The S-HOCK Dataset: Analyzing Crowds at the Stadium

Davide Conigliaro[1,3], Paolo Rota[2], Francesco Setti[3], Chiara Bassetti[3], Nicola Conci[4], Nicu Sebe[4], Marco Cristani[1],
[1]University of Verona. [2]Vienna Institute of Technology. [3]ISTC–CNR (Trento). [4]University of Trento.

The topic of crowd modeling in computer vision usually assumes a single generic typology of crowd, which is very simplistic. In this paper we adopt a taxonomy that is widely accepted in sociology, focusing on a particular category, the *spectator crowd*, which is formed by people "interested in watching something specific that they came to see" [1]. This can be found at the stadiums, amphitheaters, cinema, etc. In particular, we propose a novel dataset, the Spectators Hockey (S-HOCK), which deals with 4 hockey matches during an international tournament.

The dataset is unique in the crowd literature, and in general in the surveillance realm. The dataset analyzes the crowd at different levels of detail. At the highest level, it models the network of social connections among the public (who knows whom in the neighborhood), what is the supported team and what has been the best action in the match; all of this has been obtained by interviews at the stadium. At a medium level, spectators are localized, and information regarding the pose of their heads and body is given. Finally, at a lowest level, a fine grained specification of all the actions performed by each single person is available. This information is summarized by a large number of annotations collected over a year of work: more than 100 millions of double checked annotations. This permits potentially to deal with hundreds of tasks, some of which are documented in the full paper.

Furthermore, the dataset is multidimensional, in the sense that offers not only the view of the crowd (at different resolutions, with 4 cameras) but also on the matches. This multiplies the number of possible applications that could be assessed, investigating the reactions of the crowd to the actions of the game, opening up to applications of summarization and content analysis. Besides these figures, S-HOCK is significantly different from all the other crowd datasets, since the crowd as a whole is mostly static and the motion of each spectator is constrained within a limited space in the surrounding of his position.

Annotation	Typical Values
People detection	full body bounding box $[x, y, \text{width}, \text{height}]$
Head detection	head bounding box $[x, y, \text{width}, \text{height}]$
Head pose*	left, frontal, right, away, down
Body position	sitting, standing, (locomotion)
Posture	crossed arms, hands in pocket, crossed legs ...
Locomotion	walking, jumping (each jump), rising pelvis slightly up
Action / Interaction	waving arms, pointing toward game, applauding, ...
Supported team	the team supported in this game
Best action	the most exciting action of the game
Social relation	If he/she did know the person seated at his/her right

Table 1: Some of the annotations provided for each person and each frame of the videos.

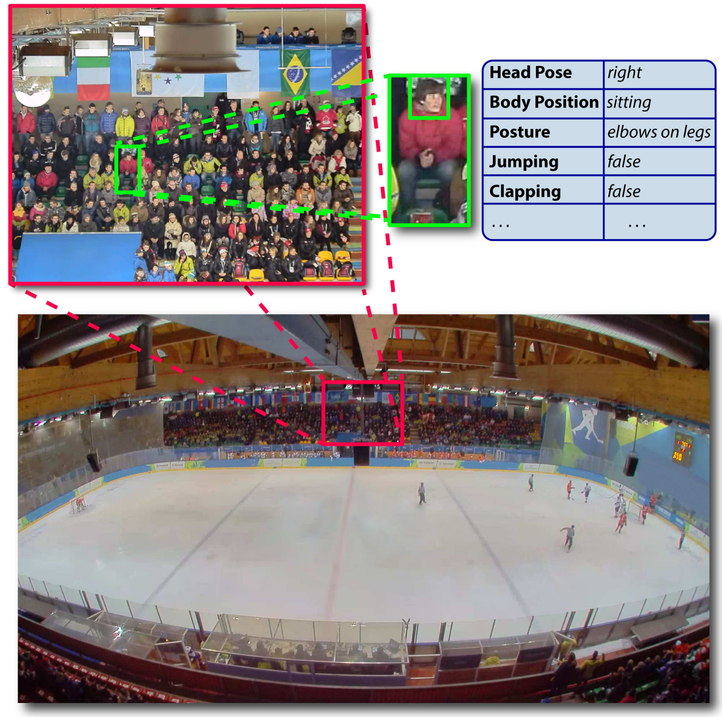

Figure 1: Example of images collected for both the spectators and the rink, plus the annotations.

Together with the annotations, in the paper we discuss issues related to low and high level detail of the crowd analysis, namely, people detection and head pose estimation for the low level analysis, and the spectator categorization for the high level analysis. For all of these applications, we define the experimental protocols, promoting future comparisons.

For people detection task we provide five different baselines, from the simplest algorithms to the state of the art method for object detection, showing how in this scenario the simplest method gets very high scores.

Regarding head pose estimation, we tested two state of the art methods which work in a low resolution domain. Furthermore, we propose two novel approaches based on Deep Learning. In particular, we evaluate the performance of the Convolutional Neural Network and the Stacked Auto-encoder Neural Network architecture. Here the results are comparable with state of the art but are obtainable at a much higher speed.

Spectator categorization is a kind of crowd segmentation, where the goal is to find the team supported by each spectator. This task is intuitively useful to segregate the different supporter teams, and individuates "hot" zones in which the two teams are mixed. We tested two state of the art methods aiming at segment pedestrian's flow by adjusting them in order to be applied on a spectator crowd. Then we propose a new framework for spectator categorization based on optical flow which outperforms the previous methods.

From the experiments we conducted, we show how standard methods for crowd analysis, which work well on state-of-the-art datasets, do not fit the type of data we are dealing with. Anyway, the performances are far from being errorless, and this witnesses the difficulty of the problem and the fact that much can be done in the future.

Concluding, the contributions of our work are:

- A novel dataset for spectator crowd, which describes at different levels of detail the crowd behavior with millions of ground truth annotations, synchronized with the game being played in the field. Crowd and game are captured with different cameras, ensuring multiple points of view;
- A set of tasks for analyzing the spectator crowd, some of them are brand new;
- A set of baselines for some of these tasks, with novel approaches which definitely overcome the standard crowd analysis algorithms.

We are confident that S-HOCK may trigger the design of novel and effective approaches for the analysis of human behavior in crowded settings.

[1] A.E. Berlonghi. Undestanding and planning for different spectator crowds. *Safety Science*, 18:239–247, 1995.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.