UNIVERSITÀ DEGLI STUDI FIRENZE

Scuola di Ingegneria

Corso di Laurea Magistrale in Ingegneria Informatica

# Improving WATSS web application with Computer Vision techniques

*Visual and Multimedia Recognition*

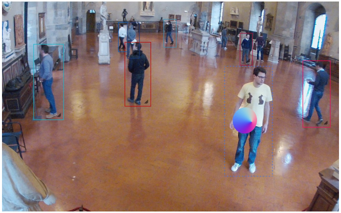
Lorenzo Cioni

Anno Accademico 2015/2016

# Introduction

WATSS, **Web Annotation Tool for Surveillance Scenarios**, is a web-based annotation tool developed for annotating dataset in surveillance systems [1].

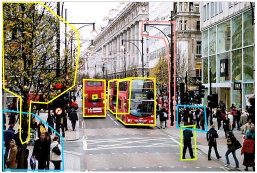
**Main goal**: improve WATSS with some **Computer Vision approaches**, in order to make easy for users to use this tool and make the annotation process more *automatic*.



# Comparative analysis of annotation tools

**LabelMe** [4]

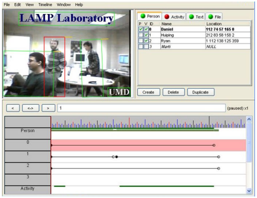
- **Web-based** tool, also released as mobile application
- Annotate scenes with **polygonal shapes**
- **Nested** objects and **occlusion** annotations
- *Zoom in* and *out* of the scene



# Comparative analysis of annotation tools

## ViPER-GT

- **Java application** tool
- Annotate scenes with **geometrical shapes**
- **Timeline** and **annotation highlighting** on time change
- Linear **interpolation** between annotations
- *Zoom in* and *out* of the scene



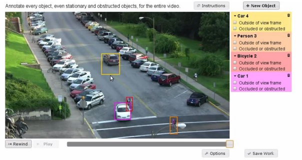
| P | V | ID | Name | Location |
|---|---|---|---|---|
| ☑ | ☑ | **0** | **Daniel** | **112 74 57 165 0** |
| □ | ☑ | 1 | Huiping | 212 83 58 158 2 |
| □ | ☑ | 2 | Ryan | 1 112 138 125 359 |
| □ | □ | *3* | *Marti* | *NULL* |

1

# Comparative analysis of annotation tools

## **VATIC** [3]

- **Online** tool
- Developed for **object detection**
- **Crowd-sourcing** to Amazon's *Mechanical Turk*
- Multiple **plugins**: *object tracking*, *sentence annotation*, etc.

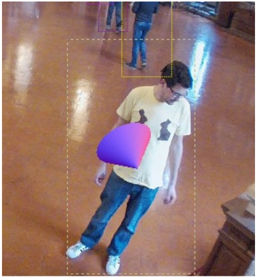


Annotate every object, even stationary and obstructed objects, for the entire video.

# Comparative analysis of annotation tools

## WATSS

- **Web-based** tool
- Annotation with bounding box
- **Occlusion** area
- Coarse **gaze** estimation
- **Groups** and **POI** under observation
- **Multiple cameras** manager

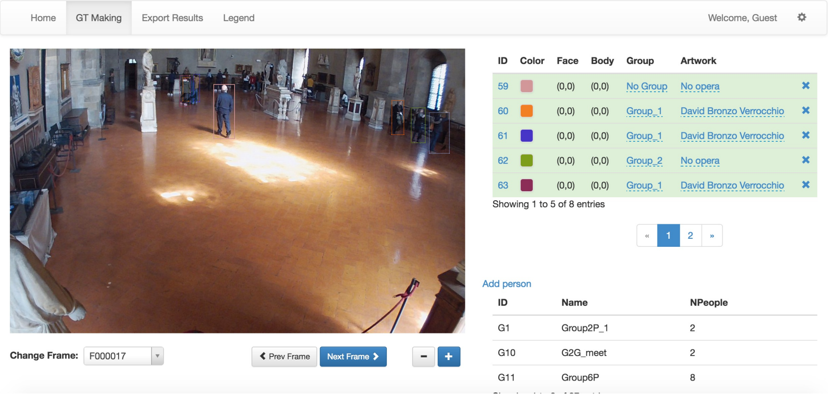


## Improvements

- **User interface** renovation
- Simpler annotation making and editing
- Video **timeline** for annotations
- Automatic **proposals** generation
- **Scene geometry-based** enhancement
- Easy **setup** process

# User interface renovation

## The **old** WATSS user interface



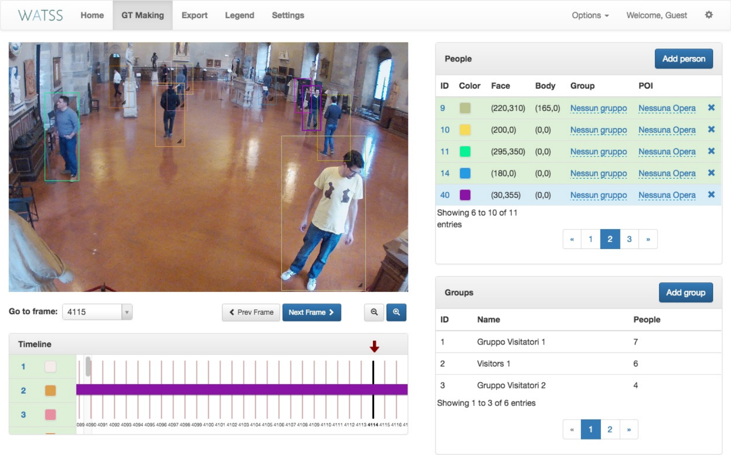
| ID | Color | Face | Body | Group | Artwork | |
|---|---|---|---|---|---|---|
| 59 | 🟥 | (0,0) | (0,0) | No Group | No opera | ✖ |
| 60 | 🟧 | (0,0) | (0,0) | Group_1 | David Bronzo Verrocchio | ✖ |
| 61 | 🟦 | (0,0) | (0,0) | Group_1 | David Bronzo Verrocchio | ✖ |
| 62 | 🟩 | (0,0) | (0,0) | Group_2 | No opera | ✖ |
| 63 | 🟥 | (0,0) | (0,0) | Group_1 | David Bronzo Verrocchio | ✖ |

Showing 1 to 5 of 8 entries

| ID | Name | NPeople |
|---|---|---|
| G1 | Group2P_1 | 2 |
| G10 | G2G_meet | 2 |
| G11 | Group6P | 8 |

# User interface renovation

## The **new** WATSS user interface



People

| ID | Color | Face | Body | Group | POI |
|---|---|---|---|---|---|
| 9 | 🟫 | (220,310) | (165,0) | Nessun gruppo | Nessuna Opera |
| 10 | 🟨 | (200,0) | (0,0) | Nessun gruppo | Nessuna Opera |
| 11 | 🟩 | (295,350) | (0,0) | Nessun gruppo | Nessuna Opera |
| 14 | 🟦 | (180,0) | (0,0) | Nessun gruppo | Nessuna Opera |
| 40 | 🟪 | (30,355) | (0,0) | Nessun gruppo | Nessuna Opera |

Showing 6 to 10 of 11 entries

« 1 2 3 »

Groups

| ID | Name | People |
|---|---|---|
| 1 | Gruppo Visitatori 1 | 7 |
| 2 | Visitors 1 | 6 |
| 3 | Gruppo Visitatori 2 | 4 |

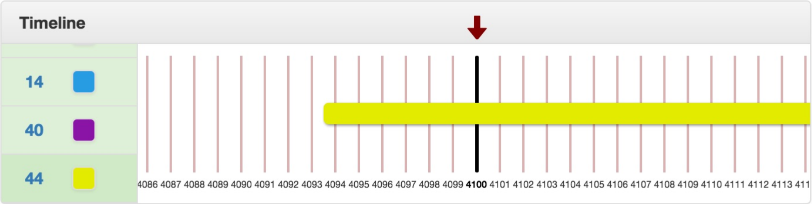
Showing 1 to 3 of 6 entries

« 1 2 »

Go to frame: 4115

# Video timeline

In the video **timeline** all the video frames are shown, coloring the ones with at least one annotated person.



Selecting a person in the list makes the timeline display its **history** highlighting frames where it is present. It is possible to navigate video frames by clicking on it.

# Proposals generation

It is possible to generate **proposals** for a person in some selected frames based on previous annotation of it using timeline: just click and drag the highlighted annotation.

Proposals generation is based on the combination of three different techniques:

- **Motion detection** using a *background subtractor*
- **Pedestrian detection** using *HOG descriptors*
- **Kalman filter** for the *motion estimation*

# Motion detection

Motion detection is based on a **background subtraction** method: moving objects are detected performing a subtraction between the current frame and a background model of the current scene, obtaining a **foreground mask**.

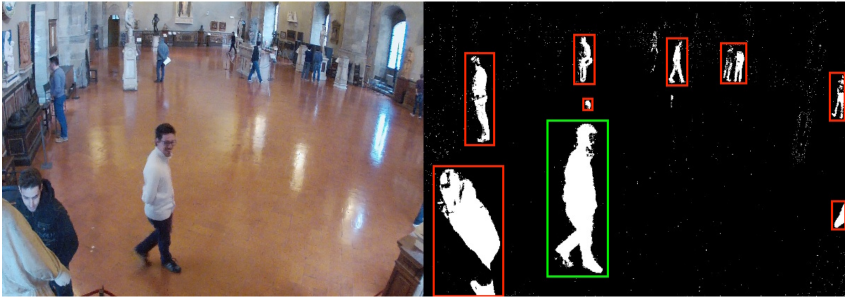
Each pixel of a frame is modeled as a **Mixture of Gaussians** and those which correspond to background colors are selected according to variance and persistence.
Pixel values that do not fit the background distributions are considered part of the foreground.

Background modeling consists of two main steps:

- **Background initialization**: background model evaluation.
- **Background update**: background model is adapted to possible changes in the scene.

# Motion detection

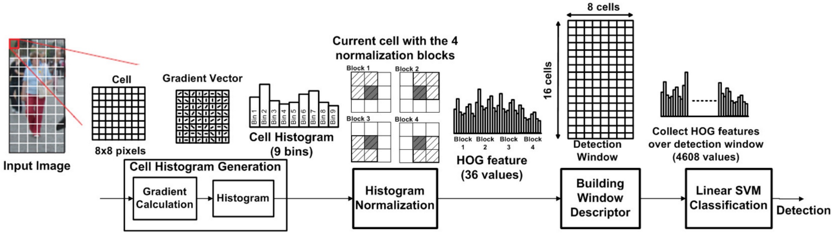
Using the foreground mask, a set of **detections** are extracted based on contours.



Given the previous frame person bounding box, those that do not *overlap* or are *inconsistent* with its dimensions are discarded.

# Pedestrian detector

The used *pedestrian detection* technique is based on **Histogram of Oriented Gradients** and SVM classifier.

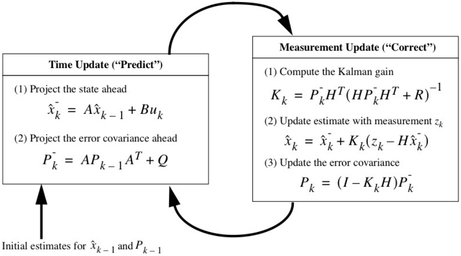


As in the previous case, detections are filtered according to the person history in scene.

Figure from Suleiman, A., Sze, V. J *Signal Process Systems* (2016)

# Kalman filter

A **Kalman filter** is an optimal estimator used for following state estimations based on a set of previous observations.



**Time Update ("Predict")**

(1) Project the state ahead

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_k$$

(2) Project the error covariance ahead

$$P_k^- = AP_{k-1}A^T + Q$$

**Measurement Update ("Correct")**

(1) Compute the Kalman gain

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1}$$

(2) Update estimate with measurement $z_k$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-)$$

(3) Update the error covariance

$$P_k = (I - K_k H)P_k^-$$

Initial estimates for $\hat{x}_{k-1}$ and $P_{k-1}$

A single state in the system considers the **coordinates** $(x, y)$ **of the person** in the current scene, using motion and pedestrian detection for updating the measure.

# Proposals generation

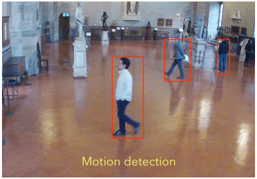
A proposal for a generic frame is the result of the combination of the above descripted methods. Each resulted bounding box is compared with the previous frame annotation in order to evaluate a **score**:

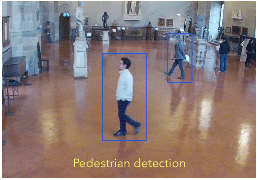
$$\mathit{score}(r) = \frac{\mathit{intersection}(r, p)}{\mathit{union}(r, p)}$$

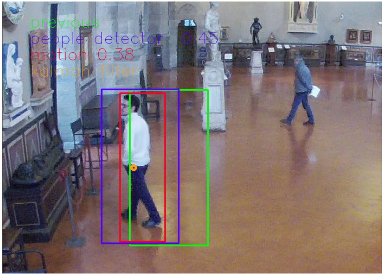
where $r$ is a resulting bounding box (*i.e. the output of the motion detector*) and $p$ is the bounding box of the previous frame.

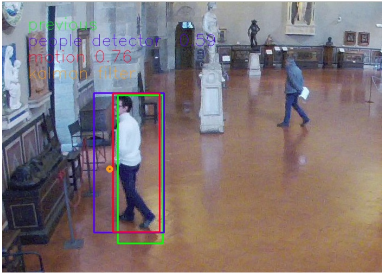
If the motion or the pedestrian detector fails, then the *Kalman filter prediction* is used as the proposal.

# Proposals generation example









# Scene geometry

In the annotation insertion step, it is possible to use the **scene geometry** knowledge in order to generate proposals based on the pointer position.

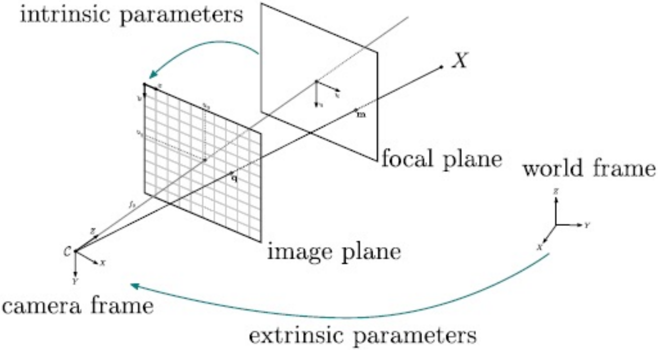
Requirements:

- **Static** cameras
- Camera **calibration parameters**
    - *Intrinsic* parameters, $\mathbf{K}$
    - *Estrinsic* parameters, rotation $\mathbf{R}$ and translation $\mathbf{t}$
    - A *cross-ratio* $\mu$, being projective invariant

# Camera calibration

Given $\mathbf{X} = (X, Y, Z, 1)$ as coordinates in **world**, and $\mathbf{x} = (x, y, 1)$ as coordinates in **scene**:

$$\mathbf{x} = K[R|t]X$$



# Human height estimation

From the camera parameters, it is possible to evaluate the **vanishing line l** and the **vanishing point v**

$$l = P * [0, 0, 1]'$$

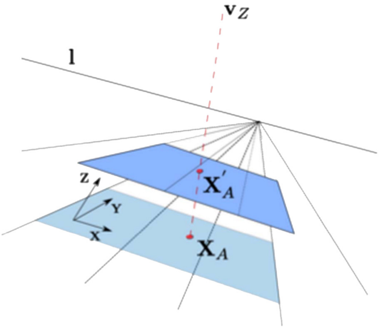
$$v = ((K')^{-1} * K^{-1}) * l$$

$$W = I + (\frac{1}{(1-\mu)} - 1) * \frac{v * l'}{v' * l}$$

Given the head position $head = (x, y, 1)$ and **W**:

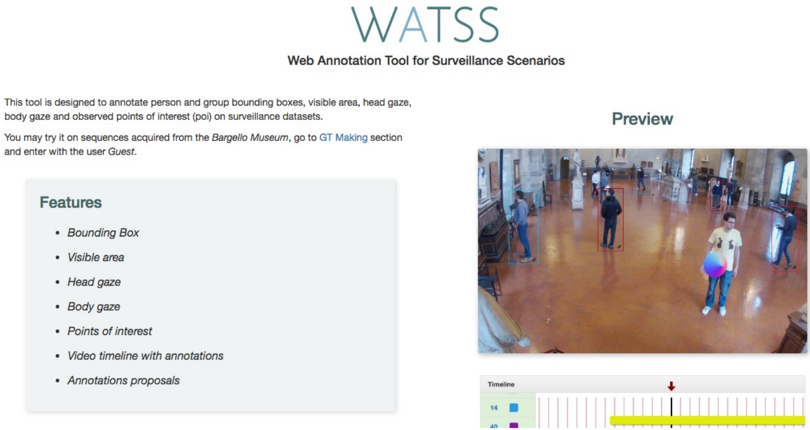
$$feet = W^{-1} * head$$

$$height = |head_y - feet_y|$$



# Demo

Time for a **demo**!



WATSS

Web Annotation Tool for Surveillance Scenarios

This tool is designed to annotate person and group bounding boxes, visible area, head gaze, body gaze and observed points of interest (poi) on surveillance datasets.

You may try it on sequences acquired from the *Bargello Museum*, go to GT Making section and enter with the user *Guest*.

## Features

- *Bounding Box*
- *Visible area*
- *Head gaze*
- *Body gaze*
- *Points of interest*
- *Video timeline with annotations*
- *Annotations proposals*

## Preview

## Conclusions

WATSS application has been improved with **new features** and a renewed **user interface**.

Some features are based on Computer Vision techniques, as automatic annotation proposals generation and scene geometry-based annotation insertion.

**Future developments**:

- Use scene geometry in proposals generation
- Using different people detectors, e.g. an **upper body detector**, for more accurate proposals
- Different types of **annotation shapes**, e.g circle, ellipsis, polygon etc.

# References

[1] F. Bartoli, L. Seidenari, G. Lisanti, S. Karaman, and A. Del Bimbo. *Watts: A web annotation tool for surveillance scenarios*. In *Proceedings of the 23rd ACM International Conference on Multimedia, MM '15*, pages 701–704, New York, NY, USA, 2015. ACM.

[2] F. Bartoli, G. Lisanti, L. Seidenari, S. Karaman, and A. Del Bimbo. *Museumvisitors: A dataset for pedestrian and group detection, gaze estimation and behavior understanding*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 19–27, 2015.

[3] C. Vondrick, D. Patterson, and D. Ramanan. *Efficiently scaling up crowdsourced video annotation*. Int. J. Comput. Vision, 101(1):184–204, Jan. 2013.

[4] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. *Labelme: A database and web-based tool for image annotation*. Int. J. Comput. Vision, 77(1-3):157–173, May 2008.

[5] A. Del Bimbo, G. Lisanti, and F. Pernici. *Scale invariant 3d multi-person tracking using a base set of bundle adjusted visual landmarks*. In Computer Vision Workshops (ICCV Work- shops), 2009 IEEE 12th International Conference on, pages 1121–1128. IEEE, 2009.