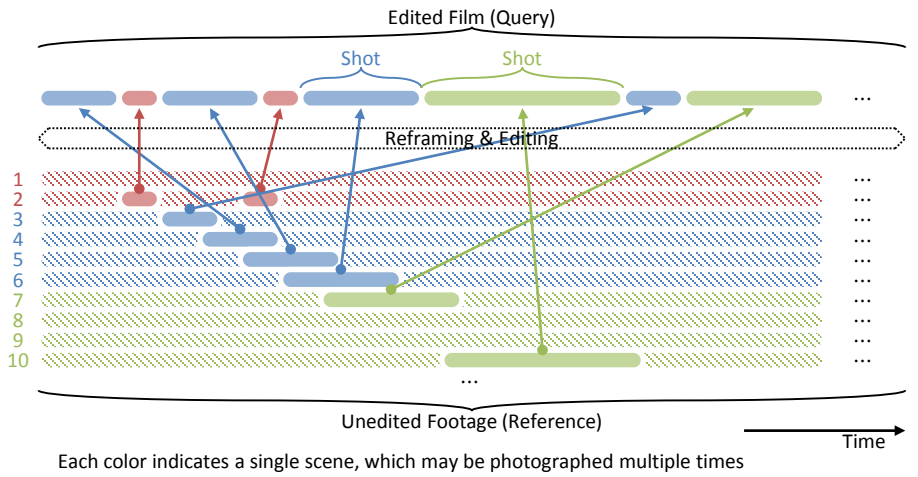
Edited Film Alignment via Selective Hough Transform and Accurate Template Matching

Xiaomeng Wu, Takahito Kawanishi, Minoru Mori, Kaoru Hiramatsu, and Kunio Kashino
NTT Communication Science Laboratories, NTT Corporation

1. Introduction



Each color indicates a single scene, which may be photographed multiple times

Task

Given a collection of unedited footage (reference) and an edited film (query), for each frame of the edited film, determine the **footage**, the **frame**, and the **spatial position** at which the frame of the edited film occurs in the collection

Video Copy Detection (VCD)
Identify the reference video (**footage**) and the temporal position (**frame**) at which the query occurs

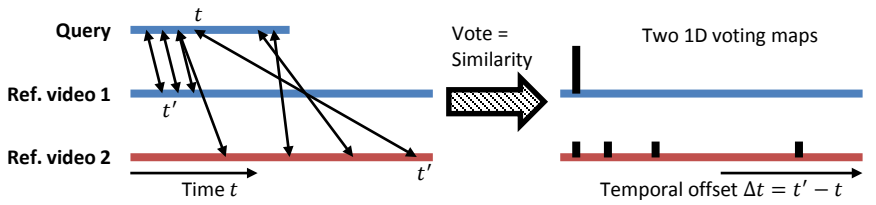
Template Matching (TM)
Find the part (**spatial position**) of an image that matches a template image

Contribution
Frame weighting: propose a measure of how discriminating each frame of a query is for robust VCD

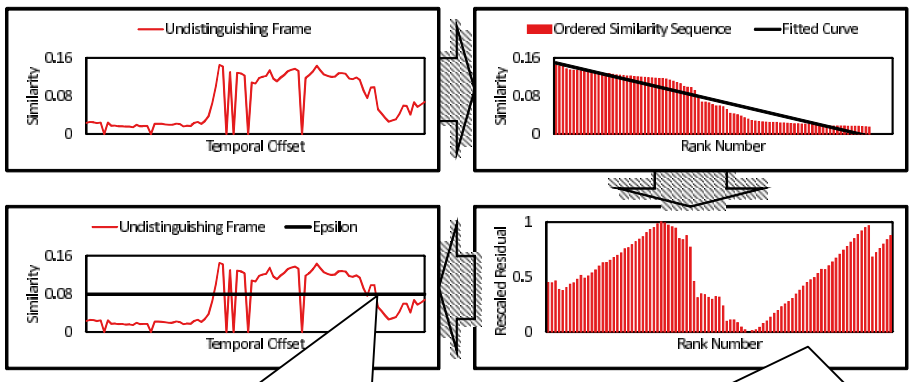
Contribution
Propose a method that encompasses local feature matching and pixel matching for accurate TM

2. Video Copy Detection

Existing Method / Temporal Hough Transform

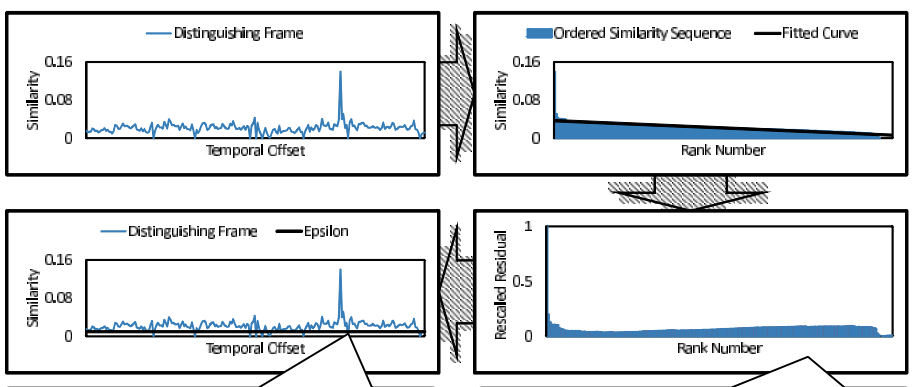


Frame Weighting Based on Similarity Sequence Analysis



A **larger** threshold, which is directly proportional to the AUC of the rescaled residual sequence, is used for **ambiguous** frames

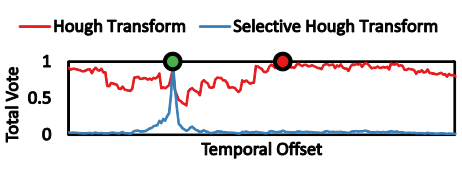
The rescaled residual sequence of the **ambiguous** frame has a much **larger** AUC



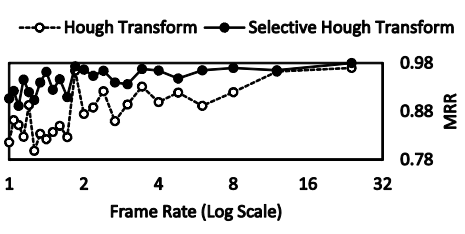
A **smaller** threshold is used for **distinguishing** frames, such that the contribution of votes of such frames can be emphasized

The rescaled residual sequence of the **distinguishing** frame has a much **smaller** AUC, because of the leftmost, distinct peak

Results on real data of 25 queries and 110 reference videos



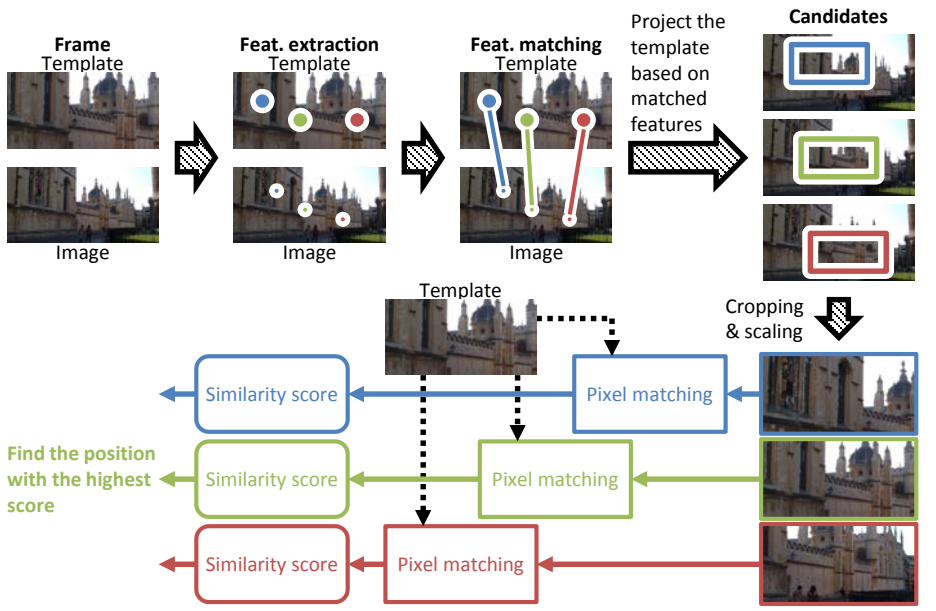
Two voting maps obtained with the Hough transform and the proposed method. The red circle is the offset detected by the Hough transform. The green circle is the ground truth.



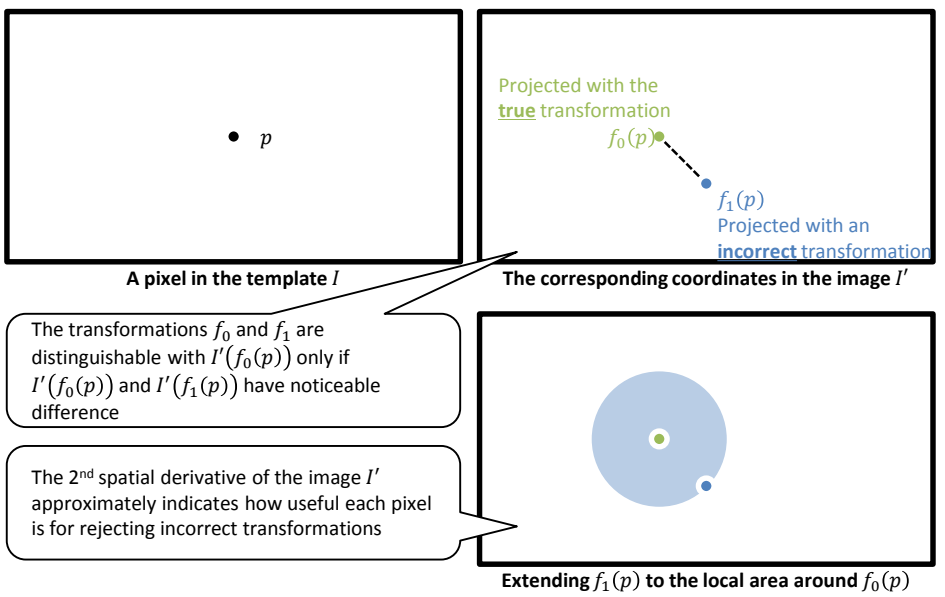
MRR: Mean Reciprocal Rank
Selective Hough Transform = Proposed method

3. Template Matching

Flowchart of Proposed Template Matching Method



Pixel Sampling



A pixel in the template I

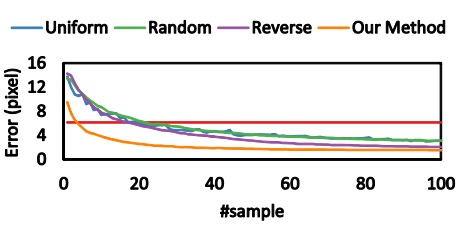
The corresponding coordinates in the image I'

The transformations f_0 and f_1 are distinguishable with $I'(f_0(p))$ only if $I'(f_0(p))$ and $I'(f_1(p))$ have noticeable difference

The 2nd spatial derivative of the image I' approximately indicates how useful each pixel is for rejecting incorrect transformations

Extending $f_1(p)$ to the local area around $f_0(p)$

Results on synthetic data of 2,088 pairs of templates and images



The red line indicates the error of template matching w/o pixel matching

By inspecting only 40 pixels, all the methods that encompass the local feature matching and pixel matching outperformed the method w/o pixel matching

Our method uses the central pixels of Hessian-affine key-points for pixel matching, which outperforms uniform and random sampling