# Tracking based Text detection and Recognition From Web Videos

Harshada Deokar

Under Guidance Of

Prof.Sawant.V.V

*harshadatd1998@gmail.com*

April 2, 2019

# Introduction

- Text in Videos
- Embedded Caption Text
- Video Text Extraction
- Text Tracking
- Tracking Based Text Detection
- Tracking Based Text Recognition

# Motivations

## Lots Of Knowledge

Text Objects embedded into Videos contains lots of knowledge to the multimedia System

## Source Of Information

Text is the Rich Source of information for Retrieval Application

## Properties

Common Text in videos contains common properties

## Multimedia Understanding

Text Extraction is Important for multimedia understanding

# Literature Survey

## Based on Bayes Theorem

## Region Matching and Text Tracking

## Feedback Information

## Over Segmentation

# Goals And Objectives

## Large Data

Alert The user about large amount of visual data present in video

## Text Extraction

To recognize and fetching out contents of the video

## User Awareness

Creating User Awareness about various text extraction methods

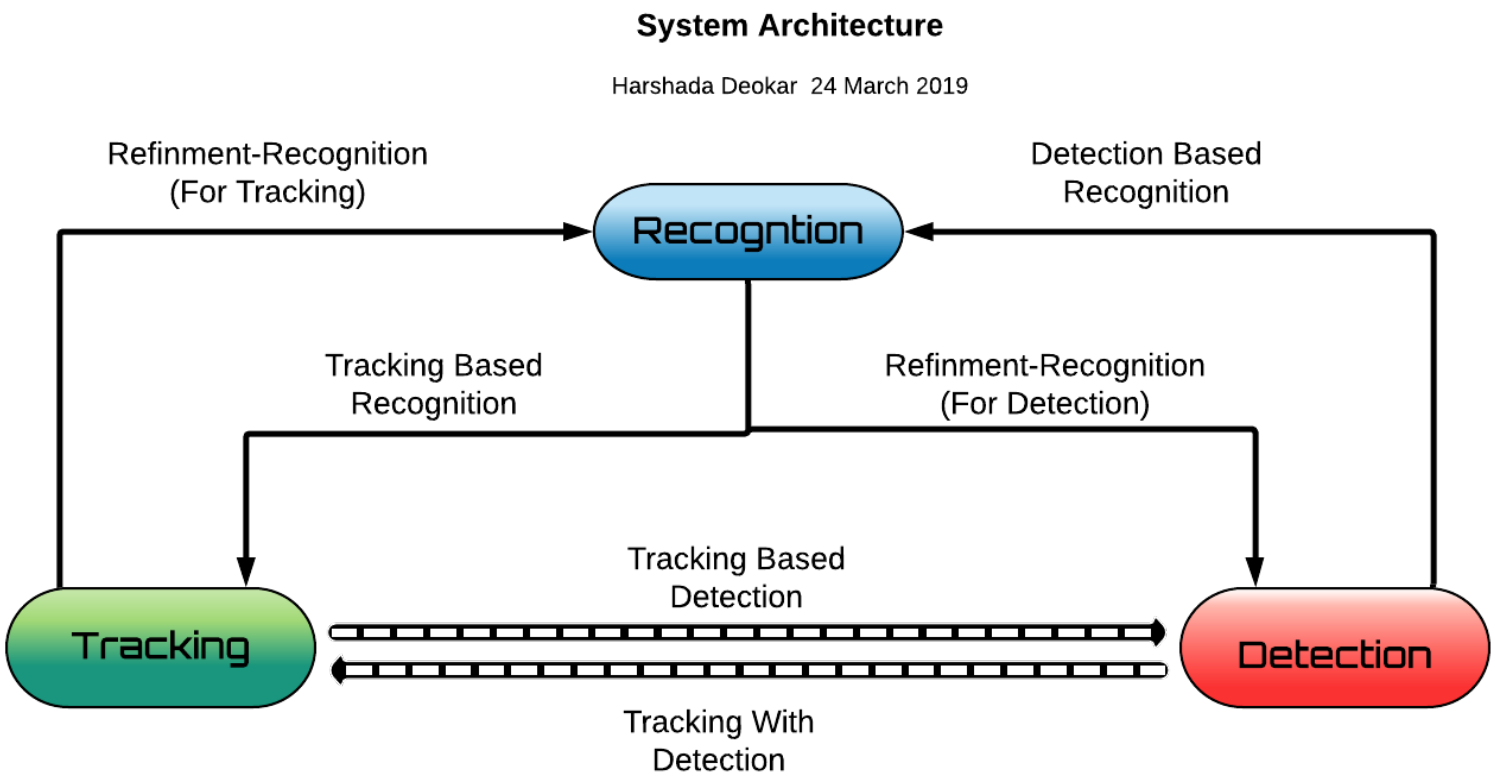
## Video Search

To Provide way for video search in faster time

# Problem Statement

*"To extract the text from web videos using tracking,text-detection and text-recognition"*

# System Architecture



# Baysian Formulation

$$D_t = \{d_{t,1}, d_{t,2} \cdots \cdots d_{t,m_t}\}$$

$$R_t = \{r_{t,1}, r_{t,2} \cdots \cdots r_{t,m_t}\}$$

$$T_t = \{T_{t,1}, T_{t,2} \cdots \cdots T_{t,n_t}\}$$

$$T_{t,i} = \{d_{t,i,1}, d_{t,i,2} \cdots \cdots d_{t,i,p_{t,i}}\}$$

# Bayesian Formulation

The target is
$X_t^* = \arg\max P\left(X_t | X_t^0, T_{t-1}^*\right)$
According to total probability theorem,

$$P(X_t | X_t^0, T_{t-1}^*) = \sum_{T_t} P(X_t | T_t, X_t^0, T_{t-1}^*) P(T_t | X_t^0, T_{t-1}^*)$$

The final detection or recognition results in frame t are
$X_t^* = \arg\max P\left(X_t | T_t^{*,0}, X_t^0, T_{t-1}^*\right)$

# Text Tracking

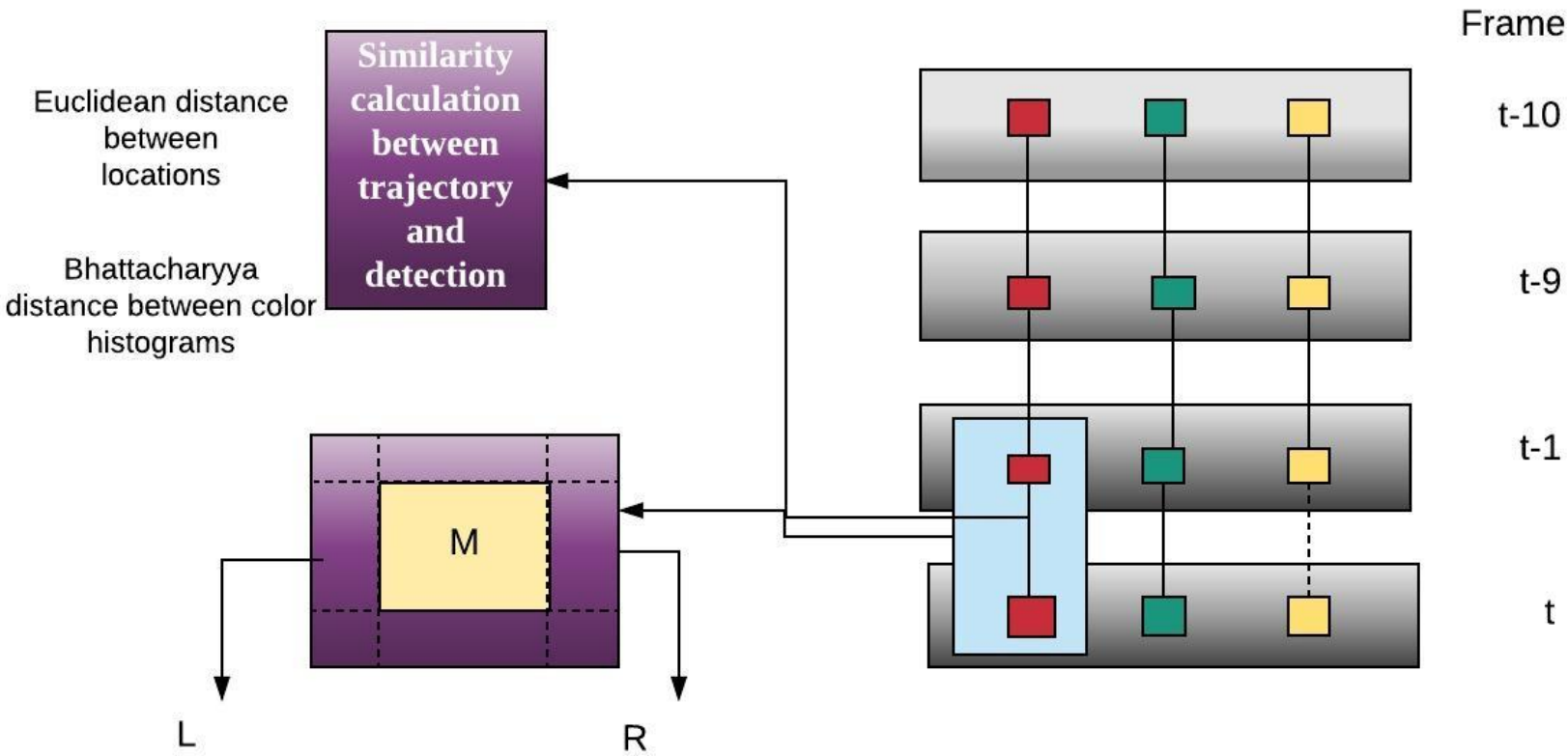



Figure: Text Tracking with tracking-by-detection

# Similarity calculation

**Similiarity:**

$$S\left(T^*_{t-1,i}, d^0_{t,j}\right) = S_a(T^*_{t-1,i}, d^0_{t,j}).S_l(T^*_{t-1,i}, d^0_{t,j})$$

**Appearance Similiarity:**

$$S_a(T^*_{t-1,i}, d^0_{t,j}) = exp\frac{(DB(H(d^*_{t-1,i,final}), H(d^0_{t,j})^2)}{2\alpha_a^2}$$

**Location Similiarity:**

$$S_l\left(T^*_{t-1,i}, d^0_{t,j}\right) = exp\left(\frac{DE\left(l^*_{t-1,i}, l^0_{i,j}\right)^2}{2\alpha_i^2}\right)$$

# Text Detection

*Detection is the task of localizing the text in each video frame with bounding boxes.*

1) Detection Revising by Tracking.
2) Trajectory Updating

# Text Recognition

Recognition is the process of segmenting.

- Temporal over-segmentation.
- Merging
- Boundary Refinement
- Voting

# Algorithm

**Agglomerative hierarchical clustering algorithm**

Let X=x1,x2,x3,.....,xn be the data points.

**Step 1:** L(0) = 0 and m = 0.

**Step 2:**Find (r), (s), according to d[(r),(s)] = min d[(i),(j)]

**Step 3:** m = m +1.Merge (r),(s) for next m. then L(m) = d[(r),(s)].

**Step 4:**Update D, The distance between the new cluster, denoted (r,s) and old (k) d[(k), (r,s)] = min (d[(k),(r)], d[(k),(s)]).

**Step 5:**If all the data points are in one cluster then stop, else repeat from **Step 2**.

# Multi-Frame Integration

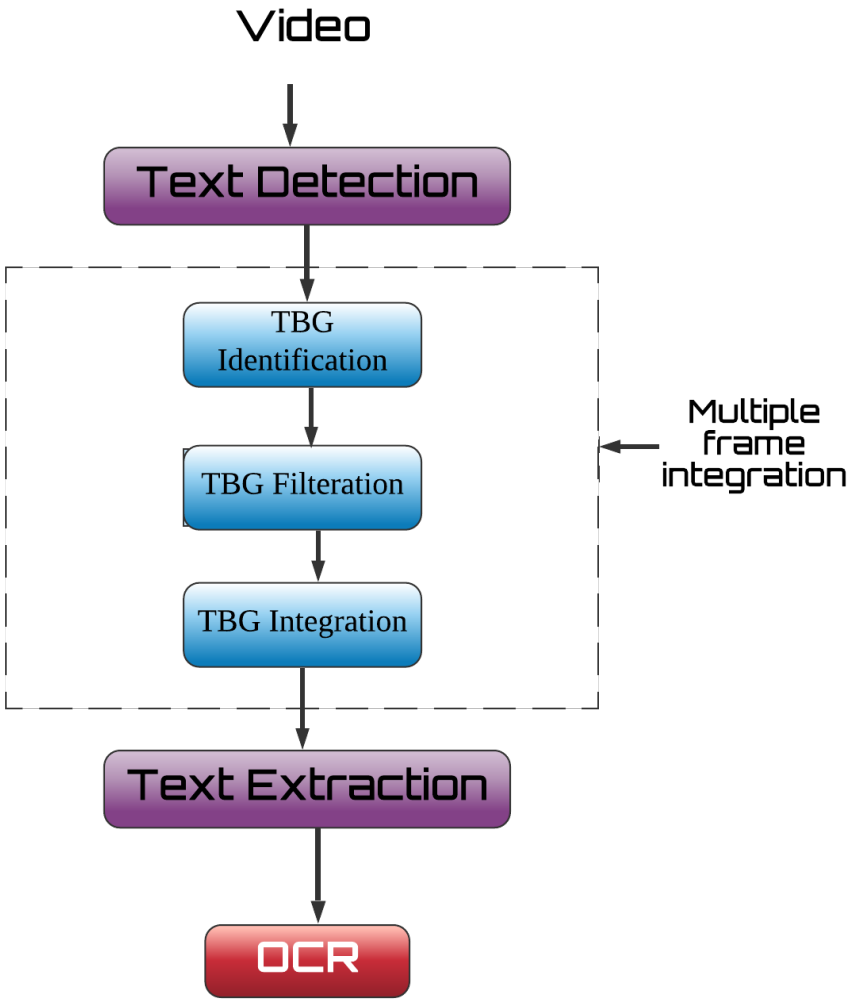



Figure: Multiframe Integration

# Conclusion

# References

Shu Tian,Xu-cheng Yin,

A Unified framework for tracking based text detection and recognition from web videos

*IEEE transactions on pattern analysis and machine intelligence,VOL.,NO.,*

x.-C.Yin,Z.-Y.Zuo,S.Tian, and C.-L.Liu

Text Detection,tracking and recognition in video:A comprehensive survey

*, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL.25 , NO.6 2016.*

Y.F.Pan,X.Hou,and C.L.Liu

A hybrid approach to detect and localize texts in natural scene images

*IEEE Transactions on Image processing,vol.20,no.3 March 2011*

Mr. Mule S. S. Mr. Holambe S. N

Detecting Text in Natural Scenes with Connected Component Clustering and Nontext Filtering

*International Research Journal of Engineering and Technology (IRJET) Sep 2016*

# The End

Thanks you for your attention

©https://github.com/HarshadaDeokar317/Text-Extraction