From Cliques to Equilibria

The Dominant-Set Approach to Clustering

Marcello Pelillo

University of Venice, Italy

Ca' Foscari University of Venice

5th Winter School on Data Analytics

November 20–22, 2020

European Centre for Living Technology

The clustering problem

Given:

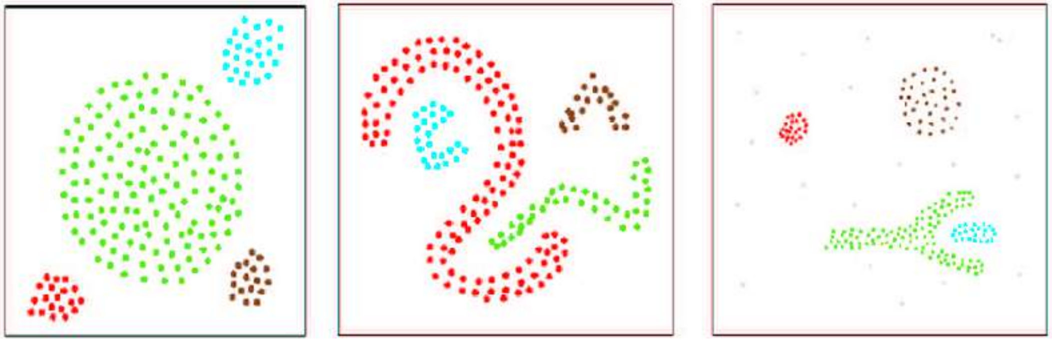
✓a set of n "objects"

✓an $n \times n$ matrix A of pairwise similarities

} = an edge-weighted graph G

Goal: <u>*Partition*</u> the vertices of the G into maximally homogeneous groups (i.e., clusters).

Usual assumption: symmetric and pairwise similarities (G is an undirected graph)



Applications

Clustering problems abound in many areas of computer science and engineering.

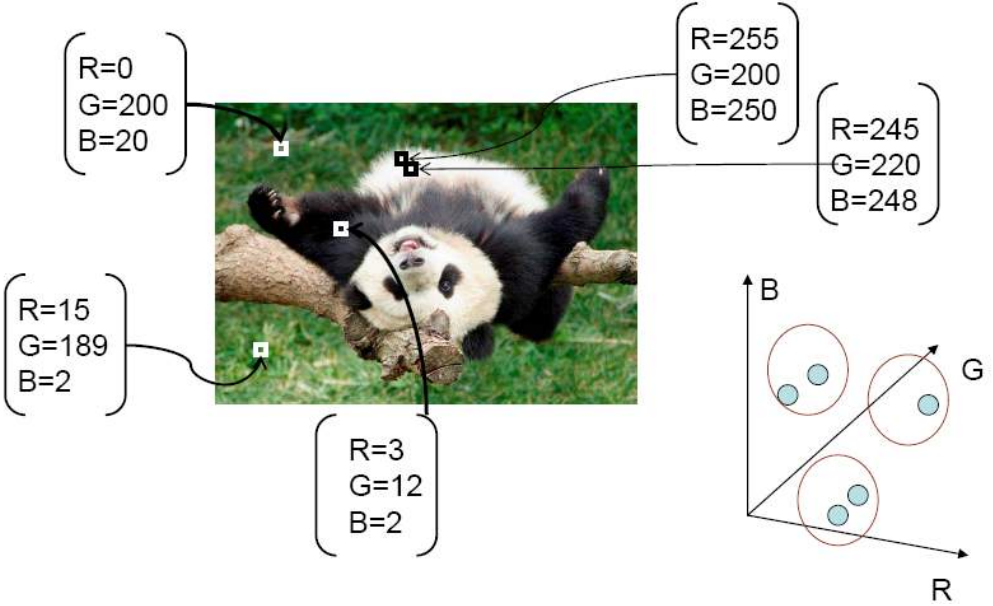
A short list of applications domains:

- Image processing and computer vision
- Computational biology and bioinformatics
- Information retrieval
- Document analysis
- Medical image analysis
- Data mining
- Signal processing
- …

For a review see, e.g., A. K. Jain, "Data clustering: 50 years beyond K-means," Pattern Recognition Letters 31(8):651-666, 2010.

Image Segmentation as Clustering

- Cluster similar pixels (features) together



Source: K. Grauman

What is a Cluster?

No universally accepted (formal) definition of a “cluster” but, informally, a cluster should satisfy two criteria:

Internal criterion

all “objects” inside a cluster should be highly similar to each other

External criterion

all “objects” outside a cluster should be highly dissimilar to the ones inside

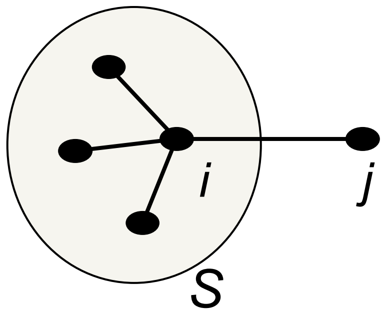
How to formalize these criteria?

Basic Definitions

Let $S \subseteq V$ be a non-empty subset of vertices, and $i \in S$.

The **(average) weighted degree** of i w.r.t. S is defined as:

$$\mathrm{awdeg}_S(i) = \frac{1}{|S|} \sum_{j \in S} a_{ij}$$



Moreover, if $j \notin S$, we define:

$$\phi_S(i,j) = a_{ij} - \mathrm{awdeg}_S(i)$$

Intuitively, $\Phi_S(i,j)$ measures the similarity between vertices j and i, with respect to the (average) similarity between vertex i and its neighbors in S.

Basic Definitions

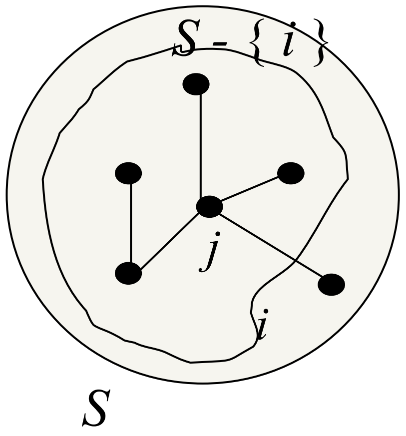
Let $S \subseteq V$ be a non-empty subset of vertices, and $i \in S$.

The **weight** of i w.r.t. S is defined as:

$$w_S(i) = \begin{cases} 1 & \text{if } |S| = 1 \\ \displaystyle\sum_{j \in S - \{i\}} \phi_{S-\{i\}}(j,i)\, w_{S-\{i\}}(j) & \text{otherwise} \end{cases}$$

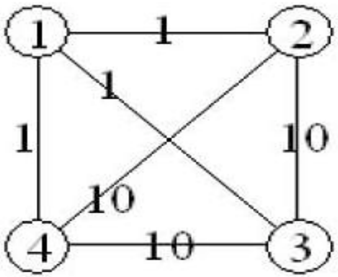
Further, the **total weight** of S is defined as:

$$W(S) = \sum_{i \in S} w_S(i)$$

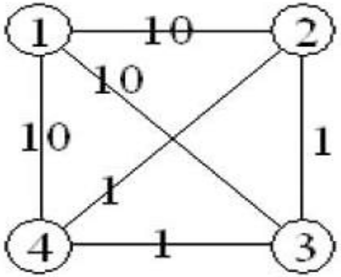


Interpretation

Intuitively, $w_S(i)$ gives us a measure of the overall (relative) similarity between vertex i and the vertices of S \ {i} with respect to the overall similarity among the vertices in S \ {i}.



$w_{\{1,2,3,4\}}(1) < 0$



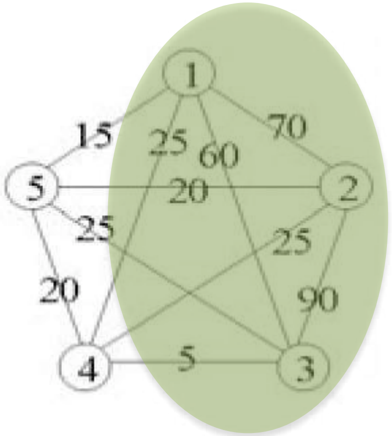
$w_{\{1,2,3,4\}}(1) > 0$

Dominant Sets

Let $S \subseteq V$ be a subset of vertices of a graph G.

S is said to be a **dominant set** if:

1. $w_S(i) > 0$, for all $i \in S$ (internal homogeneity)
2. $w_{S \cup \{i\}}(i) < 0$, for all $i \notin S$ (external homogeneity)



M. Pavan and M. Pelillo. Dominant sets and pairwise clustering (PAMI 2007)

The Many Facets of Dominant Sets

Dominant sets have intriguing connections wth:

- **Game theory**
  Nash equilibria of “clustering games”
- **Optimization theory**
  Local maximizers of (continuous) quadratic problems
- **Graph theory**
  Maximal cliques
- **Dynamical systems theory**
  Stable attractors of evolutionary game dynamics

See Rota Bulò and Pelillo (EJOR 2017) for a a review

Some Cornerstones in Game Theory

1921−1928: Emile Borel and John von Neumann give the first modern formulation of a mixed strategy along with the idea of finding minimax solutions of normal-form games.

1944, 1947: John von Neumann and Oskar Morgenstern publish *Theory of Games and Economic Behavior*.

1950−1953: In four papers John Nash made seminal contributions to both non-cooperative game theory and to bargaining theory.

1972−1982: John Maynard Smith applies game theory to biological problems thereby founding "evolutionary game theory."

late 1990's −: Development of algorithmic game theory …

Nobel prizes in economics awarded to game theorists:

1994: John Nash, John Harsanyi and Reinhard Selten: "for their pioneering analysis of equilibria in the theory of non-cooperative games"

2005: Robert Aumann and Thomas Schelling: "for having enhanced our understanding of conflict and cooperation through game-theory analysis"

2007: Leonid Hurwicz, Eric Maskin and Roger Myerson: "for having laid the foundations of mechanism design theory"

“Solving” a game

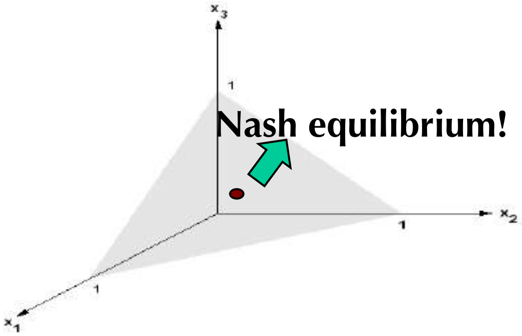
Nash equilibrium: no player has an incentive to deviate unilaterally from it.

		Player 2		
		Left	Middle	Right
Player 1	Top	3 , 1	2 , 3	10
	High	4 , 5	3 , 0	6 , 4
	Low	2 , 2	5 , 4	12 , 3
	Bottom	5 , 6	4 , 5	9 , 7

Nash equilibrium!

Mixed-straegy solutions

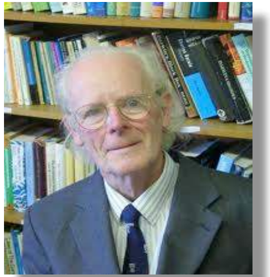
Mixed strategy = probability distribution $\mathbf{x}=(x_1,\ldots,x_n)^T$ over the set of "pure" strategies



No Nash ...ilibrium?

		You		
		Rock	Scissors	Paper
Me	Rock	0 , 0	1 , -1	-1 , 1
	Scissors	-1 , 1	0 , 0	1 , -1
	Paper	1 , -1	-1 , 1	0 , 0

Evolutionary Game Theory



"Paradoxically, it has turned out that game theory is more readily applied to biology than to the field of economic behaviour for which it was originally designed."

John Maynard Smith
Evolution and the Theory of Games (1982)

- ✓ A large population of individuals which compete for limited resources
- ✓ This kind of conflict is modeled as a two-player game, players being pairs of randomly selected population members
- ✓ Players do not behave "rationally" but act according to a pre-programmed behavioral pattern (pure strategy)
- ✓ Utility is measured in terms of Darwinian fitness, or reproductive success

Evolutionary Stable Strategy (ESS): a Nash equilibrium that is stable under small perturbations of the population structure.

Dominant Sets and ESS's

Consider the following "clustering game"

- ✓ There are 2 players (because we have pairwise affinities)
- ✓ The objects to be clustered are the pure strategies
- ✓ The (null-diagonal) payoff matrix coincides with the similarity matrix

Evolutionary stable strategies (ESS's) of the clustering game are in one-to-one correspondence with dominant sets.

Using Symmetric Affinities

Given a symmetric affinity matrix A, consider the following continuous quadratic optimization problem (QP):

$$\begin{array}{ll} \text{maximize} & f(\mathbf{x}) = \mathbf{x}' A \mathbf{x} \\ \text{subject to} & \mathbf{x} \in \Delta \end{array}$$

where Δ is the standard simplex (probability space).

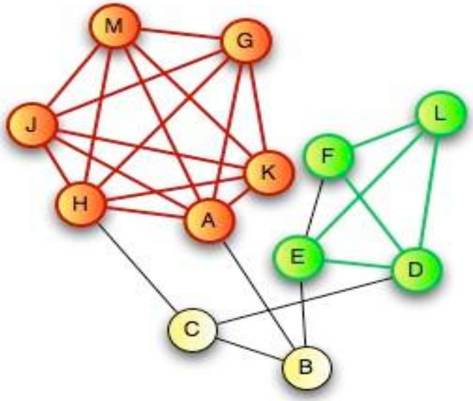
The function $f(x)$ provides a measure of cohesiveness of a cluster.

Dominant sets are in one-to-one correspondence to (strict) local solutions of QP

Using Binary (Symmetric) Affinities

Suppose the similarity matrix is a binary (0/1) matrix.

Given an unweighted undirected graph $G=(V, E)$:



A clique is a subset of mutually adjacent vertices

A maximal clique is a clique that is not contained in a larger one

ESS's are in one-to-one correspondence to maximal cliques of *G*

Note. Generalization of the Motzkin-Straus theorem (1965).

Replicator Dynamics

Let $x_i(t)$ the population share playing pure strategy i at time t. The **state** of the population at time t is: $x(t)= (x_1(t),\ldots,x_n(t))\in\Delta$.

Replicator dynamics are motivated by Darwin's principle of natural selection:

$$\frac{\dot{x}_i}{x_i} \propto \text{payoff of pure strategy i} - \text{average population payoff}$$

which yields:

$$\dot{x}_i = x_i\left[(Ax)_i - x^T Ax\right]$$

Theorem (Nachbar, 1990; Taylor and Jonker, 1978). A point $x\in\Delta$ is a Nash equilibrium if and only if x is the limit point of a replicator dynamics trajectory starting from the interior of Δ.

Furthermore, if $x\in\Delta$ is an ESS, then it is an asymptotically stable equilibrium point for the replicator dynamics.

Doubly-Symmetric Games

In a "doubly-symmetric" game, the payoff matrix A is symmetric ($A = A^T$).

Fundamental Theorem of Natural Selection (Losert and Akin, 1983). For any doubly symmetric game, the average population payoff $f(x) = x^T A x$ is strictly increasing along any non-constant trajectory of replicator dynamics, namely, $d/dt f(x(t)) \geq 0$ for all $t \geq 0$, with equality if and only if $x(t)$ is a stationary point.

Characterization of ESS's (Hofbauer and Sigmund, 1988)

For any doubly simmetric game with payoff matrix A, the following statements are equivalent:

a) $x \in \Delta^{ESS}$

b) $x \in \Delta$ is a strict local maximizer of $f(x) = x^T A x$ over the standard simplex Δ

c) $x \in \Delta$ is asymptotically stable in the replicator dynamics

Finding Dominant Sets

Discrete-time **replicator dynamics** are a popular and principled way to find dominant sets.

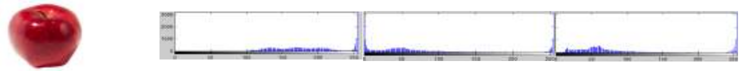
$$x_i(t+1) = x_i(t)\frac{A\big(x(t)\big)_i}{x(t)^T A x(t)}$$

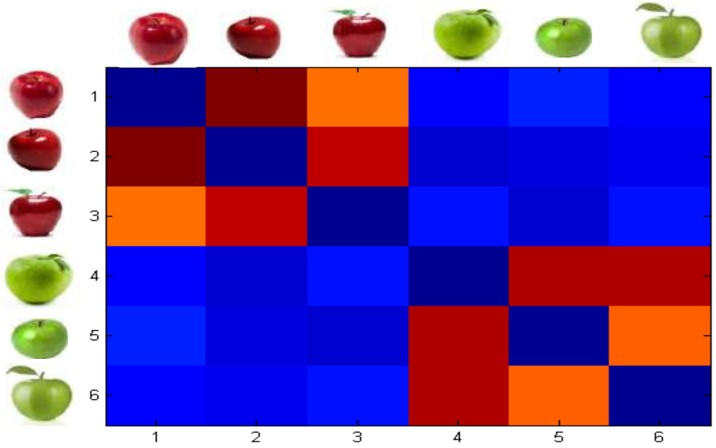
MATLAB implementation

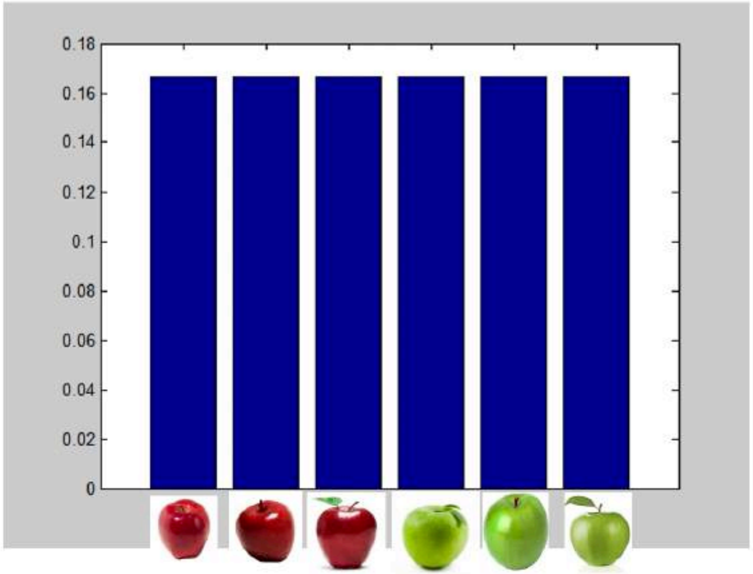
```
distance=inf;
while distance>epsilon
    old_x=x;
    x = x.*(A*x);
    x = x./sum(x);
    distance=pdist([x,old_x]');
end
```

Faster dynamics available!
(See Rota Bulò and Pelillo, 2017)

A Toy Example







Measuring Cluster Membership

The components of the converged vector **x** give us a measure of the participation of the corresponding vertices in the cluster, while the value of the objective function measures the cluster's cohesiveness.

Useful for ranking the elements in the cluster!

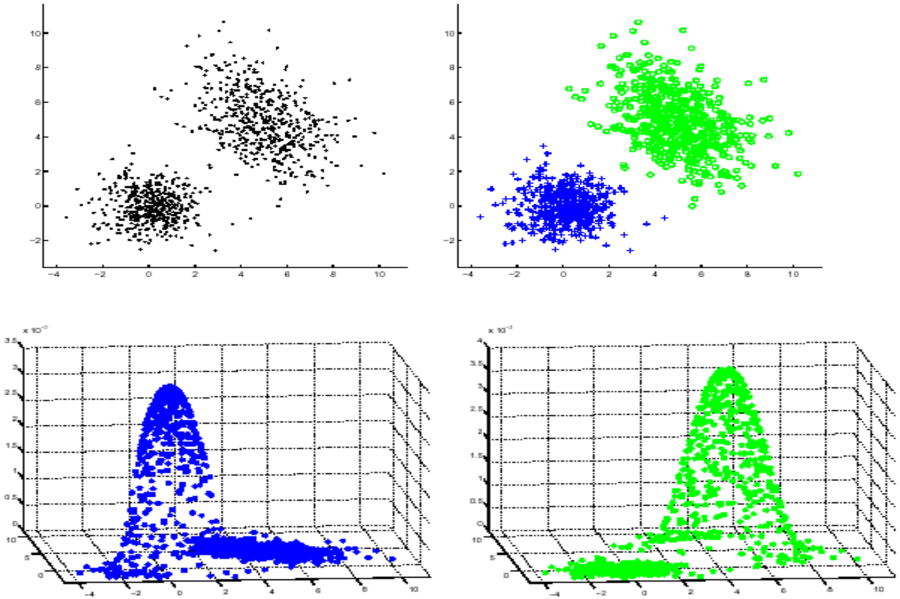
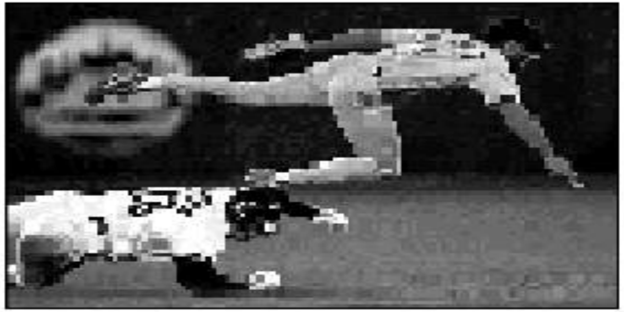
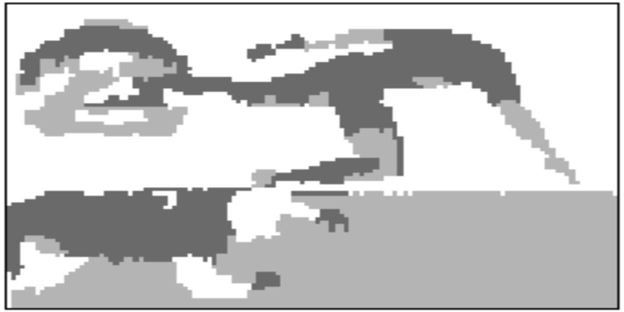
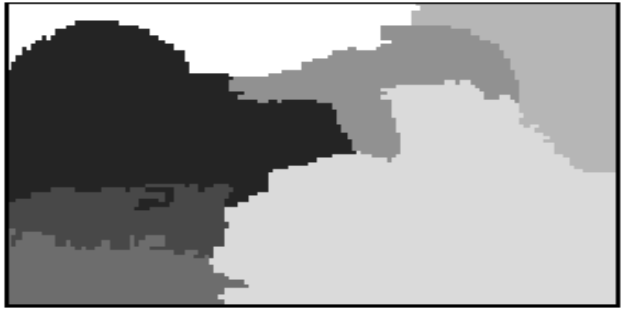


Image Segmentation



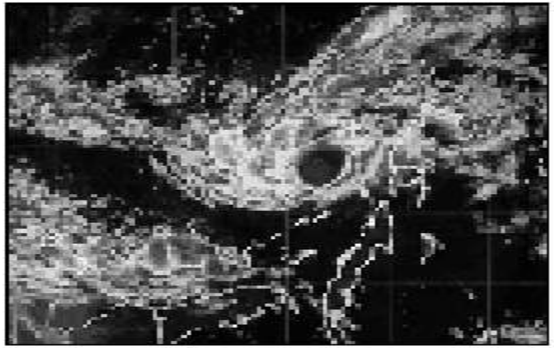


Dominant sets

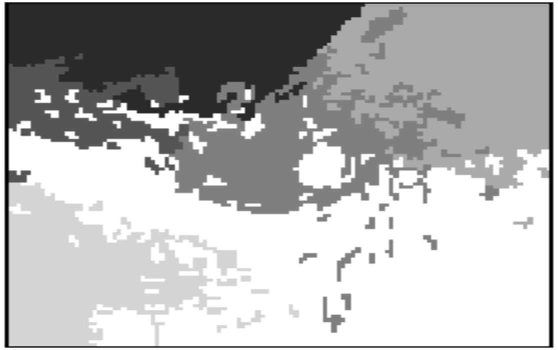


Ncut

Image Segmentation

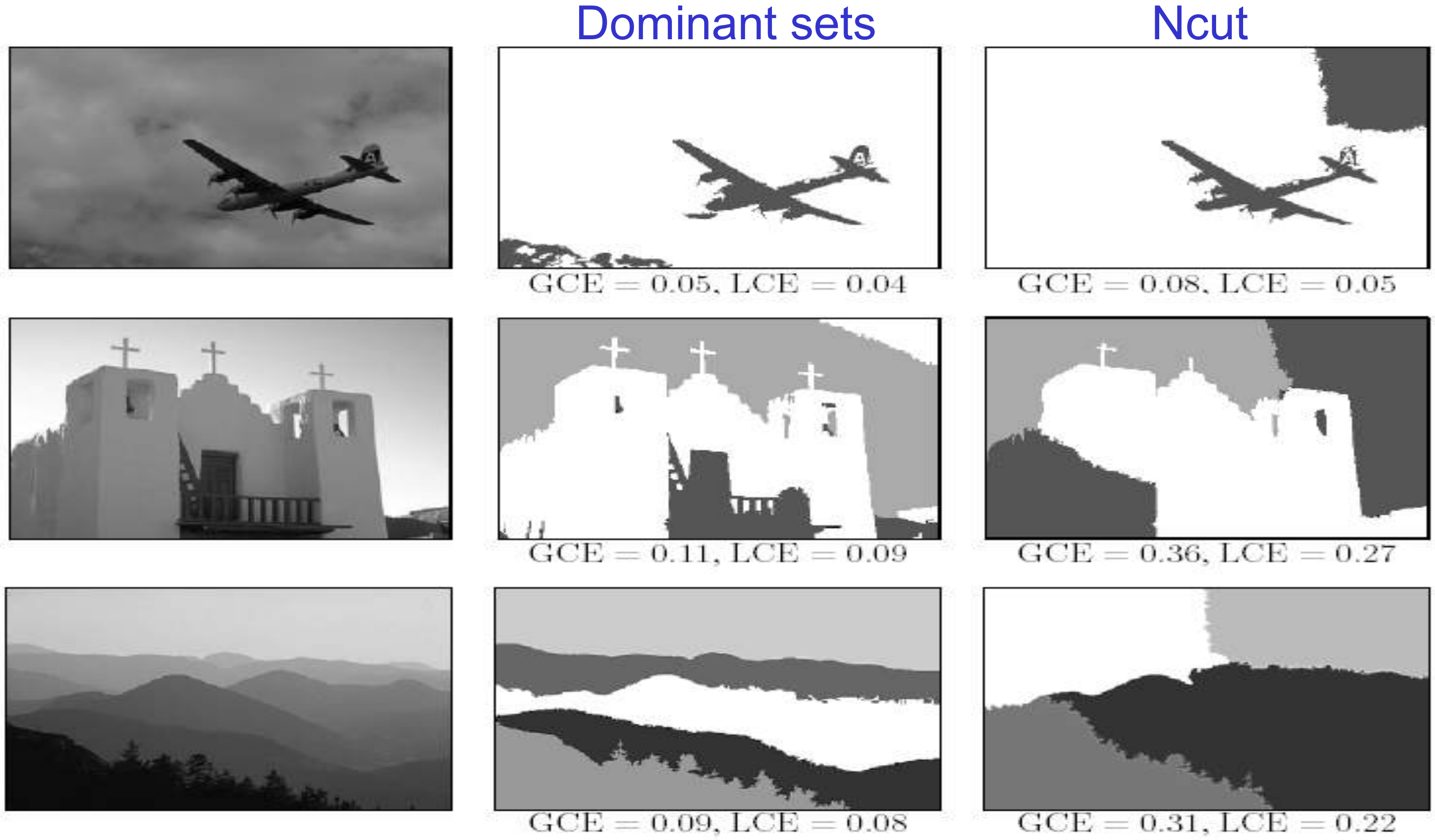


Dominant sets



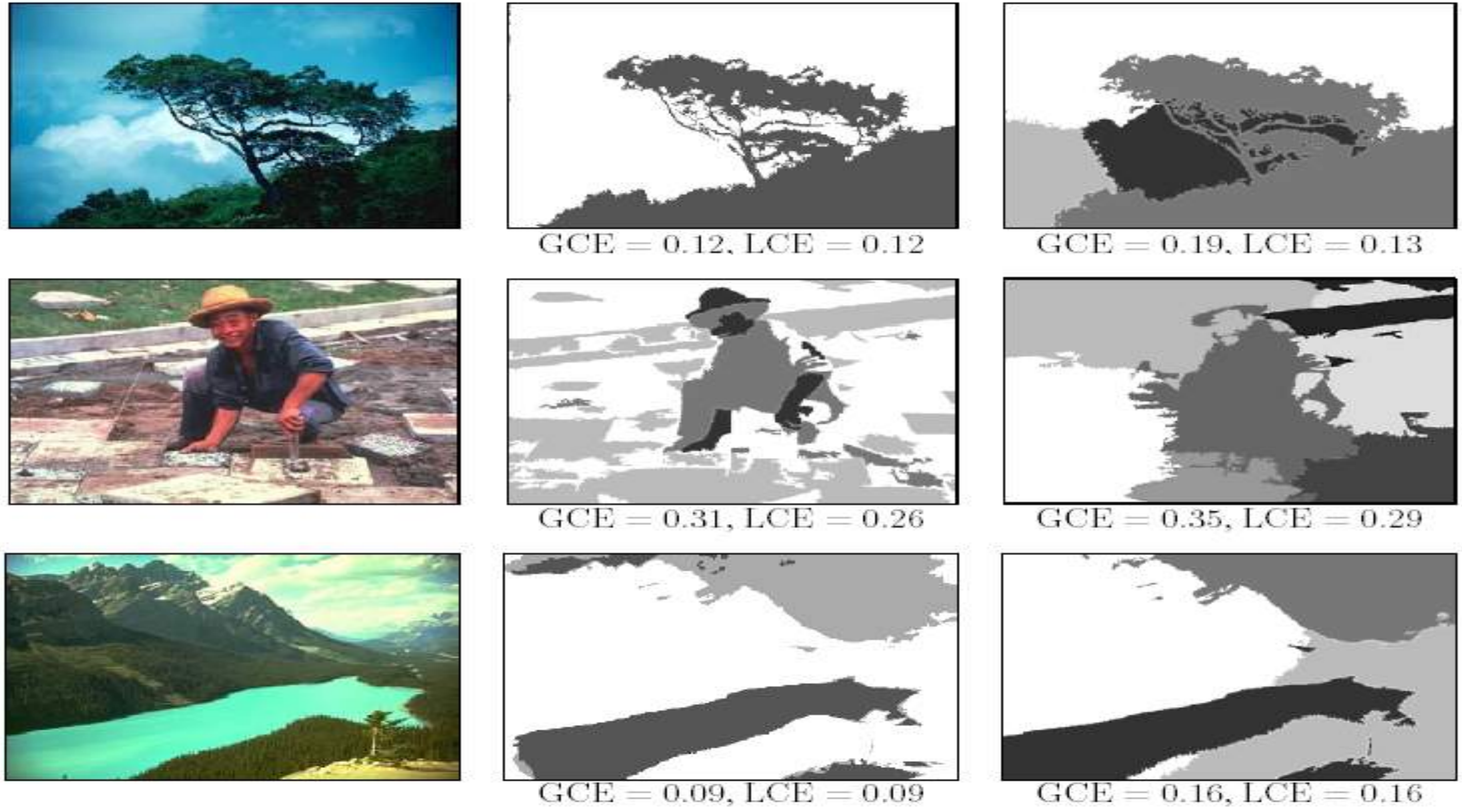
Ncut

Results on Berkeley Dataset

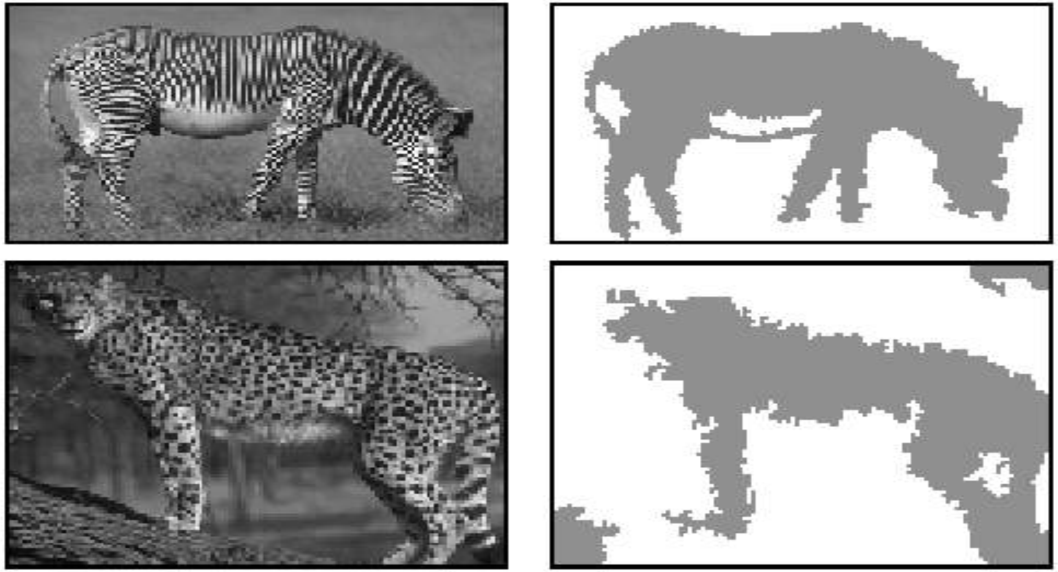


Results on Berkeley Dataset

Dominant sets | Ncut

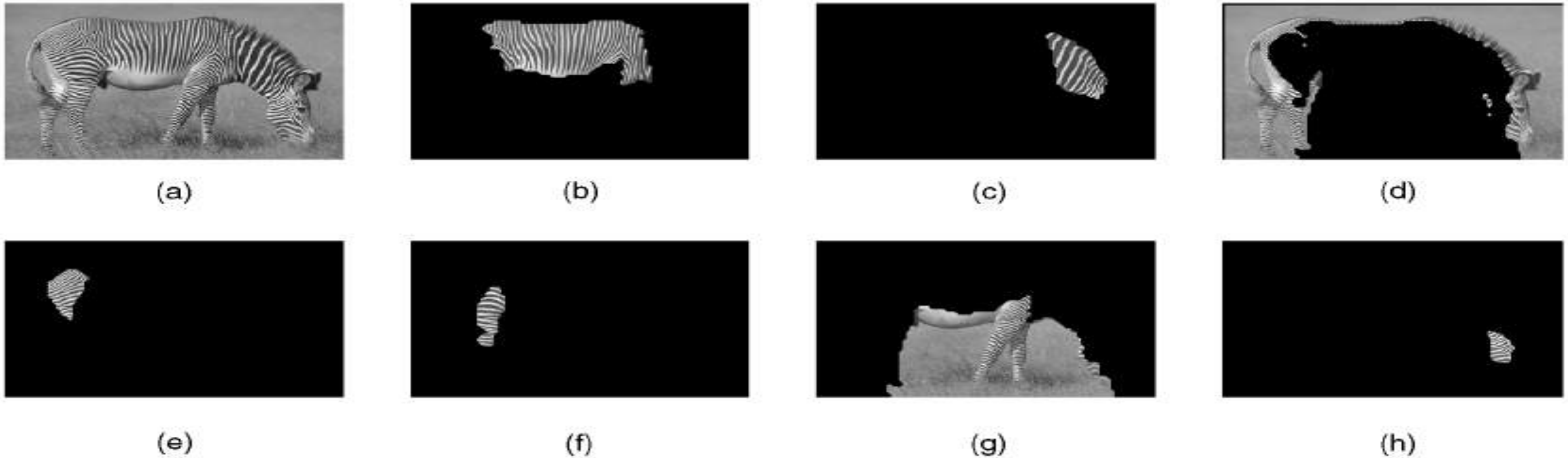


Results on Berkeley Dataset



Dominant sets

Results on Berkeley Dataset



NCut

More Computer Vision Applications

- Video segmentation
- Anomaly detection
- Video summarization
- Feature selection
- Image matching and registration
- 3D reconstruction
- Human action recognition
- Content-based image retrieval
- …

But also in neuroscience, bioinformatics, medical image analysis, etc.

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Detecting conversational groups in images and sequences: A robust game-theoretic approach☆

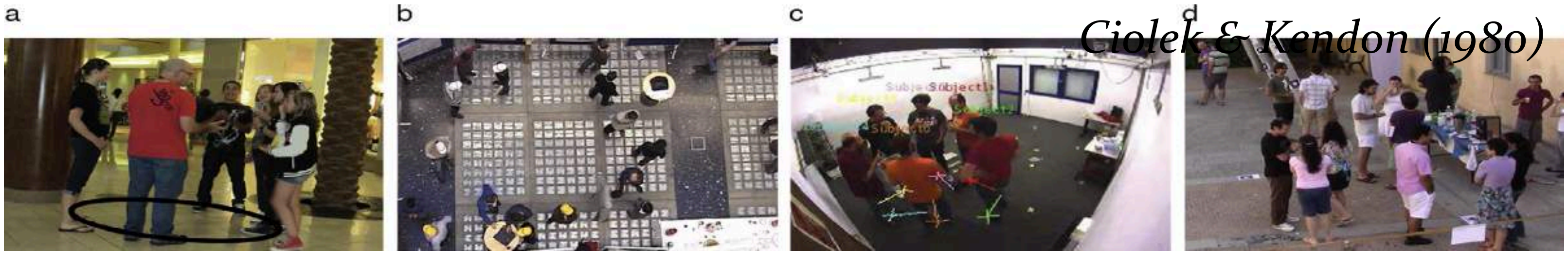
Sebastiano Vascon[a,*], Eyasu Z. Mequanint[b], Marco Cristani[a,c], Hayley Hung[d], Marcello Pelillo[b], Vittorio Murino[a,c]

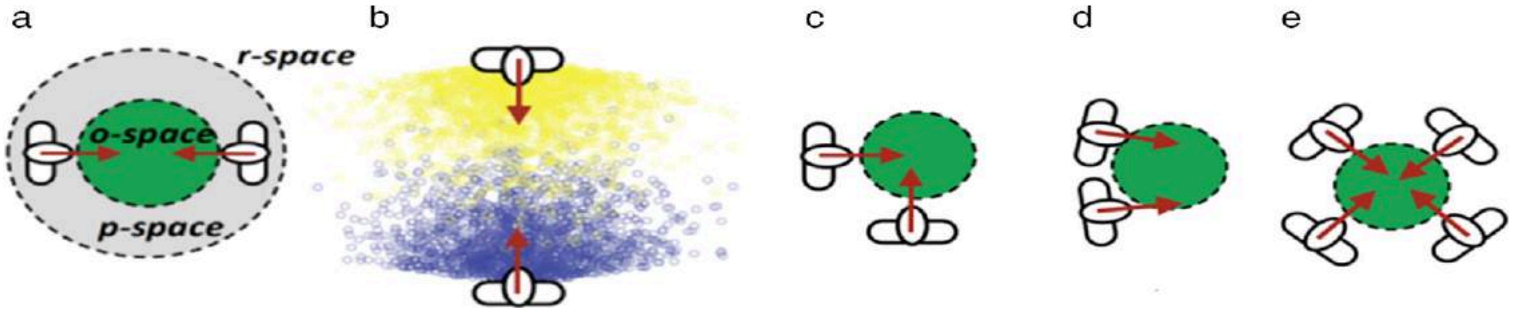
[a] Pattern Analysis and Computer Vision (PAVIS), Istituto Italiano di Tecnologia, Genova, Italy
[b] Deptartment of Environmental Sciences, Informatics and Statistics, University Ca' Foscari of Venice, Italy
[c] Deptartment of Computer Science, University of Verona, Italy
[d] Faculty of Electrical Engineering, Mathematics and Computer Science, Technical University of Delft, Netherlands

F-formations

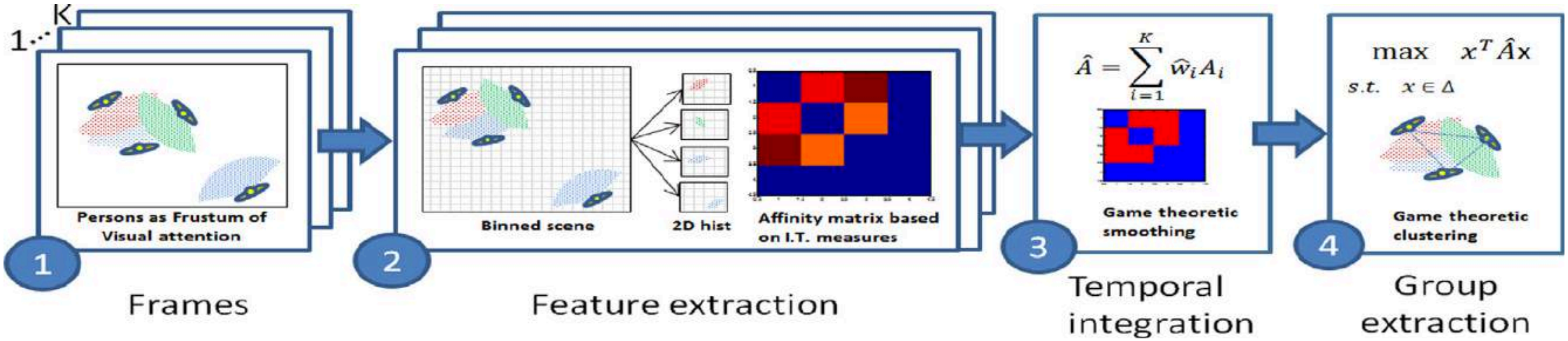
"Whenever two or more individuals in close proximity orient their bodies in such a way that each of them has an easy, direct and equal access to every other participant's transactional segment"

Ciolek & Kendon (1980)





System Architecture



Frustrum of visual attention

- A person in a scene is described by his/her position *(x,y)* and the head orientation θ
- The frustum represents the area in which a person can sustain a conversation and is defined by an aperture and by a length

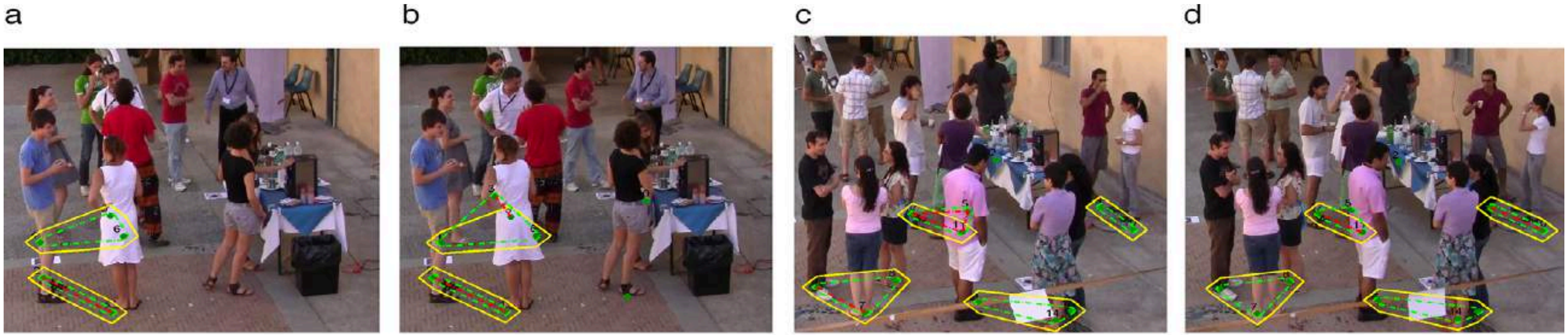
Results

Method ⋆	CoffeeBreak (S1+S2) Prec	Rec	F1	PosterData Prec	Rec	F1	Gdet Prec	Rec	F1
IRPM [61],[22]	0.60	0.41	0.49	–	–	–	–	–	–
HFF [22]	0.82	0.83	0.82	**0.93**	**0.96**	**0.94**	0.67	0.57	0.62
DS ([6], [22])∗	0.68	0.65	0.66	**0.93**	0.92	0.92	–	–	–
MULTISCALE [46]	0.82	0.77	0.80	–	–	–	–	–	–
GTCG [47] KL	0.80	0.84	0.82	0.90	0.94	0.92	**0.76**	0.75	0.75
GTCG [47] JS	0.83	**0.89**	0.86	0.92	**0.96**	**0.94**	**0.76**	**0.76**	**0.76**
R-GTCG SC	0.52	0.59	0.55	0.26	0.27	0.26	0.75	0.75	0.75
R-GTCG	**0.86**	0.88	**0.87**	0.92	**0.96**	**0.94**	**0.76**	**0.76**	**0.76**
	$\sigma = 0.2$, $l = 145$			$\sigma = 0.25$, $l = 115$			$\sigma = 0.7$ $l = 180$		

Method ⋆	**Cocktail Party** Prec	Rec	F1	**Synth** Prec	Rec	F1
IRPM [22,61]	–	–	–	0.71	0.54	0.61
HFF ([7], [46])	0.59	0.74	0.66	0.73	0.83	0.78
MULTISCALE [46]	0.69	0.74	0.71	0.86	0.94	0.90
GTCG [47] KL	0.85	0.81	0.83	**1.00**	**1.00**	**1.00**
GTCG [47] JS	0.86	**0.82**	**0.84**	**1.00**	**1.00**	**1.00**
R-GTCG SC	0.77	0.72	0.74	0.40	0.90	0.56
	$\sigma=0.6$, $l=170$			$\sigma=0.1$, $l=75$		
R-GTCG	**0.87**	**0.82**	**0.84**	**1.00**	**1.00**	**1.00**

Spectral Clustering

Results



Seq 1 Seq 1 Seq 2 Seq 2

Qualitative results on the CoffeeBreak dataset compared with the state of the art HFF.

Yellow = ground truth
Green = our method
Red = HFF.

A Game-Theoretic Approach to Hypergraph Clustering

Samuel Rota Bulò and Marcello Pelillo, *Fellow, IEEE*

Abstract—Hypergraph clustering refers to the process of extracting maximally coherent groups from a set of objects using high-order (rather than pairwise) similarities. Traditional approaches to this problem are based on the idea of partitioning the input data into a predetermined number of classes, thereby obtaining the clusters as a by-product of the partitioning process. In this paper, we offer a radically different view of the problem. In contrast to the classical approach, we attempt to provide a meaningful formalization of the very notion of a cluster and we show that game theory offers an attractive and unexplored perspective that serves our purpose well. To this end, we formulate the hypergraph clustering problem in terms of a noncooperative multiplayer "clustering game," and show that a natural notion of a cluster turns out to be equivalent to a classical (evolutionary) game-theoretic equilibrium concept. We prove that the problem of finding the equilibria of our clustering game is equivalent to locally optimizing a polynomial function over the standard simplex, and we provide a discrete-time high-order replicator dynamics to perform this optimization, based on the Baum-Eagon inequality. Experiments over synthetic as well as real-world data are presented which show the superiority of our approach over the state of the art.

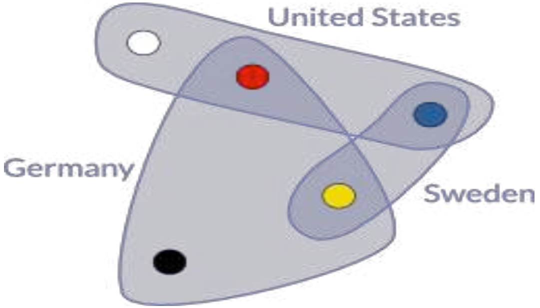
Index Terms—Hypergraph clustering, evolutionary game theory, polynomial optimization, Baum-Eagon inequality, high-order replicator dynamics

Dealing with High-Order Similarities

A (weighted) hypergraph is a triplet $H = (V, E, w)$, where

- V is a finite set of vertices
- $E \subseteq 2^V$ is the set of (hyper-)edges (where 2^V is the power set of V)
- $w : E \rightarrow R$ is a real-valued function assigning a weight to each edge

We focus on a particular class of hypergraphs, called **k-graphs**, whose edges have fixed cardinality $k \geq 2$.



A hypergraph where the vertices are flag colors and the hyperedges are flags.

Hypergraph Clustering Games

Given a weighted k-graph representing an instance of a hypergraph clustering problem, we cast it into a k-player (hypergraph) clustering game where:

- ✓ There are <u>k players</u>
- ✓ The objects to be clustered are the pure strategies
- ✓ The payoff function is proportional to the similarity of the objects/strategies selected by the players

Definition (ESS-cluster). Given an instance of a hypergraph clustering problem H = (V,E,w), an ESS-cluster of H is an ESS of the corresponding hypergraph clustering game.

Like the k=2 case, ESS-clusters do incorporate both internal and external cluster criteria (see PAMI 2013)

ESS's and Polynomial Optimization

$$f(x) = \sum_{e \in E} w(e) \prod_{j \in e} x_j$$

Nash equilibrium $\Leftrightarrow$ critical point of f(x) on Δ

ESS $\Leftrightarrow$ strict local maximizer of f(x) on Δ

Use high-order replicator equation (Baum-Eagon) to maximize f on Δ.

$$x_i \leftarrow x_i \frac{\partial f(x)}{\partial x_i} \Big/ \sum_k x_k \frac{\partial f(x)}{\partial x_k}$$

Theorem. A point x is an ESS-cluster if and only if it is an asymptotically stable equilibrium point for replicator dynamics above.

Model Fitting: Line Clustering

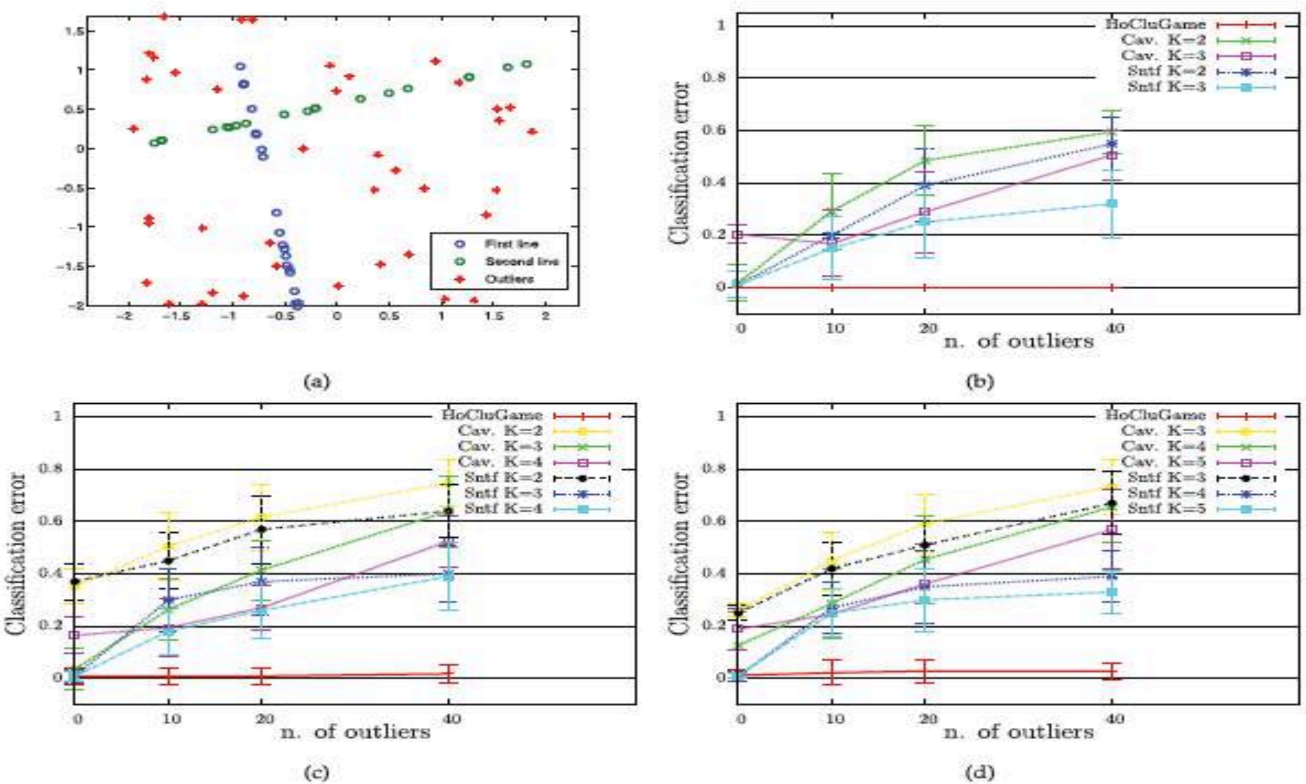


Fig. 3. Results of clustering two, three, and four lines with an increasing number of clutter points (0, 10, 20, 40). (a) Example of two 5D lines (projected in 2D) with 40 clutter points. (b) Two lines. (c) Three lines. (d) Four lines.

Illuminant-invariant Face Clustering

Images of a Lambertian object illuminated by a point light source lie in a 3D subspace (Belhumeur and Kriegman, 1998).

If we assume that four images of a face form the columns of a matrix, then $s^2_4 / (s^2_1 + \ldots + s^2_4)$ provides us with a measure of dissimilarity, s_i being the *i*th singular value of this matrix (Agarwal et al, 2005).

Average classification error and corresponding standard deviation

n. of classes:	4		5	
n. of outliers:	0	10	0	10
CAVERAGE K=3	0.26±0.09	0.40±0.10	-	-
CAVERAGE K=4	**0.03±0.04**	0.24±0.07	0.21±0.11	0.65±0.12
CAVERAGE K=5	0.13±0.05	0.12±0.05	0.07±0.07	0.41±0.09
CAVERAGE K=6	-	-	0.13±0.08	0.37±0.11
SNTF K=3	0.29±0.10	0.39±0.09	-	-
SNTF K=4	0.14±0.06	0.26±0.09	0.28±0.11	0.51±0.12
SNTF K=5	0.19±0.09	0.25±0.13	0.11±0.09	0.43±0.11
SNTF K=6	-	-	0.14±0.09	0.39±0.13
HoCluGame	0.06±0.03	**0.07±0.02**	**0.06±0.02**	**0.07±0.03**

Dominant Sets for “Constrained” Image Segmentation

Eyasu Zemene *, *Member, IEEE,* Leulseged Tesfaye Alemu *, *Member, IEEE*
and Marcello Pelillo, *Fellow, IEEE*

Abstract—Image segmentation has come a long way since the early days of computer vision, and still remains a challenging task. Modern variations of the classical (purely bottom-up) approach, involve, e.g., some form of user assistance (interactive segmentation) or ask for the simultaneous segmentation of two or more images (co-segmentation). At an abstract level, all these variants can be thought of as "constrained" versions of the original formulation, whereby the segmentation process is guided by some external source of information. In this paper, we propose a new approach to tackle this kind of problems in a unified way. Our work is based on some properties of a family of quadratic optimization problems related to *dominant sets*, a graph-theoretic notion of a cluster which generalizes the concept of a maximal clique to edge-weighted graphs. In particular, we show that by properly controlling a regularization parameter which determines the structure and the scale of the underlying problem, we are in a position to extract groups of dominant-set clusters that are constrained to contain predefined elements. In particular, we shall focus on interactive segmentation and co-segmentation (in both the unsupervised and the interactive versions). The proposed algorithm can deal naturally with several types of constraints and input modalities, including scribbles, sloppy contours and bounding boxes, and is able to robustly handle noisy annotations on the part of the user. Experiments on standard benchmark datasets show the effectiveness of our approach as compared to state-of-the-art algorithms on a variety of natural images under several input conditions and constraints.

Index Terms—Interactive segmentation, co-segmentation, dominant sets, quadratic optimization, game dynamics.

Constrained Dominant Sets

Given $S \subseteq V$ and a parameter $\alpha > 0$, define the following parameterized family of quadratic programs:

$$\begin{array}{ll} \text{maximize} & f_S^\alpha(\mathbf{x}) = \mathbf{x}'(A - \alpha \hat{I}_S)\mathbf{x} \\ \text{subject to} & \mathbf{x} \in \Delta \end{array}$$

where I_S is the diagonal matrix whose elements are set to 1 in correspondence to the vertices outside S, and to zero otherwise:

$$\hat{I}_S = \begin{pmatrix} 0 & 0 \\ 0 & I_{n-k} \end{pmatrix}$$

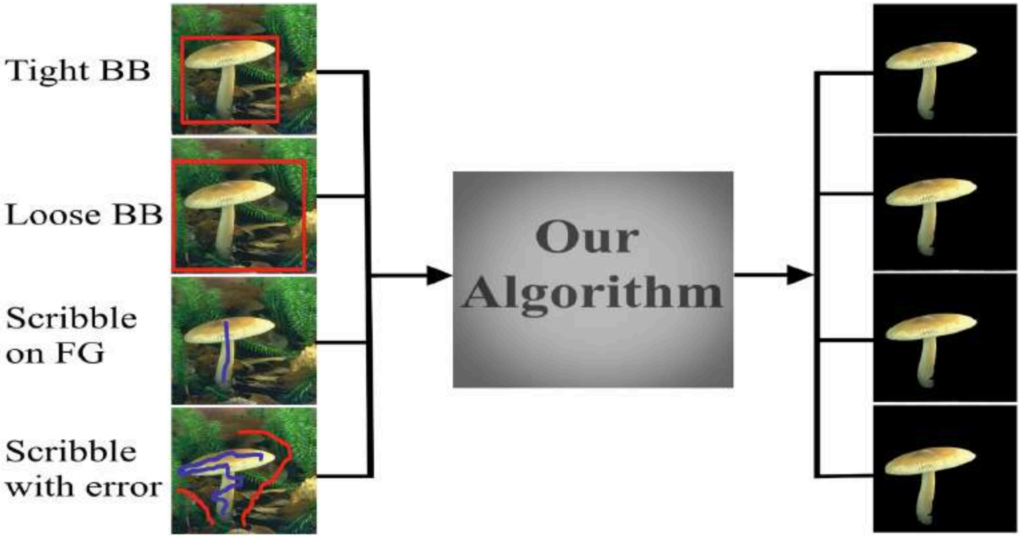
Property. By setting:

$$\alpha > \lambda_{\max}(A_{V \setminus S})$$

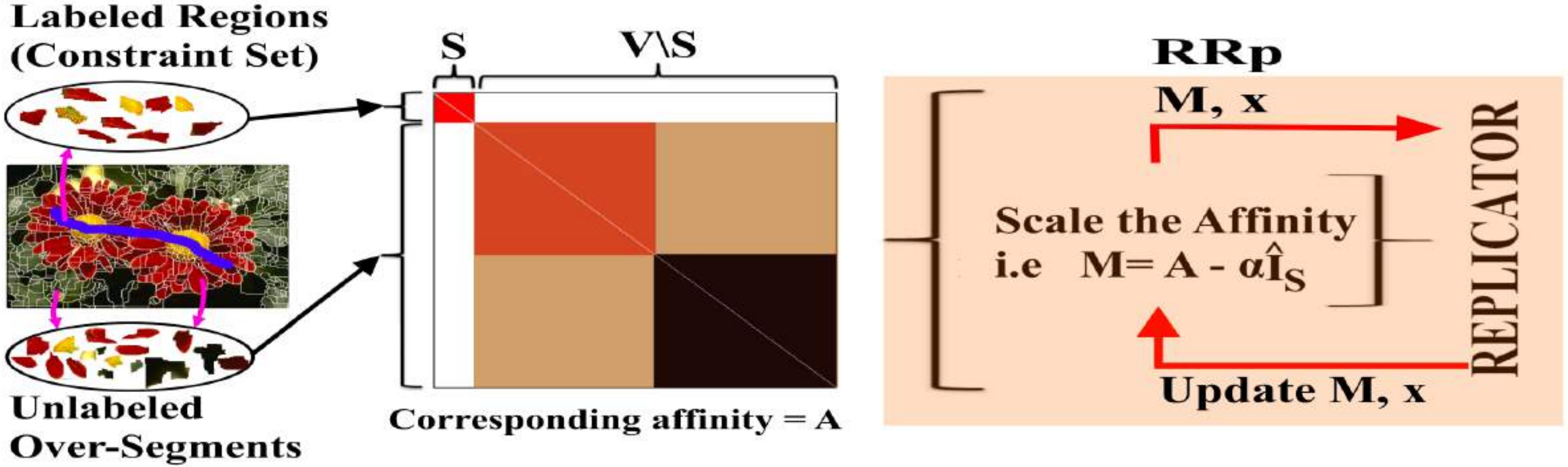
all local solutions will have a support containing elements of *S*.

Interactive Image Segmentation

Given an image and some information provided by a user, in the form of a scribble or of a bounding box, to provide as output a foreground object that best reflects the user's intent.



System Overview



Left: Over-segmented image with a user scribble (blue label).

Middle: The corresponding affinity matrix, using each over-segments as a node, showing its two parts: S, the constraint set which contains the user labels, and V n S, the part of the graph which takes the regularization parameter .

Right: RRp, starts from the barycenter and extracts the first dominant set and update x and M, for the next extraction till all the dominant sets which contain the user labeled regions are extracted.

Results

Table 1. Error rates of different scribble-based approaches on the Grab-Cut dataset

Methods	Error rate
Graph Cut [7]	6.7
Lazy Snapping [5]	6.7
Geodesic Segmentation [4]	6.8
Random Walker [33]	5.4
Transduction [34]	5.4
Geodesic Graph Cut [30]	4.8
Constrained Random Walker [31]	4.1
CDS_Self Tuning (Ours)	**3.57**
CDS_Single Sigma (Ours)	**3.80**
CDS_Best Sigma (Ours)	2.72

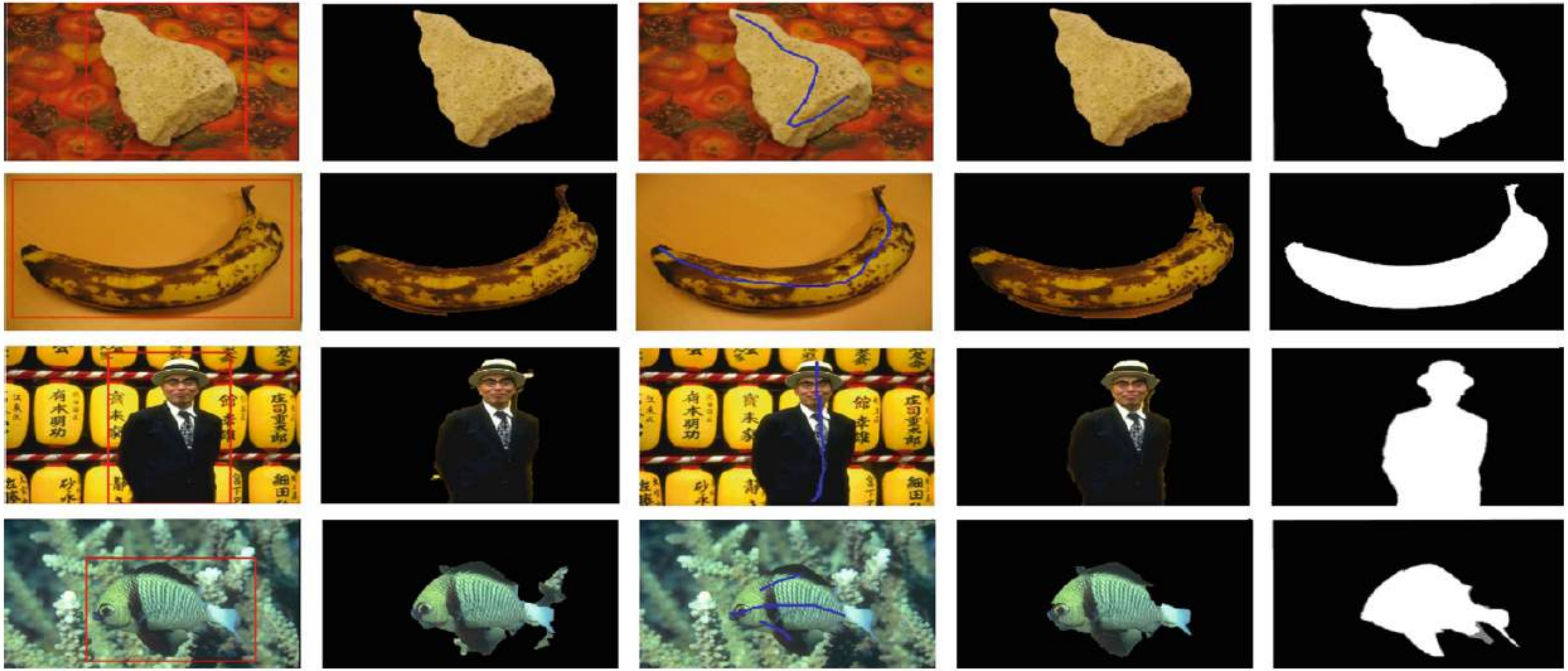
Table 2. Jaccard index of different approaches – first 5 bounding-box-based – on Berkeley dataset

Methods	Jaccard index
MILCut-Struct [3]	84
MILCut-Graph [3]	83
MILCut [3]	78
Graph Cut [1]	77
Binary Partition Trees [35]	71
Interactive Graph Cut [7]	64
Seeded Region Growing [36]	59
Simple Interactive O.E [37]	63
CDS_Self Tuning (Ours)	**93**
CDS_Single Sigma (Ours)	**93**
CDS_Best Sigma (Ours)	95

Results

Bounding box | Result | Scribble | Result | Ground truth

Results



Bounding box | Result | Scribble | Result | Ground truth

Large-scale Image Geo-Localization Using Dominant Sets

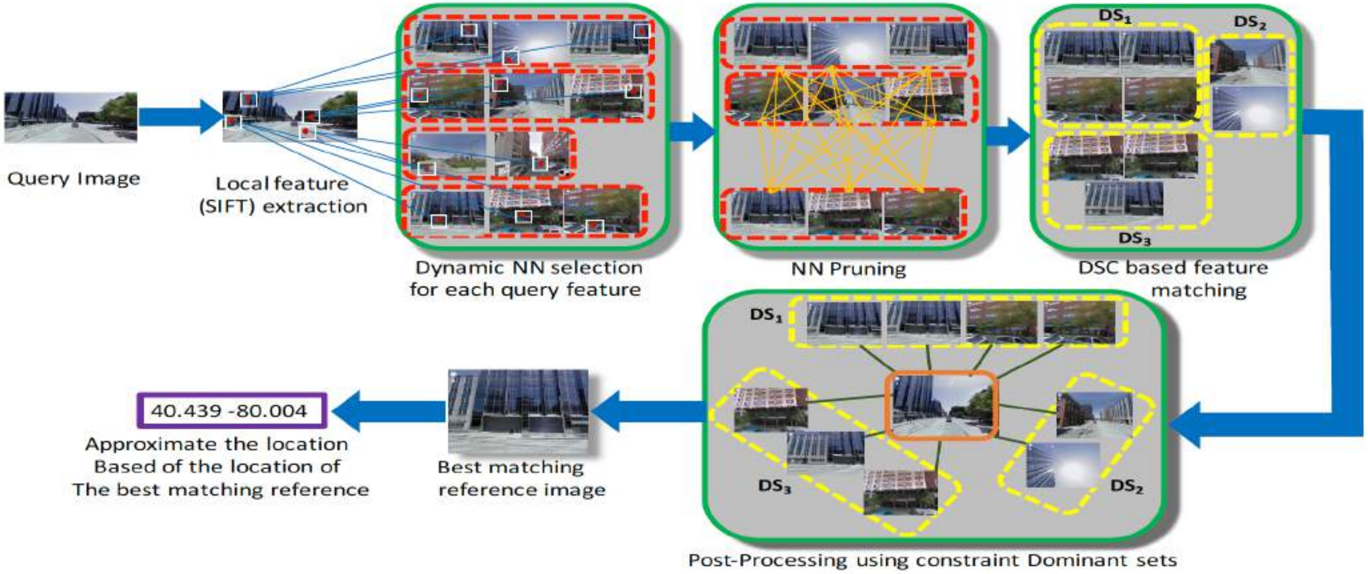
Eyasu Zemene*, *Student Member, IEEE,* Yonatan Tariku Tesfaye*, *Student Member, IEEE,* Haroon Idrees, *Member, IEEE,* Andrea Prati, *Senior member, IEEE,* Marcello Pelillo, *Fellow, IEEE,* and Mubarak Shah, *Fellow, IEEE*

Abstract—This paper presents a new approach for the challenging problem of geo-localization using image matching in a structured database of city-wide reference images with known GPS coordinates. We cast the geo-localization as a clustering problem of local image features. Akin to existing approaches to the problem, our framework builds on low-level features which allow local matching between images. For each local feature in the query image, we find its approximate nearest neighbors in the reference set. Next, we cluster the features from reference images using Dominant Set clustering, which affords several advantages over existing approaches. First, it permits variable number of nodes in the cluster, which we use to dynamically select the number of nearest neighbors for each query feature based on its discrimination value. Second, this approach is several orders of magnitude faster than existing approaches. Thus, we obtain multiple clusters (different local maximizers) and obtain a robust final solution to the problem using multiple weak solutions through constrained Dominant Set clustering on global image features, where we enforce the constraint that the query image must be included in the cluster. This second level of clustering also bypasses heuristic approaches to voting and selecting the reference image that matches to the query. We evaluate the proposed framework on an existing dataset of 102k street view images as well as a new larger dataset of 300k images, and show that it outperforms the state-of-the-art by 20% and 7%, respectively, on the two datasets.

Index Terms—Geo-localization, Dominant Set Clustering, Multiple Nearest Neighbor Feature Matching, Constrained Dominant Set

Image Geo-localization

A new approach for the problem of geo-localization using image matching in a structured database of city-wide reference images with known GPS coordinates.



200x time faster + 20% accuracy improvement w.r.t previous approach

Datasets:

- Datasets one:
  - Reference images:
    - 102K Google street view images from **Pittsburgh, PA** and **Orlando, FL**
  - Test Set:
    - 521 GPS-Tagged unconstrained images
    - Downloaded From Flickr, Panoramio, Picasa, …
- ***WorldCities*** Datasets (NEW)*:
  - Reference images:
    - 300K Google street view images
      - 14 different cities from **Europe, N. America and Australia**
  - Test Set:
    - 500 GPS-Tagged unconstrained images
    - Downloaded From Flickr, Panoramio, Picasa, …

Google Maps Street View Datasets:

For each location: 4 side views and 1 top view is collected

Side Views

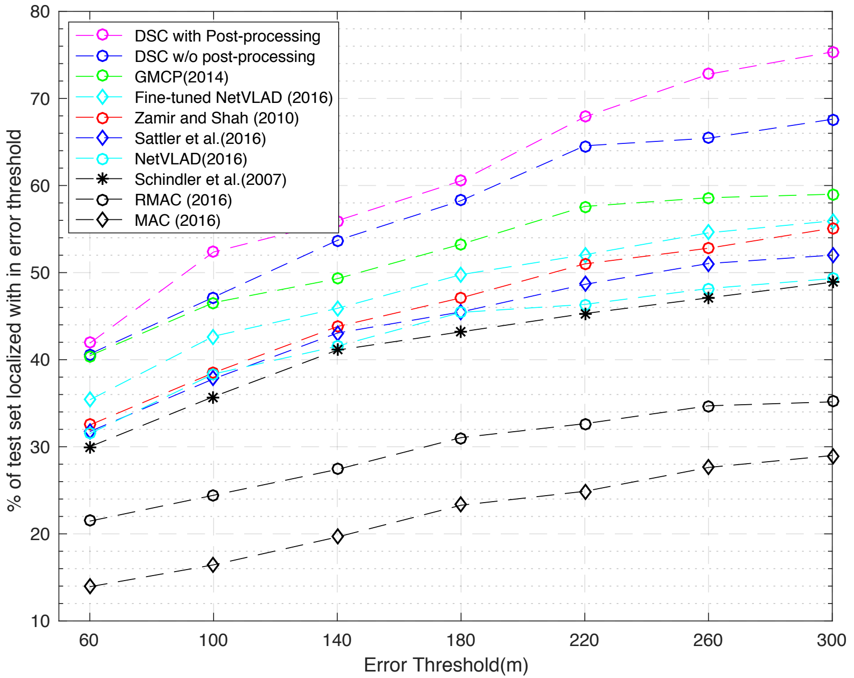
top View

Pittsburgh, PA   Longitude = 40.44146°,

Latitude=-80.0037°

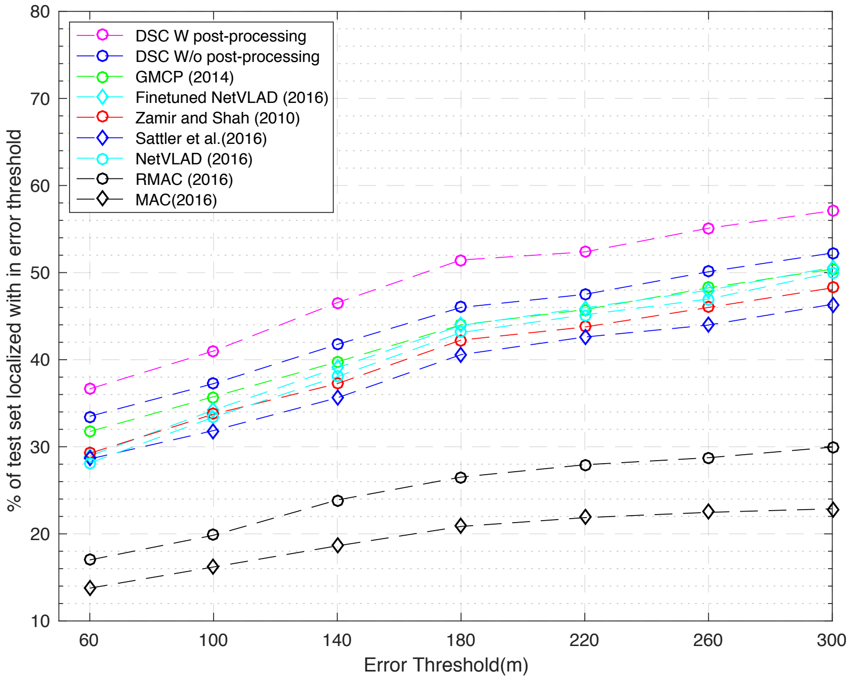
Overall Result

- Dataset 1: 102K Google street view images (Orlando and Pittsburg area)



Overall Result

- Dataset 2: WorldCities (14 different cities from Europa, North America, Australia)



Qualitative Results


Query


Match – Error: 70.01 m


Query


Match – Error: 5.4 m


Query


Match – Error: 10.4 m


Query


Match – Error: 7.5 m


Query


Match – Error: 62.7 m

International Journal of Computer Vision
https://doi.org/10.1007/s11263-019-01180-6

Multi-target Tracking in Multiple Non-overlapping Cameras Using Fast-Constrained Dominant Sets

Yonatan Tariku Tesfaye[1,2] · Eyasu Zemene[1,3] · Andrea Prati[5] · Marcello Pelillo[4] · Mubarak Shah[1]

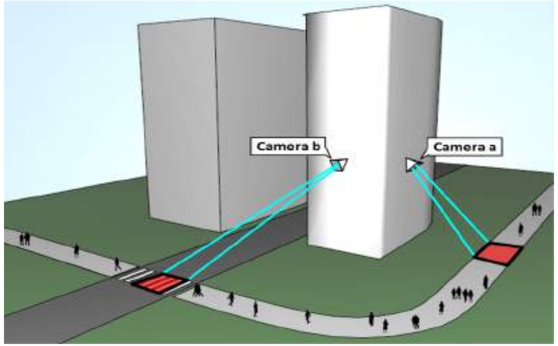
Received: 22 February 2018 / Accepted: 25 April 2019

Abstract

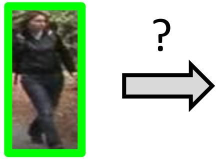
In this paper, a unified three-layer hierarchical approach for solving tracking problem in a multiple non-overlapping cameras setting is proposed. Given a video and a set of detections (obtained by any person detector), we first solve *within-camera tracking* employing the first two layers of our framework and then, in the third layer, we solve *across-camera tracking* by associating tracks of the same person in all cameras simultaneously. To best serve our purpose, we propose fast-constrained dominant set clustering (FCDSC), a novel method which is several orders of magnitude faster (close to real time) than existing methods. FCDSC is a parameterized family of quadratic programs that generalizes the standard quadratic optimization problem. In our method, we first build a graph where nodes of the graph represent short-tracklets, tracklets and tracks in the first, second and third layer of the framework, respectively. The edge weights reflect the similarity between nodes. FCDSC takes as input a constrained set, a subset of nodes from the graph which need to be included in the extracted cluster. Given a constrained set, FCDSC generates compact clusters by selecting nodes from the graph which are highly similar to each other and with elements in the constrained set. We have tested this approach on a very large and challenging dataset (namely, MOTchallenge DukeMTMC) and show that the proposed framework outperforms the state-of-the-art approaches. Even though the main focus of this paper is on multi-target tracking in non-overlapping cameras, the proposed approach can also be applied to solve *video-based person re-identification* problem. We show that when the re-identification problem is formulated as a clustering problem, FCDSC can be used in conjunction with state-of-the-art video-based re-identification algorithms, to increase their already good performances. Our experiments demonstrate the general applicability of the proposed framework for multi-target multi-camera tracking and person re-identification tasks.

Keywords Multi-target multi-camera tracking · Constrained dominant sets · Standard quadratic optimization

Person Re-identification



- Recognize an individual over different non-overlapping cameras.
- Given a gallery of person images we want to recognize (between all of them) a new observed image, called probe.



Video-based Person Re-ID

Traditional methods focus on:

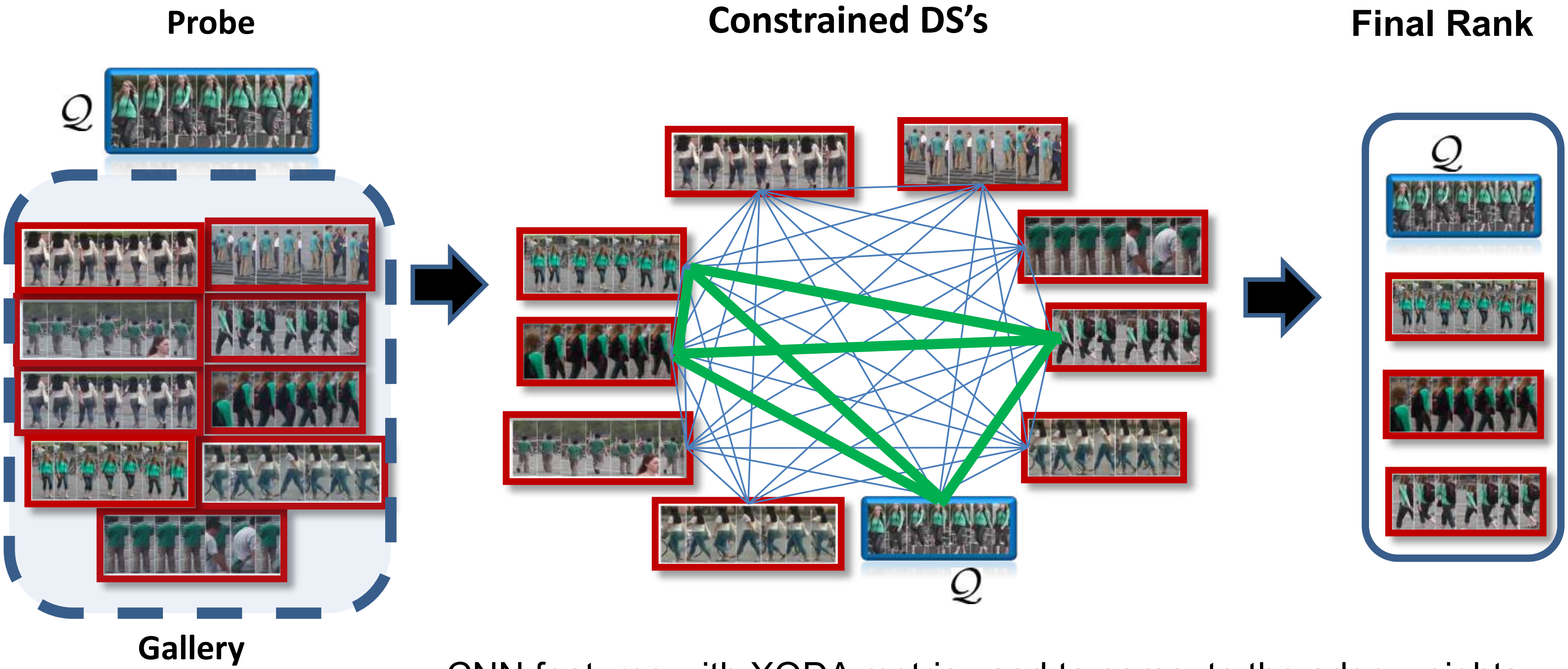
- Building better feature representation of objects
- Building a better distance metric
- Finally rank images from gallery based on the pairwise distances from the query

In our approach

- We use standard features and distance metric
- Extract constrained dominant sets for each query
- Perform ranking over shortlisted clips NOT over the whole set

We take into account **both** the relationship between query and elements in the gallery and elements in the gallery.

Re-ID with Constrained DS's



CNN features with XQDA metric used to compute the edge weights

Results on MARS Dataset

- Largest video Re-ID dataset (2016)
- 6 near-synchronized cameras
- 1,261 identities
- 3,248 distractors
- tracklets are of 25-30 frames long

Methods	rank 1
HLBP [40] + XQDA	18.60
BCov [24] + XQDA	9.20
LOMO [20] + XQDA	30.70
BoW [49] + KISSME	30.60
SDALF [8] + DVR	4.10
HOG3D [16] + KISSME	2.60
CNN + XQDA [48]	65.30
CNN + KISSME [48]	65.00
Ours	**68.22**

[8] M. Farenzena et al. Person re-identification by symmetry-driven accumulation of local features (*CVPR 2010)*

[16] A. Klaser et al. A spatio-temporal descriptor based on 3D-gradients (*BMVC 2008)*

[20] S. Liao et al. Person re-identification by local maximal occurrence representation and metric learning (*CVPR 2015)*

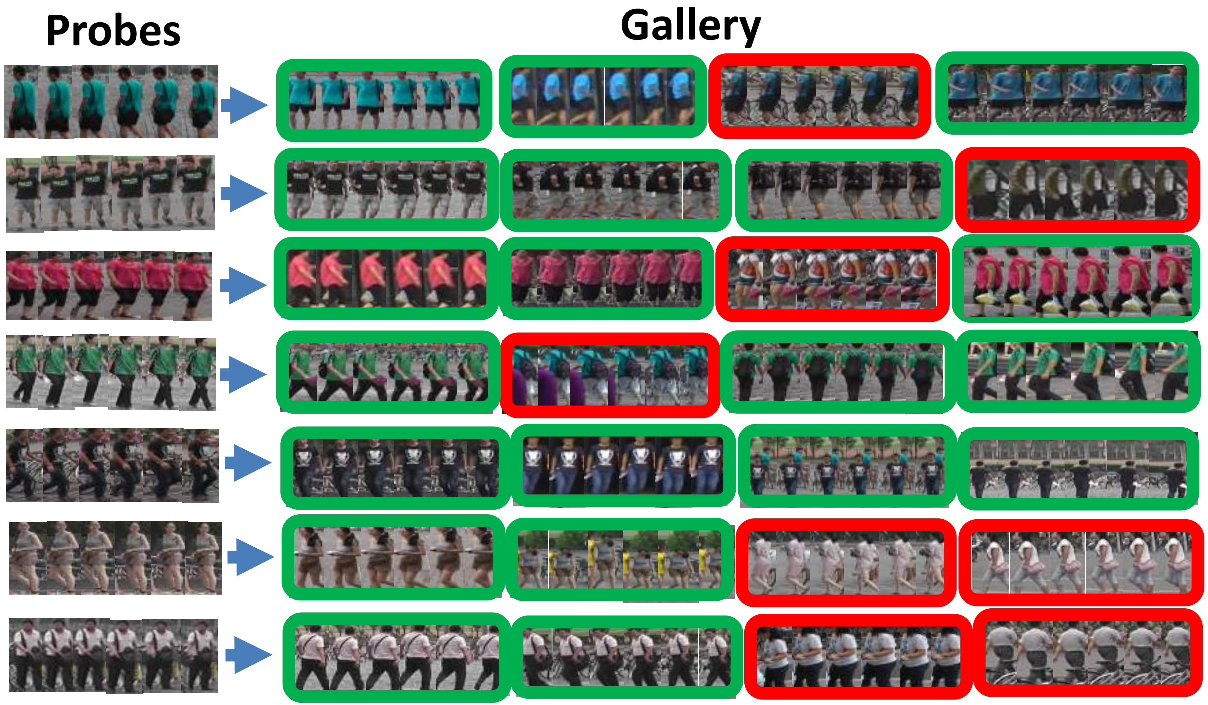
[24] B. Ma et al. Covariance descriptor based on bio-inspired features for person re-identification and face verification *(Image Vision Comput 2014)*

[40] F. Xiong et al. Person re-identification using kernel-based metric learning methods (*ECCV 2014)*

[48] L. Zheng et al. MARS: A video benchmark for large-scale person re-identification (*ECCV 2016)*

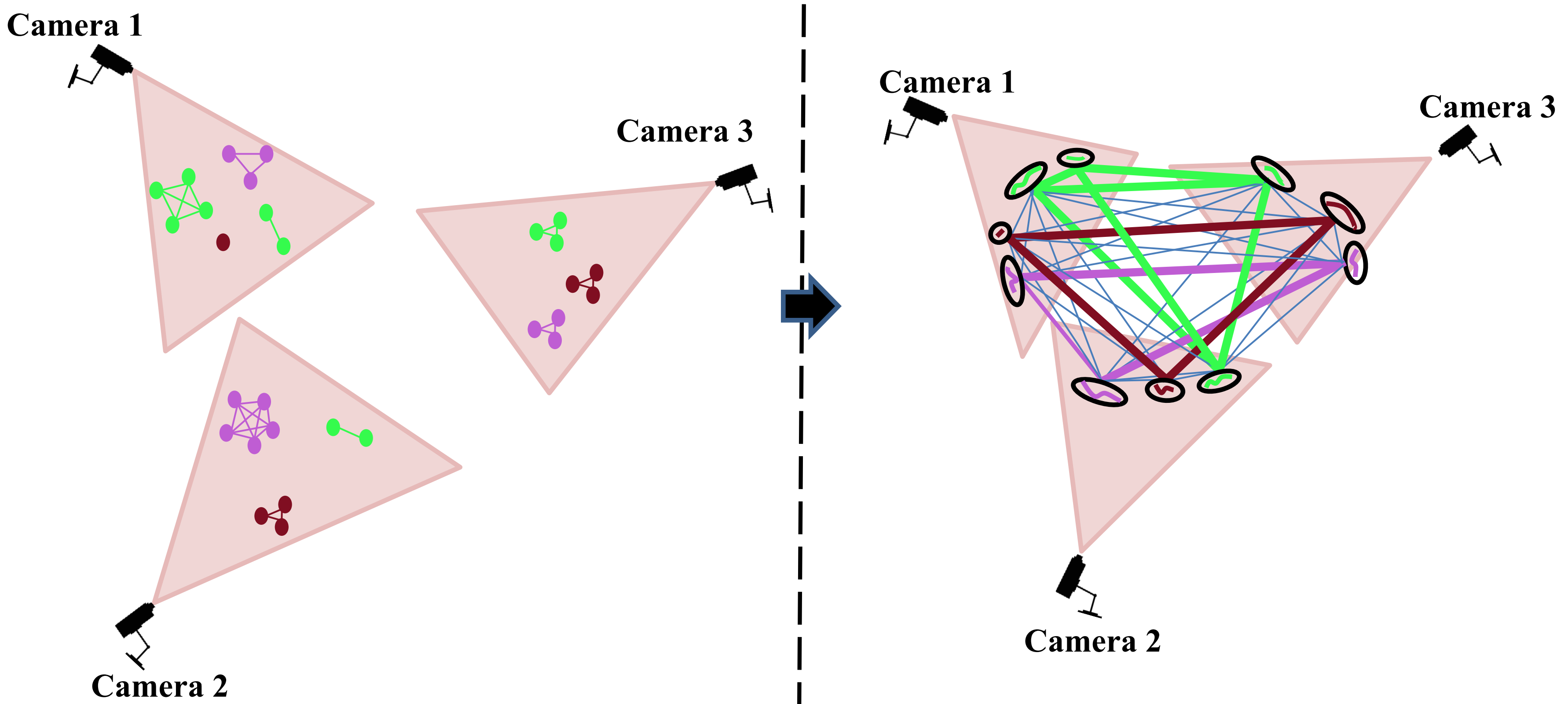
[49] L. Zheng et al. Scalable person re-identification: A benchmark *(ICCV 2015)*

Examples



The green and red boxes denote the same and different persons with the probes, respectively
Gallery images are ordered based on their membership score (highest -> lowest).

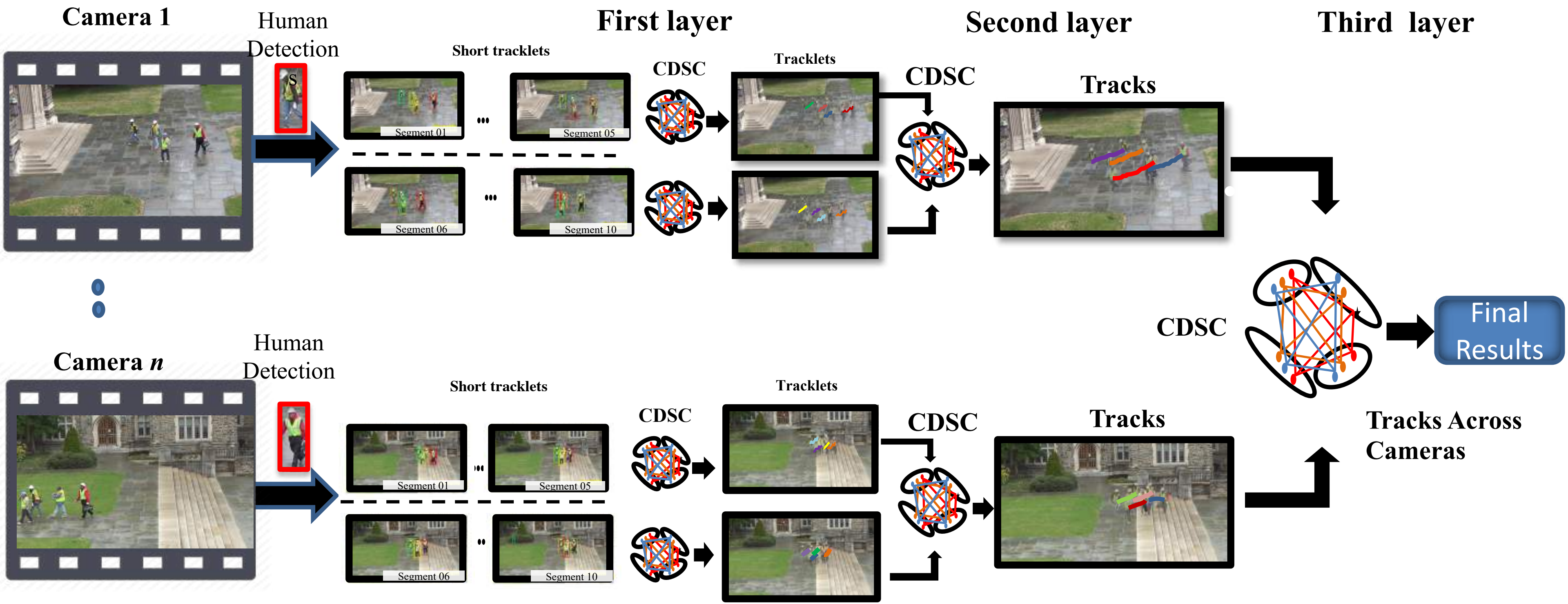
Multi-target Multi-camera Tracking



Within-camera tracking

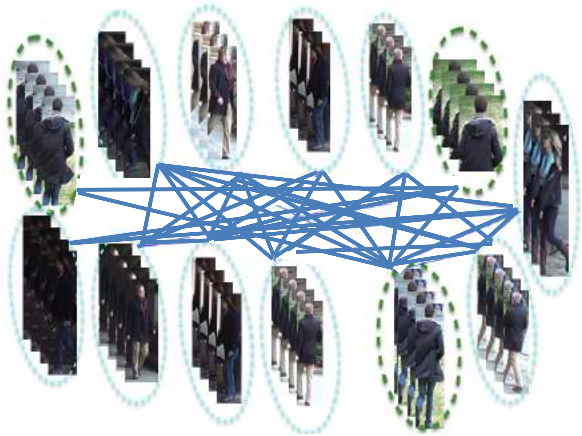
Cross-camera tracking

Pipeline



Camera 1
Human Detection
Short tracklets
Segment 01
Segment 05
Segment 06
Segment 10
First layer
CDSC
Tracklets
Second layer
CDSC
Tracks
Third layer
CDSC
Final Results
Camera n
Human Detection
Short tracklets
Segment 01
Segment 05
Segment 06
Segment 10
CDSC
Tracklets
CDSC
Tracks
Tracks Across Cameras

Layer 1: Tracklet Extraction

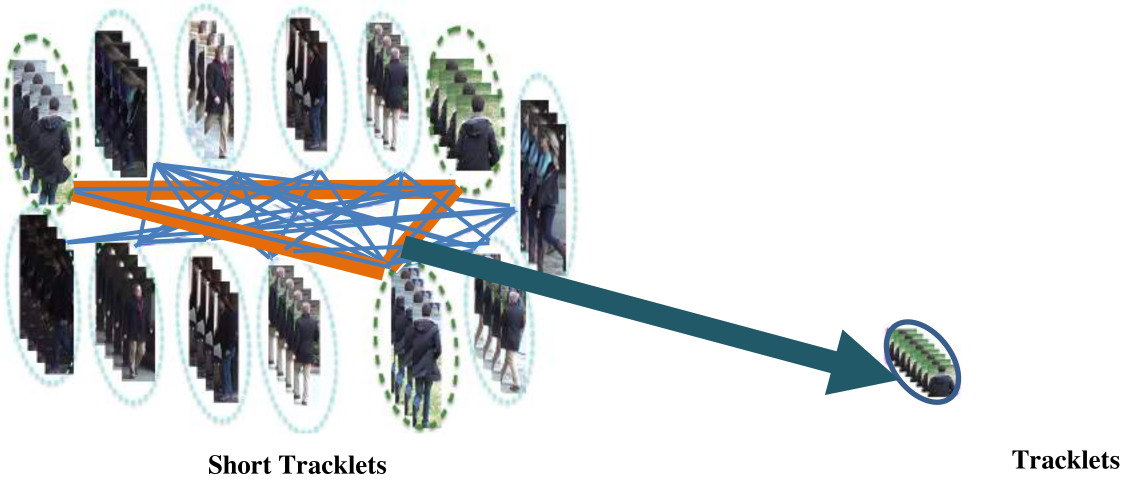


Short Tracklets

Edge weights combine appearance and motion

- Appearance = CNN features
- Motion = Constant velocity

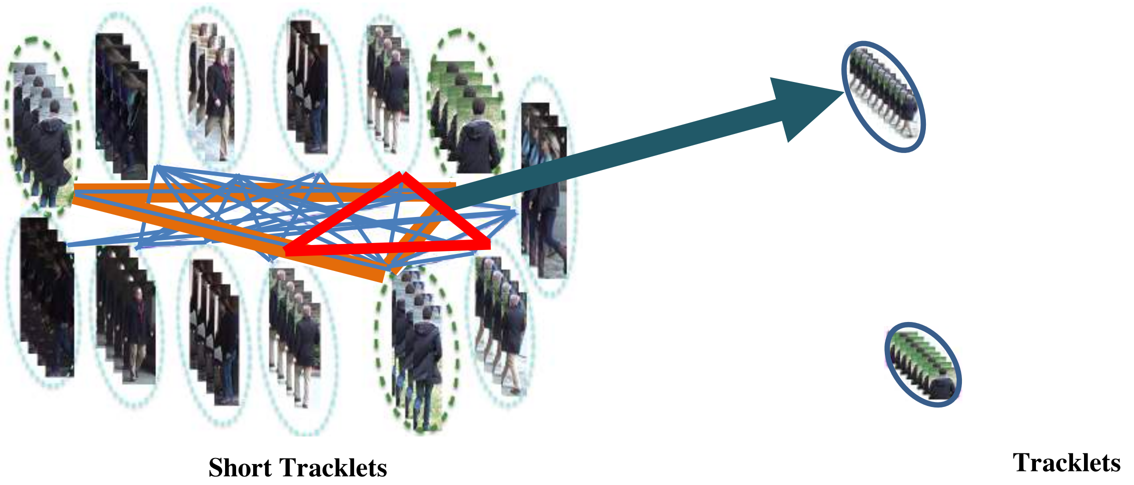
Layer 1: Tracklet Extraction



Edge weights combine appearance and motion

- Appearance = CNN features
- Motion = Constant velocity

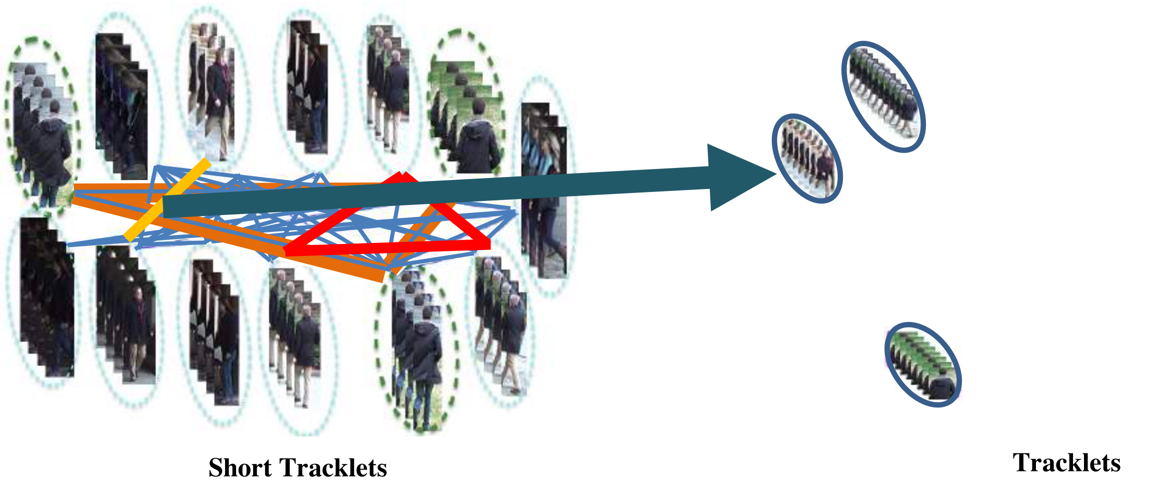
Layer 1: Tracklet Extraction



Edge weights combine appearance and motion

- Appearance = CNN features
- Motion = Constant velocity

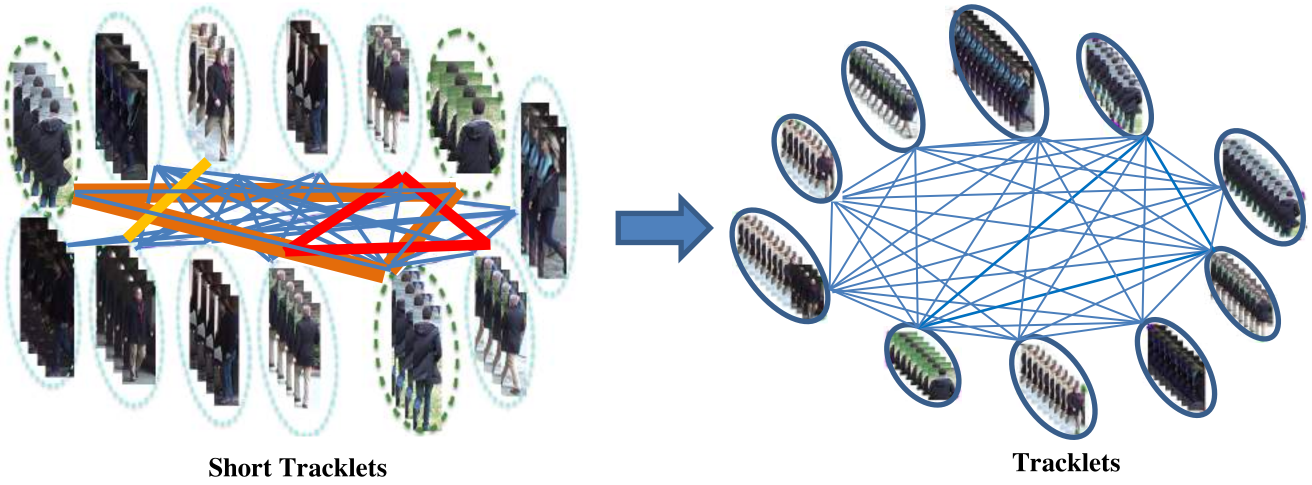
Layer 1: Tracklet Extraction



Edge weights combine appearance and motion

- Appearance = CNN features
- Motion = Constant velocity

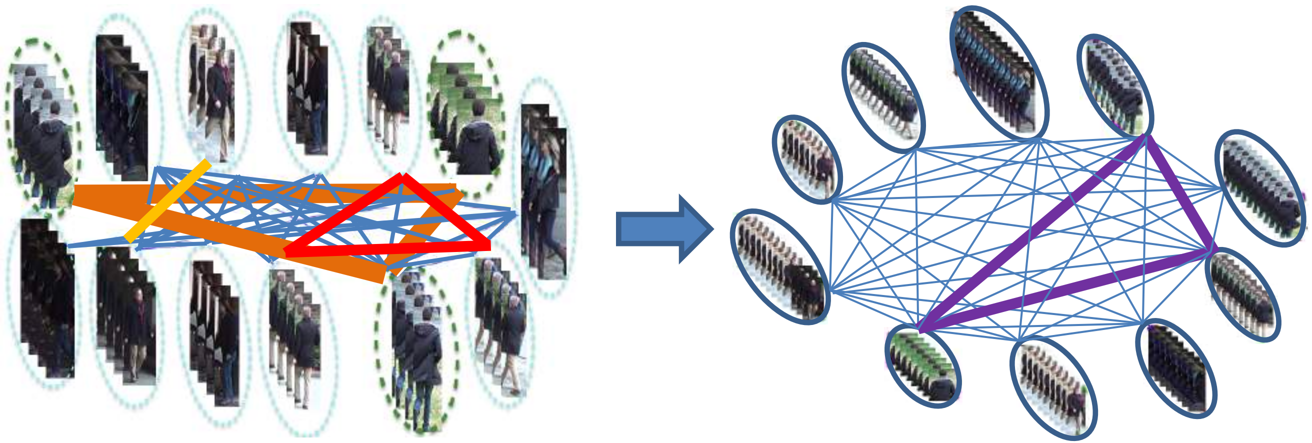
Layer 1: Tracklet Extraction



Edge weights combine appearance and motion

- Appearance = CNN features
- Motion = Constant velocity

Layer 2: Track Extraction



Short Tracklets

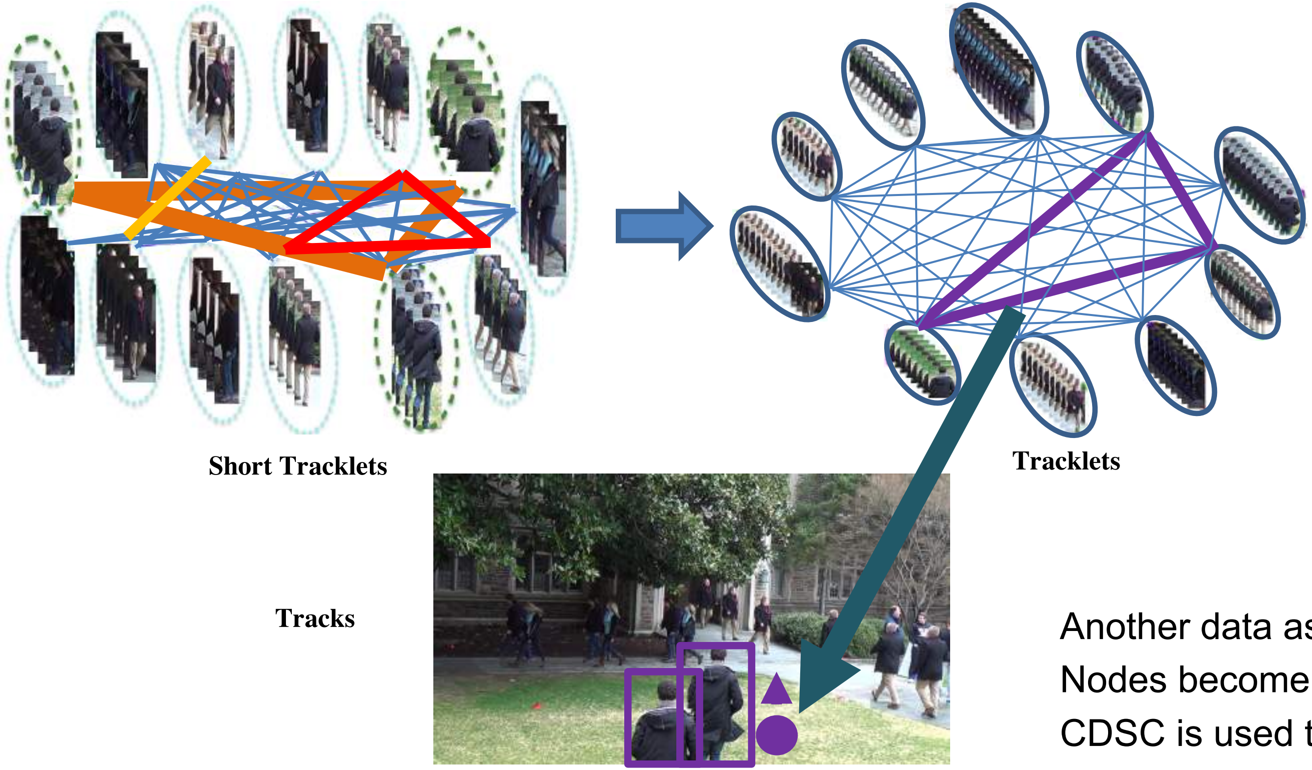
Tracklets

Another data association problem

Nodes become tracklets

CDSC is used to stitch tracklets

Layer 2: Track Extraction

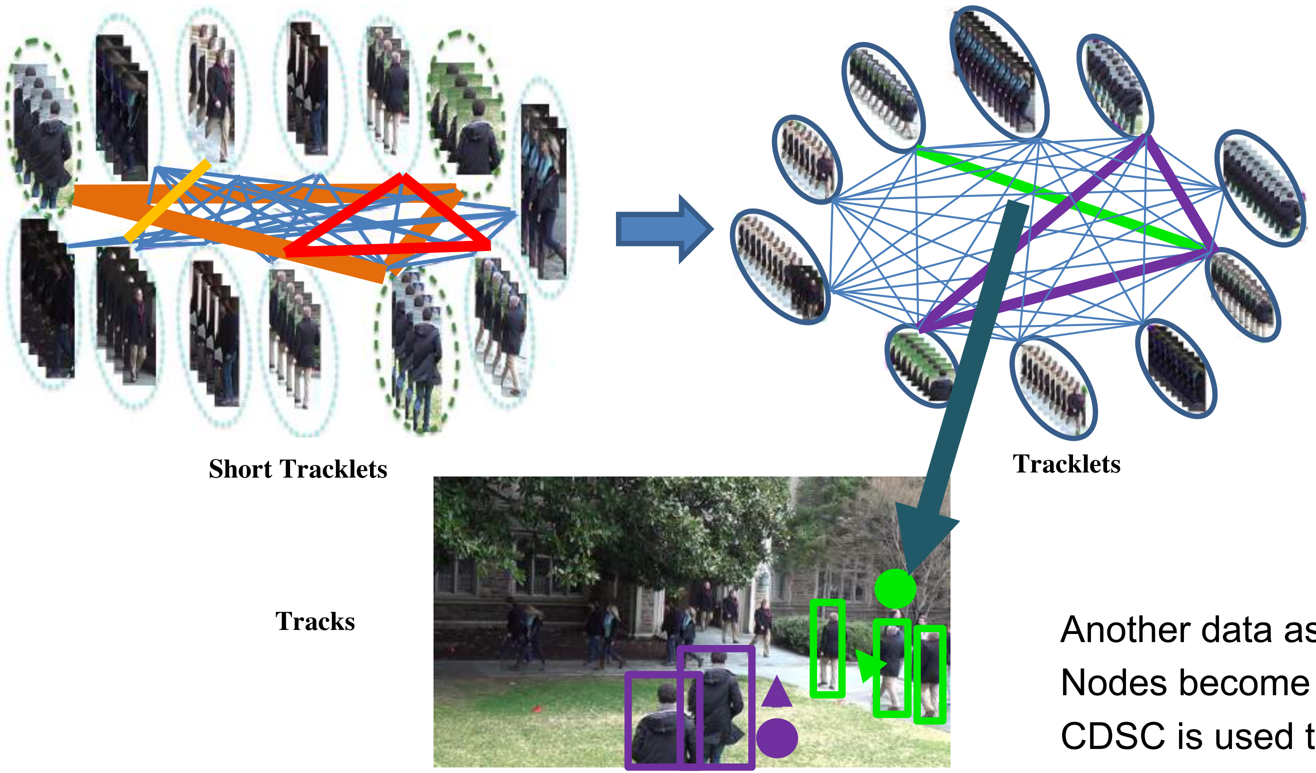


Another data association problem

Nodes become tracklets

CDSC is used to stitch tracklets

Layer 2: Track Extraction

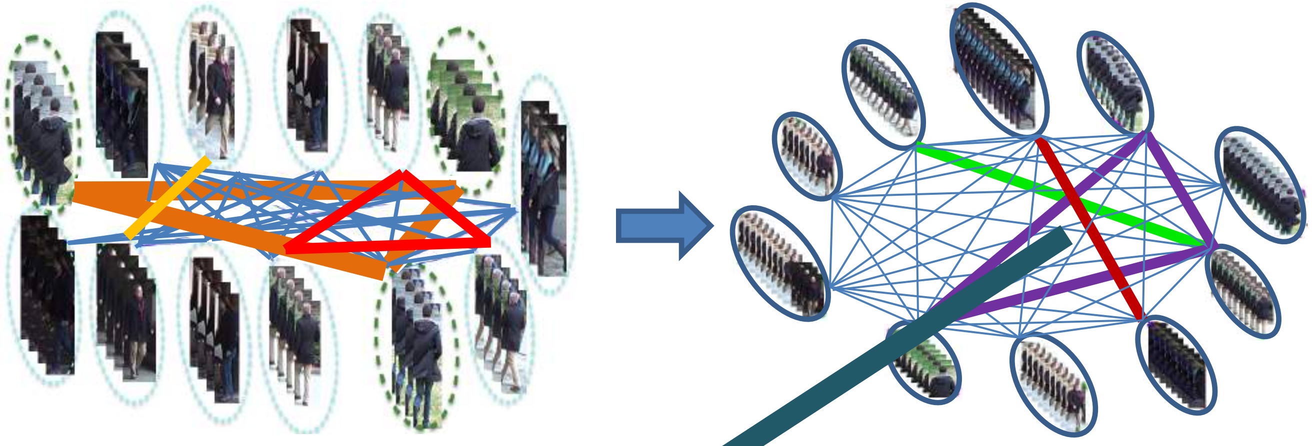


Another data association problem

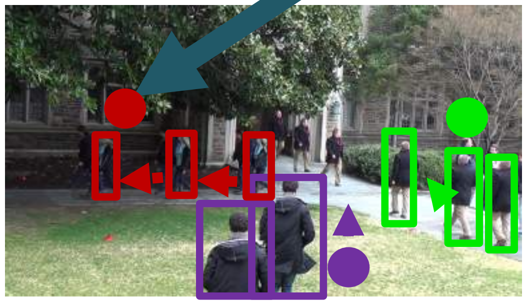
Nodes become tracklets

CDSC is used to stitch tracklets

Layer 2: Track Extraction



Tracks

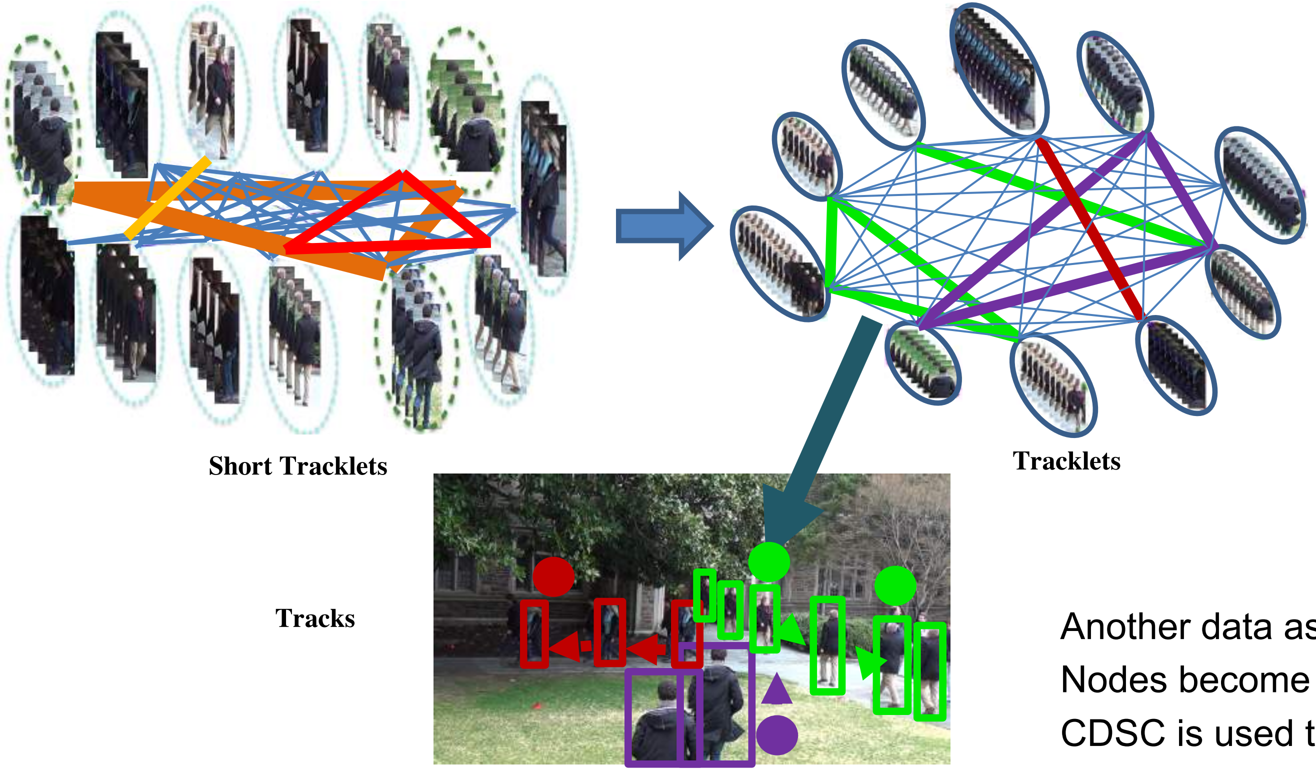


Another data association problem

Nodes become tracklets

CDSC is used to stitch tracklets

Layer 2: Track Extraction



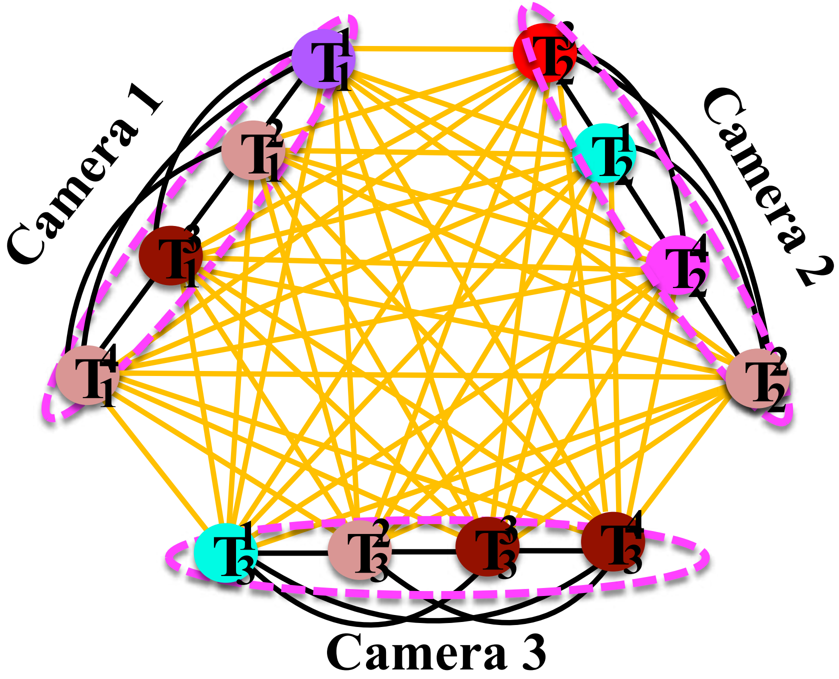
Another data association problem

Nodes become tracklets

CDSC is used to stitch tracklets

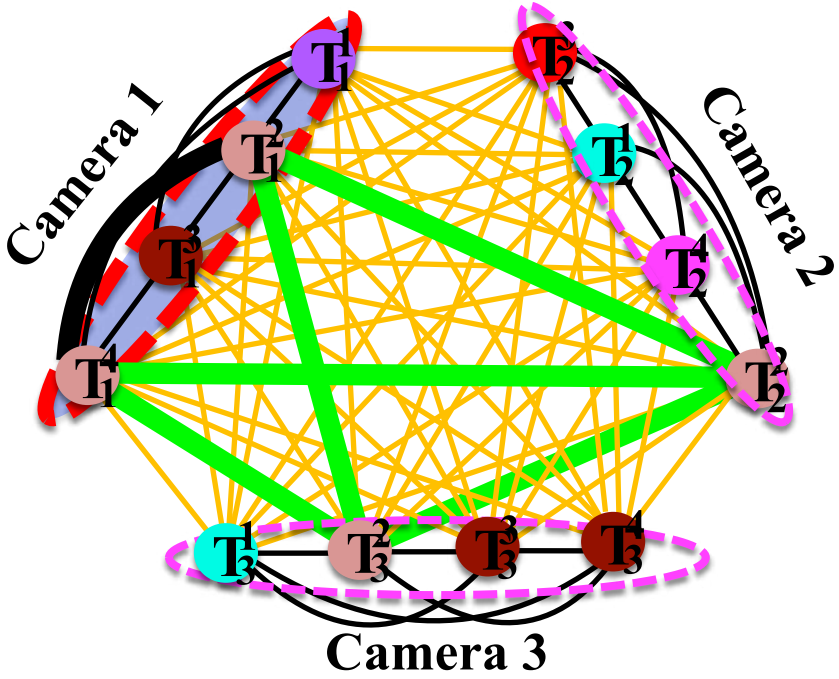
Within-Camera Tracking

Layer 3: Cross-Camera Association



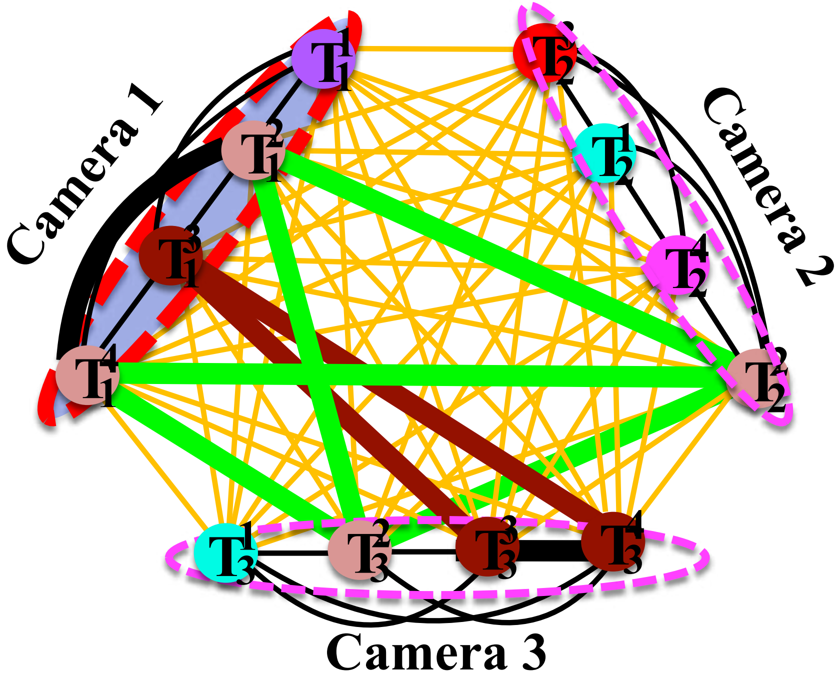
Tracks are nodes
Cameras as constraints

Layer 3: Cross-Camera Association



Tracks are nodes
Cameras as constraints

Layer 3: Cross-Camera Association



Tracks are nodes
Cameras as constraints

Results on DukeMTMC

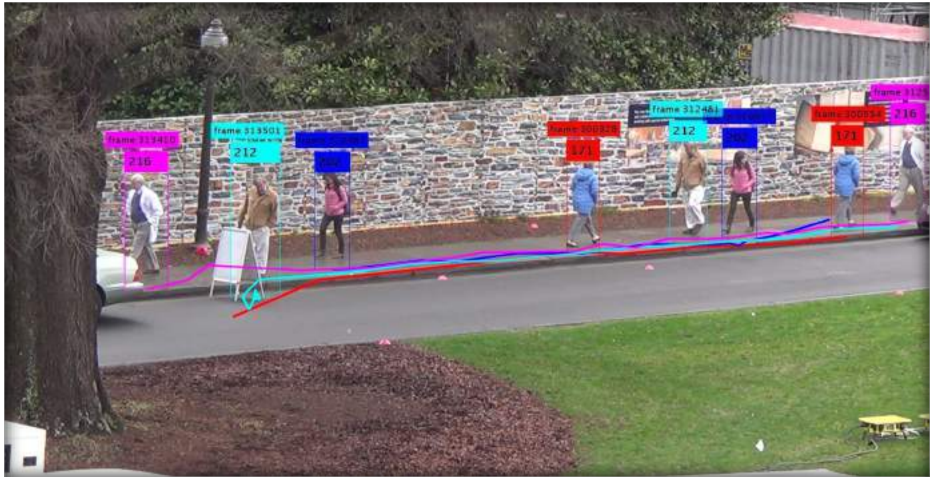
- Largest MTMC dataset (2016)
- 8 fixed synchronized cameras
- More than 2 million frames
- 0 to 54 persons per frame
- 2,700 Identities

	Methods	IDF1↑	IDP↑	IDR↑
Multi-Camera Test-easy	[33]	47.3	59.6	39.2
	[26]	32.9	41.3	27.3
	Ours	**50.9**	**63.2**	**42.6**

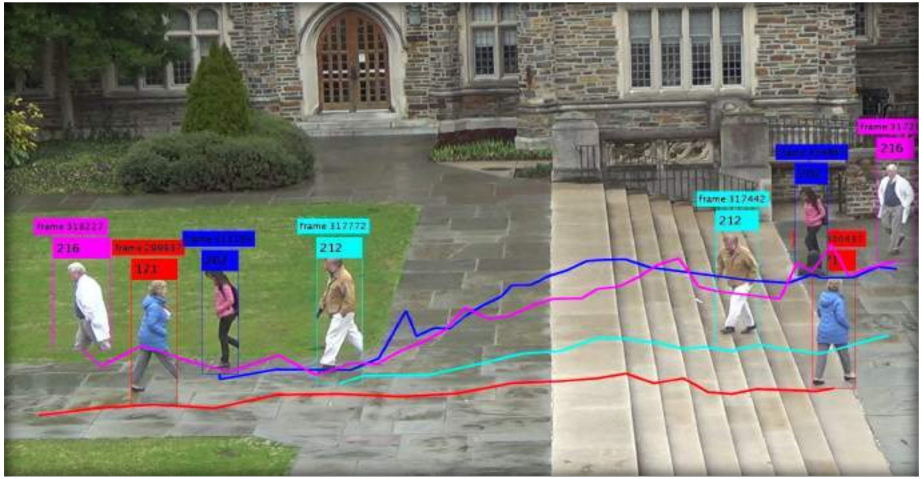
	Methods	IDF1↑	IDP↑	IDR↑
Multi-Camera Test-hard	[33]	56.2	67.0	48.4
	[26]	34.9	41.6	30.1
	Ours	**60.0**	**68.3**	**53.5**

IDP = Fraction of computed detections that are correctly identified
IDR = Fraction of ground-truth detections that are correctly identified
IDF1 = Ratio of correctly identified detections over the average number of ground-truth and computed detections

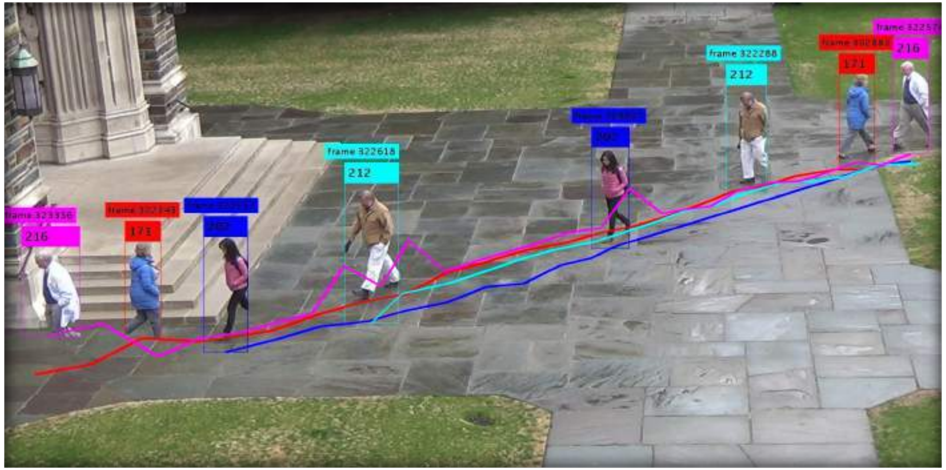
[33] E. Ristani et al. Performance measures and a data set for multi-target multi-camera tracking *(ECCV 2016)*
[26] A. Maksai et al. Non-Markovian globally consistent multi-object tracking *(ICCV 2017)*



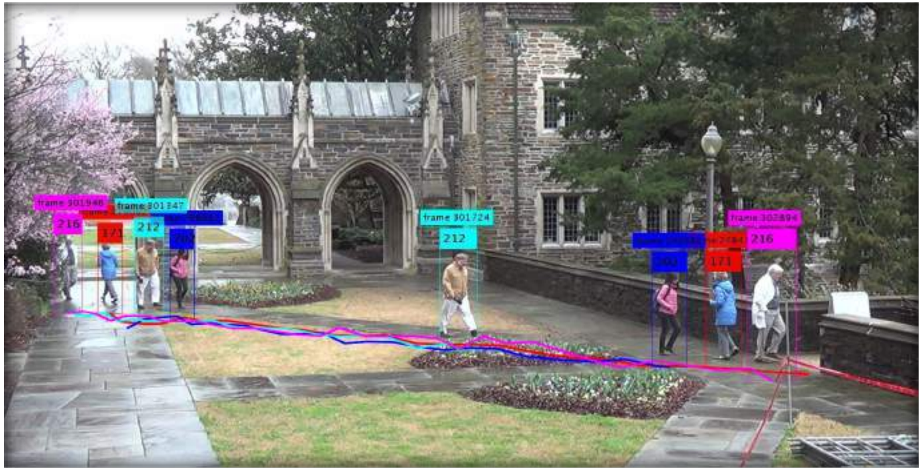
Camera 1



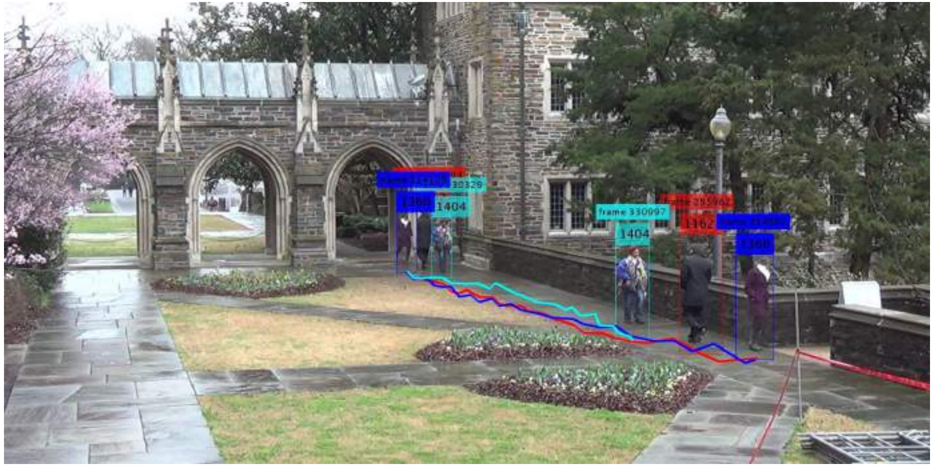
Camera 2



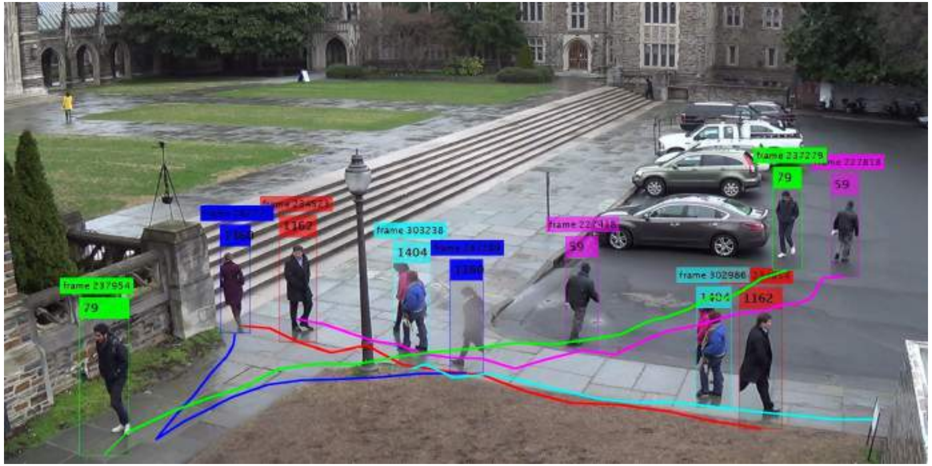
Camera 5



Camera 6



Camera 6



Camera 8

Camera 1

Camera 7

In a Nutshell

The dominant-set approach to clustering:

- ✓ does not require *a priori* knowledge on the number of clusters
- ✓ is robust against outliers
- ✓ allows to rank the cluster's elements according to "centrality"
- ✓ allows extracting overlapping clusters *(ICPR'08)*
- ✓ generalizes naturally to hypergraph clustering problems *(PAMI'13)*
- ✓ makes no assumption on the structure of the similarity matrix, (works also with asymmetric and even negative affinities)
- ✓ has a lot of applications ….

References

M. Pavan and M. Pelillo. Dominant sets and pairwise clustering. *PAMI* (2007)

S. Rota Bulò and M. Pelillo. A game-theoretic approach to hypergraph clustering. *PAMI* (2013)

E. Zemene, L. Tesfaye and M. Pelillo. Dominant sets for “constrained” image segmentation. *PAMI* (2018)

E. Zemene *et al.* Large-scale image geo-localization using dominant sets. *PAMI* (2018)

S. Vascon *et al.* Detecting conversational groups in images and sequences: A robust game-theoretic approach. *CVIU* (2016)

Y. Tariku *et al.* Multi-target tracking in multiple non-overlapping cameras using fast-constrained dominant sets. *IJCV* (to appear)

S. Rota Bulò and M. Pelillo. Dominant-set clustering: A review. *EJOR* (2017)