Media Integrity Analytics

Beyond Digital Forensics of Single Objects

Anderson Rocha (Associate Professor)

Microsoft Research Faculty Fellow
Google Faculty Research Awardee
Tan Chin Tuan Fellow
IEEE Senior Member

Reasoning for Complex Data (RECOD) Lab.
Institute of Computing,
University of Campinas (Unicamp)

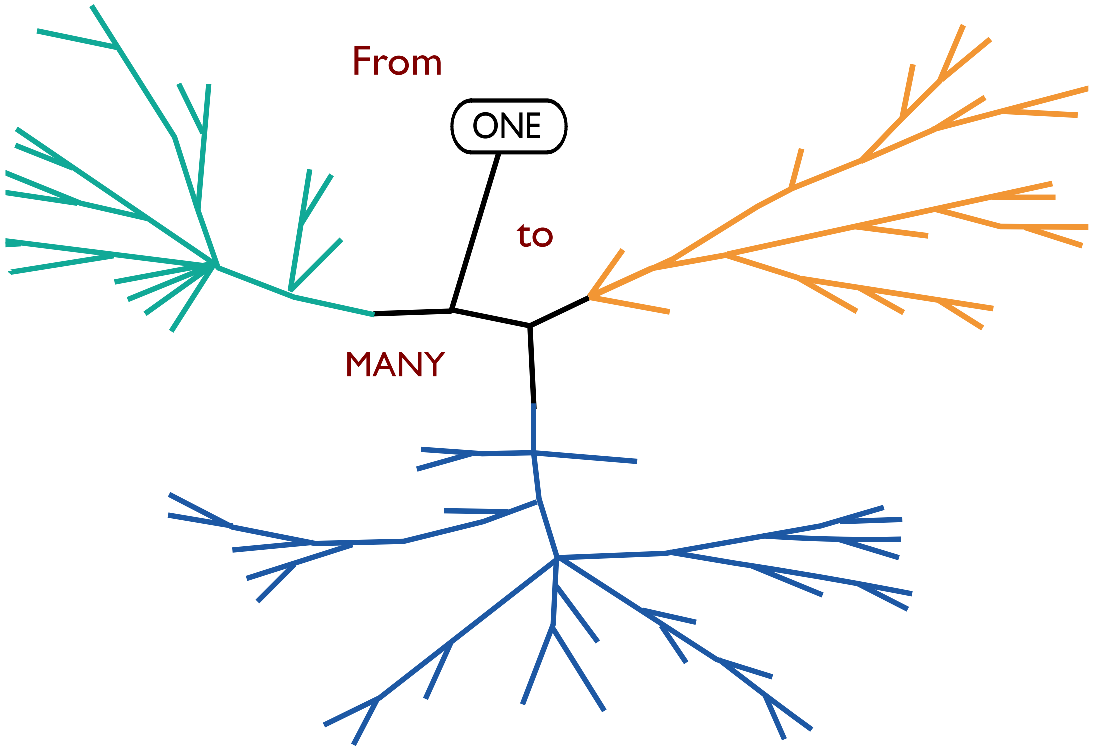
Av. Albert Einstein, 1251 - Cidade Universitária
CEP 13083-970 • Campinas/SP - Brasil

* *Joint work with several colleagues*
* *Special thanks to S. Goldenstein and M. Oikawa for helping crafting this presentation*

RECOD

Summary

- Who are we?
- Phylogeny Problem
- Image Phylogeny
- Video Phylogeny
- Current Challenges



What is this talk about?

What is this talk about?

- An important problem that has been mostly overlooked by the community.
- Has immediate applications in many areas.
- It is hard to solve.
- There is room for elegant math and different solutions.

Image Phylogeny by Minimal Spanning Trees
Z. Dias, A. Rocha, and S Goldenstein. *IEEE Transactions of Information Forensics and Security*, April 2012.

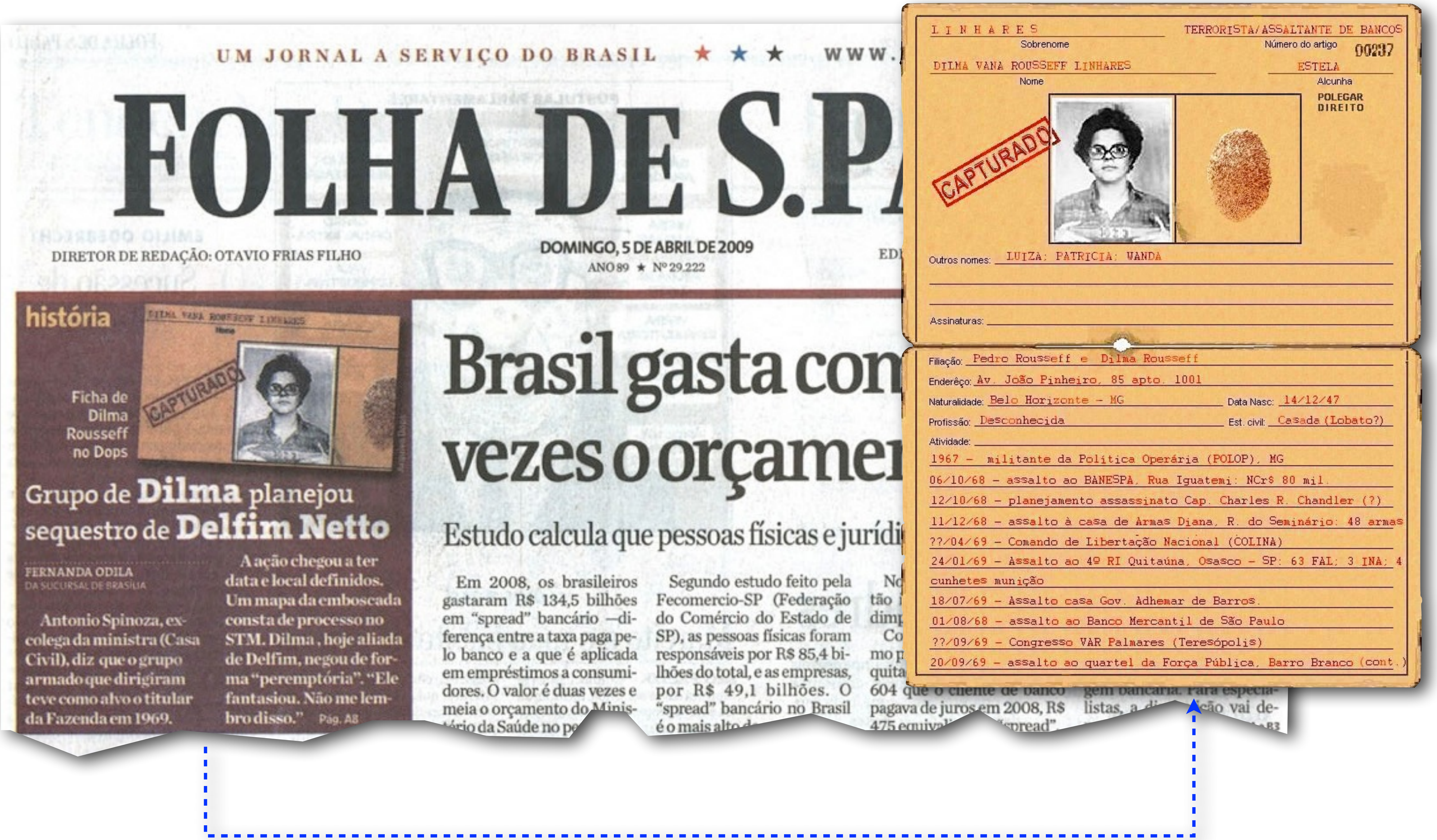
Video Phylogeny: Recovering Near-Duplicate Video Relationships
Z. Dias, A. Rocha, and S. Goldenstein. *IEEE Workshop on Information Forensics and Security (WIFS)*, 2011

How it started

- In 2009, the current Brazilian president was the president’s chief of staff, and the government pre-candidate for the 2010 presidential election.

- *Folha de SP*, a major Brazilian newspaper (think of NYT) ran an interview and article about her. They printed a “scan of her criminal records” as a political activist in the military dictatorship period (1964-1985), suggesting it as a record she engaged in violent armed activities (which she denies to this day).

High-Profile Analysis



UM JORNAL A SERVIÇO DO BRASIL ★ ★ ★ WWW.

FOLHA DE S.P

DIRETOR DE REDAÇÃO: OTAVIO FRIAS FILHO

DOMINGO, 5 DE ABRIL DE 2009
ANO 89 ★ Nº 29.222

ED

história

Ficha de Dilma Rousseff no Dops

Grupo de **Dilma** planejou sequestro de **Delfim Netto**

FERNANDA ODILA
DA SUCURSAL DE BRASÍLIA

Antonio Spinoza, ex-colega da ministra (Casa Civil), diz que o grupo armado que dirigiram teve como alvo o titular da Fazenda em 1969.

A ação chegou a ter data e local definidos. Um mapa da emboscada consta de processo no STM. Dilma , hoje aliada de Delfim, negou de forma "peremptória". "Ele fantasiou. Não me lembro disso." Pág. A8

Brasil gasta com vezes o orçamen

Estudo calcula que pessoas físicas e jurídi

Em 2008, os brasileiros gastaram R$ 134,5 bilhões em "spread" bancário —diferença entre a taxa paga pelo banco e a que é aplicada em empréstimos a consumidores. O valor é duas vezes e meia o orçamento do Ministério da Saúde no pe

Segundo estudo feito pela Fecomercio-SP (Federação do Comércio do Estado de SP), as pessoas físicas foram responsáveis por R$ 85,4 bilhões do total, e as empresas, por R$ 49,1 bilhões. O "spread" bancário no Brasil é o mais alto d

No tão i dimp Co mo p quita 604 que o cliente de banco pagava de juros em 2008, R$ 475 equivali "spread".

gem bancária. Para especialistas, a di ção vai de-

L I N H A R E S
Sobrenome

TERRORISTA/ASSALTANTE DE BANCOS
Número do artigo 00237

DILMA VANA ROUSSEFF LINHARES
Nome

ESTELA
Alcunha

POLEGAR DIREITO

CAPTURADO

Outros nomes: LUIZA; PATRICIA; WANDA

Assinaturas:

Filiação: Pedro Rousseff e Dilma Rousseff

Endereço: Av. João Pinheiro, 85 apto. 1001

Naturalidade: Belo Horizonte - MG Data Nasc: 14/12/47

Profissão: Desconhecida Est. civil: Casada (Lobato?)

Atividade:

1967 - militante da Política Operária (POLOP), MG

06/10/68 - assalto ao BANESPA, Rua Iguatemi: NCr$ 80 mil.

12/10/68 - planejamento assassinato Cap. Charles R. Chandler (?)

11/12/68 - assalto à casa de Armas Diana, R. do Seminário: 48 armas

??/04/69 - Comando de Libertação Nacional (COLINA)

24/01/69 - Assalto ao 4º RI Quitaúna, Osasco - SP: 63 FAL; 3 INA; 4 cunhetes munição

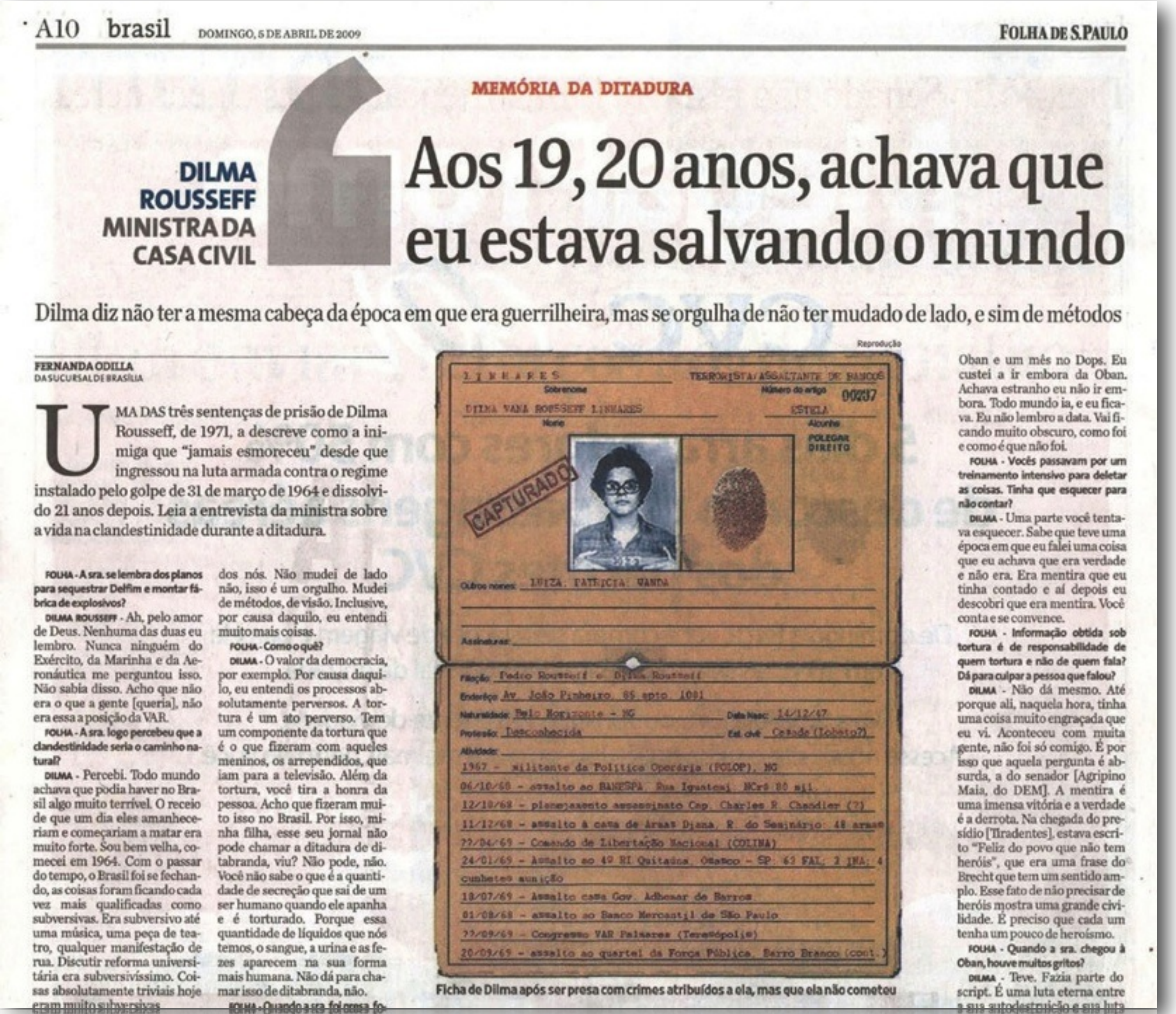
18/07/69 - Assalto casa Gov. Adhemar de Barros.

01/08/68 - assalto ao Banco Mercantil de São Paulo

??/09/69 - Congresso VAR Palmares (Teresópolis)

20/09/69 - assalto ao quartel da Força Pública, Barro Branco (cont.)

High-Profile Analysis



A10 brasil DOMINGO, 5 DE ABRIL DE 2009 FOLHA DE S.PAULO

MEMÓRIA DA DITADURA

DILMA ROUSSEFF
MINISTRA DA CASA CIVIL

"Aos 19, 20 anos, achava que eu estava salvando o mundo

Dilma diz não ter a mesma cabeça da época em que era guerrilheira, mas se orgulha de não ter mudado de lado, e sim de métodos

FERNANDA ODILLA
DA SUCURSAL DE BRASÍLIA

UMA DAS três sentenças de prisão de Dilma Rousseff, de 1971, a descreve como a inimiga que "jamais esmoreceu" desde que ingressou na luta armada contra o regime instalado pelo golpe de 31 de março de 1964 e dissolvido 21 anos depois. Leia a entrevista da ministra sobre a vida na clandestinidade durante a ditadura.

FOLHA - A sra. se lembra dos planos para sequestrar Delfim e montar fábrica de explosivos?

DILMA ROUSSEFF - Ah, pelo amor de Deus. Nenhuma das duas eu lembro. Nunca ninguém do Exército, da Marinha e da Aeronáutica me perguntou isso. Não sabia disso. Acho que não era o que a gente [queria], não era essa a posição da VAR.

FOLHA - A sra. logo percebeu que a clandestinidade seria o caminho natural?

DILMA - Percebi. Todo mundo achava que podia haver no Brasil algo muito terrível. O receio de que um dia eles amanheceriam e começariam a matar era muito forte. Sou bem velha, comecei em 1964. Com o passar do tempo, o Brasil foi se fechando, as coisas foram ficando cada vez mais qualificadas como subversivas. Era subversivo até uma música, uma peça de teatro, qualquer manifestação de rua. Discutir reforma universitária era subversivíssimo. Coisas absolutamente triviais hoje eram muito subversivas dos nós. Não mudei de lado não, isso é um orgulho. Mudei de métodos, de visão. Inclusive, por causa daquilo, eu entendi muito mais coisas.

FOLHA - Como o quê?

DILMA - O valor da democracia, por exemplo. Por causa daquilo, eu entendi os processos absolutamente perversos. A tortura é um ato perverso. Tem um componente da tortura que é o que fizeram com aqueles meninos, os arrependidos, que iam para a televisão. Além da tortura, você tira a honra da pessoa. Acho que fizeram muito isso no Brasil. Por isso, minha filha, esse seu jornal não pode chamar a ditadura de ditabranda, viu? Não pode, não. Você não sabe o que é a quantidade de secreção que sai de um ser humano quando ele apanha e é torturado. Porque essa quantidade de líquidos que nós temos, o sangue, a urina e as fezes aparecem na sua forma mais humana. Não dá para chamar isso de ditabranda, não.

FOLHA - Quando a sra. foi presa fo-

Reprodução

L I N H A R E S TERRORISTA/ASSALTANTE DE BANCOS
Sobrenome Número do artigo 00237
DILMA VANA ROUSSEFF LINHARES ESTELA
Nome Alcunha
POLEGAR DIREITO
CAPTURADO
Outros nomes: LUIZA, PATRICIA, WANDA
Assinaturas:
Filiação: Pedro Rousseff e Dilma Rousseff
Endereço: Av. João Pinheiro, 86 apto. 1001
Naturalidade: Belo Horizonte - MG Data Nasc: 14/12/47
Profissão: Desconhecida Est. civil: Casada (Lobato?)
Atividade:
1967 - militante da Política Operária (POLOP), MG
06/10/68 - assalto ao BANESPA, Rua Iguatemi, NCr$ 80 mil.
12/10/68 - planejamento assassinato Cap. Charles R. Chandler (?)
11/12/68 - assalto à casa de Armas Diana, R. do Seminário: 48 armas
??/04/69 - Comando de Libertação Nacional (COLINA)
24/01/69 - Assalto ao 4º RI Quitaúna, Osasco - SP: 63 FAL; 3 INA; 4 cunhetes munição
18/07/69 - Assalto casa Gov. Adhemar de Barros
01/08/68 - assalto ao Banco Mercantil de São Paulo
??/09/69 - Congresso VAR Palmares (Teresópolis)
20/09/69 - assalto ao quartel da Força Pública, Barro Branco (cont.)

Ficha de Dilma após ser presa com crimes atribuídos a ela, mas que ela não cometeu

Oban e um mês no Dops. Eu custei a ir embora da Oban. Achava estranho eu não ir embora. Todo mundo ia, e eu ficava. Eu não lembro a data. Vai ficando muito obscuro, como foi e como é que não foi.

FOLHA - Vocês passavam por um treinamento intensivo para deletar as coisas. Tinha que esquecer para não contar?

DILMA - Uma parte você tentava esquecer. Sabe que teve uma época em que eu falei uma coisa que eu achava que era verdade e não era. Era mentira que eu tinha contado e aí depois eu descobri que era mentira. Você conta e se convence.

FOLHA - Informação obtida sob tortura é de responsabilidade de quem tortura e não de quem fala? Dá para culpar a pessoa que falou?

DILMA - Não dá mesmo. Até porque ali, naquela hora, tinha uma coisa muito engraçada que eu vi. Aconteceu com muita gente, não foi só comigo. É por isso que aquela pergunta é absurda, a do senador [Agripino Maia, do DEM]. A mentira é uma imensa vitória e a verdade é a derrota. Na chegada do presídio [Tiradentes], estava escrito "Feliz do povo que não tem heróis", que era uma frase do Brecht que tem um sentido amplo. Esse fato de não precisar de heróis mostra uma grande civilidade. É preciso que cada um tenha um pouco de heroísmo.

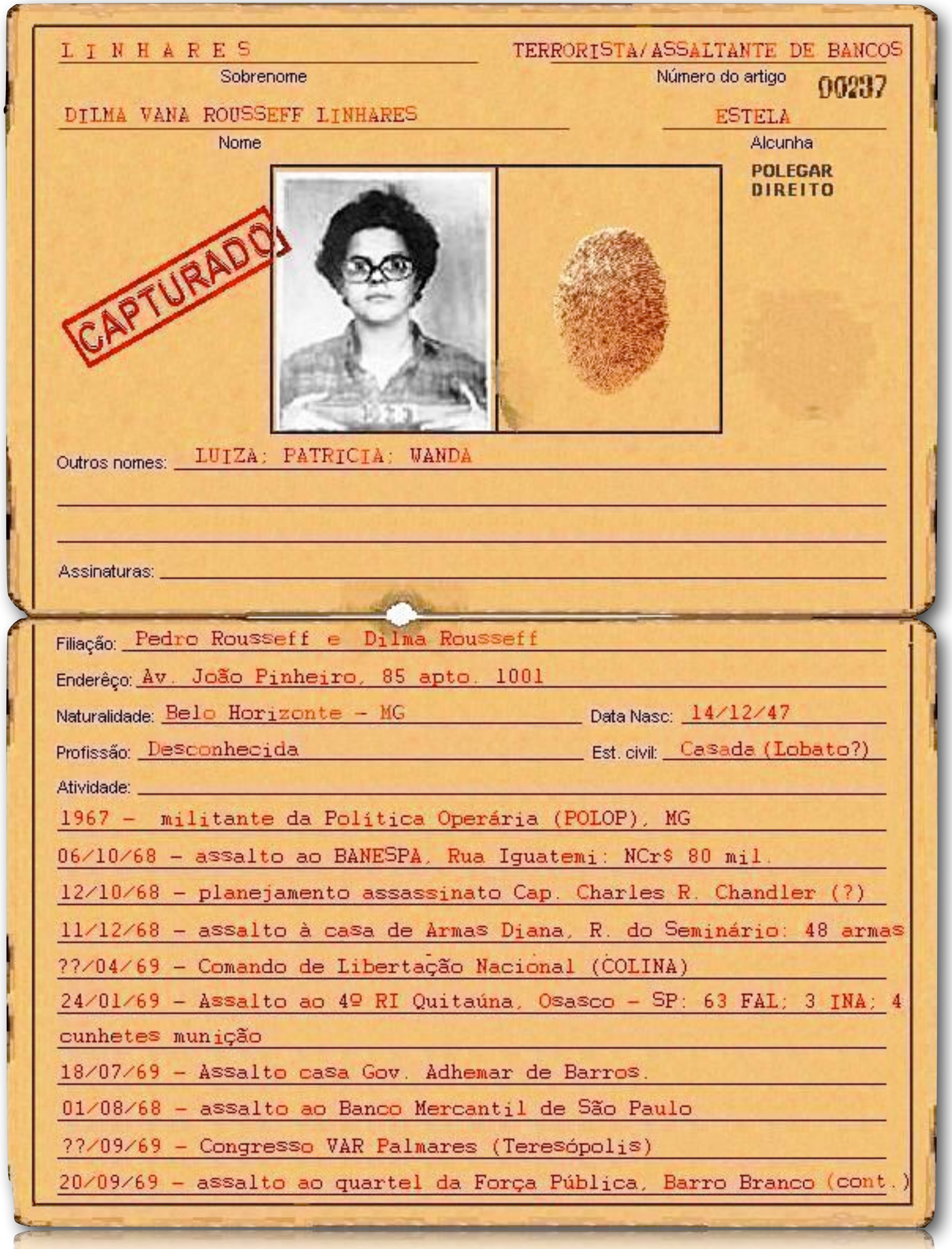
FOLHA - Quando a sra. chegou à Oban, houve muitos gritos?

DILMA - Teve. Fazia parte do script. É uma luta eterna entre a sua autodestruição e sua luta

Criminal Records?

A "scan" of her personal files maintained by the military internal security during the Brazilian military regime.

The Public Archive of SP actually hosts such a collection.



LINHARES
Sobrenome

TERRORISTA/ASSALTANTE DE BANCOS
Número do artigo 00237

DILMA VANA ROUSSEFF LINHARES
Nome

ESTELA
Alcunha

CAPTURADO

POLEGAR DIREITO

Outros nomes: LUIZA; PATRICIA; WANDA

Assinaturas:

Filiação: Pedro Rousseff e Dilma Rousseff

Enderêço: Av. João Pinheiro, 85 apto. 1001

Naturalidade: Belo Horizonte - MG   Data Nasc: 14/12/47

Profissão: Desconhecida   Est. civil: Casada (Lobato?)

Atividade:

1967 - militante da Política Operária (POLOP), MG

06/10/68 - assalto ao BANESPA, Rua Iguatemi: NCr$ 80 mil.

12/10/68 - planejamento assassinato Cap. Charles R. Chandler (?)

11/12/68 - assalto à casa de Armas Diana, R. do Seminário: 48 armas

??/04/69 - Comando de Libertação Nacional (COLINA)

24/01/69 - Assalto ao 4º RI Quitaúna, Osasco - SP: 63 FAL; 3 INA; 4 cunhetes munição

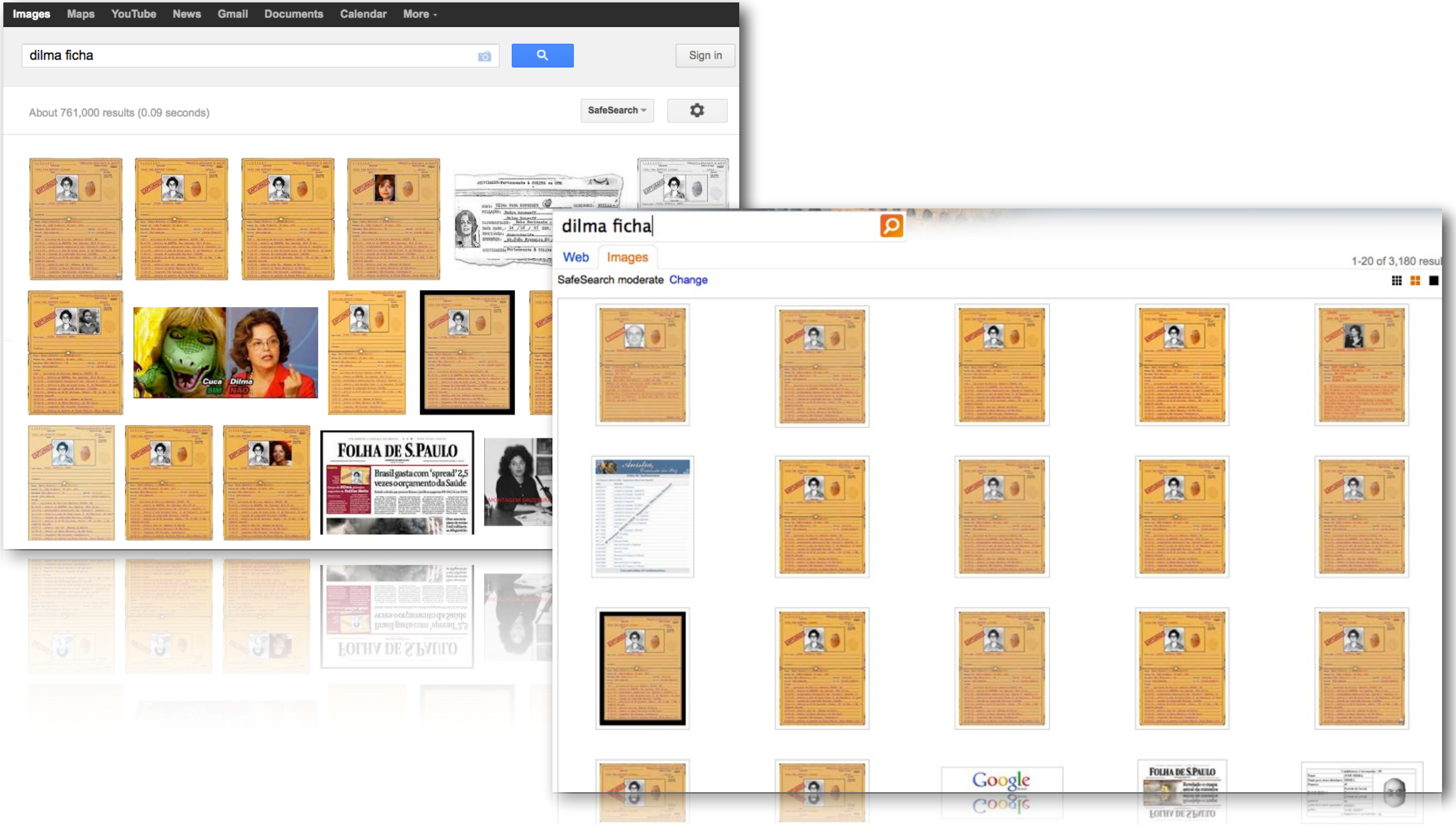
18/07/69 - Assalto casa Gov. Adhemar de Barros.

01/08/68 - assalto ao Banco Mercantil de São Paulo

??/09/69 - Congresso VAR Palmares (Teresópolis)

20/09/69 - assalto ao quartel da Força Pública, Barro Branco (cont.)

Searching the Web...



dilma ficha

About 761,000 results (0.09 seconds)

dilma ficha

1-20 of 3,180 resul

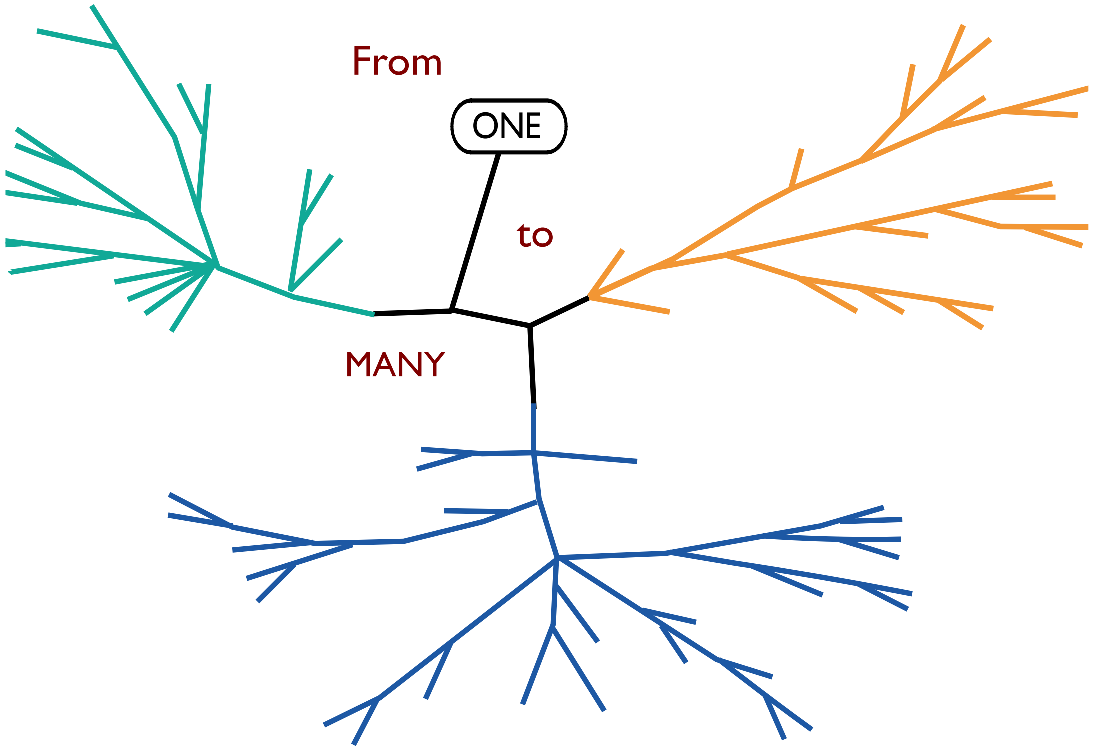
SafeSearch moderate Change

Criminal Records?

- This image was already going around the net for about six months – it is a clear fake.
- She hired us, as consultants, to provide a forensic analysis of file's authenticity that could hold on court.
- There were several versions of the image (near duplicates)
  - Which one was the original?
  - Where should we perform the analysis?

How to find the original?

- The images are “copied” around...
  - resized;
  - cropped;
  - color corrected;
  - recompressed;
  - and possibly other transformations.



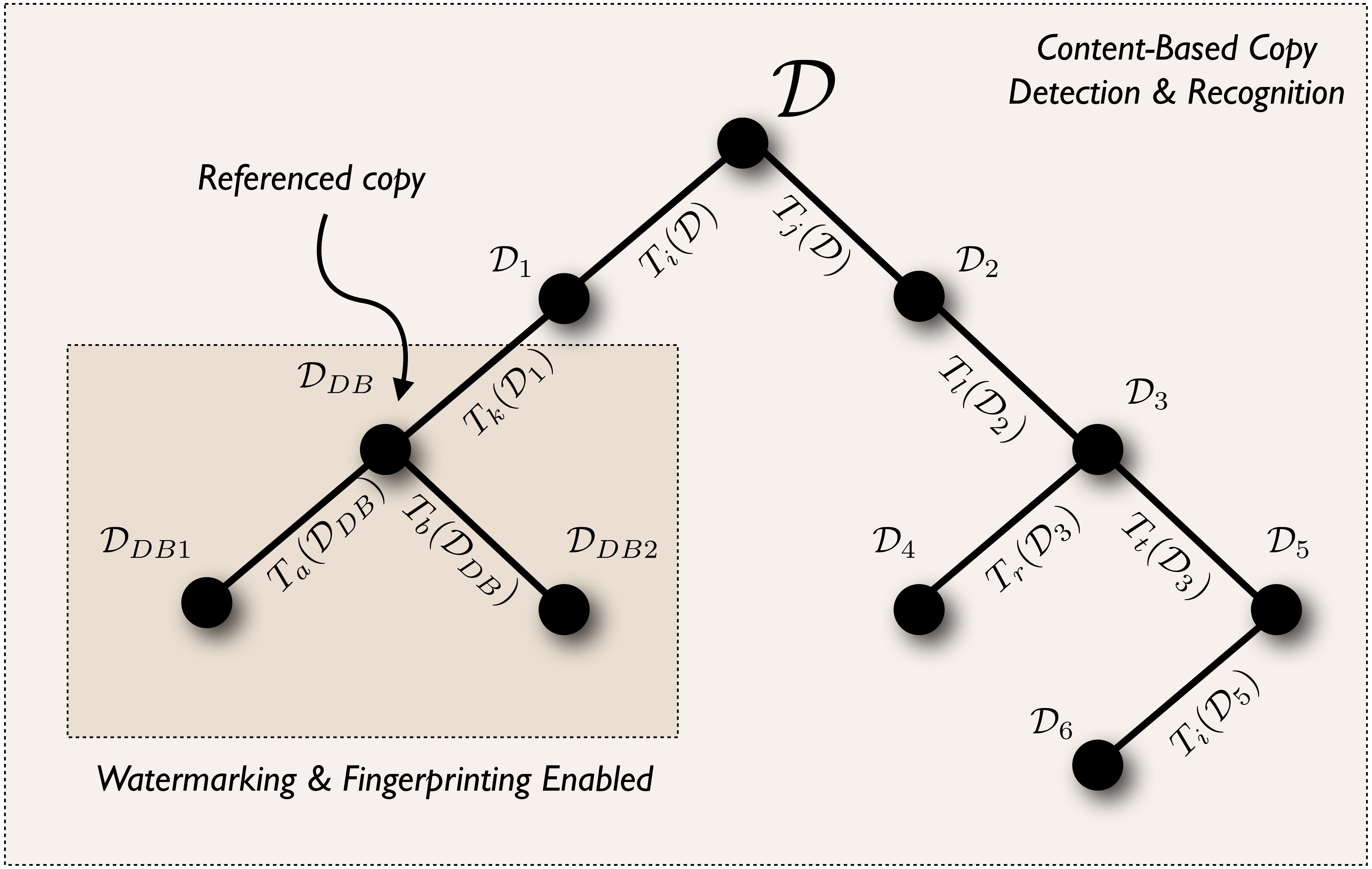
Media Phylogeny

Media Phylogeny

- Identify, among a set of near duplications, which element is the original, and the structure of generation of each near duplication.

- Tells the history of the transformations created the duplications.

Media Phylogeny



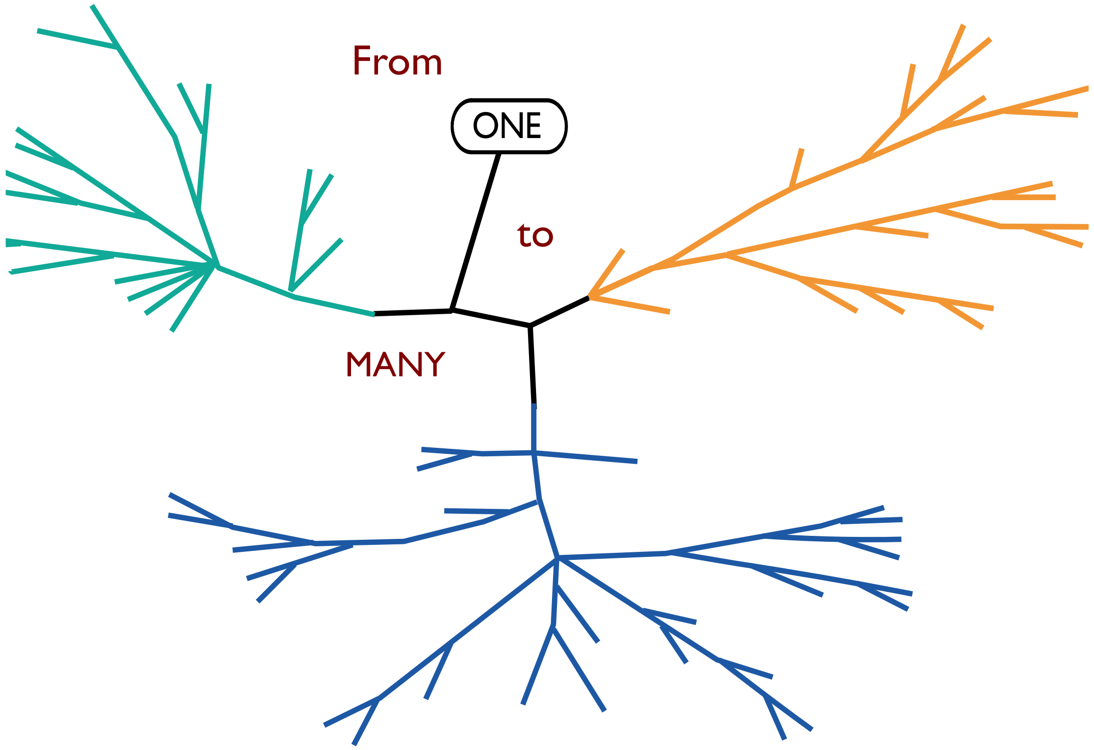


Image Phylogeny Trees
IPT

Image Phylogeny Trees: IPT

- **Security**.
- **Forensics**.
- **Copyright enforcement**.
- **News tracking services**.
- **Indexing**.

Image Phylogeny Trees: IPT

- **Security**: the modification graph provides information of suspects' behavior, and points out flow of content distribution.
- **Forensics**.
- **Copyright enforcement**.
- **News tracking services**.
- **Indexing**.

Image Phylogeny Trees: IPT

- **Security**.
- **Forensics**: analysis in the original document (root of the tree) instead of in a near duplicate.
- **Copyright enforcement**.
- **News tracking services**.
- **Indexing**.

Image Phylogeny Trees: IPT

- **Security**.
- **Forensics**.
- **Copyright enforcement**: traitor tracing without the need of source control techniques (watermarking or fingerprinting).
- **News tracking services**.
- **Indexing**.

Image Phylogeny Trees: IPT

- **Security**.
- **Forensics**.
- **Copyright enforcement**.
- **News tracking services**: the ND relationships can feed news tracking services with key elements for determining the opinion forming process across time and space.
- **Indexing**.

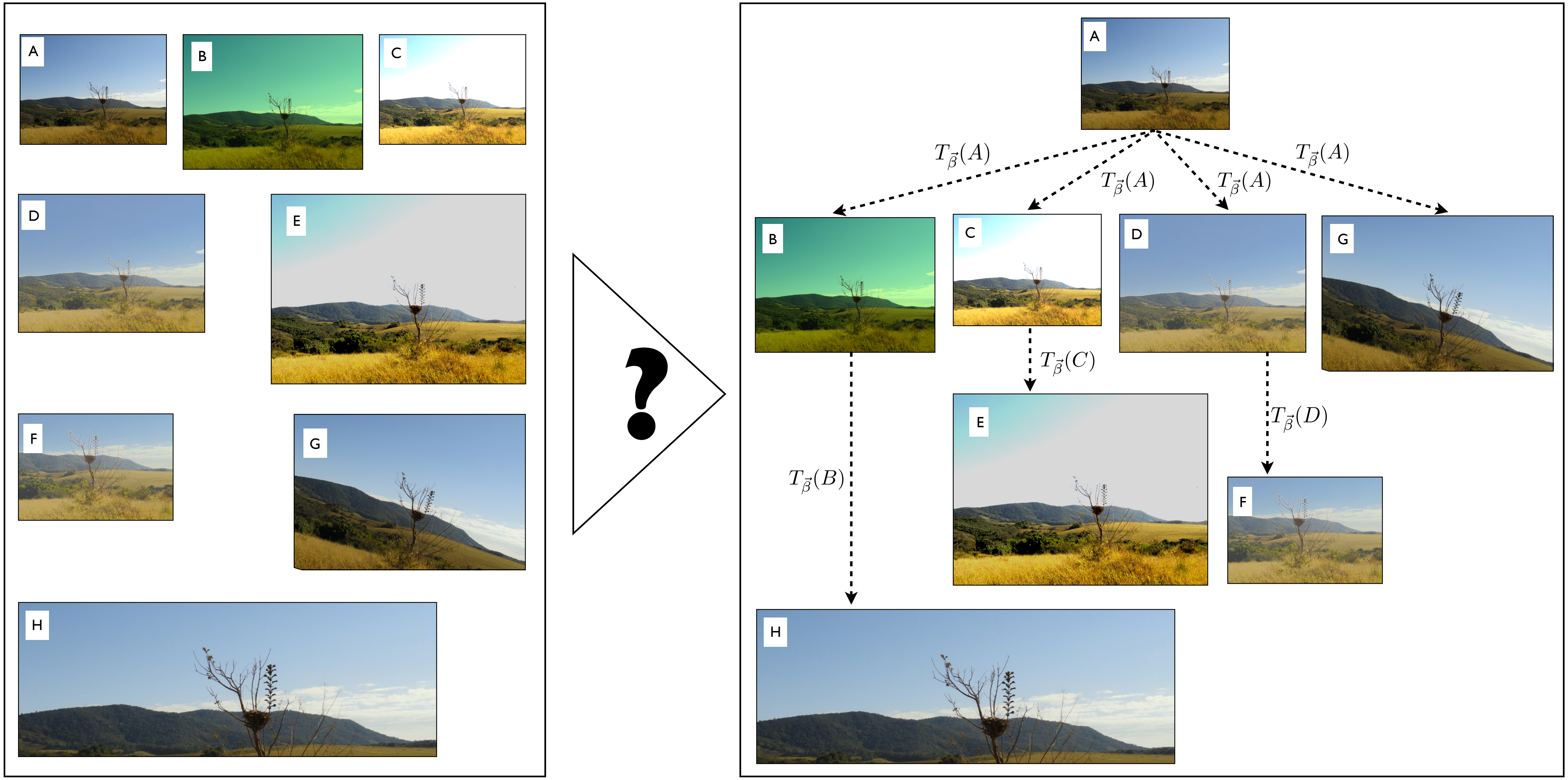
Image Phylogeny Trees: IPT

- **Security**.
- **Forensics**.
- **Copyright enforcement**.
- **News tracking services**.
- **Indexing**: tree root can give us an image from an ND set as a representative to index, store, or even further refine the ND search.
  Tree structure might help indexing and retrieving.

Our Objective

Image Near Duplicates

Phylogeny Tree



Two Subproblems

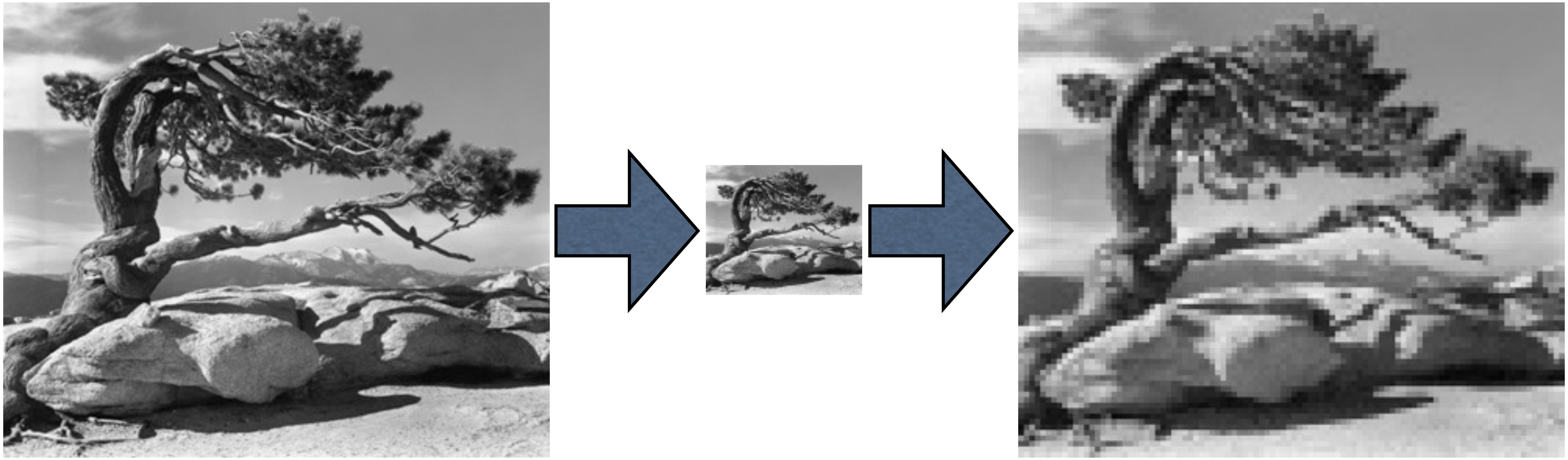
1. Define good dissimilarity functions $d(i, j)$ between images.

2. Develop algorithms that construct the Image Phylogeny Tree given a dissimilarity matrix of the images.

Dissimilarity

Dissimilarity

The dissimilarity is not a metric - we want to estimate how likely A→B and B→A.



Think about cropping, or resizing an image - these are not two-way operations.

Jeffrey Pine, Sentinel Dome, Yosemite National Park, **Ansel Adams**.

Dissimilarity

- Define a family of image transformations $T_\beta(I)$ parameterized by β.

- Let $d_\beta(i,j) = |I_j - T_\beta(I_i)|^2$ find β_{min} that minimizes $d_\beta(i,j)$

$$d(i,j) = d_{\beta_{min}}(i,j)$$

Dissimilarity

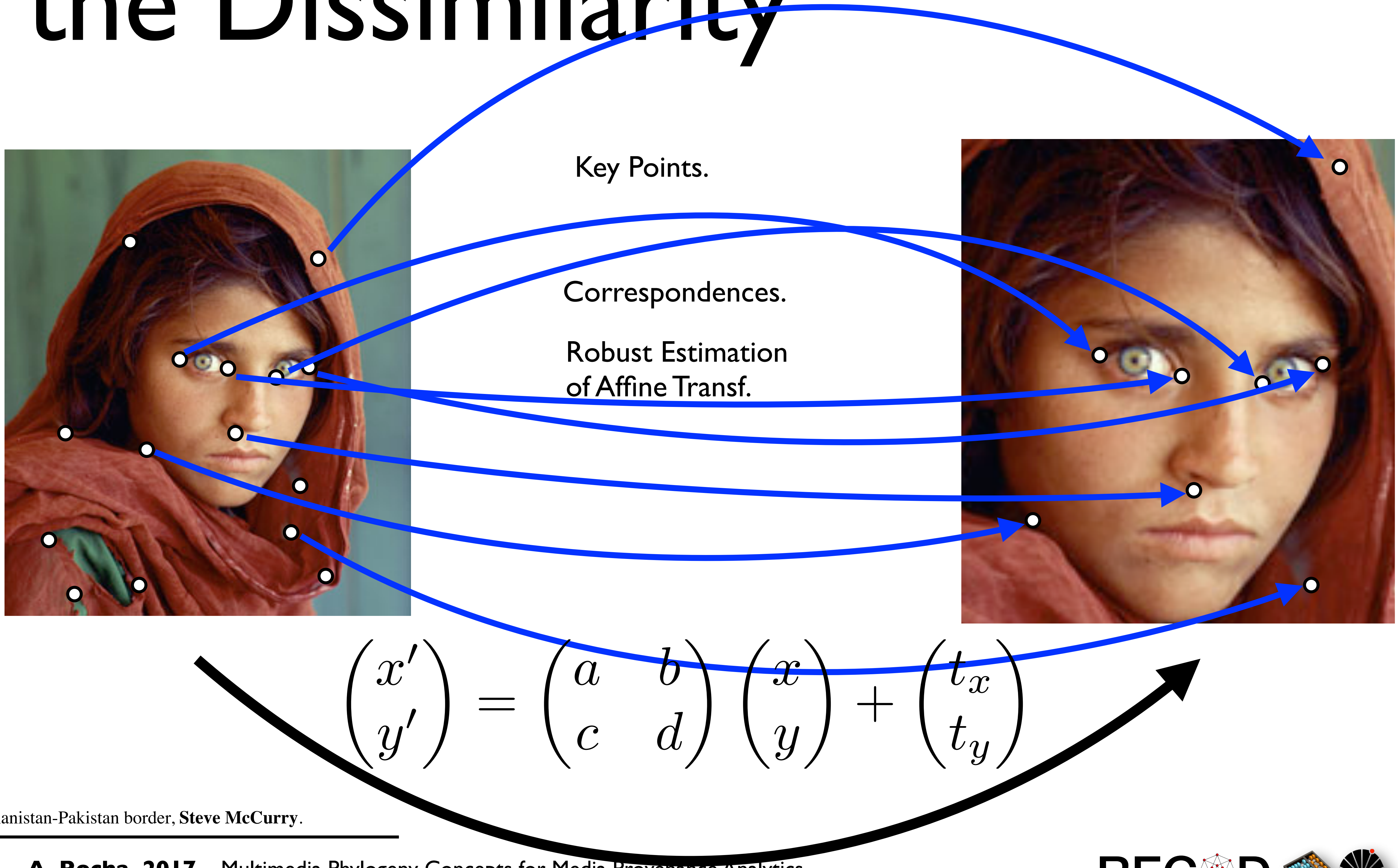
$$T_{\beta}(I) = T_{jpeg}(T_{color}(T_{spatial}(I)))$$

- We use a composition of three simple steps.
- In the general case, finding the optimum parameters of a general transformation might be a complicated optimization.

Dissimilarity

- Spatial
  - Affine Transformation
  - Cropping
- Color
  - Channel Brightness and Contrast.
- JPEG Compression
  - Quantization tables.

Spatial Transformation for the Dissimilarity



Key Points.

Correspondences.

Robust Estimation of Affine Transf.

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

Afghanistan-Pakistan border, **Steve McCurry**.

Dissimilarity: Compression

- Use the quantization table of the jpeg of B to compress T(A).

Tree Reconstruction

Tree Construction

- Local decisions of direction on pairs of images is not a good idea...

- Proposition: we want a MST.

  ...but we have a complete directed graph.

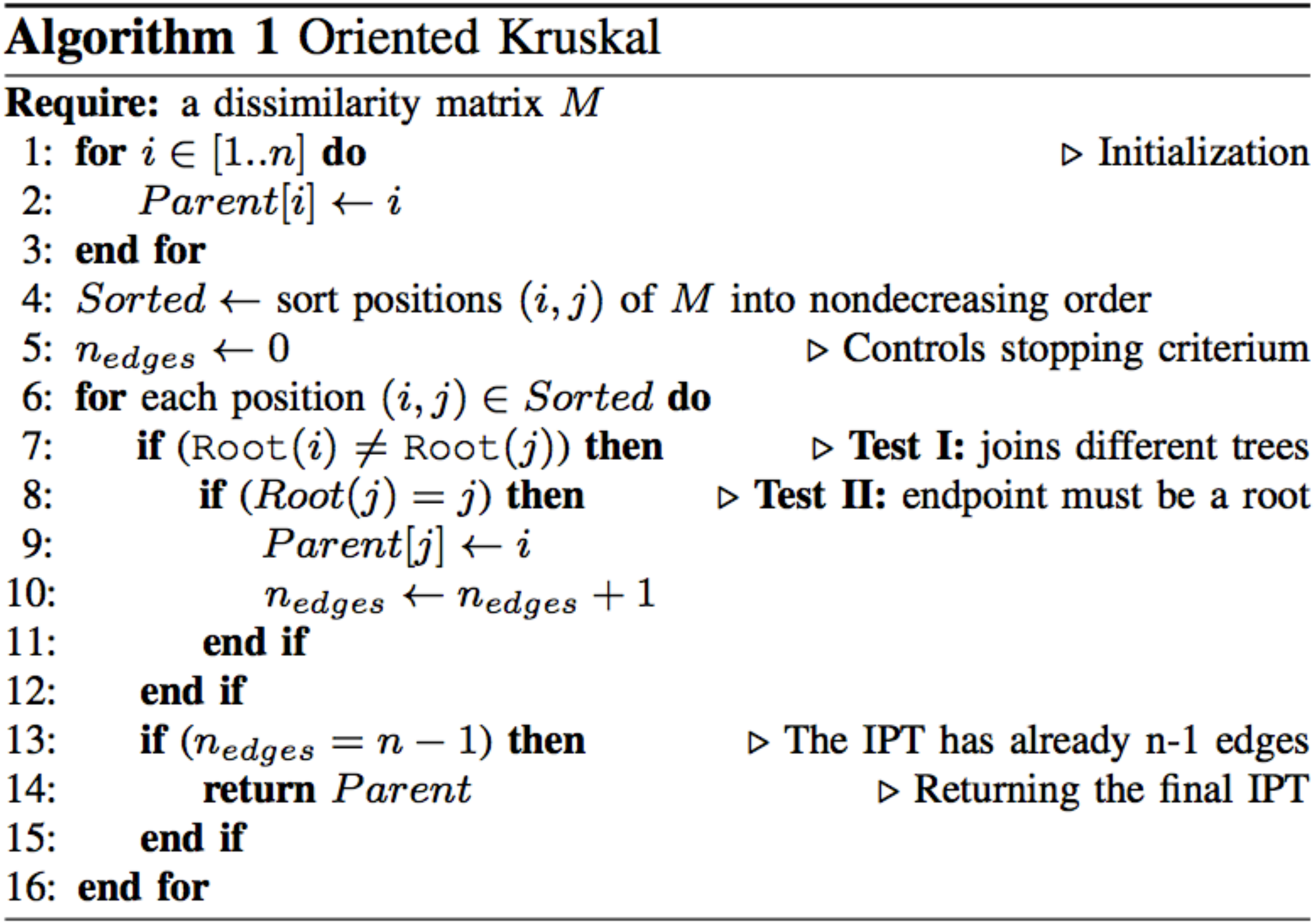
MST of directed graphs in the Literature

The *Optimum Branching* problem finds the MST of a directed graph for a given root.

In our context, it would have to be applied to each vertex as a root, and the final complexity in our scenario would be $O(n^3)$.

It also uses a Fibonacci Heap.

Oriented Kruskal



Algorithm 1 Oriented Kruskal

```
Require: a dissimilarity matrix $M$
for $i \in [1..n]$ do                                        ▷ Initialization
    $Parent[i] \leftarrow i$
end for
$Sorted \leftarrow$ sort positions $(i, j)$ of $M$ into nondecreasing order
$n_{edges} \leftarrow 0$                                     ▷ Controls stopping criterium
for each position $(i, j) \in Sorted$ do
    if $(\texttt{Root}(i) \neq \texttt{Root}(j))$ then       ▷ Test I: joins different trees
        if $(Root(j) = j)$ then                              ▷ Test II: endpoint must be a root
            $Parent[j] \leftarrow i$
            $n_{edges} \leftarrow n_{edges} + 1$
        end if
    end if
    if $(n_{edges} = n - 1)$ then                            ▷ The IPT has already n-1 edges
        return $Parent$                                      ▷ Returning the final IPT
    end if
end for
```

Oriented Kruskal

Our method runs once, and finds both the root and structure simultaneously.

It has an $O(n^2 \log n)$ complexity – we need to sort all n^2 edges of the complete graph.

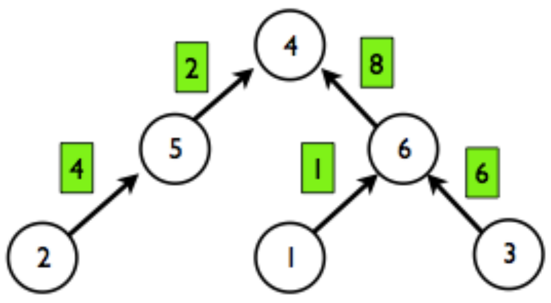
It requires the Union-Find data structure.

Construction Example

Dissimilarity Matrix

M	1	2	3	4	5	6
1	-	21	47	27	35	39
2	21	-	23	13	19	22
3	41	31	-	32	27	28
4	6	26	18	-	5	17
5	25	8	44	20	-	44
6	2	30	12	50	9	-

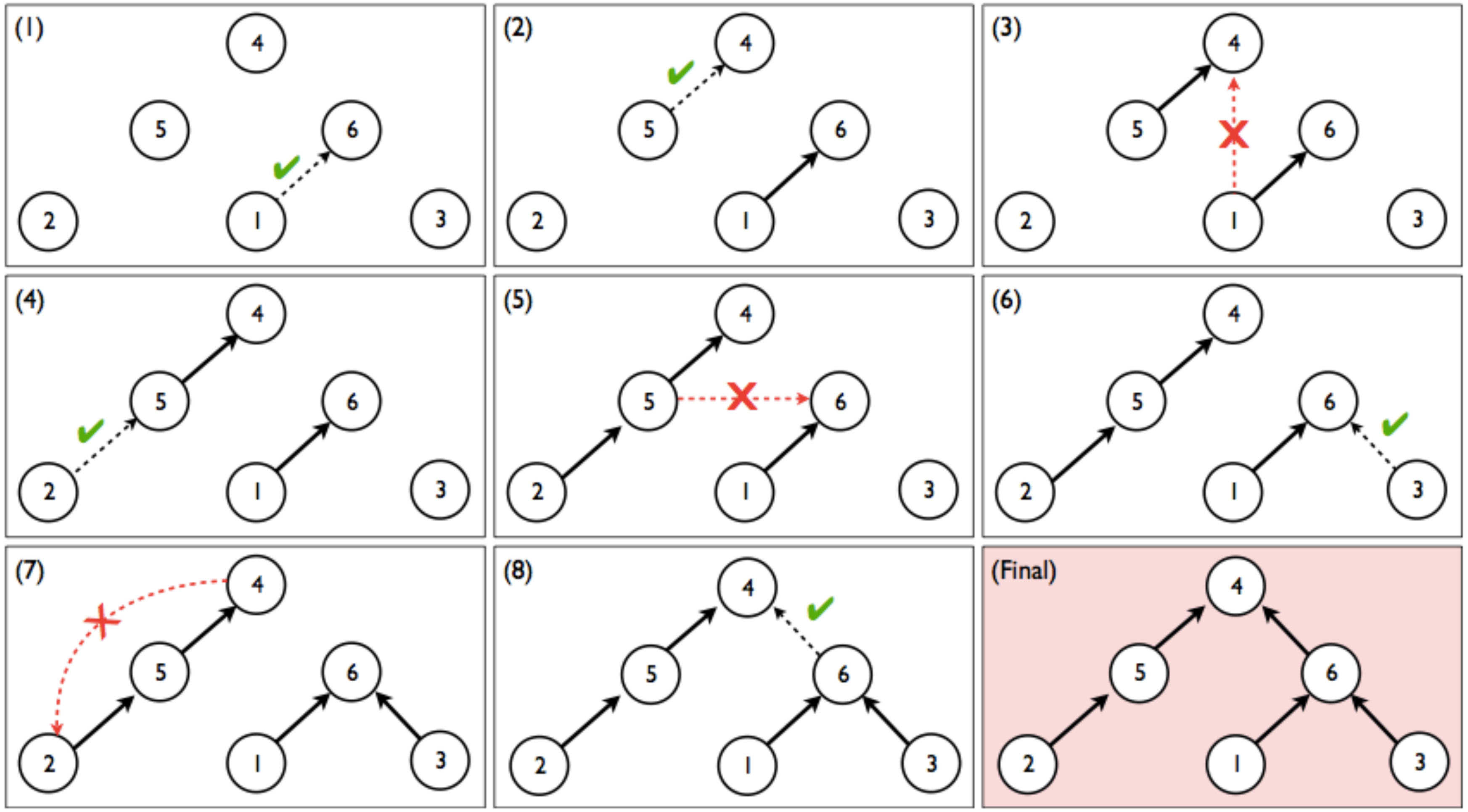
Reconstructed Tree [6, 5, 6, 4, 4, 4]

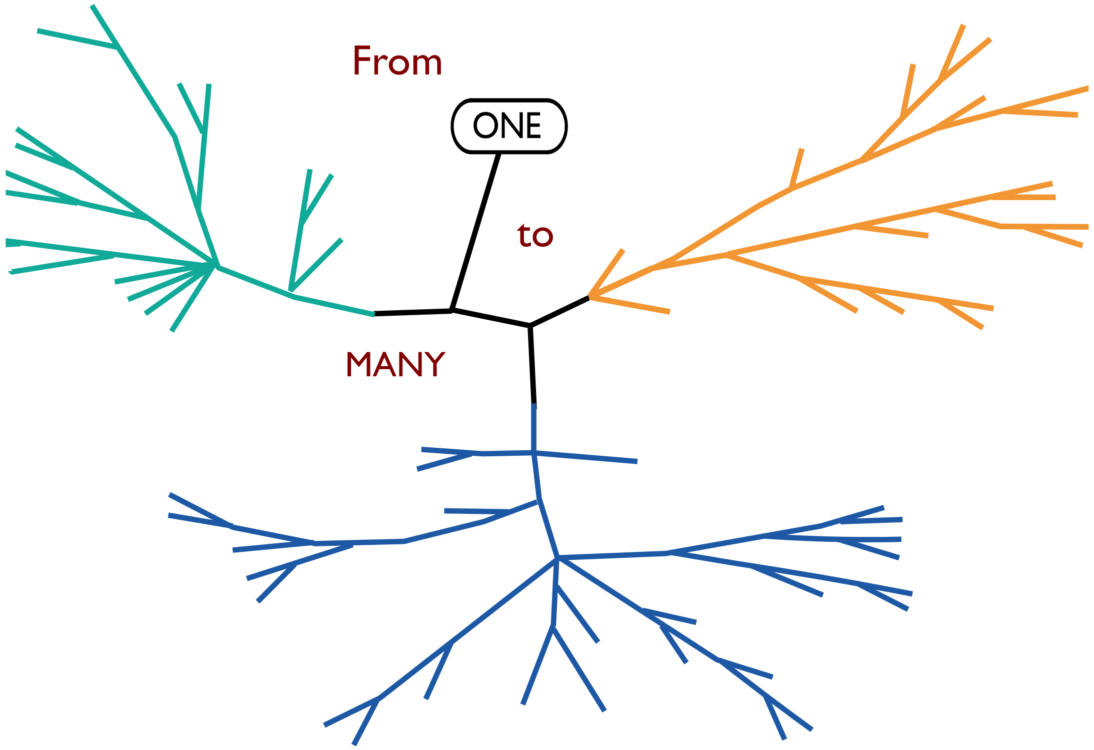


Algorithm Steps

1	M[6,1] = 2	✔	Select Edge (1, 6)
2	M[4,5] = 5	✔	Select Edge (5, 4)
3	M[4,1] = 6	✕	**Test II:** Root(1) = 6
4	M[5,2] = 8	✔	Select Edge (2, 5)
5	M[6,5] = 9	✕	**Test II:** Root(5) = 4
6	M[6,3] = 12	✔	Select Edge (3, 6)
7	M[2,4] = 13	✕	**Test I:** Root(2) = Root(4)
8	M[4,6] = 17	✔	Select Edge (6, 4)

Construction Breakdown





Evaluation

Evaluation: comparing Trees

$$\textbf{Root:}\ R(\mathrm{IPT}_1, \mathrm{IPT}_2) = \begin{cases} 1, & \text{If } \texttt{Root(IPT}_1\texttt{) = Root(IPT}_2\texttt{)} \\ 0, & \text{Otherwise} \end{cases}$$

$$\textbf{Edges:}\ E(\mathrm{IPT}_1, \mathrm{IPT}_2) = \frac{|E_1 \cap E_2|}{n-1}$$

$$\textbf{Leaves:}\ L(\mathrm{IPT}_1, \mathrm{IPT}_2) = \frac{|L_1 \cap L_2|}{|L_1 \cup L_2|}$$

$$\textbf{Ancestry:}\ A(\mathrm{IPT}_1, \mathrm{IPT}_2) = \frac{|A_1 \cap A_2|}{|A_1 \cup A_2|}$$

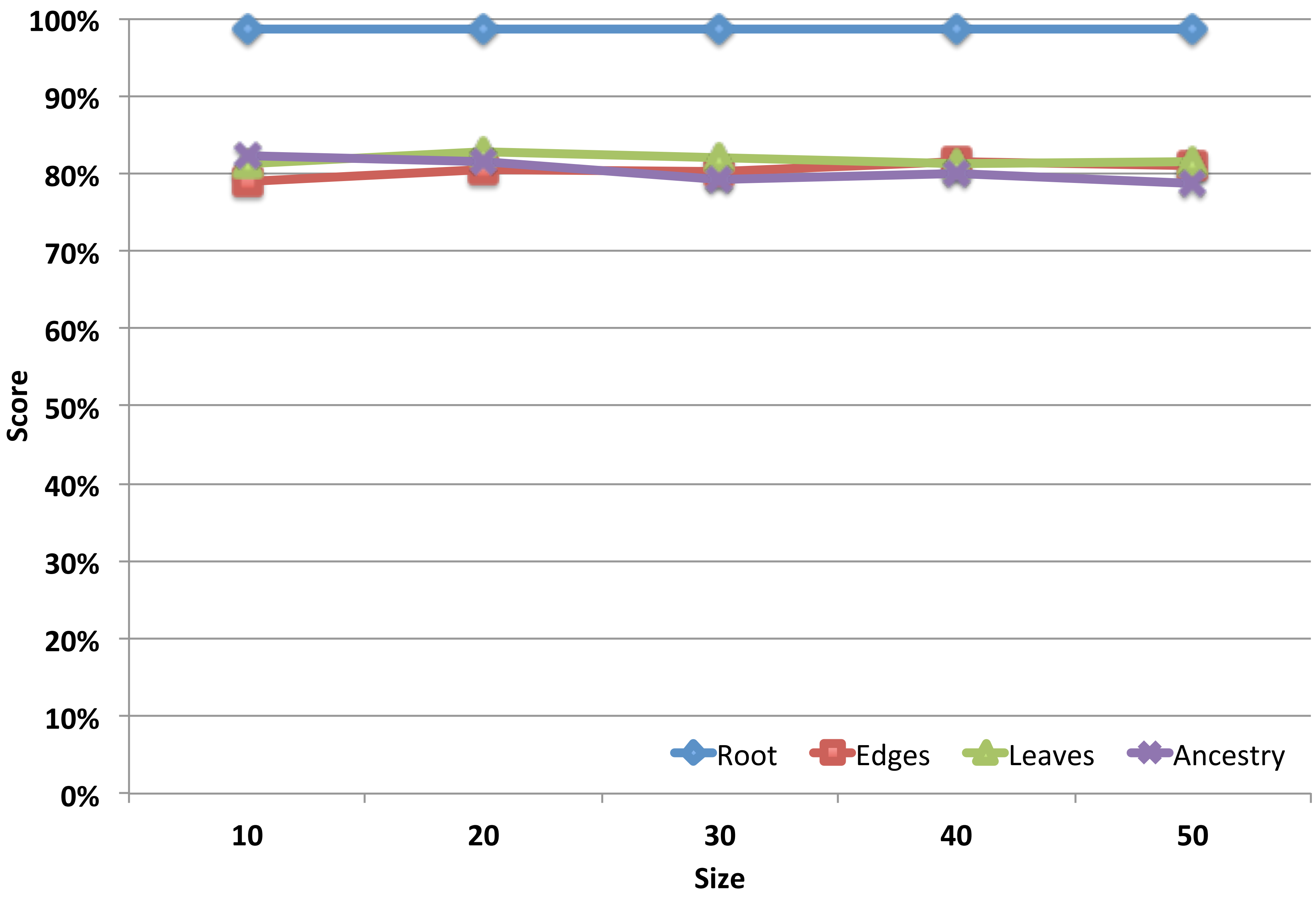
IPT Experiments

- Experimental Setup.
- Complete Trees.
- Missing Nodes.
  - Missing Root.
  - Missing Internal Nodes.
- Real ND sets from the Web.
- A first look at Forests.

Experimental Setup

- 50 raw images from UCID.
- Trees with 10, 20, 30, 40, and 50 nodes.
- For every size, 50 random tree topologies, each with 10 different random parameters.
- ND set created with affine transformation, crop, brightness-contrast-gamma on each channel and compression. We use ImageMagick.
- Dissimilarity construction with OpenCV and libjpg: affine transformation, brightness-contrast by channel, and compression.

Complete Trees

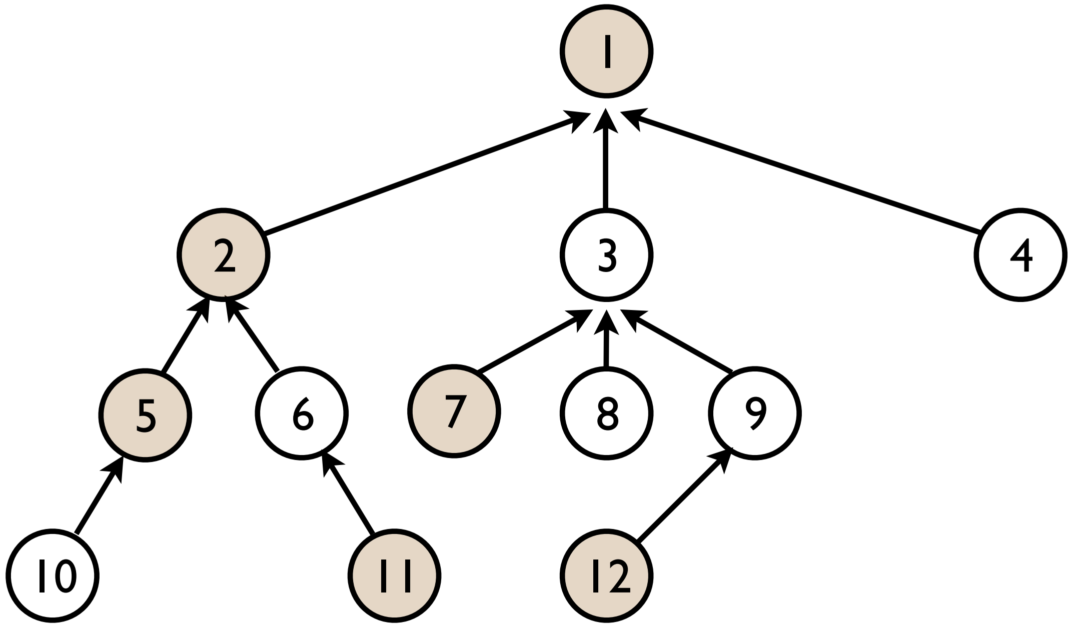


If the correct root is at depth zero, we identified the root of the tree. Here, regardless of the tree size, the average depth at which our solution finds the correct root is lower than 0.03.

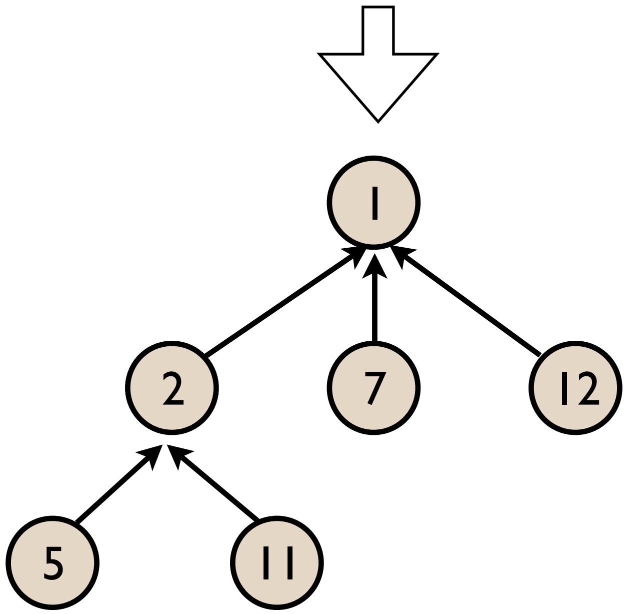
Missing Links

- On the wild, it is unrealistic to expect to have all the nodes of the tree.
- How to handle missing links?
  How do we evaluate the algorithm?

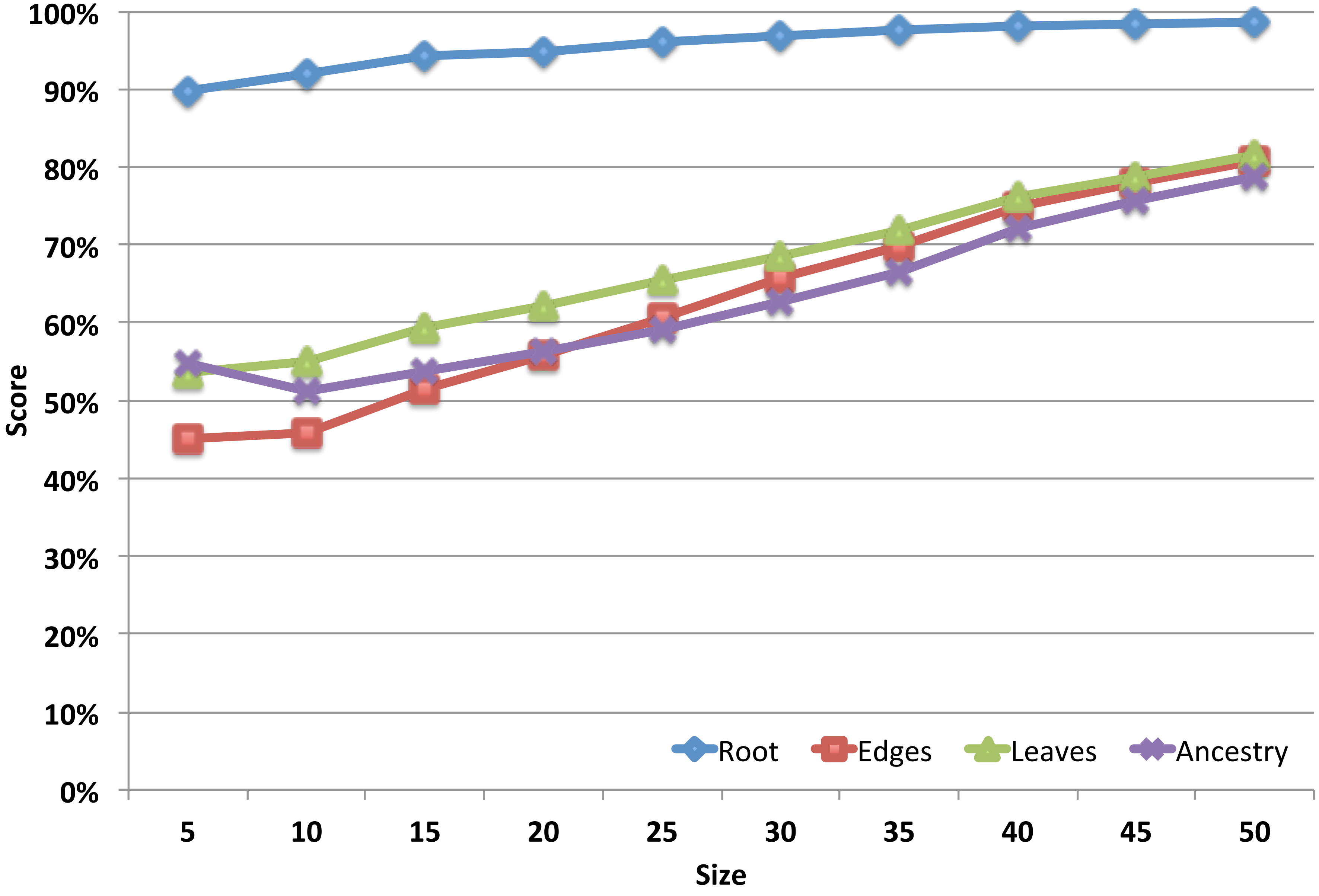
Missing Nodes



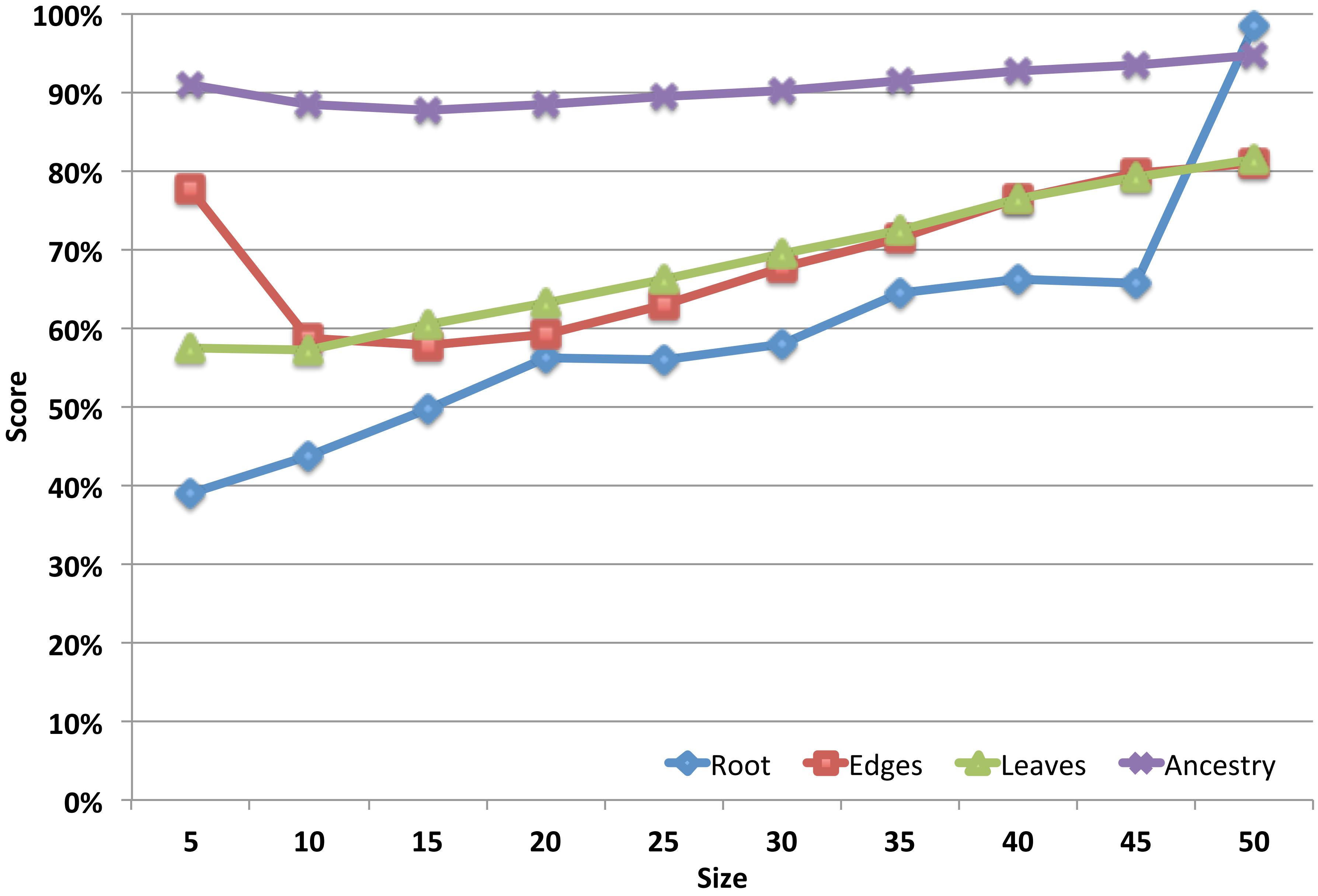
Using Ancestry Information



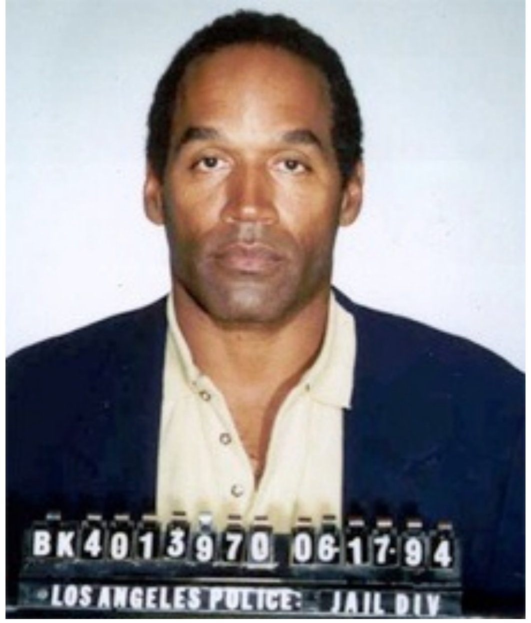
Missing only Internal Nodes

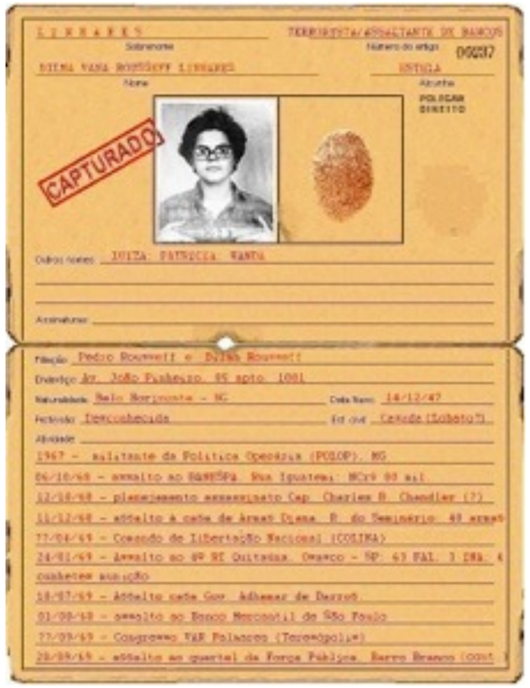


Missing Root and Internal Nodes

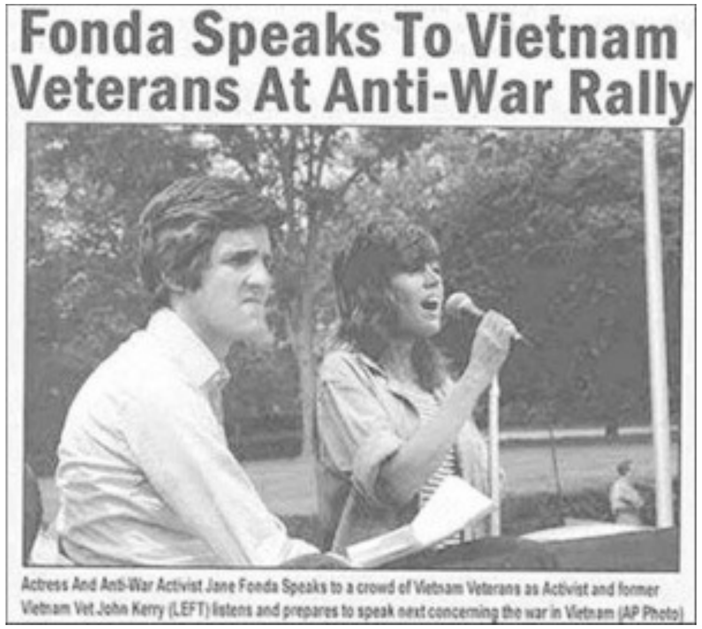


Real ND sets from Web




LINHARES — TERRORISTA/ASSALTANTE DE BANCOS
Sobrenome — Número do artigo 090217
DILMA VANA ROUSSEFF LINHARES — ESTELA
Nome — Alcunha
POLEGAR DIREITO
CAPTURADO
Outros nomes: LUIZA PATRICIA WANDA
Assinatura:
Filiação: Pedro Rousseff e Dilma Rousseff
Endereço: Av. João Pinheiro 85 apto 1001
Naturalidade: Belo Horizonte - MG — Data Nasc.: 14/12/47
Profissão: Desconhecida — Est civil: Casada (Loboto?)
Atividade:
1967 - militante da Politica Operária (POLOP), MG
06/10/68 - assalto ao BANESPA, Rua Iguatemi: NCr$ 80 mil
12/10/68 - planejamento assassinato Cap. Charles R. Chandler (?)
11/12/68 - assalto à casa de Armas Diana, R. do Seminário: 48 armas
??/04/69 - Comando de Libertação Nacional (COLINA)
24/01/69 - Assalto ao 4º RI Quitauna, Osasco - SP: 63 FAL, 3 INA, 4 [illegible]
conheceu munição
18/07/69 - Assalto casa Gov. Adhemar de Barros
01/08/69 - assalto ao Banco Mercantil de São Paulo
??/09/69 - Congresso VAR Palmares (Teresópolis)
20/09/69 - assalto ao quartel da Força Pública, Barro Branco (COST

RUBBER SOUL


Fonda Speaks To Vietnam Veterans At Anti-War Rally

Actress And Anti-War Activist Jane Fonda Speaks to a crowd of Vietnam Veterans as Activist and former Vietnam Vet John Kerry (LEFT) listens and prepares to speak next concerning the war in Vietnam (AP Photo)

How to Evaluate results?

- Since we do not know the ground truths, we evaluate the stability of reconstruction.

Er1: one if the new node IB is not a child of its generating node IA.

Er2: one if the structure of the tree relating the nodes in the original set changes with the insertion of the new node IB in the set.

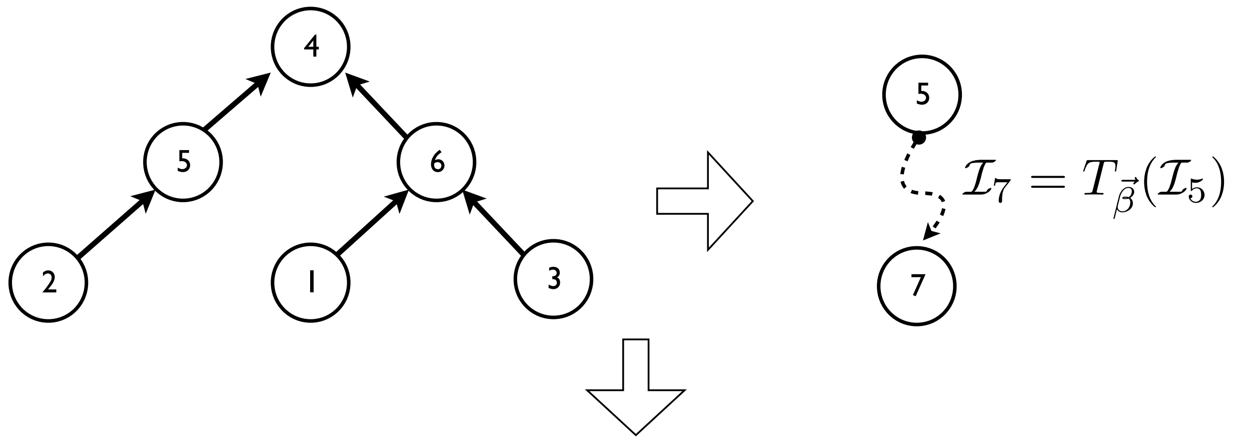
Er3: one if the new node IB appears as a father of another node on the original tree.

Er4: is one if the root of the reconstructed tree of the original set is different from the root of the reconstructed tree of the set augmented with IB.

P: one if the reconstructed tree is perfect compared to the original tree (Er1 = Er2 = 0).

Initial Tree [6, 5, 6, 4, 4, 4]

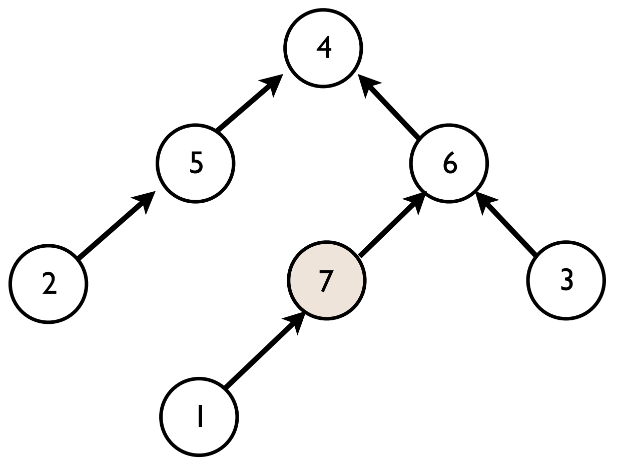
*Select one Node and **Artificially** Generate a Direct Descendant*



$\mathcal{I}_7 = T_{\vec{\beta}}(\mathcal{I}_5)$

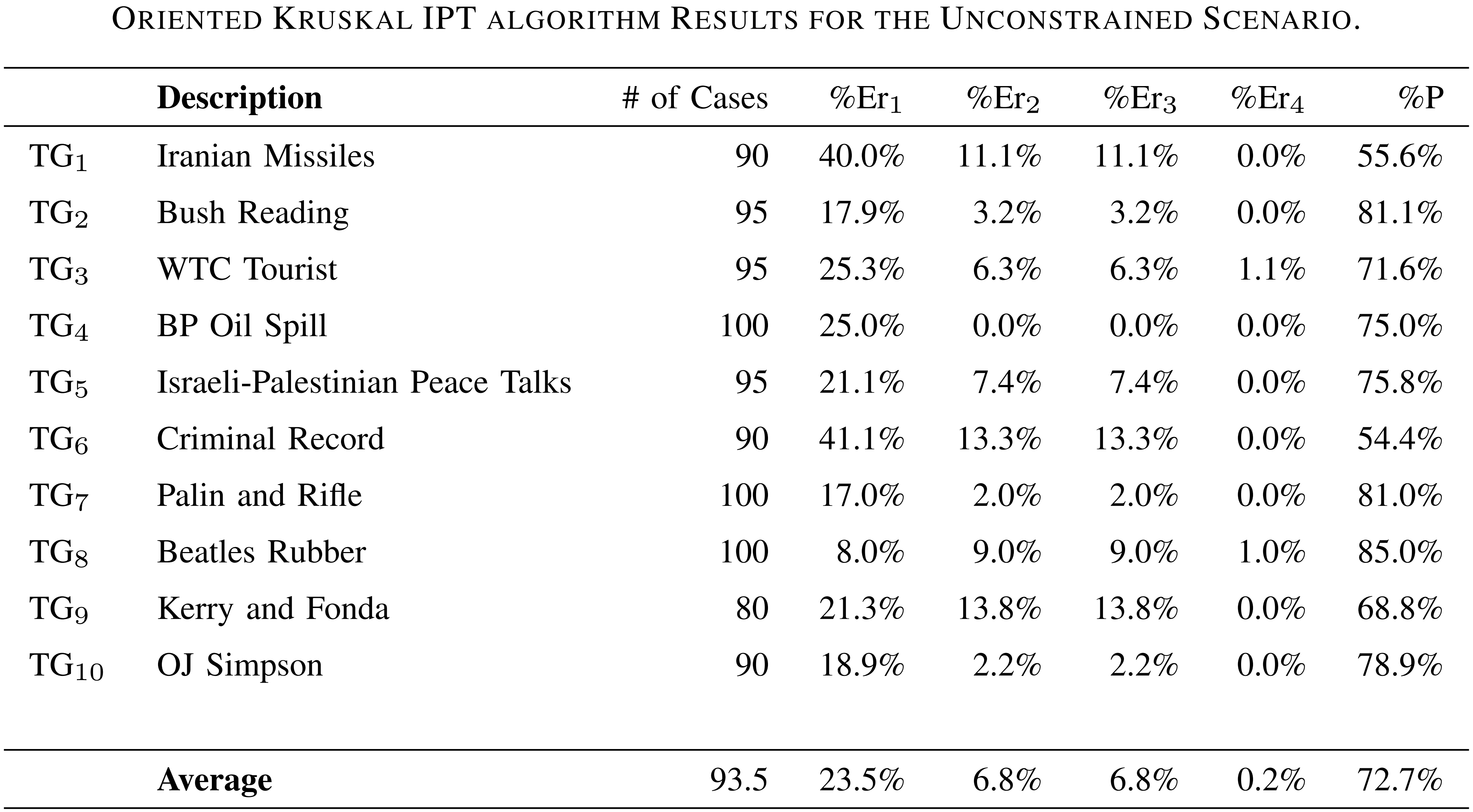
Tree After Inserting Node 7

[7, 5, 6, 4, 4, 4, 6]



Reconstruction Errors	
Er$_1$	1
Er$_2$	1
Er$_3$	1
Er$_4$	0
Success	
P	0

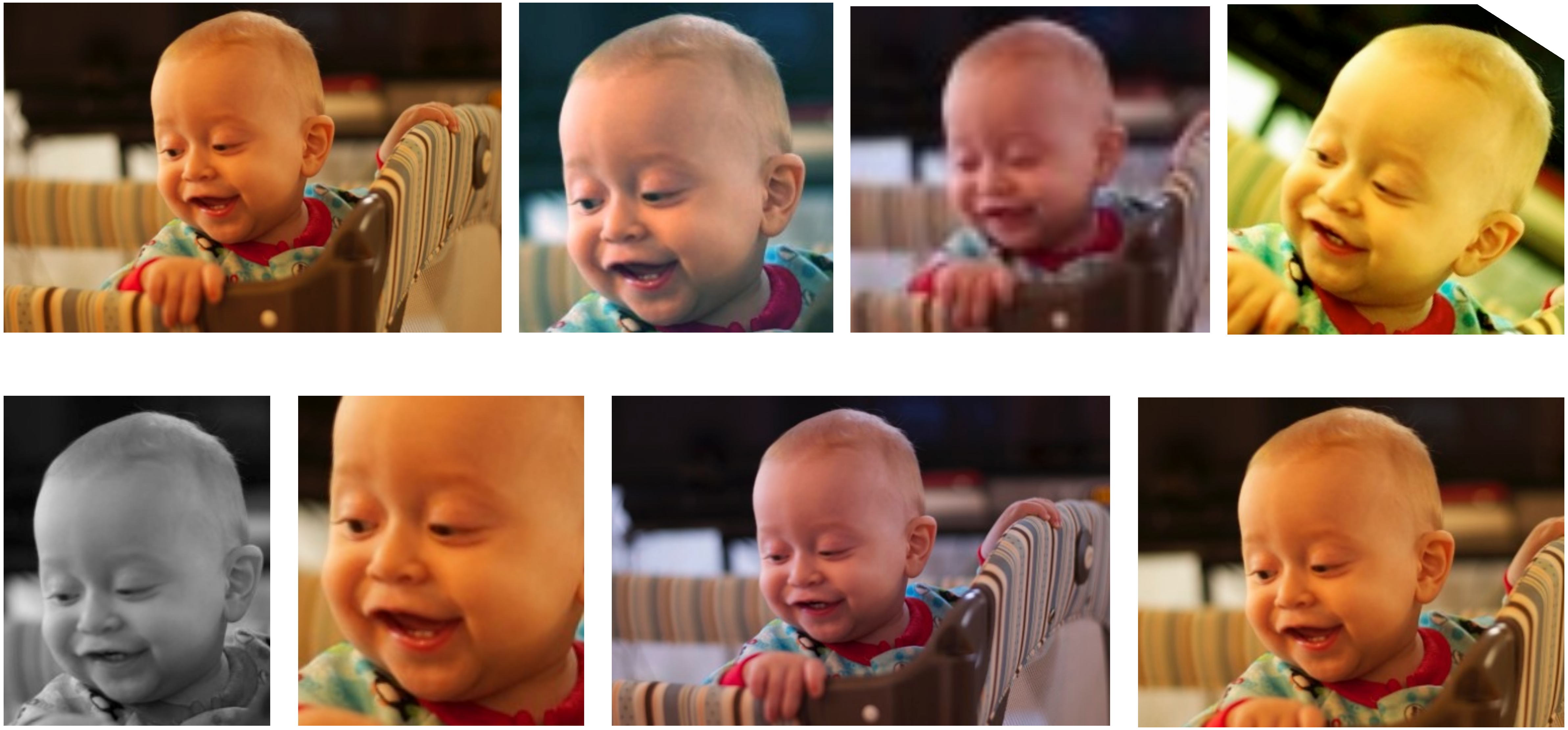
Real ND sets



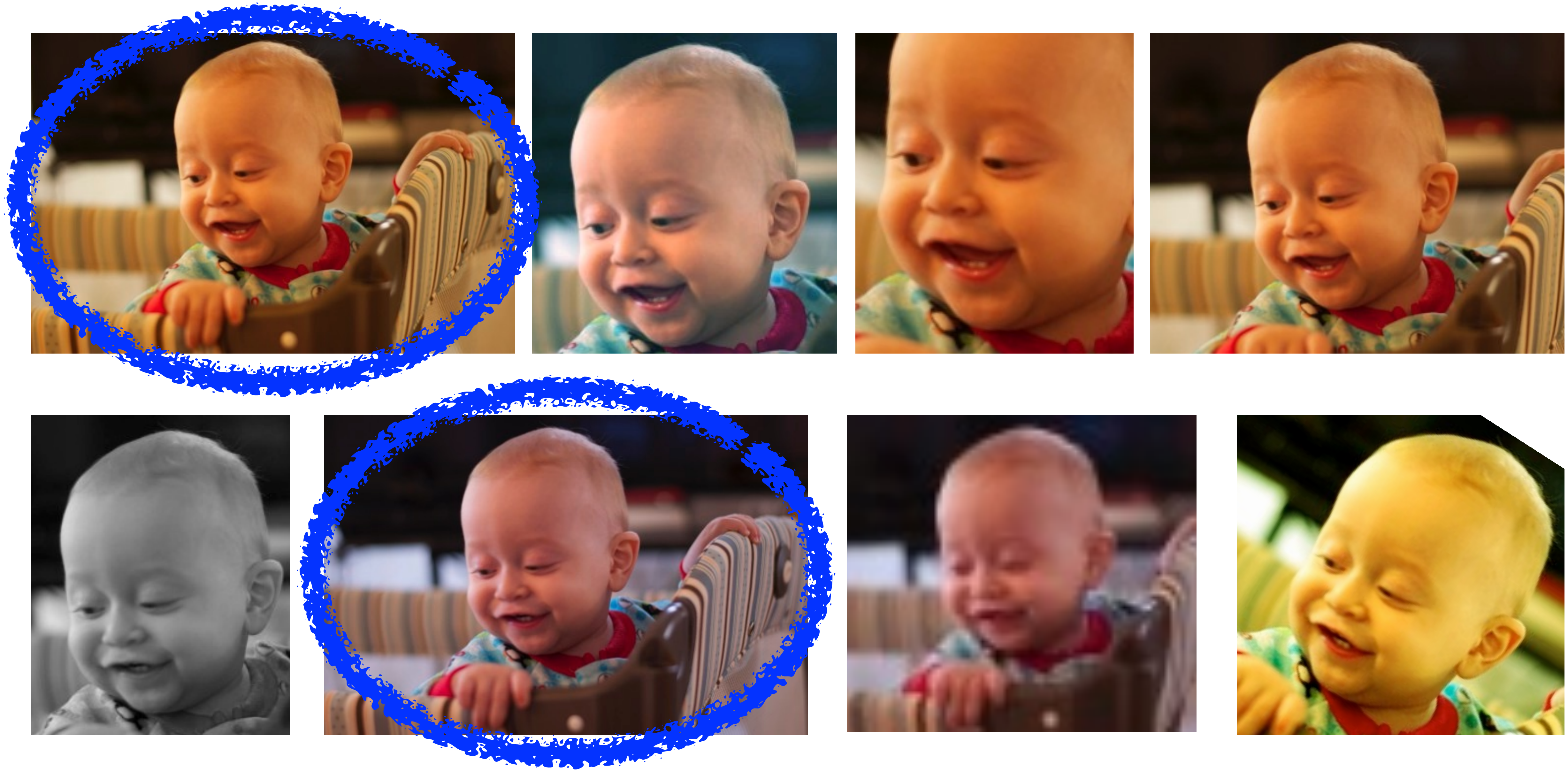
ORIENTED KRUSKAL IPT ALGORITHM RESULTS FOR THE UNCONSTRAINED SCENARIO.

	Description	# of Cases	%Er$_1$	%Er$_2$	%Er$_3$	%Er$_4$	%P
TG$_1$	Iranian Missiles	90	40.0%	11.1%	11.1%	0.0%	55.6%
TG$_2$	Bush Reading	95	17.9%	3.2%	3.2%	0.0%	81.1%
TG$_3$	WTC Tourist	95	25.3%	6.3%	6.3%	1.1%	71.6%
TG$_4$	BP Oil Spill	100	25.0%	0.0%	0.0%	0.0%	75.0%
TG$_5$	Israeli-Palestinian Peace Talks	95	21.1%	7.4%	7.4%	0.0%	75.8%
TG$_6$	Criminal Record	90	41.1%	13.3%	13.3%	0.0%	54.4%
TG$_7$	Palin and Rifle	100	17.0%	2.0%	2.0%	0.0%	81.0%
TG$_8$	Beatles Rubber	100	8.0%	9.0%	9.0%	1.0%	85.0%
TG$_9$	Kerry and Fonda	80	21.3%	13.8%	13.8%	0.0%	68.8%
TG$_{10}$	OJ Simpson	90	18.9%	2.2%	2.2%	0.0%	78.9%
	Average	93.5	23.5%	6.8%	6.8%	0.2%	72.7%

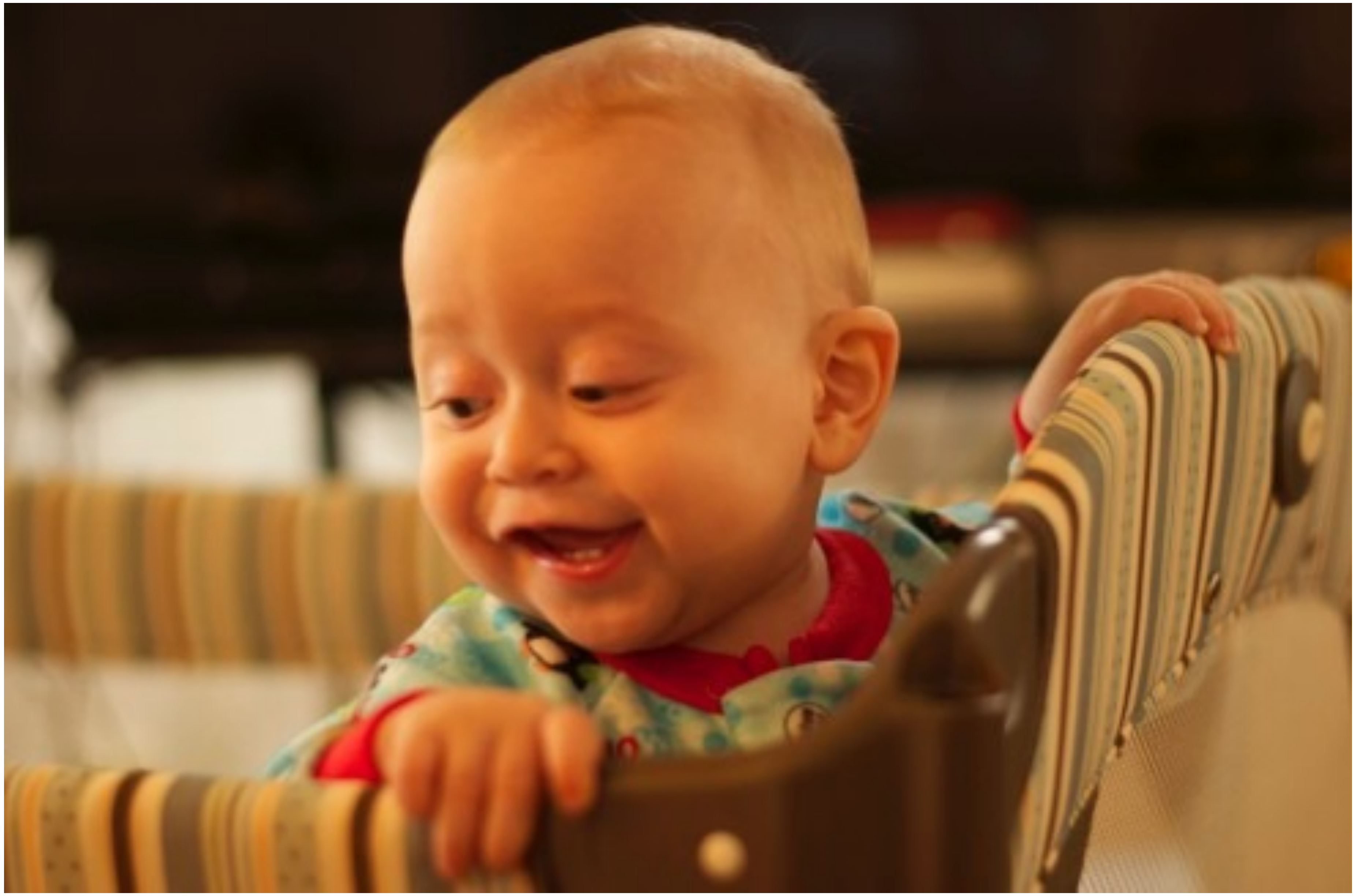
What’s up with this near duplicate set?



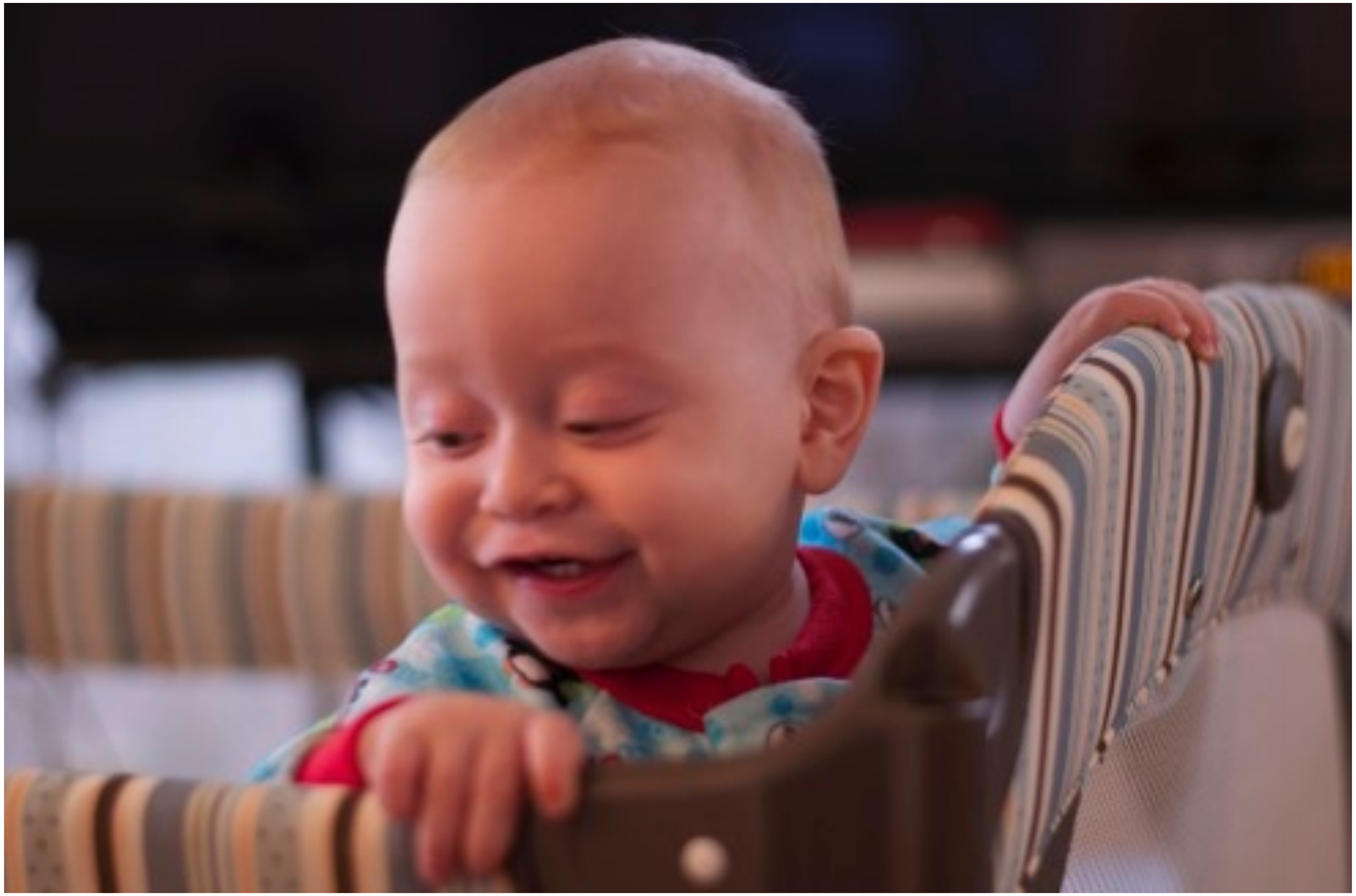
What's up with this near duplicate set?



Close-up

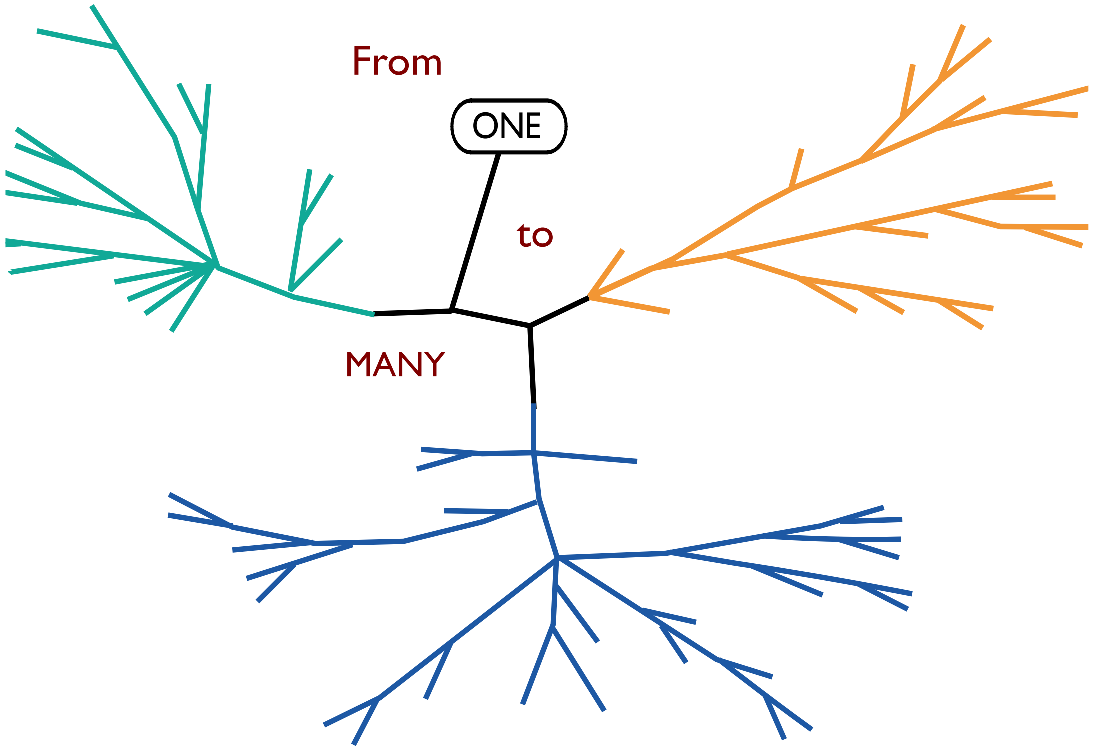


Close-up



First peek at Forests

- Forests (multiple co-existing trees) are a real case in real applications.
- Can our method be modified to find multiple trees?



Video Phylogeny Tree: VPT

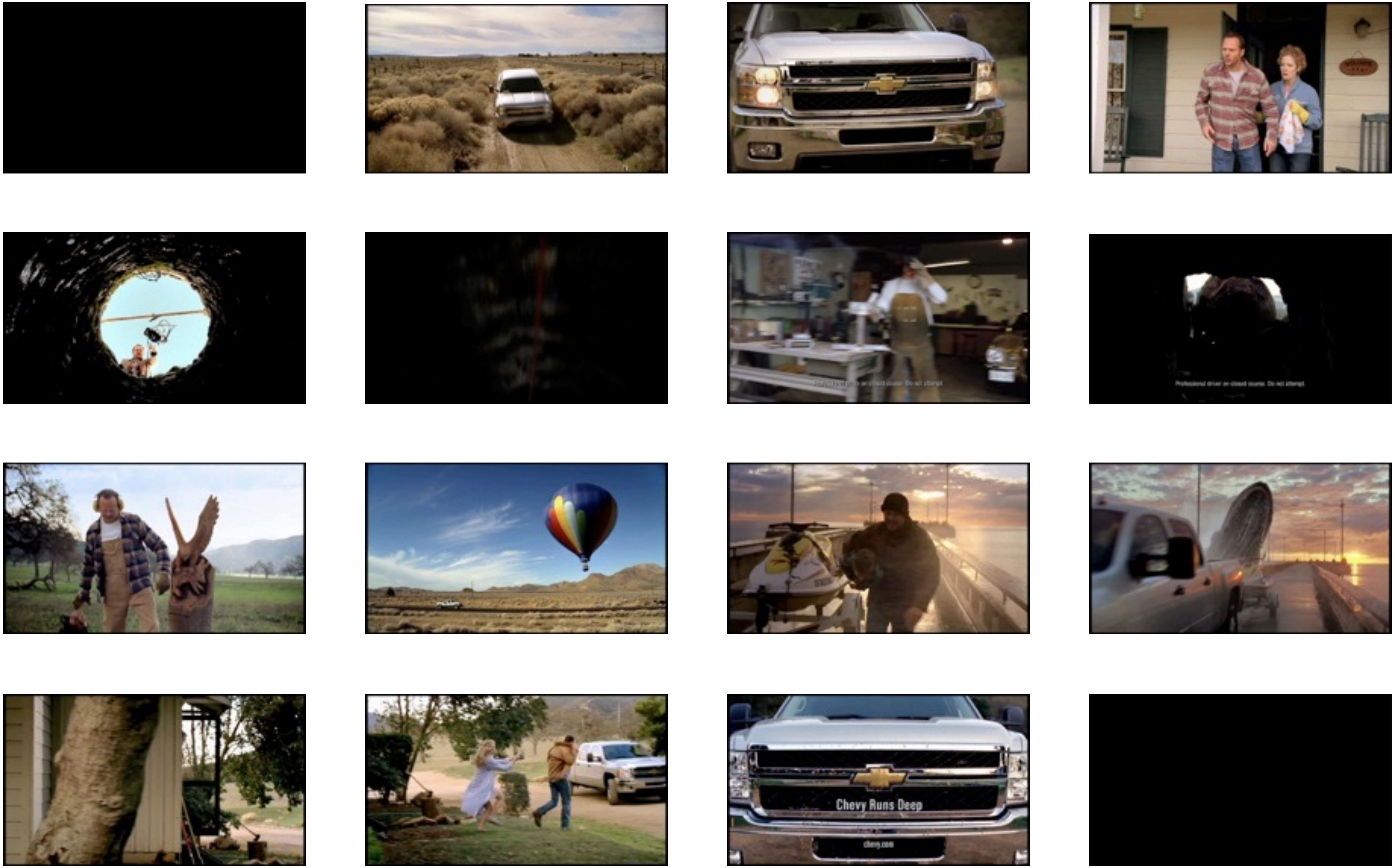
Video Phylogeny Tree

- We ignore the sound track.
- We use only static image content.

Why not get one frame, and use the IPT as the VPT?

The IPTs of frames are different along the video!!!

But why?


Professional driver on closed course. Do not attempt.
Professional driver on closed course. Do not attempt.
Chevy Runs Deep
chevy.com

Different IPTs

- Different quality over time, for example:
  - black frames,
  - blur,
  - compression artifacts,
  - dynamic range.
- Which (if any) is the right one?

A Few Approaches

- Expected result from a single frame IPT (baseline).
- Minimum dissimilarity matrix followed by IPT.
- Average dissimilarity matrix followed by IPT.
- Reconciliation Tree.

Single-Frame Expectation

- Calculate IPT on each frame.
- Calculate Expectation of metrics, but does not reconstruct a VPT (Video Phylogeny Tree).

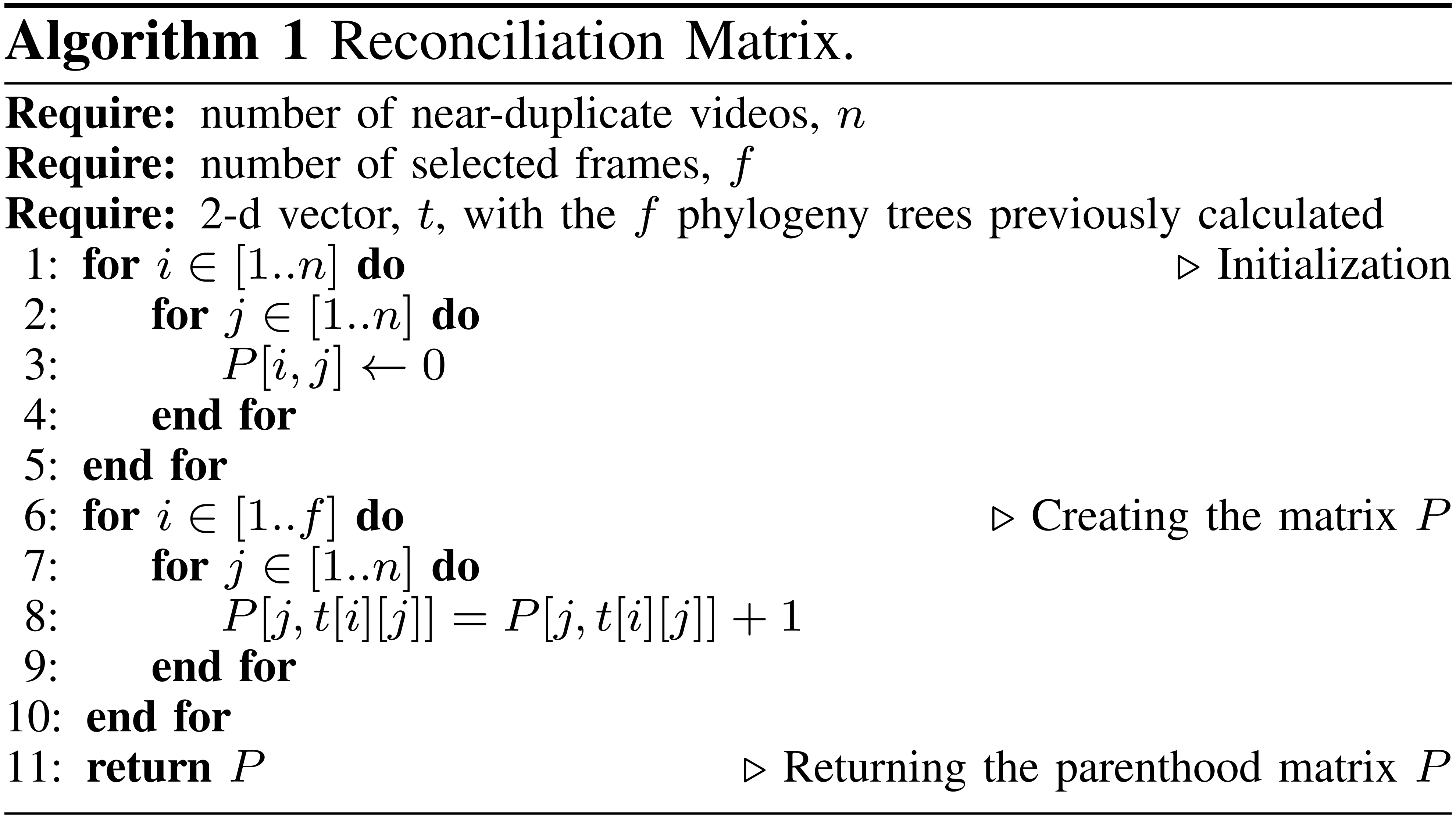
Min / Average

- Sample frames.
- Calculate Dissimilarity Matrix on each synchronized frames.
- Create a new Dissimilarity Matrix using the frame's Dissimilarities
  - min,
  - average,
  - normalized min,
  - normalized average.
- Construct VPT using oriented Kruskal on this new Matrix.

Reconciliation Approach

- Sample frames.
- Calculate IPT on each frame.
- Reconcile the frame's IPTs into the VPT.
    - Build Reconciliation Matrix.
    - Apply Tree Reconciliation Algorithm

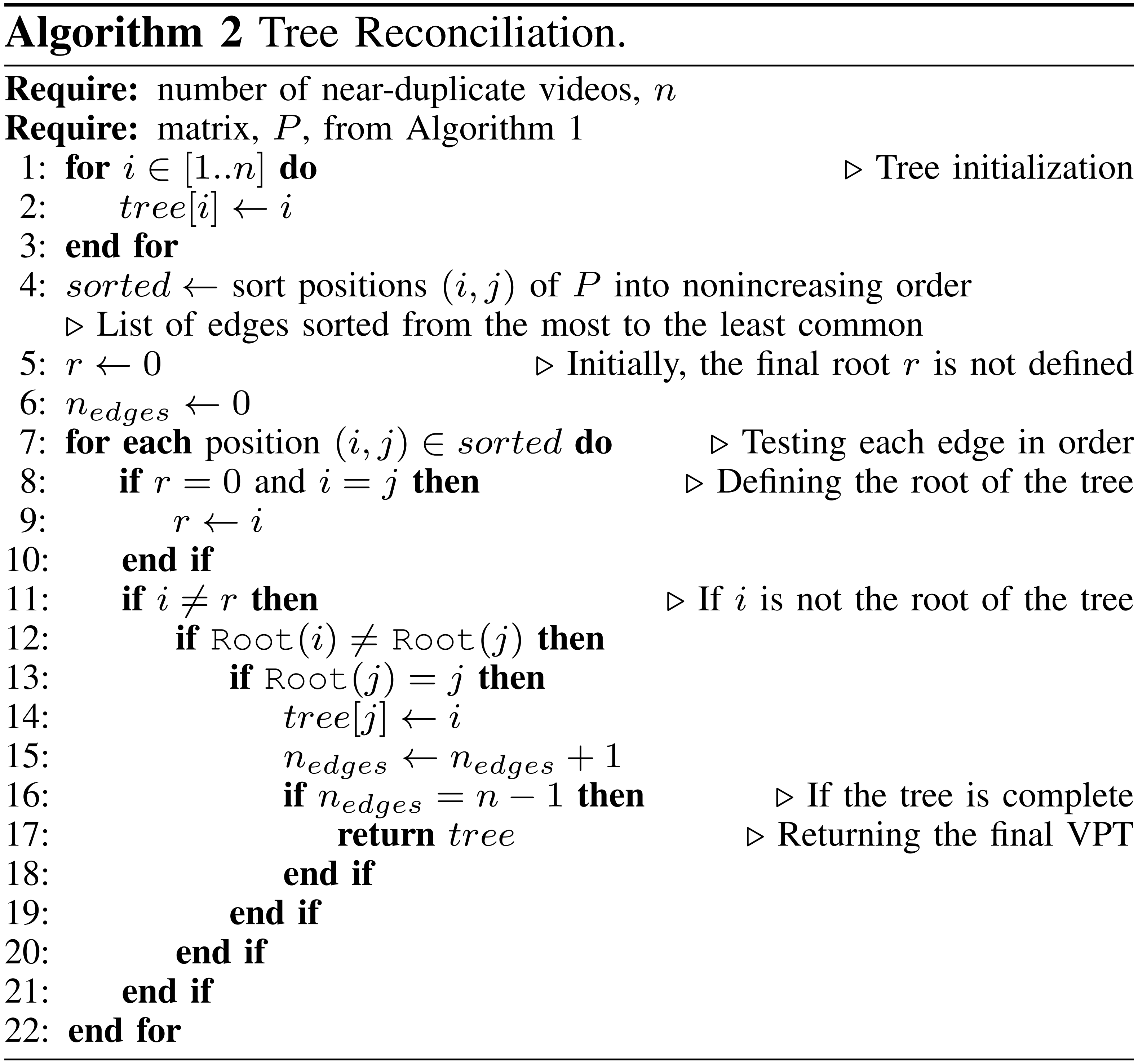
Reconciliation Matrix



Algorithm 1 Reconciliation Matrix.

```
Require: number of near-duplicate videos, $n$
Require: number of selected frames, $f$
Require: 2-d vector, $t$, with the $f$ phylogeny trees previously calculated
for $i \in [1..n]$ do                                  ▷ Initialization
    for $j \in [1..n]$ do
        $P[i, j] \leftarrow 0$
    end for
end for
for $i \in [1..f]$ do                                  ▷ Creating the matrix $P$
    for $j \in [1..n]$ do
        $P[j, t[i][j]] = P[j, t[i][j]] + 1$
    end for
end for
return $P$                                             ▷ Returning the parenthood matrix $P$
```

Tree Reconciliation Alg.



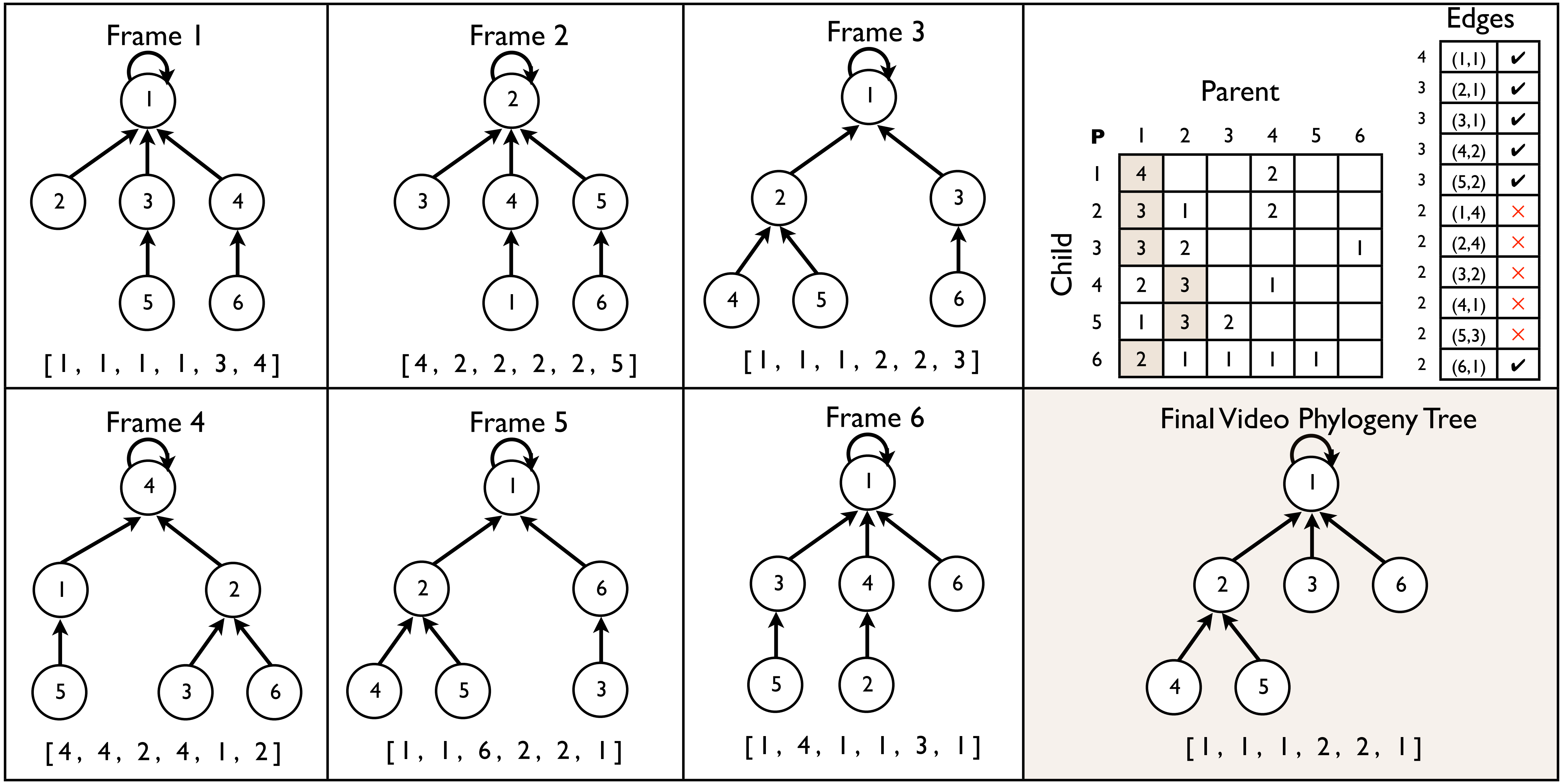
Algorithm 2 Tree Reconciliation.

```
Require: number of near-duplicate videos, $n$
Require: matrix, $P$, from Algorithm 1
for $i \in [1..n]$ do                                              ▷ Tree initialization
    $tree[i] \leftarrow i$
end for
$sorted \leftarrow$ sort positions $(i, j)$ of $P$ into nonincreasing order
    ▷ List of edges sorted from the most to the least common
$r \leftarrow 0$                                  ▷ Initially, the final root $r$ is not defined
$n_{edges} \leftarrow 0$
for each position $(i, j) \in sorted$ do                      ▷ Testing each edge in order
    if $r = 0$ and $i = j$ then                             ▷ Defining the root of the tree
        $r \leftarrow i$
    end if
    if $i \neq r$ then                                     ▷ If $i$ is not the root of the tree
        if Root$(i) \neq$ Root$(j)$ then
            if Root$(j) = j$ then
                $tree[j] \leftarrow i$
                $n_{edges} \leftarrow n_{edges} + 1$
                if $n_{edges} = n - 1$ then                    ▷ If the tree is complete
                    return $tree$                           ▷ Returning the final VPT
                end if
            end if
        end if
    end if
end for
```

Reconciliation Approach

- Sample frames.
- Calculate IPT on each frame.
- Reconcile the frame's IPTs into the VPT.
- Is this enough to achieve good results?
- What are the limitations of this approach?

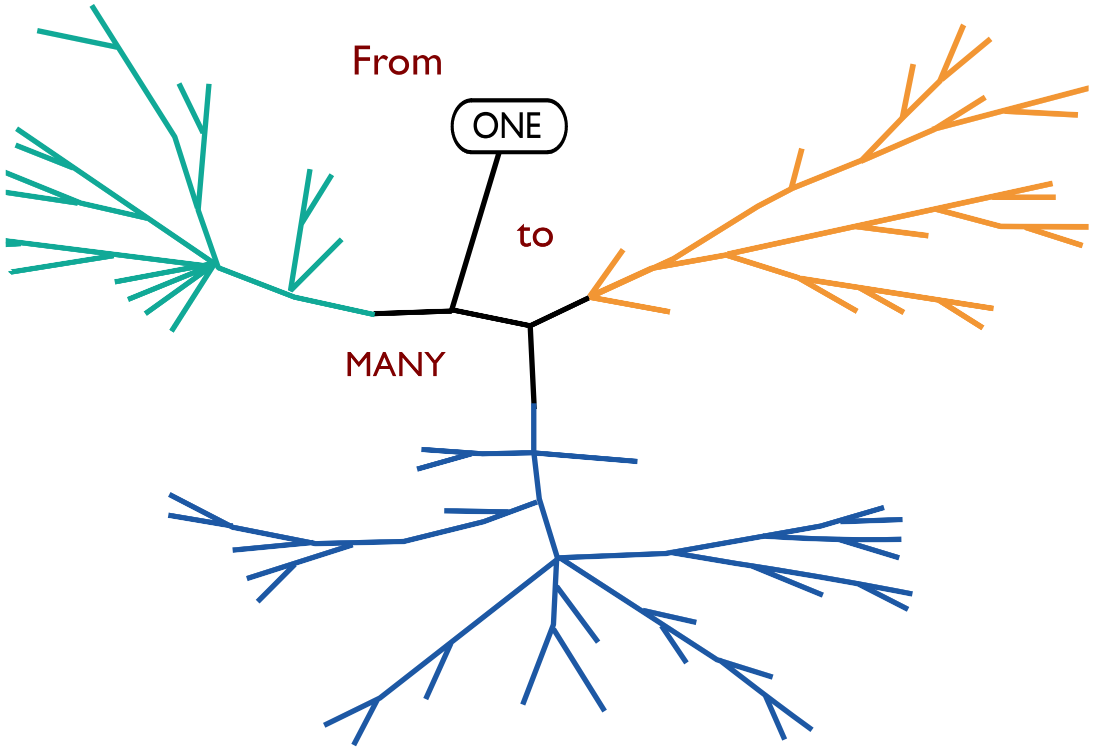
Example



P	1	2	3	4	5	6
1	4			2		
2	3	1		2		
3	3	2				1
4	2	3		1		
5	1	3	2			
6	2	1	1	1	1	

Edges

4	(1,1)	✓
3	(2,1)	✓
3	(3,1)	✓
3	(4,2)	✓
3	(5,2)	✓
2	(1,4)	✗
2	(2,4)	✗
2	(3,2)	✗
2	(4,1)	✗
2	(5,3)	✗
2	(6,1)	✓



Evaluation

Experimental results

- Limited experiments in this paper:
  - Ignored temporal cropping,
  - Ignored video compression on dissimilarities,
  - 16 Videos (Super Bowl Commercials 2011),
  - 16 trees,
  - 10 near-duplicates per tree.
- Transformations using `mencoder`.
- Sampling frames and sync by `ffmpeg`.
- Dissimilarities using `OpenCV`.

Transformations

We used `mencoder` to generate the Near-Duplicates, with these transformations and ranges:

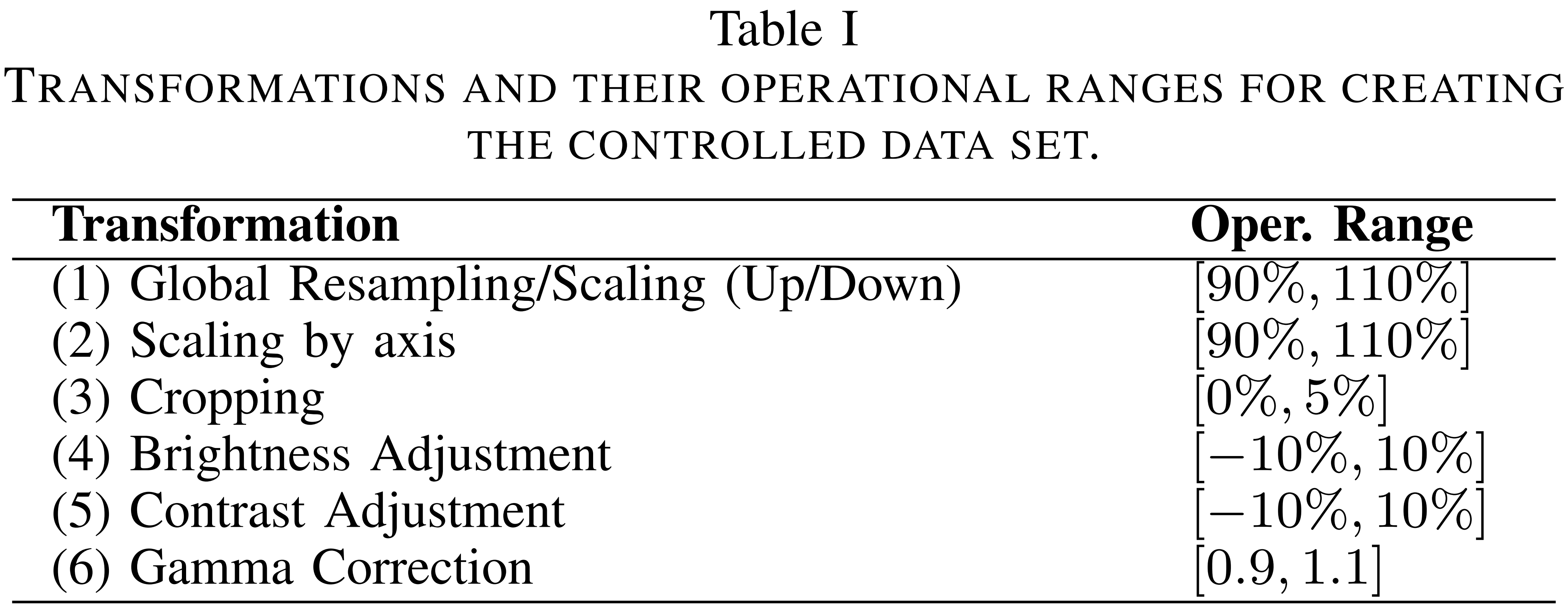


Table I
TRANSFORMATIONS AND THEIR OPERATIONAL RANGES FOR CREATING THE CONTROLLED DATA SET.

Transformation	**Oper. Range**
(1) Global Resampling/Scaling (Up/Down)	$[90\%, 110\%]$
(2) Scaling by axis	$[90\%, 110\%]$
(3) Cropping	$[0\%, 5\%]$
(4) Brightness Adjustment	$[-10\%, 10\%]$
(5) Contrast Adjustment	$[-10\%, 10\%]$
(6) Gamma Correction	$[0.9, 1.1]$

Comparing Trees

Root: $R(\mathrm{IPT}_1, \mathrm{IPT}_2) = \begin{cases} 1, & \text{If } \texttt{Root(IPT}_1\texttt{)} = \texttt{Root(IPT}_2\texttt{)} \\ 0, & \text{Otherwise} \end{cases}$

Edges: $E(\mathrm{IPT}_1, \mathrm{IPT}_2) = \frac{|E_1 \cap E_2|}{n-1}$

Leaves: $L(\mathrm{IPT}_1, \mathrm{IPT}_2) = \frac{|L_1 \cap L_2|}{|L_1 \cup L_2|}$

Ancestry: $A(\mathrm{IPT}_1, \mathrm{IPT}_2) = \frac{|A_1 \cap A_2|}{|A_1 \cup A_2|}$

Results of Approaches

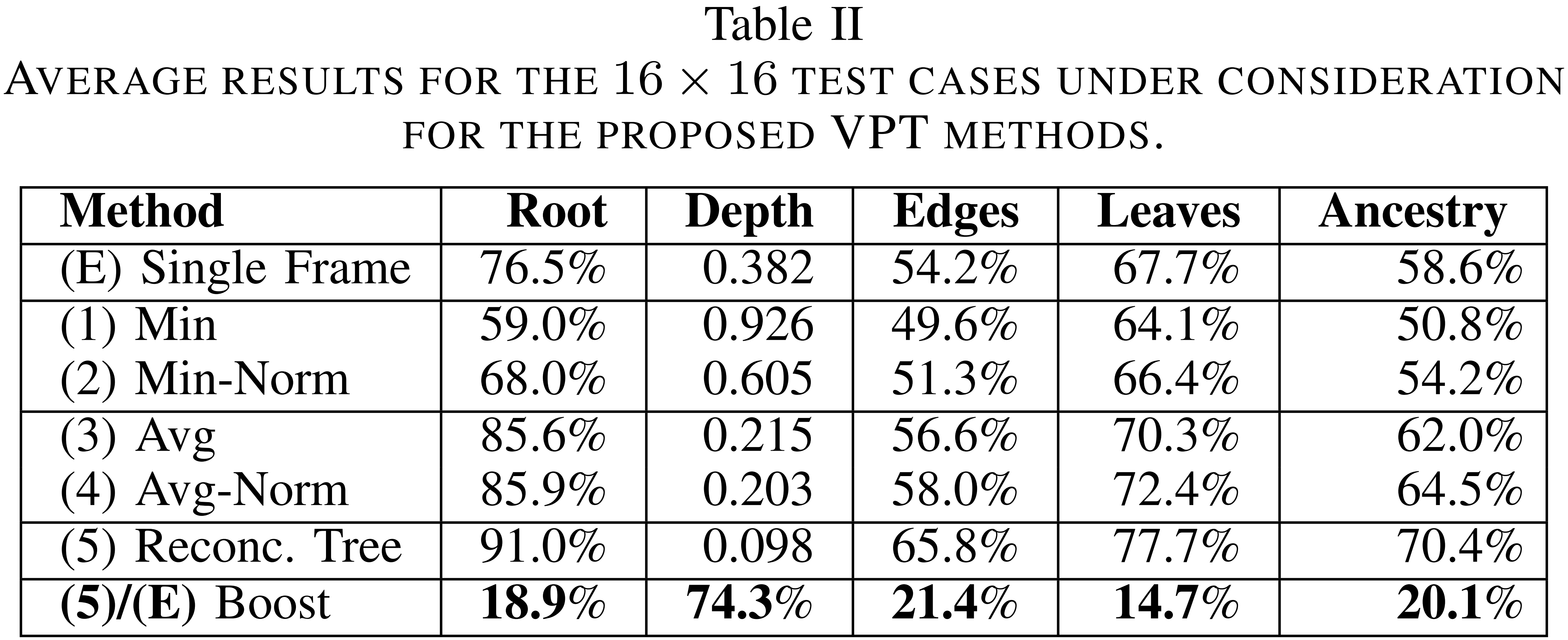


Table II
AVERAGE RESULTS FOR THE 16 × 16 TEST CASES UNDER CONSIDERATION FOR THE PROPOSED VPT METHODS.

Method	**Root**	**Depth**	**Edges**	**Leaves**	**Ancestry**
(E) Single Frame	76.5%	0.382	54.2%	67.7%	58.6%
(1) Min	59.0%	0.926	49.6%	64.1%	50.8%
(2) Min-Norm	68.0%	0.605	51.3%	66.4%	54.2%
(3) Avg	85.6%	0.215	56.6%	70.3%	62.0%
(4) Avg-Norm	85.9%	0.203	58.0%	72.4%	64.5%
(5) Reconc. Tree	91.0%	0.098	65.8%	77.7%	70.4%
(5)/(E) Boost	**18.9%**	**74.3%**	**21.4%**	**14.7%**	**20.1%**

Experimental Results

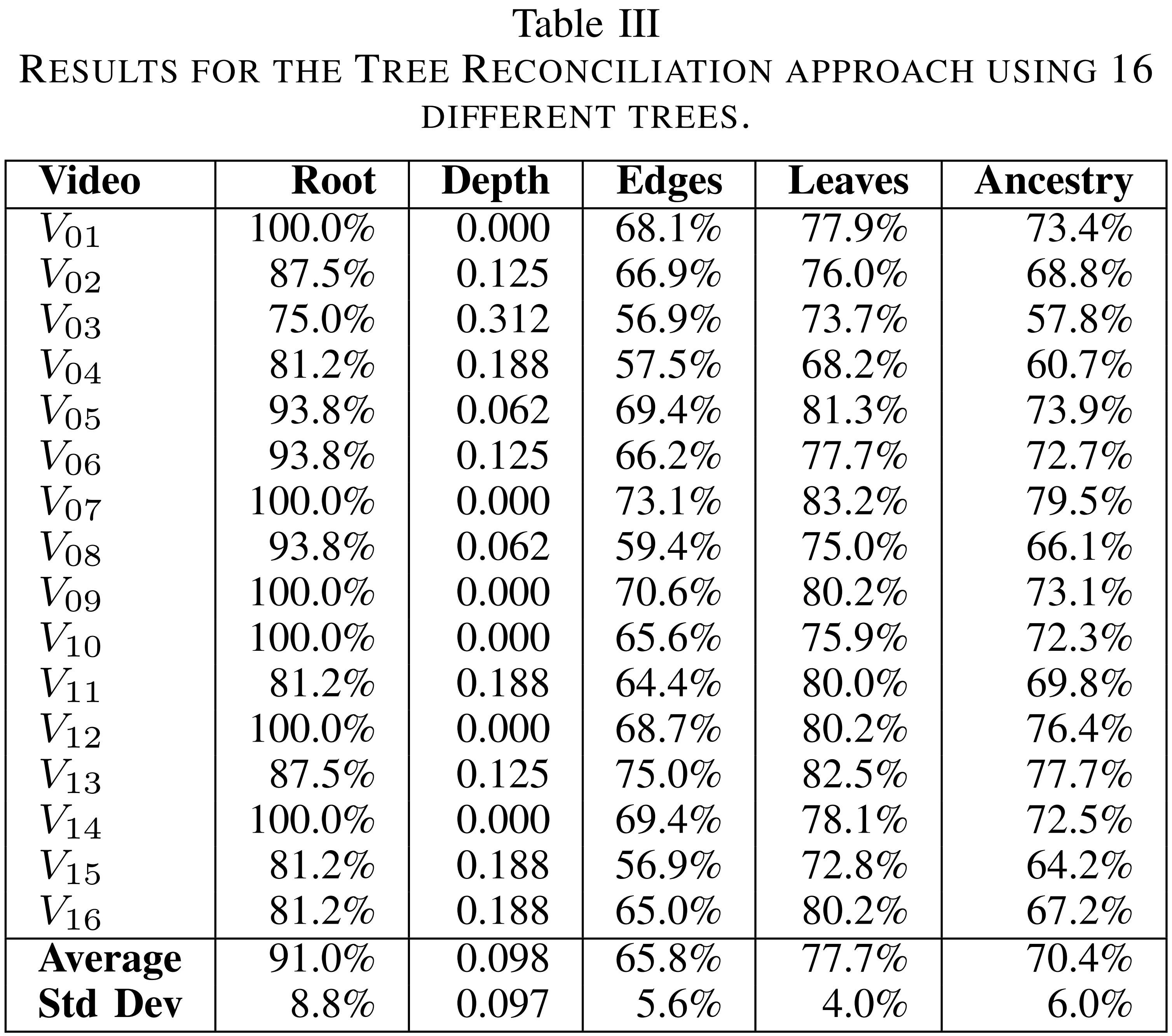
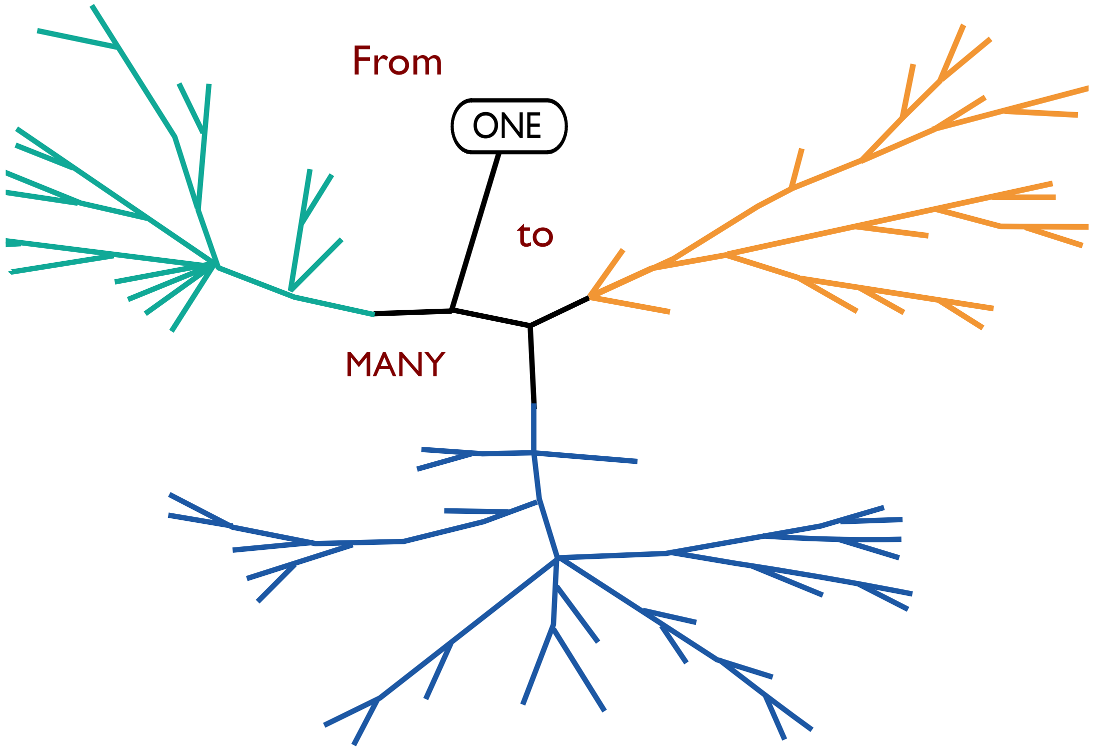


Table III
RESULTS FOR THE TREE RECONCILIATION APPROACH USING 16 DIFFERENT TREES.

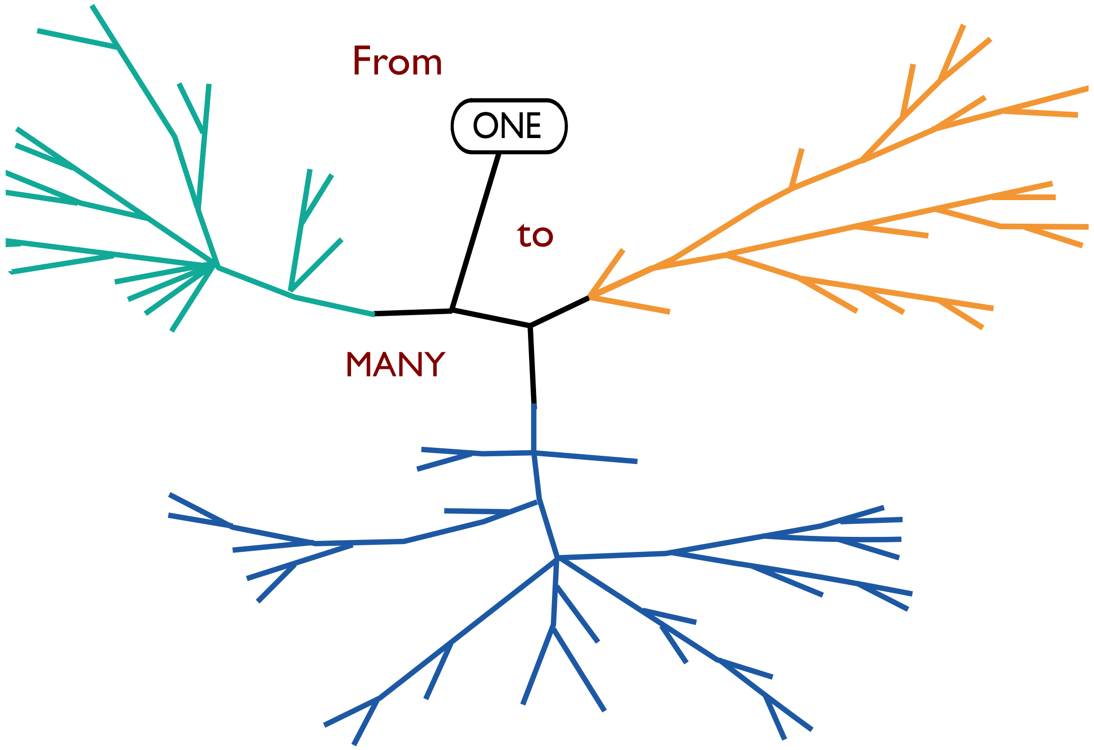
Video	Root	Depth	Edges	Leaves	Ancestry
V_{01}	100.0%	0.000	68.1%	77.9%	73.4%
V_{02}	87.5%	0.125	66.9%	76.0%	68.8%
V_{03}	75.0%	0.312	56.9%	73.7%	57.8%
V_{04}	81.2%	0.188	57.5%	68.2%	60.7%
V_{05}	93.8%	0.062	69.4%	81.3%	73.9%
V_{06}	93.8%	0.125	66.2%	77.7%	72.7%
V_{07}	100.0%	0.000	73.1%	83.2%	79.5%
V_{08}	93.8%	0.062	59.4%	75.0%	66.1%
V_{09}	100.0%	0.000	70.6%	80.2%	73.1%
V_{10}	100.0%	0.000	65.6%	75.9%	72.3%
V_{11}	81.2%	0.188	64.4%	80.0%	69.8%
V_{12}	100.0%	0.000	68.7%	80.2%	76.4%
V_{13}	87.5%	0.125	75.0%	82.5%	77.7%
V_{14}	100.0%	0.000	69.4%	78.1%	72.5%
V_{15}	81.2%	0.188	56.9%	72.8%	64.2%
V_{16}	81.2%	0.188	65.0%	80.2%	67.2%
Average	91.0%	0.098	65.8%	77.7%	70.4%
Std Dev	8.8%	0.097	5.6%	4.0%	6.0%

Limitations of frame-based VPT

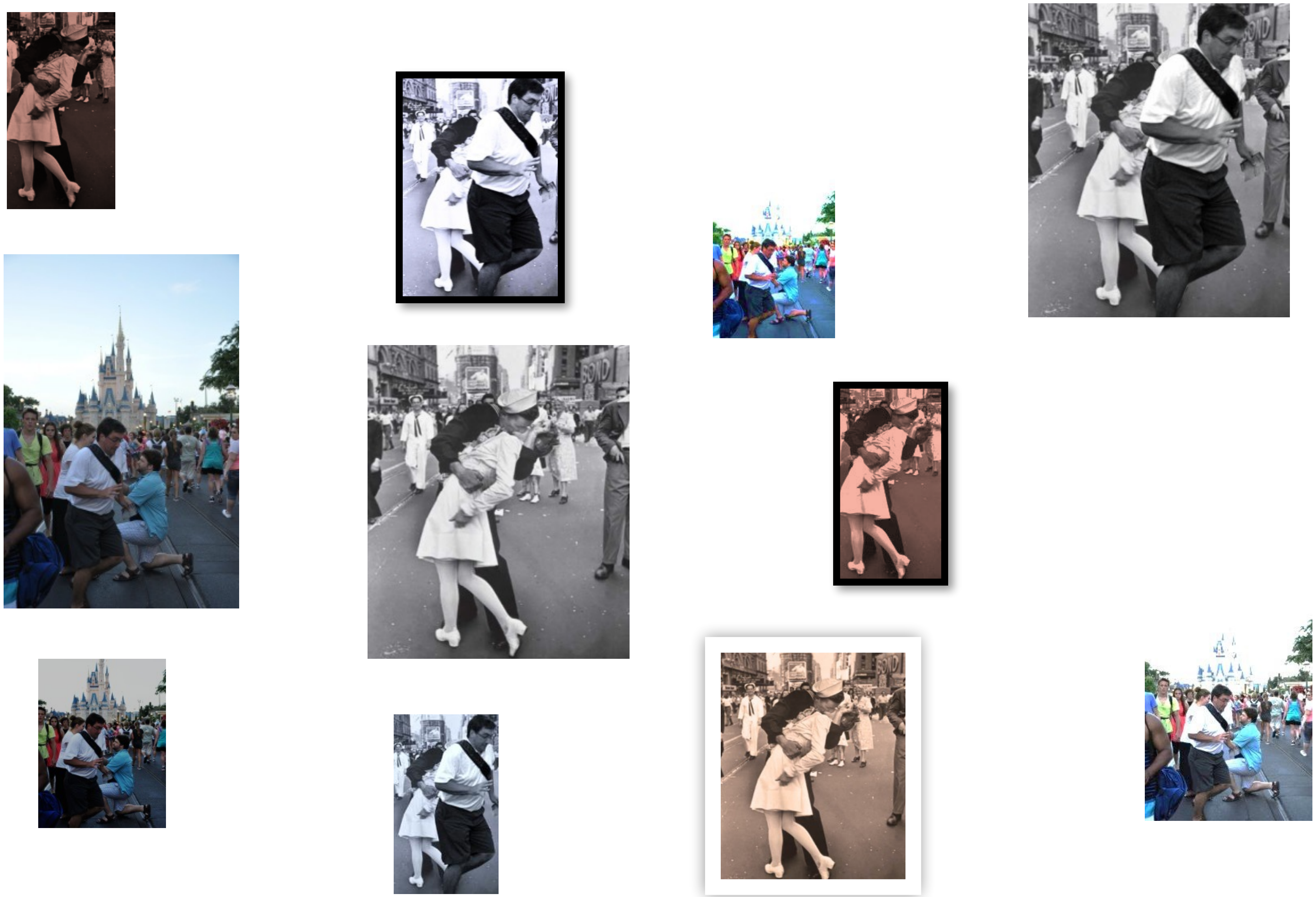
- Does not use sound.
- Ignores temporal information of the visual content.
- Requires sync frames!
  - This is actually a very complicated issue in Video, and some codecs are finickier than others.
  - If we allow change in FPS + temporal crop, it might be impossible to fulfill this requisite.



Alright... what’s next?



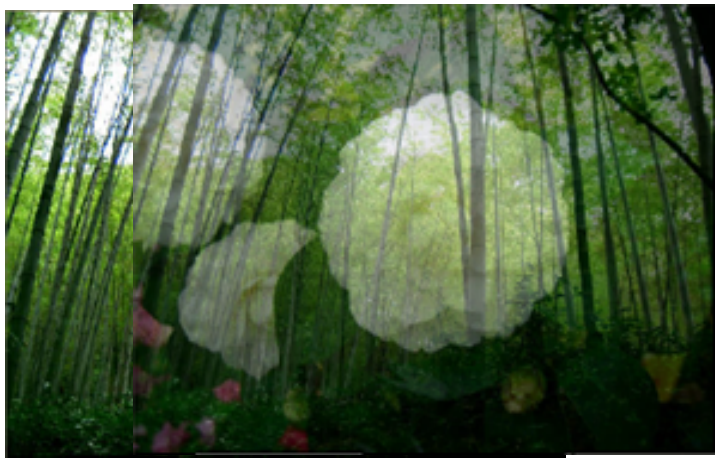
Multiple Parenting Phylogeny



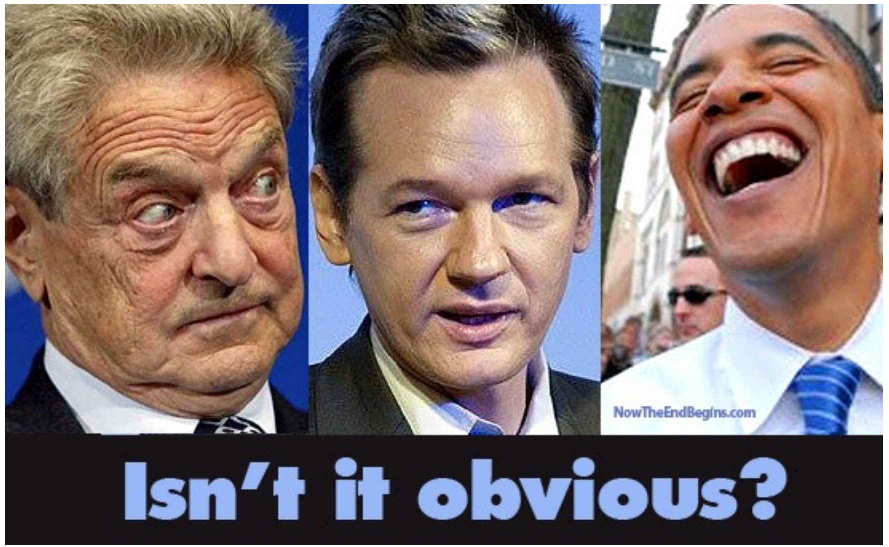
- Content from multiple images combined

Composition examples

Blending

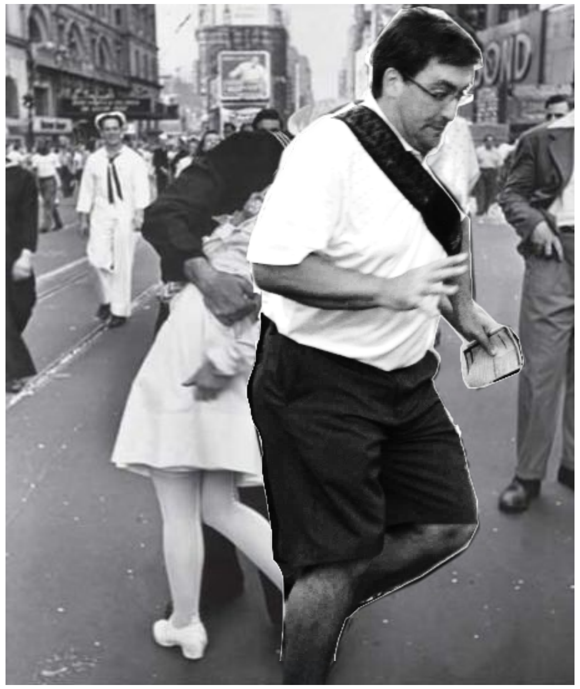


Montage



Composition examples

Splicing



Methodology

- In this work, we focus on splicing compositions;

- Three steps method

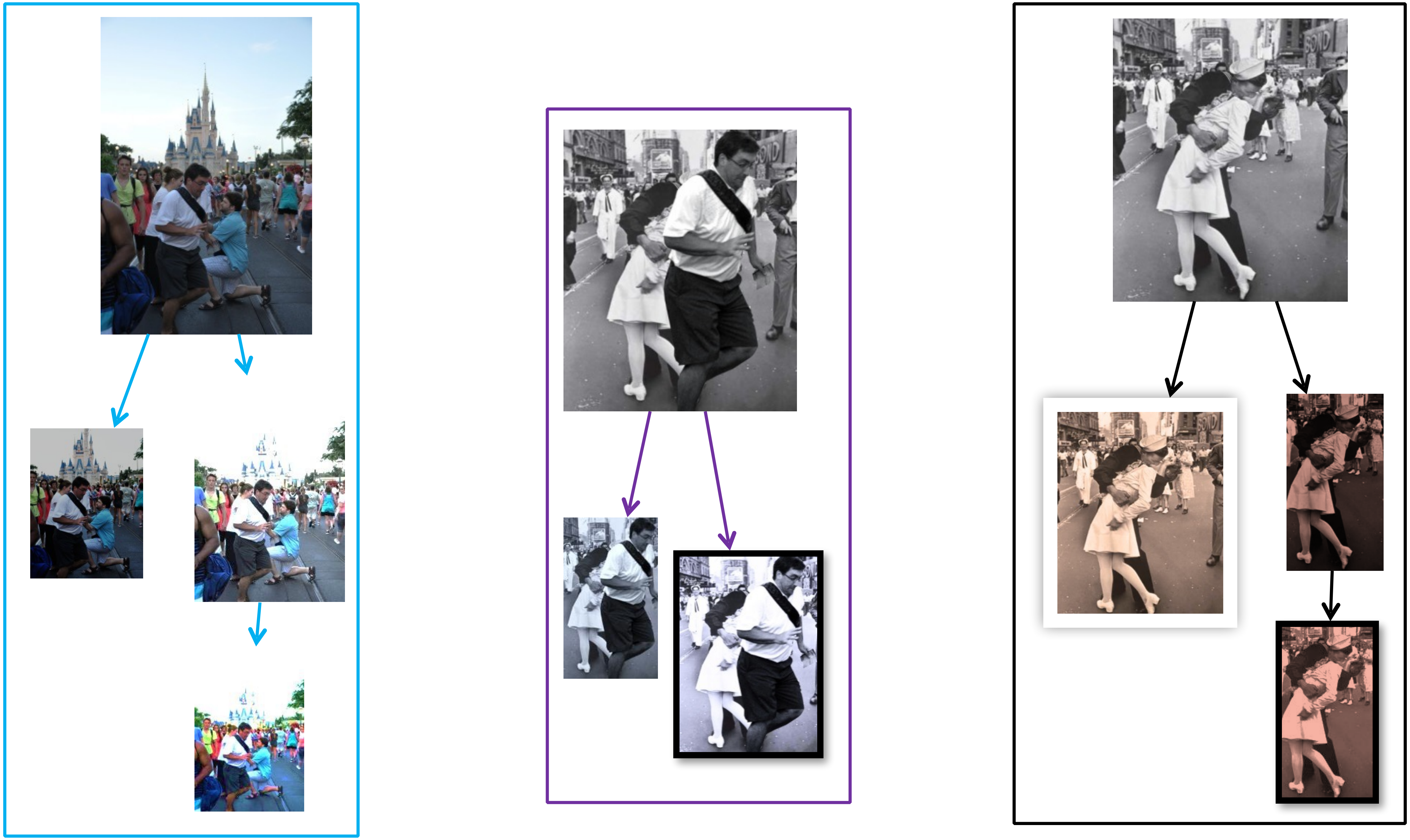
1. Group separation
   - Image Phylogeny Forests

2. Group classification
   - Finding shared content with keypoint matches

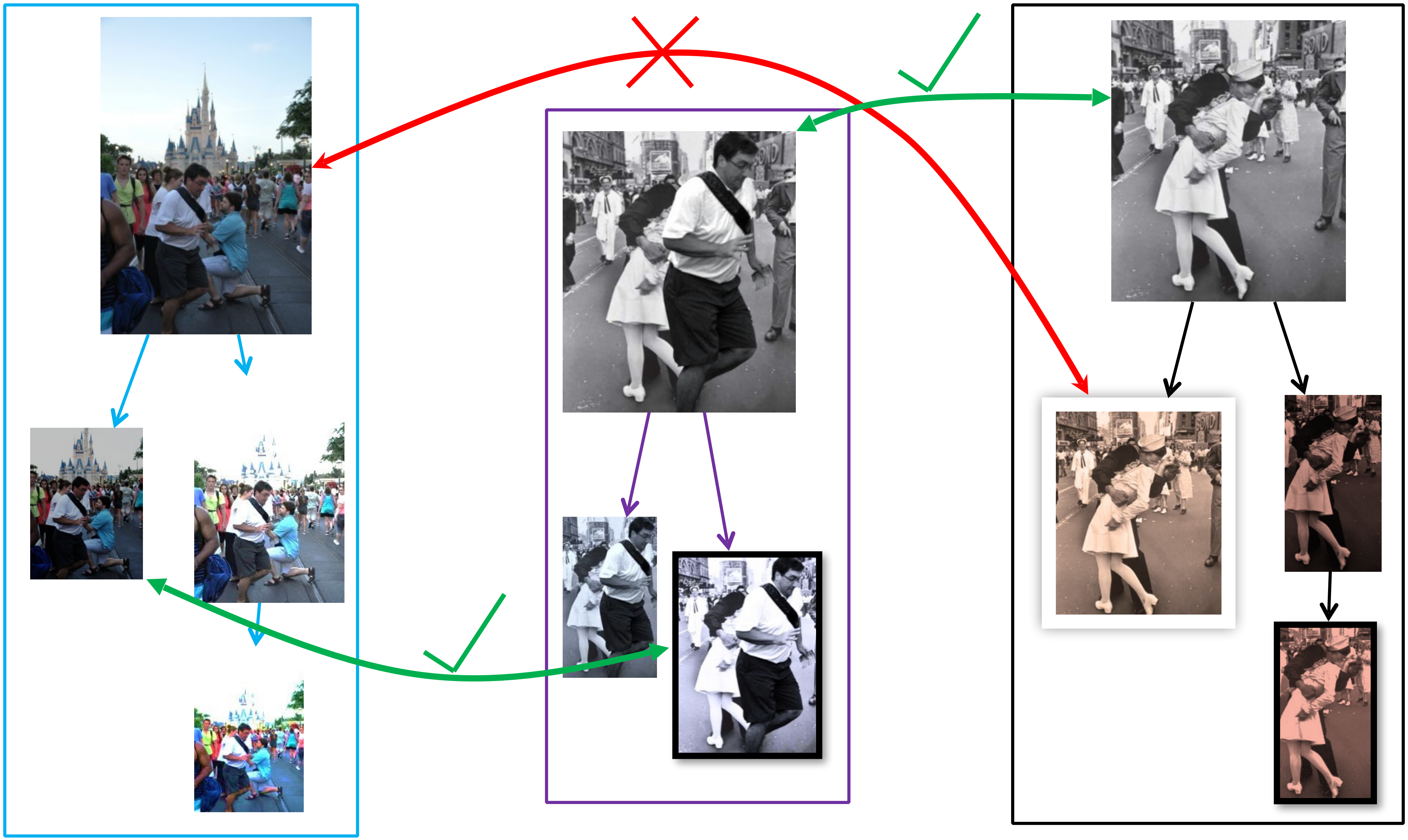
3. Finding the parents
   - Local dissimilarity

Group separation

Group separation

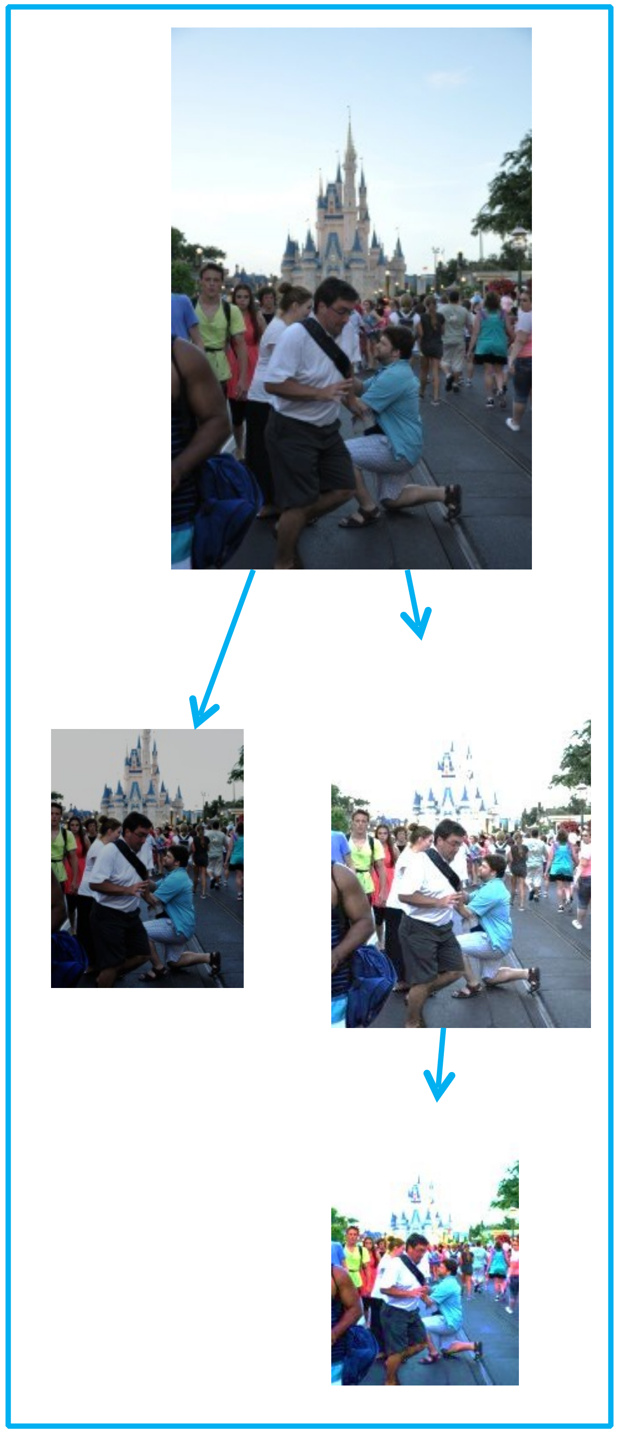


Group classification

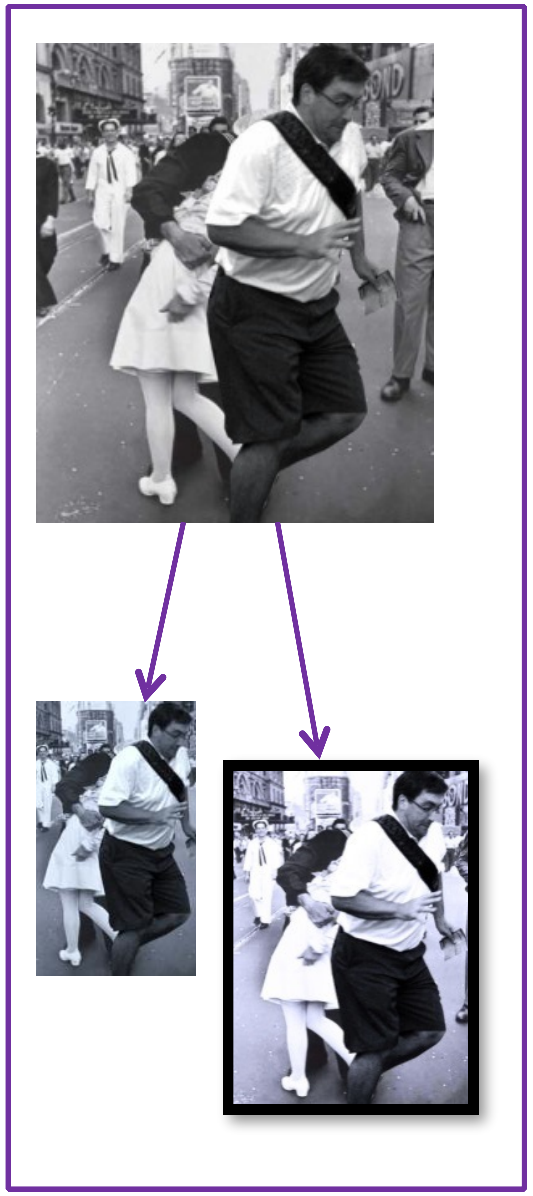


Group classification

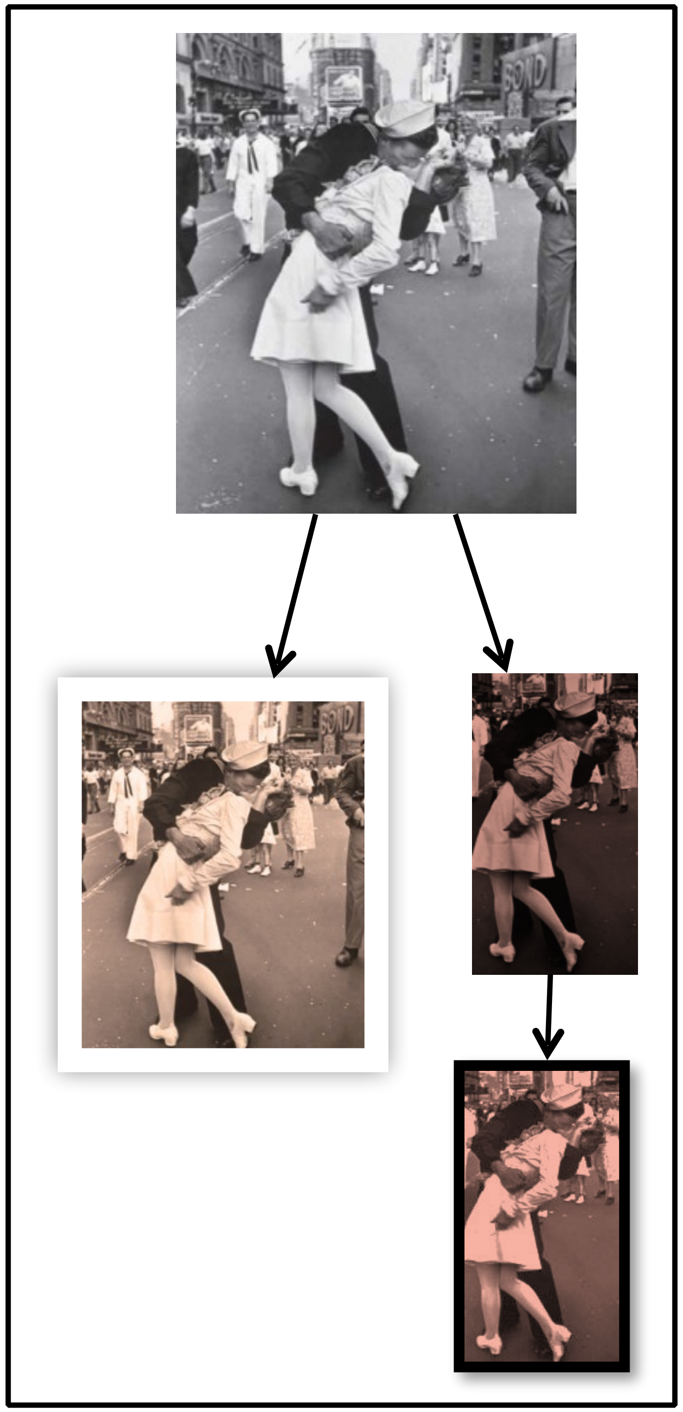
Source - Alien



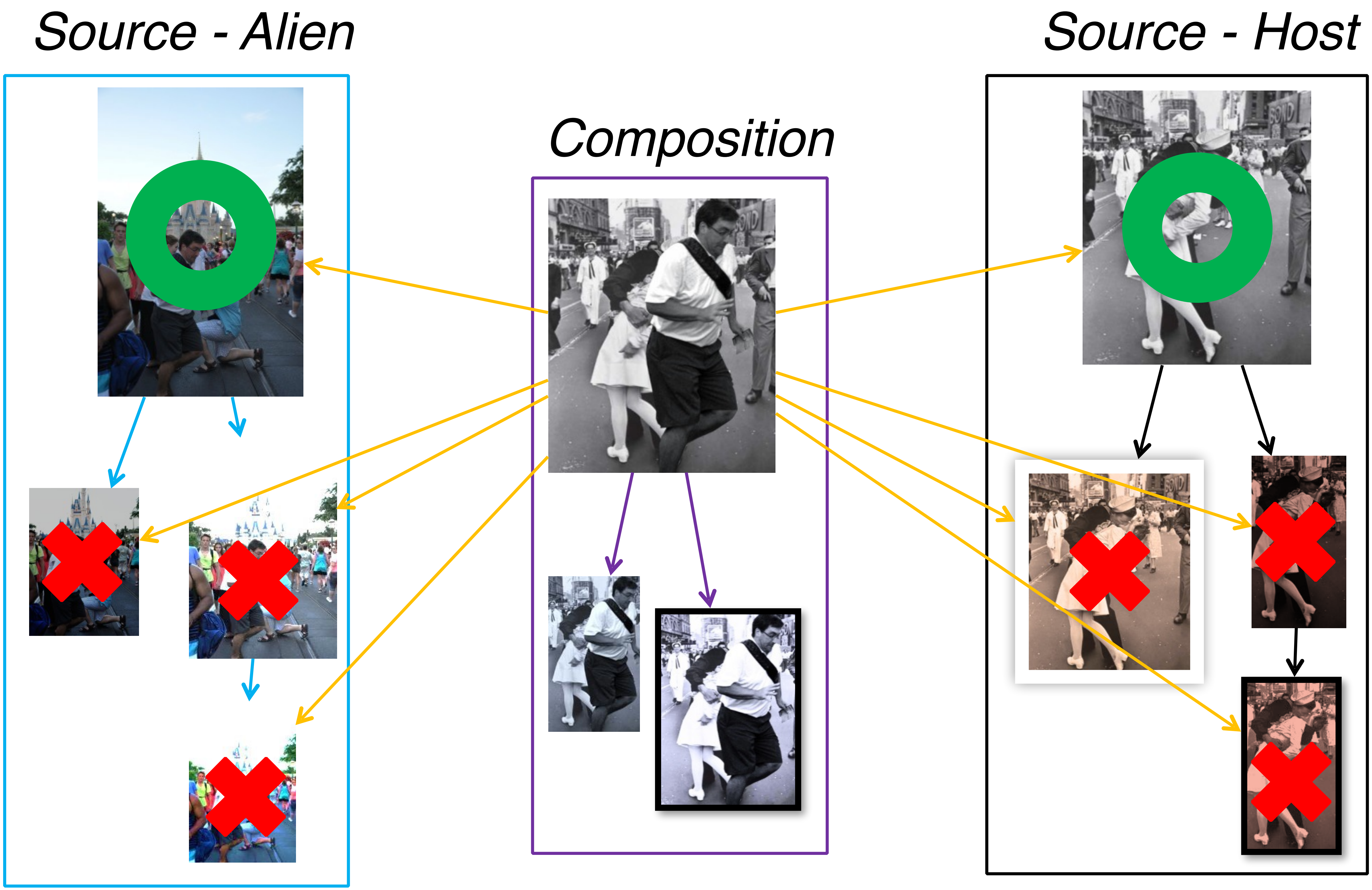
Composition



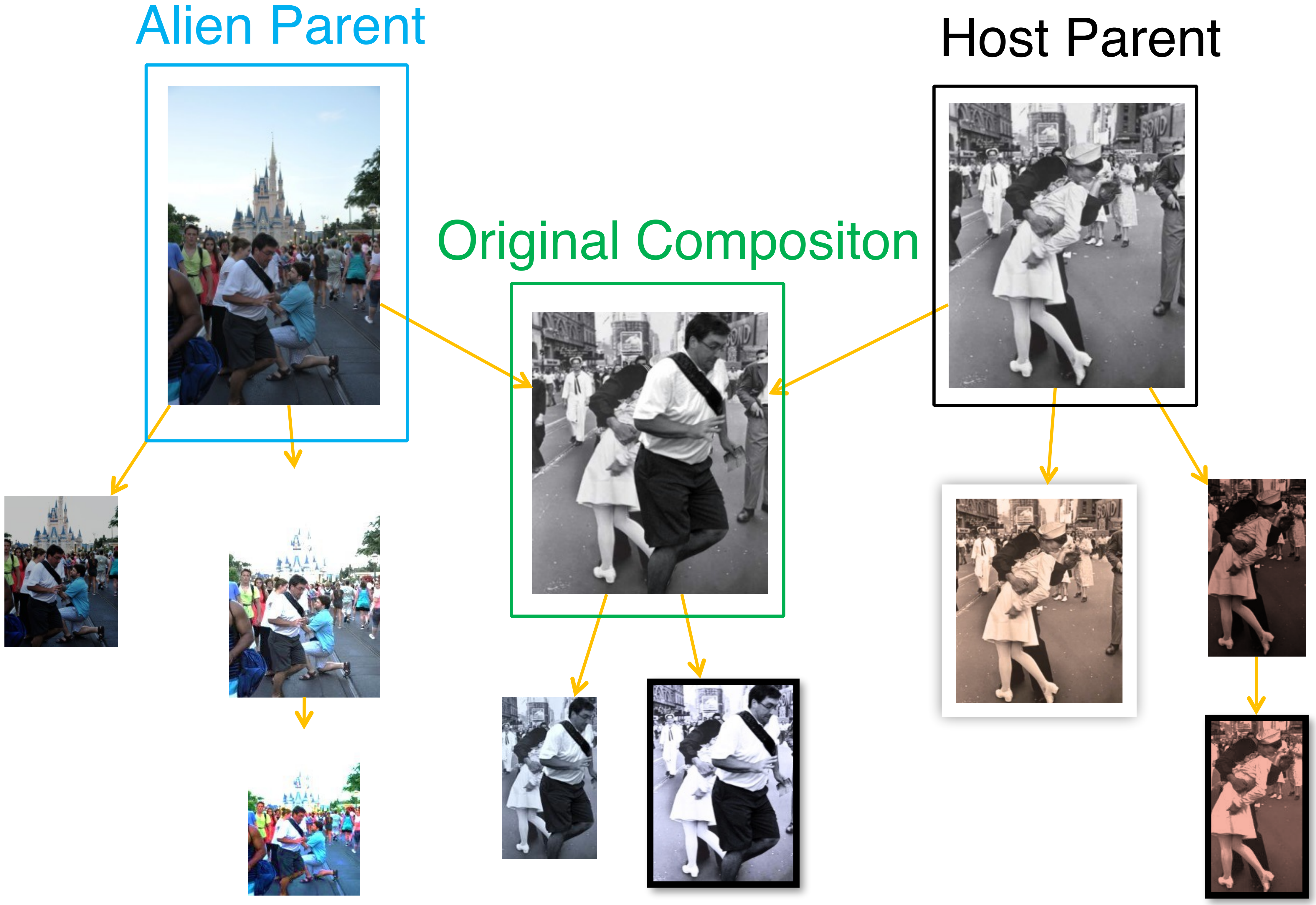
Source - Host



Finding the parents



Final Graph



Results

- The method was tested with two types of splicing compositions:
  - Easy case: Direct pasting
  - Hard case: Poisson blending

- 300 hundred test cases of each type with phylogeny trees having 25 nodes

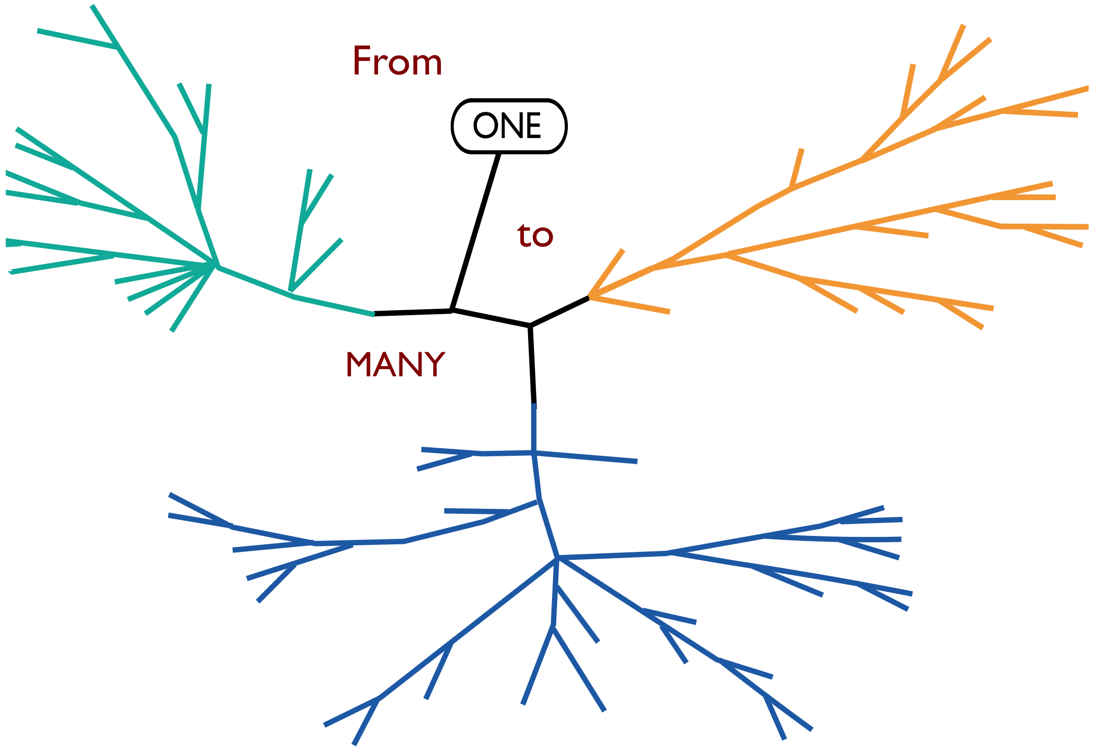
	Original Composition	**Host Parent**	**Alien Parent**
Direct Pasting	73.0%	76.0%	69.3%
Poisson Blending	66.3%	73.0%	42.0%

Contributions & Future Work

- We achieved good identification of the original composition and its host parent

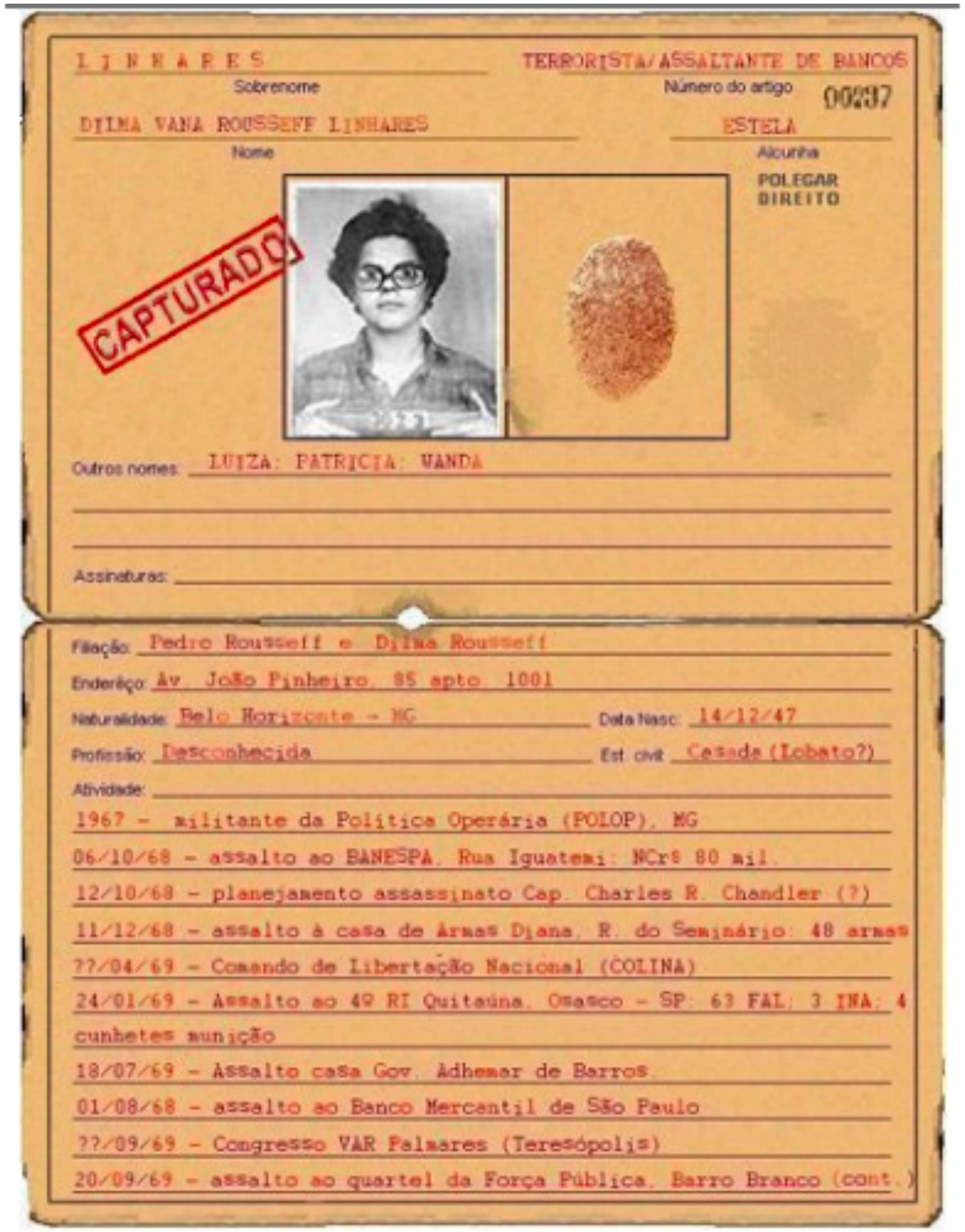
- The alien parent identification still needs improvement, specially for poisson blending

- Future work:
  - Handmade and professional compositions
  - Enhanced validation
  - Generalization of the method



What can we do with these approaches?

The Brazilian President Criminal Record



LINHARES

Sobrenome

TERRORISTA/ASSALTANTE DE BANCOS

Número do artigo 00237

DILMA VANA ROUSSEFF LINHARES

Nome

ESTELA

Alcunha

POLEGAR DIREITO

CAPTURADO

Outros nomes: LUIZA; PATRICIA; WANDA

Assinaturas:

Filiação: Pedro Rousseff e Dilma Rousseff

Endereço: Av. João Pinheiro, 85 apto. 1001

Naturalidade: Belo Horizonte - MG

Data Nasc: 14/12/47

Profissão: Desconhecida

Est. civil: Casada (Lobato?)

Atividade:

1967 - militante da Política Operária (POLOP), MG

06/10/68 - assalto ao BANESPA, Rua Iguatemi: NCr$ 80 mil.

12/10/68 - planejamento assassinato Cap. Charles R. Chandler (?)

11/12/68 - assalto à casa de Armas Diana, R. do Seminário: 48 armas

??/04/69 - Comando de Libertação Nacional (COLINA)

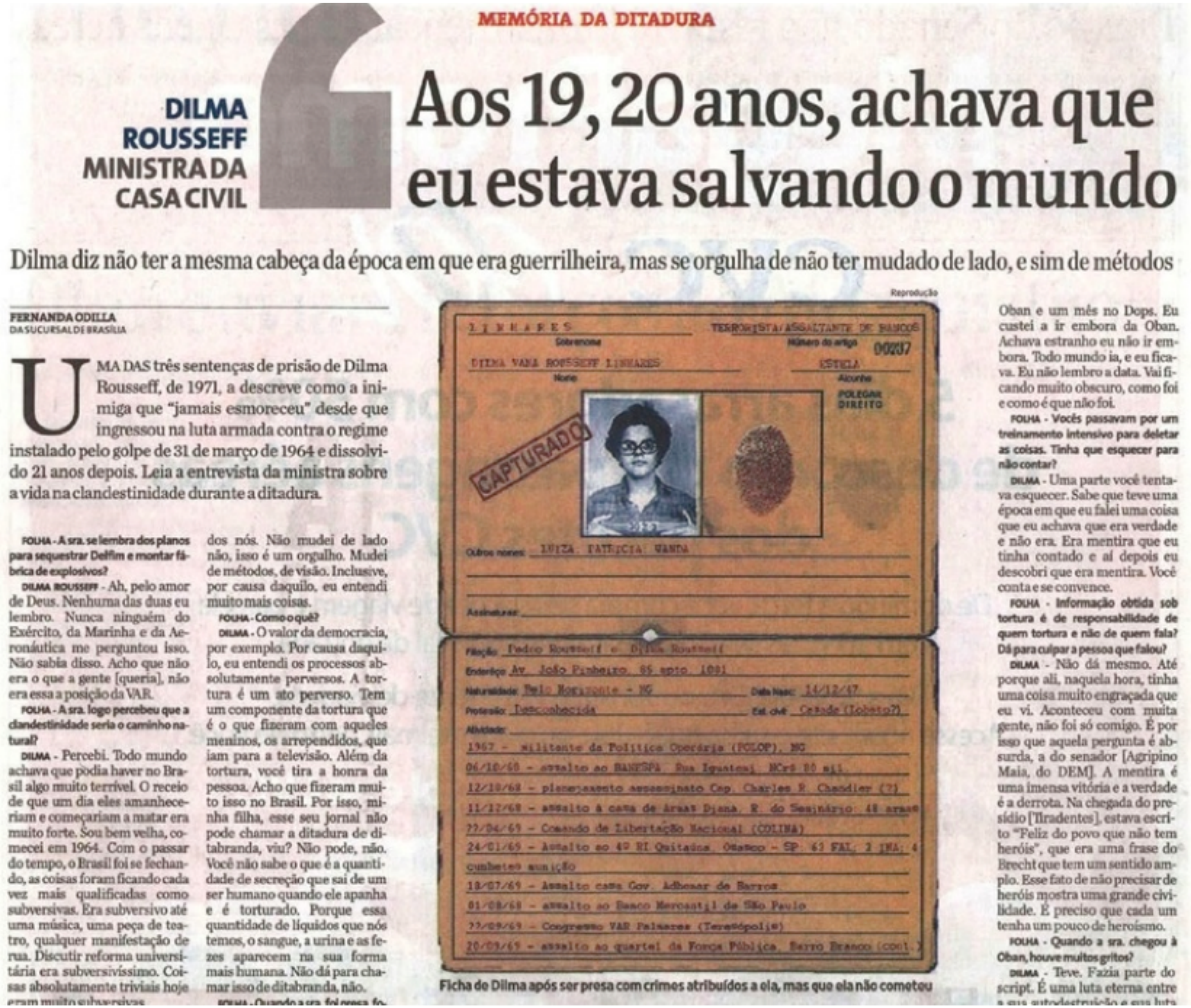
24/01/69 - Assalto ao 4º RI Quitaúna, Osasco - SP: 63 FAL; 3 INA; 4 cunhetes munição

18/07/69 - Assalto casa Gov. Adhemar de Barros.

01/08/68 - assalto ao Banco Mercantil de São Paulo

??/09/69 - Congresso VAR Palmares (Teresópolis)

20/09/69 - assalto ao quartel da Força Pública, Barro Branco (cont.)



MEMÓRIA DA DITADURA

DILMA ROUSSEFF
MINISTRA DA CASA CIVIL

"Aos 19, 20 anos, achava que eu estava salvando o mundo

Dilma diz não ter a mesma cabeça da época em que era guerrilheira, mas se orgulha de não ter mudado de lado, e sim de métodos

FERNANDA ODILLA
DA SUCURSAL DE BRASÍLIA

UMA DAS três sentenças de prisão de Dilma Rousseff, de 1971, a descreve como a inimiga que "jamais esmoreceu" desde que ingressou na luta armada contra o regime instalado pelo golpe de 31 de março de 1964 e dissolvido 21 anos depois. Leia a entrevista da ministra sobre a vida na clandestinidade durante a ditadura.

FOLHA - A sra. se lembra dos planos para sequestrar Delfim e montar fábrica de explosivos?

DILMA ROUSSEFF - Ah, pelo amor de Deus. Nenhuma das duas eu lembro. Nunca ninguém do Exército, da Marinha e da Aeronáutica me perguntou isso. Não sabia disso. Acho que não era o que a gente [queria], não era essa a posição da VAR.

FOLHA - A sra. logo percebeu que a clandestinidade seria o caminho natural?

DILMA - Percebi. Todo mundo achava que podia haver no Brasil algo muito terrível. O receio de que um dia eles amanheceriam e começariam a matar era muito forte. Sou bem velha, comecei em 1964. Com o passar do tempo, o Brasil foi se fechando, as coisas foram ficando cada vez mais qualificadas como subversivas. Era subversivo até uma música, uma peça de teatro, qualquer manifestação de rua. Discutir reforma universitária era subversivíssimo. Coisas absolutamente triviais hoje eram muito subversivas. [...] dos nós. Não mudei de lado não, isso é um orgulho. Mudei de métodos, de visão. Inclusive, por causa daquilo, eu entendi muito mais coisas.

FOLHA - Como o quê?

DILMA - O valor da democracia, por exemplo. Por causa daquilo, eu entendi os processos absolutamente perversos. A tortura é um ato perverso. Tem um componente da tortura que é o que fizeram com aqueles meninos, os arrependidos, que iam para a televisão. Além da tortura, você tira a honra da pessoa. Acho que fizeram muito isso no Brasil. Por isso, minha filha, esse seu jornal não pode chamar a ditadura de ditabranda, viu? Não pode, não. Você não sabe o que é a quantidade de secreção que sai de um ser humano quando ele apanha e é torturado. Porque essa quantidade de líquidos que nós temos, o sangue, a urina e as fezes aparecem na sua forma mais humana. Não dá para chamar isso de ditabranda, não.

FOLHA - Quando a sra. foi presa fo-

Reprodução

LINHARES
TERRORISTA/ASSALTANTE DE BANCOS
Sobrenome
Número do artigo 00237
DILMA VANA ROUSSEFF LINHARES
ESTELA
Nome
Alcunha
POLEGAR DIREITO
CAPTURADO
Outros nomes: LUIZA PATRICIA WANDA
Assinaturas:
Filiação: Pedro Rousseff e Dilma Rousseff
Endereço: Av. João Pinheiro, 85 apto. 1001
Naturalidade: Belo Horizonte - MG
Data Nasc: 14/12/47
Profissão: Desconhecida
Est. civil: Casada (Lobato?)
Atividade:
1967 - militante da Política Operária (POLOP), MG
06/10/68 - assalto ao BANESPA, Rua Iguatemi, NCr$ 80 mil.
12/10/68 - planejamento assassinato Cap. Charles R. Chandler (?)
11/12/68 - assalto à casa de Armas Diana, R. do Seminário: 48 armas
22/04/69 - Comando de Libertação Nacional (COLINA)
24/01/69 - Assalto ao 4º RI Quitaúna, Osasco - SP: 63 FAL; 3 INA; 4 cunhetes munição
18/07/69 - Assalto casa Gov. Adhemar de Barros
01/08/69 - assalto ao Banco Mercantil de São Paulo
22/09/69 - Congresso VAR Palmares (Teresópolis)
20/09/69 - assalto ao quartel da Força Pública, Barro Branco (cont.)

Ficha de Dilma após ser presa com crimes atribuídos a ela, mas que ela não cometeu

Oban e um mês no Dops. Eu custei a ir embora da Oban. Achava estranho eu não ir embora. Todo mundo ia, e eu ficava. Eu não lembro a data. Vai ficando muito obscuro, como foi e como é que não foi.

FOLHA - Vocês passavam por um treinamento intensivo para deletar as coisas. Tinha que esquecer para não contar?

DILMA - Uma parte você tentava esquecer. Sabe que teve uma época em que eu falei uma coisa que eu achava que era verdade e não era. Era mentira que eu tinha contado e aí depois eu descobri que era mentira. Você conta e se convence.

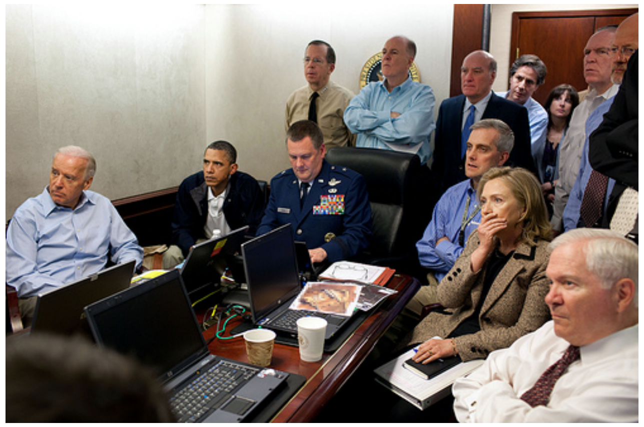
FOLHA - Informação obtida sob tortura é de responsabilidade de quem tortura e não de quem fala? Dá para culpar a pessoa que falou?

DILMA - Não dá mesmo. Até porque ali, naquela hora, tinha uma coisa muito engraçada que eu vi. Aconteceu com muita gente, não foi só comigo. É por isso que aquela pergunta é absurda, a do senador [Agripino Maia, do DEM]. A mentira é uma imensa vitória e a verdade é a derrota. Na chegada do presídio [Tiradentes], estava escrito "Feliz do povo que não tem heróis", que era uma frase do Brecht que tem um sentido amplo. Esse fato de não precisar de heróis mostra uma grande civilidade. É preciso que cada um tenha um pouco de heroísmo.

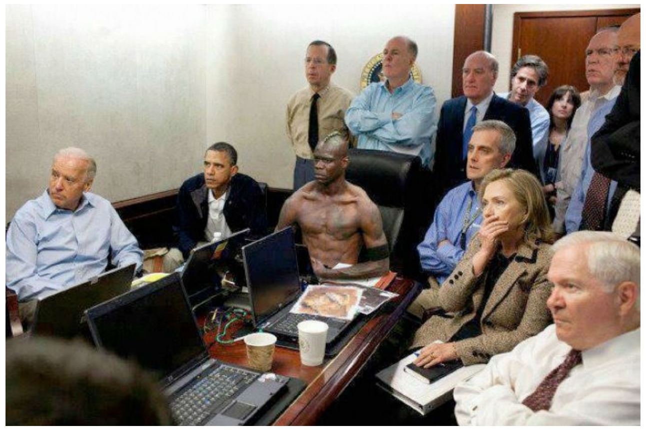
FOLHA - Quando a sra. chegou à Oban, houve muitos gritos?

DILMA - Teve. Fazia parte do script. É uma luta eterna entre a sua autodestruição e sua luta

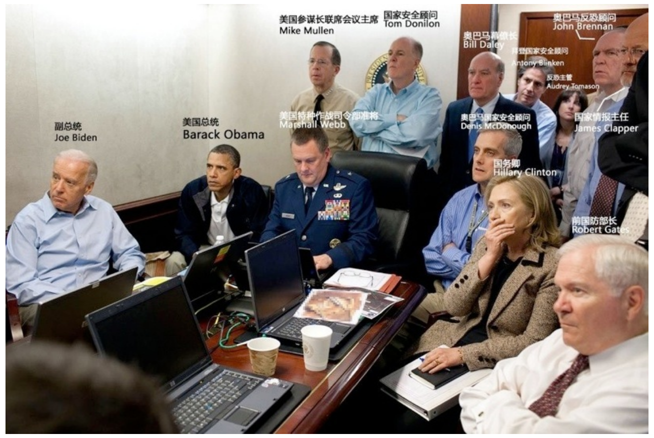
The Situation Room



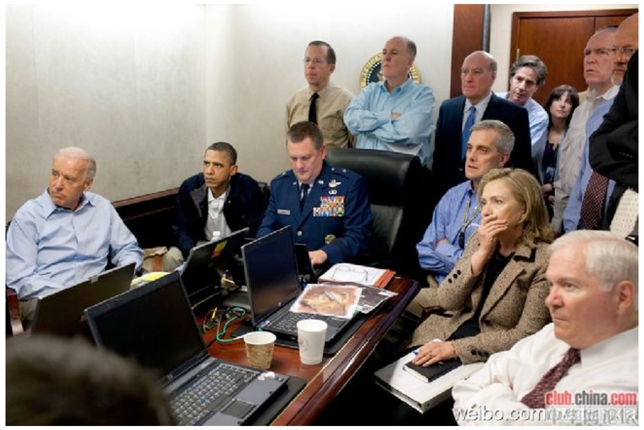
The Situation Room
(The White House version)



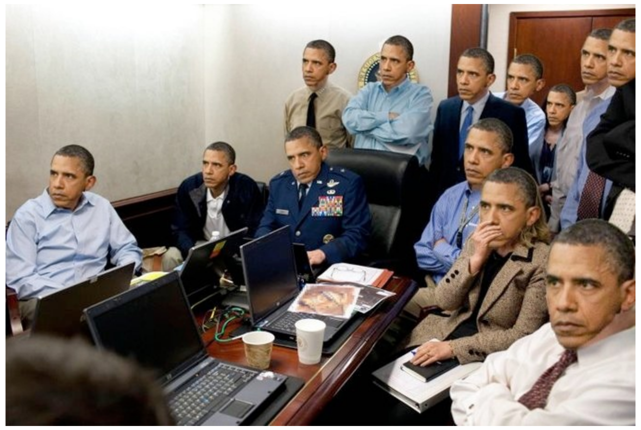
Balotelli (ID a*)



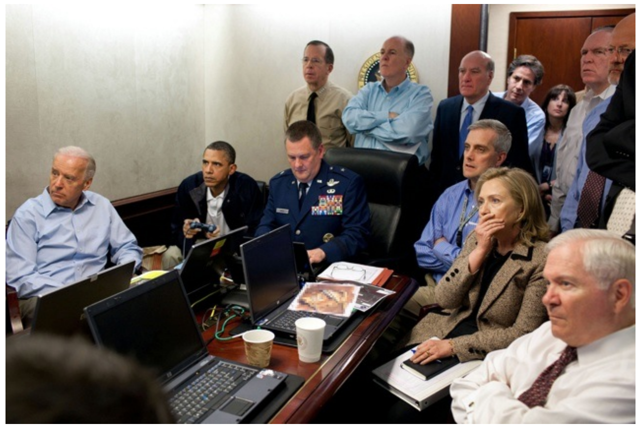
Text Overlay (ID b*)



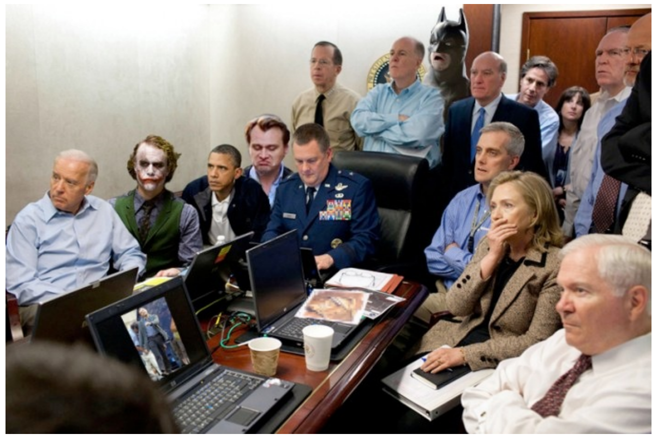
Watermarking (ID c*)



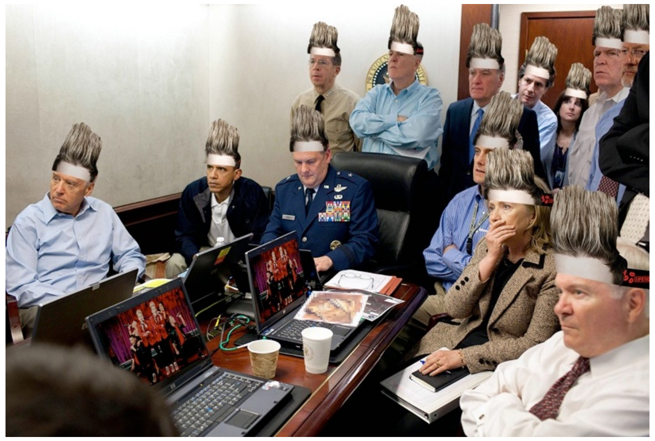
Face Swapping (ID d*)



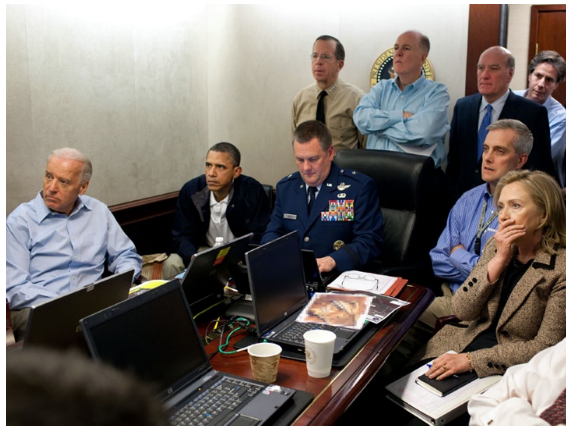
Splicing Objects (ID e*)



Splicing People (ID f*)



Splicing Objects and
Changing Content (ID g*)

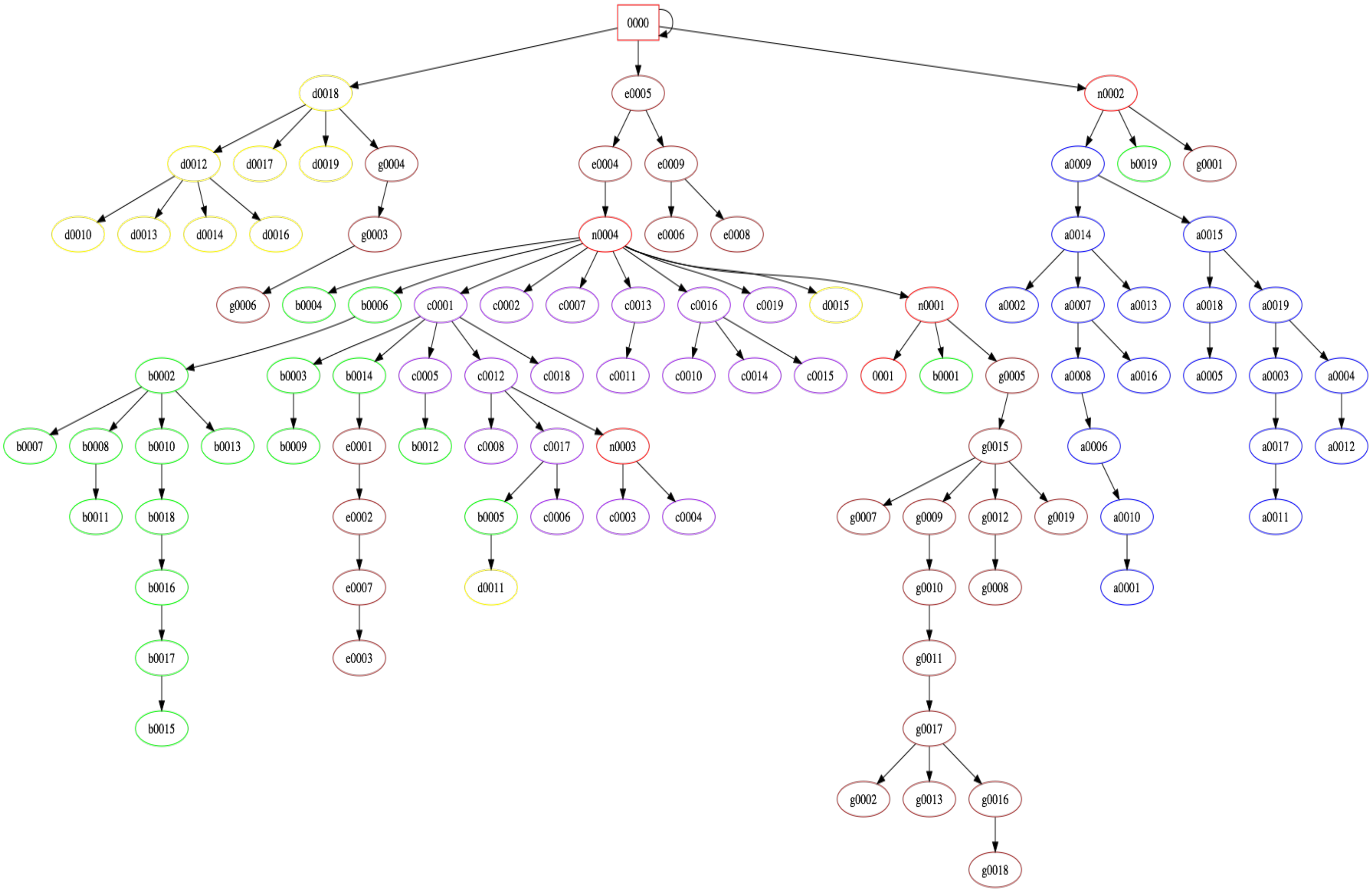


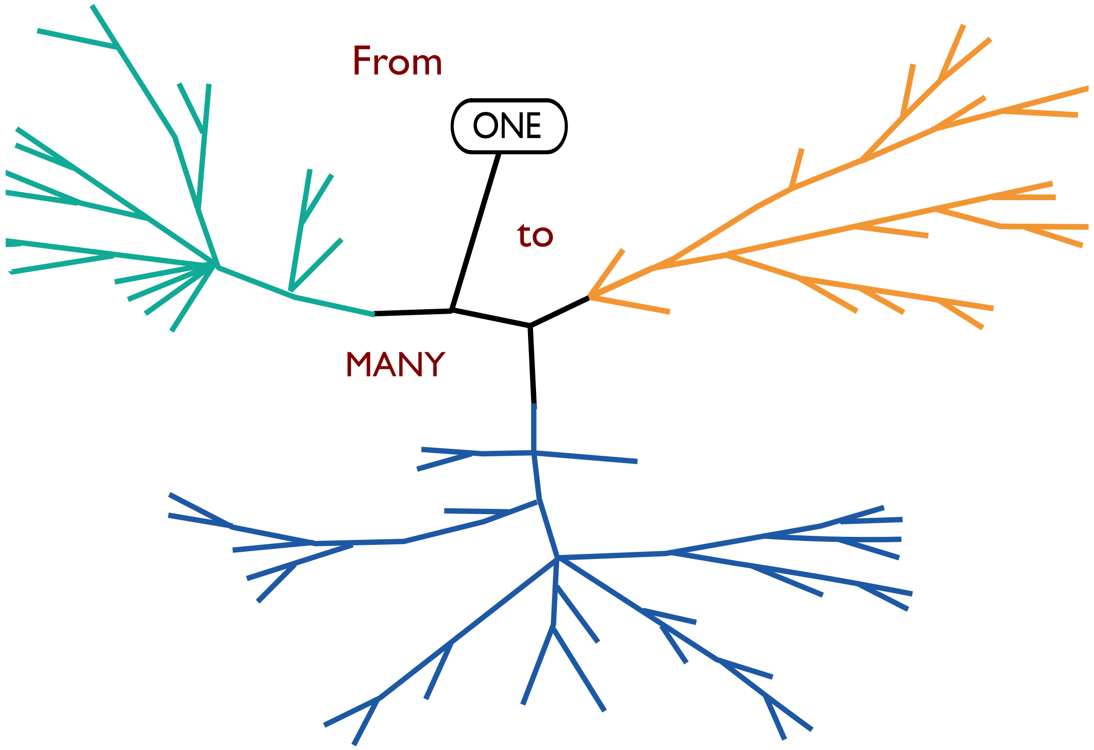
Cropping/Zoom (ID h*)

SEVENTH FRAMEWORK PROGRAMME

REWIND

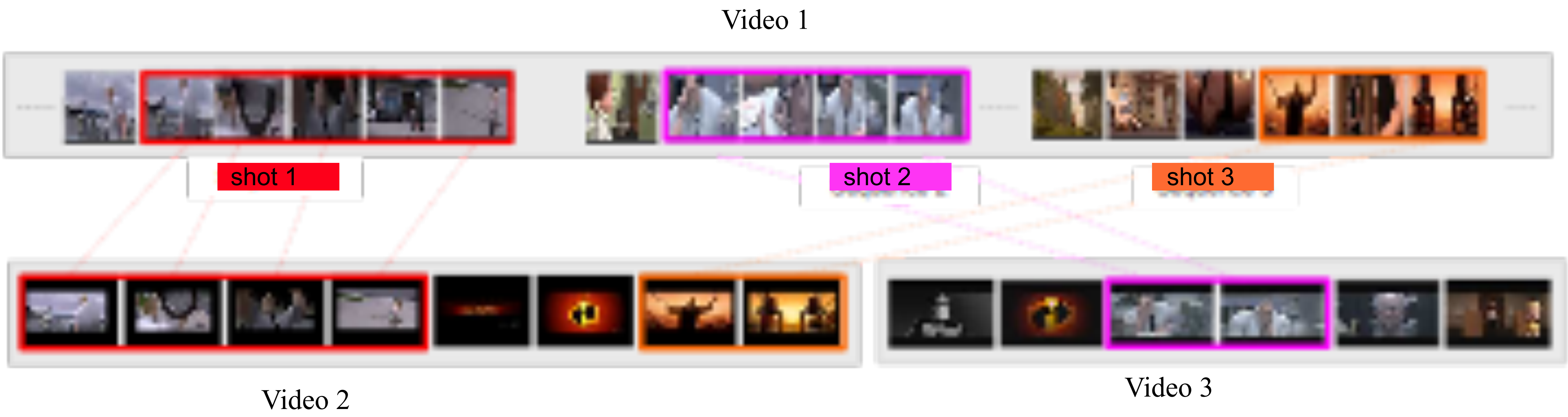
First Steps into Forests





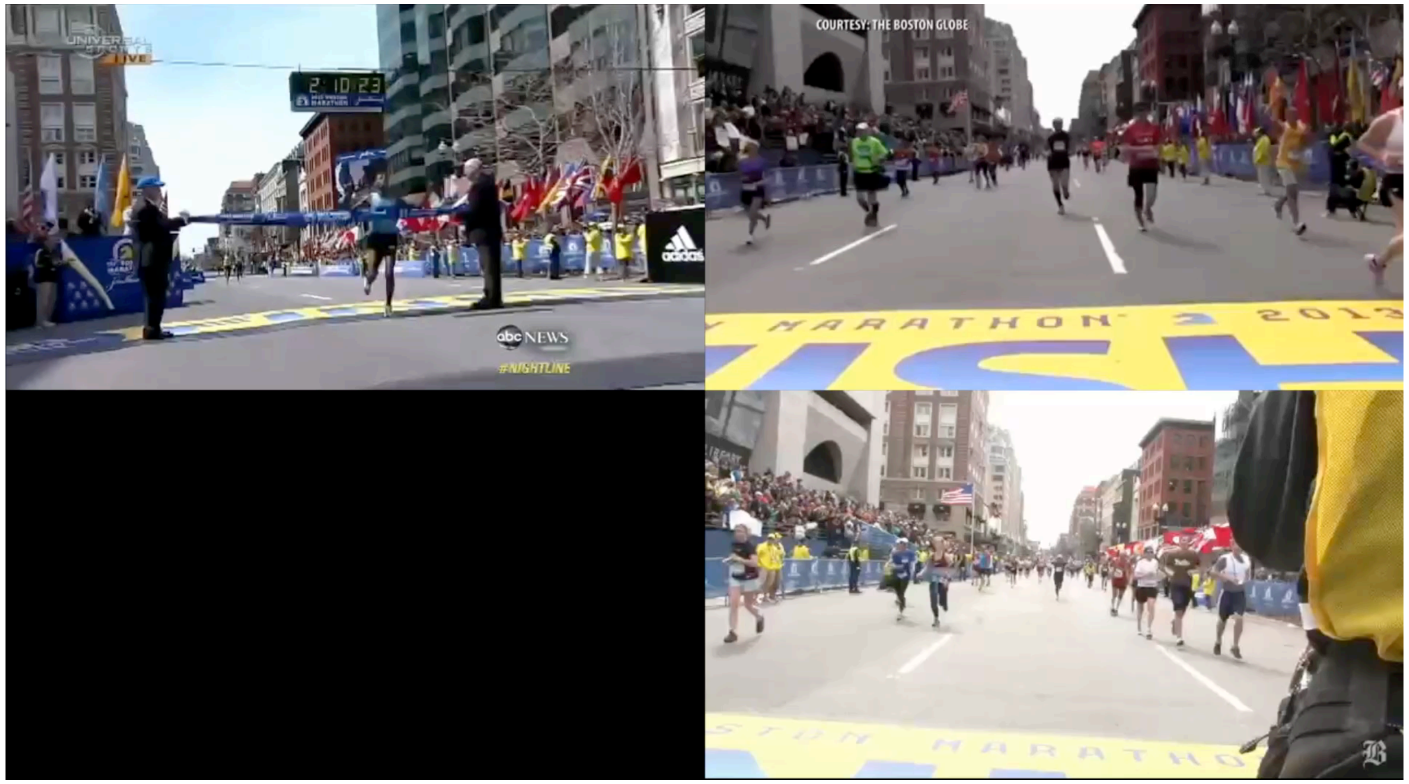
Video Phylogeny Part II :-)

Motivation



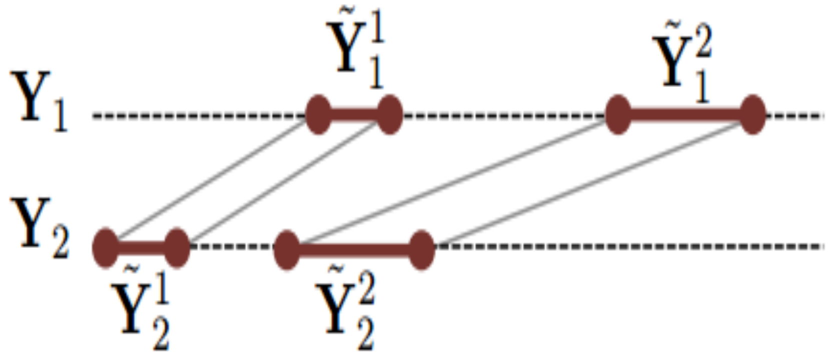
Video Phylogeny - Example

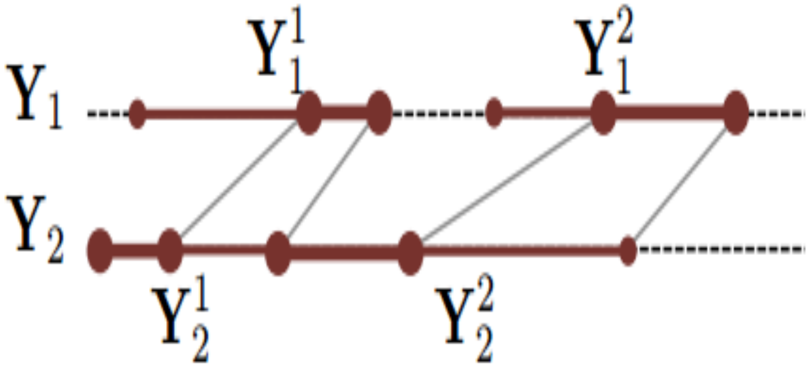
Observed sequences

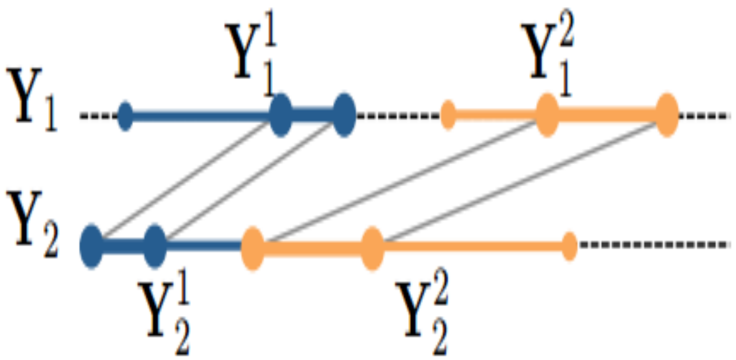


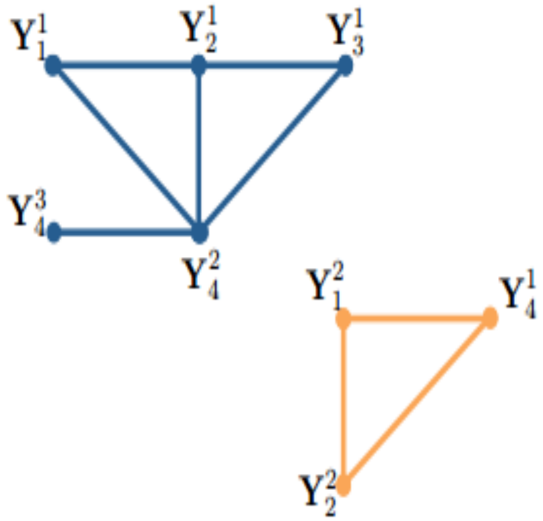
Parent sequence

Parent sequence









Sequences comparison

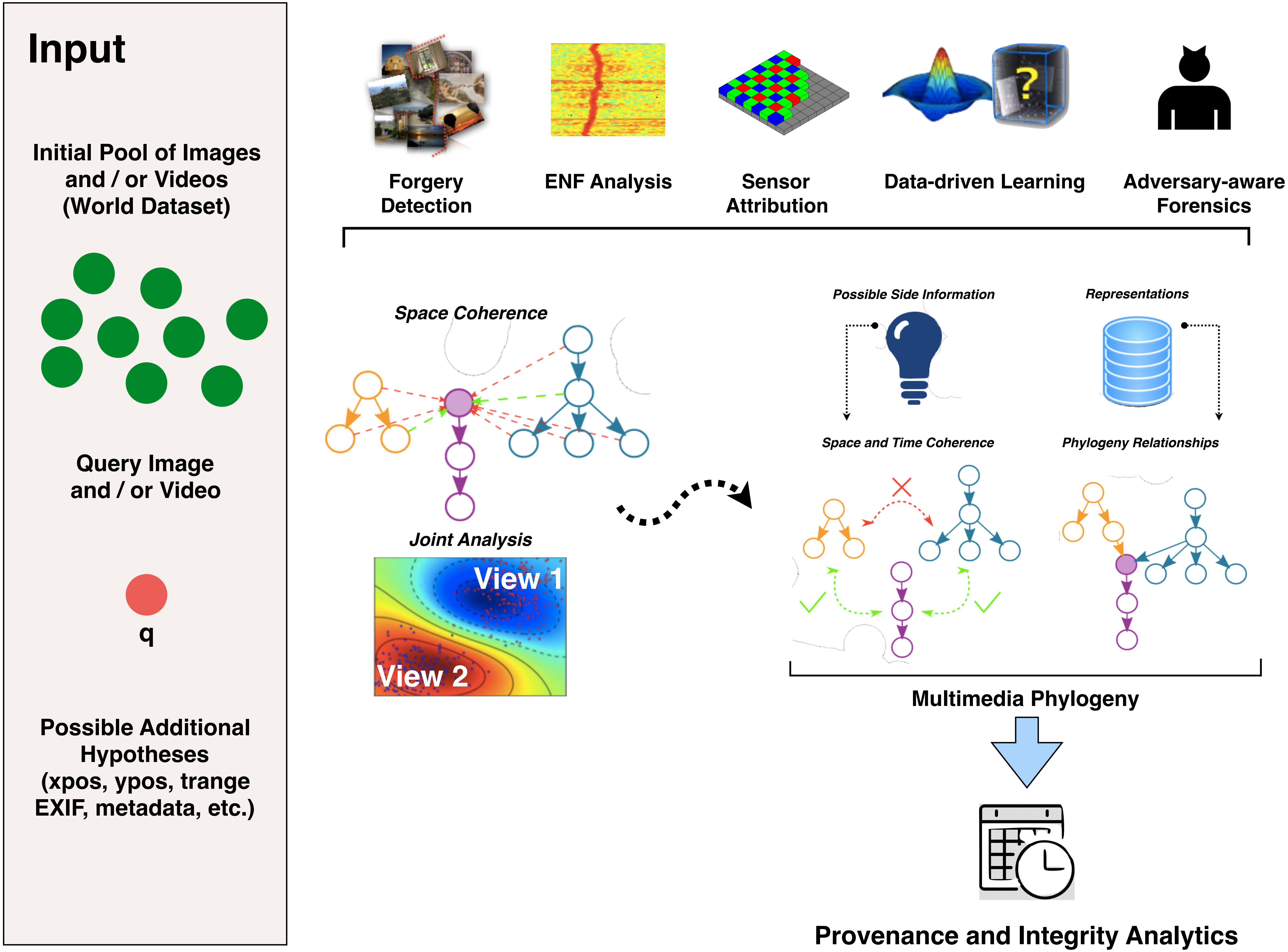
Results

- Dataset:
  - 12 standard sequences at CIF resolution
- Transformations
  - Blurring, brightness adjustment, contrast enhancement, spatial cropping, AVC/H.264 coding, logo insertion, rotation
- Results
  - Perfect parent reconstruction: 85%
  - Parent reconstruction (missing one shot): > 90%

What's next?

Media Forensics and Integrity Analytics

PURDUE UNIVERSITY POLITECNICO MILANO 1863 NYU·poly USC



Input

Initial Pool of Images and / or Videos (World Dataset)

Query Image and / or Video

q

Possible Additional Hypotheses (xpos, ypos, trange EXIF, metadata, etc.)

Forgery Detection

ENF Analysis

Sensor Attribution

Data-driven Learning

Adversary-aware Forensics

Space Coherence

Joint Analysis

Possible Side Information

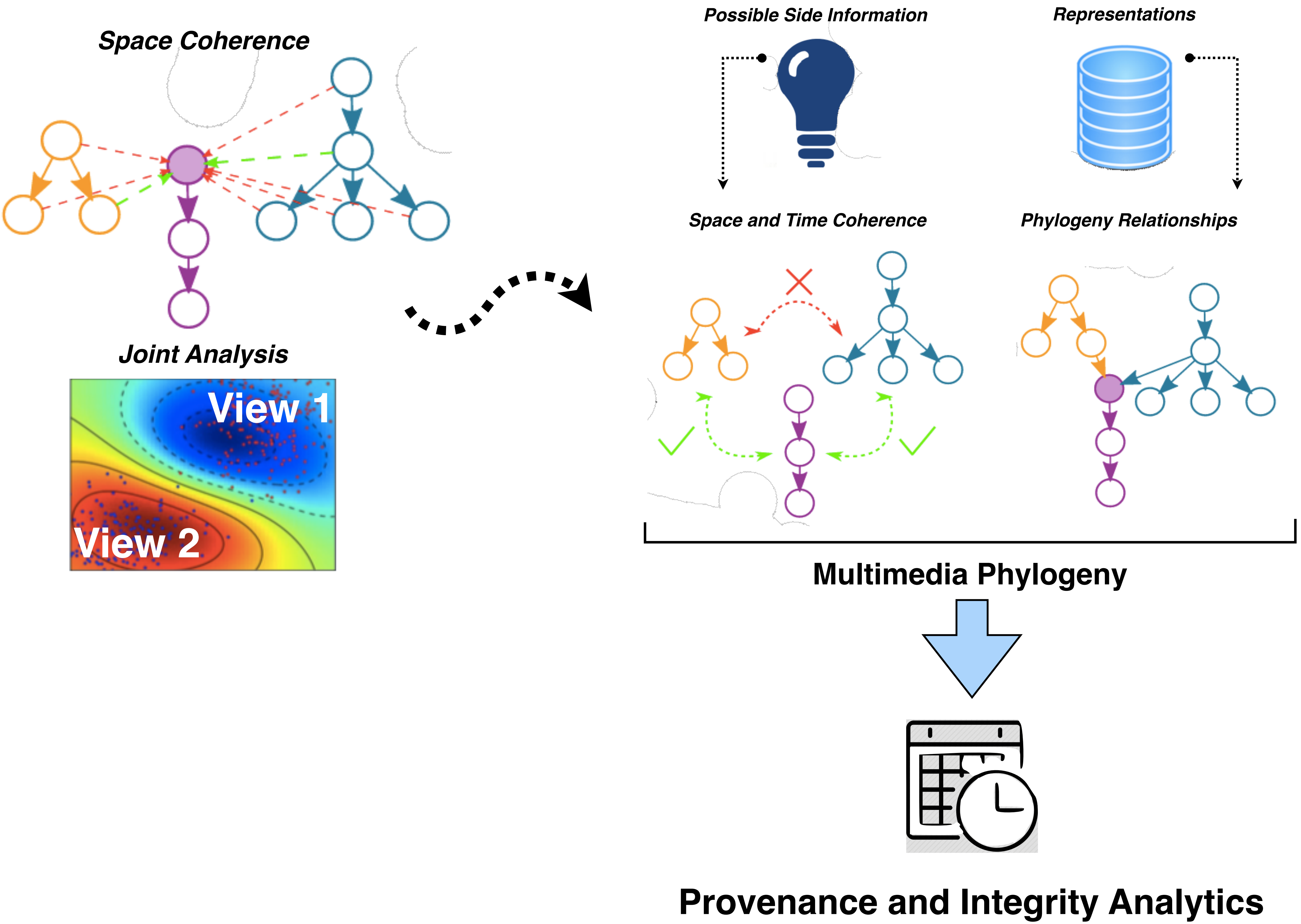
Representations

Space and Time Coherence

Phylogeny Relationships

Multimedia Phylogeny

Provenance and Integrity Analytics



Space Coherence

Joint Analysis

Possible Side Information

Representations

Space and Time Coherence

Phylogeny Relationships

Multimedia Phylogeny

Provenance and Integrity Analytics

Research Team

Research Team

Purdue Univ., U.S.

E. Delp
(Principal Investigator)

USC, U.S.

C.-C. Jay Kuo

University of Siena, Italy

M. Barni

A. Agnetis

M. Maggini

Politécnico di Milano, Italy

S. Tubaro

P. Bestagini

NYU, U.S.

N. Memon

Univ. of Notre Dame, U.S.

P. Flynn

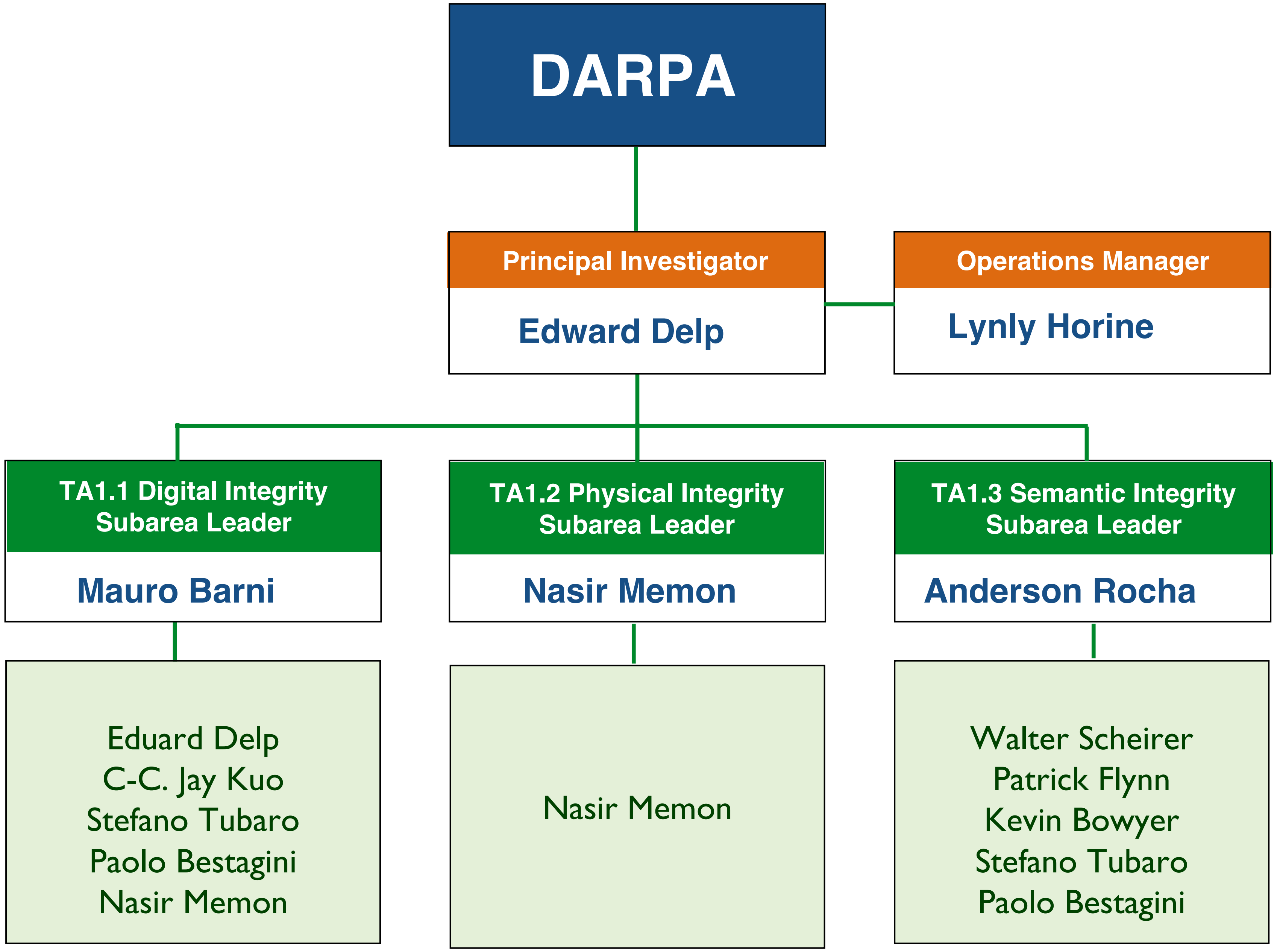
K. Bowyer

W. Scheirer

**Notre Dame, U.S. &
Unicamp, Brazil**

A. Rocha

Media Forensics Integrity Analytics



DARPA

Principal Investigator

Edward Delp

Operations Manager

Lynly Horine

TA1.1 Digital Integrity Subarea Leader

Mauro Barni

Eduard Delp
C-C. Jay Kuo
Stefano Tubaro
Paolo Bestagini
Nasir Memon

TA1.2 Physical Integrity Subarea Leader

Nasir Memon

Nasir Memon

TA1.3 Semantic Integrity Subarea Leader

Anderson Rocha

Walter Scheirer
Patrick Flynn
Kevin Bowyer
Stefano Tubaro
Paolo Bestagini

Organizational Structure

Overview

- We are working on three general areas:
  - TA1.1: Source identification, manipulation detection and localization, adversarial setups, editing suite identification, video analyses
  - TA1.2: electrical network frequency (ENF)-based video authentication
  - TA1.3: Multimedia Phylogeny and Joint Analysis

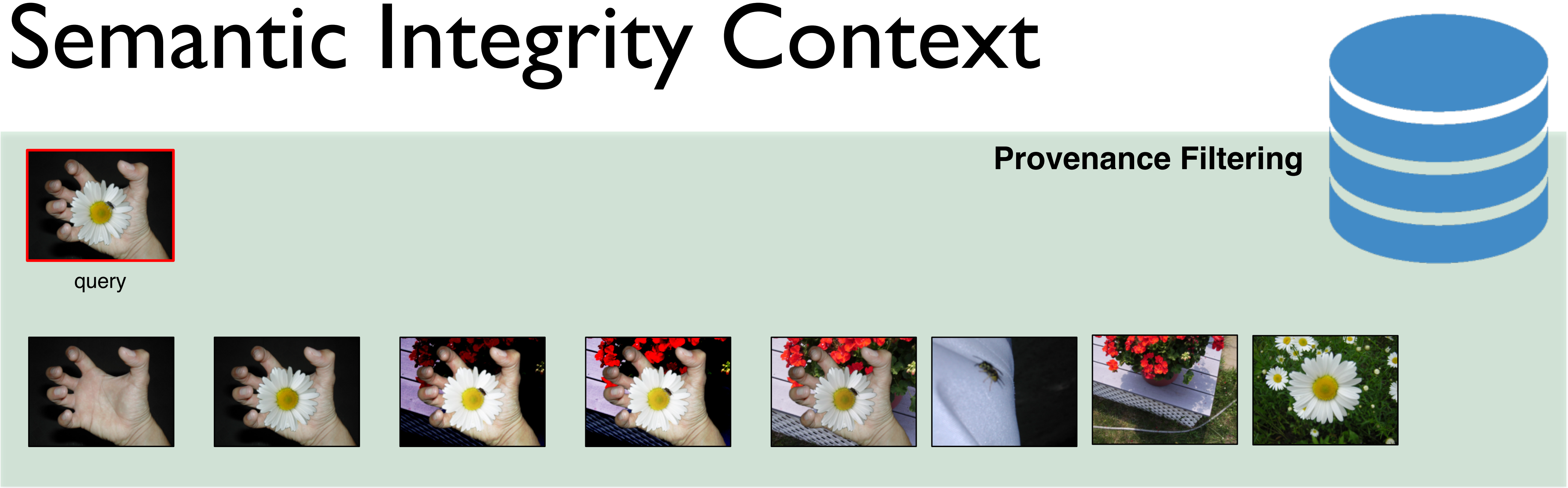
TA1.3 Goals

- **Main goal:** semantic analysis of media collections using the principles of media phylogeny to characterize media content and relationships
- What is in the scenes, their spatial coherence, and the timeline of relationships among the media objects in the pool
- **X-coherence**: space, time, and digital relationships

TA1.3 Approaches

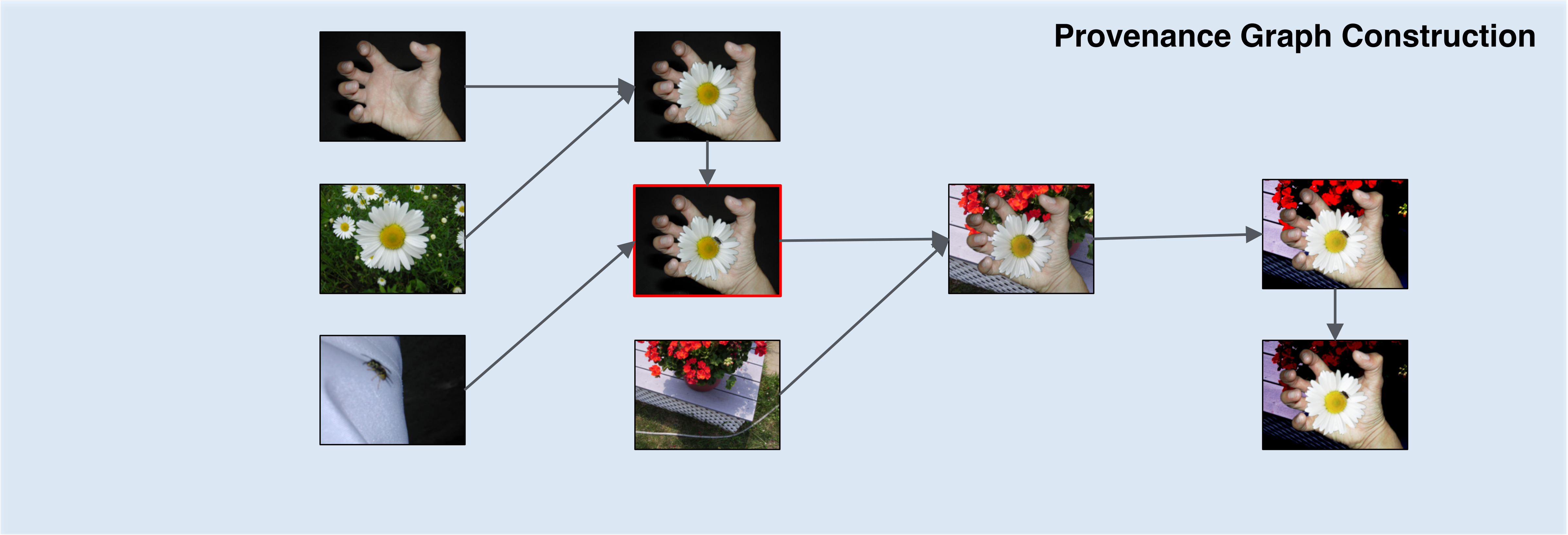
- Two main lines of action
  - Phylogenetic representations from media corpora (determination of provenance, spatial and temporal correlation among objects)
  - Semantic-level manipulation detection
- Pinpoint possible links among the objects and their processing history rather than producing an integrity indicator (e.g., TA1.1 & TA1.2)

Semantic Integrity Context



Provenance Filtering

query

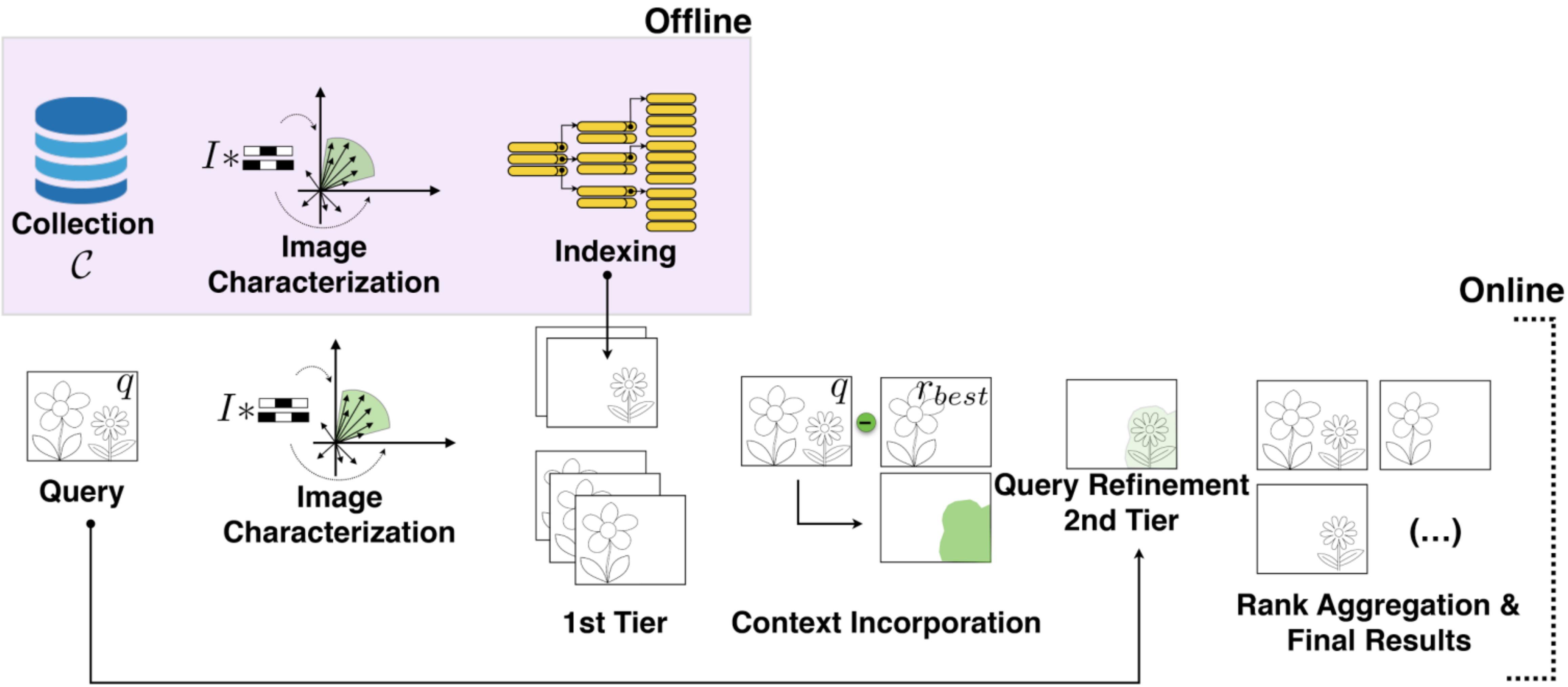


Provenance Graph Construction

ICIP Paper #1
https://arxiv.org/abs/1706.00447

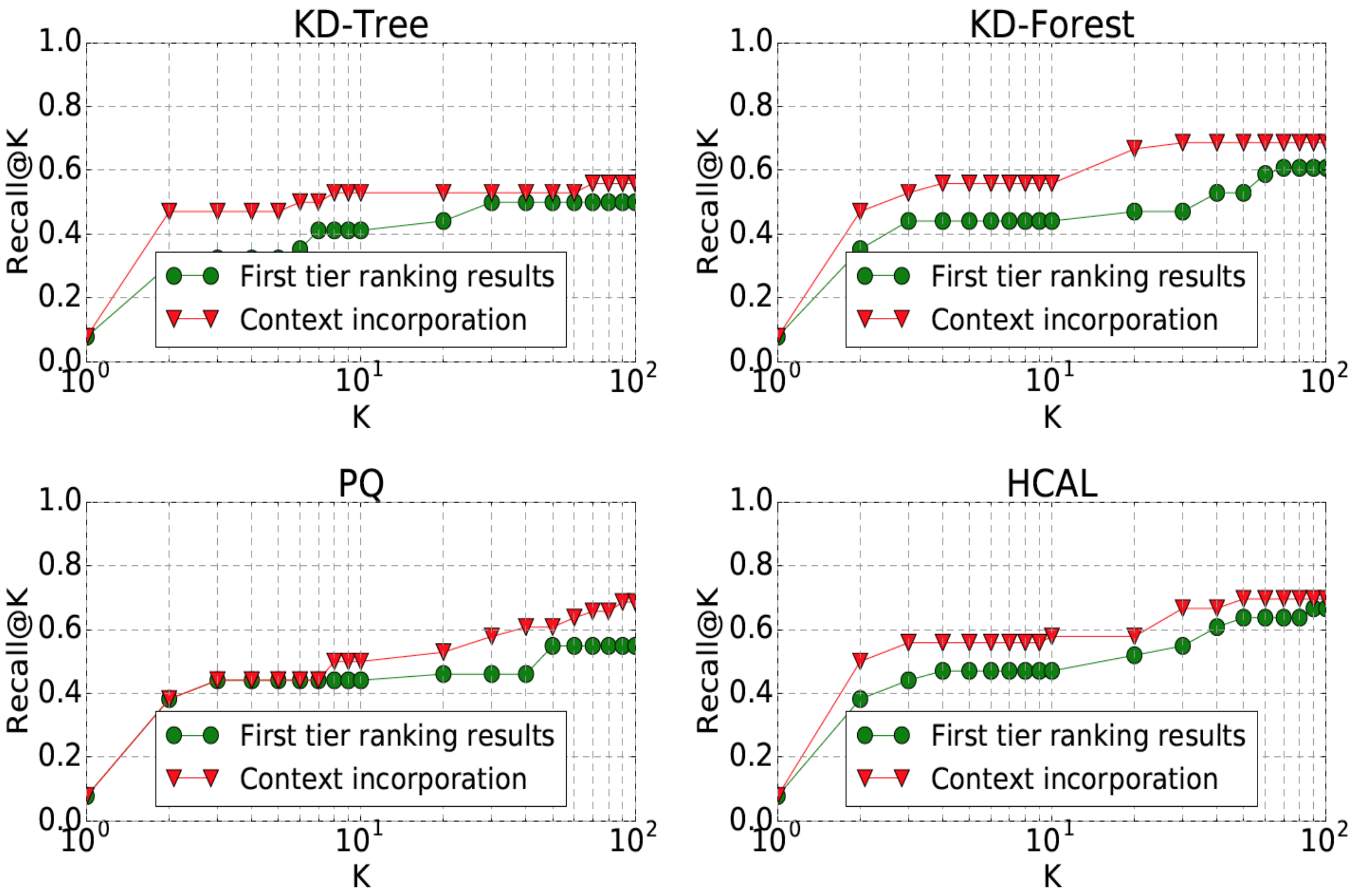
Provenance Filtering

Filtering the gallery and selecting donor candidates for a query



Provenance Filtering

Filtering the gallery and selecting donor candidates for a query



First- and second-tier results in terms of Recall@k.

The context incorporation is important regardless of the used indexing technique.

ICIP Paper #2
https://arxiv.org/abs/1705.00604

2017 ICIP
Beijing, China

Context Incorporation

Context retrieval and analysis for improved forgery detection and localization

- Collaboration with Polimi/Italy
- Robust tampering detection on large-scale datasets
- Focus on the difference between query and its donor candidates

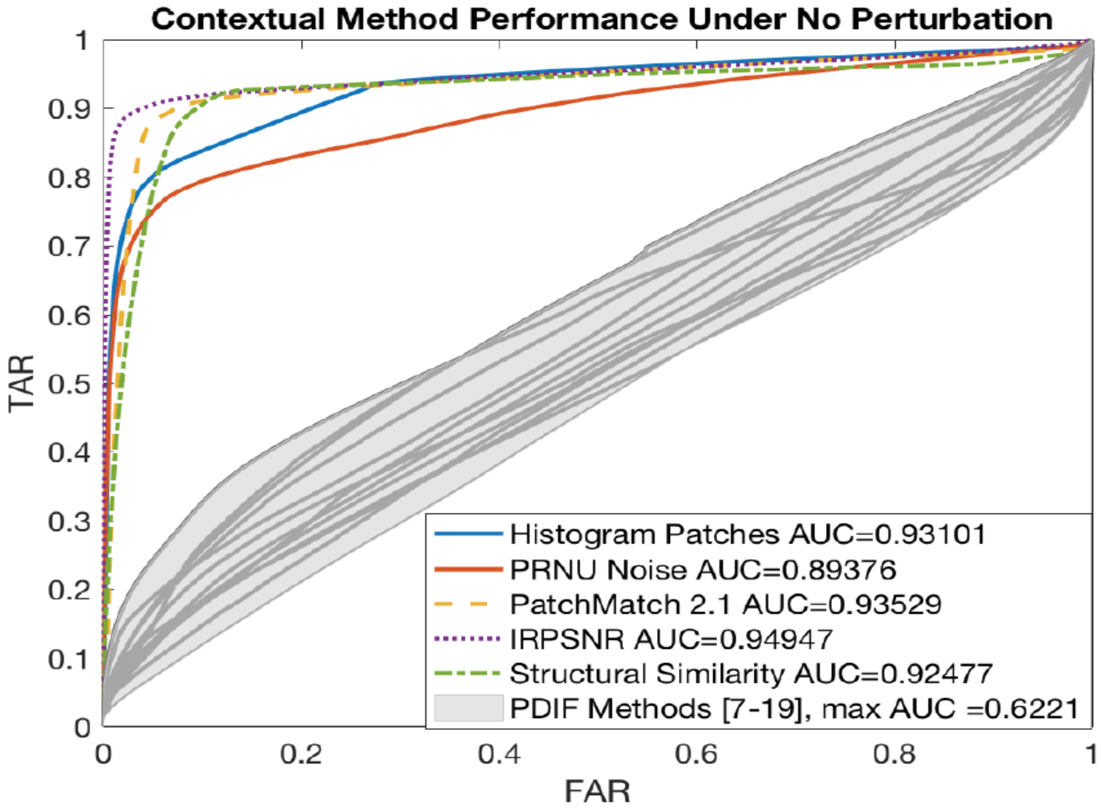
ICIP Paper #2
https://arxiv.org/abs/1705.00604

2017 ICIP
Beijing, China

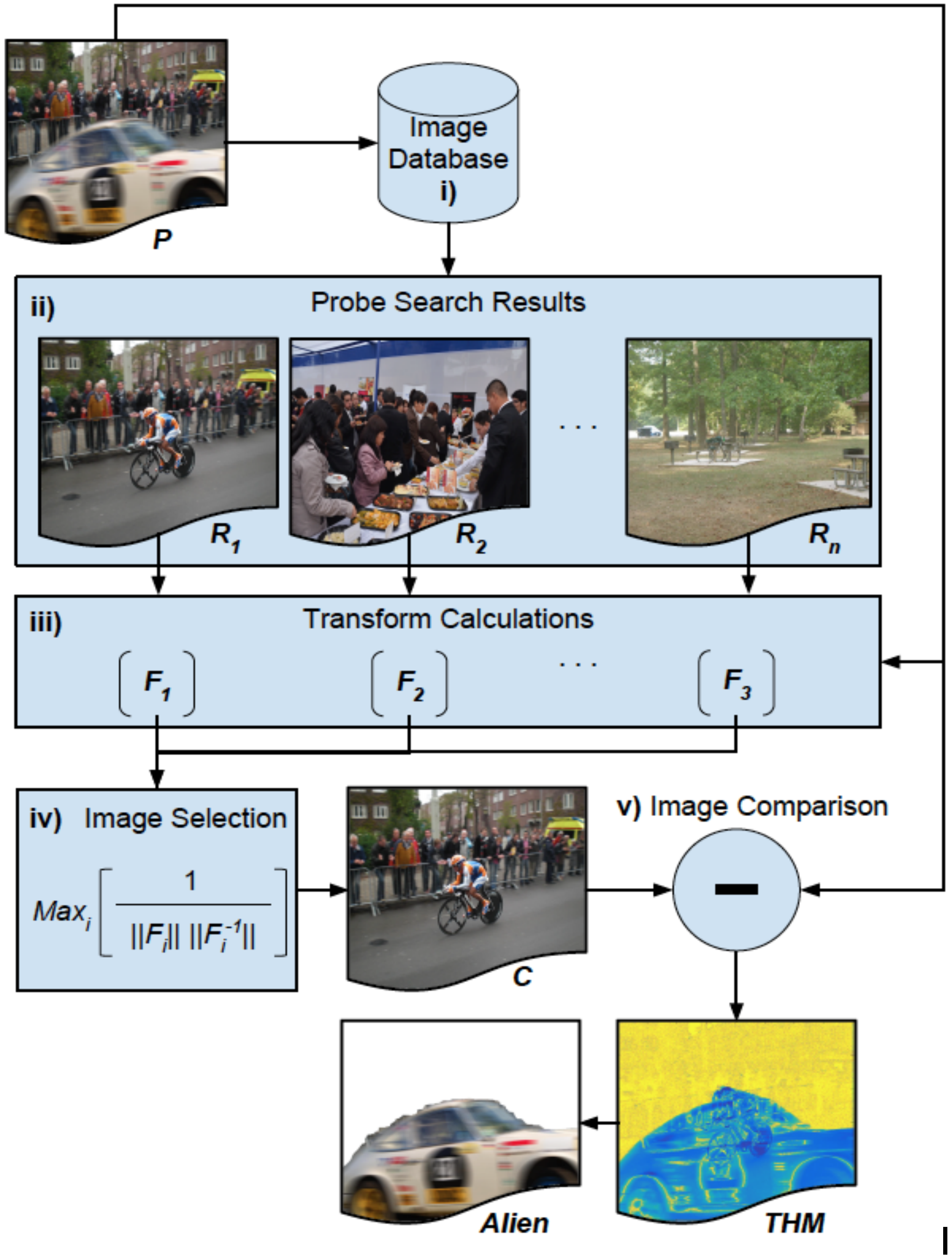
Context Incorporation

Context retrieval and analysis for improved forgery detection and localization

- THM – Tampering heat maps



* Comparison with 13 forgery detectors in the literature



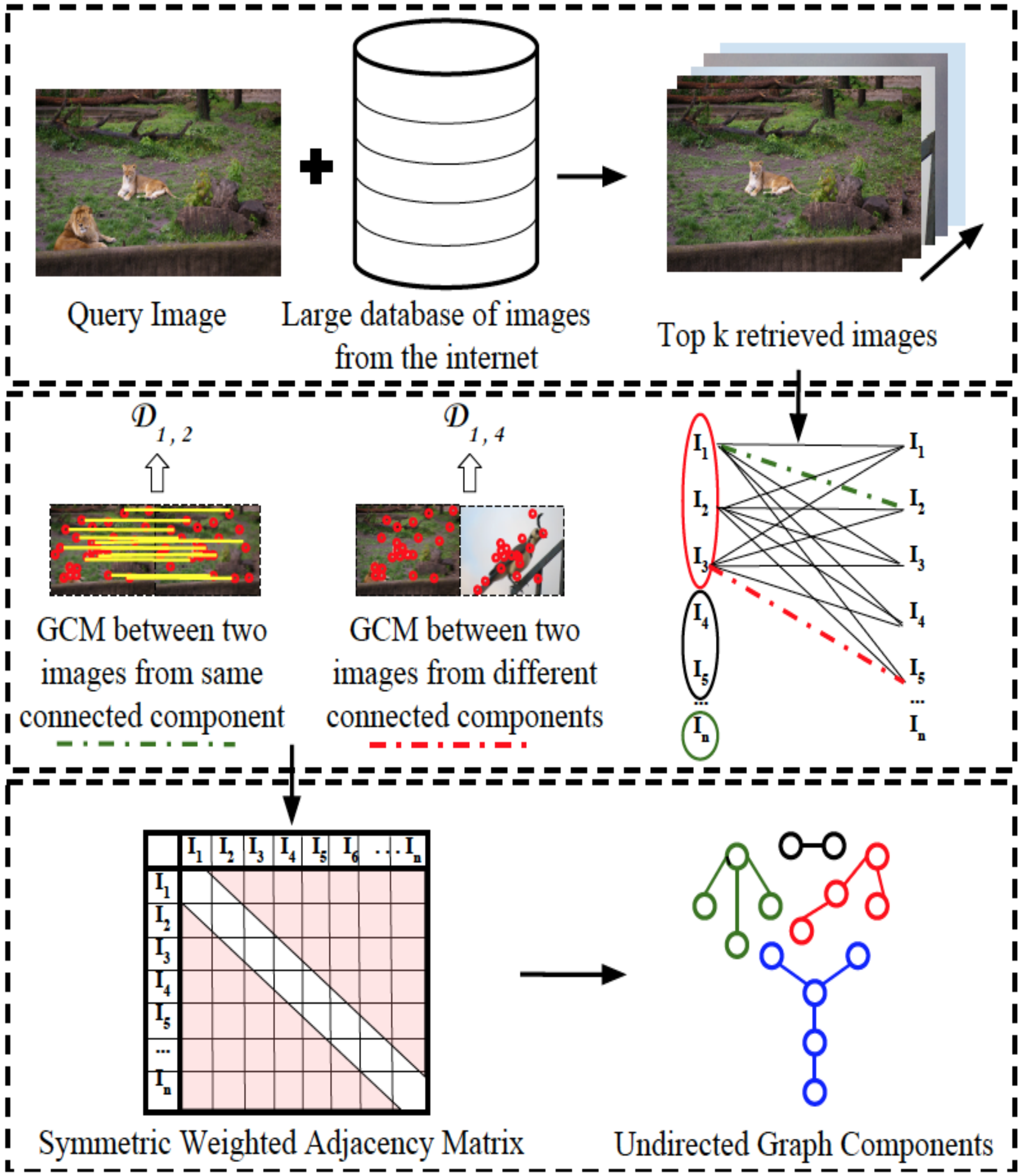
ICIP Paper #3
https://arxiv.org/abs/1705.11187

2017 ICIP
Beijing, China

U-Phylogeny

Undirected provenance graph construction in the wild

- Discovering how query and donor candidates are connected in terms of provenance
- Multiple parenting phylogeny without **literature's strong assumptions**
- Geometrical Consistency Check


Query Image
Large database of images from the internet
Top k retrieved images
$\mathcal{D}_{1,2}$
$\mathcal{D}_{1,4}$
GCM between two images from same connected component
GCM between two images from different connected components
I_1
I_2
I_3
I_4
I_5
...
I_n
I_1
I_2
I_3
I_4
I_5
I_6
. . I_n
Symmetric Weighted Adjacency Matrix
Undirected Graph Components

Phase 1 - Retrieving content related to the query from the Internet/Large Image Database

Phase 2 - Finding pair-wise dissimilarity between images

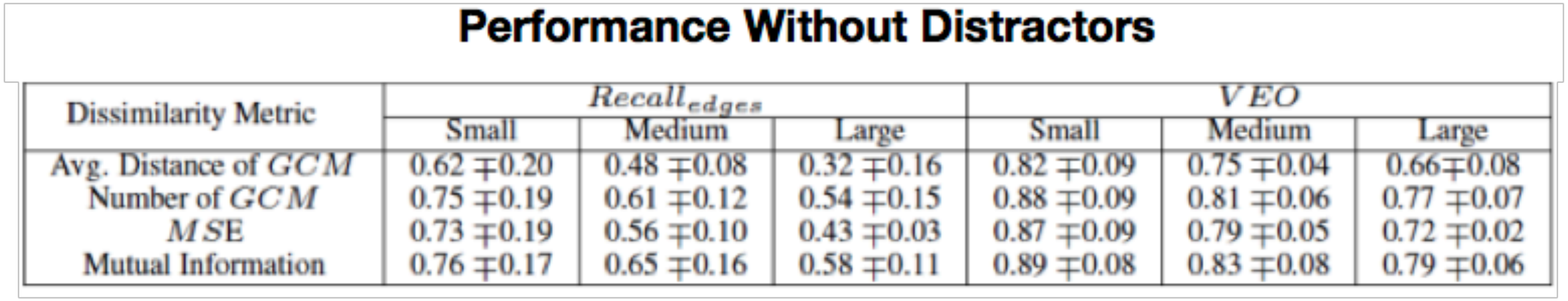
Phase 3 - Computing kxk Adjaceny Matrix and using a spanning tree algorithm to build an Undirected Graph

ICIP Paper #3
https://arxiv.org/abs/1705.11187

2017 ICIP
Beijing, China

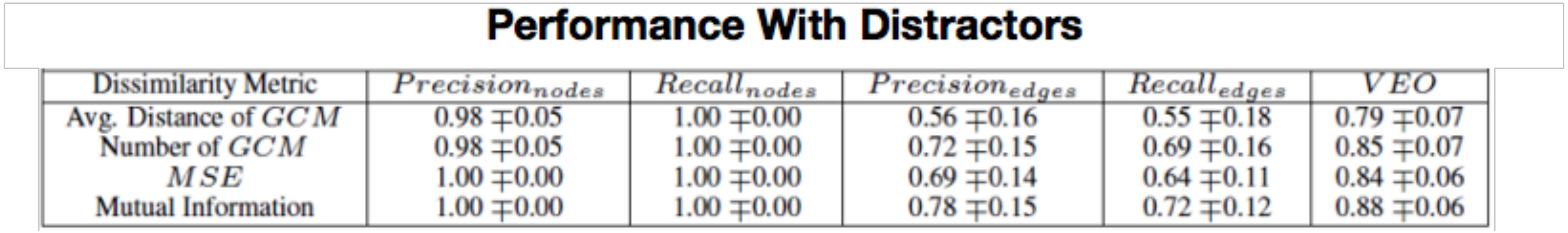
U-Phylogeny

Undirected provenance graph construction in the wild



Performance Without Distractors

Dissimilarity Metric	$Recall_{edges}$			VEO		
	Small	Medium	Large	Small	Medium	Large
Avg. Distance of GCM	0.62 ∓0.20	0.48 ∓0.08	0.32 ∓0.16	0.82 ∓0.09	0.75 ∓0.04	0.66∓0.08
Number of GCM	0.75 ∓0.19	0.61 ∓0.12	0.54 ∓0.15	0.88 ∓0.09	0.81 ∓0.06	0.77 ∓0.07
MSE	0.73 ∓0.19	0.56 ∓0.10	0.43 ∓0.03	0.87 ∓0.09	0.79 ∓0.05	0.72 ∓0.02
Mutual Information	0.76 ∓0.17	0.65 ∓0.16	0.58 ∓0.11	0.89 ∓0.08	0.83 ∓0.08	0.79 ∓0.06



Performance With Distractors

Dissimilarity Metric	$Precision_{nodes}$	$Recall_{nodes}$	$Precision_{edges}$	$Recall_{edges}$	VEO
Avg. Distance of GCM	0.98 ∓0.05	1.00 ∓0.00	0.56 ∓0.16	0.55 ∓0.18	0.79 ∓0.07
Number of GCM	0.98 ∓0.05	1.00 ∓0.00	0.72 ∓0.15	0.69 ∓0.16	0.85 ∓0.07
MSE	1.00 ∓0.00	1.00 ∓0.00	0.69 ∓0.14	0.64 ∓0.11	0.84 ∓0.06
Mutual Information	1.00 ∓0.00	1.00 ∓0.00	0.78 ∓0.15	0.72 ∓0.12	0.88 ∓0.06

Up to 25 nodes

Small: up to 12 nodes – **Medium**: from 13 up to 20 nodes – **Large**: more than 20 nodes

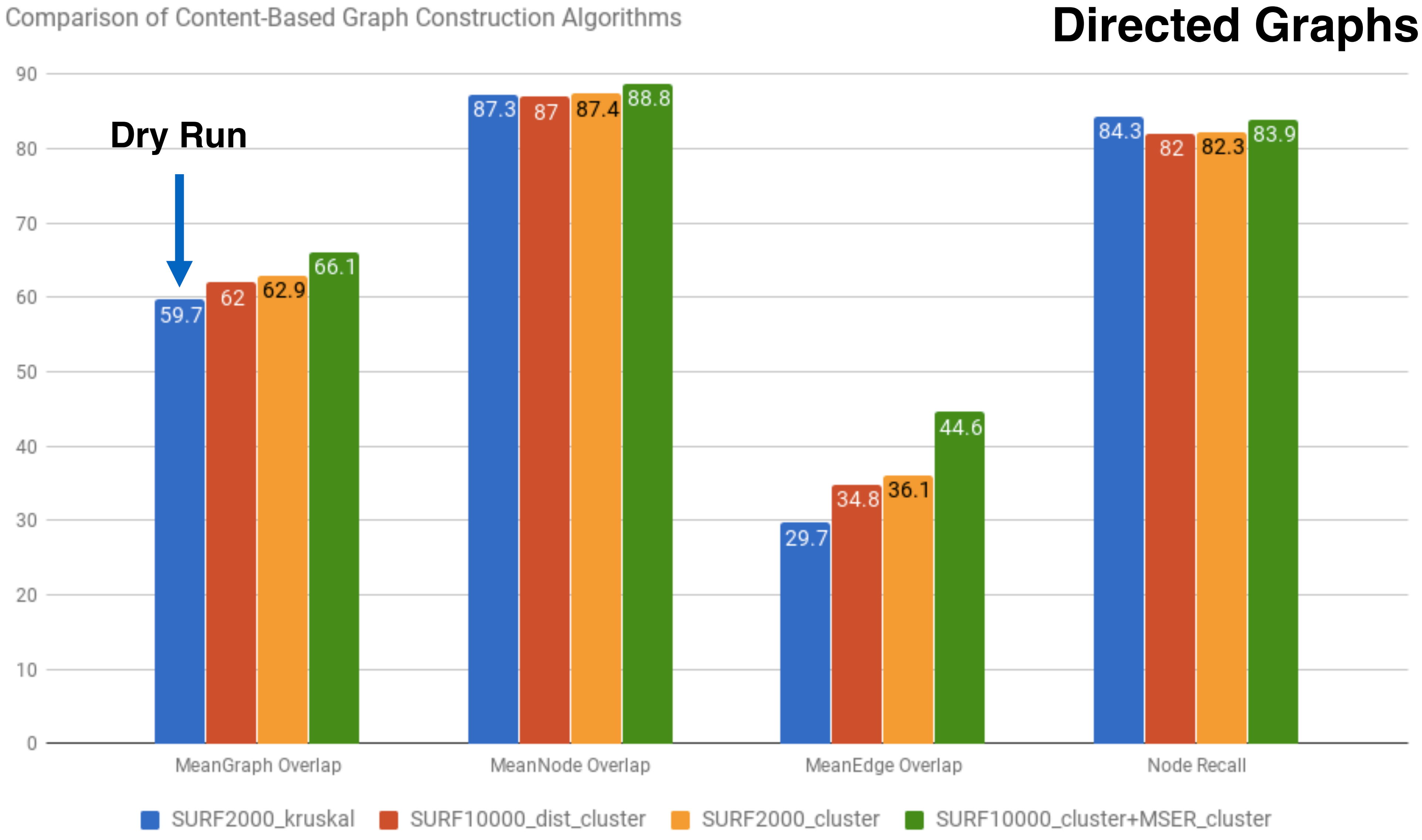
Phylogeny

Directed *provenance graph construction in the wild*

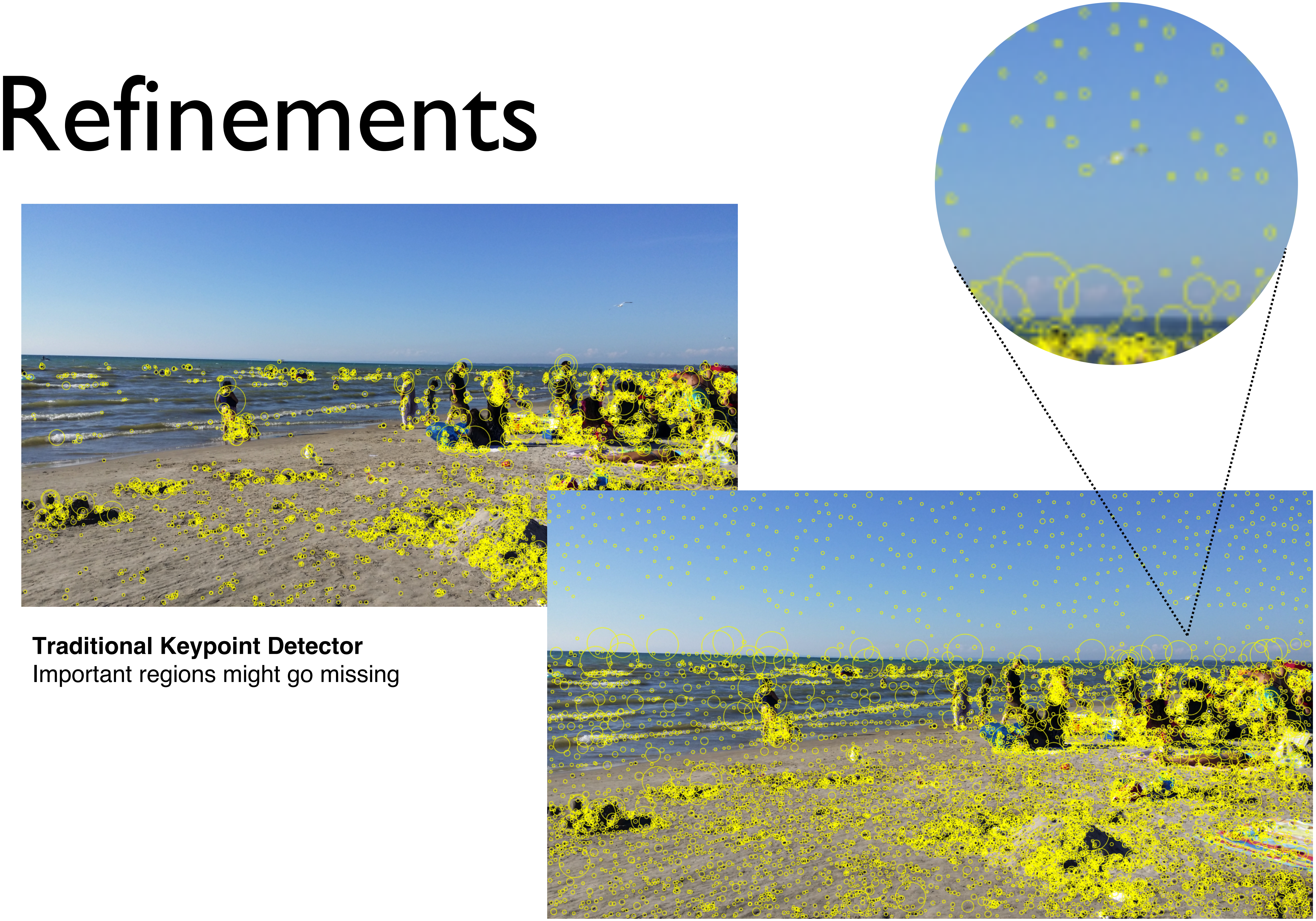
NEW

Results

Directed Graphs



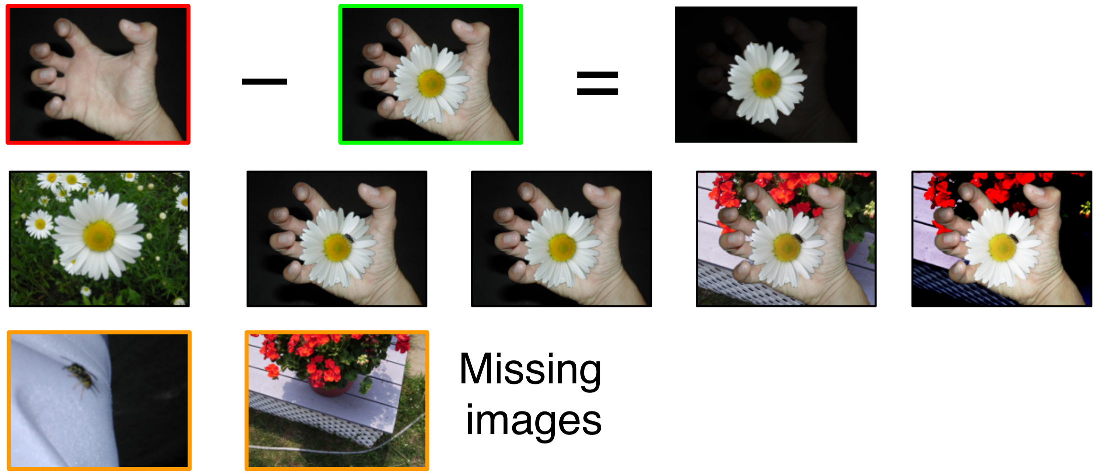
Refinements

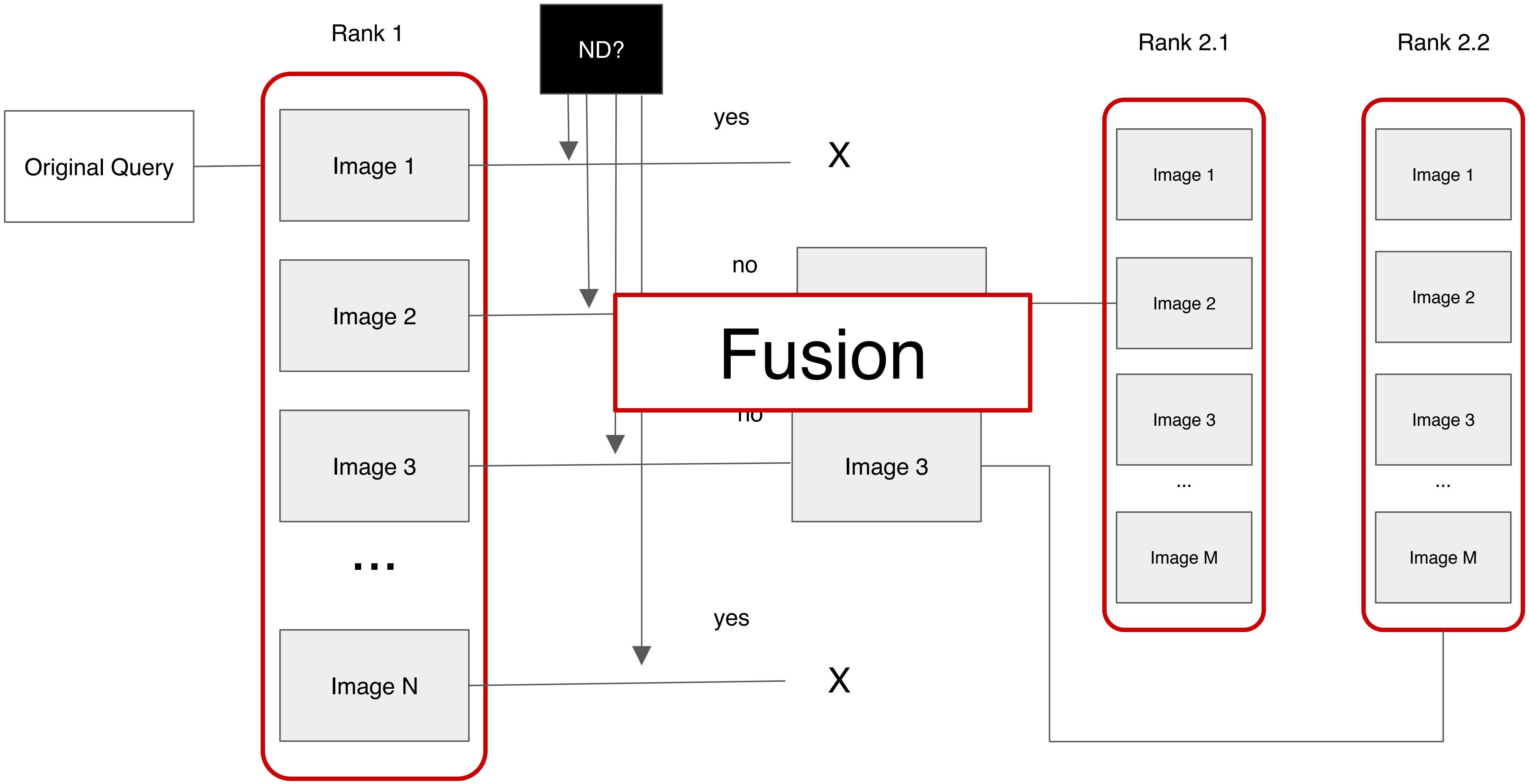


Traditional Keypoint Detector
Important regions might go missing

Refined detection with Collision Avoidance
Better balance of distinctive areas

Refinements

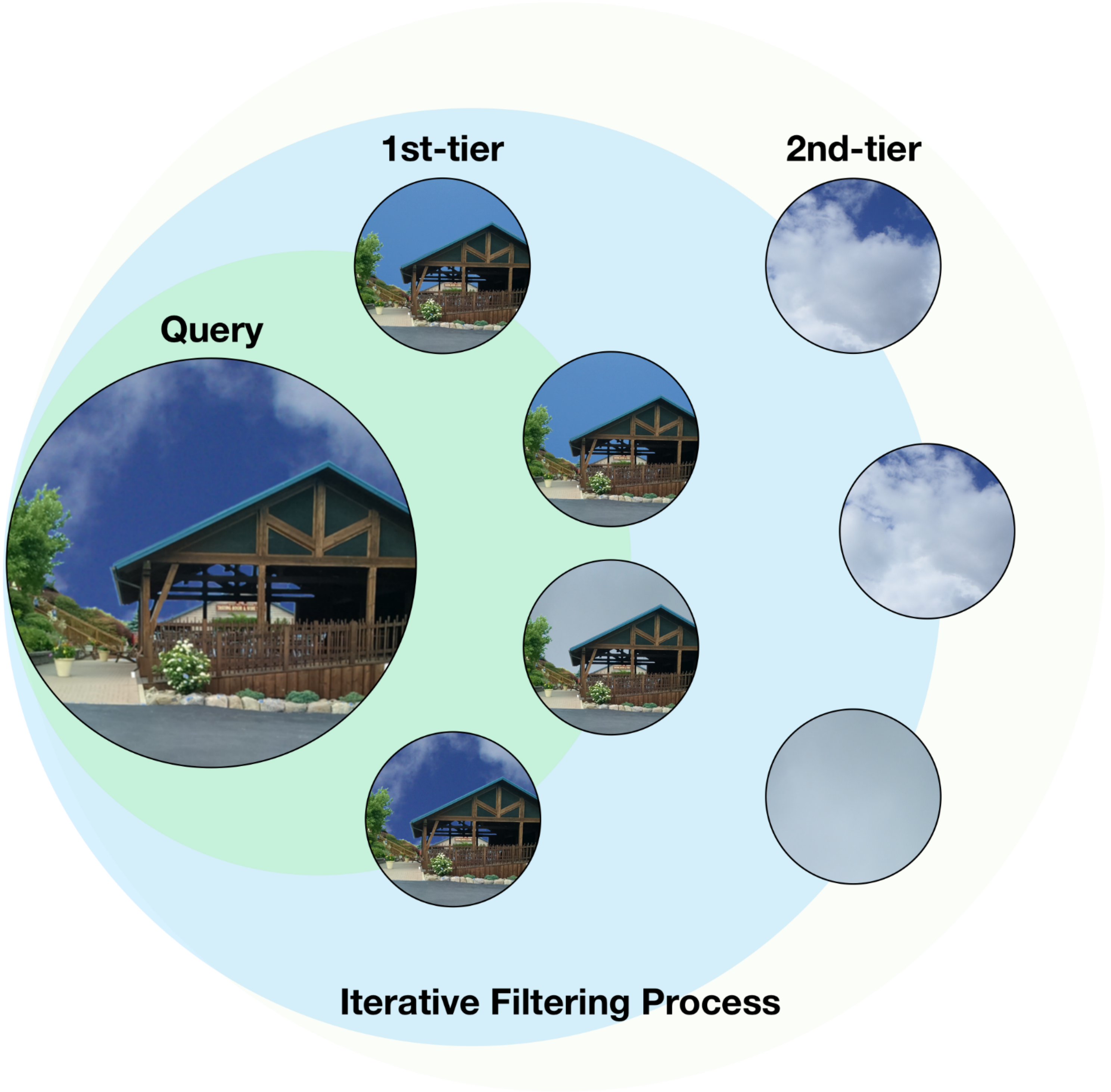




Refinements

to 92% in R@200

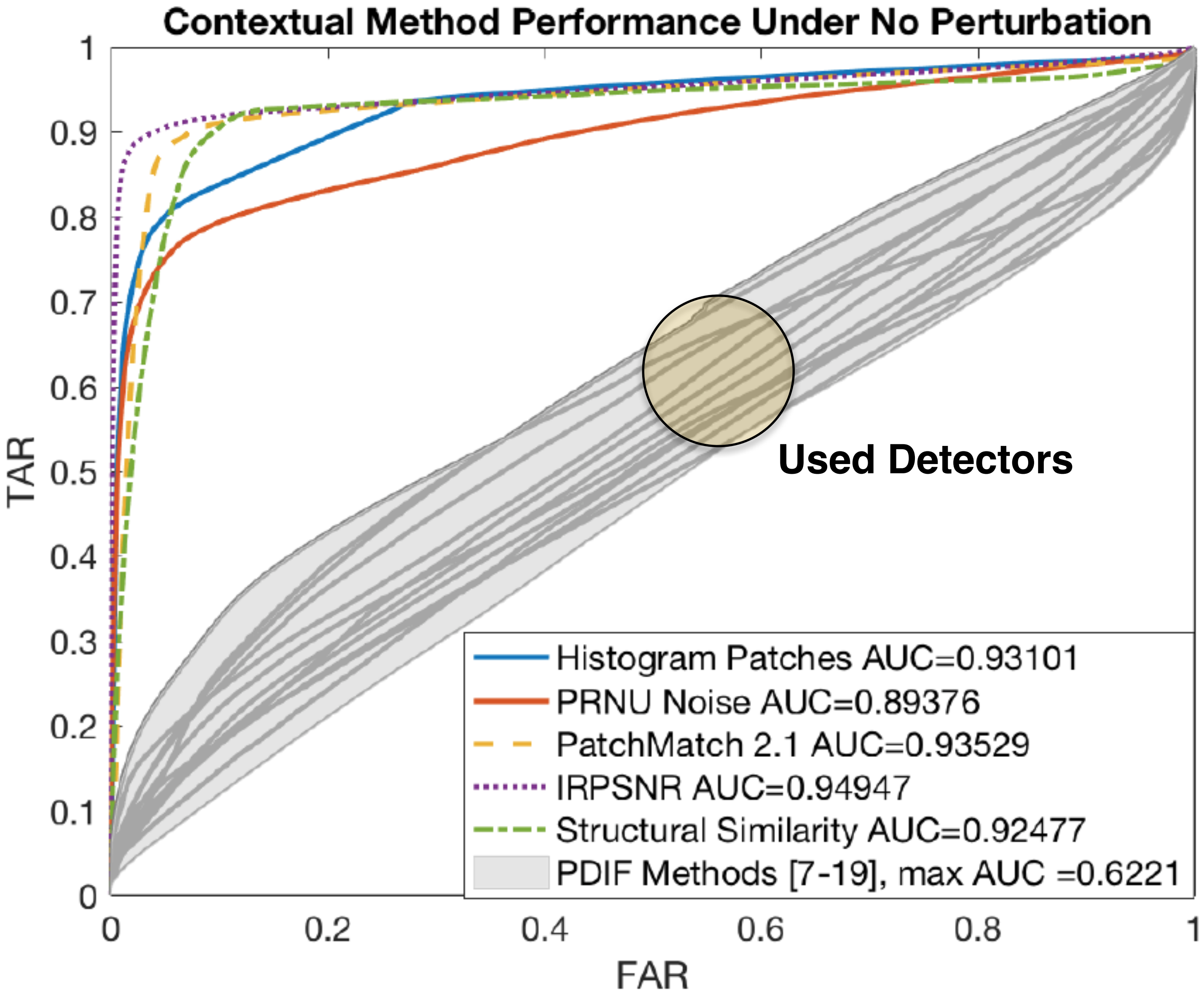
From 67%



Method	Recall@50	Recall@100	Recall@200
ICIP'17	0.6716	0.7157	0.7157
Dry-run	0.852	0.855	0.862
Iterative Filtering	**0.912**	**0.916**	**0.920**

Refinements

- Forgery detectors
- Context analysis



Significant Changes Ahead

- Scalability with GPU feature extraction & indexing
- Directionality inference from multiple cues (color, compression, illumination, mutual info, matching, etc.)
- **Context incorporation**
  - Side info (geo-tagging, date, etc.)
  - Editing/manipulation info
  - Manipulation detectors

Thank you

Media Integrity Analytics

Beyond Digital Forensics of Single Objects

Anderson Rocha (Associate Professor)

Microsoft Research Faculty Fellow
Google Faculty Research Awardee
Tan Chin Tuan Fellow
IEEE Senior Member

Reasoning for Complex Data (RECOD) Lab.
Institute of Computing,
University of Campinas (Unicamp)

Av. Albert Einstein, 1251 - Cidade Universitária
CEP 13083-970 • Campinas/SP - Brasil

** Joint work with several colleagues*
** Special thanks to S. Goldenstein and M. Oikawa for helping crafting this presentation*