Biometric Systems
Lesson 2bis – More on Performance

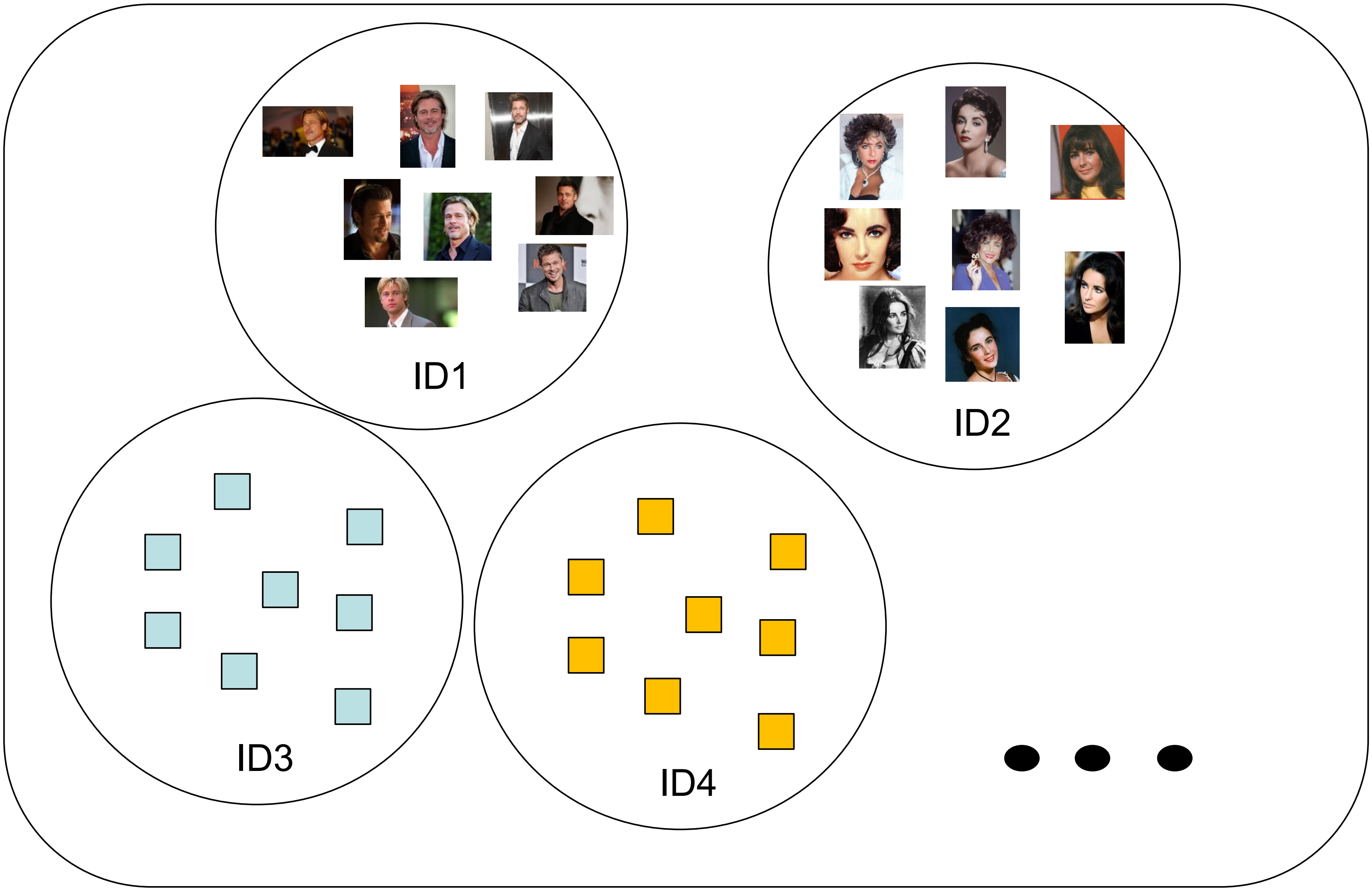
Maria De Marsico
demarsico@di.uniroma1.it

SAPIENZA Università di Roma
Dipartimento di Informatica

Offline performance evaluation

- **IMPORTANT NOTE**: the statistical (offline) analysis is carried out on suitable «static» datasets, collected to evaluate approaches to a problem, and entails a **ground truth**
- All samples are **labeled,** so that **we know** their real identities
- This is **not true in real operations**, where it could not be always possible to check the real identity of a probe, that's why it is important to have a reliable offline testing of the system accuracy to evaluate its reliability
- During performance evaluation, each sample presented **as a probe** may play the role of either a genuine sample or of an impostor one, **according to the gallery setup** for the experiment at hand and on the **possible identity claim** attributed to the probe if in verification mode

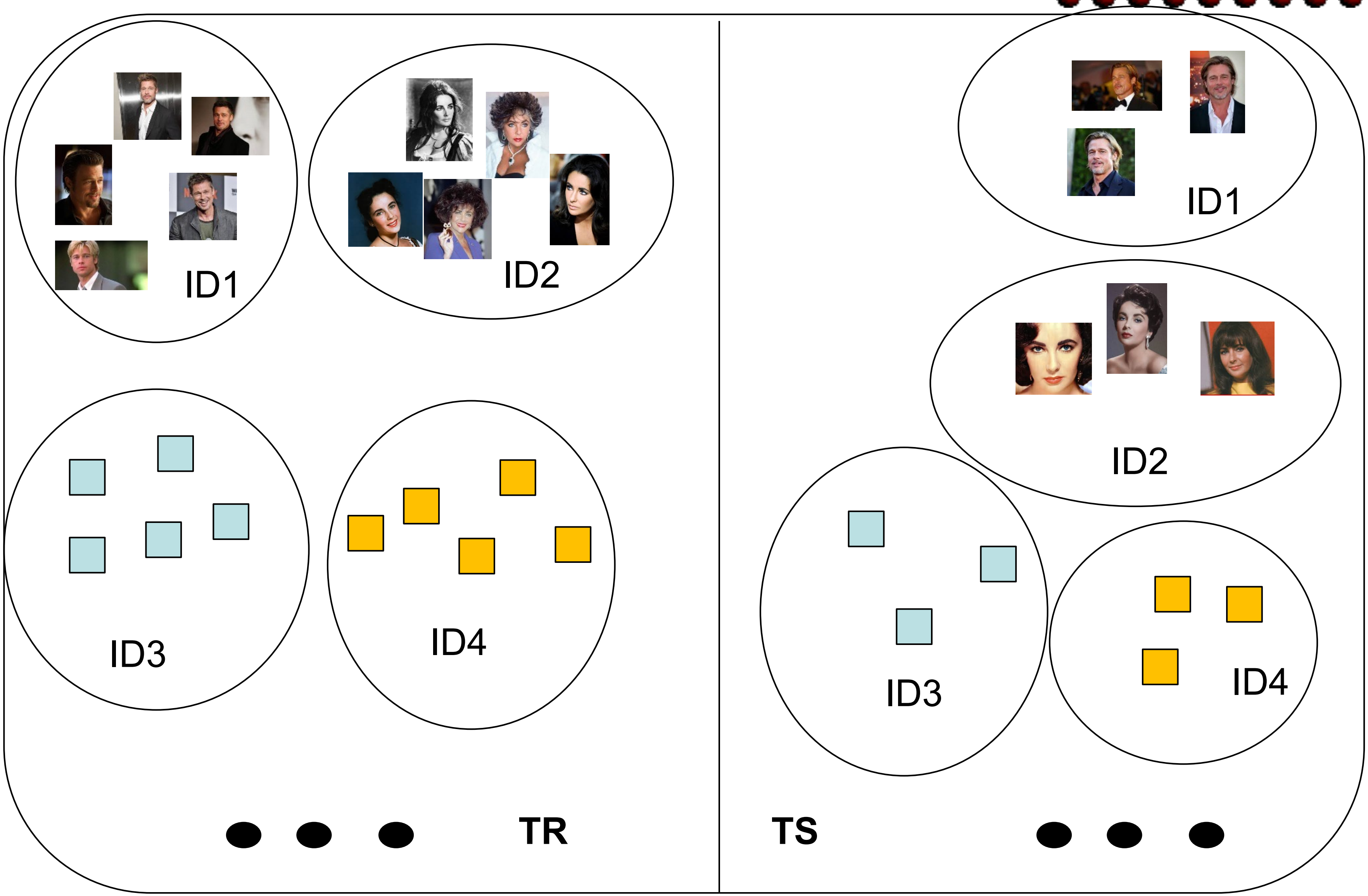
Example Dataset



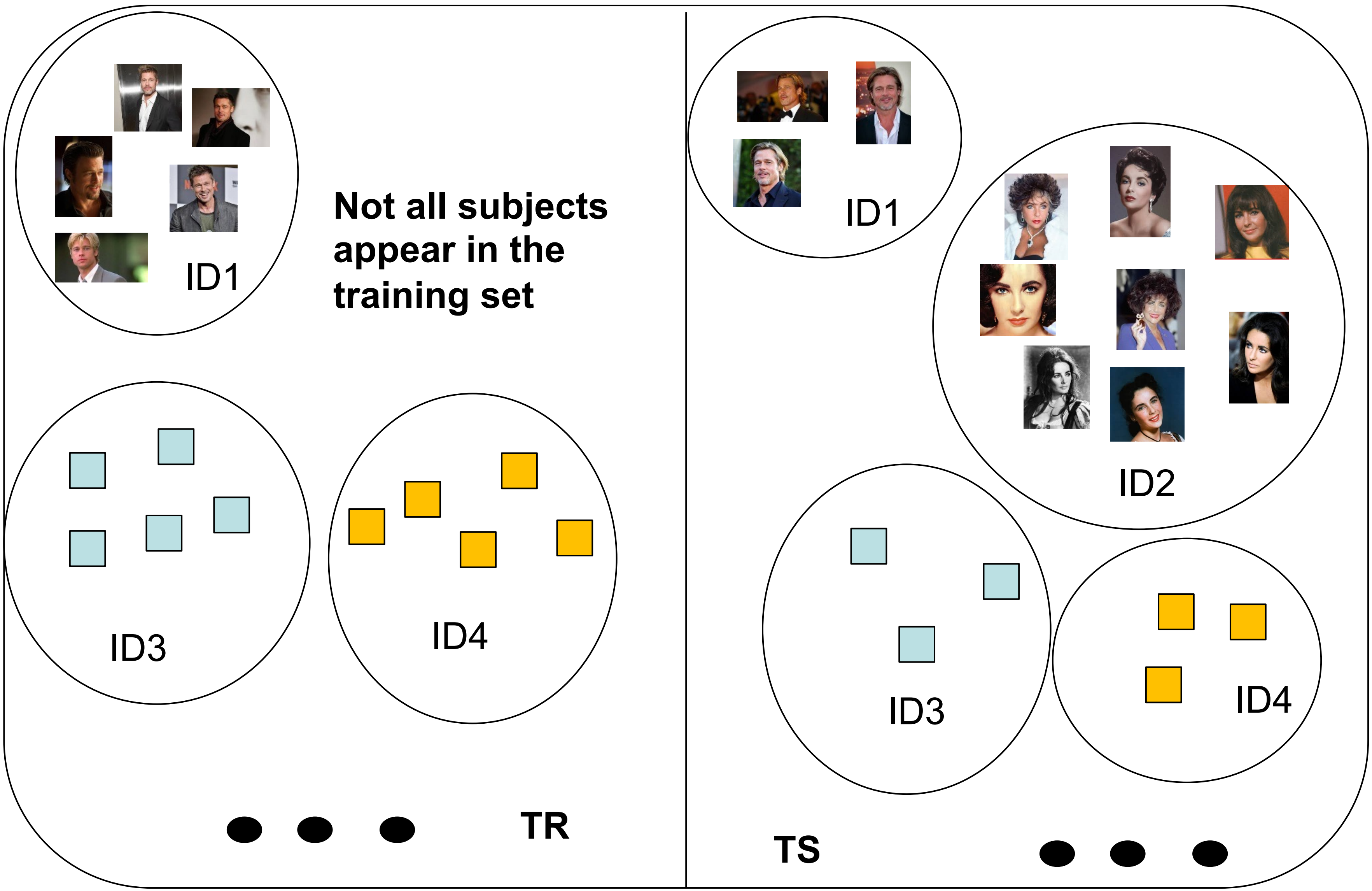
Dataset organization

- First choice: training vs testing (TR vs TS)
  - **Example**: in order to assure generalization, templates of different quality must be in cluded in the training set (those taking to non obvious recognition)
  - **Generalizability of the training outcomes depends on the choice of the training set**
  - **Some datasets suggest this division in order to assure fair comparison of methods**
  - **Recognition methods may not require training in the strict sense (machine learning) but this can be used to set up parameters for testing**
  - **Choice based on both subjects (a subject may not belong to the training set, to better test generalizability) and on samples (no overlap between TR and TS allowed!)**

TR vs TS 1



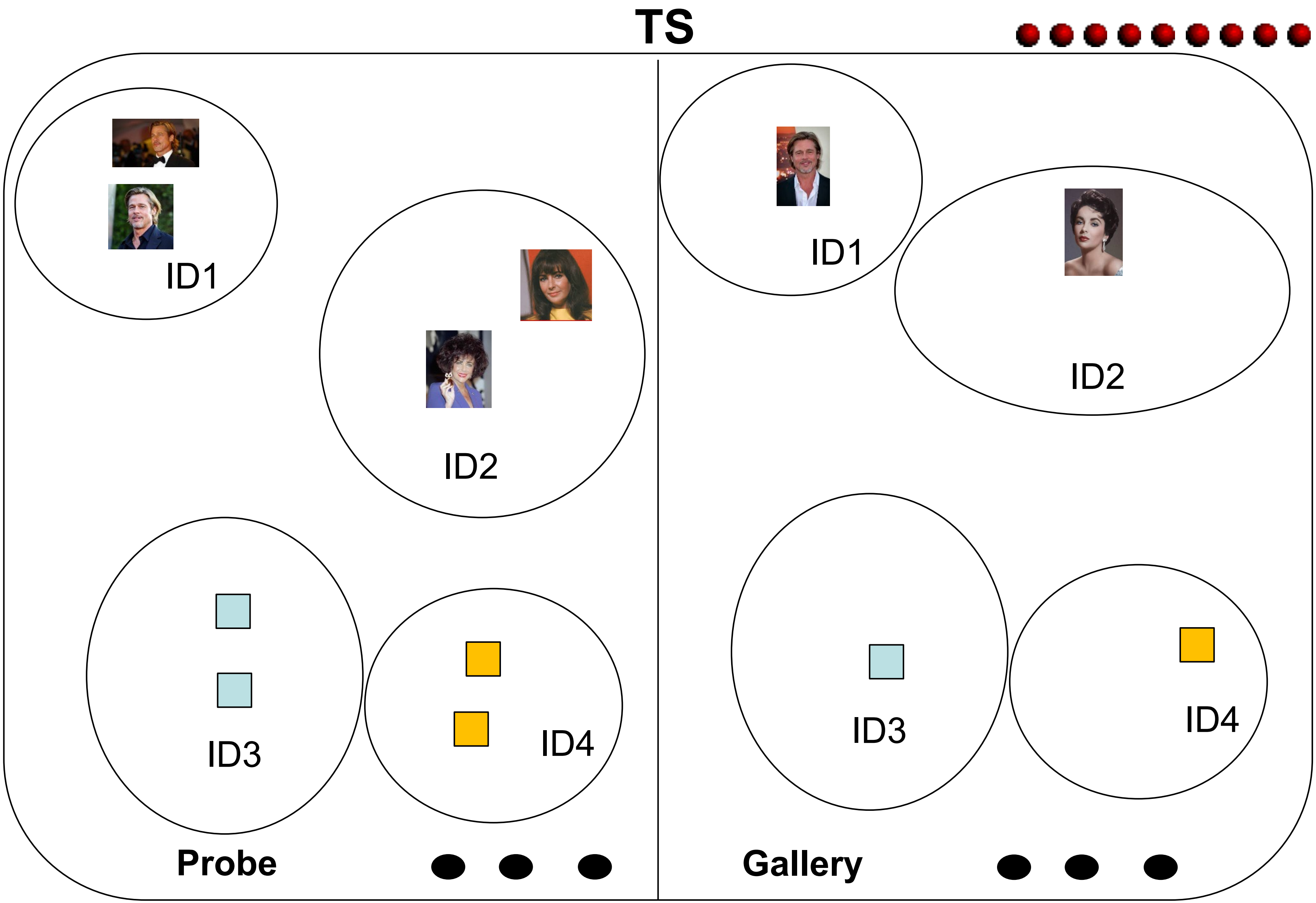
TR vs TS 2


ID1
Not all subjects appear in the training set
ID3
ID4
TR
ID1
ID2
ID3
ID4
TS

Dataset organization

- Second choice: probe set vs gallery set (P vs G)
  - **Example**: best quality templates in the gallery because enrollment is usually carried out in controlled conditions
  - **Other choices are possible**
  - **Choice based on samples (no overlap between P and G is allowed!)**
  - **Used for testing**

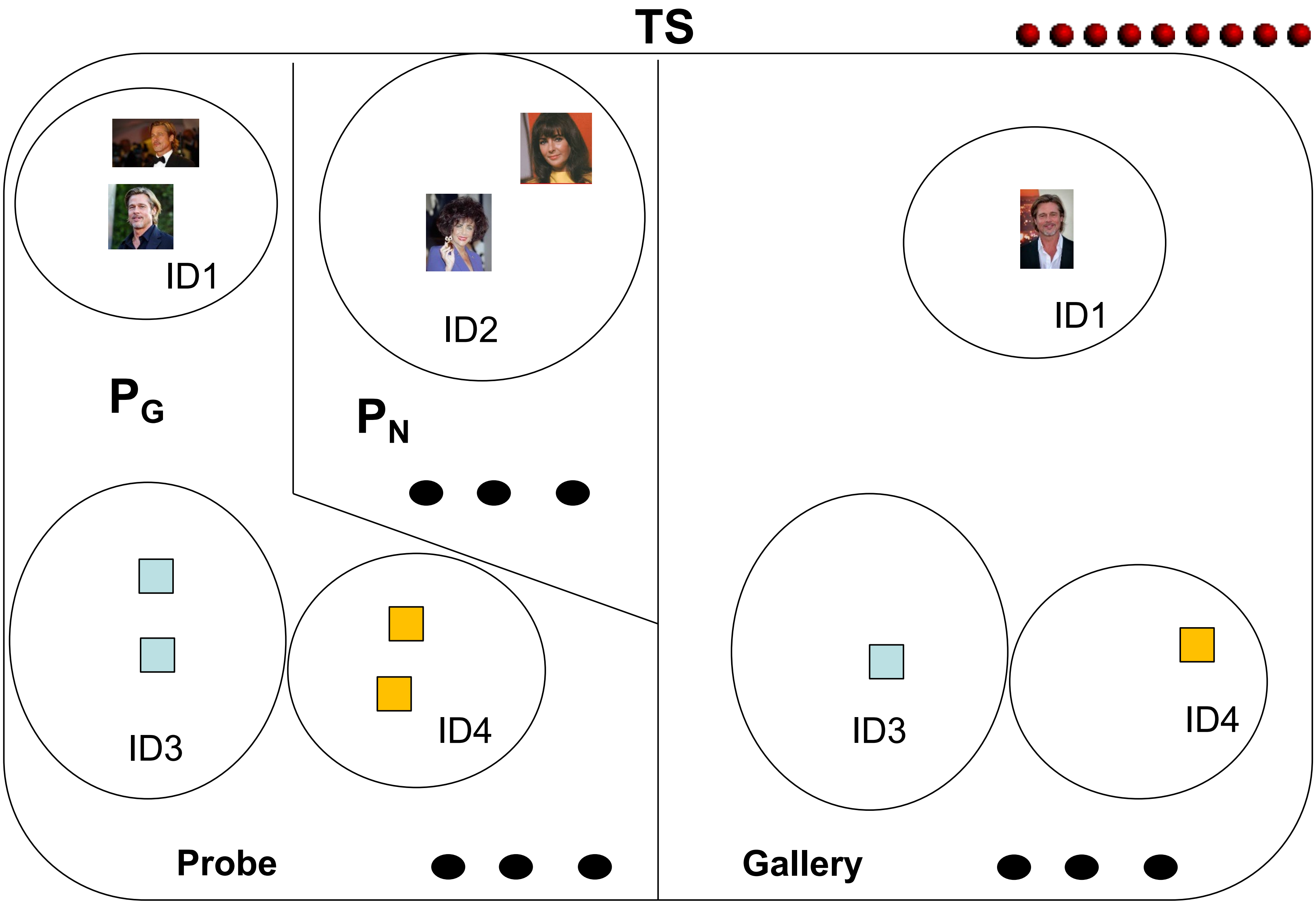
Probe vs Gallery



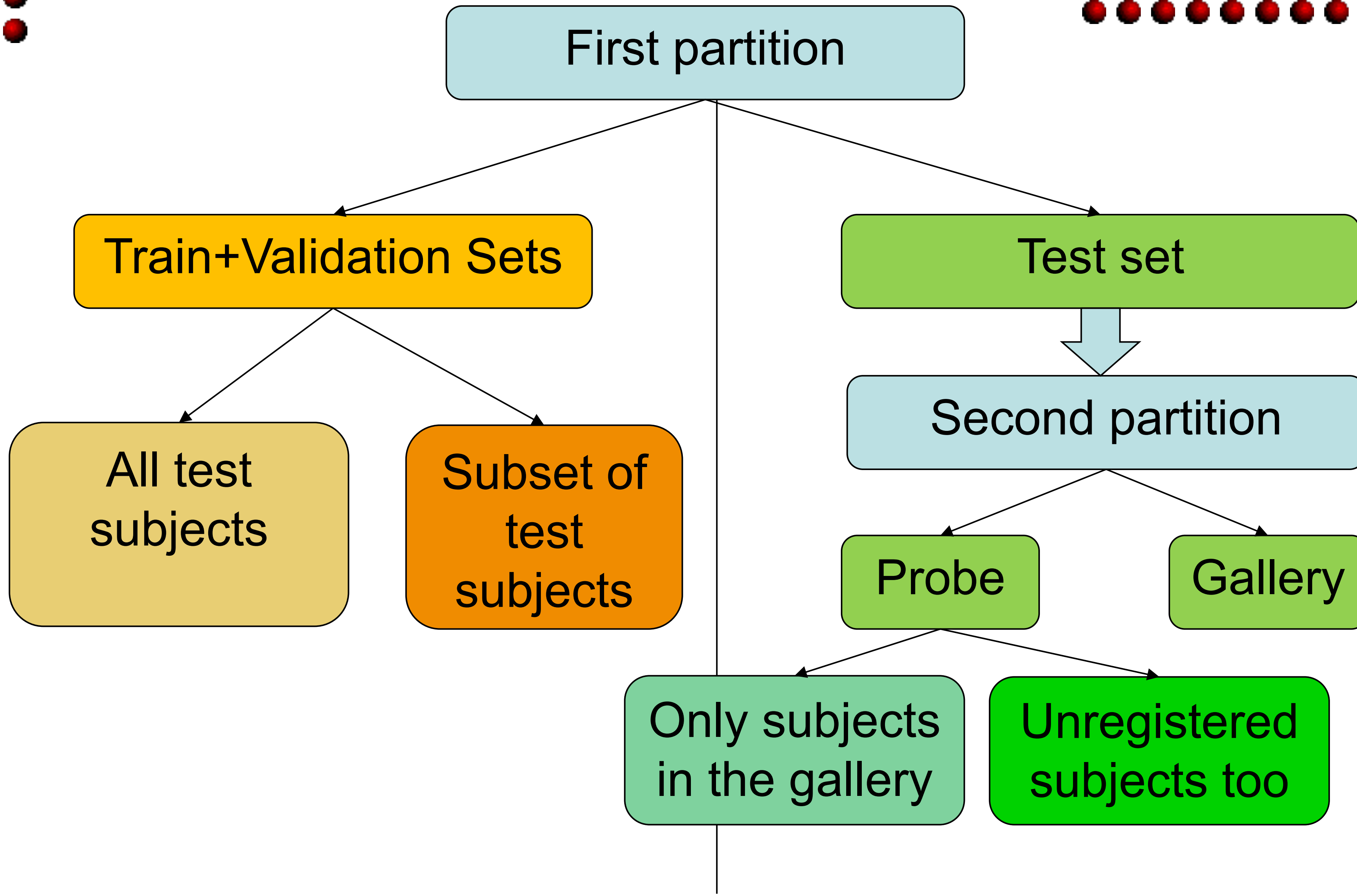
Dataset organization

- Third choice: templates in the probe set always from subjects in the gallery ($P = P_G$) vs probe sets including templates belonging to subjects not in the gallery ($P = P_G \cup P_N$)
  - **Verification:** this choice does not affect results
  - **Identification closed set:** this choice is not possible
  - **Identification open set:** this choice may influence the results according to the ratio among the number of subjects in the gallery (e.g., black or white lists may contain few subjects) and the number of total probes (e.g., all the passengers accessing an airport)
- **Choice is based on subjects**

Probe (P_G + P_N) vs Gallery



Main branches of experimental setup



Validation

- In principle, especially when training is involved, we should partition the dataset in different ways, repeat the evaluation, and take average performance, in order to avoid the bias due to a specific choice of partition elements.

Validation example

- In **K Fold cross validation**, the data is divided into k subsets. Now the holdout method is repeated k times, such that **each time, one of the k subsets is used as the test set/ validation set and the other k-1 subsets are put together to form a training set**. The error estimation is averaged over all k trials to get total effectiveness of our model. As can be seen, every data point gets to be in a validation set exactly once, and gets to be in a training set k-1 times. **This significantly reduces bias as we are using most of the data for fitting, and also significantly reduces variance as most of the data is also being used in validation set.** Interchanging the training and test sets also adds to the effectiveness of this method. **As a general rule and empirical evidence, K = 5 or 10 is generally preferred**, but nothing's fixed and it can take any value.

Probe vs All Gallery

- The validation is necessary when partitioning the dataset in different ways.
- In the following we will consider distance measures (e.g., Euclidean) since similarity entails exactly symmetric considerations.
- For each probe/gallery pair, it is possible to compute **beforehand** a PROBE-against-$\text{ALL}_{\text{GALLERY}}$ **distance matrix**, storing **all distances between pairs of templates** (a probe template vs. a gallery template).
- The distance matrix can be used for performance evaluation of all kinds of applications (verification, identification closed set, identification open set, considering both single and multiple templates per subject in the gallery).
- **Each row corresponds to a recognition operation on an incoming probe, either with (verification) or without (identification) a claim**.
- **Each probe/gallery partition produces a different distance matrix.**
- To allow a better generalization, **the distance matrices used in the following examples will not contain the numeric values but only their ascending order.**

Verification

- Only gallery templates belonging to the claimed identity are matched against the probe.
- It is not important who is in the gallery, but the claimed identity.
- Each row is labeled with the ground truth probe identity and, for verification, with the claimed identity
- In genuine matches the probe is associated with the claims of the correct identity, in impostor matches the probe is associated with the claim of a false identity (wether the subject is in the gallery or not)
- Performance evaluation can be obtained by clearly separating probe and gallery (different subsets of templates) and genuine and impostors (templates that play the role of impostors are associated with the claim of a different identity).

Example: identities A, B, C, D, E, F in the **dataset**
identities A, B, C, D in the **gallery** with a single istance
identities E, F play the role of **impostors in all cases**
d() = distance form the probe t = acceptance threshold

	Probes	A1	B1	C1	D1
ID A – Claim A	P1	1	4	2	3
ID D – Claim C	P2	4	1	3	2
ID E – Claim D	P3	4	2	1	3
ID C – Claim C	P4	…	…	…	…
ID F – Claim B	P5	…	…	…	…

- **Example** of distance matrix (only **order** of values is shown)

Each row is **a single** verification operation

P1 genuine
Cases: d(A1) <= t GA++
d(A1) > t FR ++

P2 impostor (the subject is in the gallery but claims a different identity)
Cases: d(C1) <= t FA++
d(C1) > t GR++

P3 impostor (subject not in the gallery)
Cases: d(D1) <= t FA++
d(D1) > t GR++

Example: identities A, B, C, D, E, F in the **dataset**
identities A, B, C, D in the **gallery** with multiple istances (choose best result)
identities E, F play the role of **impostors in all cases**
d() = distance form the probe t = acceptance threshold

	Probes	A1	A2	B1	B2	C1	C2	D1	D2
ID A – Claim A	P1	2	1	8	4	3	7	5	6
ID D – Claim C	P2	7	6	2	1	5	8	3	4
ID E – Claim D	P3	7	5	2	8	6	1	3	4
ID C – Claim C	P4	…	…	…	…	…	…	…	…
ID F – Claim B	P5	…	…	…	…	…	…	…	…

Example of distance matrix (only **order** of values is shown)

Each row is **a single** verification operation

P1 genuine – best match for claim A is A2
Cases: d(A2) <= t GA++
d(A2) > t FR ++

P2 impostor (the subject is in the gallery but claims a different identity) –
best match for claim C is C1
Cases: d(C1) <= t FA++
d(C1) > t GR++

P3 impostor (subject not in the gallery) – best match for claim D is D1
Cases: d(D1) <= t FA++
d(D1) > t GR++

Verification

- Having more samples per subjects decreases FRR, but may also increase FAR (more possibilities for an impostor to look similar to a genuine)

Verification

- This approach requires to carry out a sufficient number of evaluations in order to compute a reliable average result:
  - each time probe set and gallery sets are choosen in a different way,
  - each time there is a possibly different distribution of genuine probes (the claimed identity is the true one) and impostor probes (the claimed identity of the probe is not the true one, wheter the probe belongs to a gallery subject or not), i.e., how many genuine and how many impostors are considered.

- The choice of genuine and impostors can influence the outcome (we will see the Doddington zoo)

Identification – Open Set

- The probe to identify might not belong to a subject included in the gallery.
- All gallery templates are matched against the probe.
- It is important who is in the gallery, because there is no identity claim.
- Each row is labeled with the probe ground truth identity only (no claim).
- Genuine probes are those belonging to identities in the gallery, impostor probes are those belonging to identities not in the gallery.
- Performance evaluation can be obtained by clearly separating probe and gallery (different subsets of templates) and genuine and impostors (identities that play the role of impostors are not included in the gallery).

Example: identities A, B, C, D, E, F in the **dataset**
identities A, B, C, D in the **gallery** with a single istance
identities E, F play the role of **impostors**
d() = distance form the probe t = acceptance threshold

Example of distance matrix (only **order** of values is shown)

Each row is **a single** identification operation

	Probes	A1	B1	C1	D1
A	P1	1	4	2	3
D	P2	4	1	3	2
E - Impostor	P3	4	2	1	3
C	P4	…	…	…	…
F - Impostor	P5	…	…	…	…

Ordered list of distances for P1 = **d(A1)**, d(C1), d(D1), d(B1)

Cases: d(A1) <= t DI(1,t)++
d(A1) > t FR ++ (the rate will be derived from DIR(1,t) - no recording)

Ordered list of distances for P2 = d(**B1**), d(D1), d(C1), d(A1)

Cases: d(B1) <= t FR++ (the rate will be derived from DIR(1,t) - no recording)
if d(D1) <= t there is a contribution to DI(2,t), checked only if d(B1) <= t
d(B1) > t FR++ (the rate will be derived from DIR(1,t) - no recording)

Ordered list of distances for P3 = **d(C1)**, d(B1), d(D1), d(A1)

Cases: d(C1) <= t FA++
d(C1) > t GR++

Example: identities A, B, C, D, E, F in the **dataset**
identities A, B, C, D in the **gallery** with multiple istances
identities E, F play the role of **impostors**
d() = distance form the probe t = acceptance threshold

Example of distance matrix (only **order** of values is shown)

Each row is **a single** identification operation

	Probes	A1	A2	B1	B2	C1	C2	D1	D2
A	P1	2	1	8	4	3	7	5	6
D	P2	7	6	2	1	5	8	3	4
E - Impostor	P3	7	5	2	8	6	1	3	4
C	P4	…	…	…	…	…	…	…	…
F - Impostor	P5	…	…	…	…	…	…	…	…

Ordered list of distances for P1 = **d(A2)**, d(A1), d(C1), d(B2), d(D1), d(D2), d(C2), d(B1)
Cases: d(A2) <= t DI(1,t)++ (compare with previous example, A2 **closer**, more possibilities for correct recognition)
d(A2) > t FR ++ (the rate will be derived from DIR(1,t) - no recording)

Ordered list of distances for P2 = **d(B2)**, d(B1), d(D1), d(D2), d(C1), d(A2), d(A1), d(C2)
Cases: d(B2) <= t FR++ (the rate will be derived from DIR(1,t) - no recording)
→ if d(D1) <= t there is a contribution from DI(3,t) up, checked only if d(B2) <= t
d(B2) > t FR++ (the rate will be derived from DIR(1,t) - no recording)

Ordered list of distances for P3 = **d(C2)**, d(B1), d(D1), d(D2), d(A2), d(C1), d(A1), d(B2)
Cases: d(C2) <= t FA++
d(C2) > t GR++

Identification – Open Set

- This approach requires to carry out a sufficient number of evaluations in order to compute a reliable average result:
    - each time probe set and gallery sets are choosen in a different way,
    - each time there is a possibly different distribution of genuine probes (the identity of the probe is also in the gallery) and impostor probes (the identity of the probe has nt been included in the gallery), i.e., how many genuine and how many impostors are considered.

- The choice of genuine and impostors can influence the outcome (we will see the Doddington zoo)

Identification – Closed Set

- The probe to identify always belongs to a subject included in the gallery.
- All gallery templates are matched against the probe.
- Each rows is labeled with the probe ground truth identity (no claim).
- No impostor appears in the experiments. (**all identities are in the gallery**)
- **There is no acceptance threshold.**
- Performance evaluation can be obtained by clearly separating probe and gallery (different subsets of templates).

Example: identities A, B, C, D, E, F in the **dataset**
identities A, B, C, D, E, F in the **gallery** with a single istance
d() = distance form the probe

	Probes	A1	B1	C1	D1	E1	F1
A	P1	1	4	2	3	6	5
D	P2	6	1	4	2	3	5
E	P3	5	2	1	3	4	6
C	P4	…	…	…	…		
F	P5	…	…	…	…		

- **Example** of distance matrix (only **order** of values is shown)

Each row is **a single** identification operation

Ordered list of distances for P1 = **d(A1)**, d(C1), d(D1), d(B1), d(F1), d(E1)
The identity in the first place is the right one
The result contributes to CMS(1), i.e. to RR

Ordered list of distances for P2 = d(**B1**), d(D1), d(E1), d(C1), d(F1), d(A1)
The identity in the first place is not the right one
The correct identity is in the second place
The result contributes to CMS(2) and up

Ordered list of distances for P3 = d(**C1**), d(B1), d(D1), d(E1), d(A1), d(F1)
The identity in the first place is not the right one
The correct identity is in the fourth place
The result contributes to CMS(4) and up

Example: identities A, B, C, D, E, F in the **dataset**
identities A, B, C, D, E, F in the **gallery** with multiple istances
d() = distance form the probe

	Probes	**A1**	**A2**	**B1**	**B2**	**C1**	**C2**	**D1**	**D2**	**E1**	**E2**	**F1**	**F2**
A	P1	2	1	8	11	5	7	6	12	10	3	9	4
D	P2	11	3	2	7	8	10	5	1	6	12	9	4
E	P3	9	10	4	7	2	3	12	5	6	8	11	1
C	P4	…	…	…	…	…	…	…	…				
F	P5	…	…	…	…	…	…	…	…				

Example of distance matrix (only **order** of values is shown)

Each row is **a single** identification operation

Ordered distances for P1 = **d(A2)**, d(A1), d(E2), d(F2), d(C1), d(D1), d(C2), d(B1), d(F1), d(E1), d(B2), d(D2)

The identity in the first place is the right one
The result contributes to CMS(1), i.e. to RR

Ordered distances for P2 = **d(D2)**, d(B1), d(A2), d(F2), d(D1), d(E1), d(B2), d(C1), d(F1), d(C2), d(A1), d(E2)

The identity in the first place is the right one (the new template for D has improved the recognition)
The result contributes to CMS(1), i.e. to RR

Ordered distances for P3 = **d(F2)**, d(C1), d(C2), d(B1), d(D1), d(E1), d(B2), d(E2), d(A1), d(A2), d(F1), d(D2)

The identity in the first place is not the right one
The result contributes to CMS(6) (the recognition for E worsened due to the additional templates)

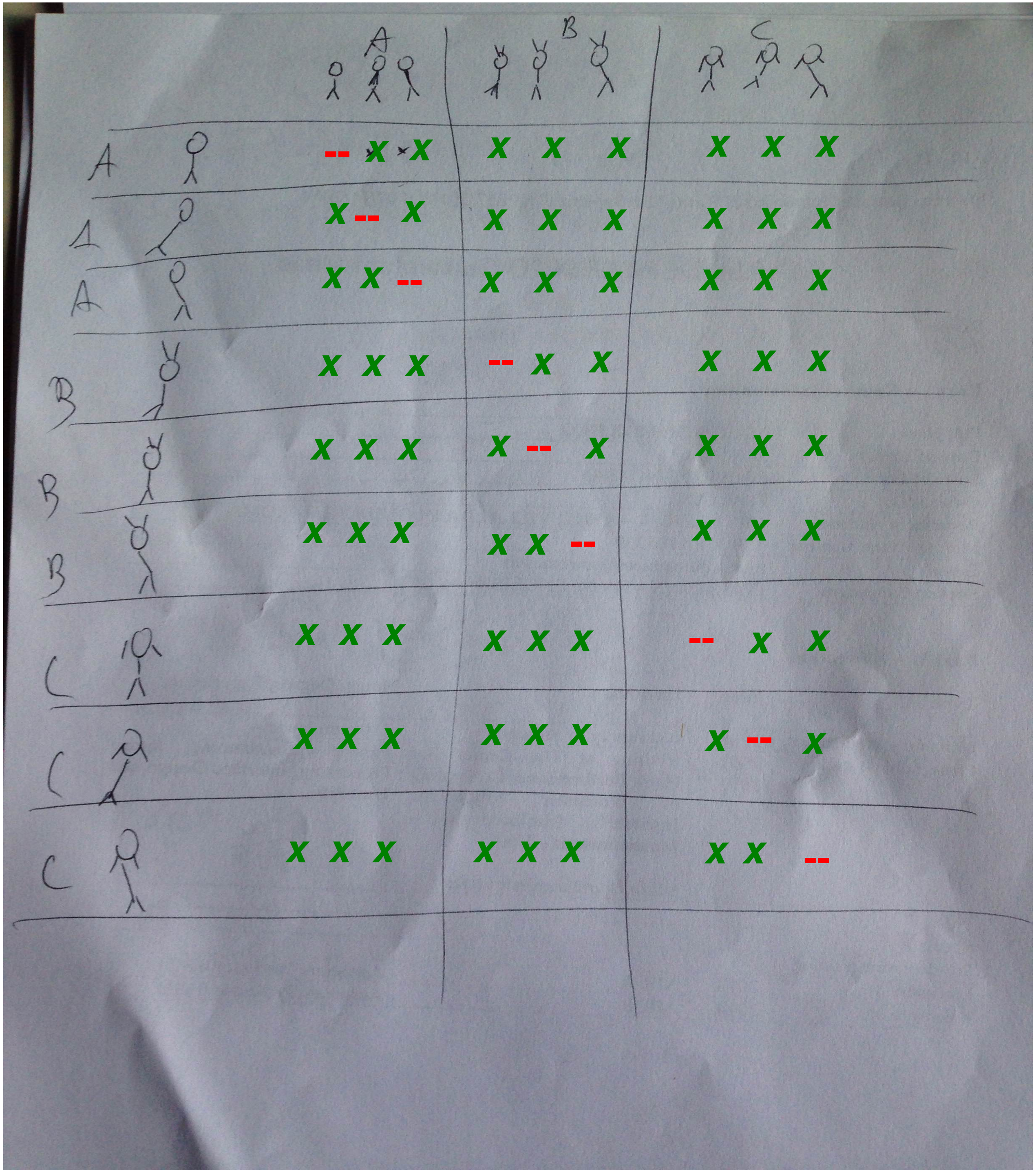
Identification – Closed Set

- This approach requires to carry out a sufficient number of evaluations in order to compute a reliable average result:

  - each time probe set and gallery sets are choosen in a different way

ALL-against-ALL

- It is possible to compute encompassing statistics using a **complete** ALL-against-ALL distance matrix (**all** templates are compared with **any other** template).
- Also in this case distances are computed once and for all, and they are used in a different way according to the setup (modality and number of gallery templates per subject)
- Each template plays in turn the role of either probe or gallery, possibly more than once (see the different cases).
- Diagonal values (comparisons of a a template with itself) are not considered.
- Cumulative averages (rates) that are computed encompass many possible partitions.



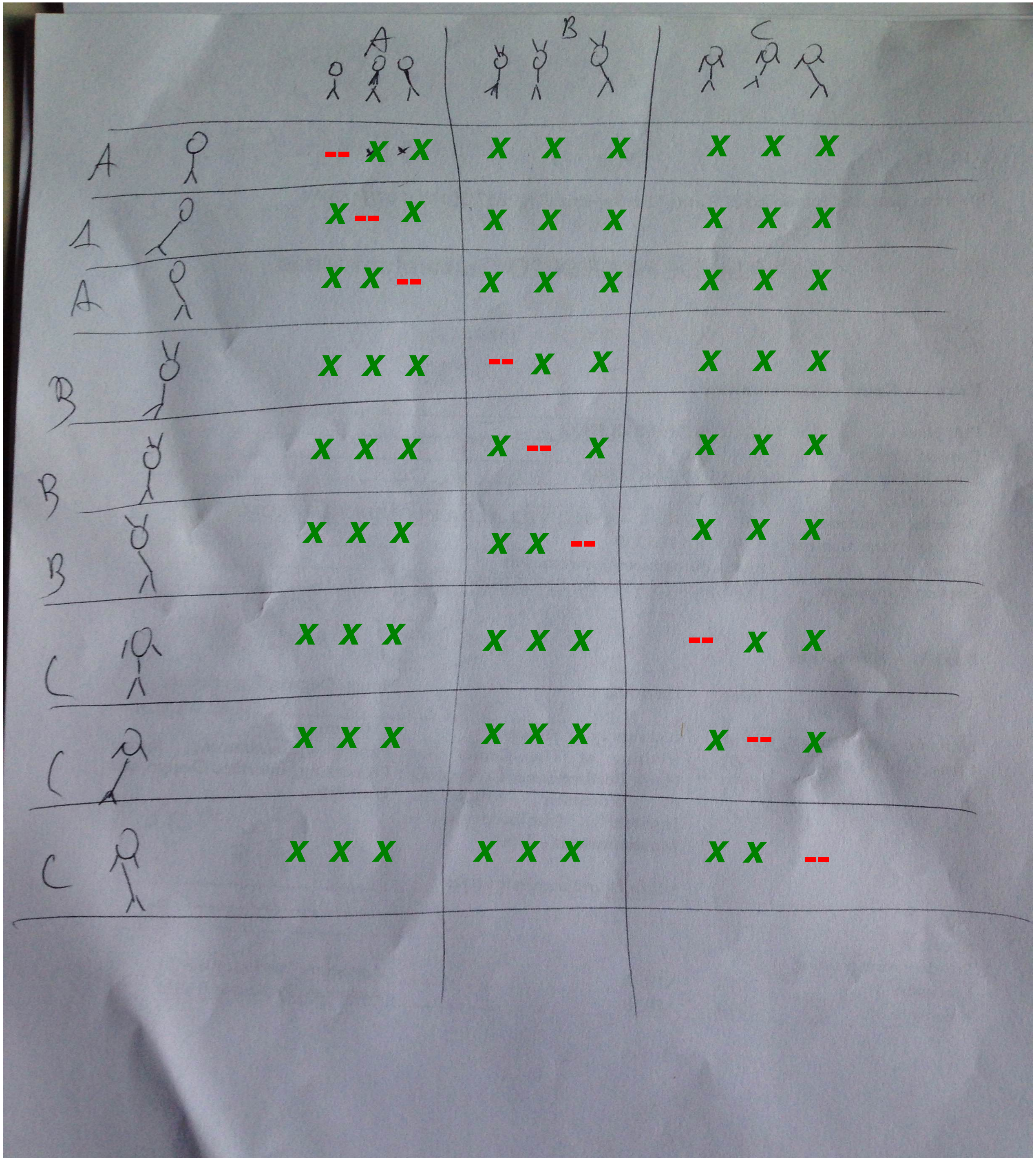
Valid matching operations

***x* = yes**

-- = no

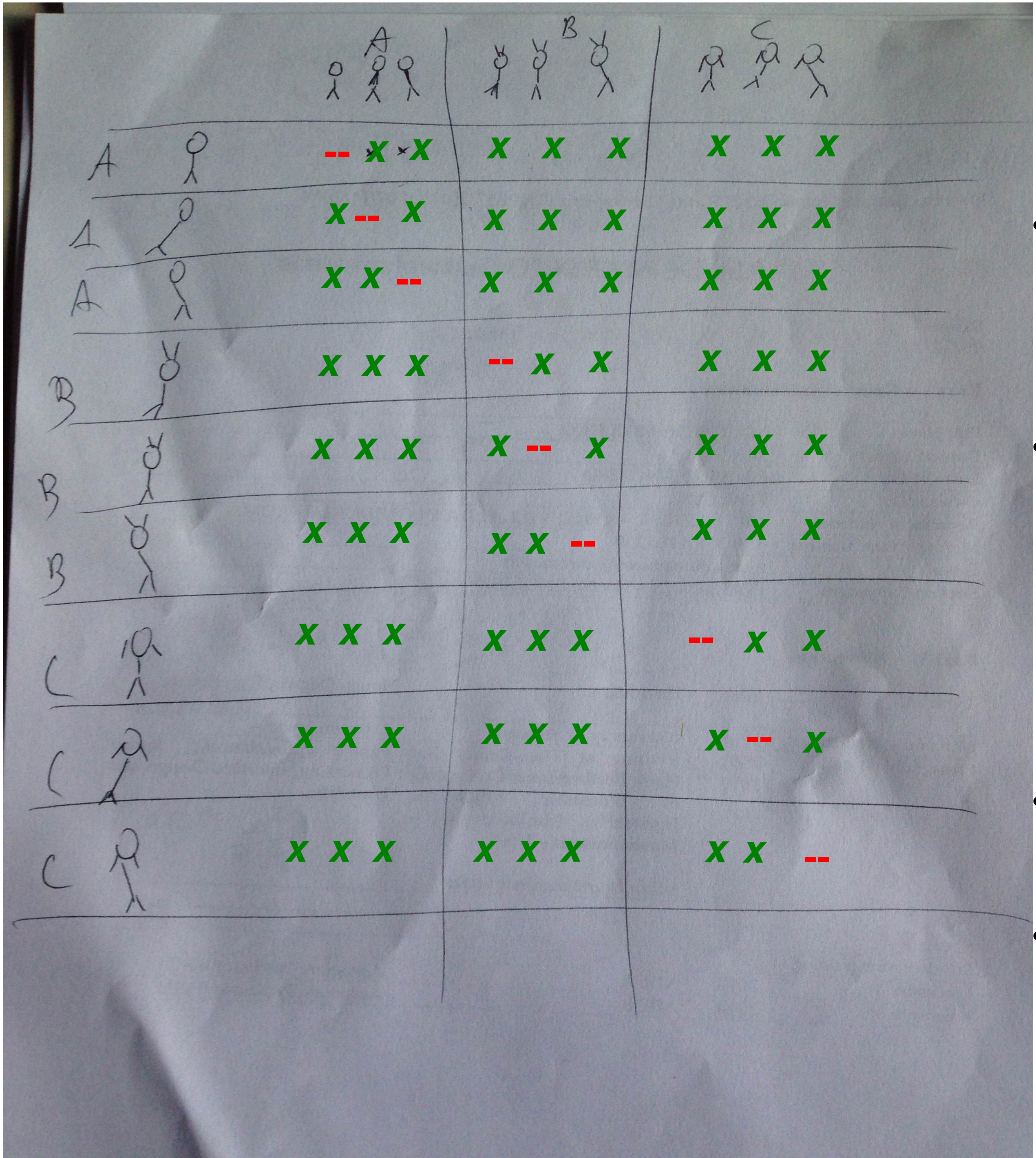
This holds **whatever** is the kind of test one is performing

The distance/similarity matrix **M** can be computed once and for all, and used in a different way depending on the kind of recognition modality one wants to test



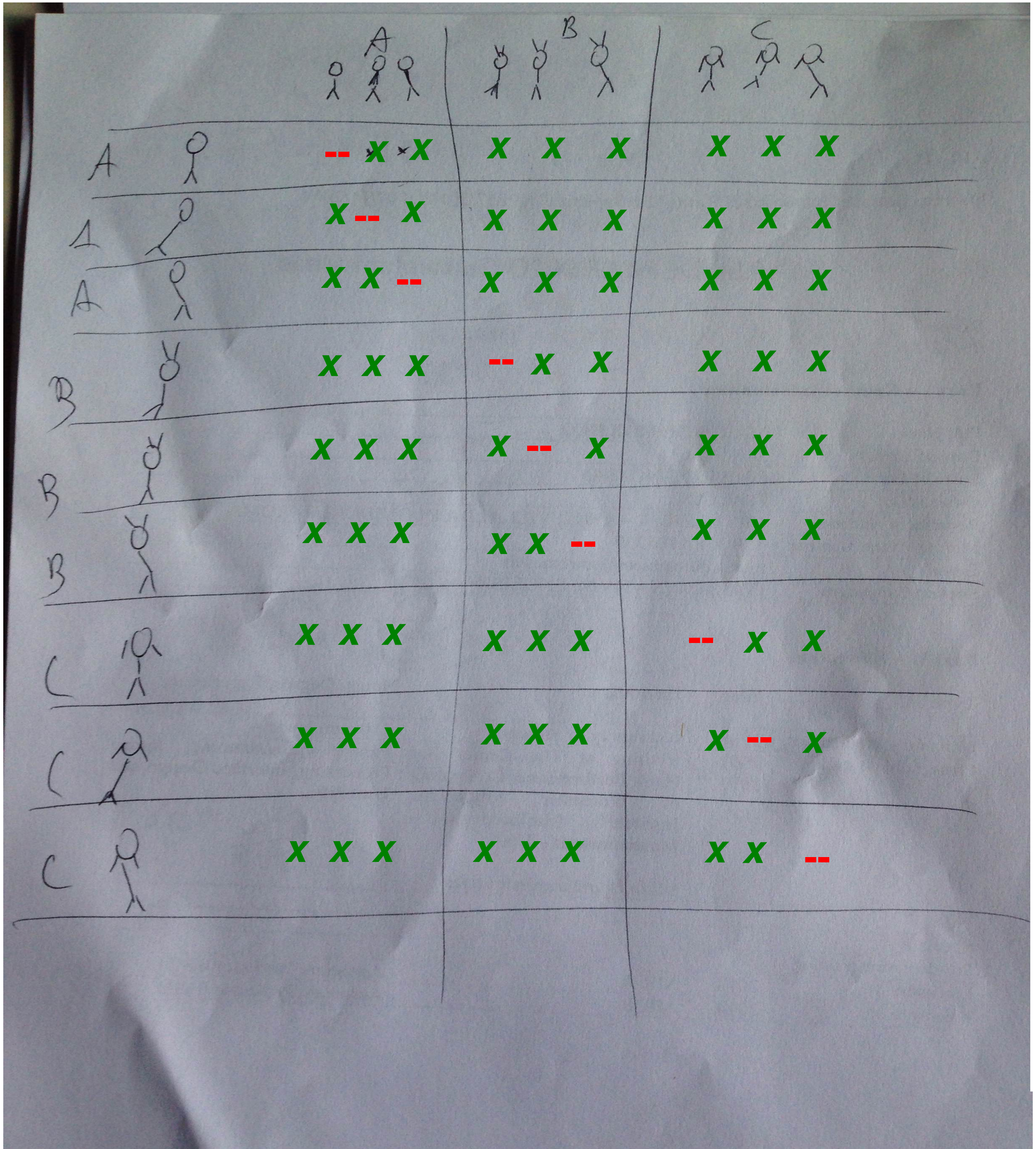
A possible strategy:
a kind of **all-against-all** computation which is suitably modified according to the recognition modality.

- Each probe (rows) is considered either as genuine or impostor in turn.
- Results are accumulated to get the final statistics.
- Each row possibly contributes more times according to the number of experiments it can possibly represent.
- For simplicity, it is possible to assume the same number of samples per subject. Extensions are straightforward.



- **Pros**: easy to program, in practice computes a kind of «average» over all possible specific distributions of genuine/impostor attempts.
- **Pros:** The number of impostors is always much higher than genuine, therefore it is possible to overstress the system to assess situations with many impostor attempts.

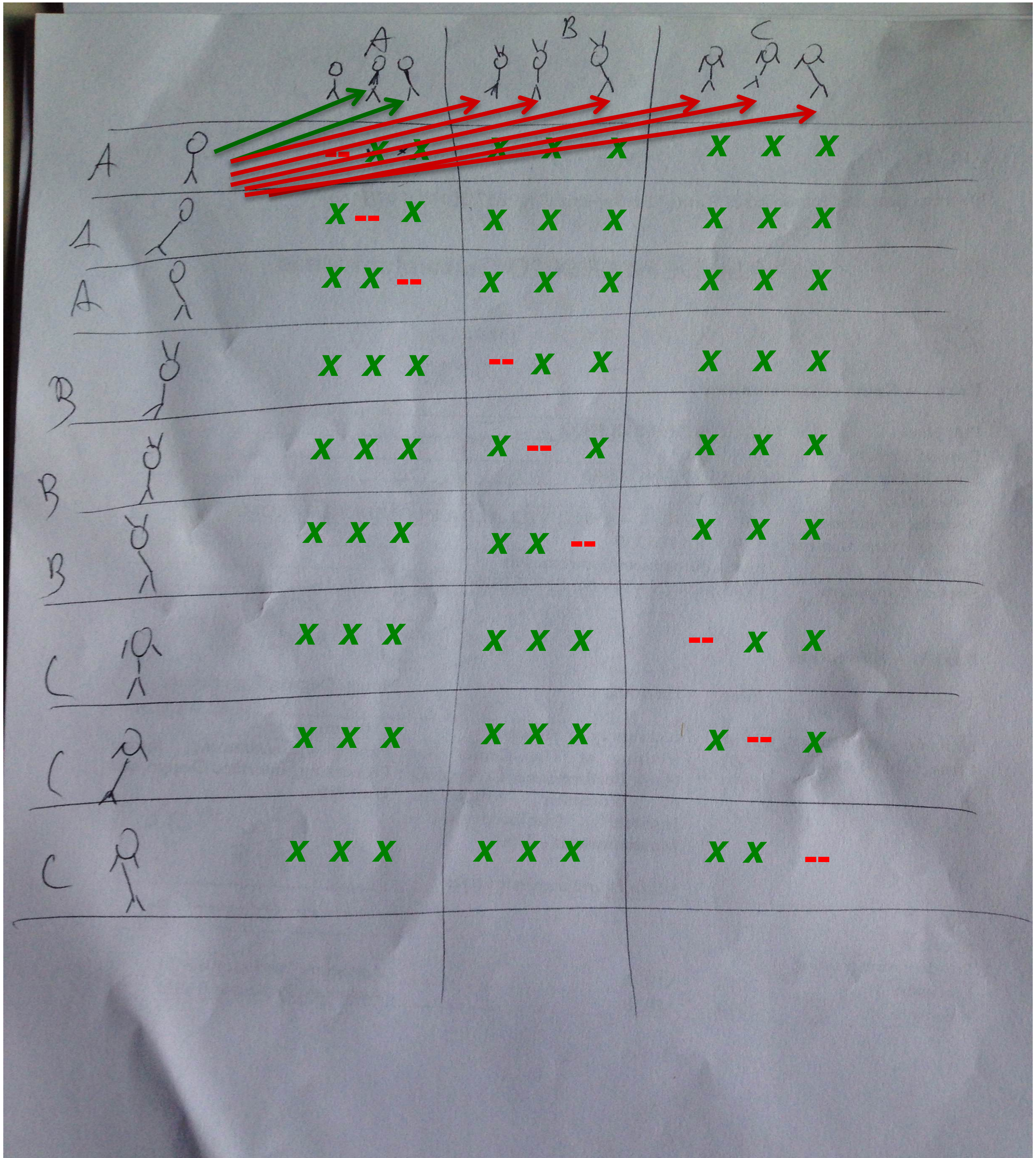
- **Cons**: the time requested to compute the full matrix may become very high.
- **Cons**: does not to analyze specific distributions of genuine/impostor attempts, that may achieve peculiar results.



- Each row is either **a single** test operation **or a set of tests**, depending on the recognition modality.

- We will further distinguish between **single-template** and **multiple-template** settings, depending on the **number of samples per subject** that are assumed to be stored in the test gallery.

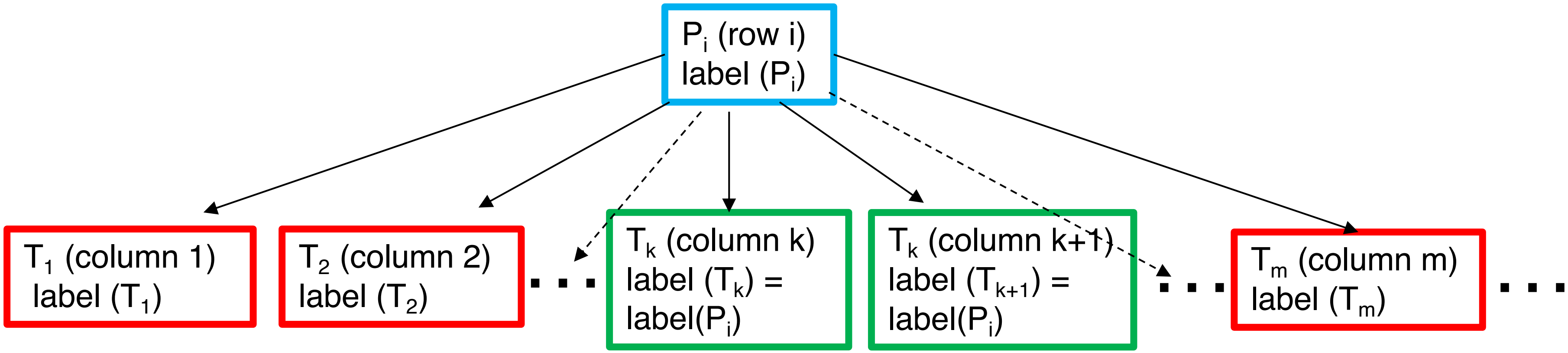
- N =number of **subjects**
- $|G|$ = total number of **samples** = number of rows/columns in the matrix
- S = number of **templates per subject** ($|G| = S \times N$)
- i = **row index** (**probe** template) and **label(i)** is the associated identity
- j = **column index** (**gallery** template) and **label(j)** is the associated identity
- label(·) = **ground truth**



- N = number of subjects
- |G| = total number of samples
- S = number of templates per subject (|G|=S×N)
- M = distance matrix

Verification single-template

Each row: a tree of several tests in parallel probe P_i against each template T_j



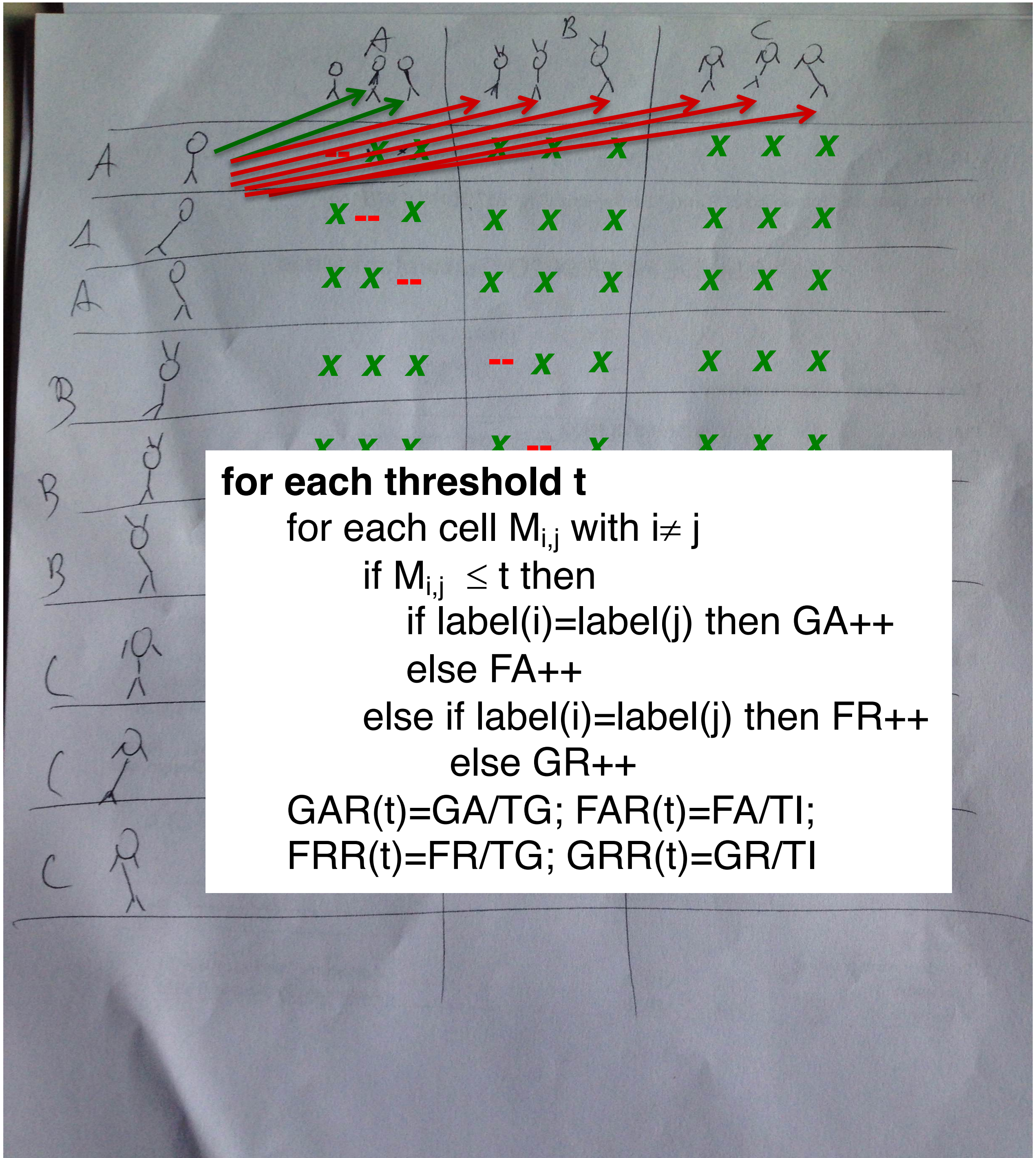
Impostor	Impostor	Genuine	Genuine	Impostor
IMP++	IMP++	GEN++	GEN++	IMP++
If dist <=th	If dist <=th	If dist <=th	If dist <=th	If dist <=th
FA++	FA++	GA++	GA++	FA++
Else	Else	Else	Else	Else
GR++	GR++	FR++	FR++	GR++

IMP = number of impostor tests; GEN = number of genuine tests
FA = number of false accepts; FR = number of false rejects
GA = number of genuine accepts; GR = number of genuine rejects

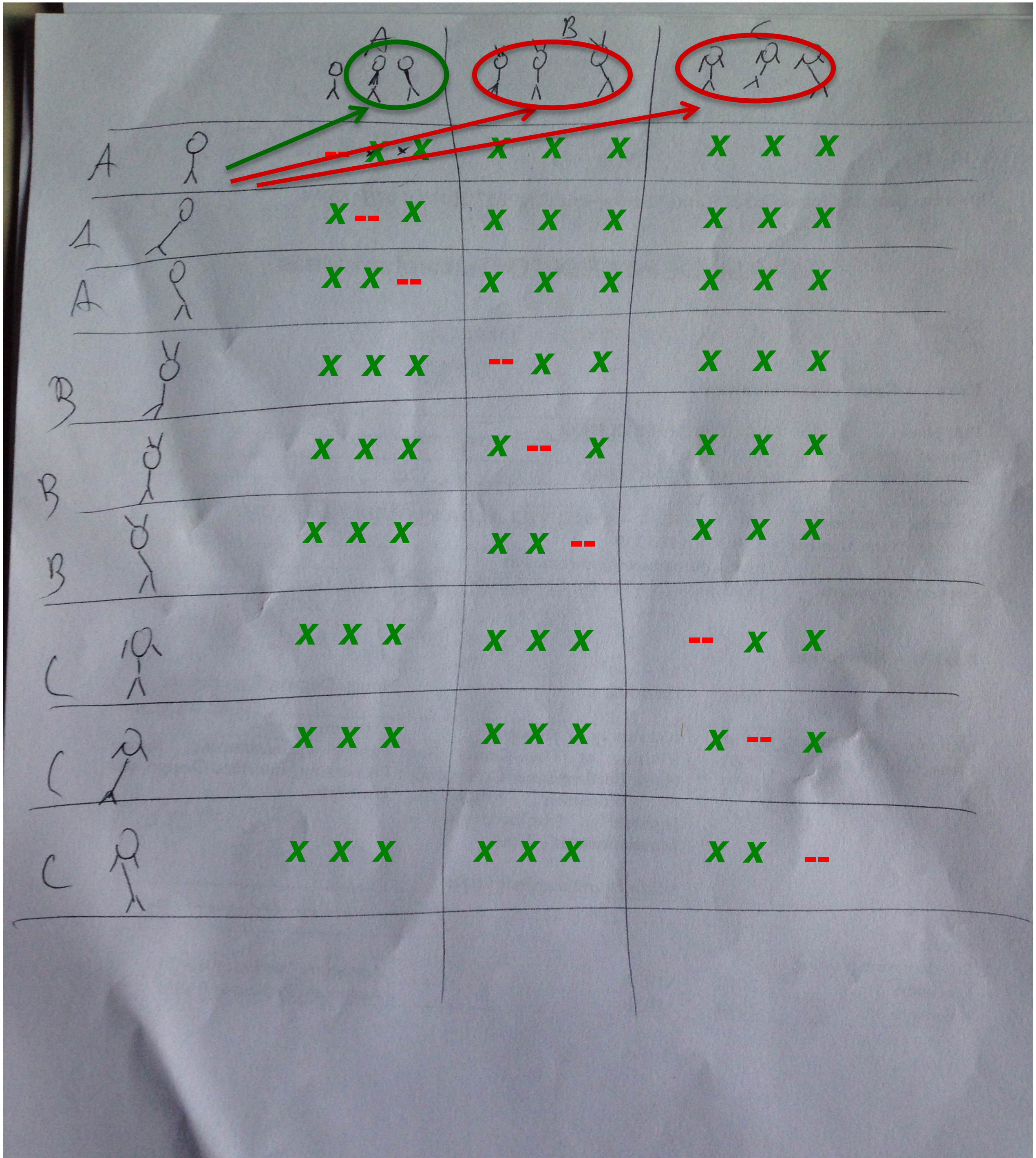
- N = number of subjects
- |G| = total number of samples
- S = number of templates per subject (|G|=S×N)
- M = distance matrix

Verification single-template

- **Each** row is a set of **|G|-1 operations**
- **Each** row contains **S-1 genuine attempts**
- **Each** row contains **(N-1) × S impostor attempts**
- **Total Genuine attempts TG = |G| × (S-1)**
- **Total Impostor attempts TI = |G| × (N-1) × S**




```
for each threshold t
    for each cell M_i,j with i≠ j
        if M_i,j ≤ t then
            if label(i)=label(j) then GA++
            else FA++
        else if label(i)=label(j) then FR++
            else GR++
    GAR(t)=GA/TG; FAR(t)=FA/TI;
    FRR(t)=FR/TG; GRR(t)=GR/TI
```



- N = number of subjects
- |G| = total number of samples
- S = number of templates per subject (|G|=S×N)
- M = distance matrix

Verification multiple-template

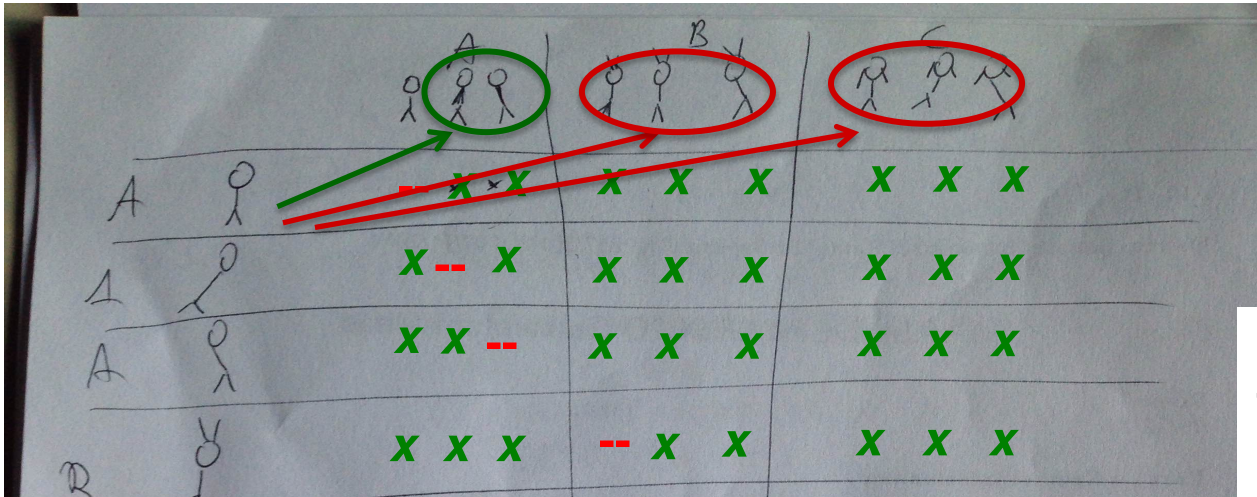
Each row: a tree of several tests in parallel probe P_i against each group of templates T_j

P_i (row i)
label (P_i)

Best T with label (T) = 1	Best T with label (T) =2	. . .	Best T with label (T) = label(P_i)	. . .	Best T with label (T) = k	. . .
Impostor	Impostor		Genuine		Impostor	
IMP++ If dist <=th FA++ Else GR++	IMP++ If dist <=th FA++ Else GR++		GEN++ If dist <=th GA++ Else FR++		IMP++ If dist <=th FA++ Else GR++	

IMP = number of impostor tests; GEN = number of genuine tests
FA = number of false accepts; FR = number of false rejects
GA = number of genuine accepts; GR = number of genuine rejects

- N = number of subjects
- |G| = total number of samples
- S = number of templates per subject (|G|=S×N)
- M = distance matrix



Verification multiple-template

for each threshold t

for each row i

for each group M_{label} of cells $M_{i,j}$ with same label(j)

excluding $M_{i,i}$

select diff = min(M_{label})

if diff $\leq$ t then

if label(i)=label(M_{label}) then GA++

else FA++

else if label(i)=label(M_{label}) then FR++

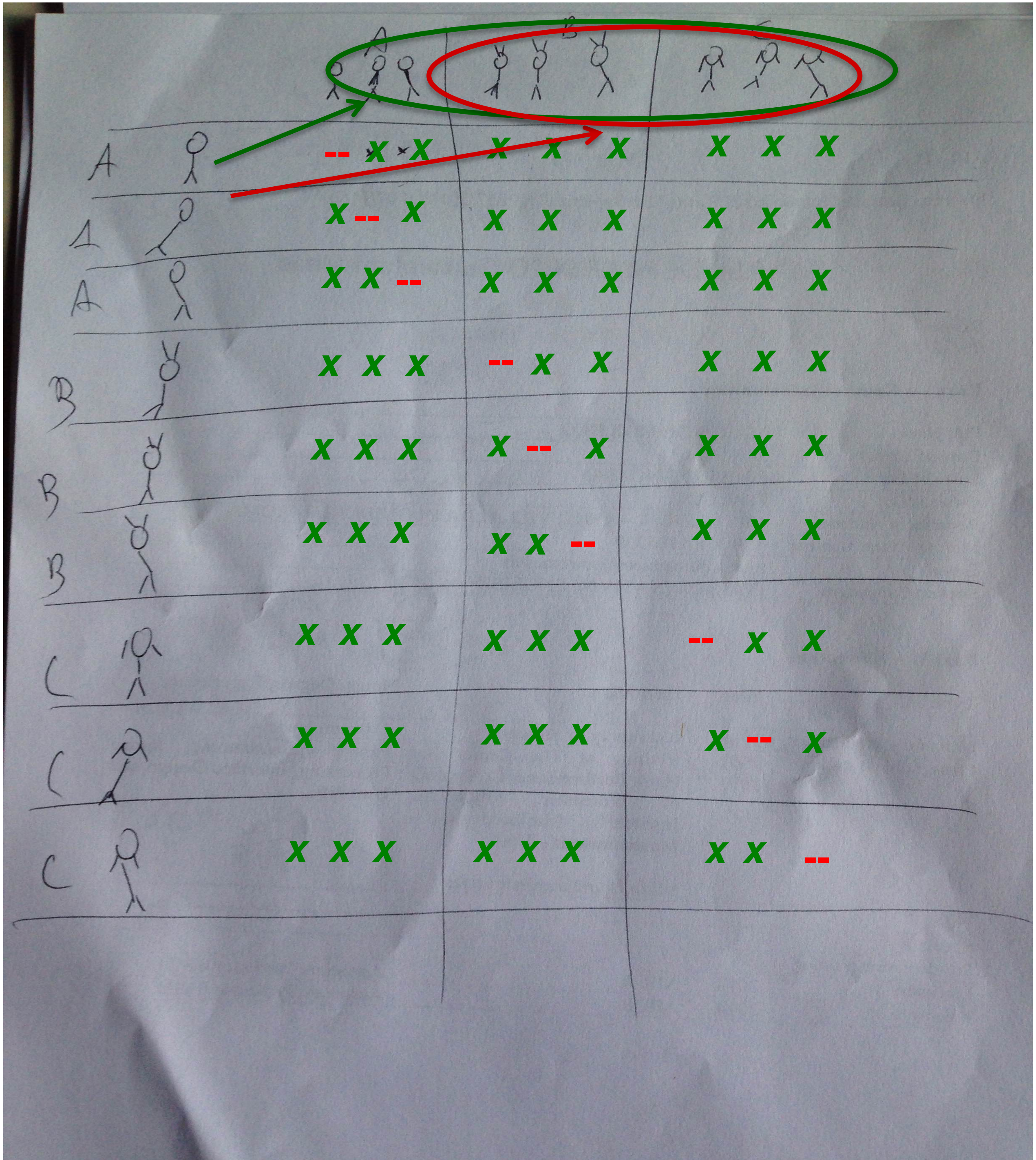
else GR++

GAR(t)=GA/TG; FAR(t)=FA/TI;

FRR(t)=FR/TG; GRR(t)=GR/TI

- **Each** row is a set of **N operations**
- **Each** row contains **1 genuine attempt**
- **Each** row contains **(N-1) impostor attempts**
- **Total Genuine attempts TG = |G|**
- **Total Impostor attempts TI = |G| × (N-1)**

- Substituting min(M_{label}) with avg(M_{label}) may increase accuracy (more restrictive) but some advantages of multiple-template approach may be lost



- N = number of subjects
- IGI = total number of samples
- S = number of templates per subject (IGI=S×N)
- M = distance matrix

To avoid having a too complicated pseudo-code, we only consider Identification open set multiple-template. Otherwise, being an identification operation, we should consider all possible sets of single templates per subject with their ordering.

Each row: a tree of two tests in parallel probe P_i either in the gallery (G) or not (NG)

P_i (row i)
label (P_i)

Label(P_i) not in G

Impostor

Label(P_i) in G

Genuine

The ordered candidate list is the same but when label(P_i) is not in the gallery all the templates with label(P_i) should not appear in the list **since only the Gallery tenplates** should be compared with the probe; therefore we do not consider them

Ordered template list. Only the label is reported and the template subscript is the list order: two possible cases

Label(P_i) not in G

~~Label(T_{l1}) =Label(P_i)~~ ~~Label(T_{l2}) =Label(P_i)~~ Label(T_{l3}) !=Label(P_i) . . .

Label(T_{l1}) !=Label(P_i) . . .

Each row: a tree of two tests in parallel probe P_i either in the gallery (G) or not (NG)

P_i (row i)
label (P_i)

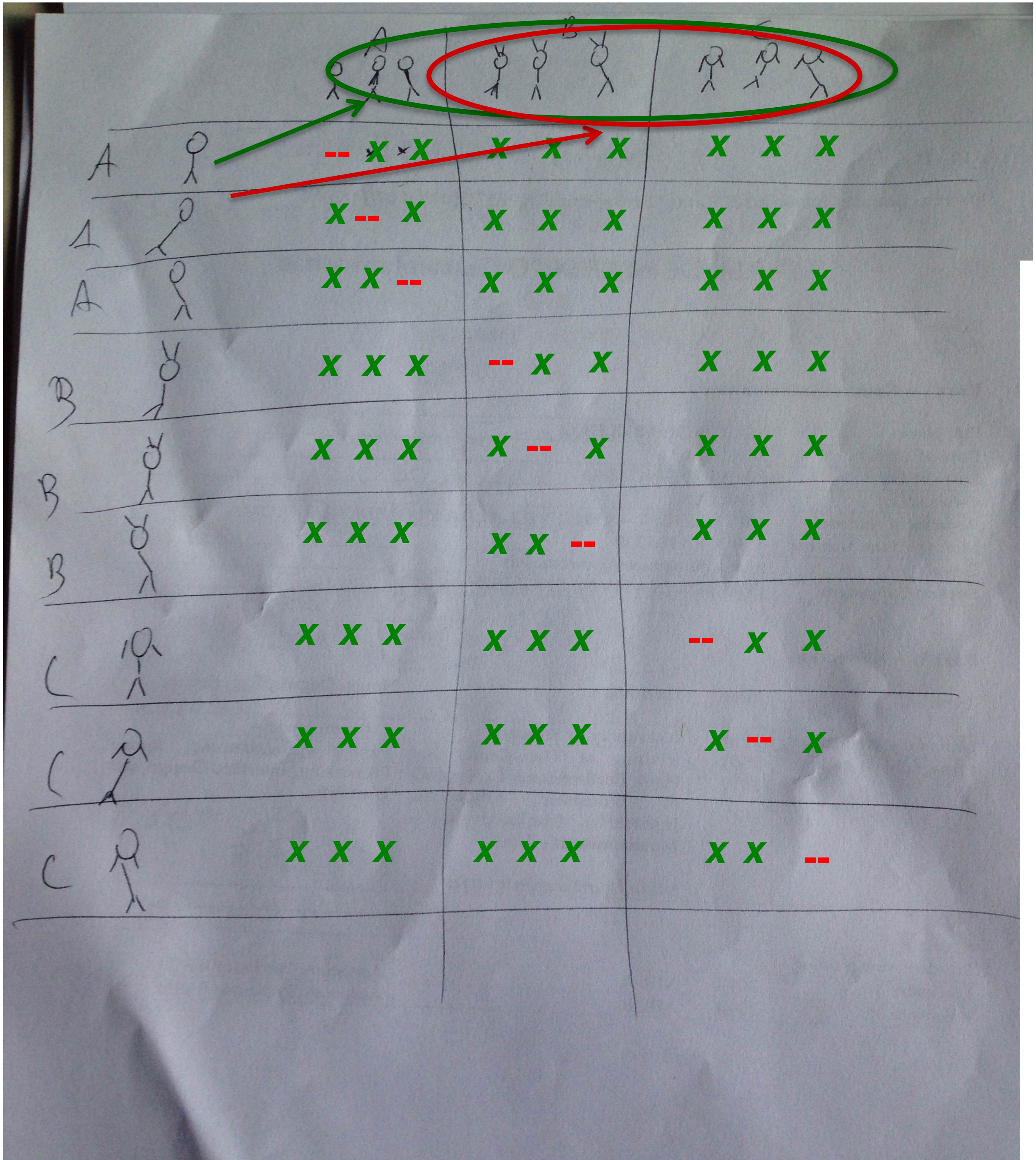
label (P_i) not in G

IMP++
If DistNG <=th
  FA++
Else
  GR++

label (P_i)in G

GEN++
If DistG <=th
  If label(T_{l1} = P_i)
    DIR(1)++
  Else FR++
Else
  FR++

DistNG= distance from the first template with different label in the ordered list
DistG = distance from the first template in the list
IMP = number of impostor tests; GEN = number of genuine tests
FA = number of false accepts; FR = number of false rejects
GA = number of genuine accepts; GR = number of genuine rejects



- N = number of subjects
- IGI = total number of samples
- S = number of templates per subject (IGI=S×N)
- M = distance matrix

To avoid having a too complicated pseudo-code, we only consider Identification open set multiple-template. Otherwise, being an identification operation, we should consider all possible sets of single templates per subject with their ordering.

- **Each** row is a set of **2 identification operations**
- **Each** row contains **1 genuine attempt (subject in the gallery)**
- **Each** row contains **1 impostor attempt (subject not in the gallery)**
- **Total Genuine attempts TG = IGI**
- **Total Impostor attempts TI = IGI**

for each threshold t

for each row i

$\{L_{i,m} \mid m=1 \ldots |G|-1\}$ =

$\{M_{i,j} \mid j=1, \ldots |G|\} \setminus M_{i,i}$ ordered by increasing value (the identical element is excluded)

if $L_{i,1} \leq t$ then (potential accept)

if label(i)=label($L_{i,1}$) then DI(t, 1)++ (genuine case detected+identified)

(parallel impostor case: jump the templates belonging to label(i) since i not in G)

find the first $L_{i,k}$ such that label($L_{i,k}$)<>label(i) AND $L_{i,k} \leq t$

if this k exists, then FA++ (the first template != label(i) has a distance $\leq t$)

else GR++ (impostor is correctly not detected)

else find the first $L_{i,k}$ such that (if genuine yet not the first, look for higher ranks)

label(i)=label($L_{i,k}$) AND $L_{i,k} \leq t$

if this k exists, then DI(t, k)++ (end of genuine)

FA++ (impostor in parallel, distance below t but different label)

(no need to jump since the first label is not the «impostor»)

else GR++ (impostor case counted directly, FR computed through DIR)

DIR(t,1)=DI/TG; FRR(t)= 1- DIR(t,1)

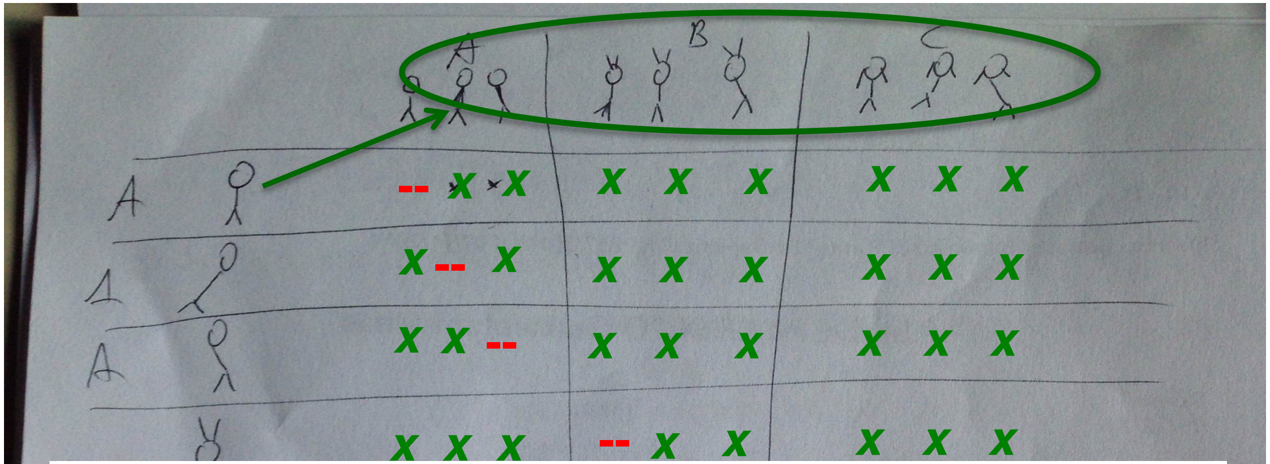
FAR(t)=FA/TI; GRR(t)=GR/TI

k=2 (higher ranks)

while DI(t, k)$\neq$0

DIR(t, k)=DI(t, k)/TG+DIR(t, k-1) (we have to compute rates)

- N =number of subjects
- IGI = total number of samples
- S = number of templates per subject (IGI=S×N)
- M = distance matrix



for each row i

$\{L_{i,m} | \ m=1 \ldots |G|-1\} =$

$\{M_{i,j} | j=1, \ldots |G|\} \setminus M_{i,i}$ ordered by increasing value

find the first $L_{i,k}$ such that

$label(i)=label(L_{i,k})$

CMS(k)++

CMS(1)=CMS(1)/TA; RR=CMS(1)

k=2

while k< IGI-1

CMS(k)=CMS(k)/TA+CMS(k-1)

To avoid having a too complicated pseudo-code, we only consider Identification open set multiple-template. Otherwise, being an identification operation, we should consider all possible sets of single templates per subject with their ordering.

- **Each** row is an **operation**
- **Each** row contains **1 genuine attempt**
- **Total Attempts TA = IGI**
- **No impostor attempt is assumed**
- **No acceptance threshold is used**

ALL-against-ALL

- It is suited for situations where we do not have a known/relevant time elapse among the different templates of a same subject.

- When there is a precise subdivision of samples into sessions **captured during non-overlapping sessions at different times**, it is more fair to create probe and gallery sets using templates from **different sessions**.

- **WHY?**

The reason is that samples of the same subject captured in the same session are **usually more similar than those from different sessions**, due to variations happening with time (e.g., slight face variations)/environmental conditions (e.g. illumination)/sensor performance (e.g., dirt deposited on a fingerprint sensor), therefore the overall performance may appear better.

All Against All – Probe vs Gallery

- Also in this case it is possible to compute extended statistics using a **complete** $\text{ALL}_{\text{PROBE}}$-against-$\text{ALL}_{\text{GALLERY}}$ distance matrix (all **probe** templates are compared with all **gallery** templates).
- Also in this case distances are computed once and for all, and they are used in a different way according to the setup (modality and number of gallery templates per subject)
- Each template plays in turn the role of either genuine/impostor or enrolled/not enrolled more than once according to the recognition application (see the different cases).
- It could be useful to create different probe/gallery partitions and average the results

All Against All – Probe vs Gallery

- Each probe (rows) is considered either as genuine/enrolled or impostor/not enrolled in turn.
- Results are accumulated to get the final statistics.
- **Each row** possibly contributes **more times** according to the number of experiments it can possibly represent.
- Again, for simplicity, it is possible to assume the same number of samples per subject. Extensions are straightforward.

All Against All – Probe vs Gallery

- **Pros**: easy to program, in practice computes a kind of «average» over all possible specific distributions of genuine/impostor (enrolled/not enrolled) attempts.
- **Pros:** The number of impostors is still much higher than genuine, therefore it is possible to overstress the system to assess situations with many impostor attempts.

- **Cons**: the time requested to compute the full matrix may become very high.
- **Cons**: does not to analyze specific distributions of genuine/impostor /enrolled/not enrolled) attempts, that may achieve peculiar results.

All Against All – Probe vs Gallery

- Also in this case, each row is either **a single** test operation **or a set of tests**, depending on the recognition modality.

- We will again distinguish between **single-template** and **multiple-template** settings, depending on the **number of samples per subject** that are assumed to be stored in the test gallery.

- N =number of **subjects (all contributing to both probe and gallery)**
- $|G|$ = total number of **gallery samples** = number of columns in the matrix
- $|P|$ = total number of **probe samples** = number of rows in the matrix
- S = number of **templates per subject.** Without loss of generality we can assume that they are evenly divided between probe and gallery ($|G| = |P| = S \times N$)

- i = **row index** (**probe** template) and **label(i)** is the associated identity
- j = **column index** (**gallery** template) and **label(j)** is the associated identity
- label($\cdot$) = **ground truth**

Verification single-template

- N = number of subjects
- IGI = total number of gallery samples
- IPI = total nuber of probe samples
- S = number of templates per subject in both probe and gallery(IGI= IPI = S×N)
- M = distance matrix

- **Each** row is a set of **IGI operations**
- **Each** row contains **S genuine attempts**
- **Each** row contains **(N-1) × S impostor attempts**
- **Total Genuine attempts TG = IPI × (S)**
- **Total Impostor attempts TI = IPI × (N-1) × S**

for each threshold t
for each cell $M_{i,j}$
if $M_{i,j} \leq t$ then
if label(i)=label(j) then GA++
else FA++
else if label(i)=label(j) then FR++
else GR++
GAR(t)=GA/TG; FAR(t)=FA/TI;
FRR(t)=FR/TG; GRR(t)=GR/TI

Verification multiple-template

- N = number of subjects
- |G| = total number of gallery samples
- |P| = total nuber of probe samples
- S = number of templates per subject in both probe and gallery(|G|= |P| = S×N)
- M = distance matrix

for each threshold t
    for each row i
        for each group M_{label} of cells $M_{i,j}$ with same label(j)
            select diff = min(M_{label})
            if diff $\leq$ t then
                if label(i)=label(M_{label}) then GA++
                else FA++
            else if label(i)=label(M_{label}) then FR++
                    else GR++
    GAR(t)=GA/TG; FAR(t)=FA/TI;
    FRR(t)=FR/TG; GRR(t)=GR/TI

- **Each** row is a set of **N operations**
- **Each** row contains **1 genuine attempt**
- **Each** row contains **(N-1) impostor attempts**
- **Total Genuine attempts TG = |P|**
- **Total Impostor attempts TI = |P| × (N-1)**

Substituting min(M_{label}) with avg(M_{label}) may increase accuracy (more restrictive) but some advantages of multiple-template approach may be lost

Identification open-set

- N = number of subjects
- IGI = total number of gallery samples
- IPI = total nuber of probe samples
- S = number of templates per subject in both probe and gallery(IGI= IPI = S×N)
- M = distance matrix

To avoid having a too complicated pseudo-code, we only consider Identification open set multiple-template. Otherwise, being an identification operation, we should consider all possible sets of single templates per subject with their ordering

- **Each** row is a set of **2 identification operations**
- **Each** row contains **1 «genuine» attempt (subject enrolled in the gallery)**
- **Each** row contains **1 «impostor» attempt (subject not enrolled in the gallery)**
- **Total Genuine attempts TG = IPI**
- **Total Impostor attempts TI = IPI**

for each threshold t

for each row i

$\{L_{i,m} | m=1 \dots |G|\} =$

$\{M_{i,j} | j=1, \dots |G|\}$ ordered by increasing value

if $L_{i,1} \leq t$ then (potential accept)

if label(i)=label($L_{i,1}$) then DI(t, 1)++ (label(i) enrolled)

find the first $L_{i,k}$ such that label($L_{i,k}$)<>label(i) AND $L_{i,k} \leq t$

if this k exists, then FA++ (label(i) not enrolled, jump it)

else find the first $L_{i,k}$ such that (if genuine yet not the first, look for higher ranks)

label(i)=label($L_{i,k}$) AND $L_{i,k} \leq t$

if this k exists, then DI(t, k)++ (end of genuine)

FA++ (impostor in parallel, distance below t but different label)

(no need to jump since the first label is not the «impostor»)

else GR++ (impostor case counted directly, FR computed through DIR)

DIR(t,1)=DI/TG; FRR(t)= 1- DIR(t,1)

FAR(t)=FA/TI; GRR(t)=GR/TI

k=2 (higher ranks)

while DI(t, k)$\neq$0

DIR(t, k)=DI(t, k)/TG+DIR(t, k-1) (we have to compute rates)

Identification closed-set

- N = number of subjects
- |G| = total number of gallery samples
- |P| = total nuber of probe samples
- S = number of templates per subject in both probe and gallery(|G|= |P| = S×N)
- M = distance matrix

for each row i

$\{L_{i,m} | \ m=1 \ \ldots \ |G|\}$ =

$\{M_{i,j} \ | j=1, \ldots |G|\}$ ordered by increasing value

find the first $L_{i,k}$ such that

label(i)=label($L_{i,k}$)

CMS(k)++

CMS(1)=CMS(1)/TA; RR=CMS(1)

k=2

while k< |G|

CMS(k)=CMS(k)/TA+CMS(k-1)

To avoid having a too complicated pseudo-code, we only consider Identification closed set multiple-template. Otherwise, being an identification operation, we should consider all possible sets of single templates per subject with their ordering.

- **Each** row is an **operation**
- **Each** row contains **1 genuine attempt**
- **Total Attempts TA = |P|**
- **No impostor attempt is assumed**
- **No acceptance threshold is used**

Why is it so important to carry out the correct calculation, especially when rates are involved?

Let us consider the following example, related to verification mode.

There are 100 probes. Let us assume that the system erroneously accepts 10 impostors, and erroneously rejects 10 genuine users.

If we compute FAR and FRR with respect to the total number of probles, we would get FAR=FRR=10/100

What is missing in this resoning?

Why is it so important to carry out the correct calculation, especially when rates are involved?

Let us continue the example.

There are 100 probes. Let us assume that the system erroneously accepts 10 impostors, and erroneously rejects 10 genuine users.

Let us further assume that the impostor probes are actually 10, and genuine probes are actually 90.

According to the previous computation, we still get FRR=FAR=10/100..

Does this sound correct?

Why is it so important to carry out the correct calculation, especially when rates are involved?

Let us continue the example.

There are 100 probes. Let us assume that the system erroneously accepts 10 impostors, and erroneously rejects 10 genuine users.

Let us rather assume that the impostor probes are actually 50, and genuine probes are actually 50.

According to the previous computation, we still get FRR=FAR=10/100..

Does this sound correct?

Why is it so important to carry out the correct calculation, especially when rates are involved?

Let us continue the example.

There are 100 probes. Let us assume that the system erroneously accepts 10 impostors, and erroneously rejects 10 genuine users.

Let us rather assume that the impostor probes are actually 90, and genuine probes are actually 10.

According to the previous computation, we still get FRR=FAR=10/100..

Does this sound correct?

Why is it so important to carry out the correct calculation, especially when rates are involved?

There are 100 probes. Let us assume that the system erroneously accepts 10 impostors, and erroneously rejects 10 genuine users.

In the previous cases, the system makes a different «amounts» of error and in different «directions» !

We also have to consider the real number od impostors and the real number of users! The error rate is the number of incorrect responses vs. thoserelated to the same kind of users.

Why is it so important to carry out the correct calculation, especially when rates are involved?

Probes	Genuine	Impostors	FR	FA	FRR	FAR
100	90	10	10	10	10/90=0.11	10/10=1 (!!!)
100	50	50	10	10	10/50=0.2	10/50=0.2
100	10	90	10	10	10/10=1 (!!!)	10/90=0.11

It is clear that, given the same total number of probes and the same number of incorrect answers, the three situations are very different!

NOTE: In ALL the above cases the Accuracy=(Number of correct answers)/(Total number of answers) is 80%! We cannot be satisfied with this result because we need to evaluate the main «direction» of errors!