Tracking

Hao Guan (管皓)

School of Computer Science
Fudan University
2014-09-29
Tracking in Multimedia
Multimedia

• Video
• Audio
Video Tracking

• Use your eyes
Audio Tracking

• Use your ears
Video Tracking

Object Tracking

Tracking

Visual Tracking

Video Tracking
Definition

Given a bounding box defining the initial position of an object in a single frame, automatically determine the object’s bounding box in the following frames or indicate the object is not visible.
Why important?

• An important Mid-level of a vision system
Why important?

• One of the most practical areas of CV
Why difficult?

- Illumination

- Occlusion
Why difficult?

- Pose variants

- Clutter
Why difficult?

• Scale variant
Why difficult?

- Fast Motion
Categories

- Single camera
- Multiple camera
- Re-identification
Categories

- Static camera
- moving camera
Categories

- Single Object
- Multiple Object
Categories

- Visible
- Infrared
Categories

- Rigid Object
- Non-rigid Object
Evaluation

- Center Location Error

average Euclidean distance between the center of the tracked target and the ground truth in all the frames of one video.
Evaluation

• Success Rate

The success rate is the ratio of the frames whose scores are larger than a given threshold.

\[ \text{score} = \frac{R_t \cap R_g}{R_t \cup R_g} \]
The State-of-the-art trackers

• Tracking by detection is becoming popular.

This stems directly from the development of powerful discriminative methods in machine learning and their application to detection with offline training.

The discriminative trackers try to differentiate the target from the background by taking tracking as a binary classification problem.
Real-Time Compressive Tracking (CT)

• Core idea
  Facilitate an efficient project from the image feature space to a low-dimensional compressed space.

• Theoretical basis
  Compressive sensing theories
  A small number of randomly generated linear measurements can preserve most of the salient information and almost perfect reconstruct the signal.
Feature Extraction

- Dimension reduction
Updating the classifier at the $t$-th frame

Positive and negative samples are used to train a Naïve Bayes Classifier
Tracking at the \((t+1)\)-th frame

The sample which has the highest score will be the tracked position.

\[
H(v) = \log \left( \frac{\prod_{i=1}^{n} p(v_i | y = 1)p(y = 1)}{\prod_{i=1}^{n} p(v_i | y = 0)p(y = 0)} \right) = \sum_{i=1}^{n} \log \left( \frac{p(v_i | y = 1)}{p(v_i | y = 0)} \right)
\]
Tracking at the \((t+1)\)-th frame

\[
H(v) = \log \left( \frac{\prod_{i=1}^{n} p(v_i|y = 1)p(y = 1)}{\prod_{i=1}^{n} p(v_i|y = 0)p(y = 0)} \right) = \sum_{i=1}^{n} \log \left( \frac{p(v_i|y = 1)}{p(v_i|y = 0)} \right)
\]

\[
p(v_i|y = 1) \sim N(\mu_i^1, \sigma_i^1), \quad p(v_i|y = 0) \sim N(\mu_i^0, \sigma_i^0).
\]
Experiments
Experiments
Experiments
Tracking-Learning-Detection (TLD)

- Core idea

Combination of motion tracking and object detection. Focused on long-term tracking.

Use detector to relocate the tracker.
Framework Overview

- re-initialization
- Output
- update
- tracking
- detection
- Video
Tracking Module of TLD

- Frame difference
- Background subtraction
- Optical flow
Optical Flow

• A classical, common, successful method
Optical Flow

- A classical, common, successful method
Median Flow for TLD

• Assumption

A good tracker should have forward-backward consistency.
Median Flow for TLD

- Features can be simple
- Handle scale variants
- Simple and efficient
- Sensitive to illumination variants
- Lack of self-learning and update
Median Flow for TLD
Median Flow for TLD
Median Flow for TLD
Detection Module of TLD

- Full frame scan
  - Detector
    - Candidate windows
  - Variance filter
    - Random Ferns
      - Nearest Neighborhoods
Variance Filter

Remove patches that are smooth.

Effective for the images that have smooth background.
Variance Filter

When background is more complicated, the effect will get worse.
Variance Filter

When object is similar to the background, the effect will get worse.
Random Ferns Classifier

- The core part of the detector of TLD.
- Different from Random Forests:
  - same criterion for every layer, become linear
Feature Extraction

LBP features
Select two points A and B randomly from one patch, compare their intensity, if I(A) > I(B), then the feature value is 1, else is 0.

Example: one patch passes a fern with 5 nodes.
Random Ferns Classifier

If one fern has $s$ nodes, then there will be $1+2^s$ feature values.

The samples of one class pass the fern, we can get the histogram of the prior probability.
Random Ferns Classifier

Samples of different classes pass the fern, we can get corresponding histograms.
Random Ferns Classifier

When a new patch passes the fern, if its feature is 00011 (3) for example, then find the max posterior probability from the given distribution.
Random Ferns Classifier

We usually use a few ferns to form a random ferns classifier. Each fern has a vote.
Random Ferns Classifier

In TLD, 13 nodes each fern, 10 ferns.
Nearest Neighborhood

• Compute the similarity with object models.
Integration

tracking
detection
Conclusions & Future Directions

• There is not a perfect tracker.
  Select the most suitable one according to the application.

• New discriminative features.

• Dynamic and motion analysis.
Conclusions & Future Directions

• Depth information from multi-views.
• Re-identification.
• Integration of Video & Audio tracking
References

Thank you