ASSISTED MEDIA FILTERING

SAM Z. GLASSENBERG, DAVID LOWSKY, SAM PEARLMAN, SEBASTIAN THRUN
Stanford University Department of Computer Science
E-mail: glassenb@stanford.edu, dlowsky@stanford.edu, sampearl@stanford.edu, thrun@stanford.edu

Media broadcasts often require personally-identifiable visual information to be obfuscated to preserve anonymity of witnesses, suspects, and minors. Currently, this process requires a manual post-processing step, incurring a significant delay that prevents content from being televised live. The media filtering problem imposes unique constraints on the object recognition problem, particularly the inability to pre-train the system from multiple views of the target object. Using scale-invariant feature transforms, a clustering system that can automatically identify and obfuscate the target object in subsequent video frames was developed. Using dynamic learning of target features along with "target" and "decoy" feature databases and a weighted voting scheme, the system maintains awareness of similar subjects, avoiding obfuscation of incorrect objects while tracking the target. The system uses an ellipsoid approximation of the object to track through 3D rotation and an active contour correction of its projection onto the image plane to determine the feature learning region in each frame. The system demonstrates the effectiveness of using SIFT features to track human faces as well as false objects. The system generates smooth, continuous movement of the obfuscation region across frames.

1 Introduction

Live media broadcasts often require personally-identifiable visual information to be obfuscated to preserve anonymity (faces, license plates, addresses, etc.). This is especially true when video footage includes minors or witnesses. To do so for a stationary interview is simple. However, if motion relative to the camera is involved, a painstaking human post-processing step is currently required before broadcast: the subject must be manually censored. This results in a significant delay and prevents many programs from live airing.

The right and desire of the viewing public to see events broadcast in real-time must be balanced with an individual's right to privacy. This conflict is most pronounced in broadcasts of sensitive events involving legal culpability such as police actions and court proceedings. Viewers demand live broadcast of these events, yet the privacy of suspects, victims and witnesses must be preserved. It has been concluded that "Broadcasting the identity of a crime victim most often only adds to the person's grief, anguish and trauma" [CBC03], while broadcasting the identity of a suspect can jeopardize the fairness of criminal proceedings. Governments have sought reasonable compromise [TEXAS02], but the conflict remains, exacerbated by the fact that preservation of anonymity demands a broadcast delay of minutes to hours.

Figure 1: Pixilated image of a 13-year-old murder suspect turning himself in to the police (the youth's face has been obscured because he is a juvenile)

Preservation of privacy is not necessarily guaranteed by a system limited to facial occlusion. Additional scenes may require obfuscation of other personally-identifiable information (PII), such as license plates or addresses. By enabling a camera-operator to identify PII while filming, automatic object obfuscation could begin, allowing the output of such a system to be broadcast live. A courtroom scene could be broadcast live without the risk of privacy violation if the camera or protected subject moved.

In the live broadcast scenario, the primary objective is to enable a camera operator to rapidly locate and specify an object in the first frame, then automatically track and obfuscate the object throughout the stream. This scenario presents some unique conditions that differentiate it from other face and object recognition tasks:

- No pre-training. Data about the target object must be acquired immediately prior to broadcast in the first set of frames. Only one view of the complex 3D object is available. Target objects are present at the start of broadcast, and can be specified by a camera operator using an integrated interface.
- Scene transformation. In live broadcast, the camera and object position and orientation are independently dynamic and unpredictable. Tracking must be invariant to translation, rotation, scale and lighting changes.
- Rotation. The features in subsequent frames may reflect a completely different region of the primary target object due to rotation.
- Reacquisition. When the target object is temporarily occluded, out-of-focus, rotated out-of-view or out of the scene in subsequent frames, no obfuscation is required. In all cases reacquisition must occur immediately and obfuscation resumed once the object re-enters the visible scene.
- Differentiation. Similar objects can enter and leave the scene throughout this process, yet the system should consistently track only the target object.
- Temporal coherence. While movement in the scene may be rapid, reasonable temporal coherence between object features can be assumed. Data is recorded at 30 frames-per-second, and it can be assumed that the camera operator will be professional and deliberate.
- 2D result. The end result need not be a full 3D reconstruction or 3D transformation. The goal is the obfuscation of the element's identifiable features in image-space.

Existing solutions deal separately with variations of two basic problems: tracking and object recognition. While these methods may be partially applicable, our specific global tracking problem differs in several significant ways from the problems addressed by existing solutions.