Discovering Event Media Semantics using Games with a Hidden Purpose

Francesco G.B. De Natale
DISI - University of Trento

ACM-Multimedia Workshop on
Event-based Media Integration and Processing
Barcelona, Oct. 21-22, 2013

Acknowledgements

- Have collaborated to this work:
  - Nicu Sebe
  - Andrea Rosani
  - Giulia Boato
  - Oscar Fanelli
  - Silvia Modenese

- *Part of this activity has been developed with the support of the EIT-ICTLabs Project "S-MAX", funded by the European Union*

Outline

- Event detection from media
  - Event vs. media
  - Event discovery
  - Event-based media classification
- Crowdsourcing and Gamification
  - Exploiting humans in complex tasks
  - Games with a hidden purpose
- Event Saliency detection
  - A game-based approach using masking and discovery
- Event media association
  - A game-based approach using domino
- Conclusions

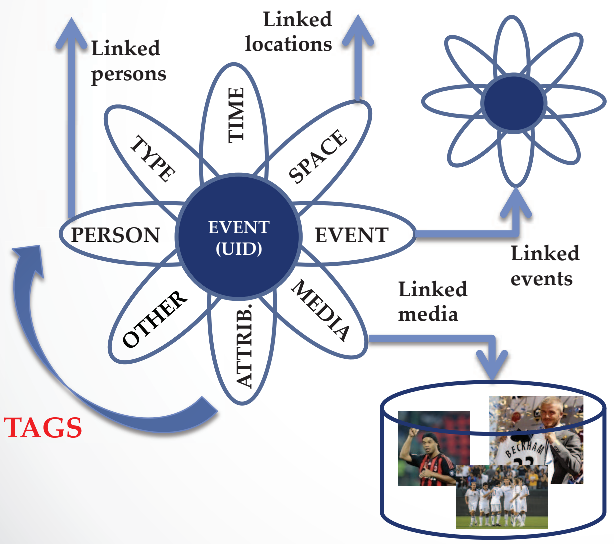
Can you guess which event is it?

Events and Media

- Whatever they are everyday personal experiences or large social happenings, events mark our lives and memories
- Events are more and more associated to massive quantities of media: photos, videos, audio recordings, tweets, web pages, etc.
- Event-based indexing of media makes possible a number of applications, such as:
  - Faceted search based on event-related tags
  - Creating event/media networks (graphs)
  - Browsing media following intuitive event-related links
  - Easily creating summaries, stories, etc.

Event models

- A typical approach to associate media to events is to define appropriate event models



media become the experiential dimension of events!

www.glocal-project.eu

Media vs Event Models

- Some nice problems related to the above model are the following:
  - How to automatically extract facets (tags) from media:
    - Identify the location through landmarks or similarity
    - Identify people by face detection and recognition
    - Detect significant objects
    - Understand the event/sub-event type
  - How to associate new media to an event
    - Recognize and link related events
    - Gather relevant images from the network (similarity, relatedness)
    - Cluster and/or classify relevant media
  - How to automatically instantiate a model over a given media collection
    - Discover the event type and instantiate the relevant event structure
    - Cluster media in sub-events and label them accordingly

P.Andrews, F.De Natale, S.Buschbeck, A.Jameson, K.Bischoff, C.Firan, C.Niederée, V.Mezaris, S.Nikolopoulos, V.Murdock, and A.Rae, "GLOCAL: event-based retrieval of networked media," *Proc. 21st ACM Intl. Conf. on World Wide Web, 2012*

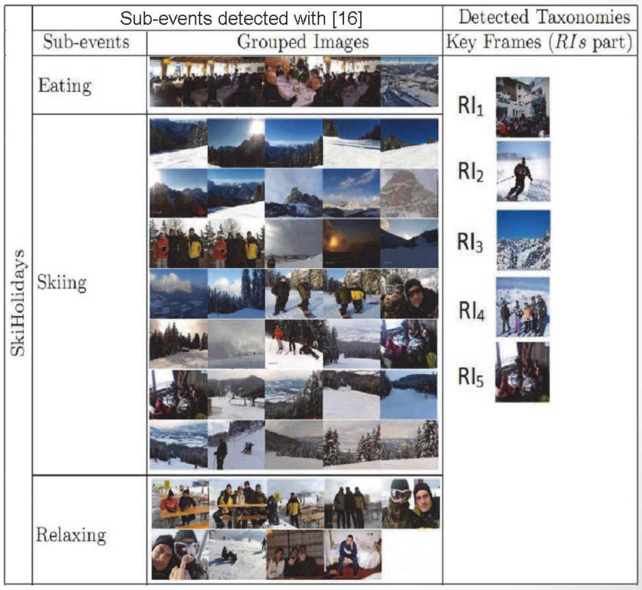
Media @ Event: the gap

- All these problems share the same major technological challenge:
  - How to extract event semantics from media → semantic gap
- Two main approaches (*):
  - Extract visual concepts and then correlate visual concepts to events
    - More complex (cascade classifiers, managing uncertainty)
    - Explicit knowledge vs. learning
    - May work on single images
  - Directly correlate visual content to events
    - Requires that the media contains a clear event fingerprint
    - Often unsuitable for single images, works better and faster on media collections (exploiting event-structure fingerprint)

(*) assumption: using visual content only

Example 1: event-based clustering of photo albums

- Automatic clustering/annotation of photo collections
  - Collect event samples
  - Detect implicit structures
  - Extract fingerprints
  - Match with new event



M. Dao, D. Dang-nguyen, F. De Natale, "Robust event discovery from photo collections using Signature Image Bases (SIBs)" in MULTIMEDIA TOOLS AND APPLICATIONS, 2012

Example 2: Automatic event detection

- Associating single media items to an event class based on visual content
  - Don't trust on individual visual concepts
  - Build "visual concept vectors"
  - Apply Mixture Subclass Discriminant Analysis (MSDA) to learn event classes
  - Nearest neighbor media-event association

"leaders meeting"

Airplane-flying → 0.51
Animal → 0.33
…

$$\begin{bmatrix} 0.51 \\ 0.33 \\ \cdots \\ 0.91 \\ \cdots \\ 0.47 \end{bmatrix}$$

"demostration"

N. Gkalelis, V. Mezaris, I. Kompatsiaris, "High-level event detection in video exploiting discriminant concepts", Proc. 9th Int. Workshop on Content-Based Multimedia Indexing, 2011

But… what are the limits?

- Visual concept detection still performs poorly
  - Single concept detection is still highly unreliable
  - Multiple-concepts may increase evidence but uncertainty remains high
- Limits of event-media models
  - Linking visual concepts with events is not trivial (a-priori knowledge may be biasing)
  - Learning is a possibility, but the variety of events is very large
  - Subjectivity, cultural differences, etc. introduce further fuzziness

- Example: MediaEval Social Event Detection
  - SED task 2013: classification of various event types from a single image
  - Average performance over 10 research groups around 30% (13-50%)

A possible way out: Human computing

- In human computing the idea is to put the man in the loop to solve tasks that are out of reach for current technologies
- Among the most interesting approaches:
  - Crowdsourcing:
    - it is the outsourcing of microtasks to crowds
    - main incentive is economic revenue (workers are paid)
    - based on platforms (Mechanical Turk, Crowdflower, Microworkers)
  - Gamification:
    - it is still an outsourcing of tasks to crowds or communities, BUT
    - the real task is hidden within a gaming environment
    - incentives is mainly reputation or entertainment (but also points, next levels, status, virtual or real goods, etc.)
    - there is a social dimension (among players)
    - could be performed through crowdsourcing platforms but also other means (Web, mobile platforms, etc.)

Hiding the real purpose

- Often it is useful to hide the real purpose of a task
  - This is not necessarily done to exploit human power at low cost, but mainly to achieve better and less biased results
  - Example: the reCAPTCHA tool

Gamification

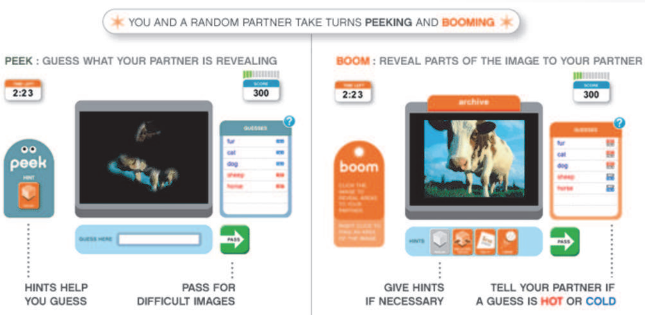
- Luis von Ahn coined the term Games With A Purpose (GWAP) to define a type of crowdsourcing game that is:
  - fun and engaging
  - includes a task that can only be completed by humans
  - has a goal that is hidden from the player
- He also introduced the ESP Game:
  - Two players are presented the same image
  - They have to agree on a common label
  - The game includes mechanisms to incentivize users and avoid cheating

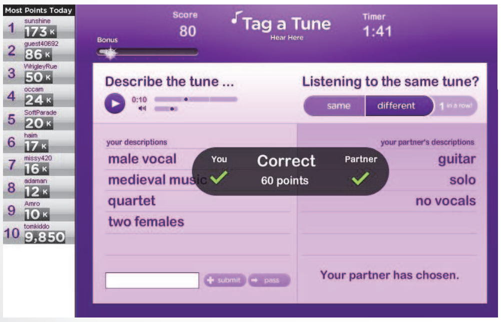
L. Von Ahn and L. Dabbish, "Designing games with a purpose", Communications of the ACM 51, 8, Aug. 2008

L. Von Ahn and L. Dabbish. Labeling images with a computer game. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 319–326. ACM, 2004

Gamification

- Other GWAP mechanisms in the field of media:
  - TagaTune (tagging music)
  - Peekaboom (a player discovers part of object disclosed by the other user)
  - Yahoo's Video Tag Game
  - WhoKnows and RISQ!
  - …





E. Law and L. von Ahn, "Input-agreement: a new mechanism for collecting data using human computation games", CHI 2009

L. Von Ahn, R. Liu, and M. Blum, "Peekaboom: A game for locating objects in images", ACM SIGCHI conference on Human Factors in computing systems 2006

R. van Zwol, L. Garcia Pueyo, G. Ramirez, B. Sigurbjörnsson, and M. Labad, "Video tag game", WWW Conference 2009

N. Ludwig, M. Knuth, J. Waitelonis, and H. Sack, "WhoKnows? - Evaluating Linked Data Heuristics with a Quiz that Cleans Up DBpedia", ESWC 2011

GWAP as a tool to learn event semantics in media

- Based on the above considerations, we considered exploiting games to gain knowledge about some complex problems, such as:
  - Linking event types with their visual representation
  - Understanding links between similar events
  - Learning which are the most significant visual concepts that characterize an event
- In our idea, games are not a way to solve a problem but to learn from humans how to do that
  - Creating a ground truth thanks to human classification
  - Capturing concepts and semantics
- It is very important that the real goal is concealed by the game, to avoid cultural/personal bias

Event masking game: an event saliency detector

- Problem: what is really important in an image to understand the underlying event?

Back to the joke in the first slide:

Which event is it?...

Event masking game: an event saliency detector

- Problem: what is really important in an image to understand the underlying event?

Back to the joke in the first slide:

A wedding!
Hard to guess…

EventMask game: an event saliency detector

- The idea is to define a kind of "event-saliency"
  - What is really important to understand the event depicted by an image?
  - If known, this provides a basis for learning and recognizing event media
- Event saliency is different from visual saliency!
  - Although the two things may partially coincide, the idea is very different:
    - Visual saliency detects the parts of the image that attract attention (depend on color, contrast, foreground, position)
    - Event saliency should detect all the visual contents that can lead to a recognition of the event, even if it is in the background or represented in a marginal detail
- Event saliency is very difficult to extract automatically without a-priori knowledge

Event saliency vs. Visual Saliency

- The relationship between visual and event saliency is not always evident
  - Given the attitude of photographers to attract the attention on most important things sometimes they may partially overlap BUT…
  - often important event related information is on details or background

Event saliency

Visual saliency

EventMask: the game

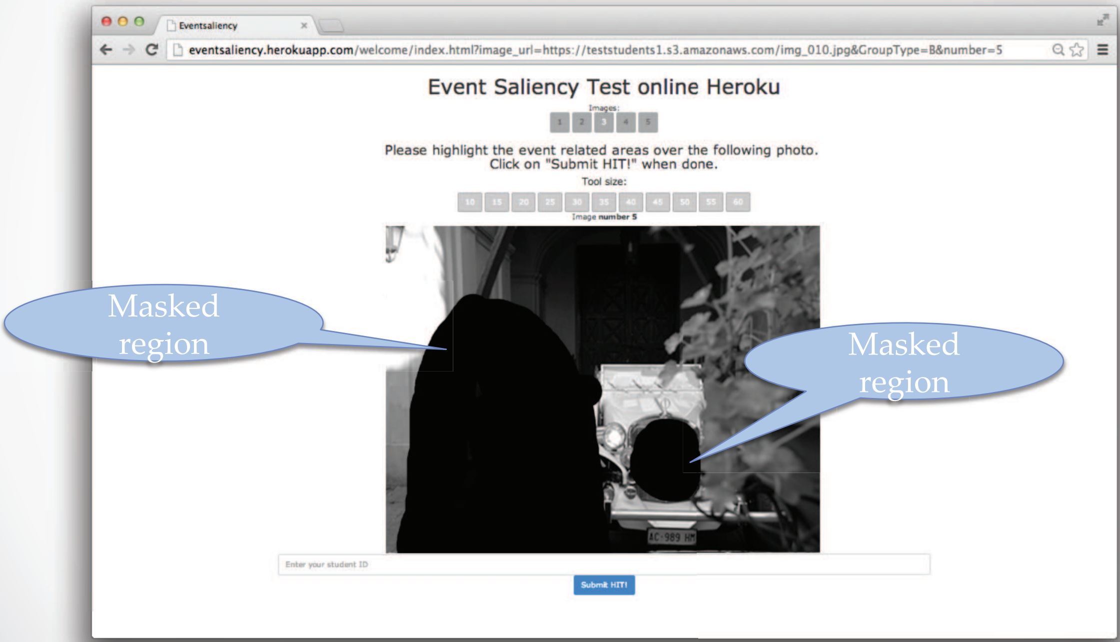
- EventMask is in the form of an inversion problem:
  - We don't ask people to indicate what's important for them to recognize an event, but to hide it to other people
  - It is competitive: to win they have to avoid that the other person recognize the event (this is also the game incentive)
  - Fundamental is to define a scoring system that prevents cheating
- Some similarity with PeekaBoom, but both purpose and mechanism are very different
- The game has two roles: masking and discovering
  - Mask players are presented a few images related to events, and have to hide parts of them so as to make the event unrecognizable
  - Discover players are presented the images masked by other players and have to select the relevant event from a list of possibilities
- The event saliency map is generated from a suitable combination of the above results

The game: masking role

- MASKING ROLE
  - The first player has the masking role
  - Players are presented a few images related to different events: they have to hide parts of them so that the event is no more recognizable
  - His score is inverse proportional to the masked area *(to discourage cheating or hiding the whole image)*
  - Afterwards his score will be diminished proportionally to the ability of adversaries to discover the event from the masked image (*to discourage cheating or leaving the whole image uncovered*)

The game: masking role

The game: masking role

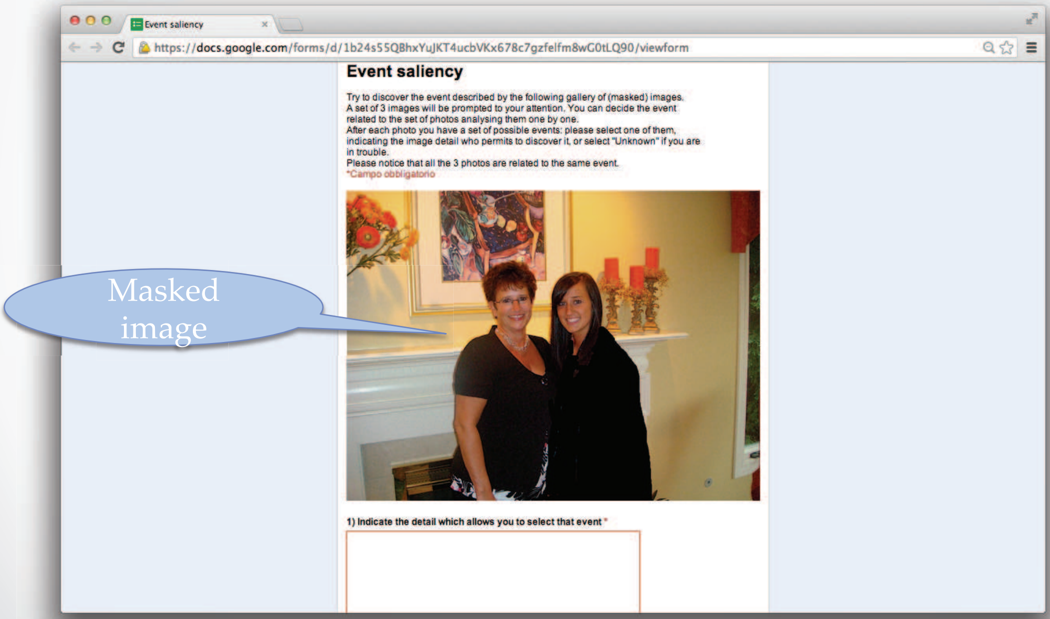


The game: discovery role

- Discovery role
  - The second player (adversary) has the discovery role
  - *Players are requested to evaluate a certain number of images (different* from the ones they masked).
  - They should assign every image to known set of potential events. Correct classification increases score of discoverer and reduce the score of the player who masked that image
  - They can choose not to classify an image, if they are not sure: misclassifications reduce the score (*to avoid cheating*)
  - This role is enabled only after the player acted as a masking player for a while

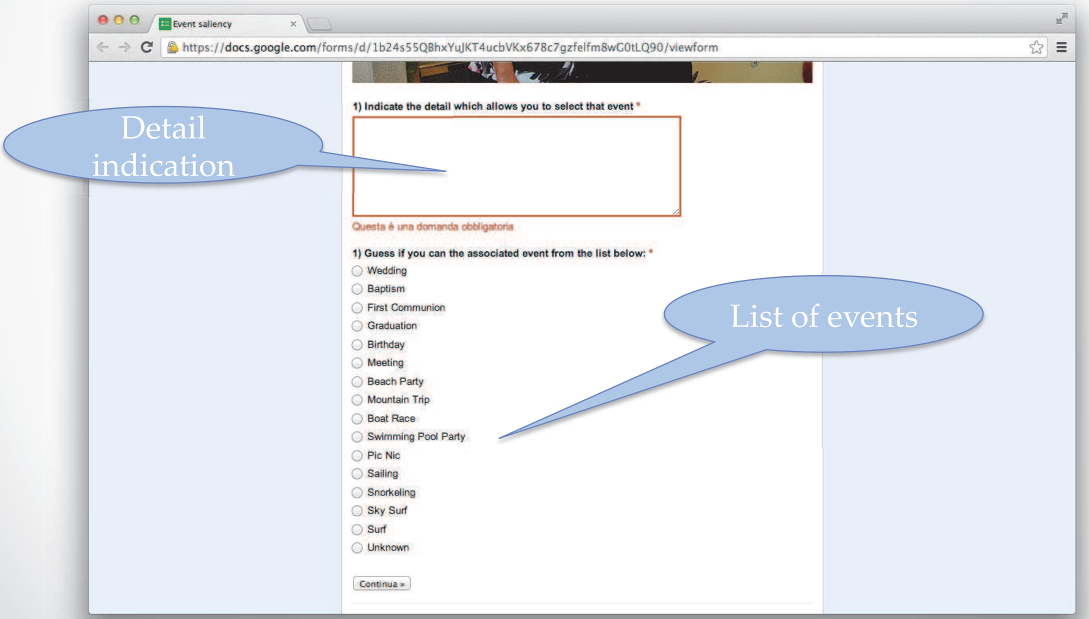
The game: discovery role

- Unmasking procedure



The game: discovery role

- Unmasking procedure



The game: map generation

- Event saliency maps are obtained taking into account both the masking and unmasking results
  - After masking, j_{th} image is associated to a set of masks M_{ij} for each j_{th} masking player:

$$M_{ij}(x,y)=\begin{cases} 1 & pixel\ (x,y)\ is\ marked \\ 0 & pixel\ (x,y)\ is\ unmarked \end{cases}$$

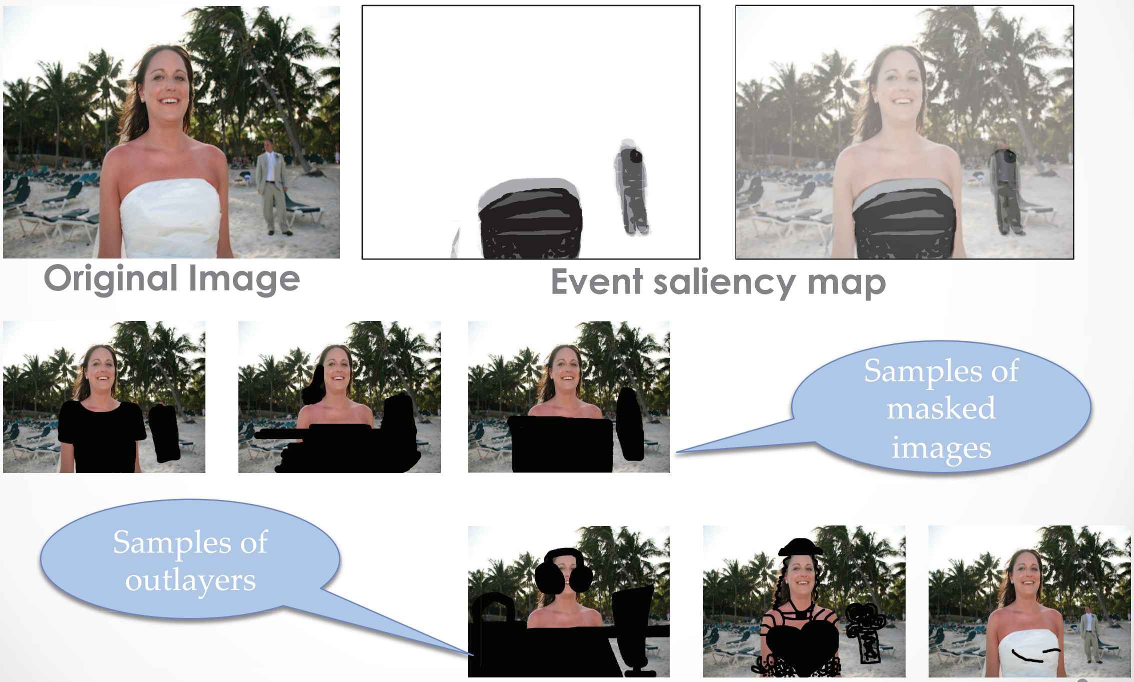
  - After discovery, each map is converted into M_{ij}^* such that:

$$M_{ij}^*(x,y)=\begin{cases} P_D/T_D & if\ M_{ij}(x,y)=0\ AND\ event\ discovered\ by\ P_D\ \ players\ out\ of\ T_D \\ 1 & if\ M_{ij}(x,y)=1 \end{cases}$$

  - Finally, the j_{th} maps are multiplied pixel-wise and normalized to [0,1]

$$M_i(x,y)=\prod_j M_{ij}^*(x,y)$$

Masking game: results



Original Image

Event saliency map

Samples of masked images

Samples of outlayers

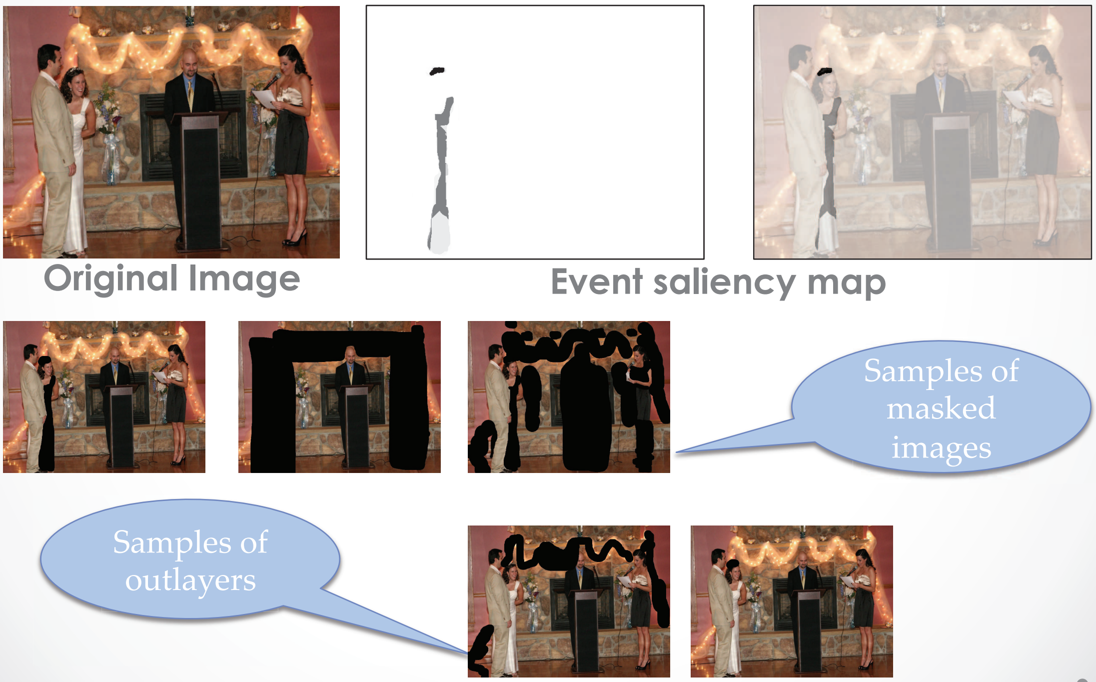
Masking game: results

Original Image

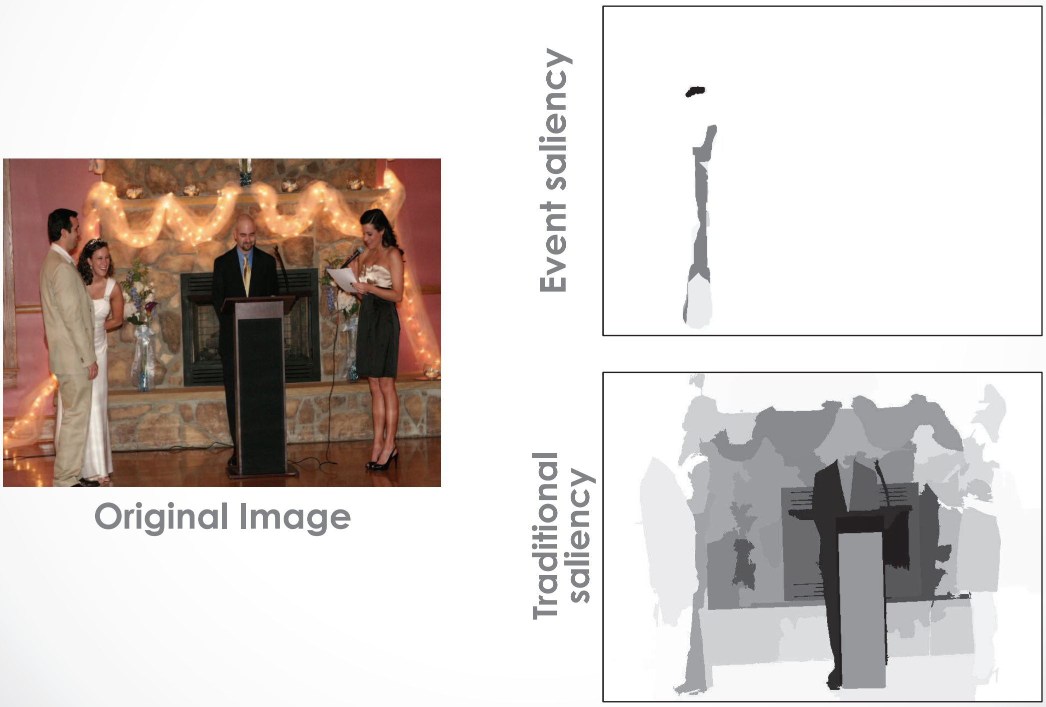
Event saliency

Traditional saliency

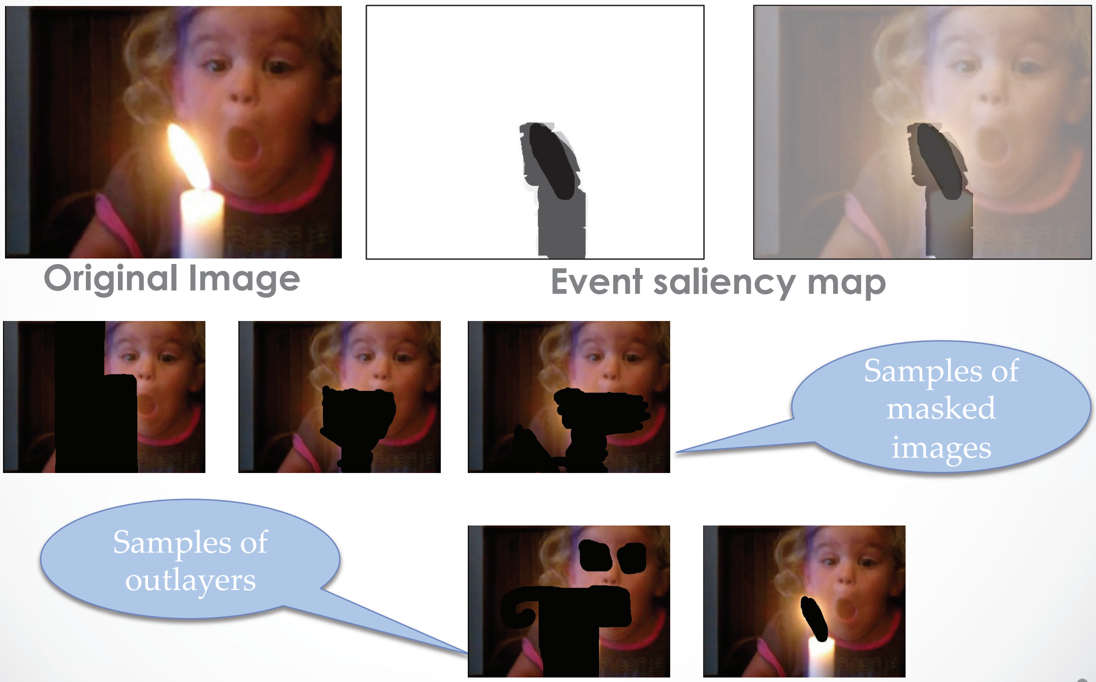
Masking game: results



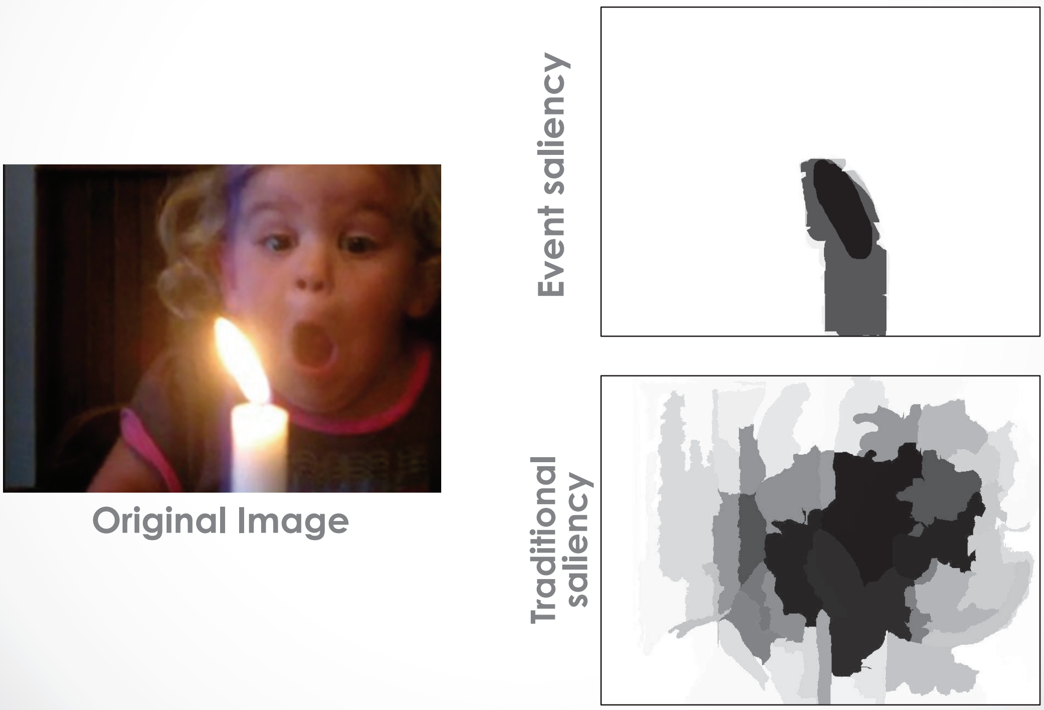
Masking game: results



Masking game: results

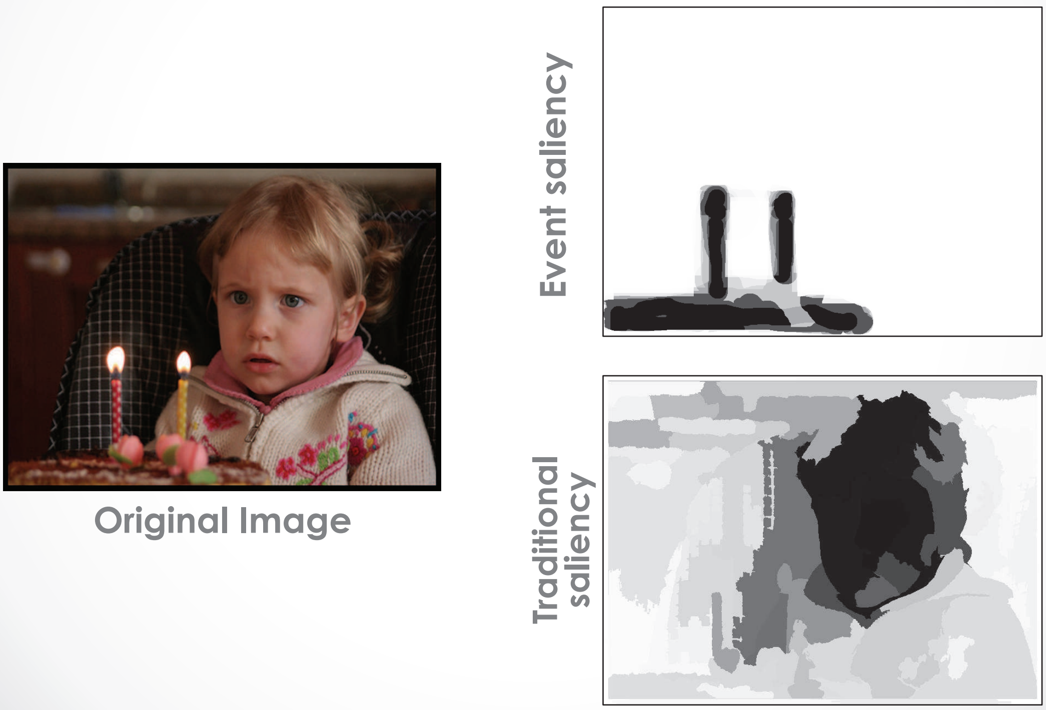


Masking game: results



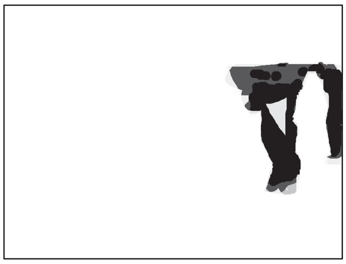
Masking game: results

Masking game: results



Masking game: results

Original Image



Event saliency map

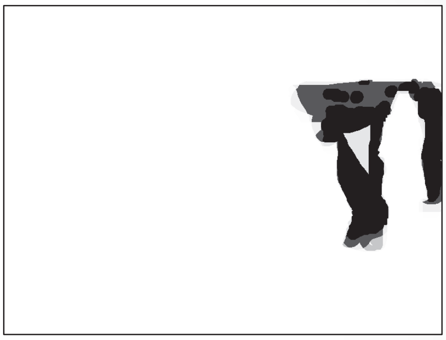
Samples of masked images

Samples of outlayers

Masking game: results

Original Image

Event saliency



Traditional saliency

Masking game: results

Original Image

Event saliency map

Samples of masked images

Samples of outlayers

Masking game: results

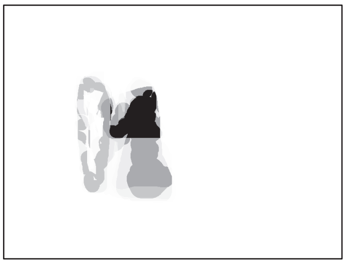
Original Image

Event saliency

Traditional saliency

Masking game: results

Original Image



Event saliency map

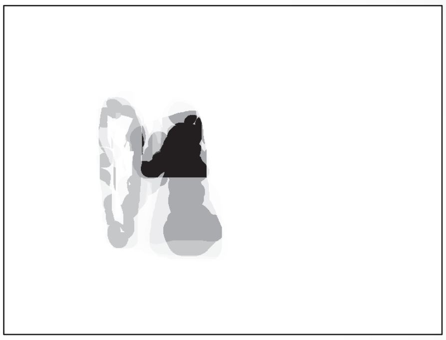
Samples of masked images

Samples of outlayers

Masking game: results

Original Image

Event saliency



Traditional saliency

Domino game: extending annotation across media

- Another significant problem in event media analysis is associating related events
  - Sub-events of the same event
  - Events that are linked by some common facet (who, what, where, when)
- Automatic tools make use of explicit annotation or visual similarity to capture similarities or relatedness among different event media
  - Time is a significant source of information, often available
  - Location is also relevant, even if more rarely available
  - Visual content can be used but introduces more uncertainty

Domino game: extending annotation across media

- Also in this case, gamification may help building and learning relationships among event-media
- A simple approach would be to ask users to match related media, specifying the reason of the link:
  - E.g., linking two images that show the same person, or that refer to the same or to a similar event, or that happened in the same place
- A good mechanism to implement this linking procedure and make it competitive is concatenation: the domino game

The Domino Game

- General description
  - Objective: given two data collections, one (partially) annotated and the other non annotated, extend the annotation of the first to the second according to human-suggested concept links
  - How-to: the linking is driven by a domino game, where players can attach a tile to the domino if the right face of the last tile of the domino matches the left face of the player's tile by some attribute (e.g., similar geo-location, same person appearing, etc.).
  - Scope of the game:
    - Explicit: finish the tiles before the adversary and win
    - Hidden: tag as much images as possible

The Domino Game

- Construction of domino tiles
  - Tiles are automatically created by randomly taking images from the two datasets (annotated vs. non-annotated). Annotated images stay always on the left side

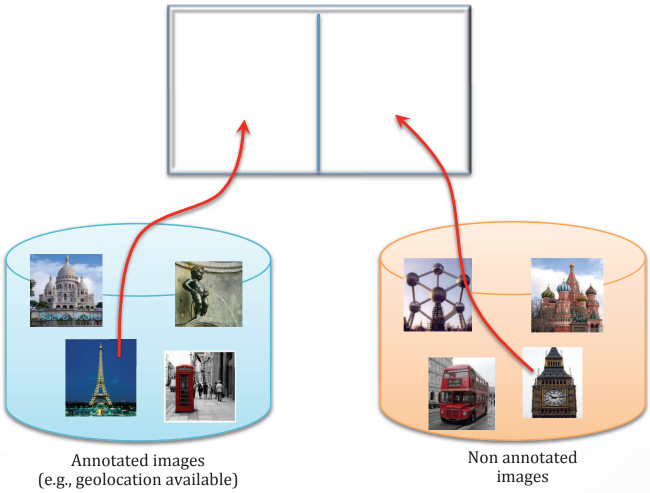


Fig. 1 – Construction of tiles

The Domino Game

- Domino rules
  - While inserting a tile, players automatically insert a new annotated image in the front of the queue, for the other player to continue.
  - Verification of the correctness is left to the other user, who can refuse the tile if the link is wrong (*to avoid cheating*).
  - Winner: wins the player who remains without tiles or has placed more tiles when time is over

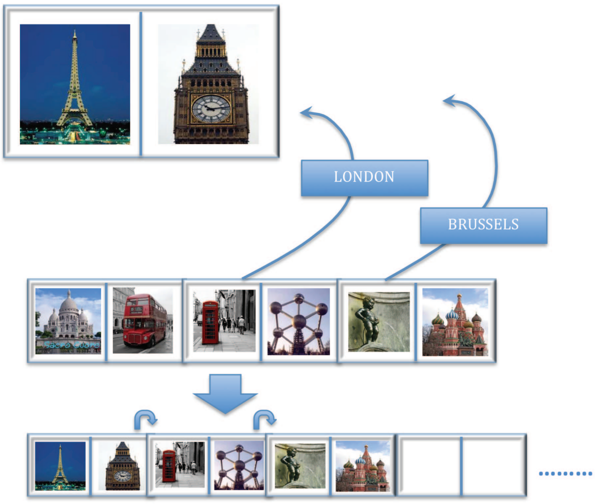
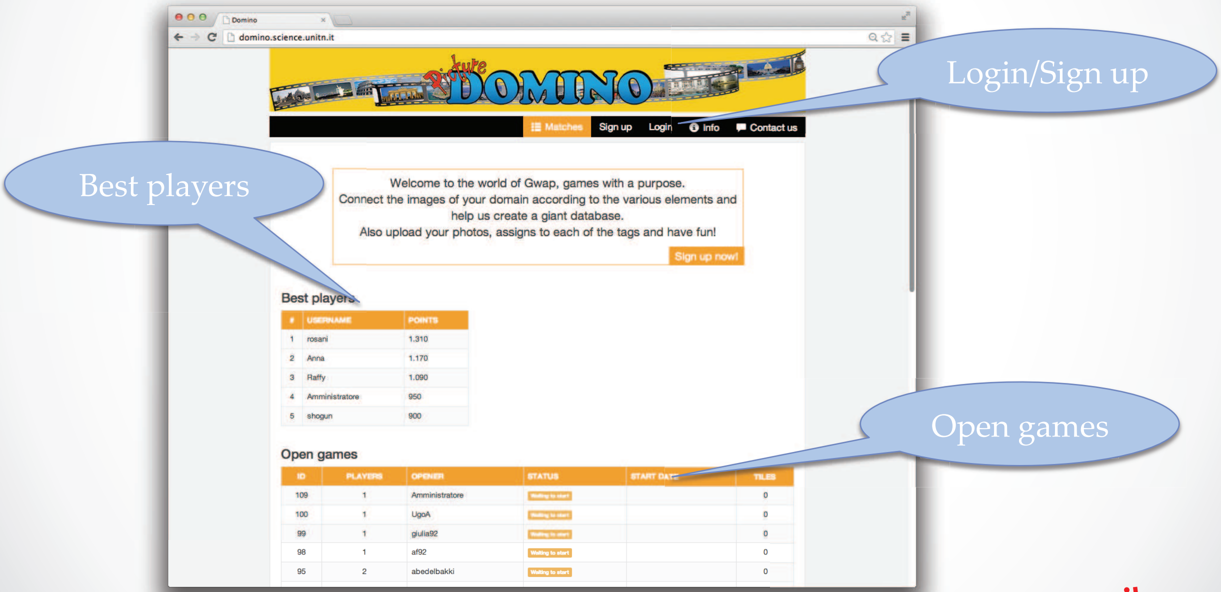


Fig. 2 – Playing Domino

The Domino Game

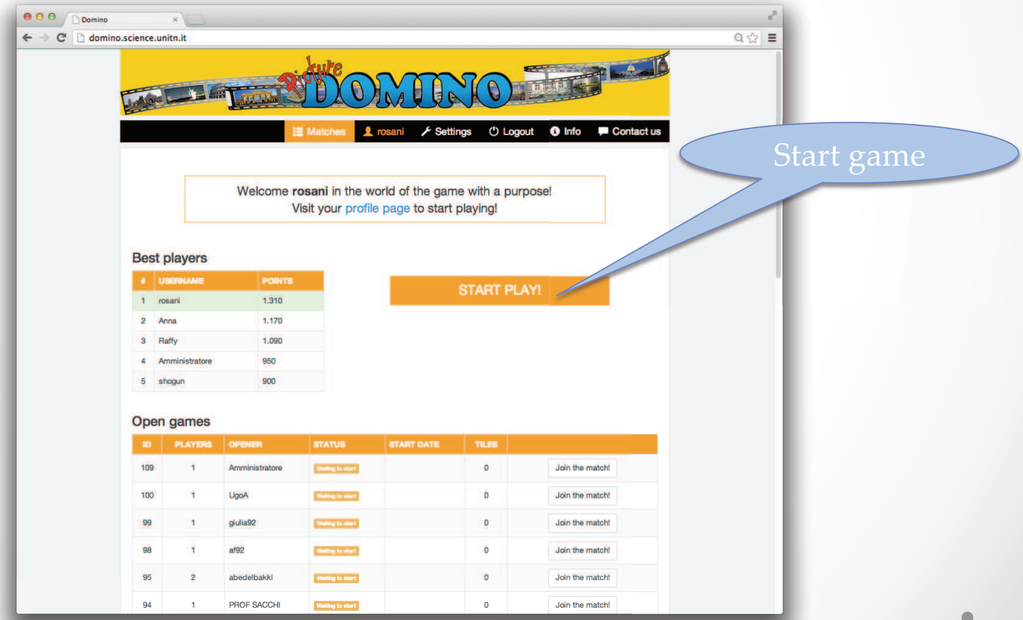
- Web interface: main page



http://domino.science.unitn.it

The Domino Game

- Web interface: user page

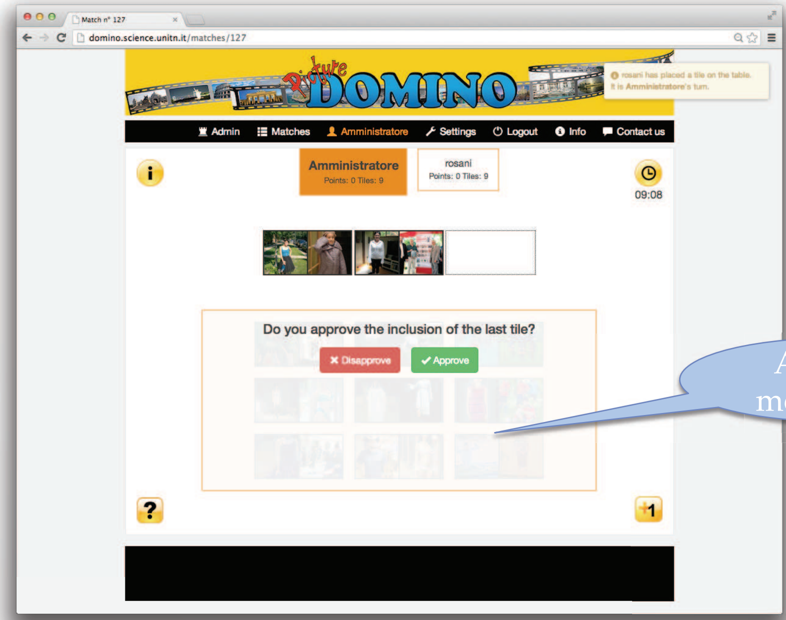


The Domino Game

- Web interface: game page

The Domino Game

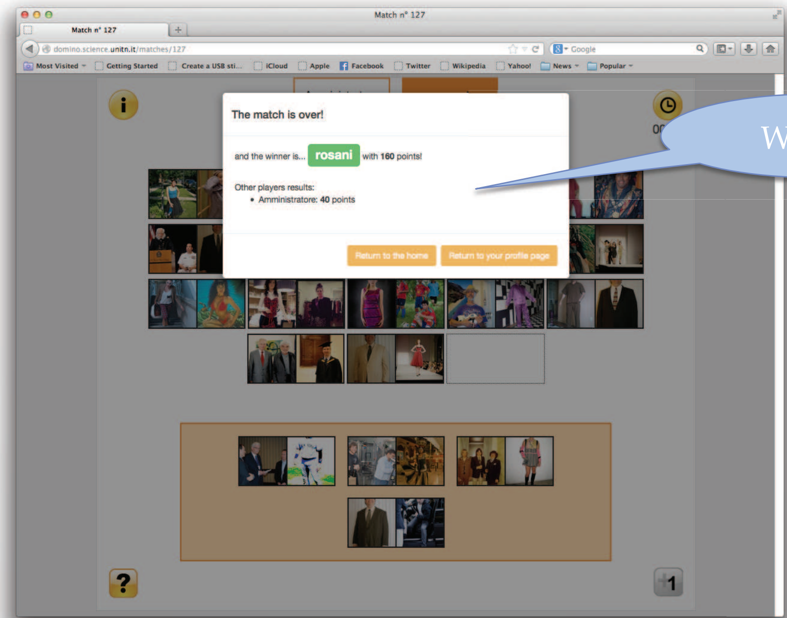
- Web interface: game page



Approve mechanism

The Domino Game

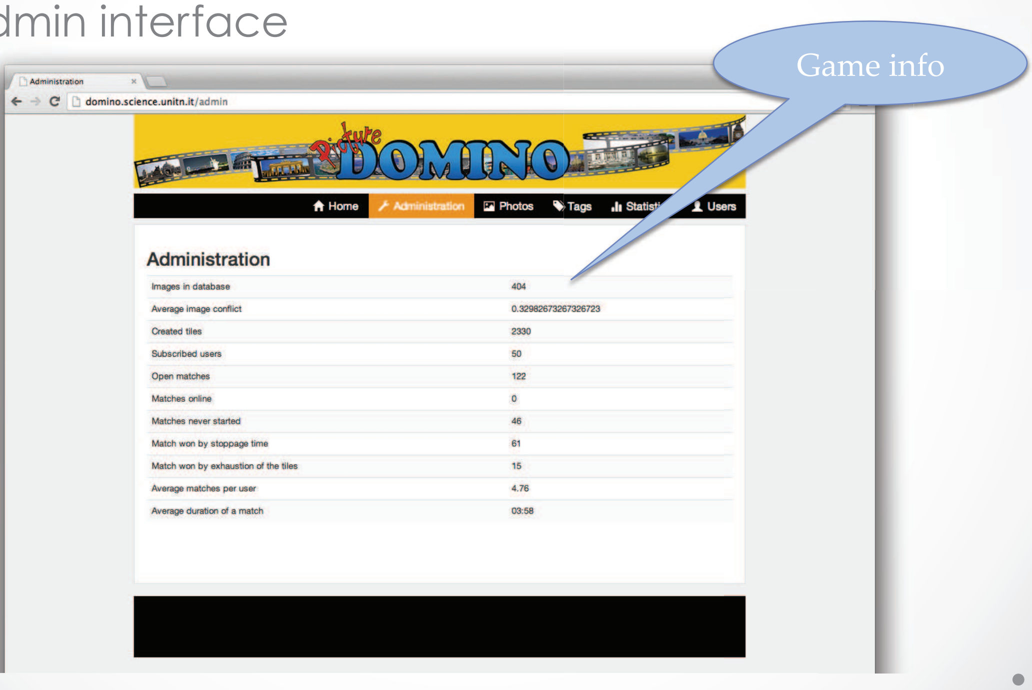
- Web interface: game page



Winner!

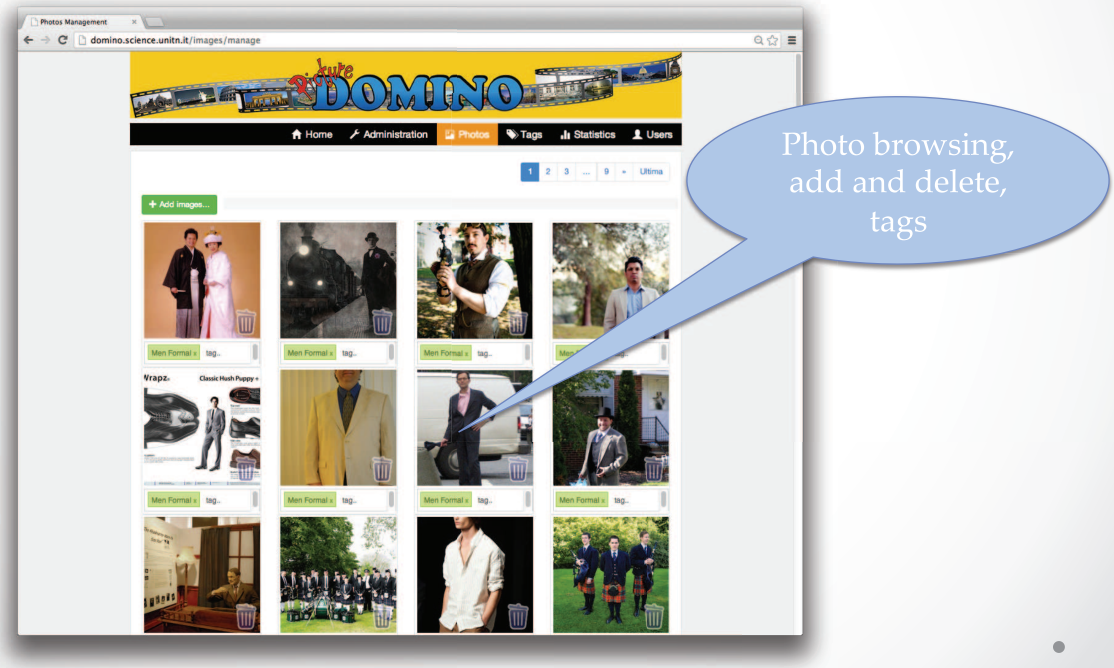
The Domino Game

- Admin interface



The Domino Game

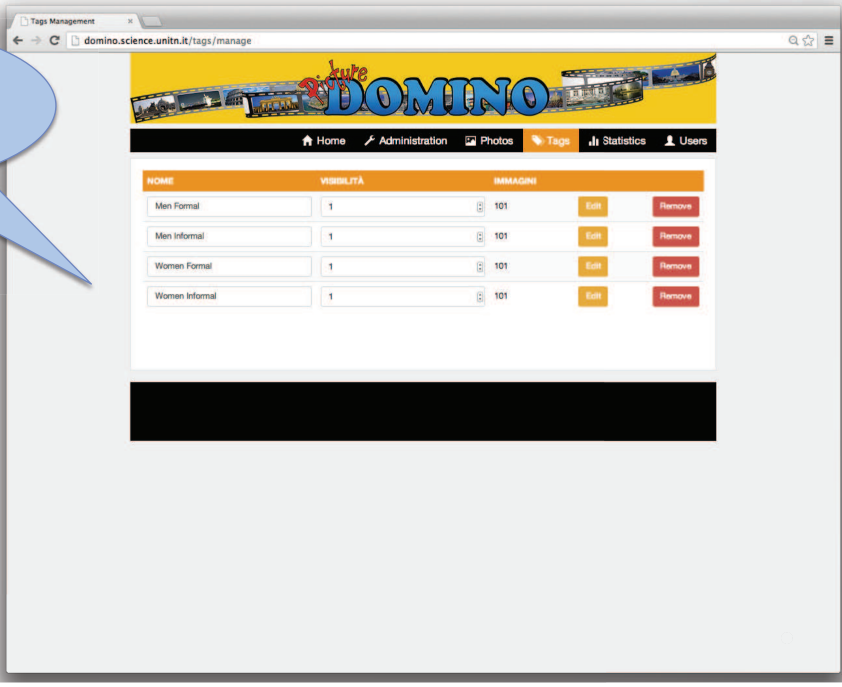
- Admin interface



The Domino Game

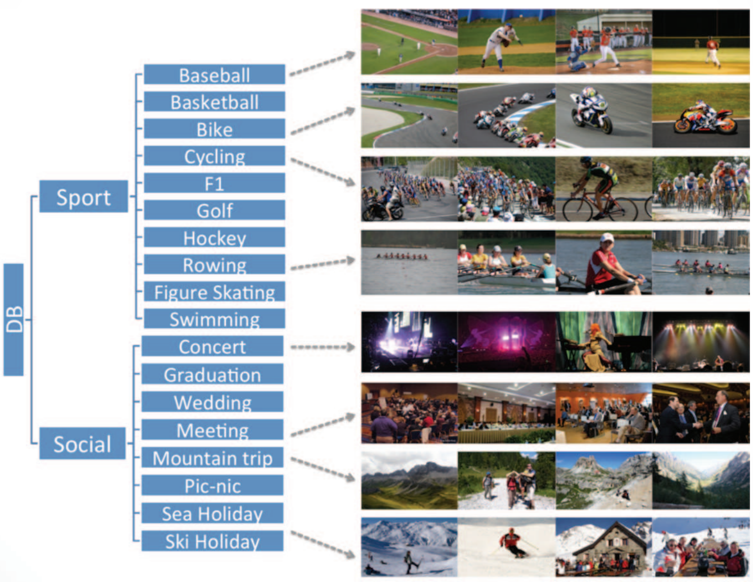
- Admin interface

Tag management



Domino Game: Datasets

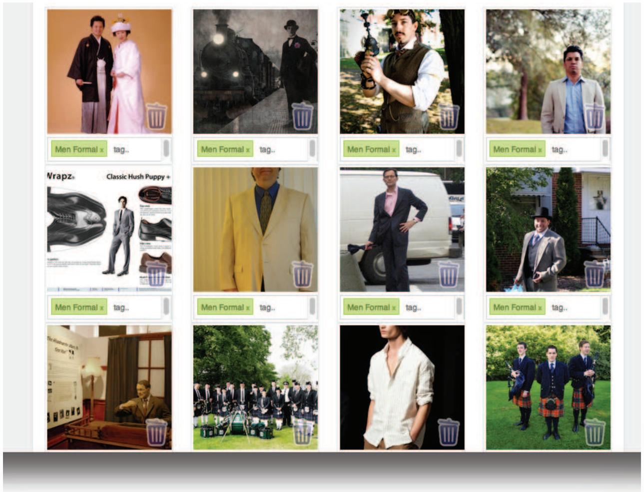
- Dataset EiMM



Riccardo Mattivi, Jasper Uijlings, Francesco G.B. De Natale, and Nicu Sebe. 2011. Exploitation of time constraints for (sub-)event recognition. In Proceedings of the 2011 joint ACM workshop on Modeling and representing events (J-MRE '11)

Domino Game: Datasets

- Dataset Fashion
  - 10.000+ images
  - 4 tags (over 100+) : men formal/informal, women formal/informal



B. Loni, M. Menendez, M. Georgescu, L. Galli, C. Massari, M. Melenhorst, M. Larson, I. Altingovde, R. Vliegendhart, D. Martinenghi Fashion-focused Creative Commons Social dataset ACM Multimedia Systems (MMSys 2013), 2013.

Domino Game: results

- Preliminary results available on Fashion dataset, EiMM dataset under test
  - About 75% images are correctly classified
  - Remaining 30% show conflicting classification: majority mechanisms can improve the performance, BUT…

    … some images are difficult to classify even for human beings

Domino Game: what's next

- Almost unlimited applications (changing datasets, linking models, etc.)
- Extensible to more complex situations (e.g., possibility of creating branches, tiling, etc.)
- Suitable for mobile applications (smartphone version under development)
- Other incentivization models (besides entertainment) under definition:
  - Building stories
  - Sharing media about similar events, hobbies, experience, …

Conclusions

- Current tools for automatic event-media analysis are still unreliable
  - Capturing humans' semantics is still a major challenge
  - Transferring such knowledge in a new generation of automatic tools is the next step
- The different alternatives to do this are:
  - Explicit knowledge (expert systems, taxonomies, ontologies, models, etc.)
  - Implicit knowledge (machine learning)
  - Mixed approach (capturing human knowledge in an indirect way)
- Gamification can be an effective way to implement the mixed approach
  - It may involve crowds, communities, social networks, etc.
  - It may guarantee unbiased, volunteering user support
  - Appropriate usage of acquired knowledge is important as well