Project ORUSSI
(Optimal Road sUrveillance System based on Scalable vIdeo)

The growing mobility of people and goods has a very high societal cost in terms of traffic congestion and of fatalities and injured people every year. The management of a road network needs efficient ways for assessment at minimal costs. Road monitoring is a relevant part of road management, especially for safety, optimal traffic flow and for investigating new sustainable transport patterns. On the road side, there are several technologies used for collecting detection and surveillance information: sophisticated automated systems such as in-roadway or over-roadway sensors, closed circuit television (CCTV) system for viewing real-time video images of the roadway or road weather information systems for monitoring pavement and weather. Current monitoring systems based on video lack of optimal usage of networks and are difficult to be extended efficiently.

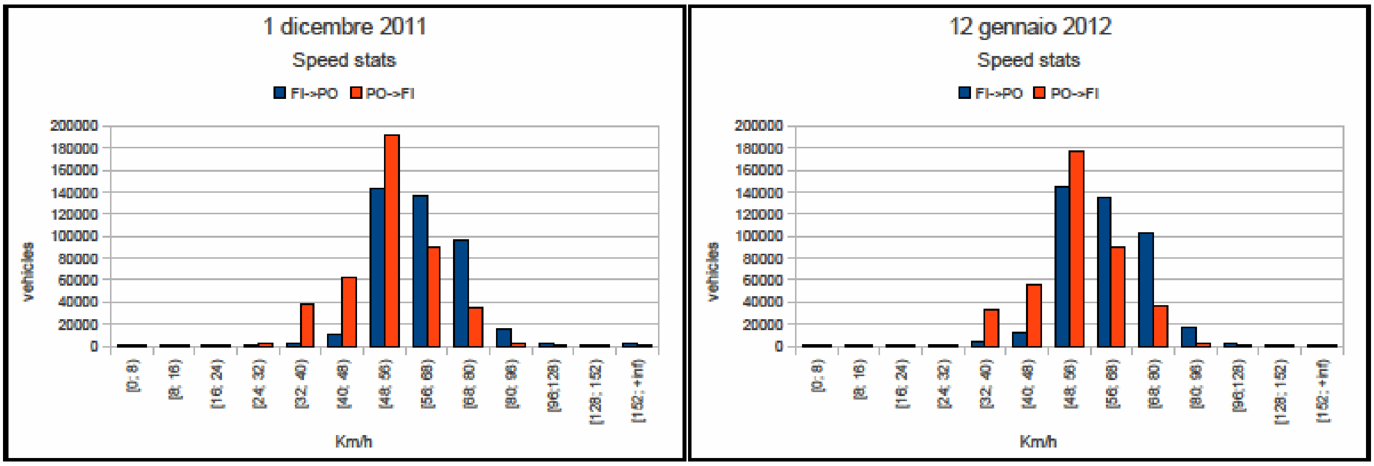
Project ORUSSI (Optimal Road sUrveillance System based on Scalable vIdeo)[1] focuses on road monitoring through a network of roadside sensors (mainly cameras) that can be dynamically deployed and added to the surveillance systems in an efficient way. The main objective of the project is to develop an optimized platform offering innovative real-time media (video and data) applications for road monitoring in real scenarios. The project will develop a novel platform based
on the synergetic bundling of current research results in the field of semantic transcoding, the recently approved standard Scalable Video Coding standard (SVC), wireless communication and roadside equipment.
In the scope of this project we present embedded solutions for video analysis applied to road surveillance. Our solution exploits the Axis(tm) Developer Platform which is designed to develop custom image analysis applications on video servers and smart cameras equipped with a low powered and low-performance processor (ARTPEC-3). The development on this platform presents several challenges since ARTPEC-3 CPU does not employ any FPU, and there are no high-level computer vision or image analysis libraries. Our embedded video analysis solutions aim at computing low cost visual features for typical road surveillance applications like adaptive video compression and vehicle monitoring.

In the course of this project we also collected and annotated a large dataset of road surveillance video streams. We gathered around four days of video at different resolutions (spanning from 320x240 to 1280x800), with different lighting and weather conditions. We annotated the dataset for two tasks: car counting, and vehicle class recognition. For the first task the whole dataset is annotated; for the latter we labeled 5000 vehicle images dividing them in 4 classes: cars, motorbikes, trucks, unknown. Since vehicles are extracted with an automatic procedure and this may fail in extracting a well defined vehicle we label frames with partial vehicles or containing one or more vehicles as such.

The prototyped application has been deployed in a real-world installation, in collaboration with the municipality of Prato. The system (which has been up and running for about 10 months so far) is being used to monitor the traffic in a nodal intersection, continuously providing information about vehicle entering and exiting the city at every hour of day and night. Further experimental installations have been planned, also in conjunction with speed traps that will allow to evaluate the performance of the video-based speed

[1] Project ORUSSI is funded on the ERA-SME program (6th call), and is a joint collaboration between the MICC department of the University of Florence and Magenta srl.

estimation. The histograms displayed in the figure below show speed statistics collected on one day of operation.



Many critical video streaming applications require transmission of many streams over limited bandwidth. These applications have several things in common, among them the need to deliver reasonably high quality video from multiple cameras spread over large areas and to accomplish this using limited bandwidth. Using a combination of low-level features with low computational cost, we show how it is possible to control the quality of video compression so that semantically meaningful elements of the scene are encoded with higher fidelity, while background elements are allocated fewer bits in the transmitted representation. Our approach is based on adaptive smoothing of individual video frames so that image features highly correlated to semantically interesting objects are preserved. Using only low-level image features on individual frames, this adaptive smoothing can be seamlessly inserted into a video coding pipeline as a pre-processing state.

We will also present a distributed solution to control network cameras and video analysis servers in order to tune computer vision algorithms and camera setups in the field by evaluating the results and the camera setups directly in the scene. Our application is a front-end for a SOAP based service and can automatically generate the correct widget to control algorithms parameters based on the service offered by the SOAP back-end. The application is based built with modern web technologies and has been easily ported on mobile devices like tablets and smart-phones. The following picture shows the architecture of the overall system.