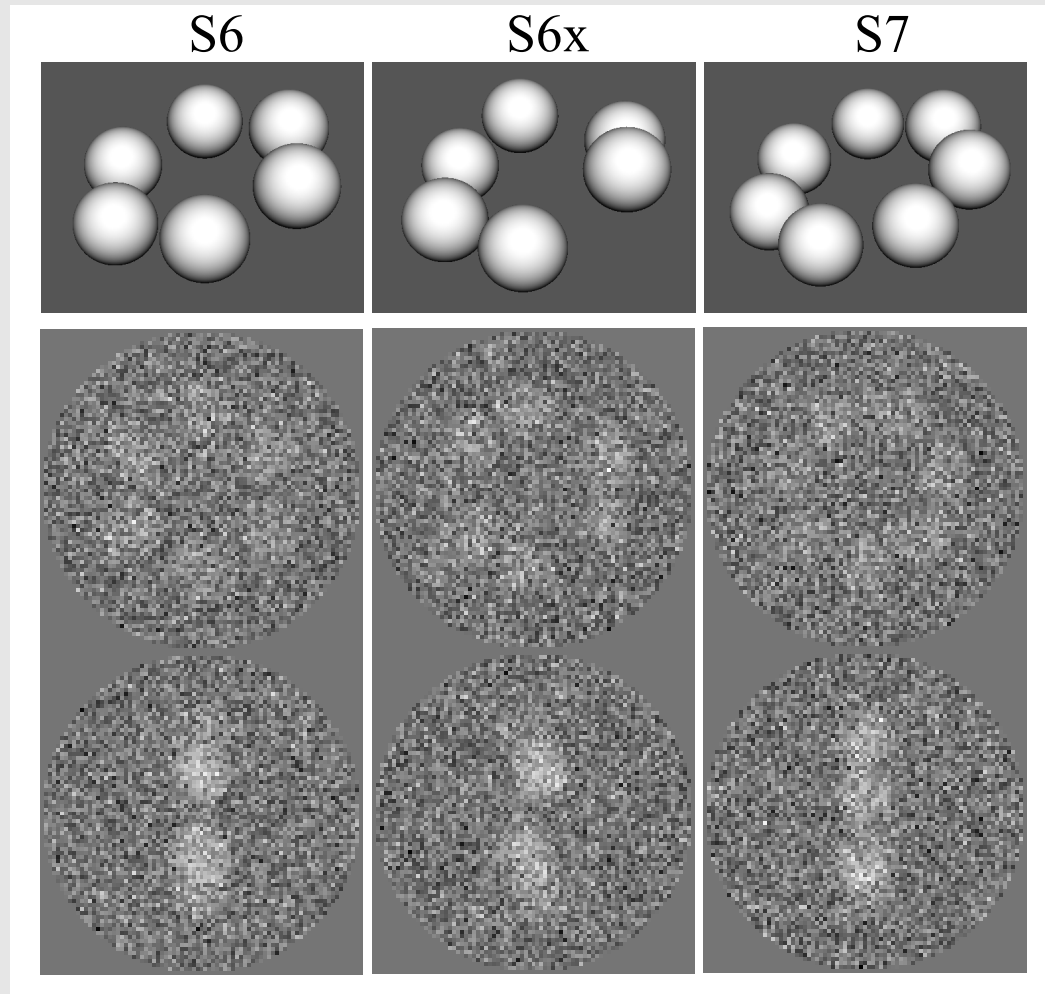
Introduction/Background

Classification Based Approach (Cont.)



Introduction/Background

Classification Based Approach (Cont.)



Introduction/Background

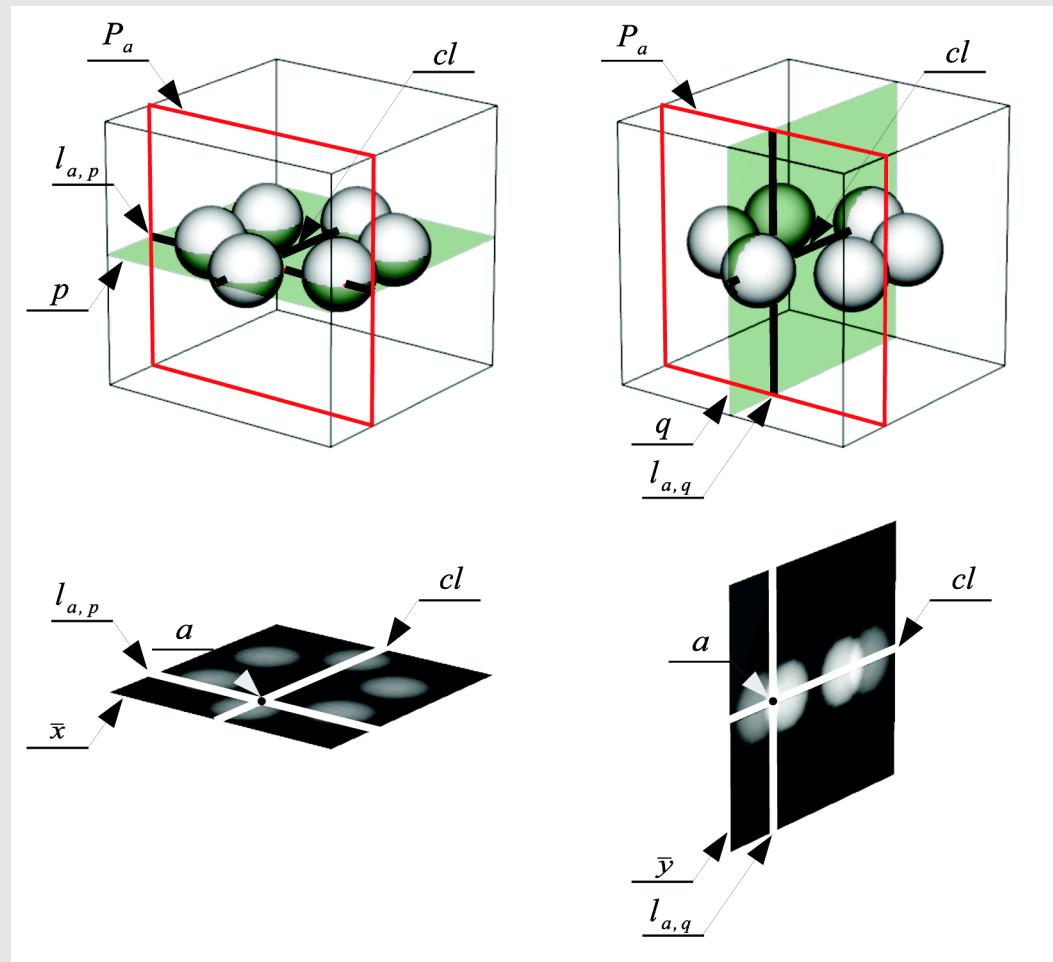
Objectives

- Demonstrate feasibility of classification based approach.
- Develop a reconstruction procedure that is capable of handling a wide variety of reconstruction problems, including those for which no prior knowledge is available.
- By utilizing mathematical properties of the projection images and combinatorial optimization techniques, construct an appropriate unsupervised image classification procedure.
- Demonstrate that an implementation of the proposed method, efficient enough to handle classification problems encountered in 3D-EM, is possible.

Projection Image Dissimilarity Measure

Projection Image Dissimilarity Measure

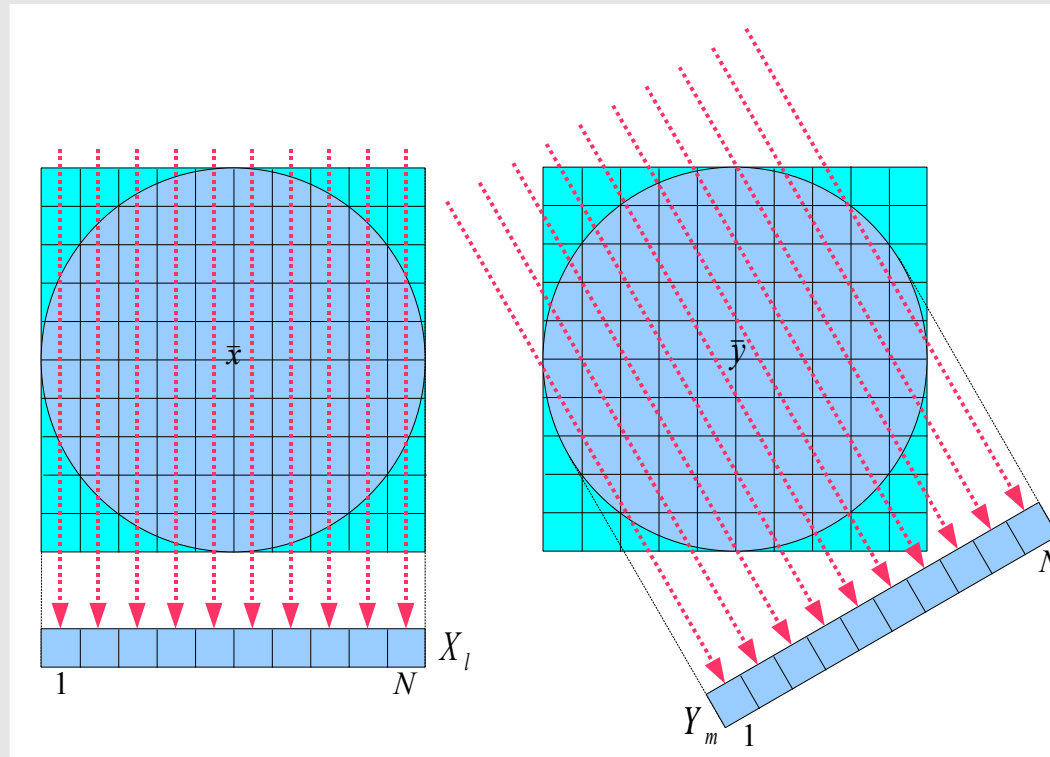
Mathematical Background



Two projections and of object S_6 .

Projection Image Dissimilarity Measure

Application to EM Projection Images



The 10 x 10 image and its two 1D projections.

Circular mask and 1D projections of 2D image (the values of pixels with centers outside of the masking circle are set to zero).

Projection Image Dissimilarity Measure

Definition

- Let L be the number of evenly distributed lines at which we will look in each projection plane p , we index them by $l, 1 \leq l \leq L$.
- On each of them we pick N points (these points are picked at matching distances).
- For each projection image \bar{x} and for each such line l we define an N -dimensional vector X_l whose n -th component (for $1 \leq n \leq N$) is the estimated line integral in the projection image along the line perpendicular to l going through the n -th point.
- If errors due to noise and discretization are ignored, then two projection images \bar{x} and \bar{y} of the same 3D object must have identical vectors X_l and Y_m for some pair of indexes l and m .

Projection Image Dissimilarity Measure

Definition

In reality, due to discretization error and noise, there is practically no pair of indexes l and m for which vectors X_l and Y_m are identical.

However there is an increased probability of finding two ‘similar’ vectors X_l and Y_m , if the projections \bar{x} and \bar{y} came from the same object.

Let us assume that ‘dissimilarity’ of vectors can be measured by a function s that returns 0 given a pair of identical vectors and a positive value indicative of the differences between the vectors otherwise.

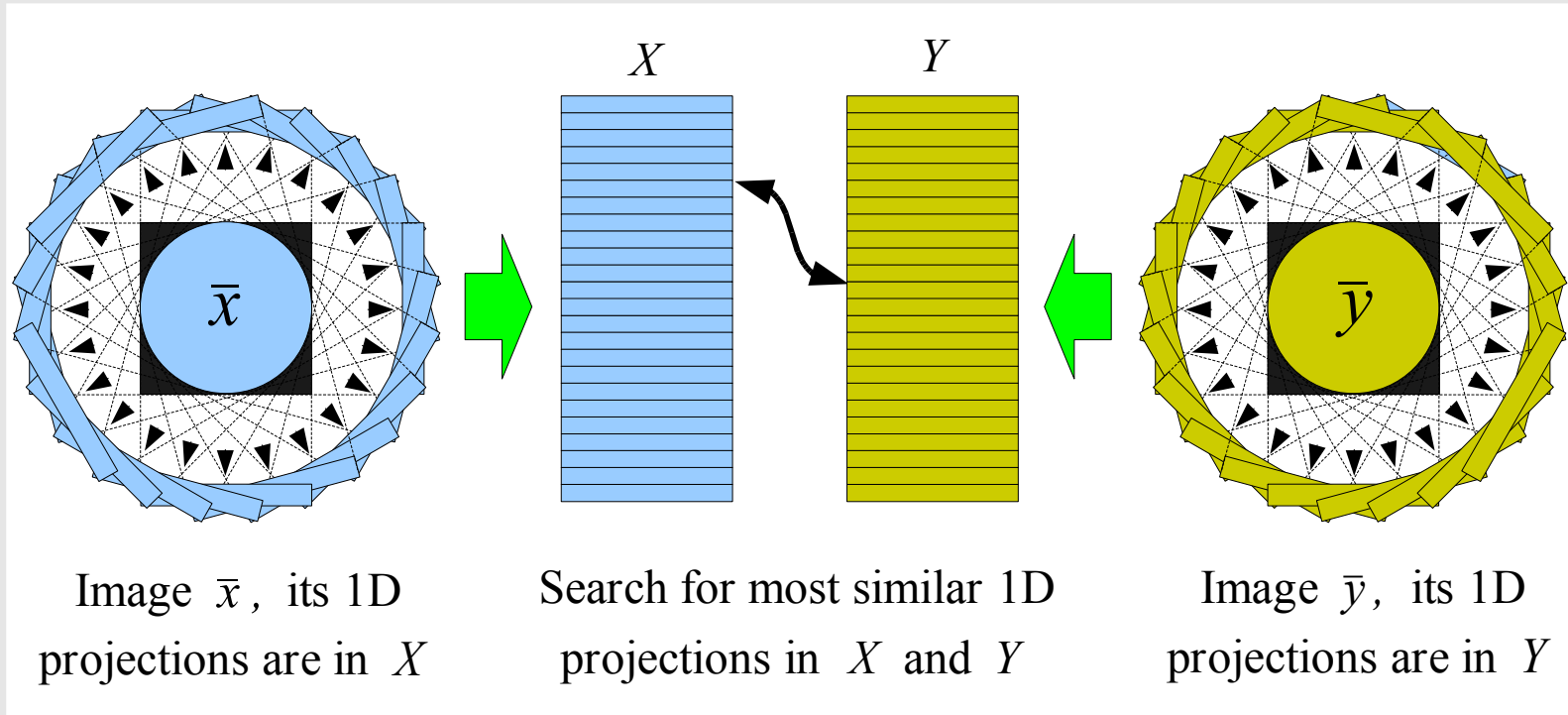
Definition

We define the dissimilarity of any two projection images \bar{x} and \bar{y} as

$$\bar{s}(\bar{x}, \bar{y}) = \min_{1 \leq l, m \leq L} s(X_l, Y_m).$$

Projection Image Dissimilarity Measure

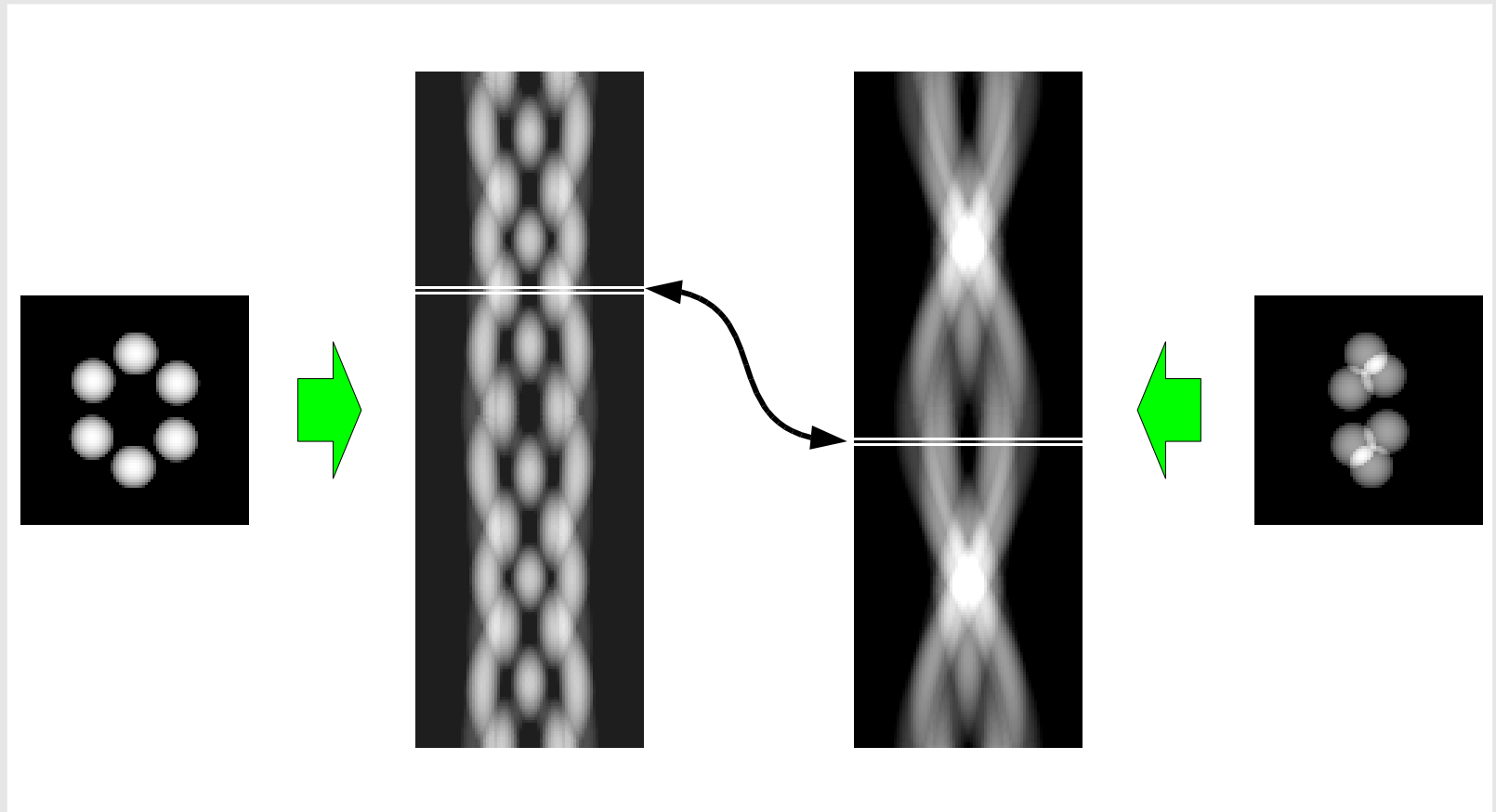
Application to EM Projection Images



Process of calculating the value of dissimilarity measure for two images \bar{x} and \bar{y} .

Projection Image Dissimilarity Measure

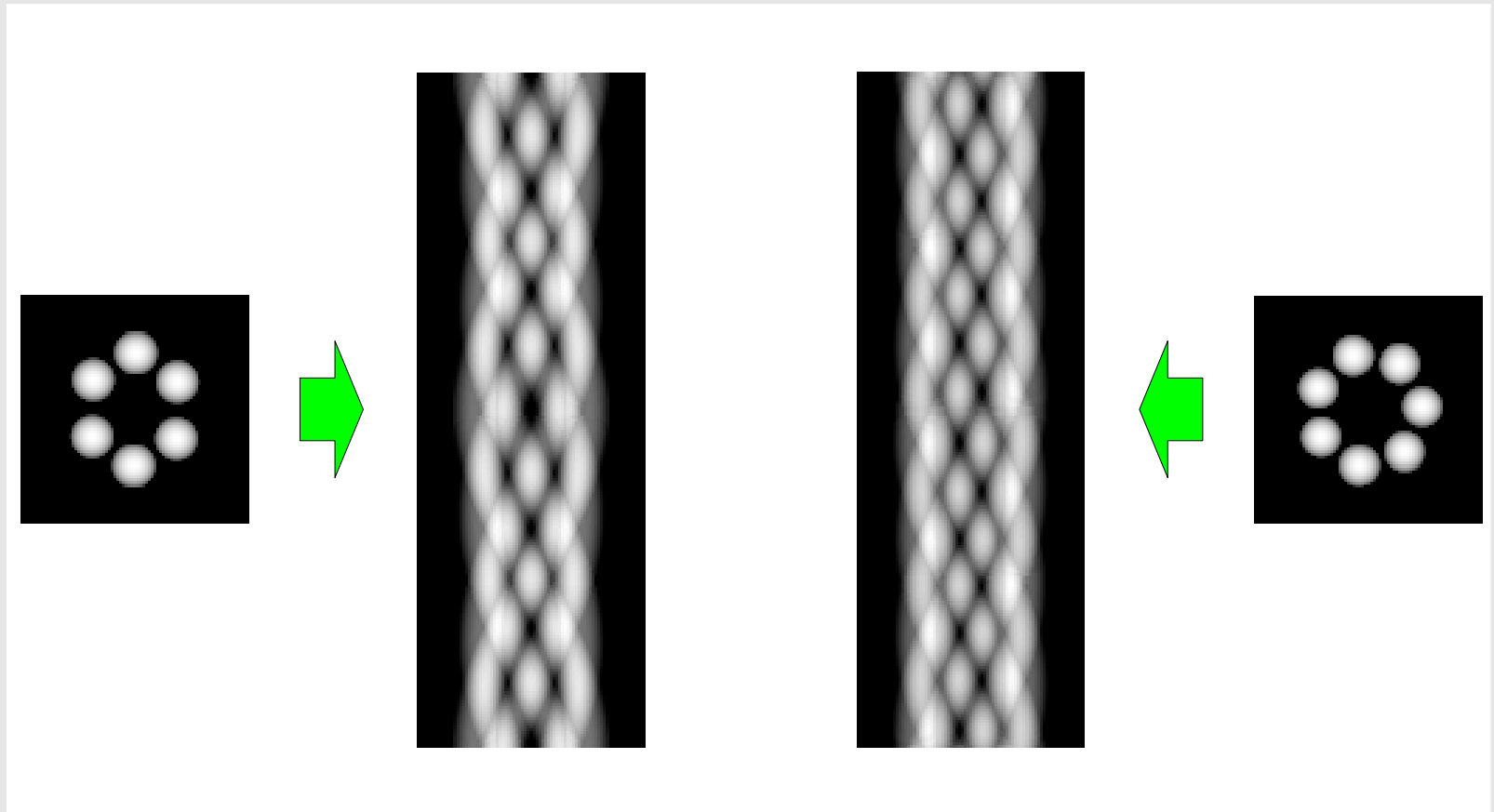
Application to EM Projection Images



Matching line in the sinograms of two noiseless projections images that originate from the same 3D object.

Projection Image Dissimilarity Measure

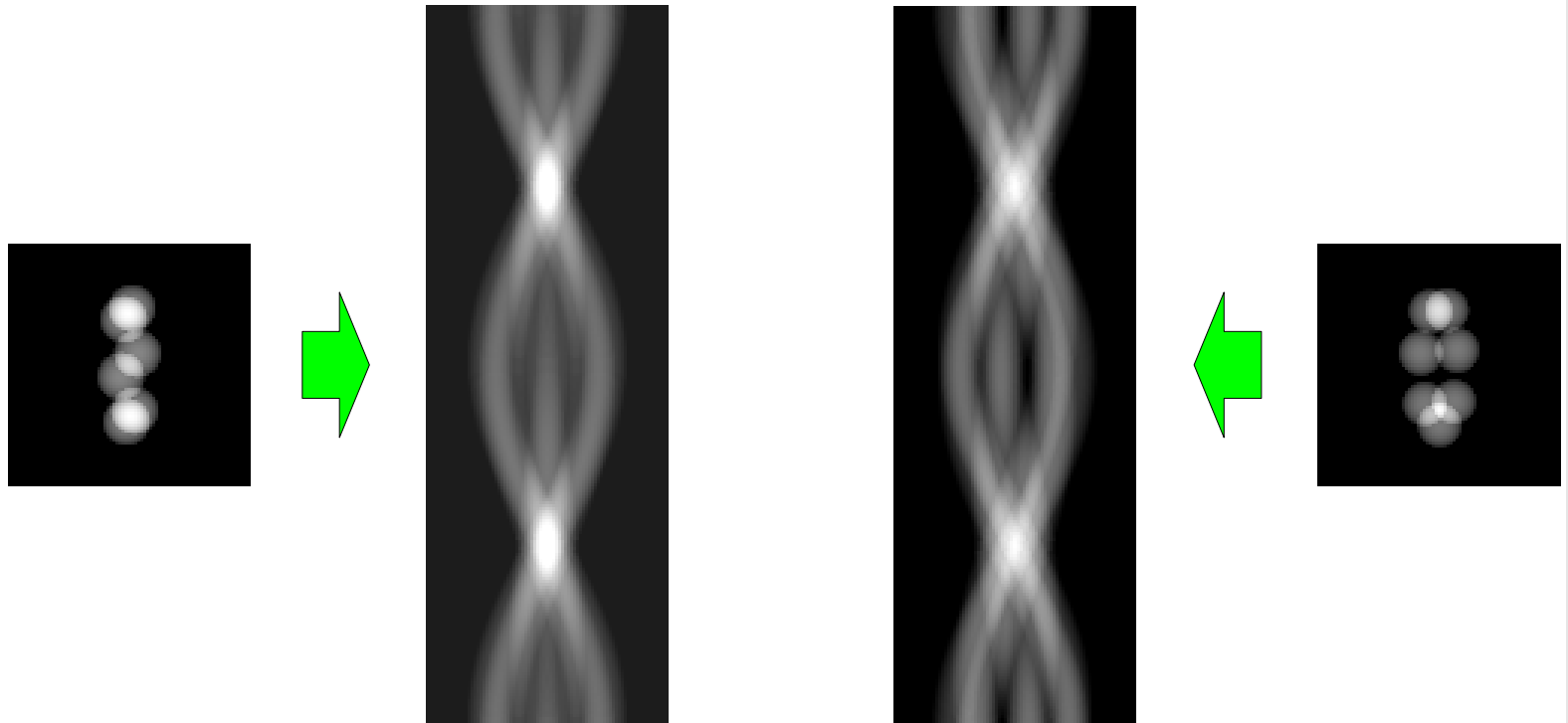
Application to EM Projection Images



The sinograms of two noiseless projections images that originate from different 3D object.

Projection Image Dissimilarity Measure

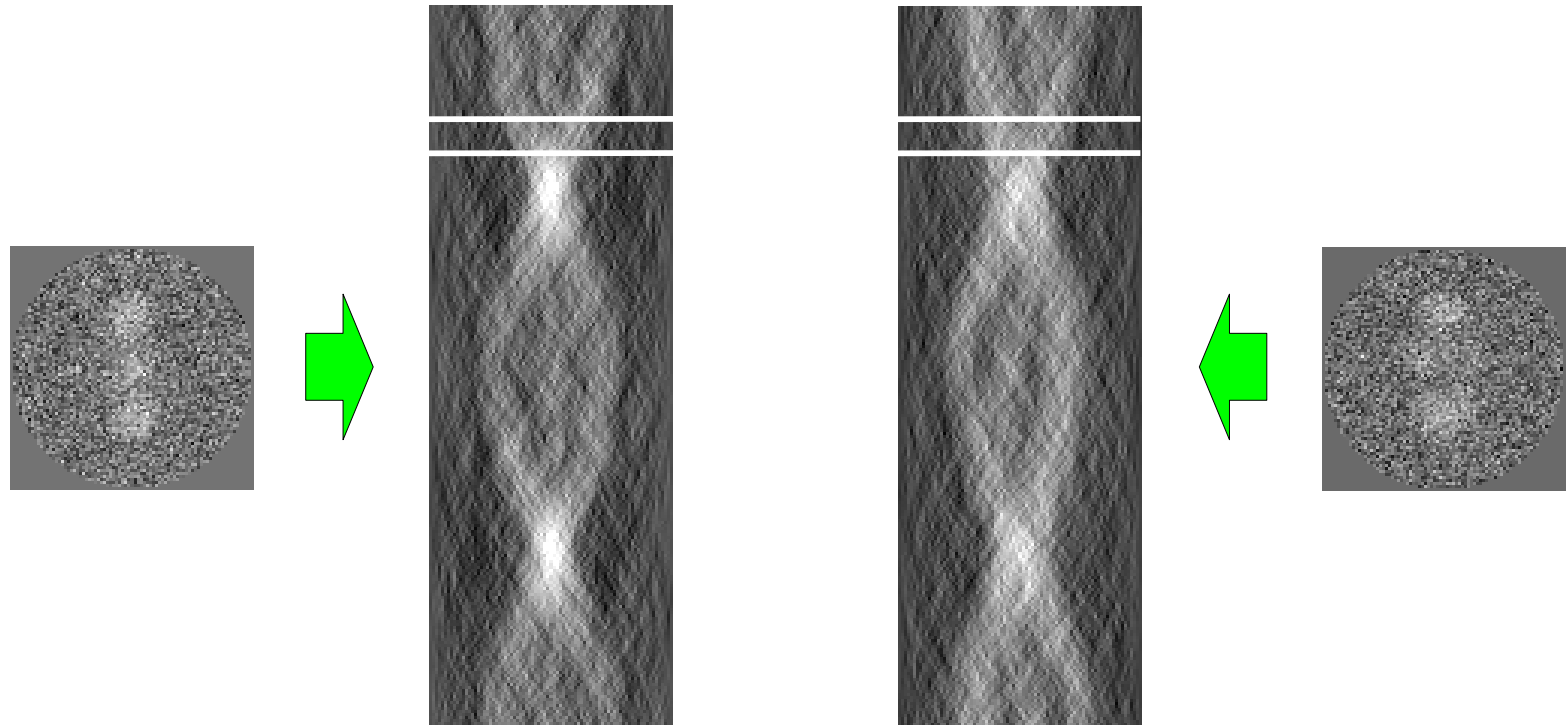
Application to EM Projection Images



The sinograms of two noiseless projections images that originate from different 3D object.

Projection Image Dissimilarity Measure

Application to EM Projection Images



The sinograms of two noisy projections images that originate from different 3D objects

Projection Image Dissimilarity Measure

Application to EM Projection Images

Vector Dissimilarity Measure

For current work

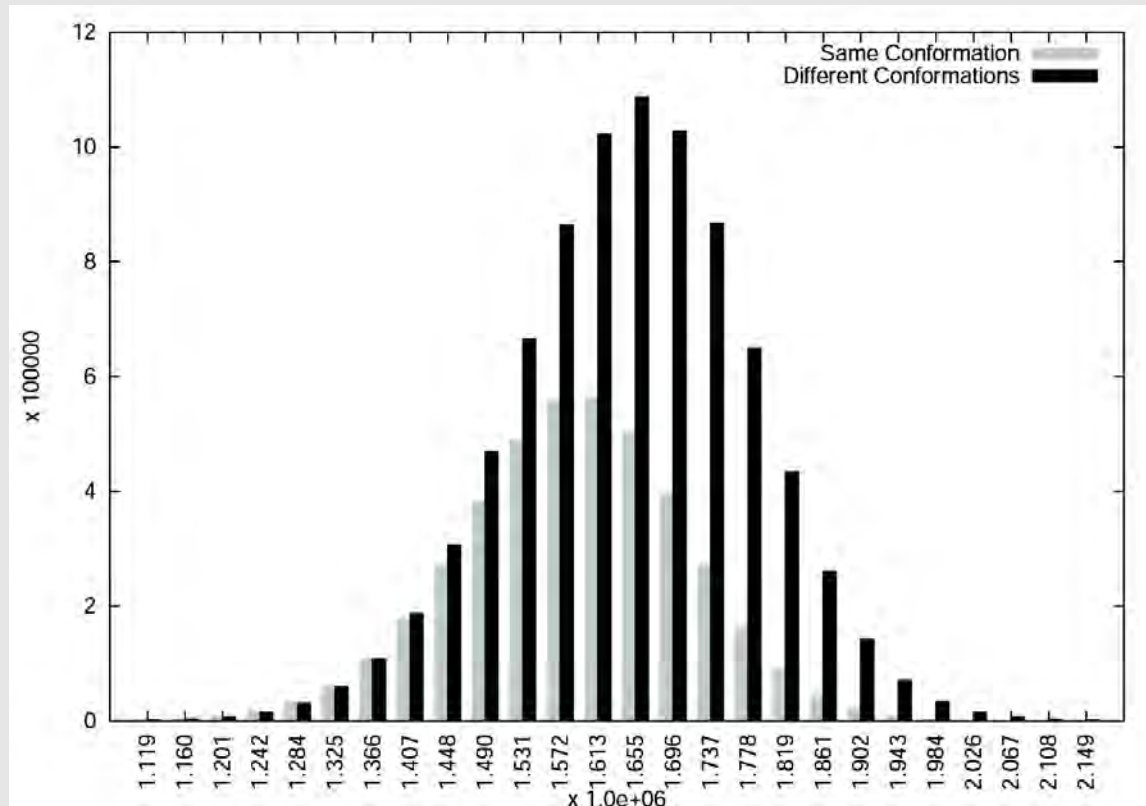
$$s(x, y) = \|x - y\|^2$$

(the squared 2-norm of the difference) was chosen.

Projection Image Classification as Optimization Problem

Projection Image Classification as Optimization Problem

Similarity of EM Projection Images



Histograms of distances between pairs of projection images in a heterogeneous set for the pairs originating from the same and from different conformations.

Projection Image Classification as Optimization Problem

Formal Statement of the Optimization Problem

Definition

Let V denote the heterogeneous projection set. For any positive integer K , a K -partition A of V is a set $\{A_1, \dots, A_K\}$ of K nonempty subsets of V such that the union of these subsets is the whole of V and no two subsets have any element in common.

Projection Image Classification as Optimization Problem

Formal Statement of the Optimization Problem

GIVEN a set V of 2D projections and a positive integer K ,

FIND a K -partition $A = \{A_1, \dots, A_K\}$ of V ,

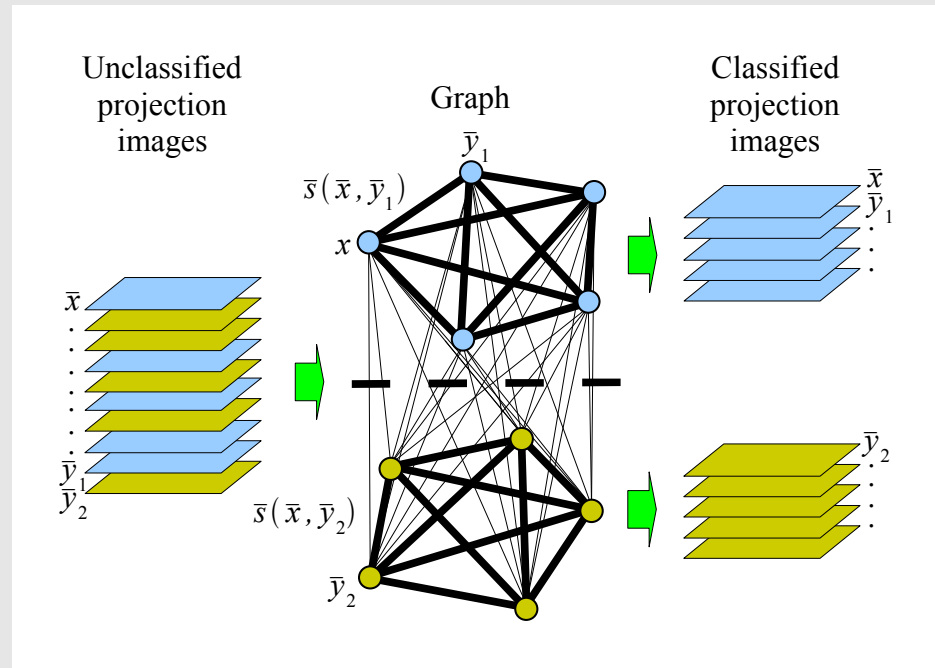
SUCH THAT

$$\sum_{k=1}^K \sum_{\bar{x}, \bar{y} \in A_k} \bar{s}(\bar{x}, \bar{y}) \quad (3.1)$$

is as small as possible.

Projection Image Classification as Optimization Problem

Graph Theoretical Interpretation



Classification by graph cutting.

Images represented by nodes of each graph component belong to the same class.

Projection Image Classification as Optimization Problem

Graph Theoretical Interpretation

- When the projections in V are represented by nodes of a complete weighted graph G , and the weight of the edge between nodes \bar{x} and \bar{y} is the distance $\bar{s}(\bar{x}, \bar{y})$, then the edges between the nodes representing projections of the same object are more likely to have lower weights.
- The problem of separating the homogeneous subsets of a heterogeneous projection sets becomes a graph cutting problem, in which the objective is to find a separation of the graph G into K complete subgraphs G_1, \dots, G_K such that the sum of all edge weights in the subgraphs G_1, \dots, G_K is minimal.
- This problem is known as Max k-Cut, and in case $K = 2$ it is equivalent to the maximum capacity cut problem.

Projection Image Classification as Optimization Problem

Computational Complexity

- Both Max k-Cut and maximum capacity cut problems have been shown to be NP-complete.
- It also has been demonstrated that finding even approximately optimal solution to the Max k-Cut is NP-complete.
- The estimated run time for solving the 5,000 node instance of the graph cutting problem using DSDP algorithm is approximately one month.
- However, an efficient algorithm capable of producing good (from our classification problem perspective) estimates of Max k-Cuts for graphs originating from 3D-EM can be constructed!

Construction of the Distance Graph

Construction of the Distance Graph

Construction Cost

- Since the topology of our graphs is fixed, the process of constructing them is simple (only the weights of the edges must be calculated).
- However, the number of edge weights that need to be calculated is large (for a graph with 5,000 nodes, 12,497,500 edge weights must be calculated).
- A significant amount of computer time must be dedicated to calculating edge weights in a realistically sized graph. (without optimizations) it takes 24 hours on a single processor (Intel Xeon 1.7 GHz) to construct a graph for a data set that contains 5,000 images.
- Since the calculations of edge weights between different nodes of the graph are mutually independent, the task of constructing the graph can be easily parallelized. However, the cost of constructing such graphs increases proportionally to the square of the number of projection images.
- For larger datasets that contain tens of thousands of projection images, significant resources are required to construct the corresponding graphs.

Graph Cutting Algorithm

Graph Cutting Algorithm

Concept

- Initial graph cut (partitioning) is generated randomly.
- In each step of the algorithm reclassification of each node (2D projection) is considered.
- A new value of the objective function is calculated for each reclassification.
- Best or least harmful reclassification which is not prohibited by the taboo list is selected and executed.
- Reclassified node along with better of two objective function values (before and after reclassification) is used to update taboo list.
- Algorithm stops after executing specified number of steps.

Graph Cutting Algorithm

Concept

Taboo list operation - checking

- Reclassification is allowed if affected node is not on the list.
- The reclassification of the node is prohibited if it results in the value of the objective function worst than recorded for this node on the list.

Graph Cutting Algorithm

Concept

Taboo list operation - updating

- If 2D projection is already on the list the objective function value associated with this projection is updated.
- Otherwise, 2D projection with associated value is placed at the end of the list and if the list is full causes removal of the first projection from the list.

Graph Cutting Algorithm

Parameters

K : Number of classes

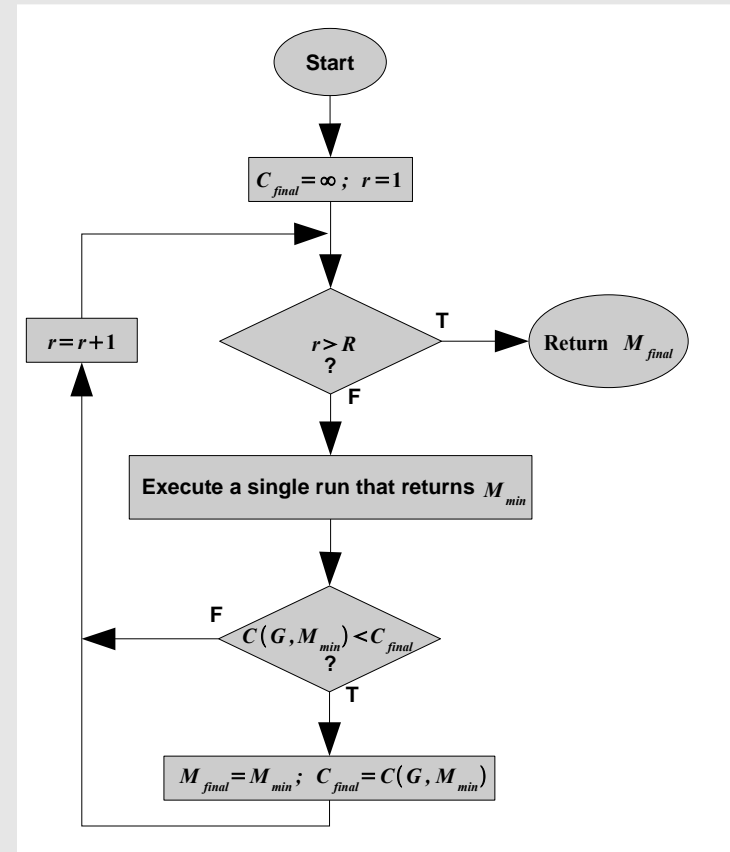
I : Number of iterations

t : Length of tabu list

Graph Cutting Algorithm

Multiple Runs

- The cut produced by our algorithm is an approximation of the Max k-Cut that depends on the initial random classification of the nodes.
- The chances of finding a good approximation of the Max k-Cut can be significantly increased by running the core algorithm several times.
- Since each of the runs starts from different randomly selected initial cut, the likelihood that all of them are many reassignments away from a good approximations decreases.



Evaluation

Evaluation

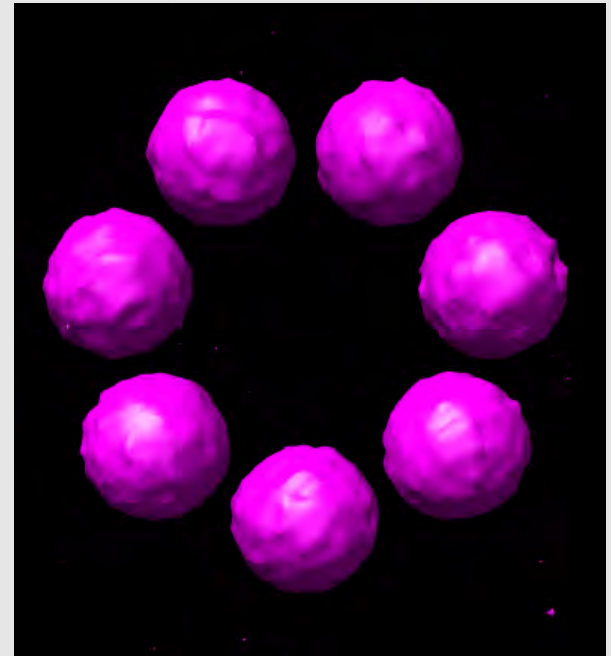
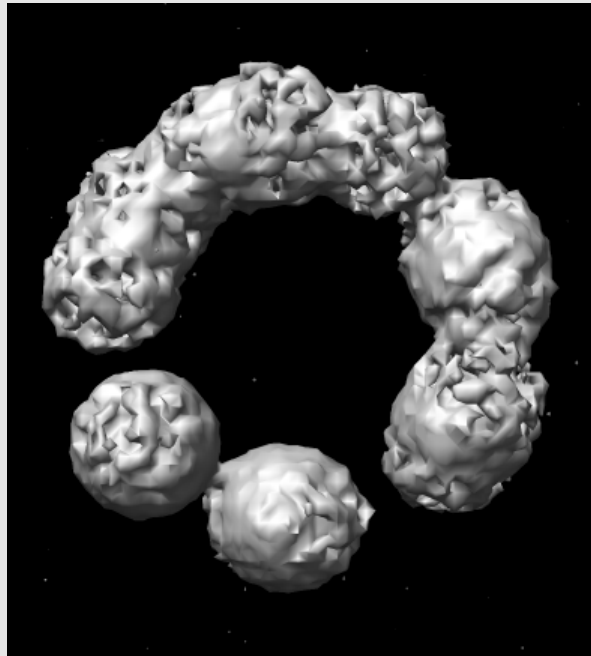
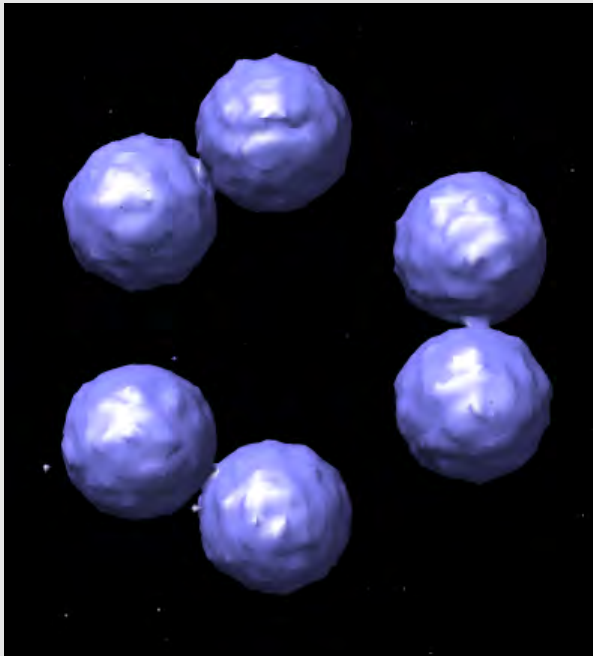
Experiments with Aligned Projection Images

Datasets

- Randomly selected projections of 2 or 3 objects (S6, S6x, S7)
- Representation ratios: 50:50, 35:65, 20:80, 33:33:33
- SNR = 0.1
- Perfectly alligned images

Evaluation

Experiments with Aligned Projection Images (50:50)

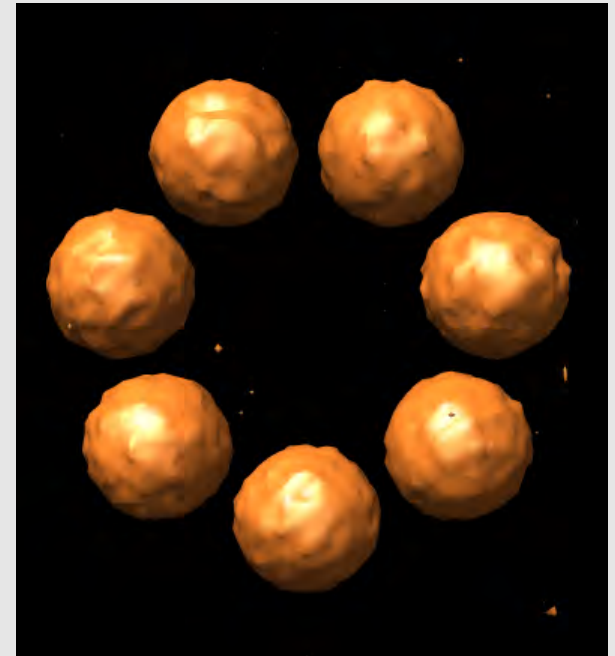
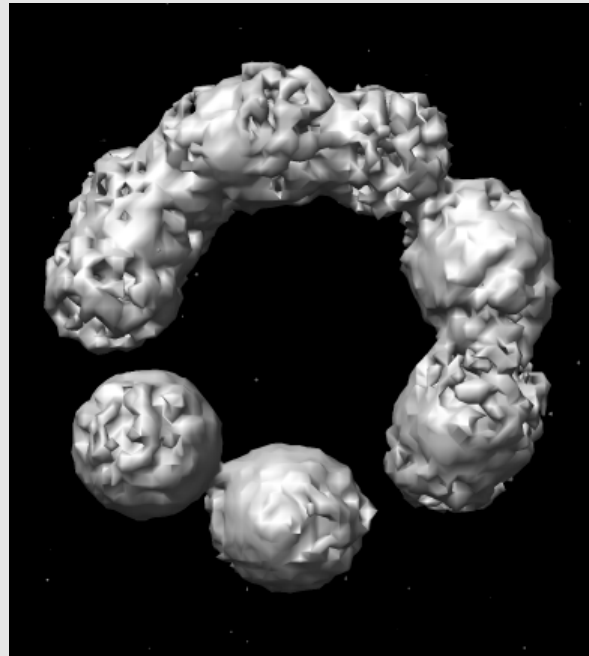
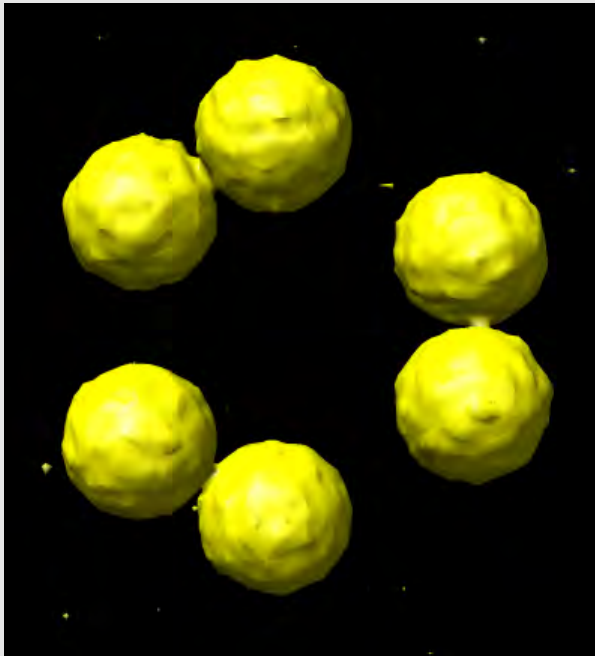


Center: 3D model obtained by reconstructing from heterogeneous projection set that contains aligned projection images of objects S6x and S7.

Left, Right: 3D models obtained by reconstructing from the aligned projection images of objects S6x, S7 classified by the proposed method.

Evaluation

Experiments with Aligned Projection Images (50:50)

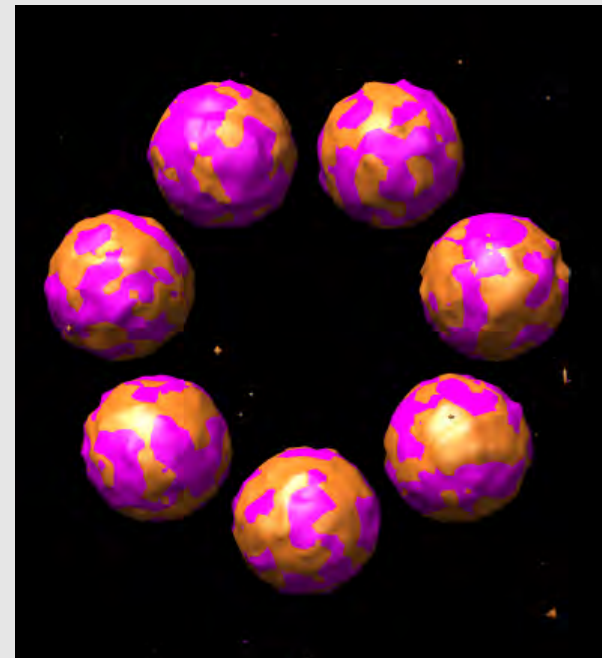
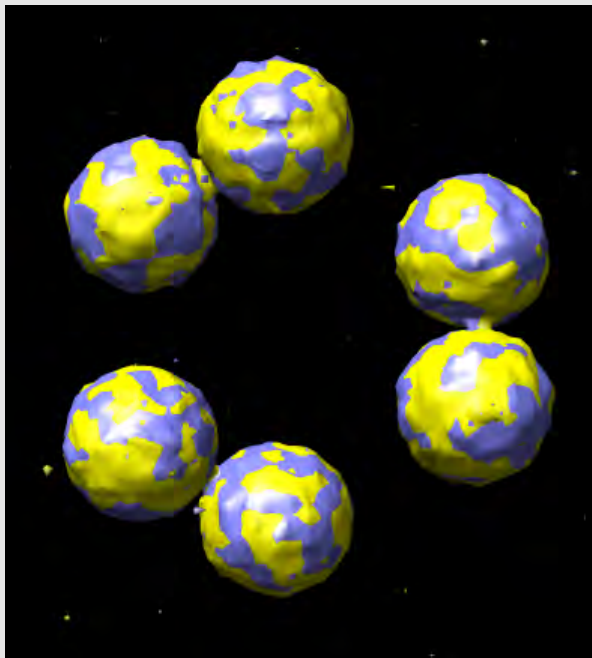


Center: 3D model obtained by reconstructing from heterogeneous projection set that contains aligned projection images of objects S6x and S7.

Left, Right: 3D models obtained by reconstructing from perfectly classified aligned projection images of objects S6x and S7.

Evaluation

Experiments with Aligned Projection Images (50:50)



Differences between 3D models obtained by reconstructing from perfectly classified aligned projection images of objects S6x, S7 and corresponding 3D models obtained by reconstructing from these images classified by the proposed method.

Evaluation

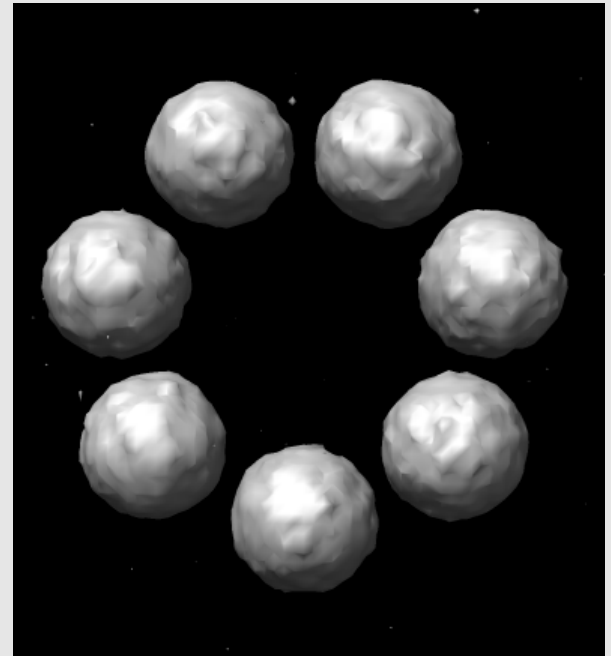
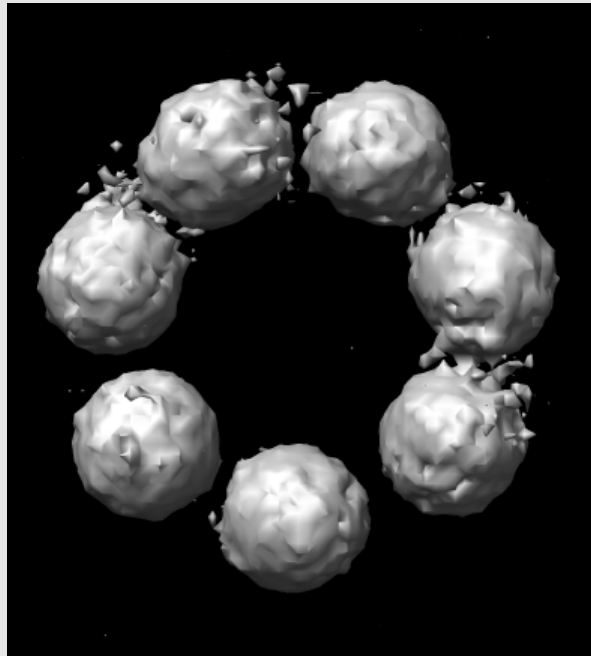
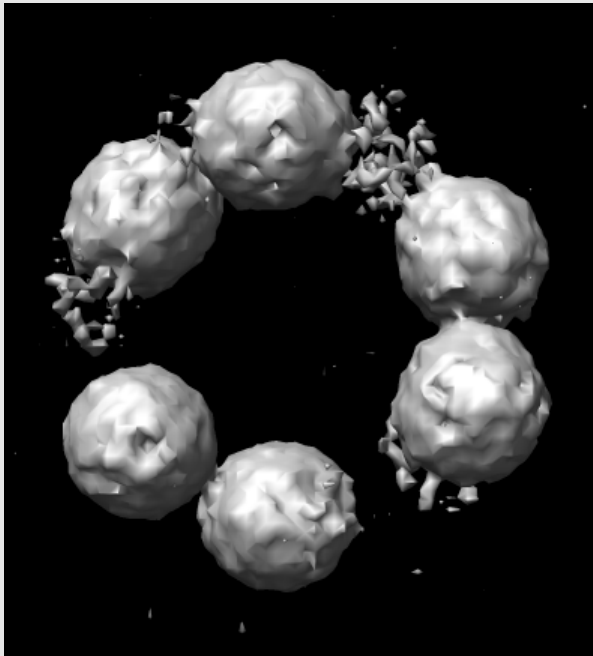
Experiments with Aligned Projection Images (50:50)

Projections of object	No of projections assigned to	
	Class 1	Class 2
S6x	33	2467
S7	2499	1

Example of the results from the two-class classification experiments with conformation representation ratio 50:50.

Evaluation

Experiments with Aligned Projection Images (35:65)

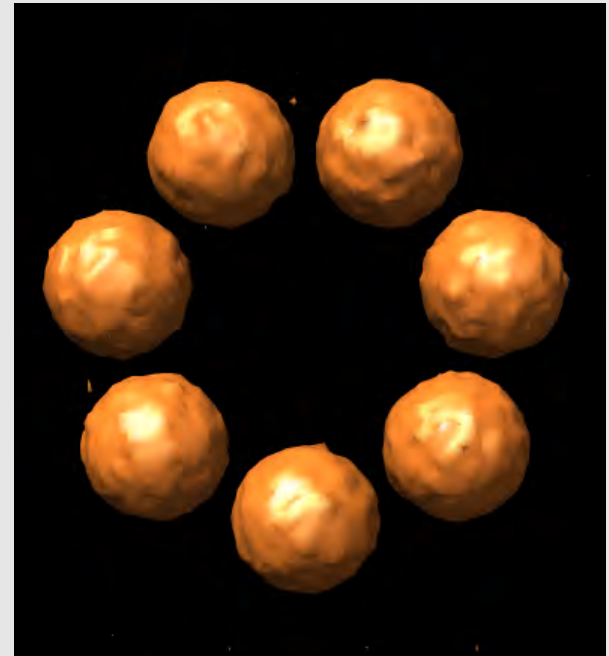
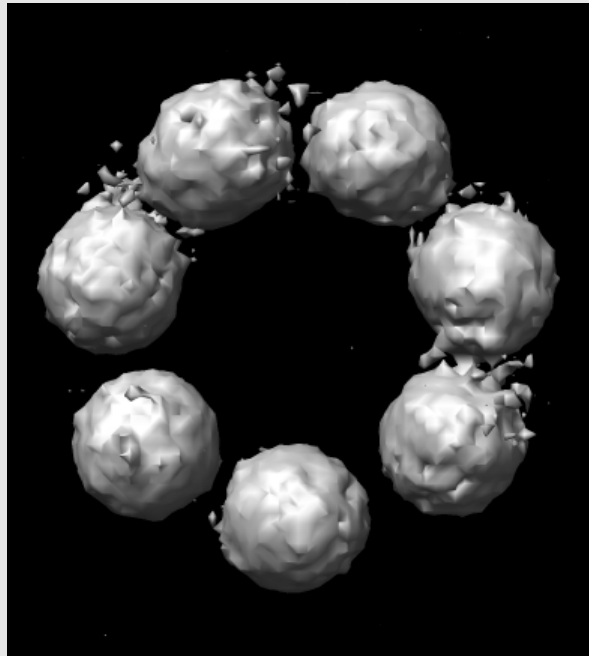
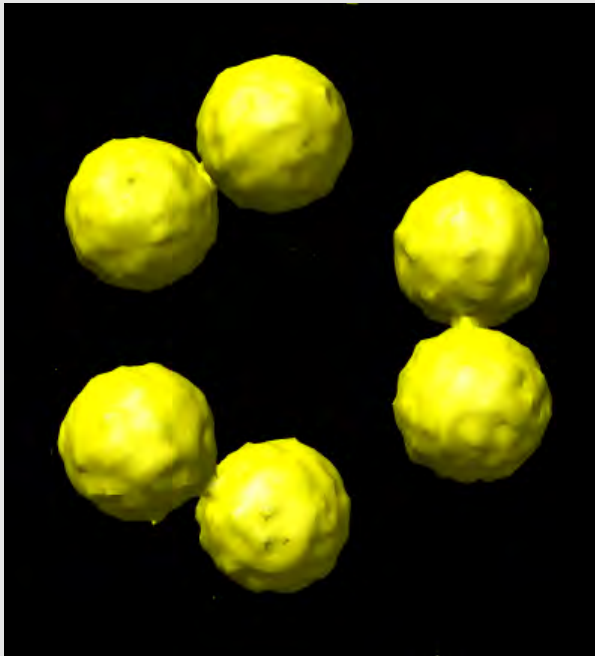


Center: 3D model obtained by reconstructing from heterogeneous projection set that contains aligned projection images of objects S6x and S7.

Left, Right: 3D models obtained by reconstructing from the aligned projection images of objects S6x, S7 classified by the proposed method.

Evaluation

Experiments with Aligned Projection Images (35:65)



Center: 3D model obtained by reconstructing from heterogeneous projection set that contains aligned projection images of objects S6x and S7.

Left, Right: 3D models obtained by reconstructing from perfectly classified aligned projection images of objects S6x and S7.

Evaluation

Experiments with Aligned Projection Images (35:65)

Projections of object	No of projections assigned to	
	Class 1	Class 2
S6x	0	1750
S7	2559	691

Example of the results from the two-class classification experiments with conformation representation ratio 35:65.

Evaluation

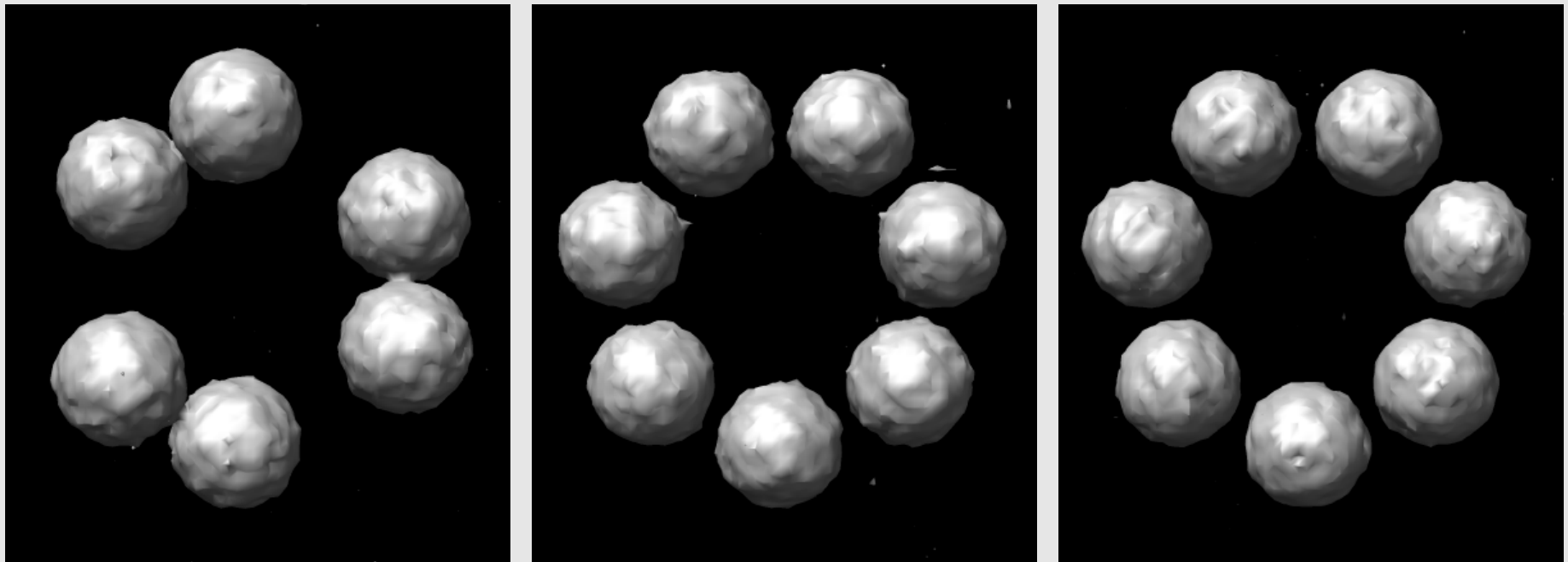
Experiments with Aligned Projection Images (35:65)

Projections of object	No of projections assigned to		
	Class 1	Class 2	Class 3
S6x	18	95	1637
S7	1674	1575	1

Example of the results from the three-class classification experiment with conformation representation ratio 35:65.

Evaluation

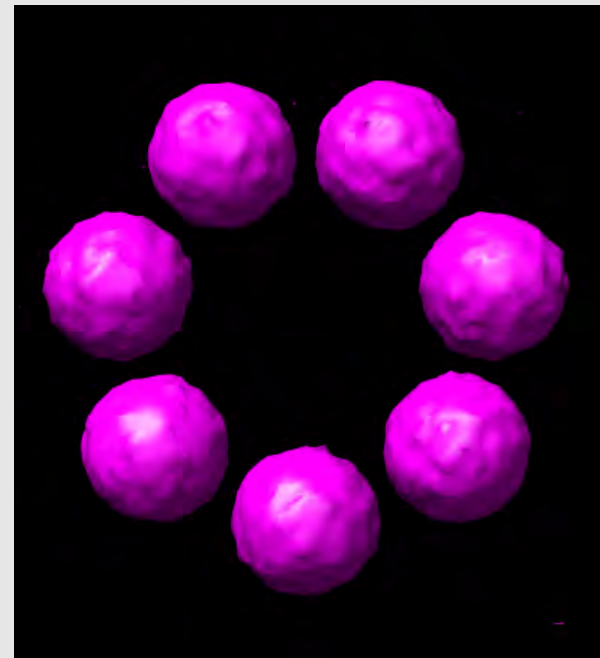
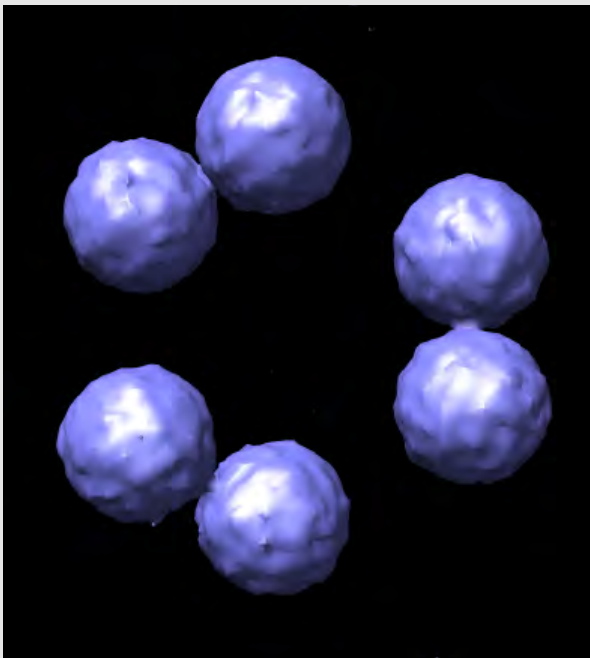
Experiments with Aligned Projection Images (35:65)



3D models obtained by reconstructing from the aligned projection images of objects S6x, S7 classified by the proposed method into three classes.

Evaluation

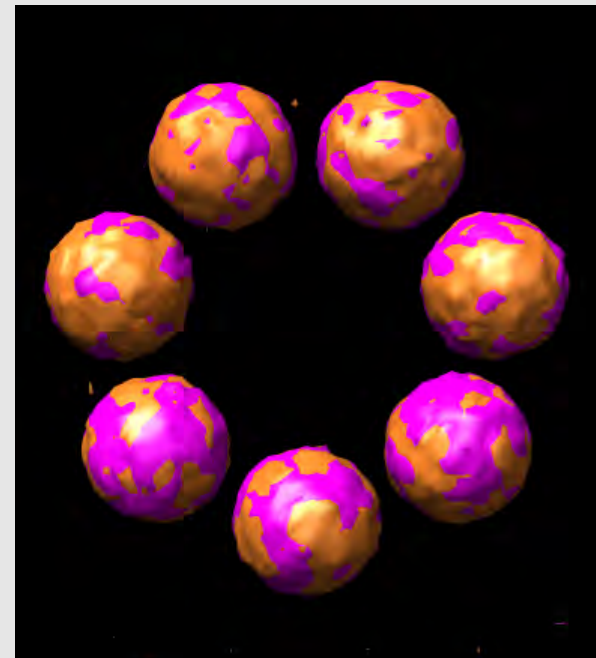
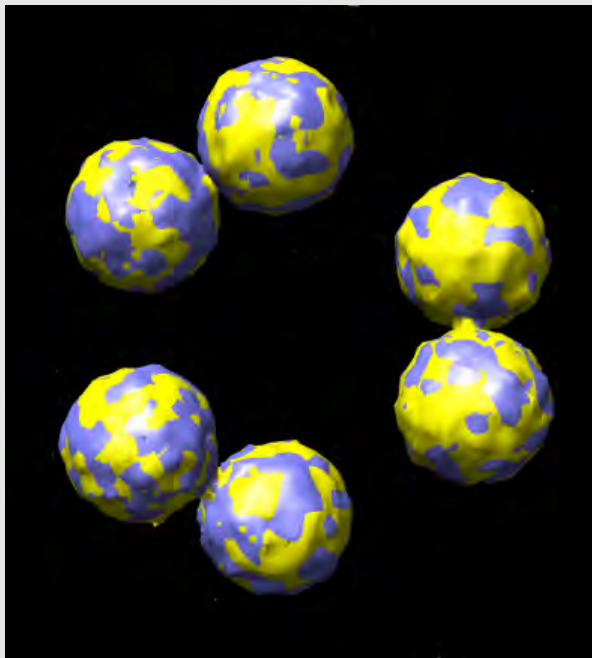
Experiments with Aligned Projection Images (35:65)



3D models obtained by reconstructing from the aligned projection images of objects S6x, S7 classified by the proposed method into three classes. The classes corresponding to the same object were merged.

Evaluation

Experiments with Aligned Projection Images (35:65)



Differences between 3D models obtained by reconstructing from perfectly classified aligned projection images of objects S6x, S7 and corresponding 3D models obtained by reconstructing from these images classified by the proposed method.

Evaluation

Experiments with Aligned Projection Images (20:80)

Projections of object	No of projections assigned to	
	Class 1	Class 2
S6x	8	992
S7	2535	1465

Example of the results from the two-class classification experiments with conformation representation ratio 20:80.

Evaluation

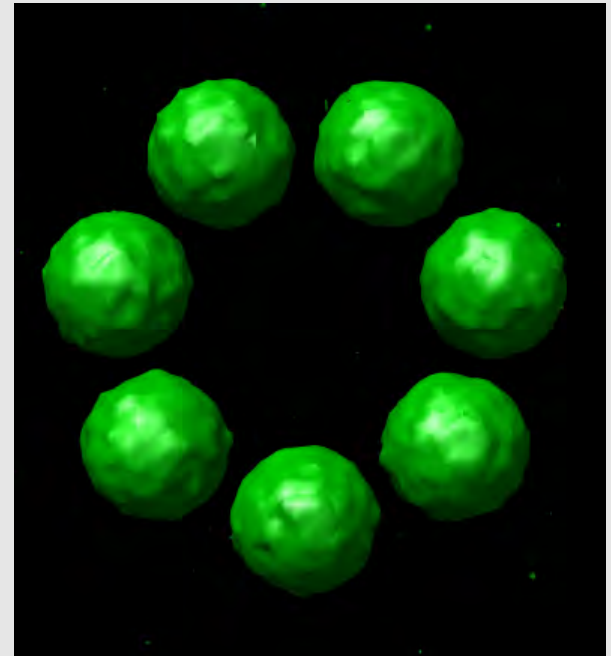
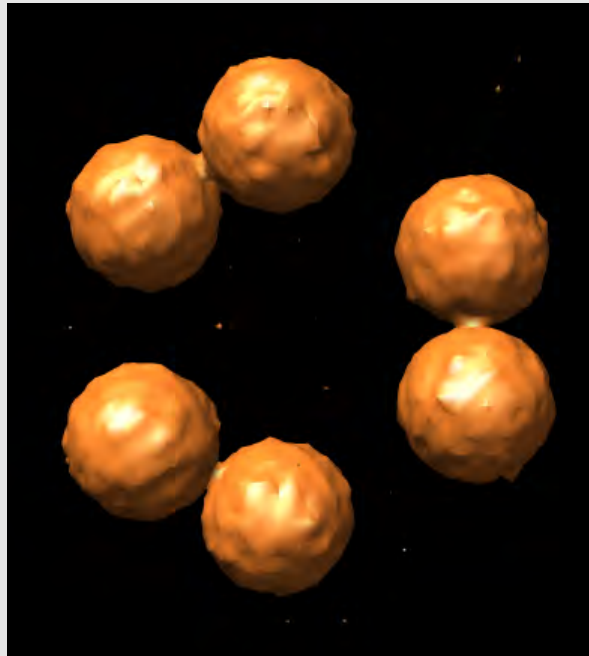
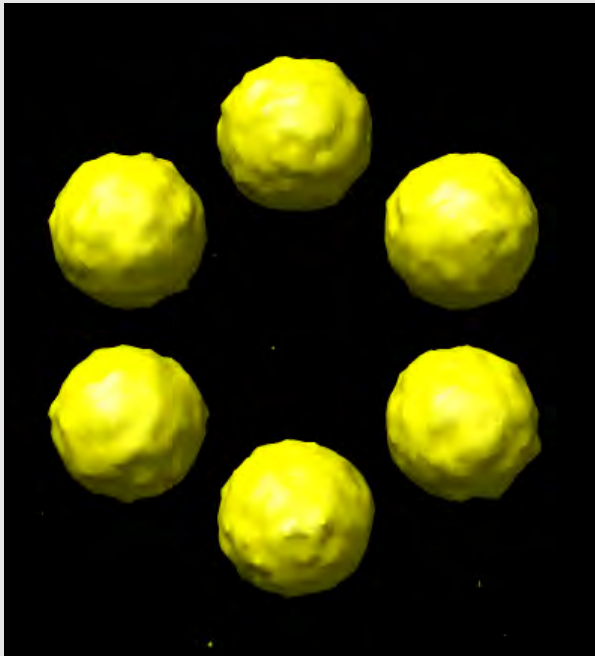
Experiments with Aligned Projection Images (20:80)

Projections of object	No of projections assigned to				
	Class 1	Class 2	Class 3	Class 4	Class 5
S6x	958	1	3	35	3
S7	9	1015	1004	961	1011

Example of the results from the five-class classification experiment with conformation representation ratio 20:80.

Evaluation

Experiments with Aligned Projection Images (33:33:33)



3D models obtained by reconstructing from the aligned projection images of objects S6, S6x, S7 classified by the proposed method into three classes.

Evaluation

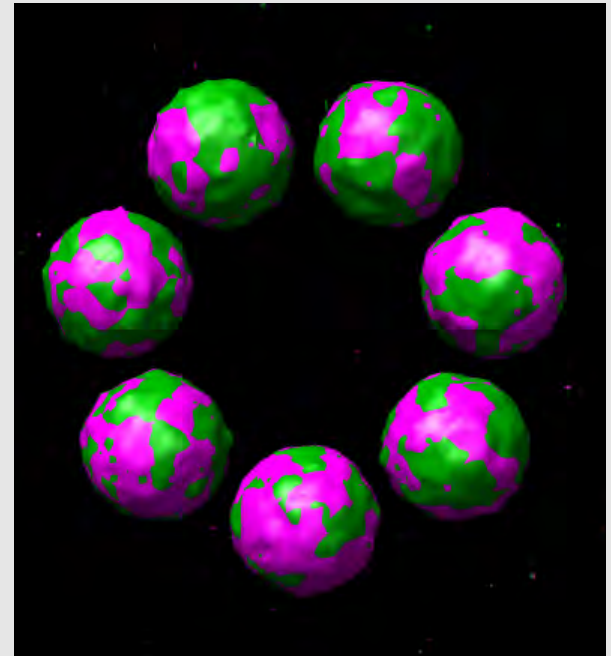
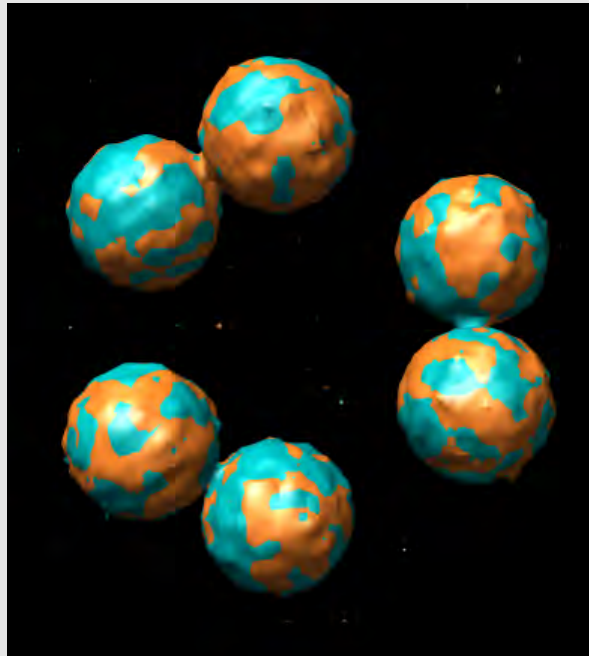
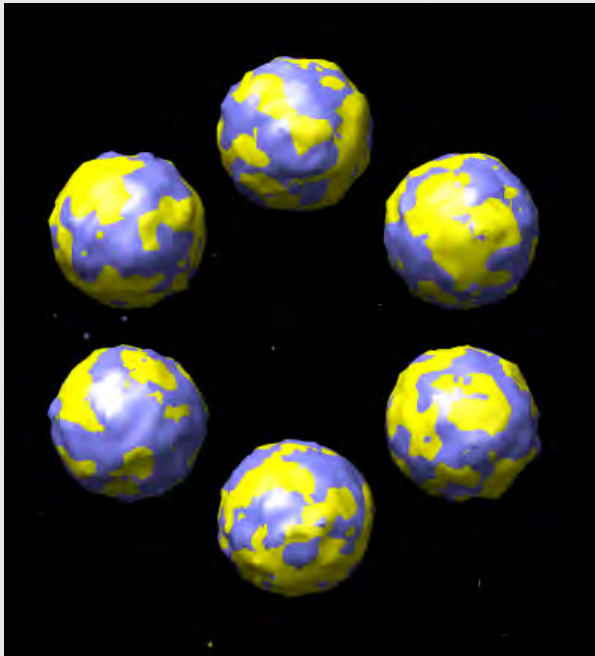
Experiments with Aligned Projection Images (33:33:33)

Projections of object	No of projections assigned to		
	Class 1	Class 2	Class 3
S6	24	1637	6
S6x	7	29	1631
S7	1654	12	0

Example of the results from the three-class classification experiment with three equally-represented conformations.

Evaluation

Experiments with Aligned Projection Images (33:33:33)



Differences between 3D models obtained by reconstructing from perfectly classified aligned projection images of objects S6, S6x, S7 and corresponding 3D models obtained by reconstructing from these images classified by the proposed method.

Evaluation

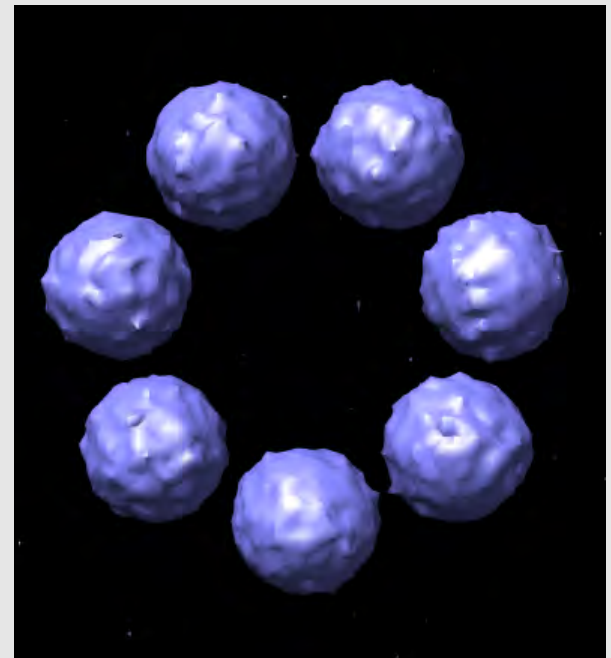
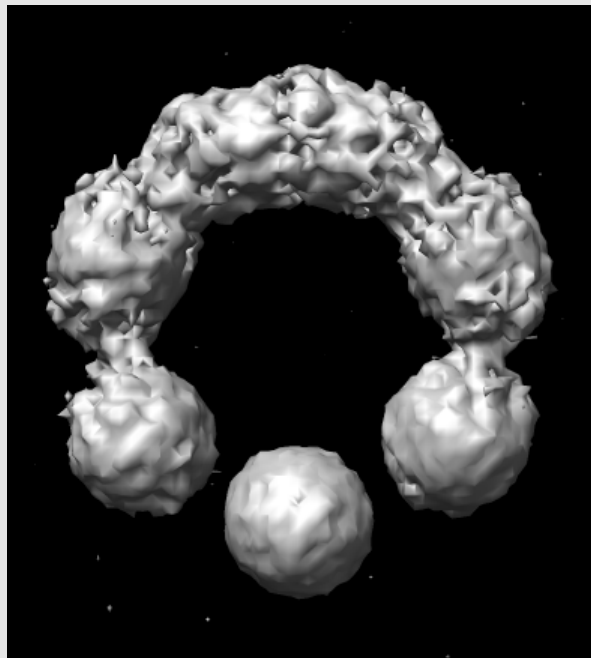
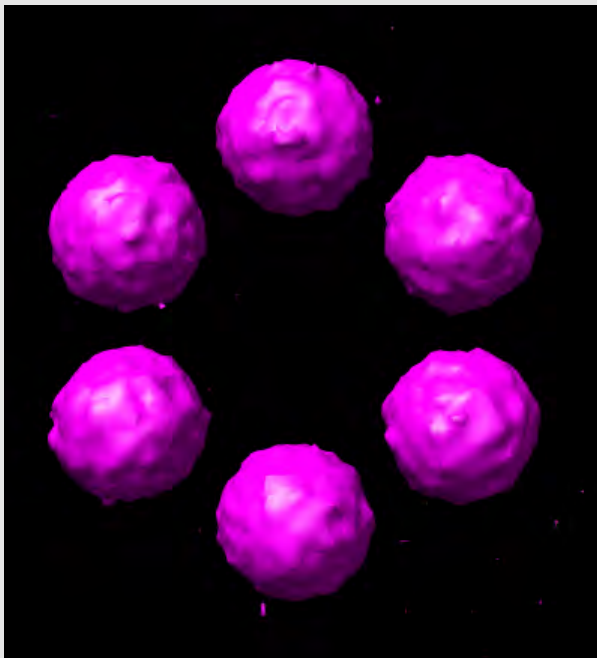
Experiments with Misaligned Projection Images

Datasets

- Randomly selected projections of 2 or 3 objects (S6, S6x, S7)
- Representation ratios: 50:50, 35:65, 20:80, 33:33:33
- SNR = 0.1
- Misaligned images

Evaluation

Experiments with Misaligned Projection Images (50:50)

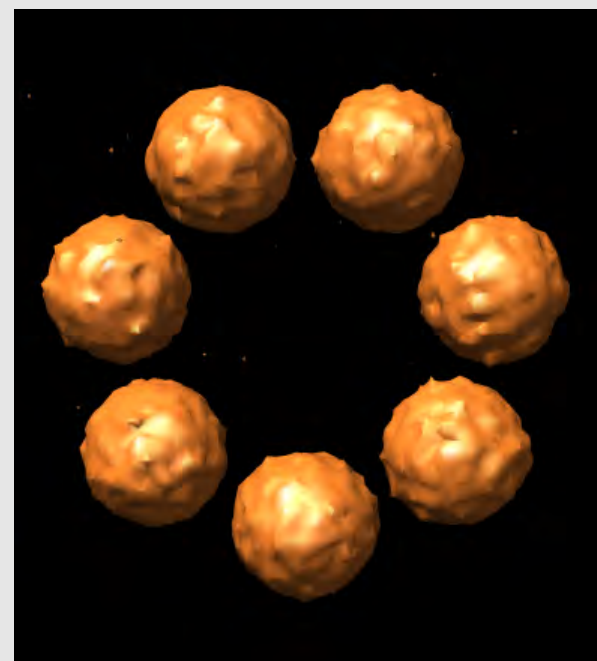
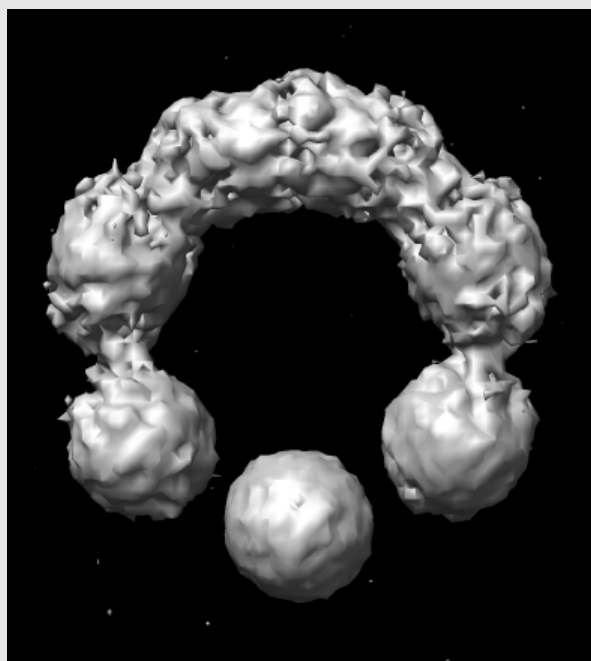
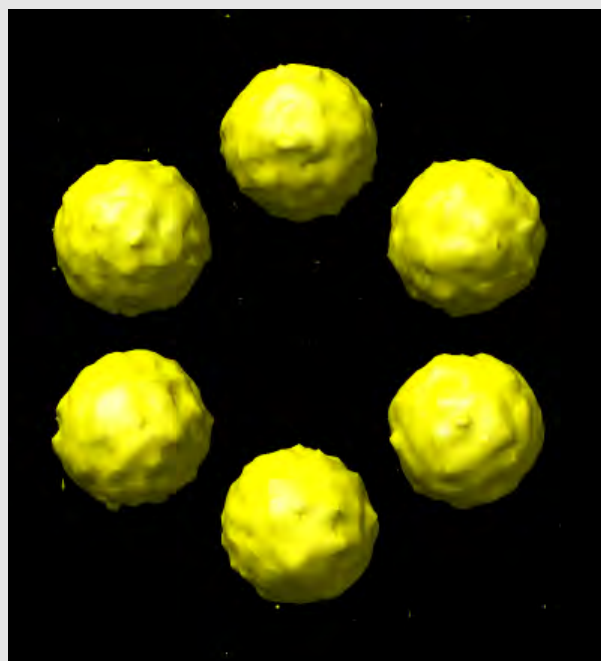


Center: 3D model obtained by reconstructing from heterogeneous projection set that contains misaligned projection images of objects S6 and S7.

Left, Right: 3D models obtained by reconstructing from the misaligned projection images of objects S6, S7 classified by the proposed method.

Evaluation

Experiments with Misaligned Projection Images (50:50)

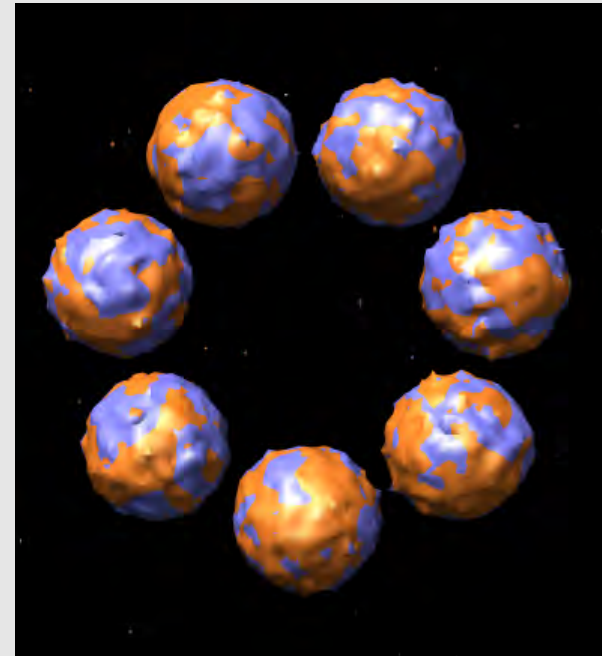
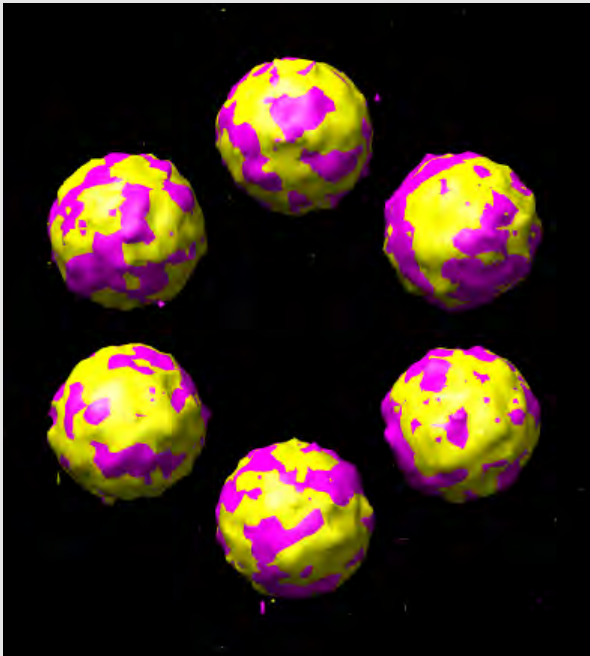


Center: 3D model obtained by reconstructing from heterogeneous projection set that contains misaligned projection images of objects S6 and S7.

Left, Right: 3D models obtained by reconstructing from perfectly classified misaligned projection images of objects S6 and S7.

Evaluation

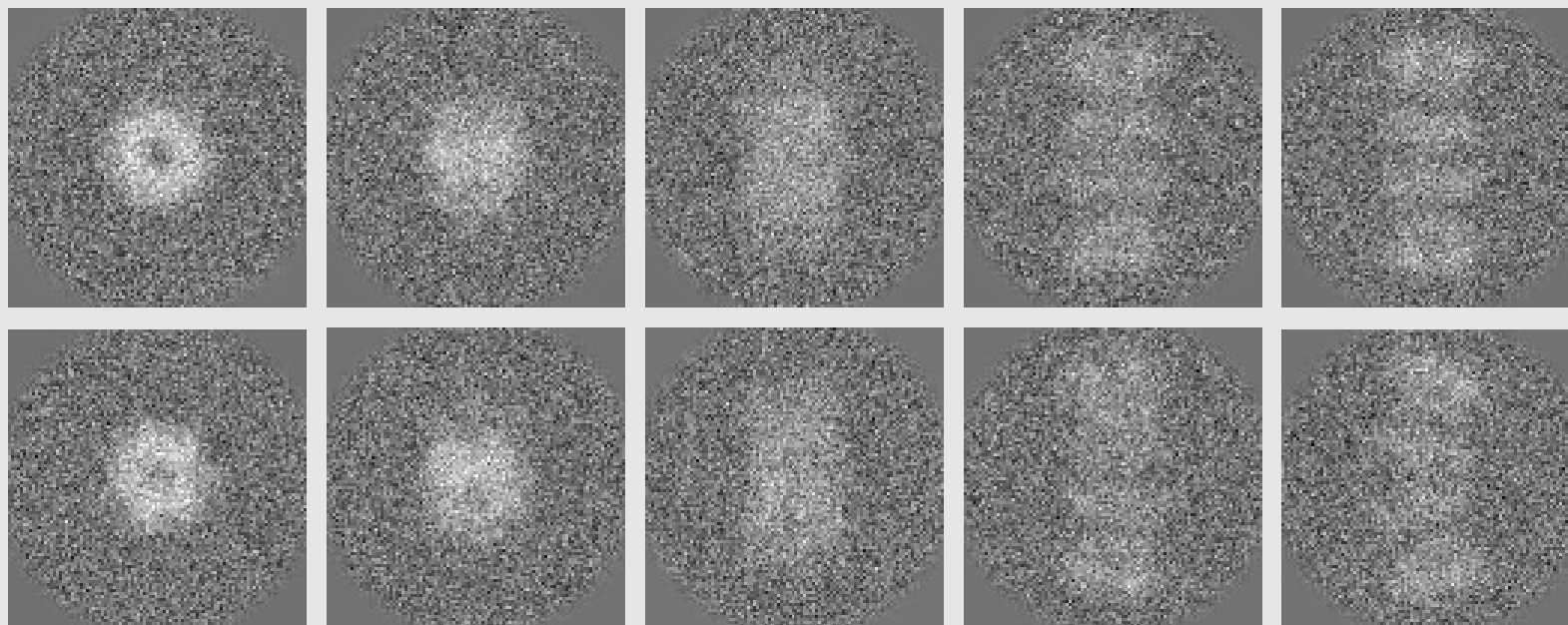
Experiments with Misaligned Projection Images (50:50)



Differences between 3D models obtained by reconstructing from perfectly classified misaligned projection images of objects S6, S7 and corresponding 3D models obtained by reconstructing from these images classified by the proposed method.

Evaluation

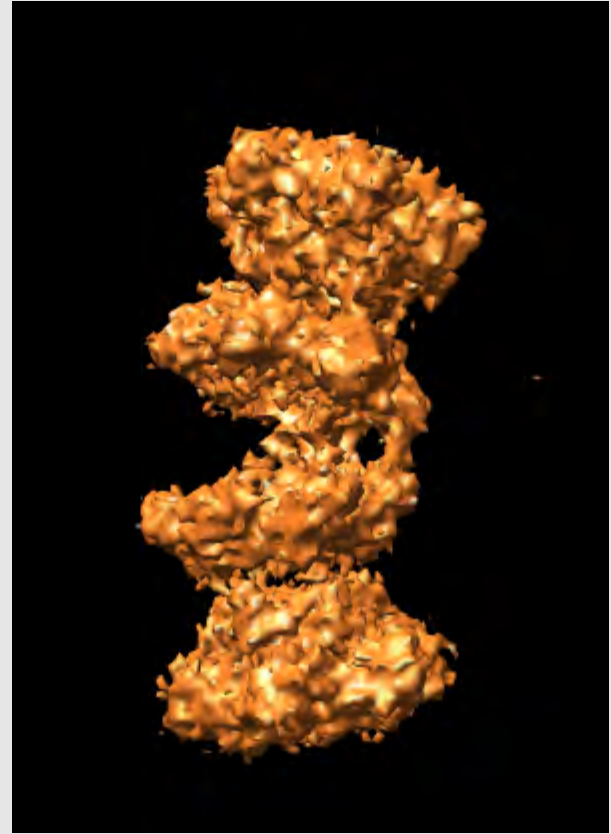
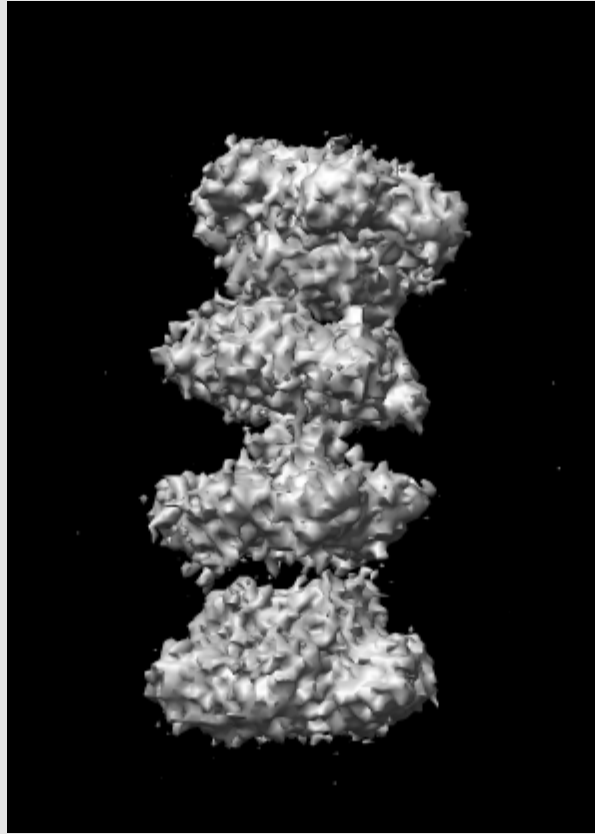
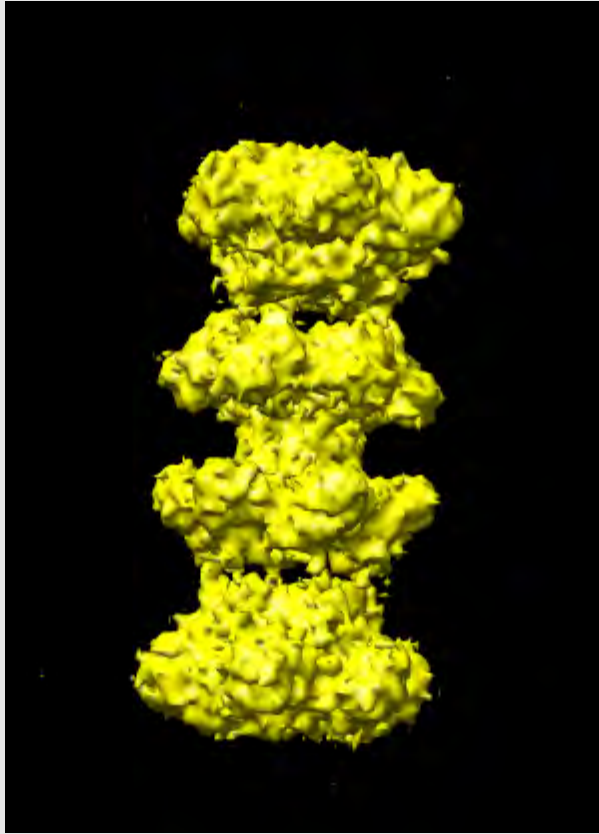
A Case Study Involving Externally Obtained Projection Data



Examples of Simian Virus 40 large T-antigen projection images.

Evaluation

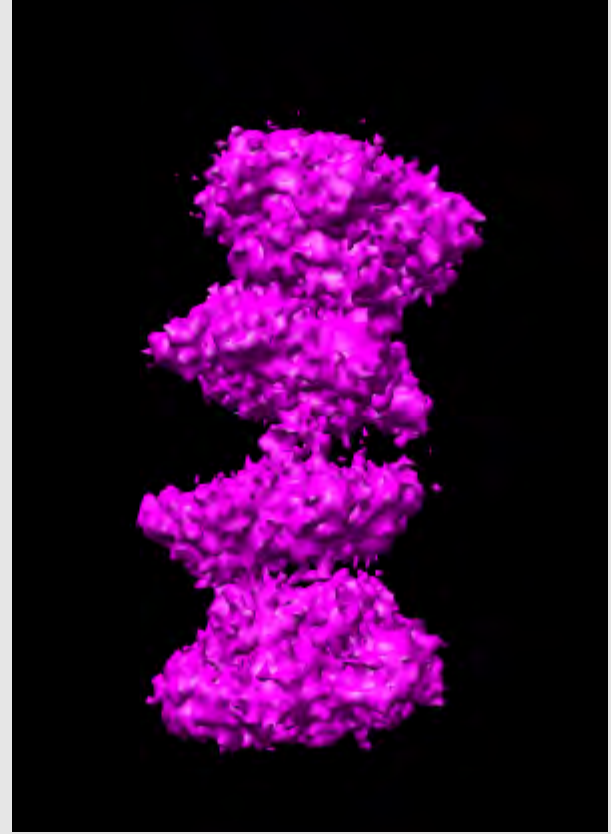
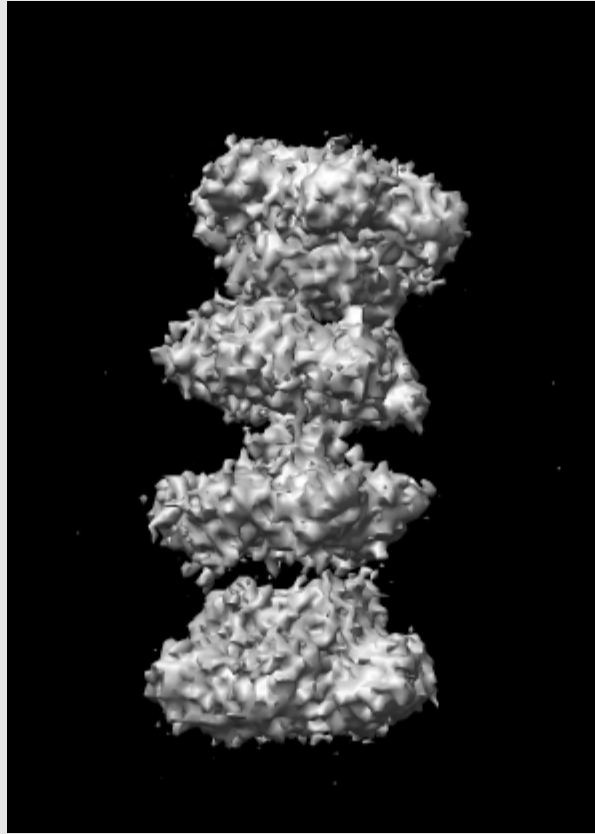
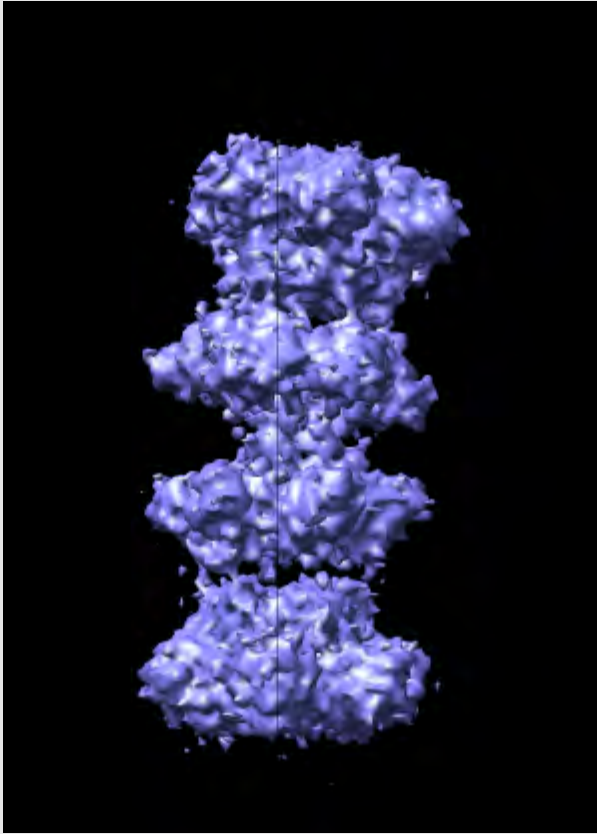
A Case Study Involving Externally Obtained Projection Data



Center: 3D model obtained by reconstructing from heterogeneous projection set.
Left, Right: 3D models obtained by reconstructing from perfectly classified projection images.

Evaluation

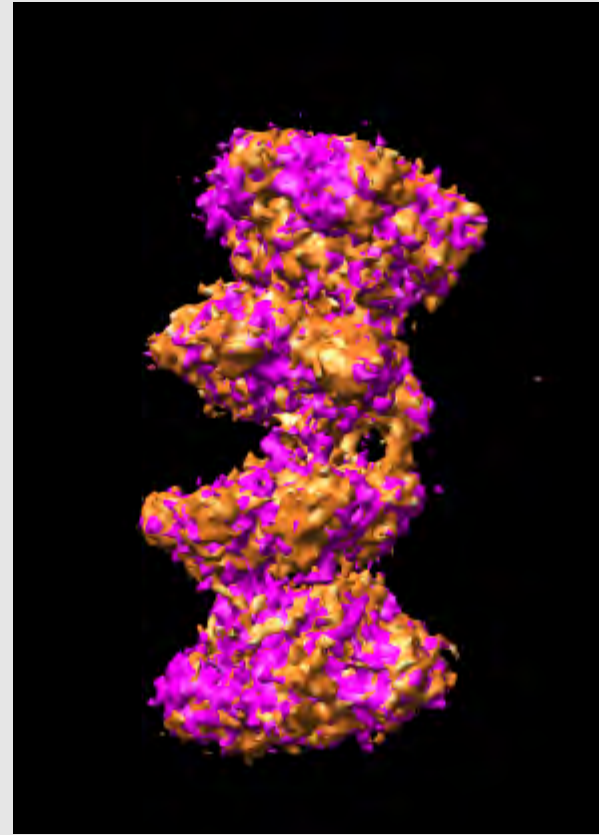
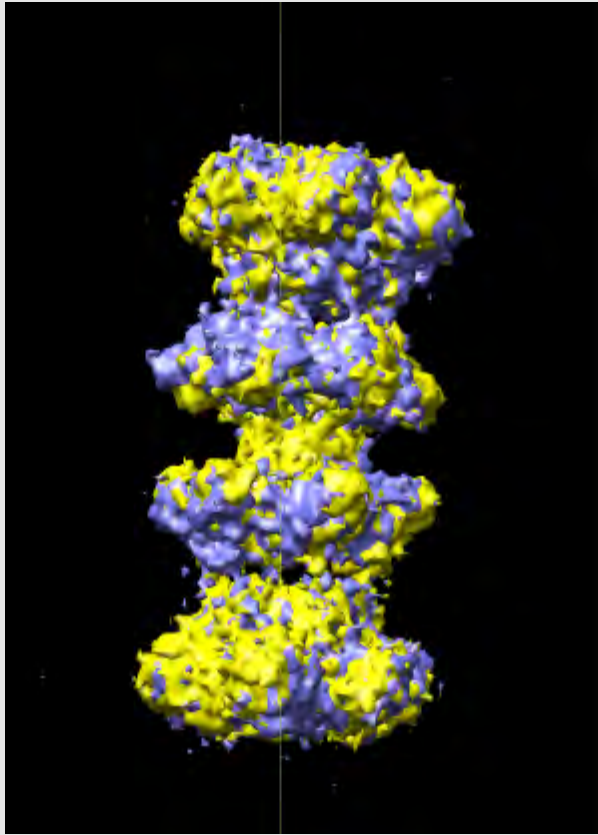
A Case Study Involving Externally Obtained Projection Data



Center: 3D model obtained by reconstructing from heterogeneous projection set.
Left, Right: 3D models obtained by reconstructing from the projection images classified by the proposed method.

Evaluation

A Case Study Involving Externally Obtained Projection Data



Differences between 3D models obtained by reconstructing from perfectly classified projection images and corresponding 3D models obtained by reconstructing from the images classified by the proposed method.

Conclusions

Conclusions

Contributions

- Proposed an optimization based unsupervised classification procedure to separate homogeneous subsets of heterogeneous projection sets was proposed.
- Demonstrated that incorporated into a heterogeneous reconstruction procedure, proposed method produces representative 3D models of various conformations represented in heterogeneous projection set.
- Proposed a new dissimilarity measure, specifically designed to deal with 2D projections of 3D objects.
- Constructed an algorithm that efficiently finds good (from the classification perspective) approximate Max k-Cuts for graphs that represent instances of heterogeneous projection sets.