Content-Based Video Copy Detection

Savaş Özkan
MSc Thesis Defense, Electric-Electronics Engineering
Supervisor: Prof. Dr. Gözde Bozdağlı Akar
23/06/2014
Outline

• Motivation
  • Definition of the problem
  • What are the solutions?
  • Essential requisites.

• Introduction
  • Developed system;
    • Local feature extraction.
    • Indexing with quantization-based scheme.
    • Constituting geometric verification.
    • Ranking and Estimation of copy video.
Outline

- Feature Extraction
  - Motivation.
  - Methods;
    - Spatial Feature
      - Interest Point Detector (Hessian-Laplacian)
      - Descriptors (SIFT, Opponent SIFT, F-SIFT, SURF)
    - Spatio-Temporal Feature
      - Interest Point Detector and Tracking (Dense sampling + Optical Flow)
      - Descriptors (HoG, MBH)

- Quantization-based Indexing
  - What are the positive & negative aspects of the concept?
  - Methods;
    - Bag-of-Word (BoW)
    - Hamming Embedding (HE)
    - Product Quantization (PQ)
  - Extra Improvements
    - Inverted Index Structure and Term Frequency-Inverted Document Frequency (Tf-Idf).
Outline

• Geometric Verification
  • The reason of utilizing this scheme.
  • Methods;
    • Weak Geometric Consistency
      • For spatial domain;
        • Weak Geometric Consistency with Scale Distribution (S-WGC)
        • Weak Geometric Consistency with Translation Distribution (T-WGC)
      • For spatio-temporal domain;
        • Trajectory-Based Weak Geometric Consistency (ST-WGC) (Novel Concept)
    • Local Neighboring Relation Signature
      • Visual Group Binary Signature (VGBS) (Novel Concept)
Outline

• Experimental Results
  • TRECVID 2009 CBCD Dataset.
  • Performance metrics.
  • Ranking.
  • Performance results.
  • Discussion and Observations.

• Conclusion & Future Work
  • Novelities and contributions of this study.
  • Academic publications.
  • The objectives that we will investigate in future.
Motivation

• Definition of the problem
  • Copyright infringements have been increased with recent technological development;
    • effective video compression,
    • internet bandwidth availability for average user,
    • increasing popularity of social and video sharing websites.
  
• Considering the amount of circulated video data over internet, it is impossible to handle by human-based intervention.
Motivation

- Definition of the problem
  - Is it real hard to find copy video in an archive?

*Youtube Search: Rise of the Planet of The Apes*
Motivation

• Definition of the problem
  • Is it really hard to find copy video in an archive?
    • Actually, It is a bit more complicated.

*Youtube Search: Rise of the Planet of The Apes*
Motivation

• Definition of the problem

  • Hence, the primary goal of an automatic copy detection is to find the source of the query video with accurate time in a large reference archive.
Motivation

• What are the solutions?

  • There exist two complimentary approaches namely Digital Watermarking and Content-Based Copy Detection

  • Digital Watermarking [1];
    • The aim is to embed robust signatures into video content for future copy search.
    • However;
      • Signatures should be pre-modeled against deformations.
      • This increases storage size alongside of video data.
      • It cannot be applied on currently circulated videos.

Motivation

• What are the solutions?

  • Content-Based Copy Detection [1,2,3,...];
    • The aim is to obtain robust signatures from video leveraging visual, temporal and audio contents.

  • It consists two stages;
    • Offline stage; Robust and distinctive signatures are obtained and an archive is constructed for future search.
    • Online stage; Reference and query signatures are compared and decision is made about source of the query video.

Motivation

• Essential requisites.

  • Succeeding copy detection can be critized around three essential requisites;
    • *Low computational complexity and instance comparison.*
      • Reference archive is huge and user wants to see instance response in query-by-example cases.
    • *Low memory requirement.*
      • Large archive should be stored effective in order to run the developed system on average personal computer.
  
  • *High precise detection accuracy.*
    • High true detection rate with low false alarm.
Introduction

• Developed system;

  • Our developed system consists of four main steps;
    • *Local feature extraction.*
      • Content representation with local spatial and spatio-temporal descriptors.

    • *Indexing with quantization-based scheme.*
      • Indexing feature vector with small indices.

    • *Constituting geometric verification.*
      • Utilization of geometric consistency and geometric relation to refine scores.

    • *Ranking and Estimation of copy video.*
      • Searching query video in reference archive and making a judgment about video.
Introduction

• Developed system *(Ranking and Estimation of copy video)*
  • We have two test setups for spatial and spatio-temporal domains.

• For spatial domain;
Introduction

- Developed system (*Ranking and Estimation of copy video*)

- For spatio-temporal domain;
Feature Extraction

• Motivation.
  • Digital image consists of pixels and they are useless when we use them signly.

• There are two main representation schemes in literature;
  • Global Feature Representation [1,2];
    • It extracts couple of color, motion or texture information from whole an image as a single vector.
    • Ease to compute.
    • They are sensitive to geometric transformations like scale and orientation changes.

Feature Extraction

• Motivation.

  • Local Feature Representation [1,2];
    • Robust to occlusion clutter and geometric transformation in addition to illustration changes.
    • Since it represents local content, initially local stable regions should be found which we call interest points or interest regions.
    • Then, these local regions are represented using color, motion or texture information.

Feature Extraction

• Methods.

  • In this work, we have reserved feature extraction into two sections as;
    • Spatial Feature.
    • Spatio-Temporal Feature.
Feature Extraction

• Methods (*Spatial Feature*)

  • It can be distinguished into two as interest point detection and description.

  • Interest Point Detection;
    • Hessian Laplacian [1] interest point detector is utilized.

  • Local Descriptor;
    • SIFT [2], Opponent SIFT [3], Flip invariant SIFT [4] and SURF [5] descriptors are implemented.

Feature Extraction

- Methods (*Spatial Feature*)
  - Interest Point Detection (*Hessian Laplacian [1,2]*)
    - Hessian matrix measures the curvature on a point.
    - Scale-space is constructed with various gaussian kernels and hessian matrixes are computed.

\[
L(x, y, \sigma_k) = I(x, y) * G(x, y, \sigma_k)
\]

\[
H(x, y, \sigma_k) = \begin{bmatrix}
L_{xx}(x, y, \sigma_k) & L_{xy}(x, y, \sigma_k) \\
L_{yx}(x, y, \sigma_k) & L_{yy}(x, y, \sigma_k)
\end{bmatrix}
\]

Feature Extraction

- Methods (Spatial Feature)
  - Interest Point Detection (Hessian Laplacian [1,2])
    - If the point is greater value among neighbors and a threshold, it is selected as interest point.
    - In order to refine scale characteristic with more proper one, laplacian function is incorporated in scale-space.

\[
\text{Lap}(x, y, \sigma_k) = \sigma_k^2 \left[ L_{xx}(x, y, \sigma_k) + L_{yy}(x, y, \sigma_k) \right]
\]

Feature Extraction

• Methods (*Spatial Feature*)

  • Local Descriptor
    • A circular region is defined using coordinate, scale and orientation characteristics of point.
Feature Extraction

- Methods (*Spatial Feature*)
  - Local Descriptor
    - A circular region is defined using coordinate, scale and orientation characteristics of point.
Feature Extraction

• Methods (Spatial Feature)

  • Local Descriptor (SIFT [1])

    • Orientation assignment: Dominant gradient orientation in circular region is accepted as the orientation characteristic.

    • Keypoint descriptor: Gradient magnitudes are accumulated as a histogram according to their location and orientation parameters.

    • Since sub-region size is $4 \times 4$ and length of orientation histogram is 8, final vector dimension is 128.

Feature Extraction

• Methods (Spatial Feature)

  • Local Descriptor (Opponent SIFT [1])
    • Follows similar steps as in SIFT descriptor.
    • Differently, instead of grayscale image, color content is employed.
    • $O_3$ intensity value in grayscale image.
    • $O_1$ and $O_2$ contain color information.
    • These color channels are invariant to illumination changes.

    \[
    \begin{pmatrix}
    O_1 \\
    O_2 \\
    O_3
    \end{pmatrix} = \begin{pmatrix}
    \frac{R - G}{\sqrt{2}} \\
    \frac{R + G - 2B}{\sqrt{6}} \\
    \frac{R + G + B}{\sqrt{3}}
    \end{pmatrix}
    \]

  • Final vector dimension is equal to $3 \times 128 = 384$.

Feature Extraction

• Methods (Spatial Feature)

  • Local Descriptor (F-SIFT [1])
    • SIFT descriptor is invariant to scale and orientation changes, but not to a flip transformation.
    • The aim is to make SIFT descriptor robust against flip transformation preserving its originality.
    • Dominant curl computation [1] is employed.

Feature Extraction

• Methods (Spatial Feature)

  • Local Descriptor (F-SIFT [1])
    • $C$ parameter’s sign determines possible direction of concatenation in clockwise or counter clockwise manner.
    • According to sign, region is flip or not and SIFT descriptor is computed.

\[
C = \sum_{(x,y) \in I} \left( \sqrt{\frac{\partial I(x,y)^2}{\partial x} + \frac{\partial I(x,y)^2}{\partial y}} \right) \times \cos(\theta_r(x,y))
\]

\[
\frac{\partial I(x,y)}{\partial x} = I(x - 1, y) - I(x + 1, y)
\]

\[
\frac{\partial (x,y)}{\partial y} = I(x, y - 1) - I(x, y + 1)
\]

\[
\theta_r(x,y) = \theta(x,y) - \tan^{-1}\left(\frac{y}{x}\right)
\]

\[
\theta(x,y) = \tan^{-1}\left(\frac{I(x, y - 1) - I(x, y + 1)}{I(x - 1, y) - I(x + 1, y)}\right)
\]

Feature Extraction

• Methods (*Spatial Feature*)

  • Local Descriptor (*SURF [1]*)
    • Orientation assignment: Haar response is calculated in x and y directions and similarly dominant orientation is accepted.

    • Keypoint descriptor: Horizontal $d_x$ and vertical $d_y$ wavelet responses are summed up and a vector $v = (\Sigma d_x, \Sigma d_y, \Sigma|d_x|, \Sigma|d_y|)$.

    • To insert to localization information, region is divided into $4 \times 4$ sub-regions. Final vector dimension is 64.

Feature Extraction

• Methods (*Spatio-Temporal Feature*)
  • Joint usage of spatial and temporal contents would have more distinctive information.

  • It can be distinguished into two parts as;
    • Interest point detection + Tracking
    • Description.

  • Interest Point Detection+Tracking;
    • Densely sampled points are tracked through in time with optical flow [1].

  • Local Descriptor;
    • Histogram of Orientated Gradient (HoG) [2] and Motion Boundary Histogram (MBH) [1,3]

Feature Extraction

Methods (Spatio-Temporal Feature)

- Interest Point Detection+Tracking;
  - Densely sampled trajectory estimation [1] can capture the foreground motion with high precision.
  - Thus, it increases the distinctive power of the representation.

- Candidate points are sampled on frames with different scales.
- Points should be stable for more correct tracking.
- The points on homogeneous areas are eliminated [2].

Feature Extraction

• Methods (Spatio-Temporal Feature)

  • Interest Point Detection+Tracking;
    • Farneback optical motion field [1] is computed.
    • Reduce to sensitivity of motion, $3 \times 3$ median filter is applied.
    • Since static scenes do not contain motion information, the trajectories with small spatial variations are eliminated.

  • At the end; $N \times N \times L$ space-time volume is estimated.
  • To insert location information, spatial and time domains are partitioned into $n_{xy} = 2$ and $n_t = 2$ respectively.

Feature Extraction

- Methods (*Spatio-Temporal Feature*)

  - Local Descriptor (*Histogram of Orientated Gradient (HoG)*) [1]
    - Firstly, proposed for human recognition.
    - Similar to SIFT, 8 bin orientation histogram is weighted with gradient magnitudes.
    - Final vector dimension is $2 \times 2 \times 2 \times 8 = 64$.

Feature Extraction

• Methods \( (\text{Spatio}-\text{Temporal Feature}) \)
  
  • Local Descriptor \( (\text{Motion Boundary Histogram (MBH)}) \) [1,2]
    • Motion on frame, consists of foreground, background and camera motions.
    • Camera motion reduces discriminative power of the scene.
    • Similar idea in gradient estimation, vertical and horizontal derivatives are computed in \( 3 \times 3 \) window and 8 bin histograms are computed for each axis.
    • Final vector dimension is \( 2 \times 2 \times 2 \times 2 \times 8 = 128 \).

Quantization-based Indexing

- What are the positive & negative aspects of the concept?
  - Comparing local descriptors according to their feature vectors is impossible for particularly large database.
  - Idea is to map a feature vector into small indices.

- Pos;
  - Enables to compare feature vectors instantly.
  - Ease to implement. (No complicated mathematical transformation.)

- Neg;
  - Since this is an unsupervised problem, clustering size affects trade-off.
  - Loss some information about feature vector during mapping.
Quantization-based Indexing

- Methods.

  - In this work, we have investigate three methods as;
    - Bag-of-word [1]
    - Hamming Embedding [2]
    - Product Quantization [3]

Quantization-based Indexing

• Methods (*Bag-of-word* [1])
  • Firstly, unveiled for text retrieval.
  • Each feature vector is quantized on a pre-clustered space (*visual codebook*) according to closest distance to cluster centers.

Quantization-based Indexing

- Methods (*Bag-of-word* [1])

  - Since there is single index value $q_c(v)$, similarity score between query and reference signatures is equal to;

    $$s_{Bow}(v^r, v^q) = \begin{cases} 1.0, & \text{if } q_c(v^r) == q_c(v^q) \\ 0.0, & \text{otherwise} \end{cases}$$

Quantization-based Indexing

• Methods (*Hamming Embedding* [1])
  • Cluster size is a critical parameter;
    • For small value, residing noisy version of vector into same cluster is high, however irrelevant vector can also have same cluster id.
    • Conversely, for high value, precise cluster id assignment can be made, but possibility of assigning noisy version of a vector is low.

  • In this method, location of a vector inside cluster center is encoded with additional binary signature.

Quantization-based Indexing

• Methods (*Hamming Embedding* [1])

  • Besides fast comparability, the beauty of this binary signature is translation between sub-regions can be permitted with hamming distance.

\[ H_{he}(v^r, v^q) = \sum_{i=1}^{d_b} |b_i(v^r) - b_i(v^q)| \]

  • In order to compute this binary signature, visual codebook should be updated.

Quantization-based Indexing

- Methods (*Hamming Embedding* [1])

  - Learning stage;
    - A $d \times d_b$ projection matrix $P$ is generated. (Random Gaussian + QR factorization)
    - Each sample in corpus projected using $P$.
    - Median values $\tau_{q_c(v),i}$ where $i = 1 \ldots d_b$ for each cluster are stored.

  - Assigning stage;
    - Assign closest center $q_c(v)$.
    - Project feature vector to $z = z_{q_c(v),1}, z_{q_c(v),2}, \ldots, z_{q_c(v),d_b}$.
    - Compute binary by comparing median values.

\[
b_i(v) = \begin{cases} 
1, & \text{if } z_{q_c(v),i} > \tau_{q_c(v),i} \\
0, & \text{otherwise}
\end{cases}
\]

Quantization-based Indexing

- Methods (*Hamming Embedding* [1])

- For similarity score;

\[
s_{HE}(v^r, v^q) = \begin{cases} 
1.0 - \frac{H_{he}(v^r, v^q)}{h_t}, & \text{if } q_c(v^r) = q_c(v^q) \\
0.0, & \text{if } H_{he}(v^r, v^q) < h_t \\
\text{otherwise} & 
\end{cases}
\]

Quantization-based Indexing

- **Methods** *(Product Quantization [1])*
  - High cluster size increases bit rate per component of a vector.
  
  - The purpose of this method is to increase bit rate per component of a vector by splitting the vector into \( m \) uniform sub-vectors.
  
  - Then, for each sub-vector, quantization step is done with cluster size \( K^* \)
    
    \[
    \{v_1, v_2, \ldots, v_m\} \rightarrow \{q_1(v_1), q_2(v_2), \ldots, q_m(v_m)\}
    \]
  
  - Final approximate cluster size is \((K^*)^m\).

Quantization-based Indexing

• Methods (*Product Quantization* [1])

  • In image retrieval, similar to hamming embedding, a small code is added that encodes residual error.

  \[ r(v) = v - q_c(v) \]

  • This residual error is quantized with product quantizer;

  \[ \hat{v} = q_c(v) - q_p(v - q_c(v)) \]

Quantization-based Indexing

- Methods (*Product Quantization* [1])

  - For similarity score, we apply two constraints;
    - Quantized indices should be equal. \( q_c(v^r) = q_c(v^q) \)
    - Order of quantized query sub-residue \( q_m(r(v^q_m)) \) in nearest neighbor of quantized reference sub-residue \( q_m(r(v^r_m)) \) should be up to a threshold \( \tau_{pq} \).

\[
s_{HE}(v^r, v^q) = \frac{1}{M} \sum_{1\leq m \leq M} 1.0 - \frac{NN_m(q_m(r(v^r_m))) - q_m(r(v^q_m)))}{\tau_{pq}}
\]

Quantization-based Indexing

• Extra Improvements

  • Inverted Index Structure [1]
    • Indexing descriptors according to their indices.
    • This allows us to compare the descriptors with same indices and reduce search space.

  • Term-Frequency – Inverted Document Frequency [1]
    • Descriptors are weighted according to their occurrence frequencies.
    • More frequent ones have less information, less frequent ones have more information.

Geometric Verification

• The reason of utilizing this scheme.
  
  • Joint usage of local descriptor and quantization-based indexing discards geometric consistency among interest points.

• Frequently, it can be reintroduced with;
  
  • adding a post-processing stage that eliminates outliers. (Weak geometric consistency)
  
  • encoding local neighboring relation of interest points by a signature. (Local Neighbor Relation.)
Geometric Verification

• Methods (Weak geometric consistency [1,2])
  
  • Simply, this is a filtering stage to refine true local matches according to dominant geometric transformation.
  
  • RANSAC [3] is the most famous one. But computational complexity is too high.
  
  • Simpler algorithm is proposed.
    • Instead of verifying exact geometric transformation, an approximate geometric characteristic can be obtained by parameter differences.

Geometric Verification

• Methods (*Weak geometric consistency*)

  • For spatial domain;
    • Weak geometric consistency with scale distribution [1].
    • Weak geometric consistency with translation distribution [2].

  • For spatio-temporal domain;
    • Trajectory-based weak geometric consistency (*Novel Concept*)

---

Geometric Verification

- Methods (*Weak geometric consistency*)

  - For spatial domain;
    - To fit a geometric model on local descriptors, scale, orientation and translation parameters can be used.
    - Quantized orientation parameters for local query and reference points should be identical.

\[
q_{\theta^r} = \frac{\theta^r}{q_{s_{step}}}
\]

\[
q_{\theta^r} = = q_{\theta^q}
\]
Geometric Verification

- Methods (*Weak geometric consistency*)
  - For spatial domain (*Weak geometric consistency with scale distribution* [1])
    - The assumption is when an image undergoes rotation and scale changes, local descriptors on this image affect same amount.
    - The geometric consistency can be built on seeking a global distributions on scale and rotation differences.
    - A scale difference histogram is constructed and peak value of histogram is accepted as scale characteristic of two frames.

\[
q_{sr} = \log_2 s^r \\
\tilde{s} = q_{sr} - q_{sq}
\]

Geometric Verification

• Methods (Weak geometric consistency)
  
  • For spatial domain (Weak geometric consistency with translation distribution [1])
    
    • Translation has more discriminative geometric clues compare to scale.
    • Manhattan distance (Decrease complexity)
    • In order to investigate joint characteristic in scale and translation changes, 2D distribution histogram is utilized.

\[
\begin{bmatrix}
    x^q \\
    y^q
\end{bmatrix} = s \times \begin{bmatrix}
    \cos \theta \\
    -\sin \theta
\end{bmatrix} \times \begin{bmatrix}
    x^r \\
    y^r
\end{bmatrix} + \begin{bmatrix}
    t_x \\
    t_y
\end{bmatrix}
\]

\[
\tilde{s} = 2(q_x - x^q) \quad \begin{bmatrix}
    \tilde{x}^q \\
    \tilde{y}^q
\end{bmatrix} = \tilde{s} \times \begin{bmatrix}
    x^r \\
    y^r
\end{bmatrix}
\]

\[
\tilde{c} = |x^q - \tilde{x}^q| + |y^q - \tilde{y}^q|
\]

Geometric Verification

• Methods (*Weak geometric consistency*)

  • For spatial domain (*Weak geometric consistency with translation distribution* [1])
    
    • Unlike scale distribution, translation distribution is not invariant to flip transformation.
    • Vertical flip deforms $x$ coordinate as $width - x$.[2]

    $\begin{bmatrix} width - \bar{x}^q \\ y^q \end{bmatrix} = \hat{s} \times \begin{bmatrix} \bar{x}^r \\ y^r \end{bmatrix}$

    $t = |x^q + \hat{s} \times x^r - width| + |y^q - \hat{s} \times y^r|$

    $t = |x^q + \hat{s} \times x^r| + |y^q - \hat{s} \times y^r|$

  • Due to the disjoint relation of original and flip versions, we have utilized two 2D histograms.

Geometric Verification

• Methods (*Weak geometric consistency*)

  • For spatio-temporal domain [1] (*Trajectory-based weak geometric consistency*)

    *(Novel Concept)*

    • Trajectories are computed on consecutive frames.
    • Spatial variations and means in time can be employed as geometric clues.

  • Proposed method considers two relations
    • Spatial variations
    • Spatial means

\[
\begin{align*}
\mu_x &= \frac{1}{L} \sum_{i=1}^{L} x_i \\
\mu_y &= \frac{1}{L} \sum_{i=1}^{L} y_i \\
\sigma_x &= \frac{1}{L} \sum_{i=1}^{L} (x_i - \mu_x)^2 \\
\sigma_y &= \frac{1}{L} \sum_{i=1}^{L} (y_i - \mu_y)^2
\end{align*}
\]

Geometric Verification

• Methods (Weak geometric consistency)

  • For spatio-temporal domain [1] (Trajectory-based weak geometric consistency)
    (Novel Concept)

    • Spatial variations;
      • If pairs are identical, spatial variations should be roughly proportional with scale difference.
    • Manhattan distance (Decrease complexity.)

\[
\sigma_{x,y}^q - \bar{s} \times \sigma_{x,y}^q < \tau_\sigma
\]
\[
\bar{s} = \sqrt{2} \times (s^q - s^r)
\]
\[
\sigma_{x,y}^r = \sigma_x^r + \sigma_y^r
\]
\[
\sigma_{x,y}^q = \sigma_x^q + \sigma_y^q
\]

Geometric Verification

• Methods (*Weak geometric consistency*)

  • For spatio-temporal domain [1] (*Trajectory-based weak geometric consistency*)
    *(Novel Concept)*

  • Spatial means;
    • Geometric transformation is reintroduced for spatial means for reference and query.

  • Manhattan distance (Decrease complexity.)

  • 2D distribution histogram is utilized for translation mean and scale differences.

\[
\begin{align*}
\begin{bmatrix}
\mu_x^q \\
\mu_y^q
\end{bmatrix} &= \hat{s} \times \begin{bmatrix}
\mu_x^r \\
\mu_y^r
\end{bmatrix} + \begin{bmatrix}
t_x \\
t_y
\end{bmatrix} \\
\hat{s} &= \sqrt{2} \times (s^q - s^n) \\
t_{\mu} &= |\mu_{x,y}^q - \hat{s} \times \mu_{x,y}^r| \\
\mu_{x,y}^q &= \mu_x^q + \mu_y^q \\
\mu_{x,y}^r &= \mu_x^r + \mu_y^r
\end{align*}
\]

Geometric Verification

• Methods (*Local Neighboring Relation*)

  • The aim is to encode a geometric signature using neighboring relation among local interest points.

  • This should be robust and ease to compare for large dataset.

• Visual Group Binary Signature [1] (*Novel Concept*)
  • It merely checks existent or non-existent of interest points in neighborhood area and generates a binary string.
  • Instead of complex voting scheme, similarity of local descriptors can be obtain by bitwise comparisons.

Geometric Verification

- Methods (Local Neighboring Relation)

  - Visual Group Binary Signature [1] (Novel Concept)
    - Circular region is defined around a central point.
    - Circular region is divided into sub-partitions \( G_{i,j}^k \) in scale and angular domains.
    - Binary signature is computed:
      \[
      b_{vg}^k = \{ b_p(G_{1,1}^k), \ldots, b_p(G_{N_{\delta \theta}, N_{\delta s}}^k) \}
      \]

      \[
      b_p(G_{i,j}^k) = \begin{cases} 
      1 & \text{if any interest point exist in } G_{i,j}^k \\
      0 & \text{otherwise}
      \end{cases}
      \]

  - Similarity score:
    \[
    s_{vg}(b_{vg}^k, b_{vg}^l) = \frac{1}{N_{\text{norm}}} \sum_{1 \leq i \leq N_{\delta \theta}} \sum_{1 \leq j \leq N_{\delta s}} b_p(G_{i,j}^k) \times b_p(G_{i,j}^l)
    \]

Experimental Results

- **TRECVID 2009 CBCD Dataset [1].**
  - Consists of 400 hours reference videos and 1407 query videos.
  - 937 query videos are in this reference videos.

- 7 types of attack models
  - T2: Picture-in-picture.
  - T3: Insertion of pattern.
  - T4: Strong re-encoding.
  - T5: Change of gamma.
  - T6: Decrease in quality.
  - T8: Post processing.
  - T10: Combination of 5 attacks.

Experimental Results

• Performance metrics.
  • Recall, precision and f1-score are calculated.

\[
\text{recall} = \frac{|\{\text{relevant document}\} \cap \{\text{retrieved document}\}|}{|\{\text{relevant document}\}|}
\]

\[
\text{precision} = \frac{|\{\text{relevant document}\} \cap \{\text{retrieved document}\}|}{|\{\text{retrieved document}\}|}
\]

\[
\text{f1score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}
\]
Experimental Results

- **Ranking**
  - Query video is searched in reference archive using their coherent content characteristics.

- Reliability score between two videos is proportional with number of interest point correspondences on videos.

- Since videos are represented with sampled frames, similarity computed one-by-one frame comparisons.
Experimental Results

• Ranking

  • The ranking procedure is;
    • Initiate a score vector whose size is equal to reference video duration.
    • For each query frame, compute similarity score on each reference frame.
    • Add similarity score to corresponding bin of score vector.
    • Maximum value indicates reference video with time location.

  • This procedure is repeated for all reference videos.

  • In order to say, query video is copy, first highest score must be at least twice of second highest score among score vectors.
Experimental Results

• Performance results.
  • We have tested performances of several combinations around local descriptors.

  • For each spatial domain local descriptor;
    • Bag-of-word.
    • Hamming Embedding.
    • Product Quantization.
    • Visual Group Binary Signature.
    • Weak Geometric Consistency with Scale Distribution.
    • Weak Geometric Consistency with Translation Distribution.

  • For each spatio-temporal domain local descriptor;
    • Bag-of-word
    • Hamming Embedding
    • Product Quantization
    • Trajectory-based Weak Geometric Consistency.
Experimental Results

- Performance results.
  - From the results, combination of Flip-invariant SIFT, Hamming Embedding, Visual Group Binary Signature and Weak Geometric Consistency with Translation Distribution yields overall best results in all performance metrics.

<table>
<thead>
<tr>
<th>Baseline</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T8</th>
<th>T10</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoW+SWGC</td>
<td>0.4029</td>
<td>0.7761</td>
<td>0.5373</td>
<td>0.8208</td>
<td>0.8731</td>
<td>0.4552</td>
<td>0.3731</td>
<td>0.6055</td>
</tr>
<tr>
<td>BoW+TWGC</td>
<td>0.6044</td>
<td>0.8955</td>
<td>0.7910</td>
<td>0.9253</td>
<td>0.9701</td>
<td>0.7248</td>
<td>0.6119</td>
<td>0.7889</td>
</tr>
<tr>
<td>BoW+SWGC+VGBS</td>
<td>0.4850</td>
<td>0.8805</td>
<td>0.6940</td>
<td>0.8731</td>
<td>0.9477</td>
<td>0.5298</td>
<td>0.4925</td>
<td>0.7004</td>
</tr>
<tr>
<td>BoW+TWGC+VGBS</td>
<td>0.6044</td>
<td>0.9253</td>
<td>0.7985</td>
<td>0.9402</td>
<td>0.9776</td>
<td>0.7248</td>
<td>0.6194</td>
<td>0.7985</td>
</tr>
<tr>
<td>HE+SWGC</td>
<td>0.5970</td>
<td>0.8880</td>
<td>0.7238</td>
<td>0.9029</td>
<td>0.9552</td>
<td>0.5970</td>
<td>0.5597</td>
<td>0.7462</td>
</tr>
<tr>
<td>HE+TWGC</td>
<td>0.7313</td>
<td>0.9328</td>
<td>0.8582</td>
<td>0.9552</td>
<td>0.9850</td>
<td>0.8059</td>
<td>0.6791</td>
<td>0.8496</td>
</tr>
<tr>
<td>HE+SWGC+VGBS</td>
<td>0.6791</td>
<td>0.9402</td>
<td>0.8059</td>
<td>0.9477</td>
<td>0.9850</td>
<td>0.7014</td>
<td>0.6343</td>
<td>0.8134</td>
</tr>
<tr>
<td>HE+TWGC+VGBS</td>
<td>0.7313</td>
<td>0.9552</td>
<td>0.8432</td>
<td>0.9552</td>
<td>0.9850</td>
<td>0.8059</td>
<td>0.6865</td>
<td>0.8516</td>
</tr>
<tr>
<td>PQ+SWGC</td>
<td>0.6417</td>
<td>0.9104</td>
<td>0.7388</td>
<td>0.9029</td>
<td>0.9477</td>
<td>0.5820</td>
<td>0.5149</td>
<td>0.7484</td>
</tr>
<tr>
<td>PQ+TWGC</td>
<td>0.6940</td>
<td>0.9328</td>
<td>0.8432</td>
<td>0.9477</td>
<td>0.9776</td>
<td>0.7611</td>
<td>0.6492</td>
<td>0.8294</td>
</tr>
<tr>
<td>PQ+SWGC+VGBS</td>
<td>0.6791</td>
<td>0.9253</td>
<td>0.7910</td>
<td>0.9253</td>
<td>0.9701</td>
<td>0.6343</td>
<td>0.5671</td>
<td>0.7846</td>
</tr>
<tr>
<td>PQ+TWGC+VGBS</td>
<td>0.7014</td>
<td>0.9328</td>
<td>0.8358</td>
<td>0.9552</td>
<td>0.9776</td>
<td>0.7761</td>
<td>0.6492</td>
<td>0.8326</td>
</tr>
</tbody>
</table>
## Experimental Results

<table>
<thead>
<tr>
<th>Baseline</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T8</th>
<th>T10</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoW+SWGC</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9863</td>
<td>0.9909</td>
<td>0.9915</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9947</td>
</tr>
<tr>
<td>BoW+TWGC</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9906</td>
<td>0.992</td>
<td>0.9923</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9959</td>
</tr>
<tr>
<td>BoW+SWGC+VGBS</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9893</td>
<td>0.9915</td>
<td>0.9921</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9954</td>
</tr>
<tr>
<td>BoW+TWGC+VGBS</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9907</td>
<td>0.9921</td>
<td>0.9924</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9959</td>
</tr>
<tr>
<td>HE+SWGC</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9897</td>
<td>0.9918</td>
<td>0.9922</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9957</td>
</tr>
<tr>
<td>HE+TWGC</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9913</td>
<td>0.9922</td>
<td>0.9924</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9962</td>
</tr>
<tr>
<td>HE+SWGC+VGBS</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9908</td>
<td>0.9921</td>
<td>0.9924</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9960</td>
</tr>
<tr>
<td>HE+TWGC+VGBS</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9912</td>
<td>0.9922</td>
<td>0.9924</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9962</td>
</tr>
<tr>
<td>PQ+SWGC</td>
<td>1.0</td>
<td>1.0</td>
<td>0.99</td>
<td>0.9918</td>
<td>0.9921</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9957</td>
</tr>
<tr>
<td>PQ+TWGC</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9912</td>
<td>0.9921</td>
<td>0.9924</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9961</td>
</tr>
<tr>
<td>PQ+SWGC+VGBS</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9906</td>
<td>0.992</td>
<td>0.9923</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9959</td>
</tr>
<tr>
<td>PQ+TWGC+VGBS</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9911</td>
<td>0.9922</td>
<td>0.9924</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9961</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Baseline</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T8</th>
<th>T10</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoW+SWGC</td>
<td>0.5744</td>
<td>0.8739</td>
<td>0.6956</td>
<td>0.8979</td>
<td>0.9285</td>
<td>0.6256</td>
<td>0.5434</td>
<td>0.7581</td>
</tr>
<tr>
<td>BoW+TWGC</td>
<td>0.7534</td>
<td>0.9444</td>
<td>0.8796</td>
<td>0.9575</td>
<td>0.9811</td>
<td>0.8938</td>
<td>0.7592</td>
<td>0.8804</td>
</tr>
<tr>
<td>BoW+SWGC+VGBS</td>
<td>0.6332</td>
<td>0.9365</td>
<td>0.8157</td>
<td>0.9285</td>
<td>0.9694</td>
<td>0.6926</td>
<td>0.66</td>
<td>0.8222</td>
</tr>
<tr>
<td>BoW+TWGC+VGBS</td>
<td>0.7534</td>
<td>0.9612</td>
<td>0.8842</td>
<td>0.9655</td>
<td>0.9849</td>
<td>0.8404</td>
<td>0.7649</td>
<td>0.8833</td>
</tr>
<tr>
<td>HE+SWGC</td>
<td>0.7476</td>
<td>0.9407</td>
<td>0.8362</td>
<td>0.9453</td>
<td>0.9733</td>
<td>0.7476</td>
<td>0.7177</td>
<td>0.8531</td>
</tr>
<tr>
<td>HE+TWGC</td>
<td>0.8448</td>
<td>0.9652</td>
<td>0.92</td>
<td>0.9733</td>
<td>0.9887</td>
<td>0.8825</td>
<td>0.8088</td>
<td>0.9171</td>
</tr>
<tr>
<td>HE+SWGC+VGBS</td>
<td>0.8088</td>
<td>0.9692</td>
<td>0.8888</td>
<td>0.9694</td>
<td>0.9887</td>
<td>0.8245</td>
<td>0.7762</td>
<td>0.8955</td>
</tr>
<tr>
<td>HE+TWGC+VGBS</td>
<td>0.8448</td>
<td>0.9770</td>
<td>0.9112</td>
<td>0.9733</td>
<td>0.9887</td>
<td>0.8825</td>
<td>0.8141</td>
<td>0.9182</td>
</tr>
<tr>
<td>PQ+SWGC</td>
<td>0.7818</td>
<td>0.9531</td>
<td>0.8461</td>
<td>0.9453</td>
<td>0.9694</td>
<td>0.7358</td>
<td>0.6798</td>
<td>0.8545</td>
</tr>
<tr>
<td>PQ+TWGC</td>
<td>0.8193</td>
<td>0.9652</td>
<td>0.9112</td>
<td>0.9694</td>
<td>0.9849</td>
<td>0.8644</td>
<td>0.7873</td>
<td>0.9051</td>
</tr>
<tr>
<td>PQ+SWGC+VGBS</td>
<td>0.8088</td>
<td>0.9612</td>
<td>0.8706</td>
<td>0.9575</td>
<td>0.9811</td>
<td>0.7762</td>
<td>0.7238</td>
<td>0.8777</td>
</tr>
<tr>
<td>PQ+TWGC+VGBS</td>
<td>0.8245</td>
<td>0.9652</td>
<td>0.9068</td>
<td>0.9733</td>
<td>0.9849</td>
<td>0.8739</td>
<td>0.7872</td>
<td>0.9070</td>
</tr>
</tbody>
</table>
Experimental Results

- Performance results.
  - Comparison time for spatial and spatio-temporal descriptor models.
  - Time in second yields for comparing 1 second query video with 100 hours of reference database.

<table>
<thead>
<tr>
<th>Feature Model</th>
<th>BoW+SWGC</th>
<th>BoW+TWGC</th>
<th>BoW+SWGC +VGBS</th>
<th>BoW+TWGC +VGBS</th>
<th>HE+SWGC</th>
<th>HE+TWGC</th>
<th>HE+SWGC +VGBS</th>
<th>HE+TWGC +VGBS</th>
<th>PQ+SWGC</th>
<th>PQ+TWGC</th>
<th>PQ+SWGC +VGBS</th>
<th>PQ+TWGC +VGBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial</td>
<td>0.04 sn</td>
<td>0.1 sn</td>
<td>0.06 sn</td>
<td>0.12 sn</td>
<td>0.05 sn</td>
<td>0.11 sn</td>
<td>0.07 sn</td>
<td>0.13 sn</td>
<td>0.07 sn</td>
<td>0.12 sn</td>
<td>0.09 sn</td>
<td>0.14 sn</td>
</tr>
<tr>
<td>Spatio-Temporal</td>
<td>0.01 sn</td>
<td>0.05 sn</td>
<td>0.02 sn</td>
<td>0.07 sn</td>
<td>0.03 sn</td>
<td>0.09 sn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experimental Results

- Discussion and Observations.
  - Hamming embedding and product quantization give better results compare to bag-of-word.
  - Computing similarity score in local content signature with soft-assignment increases detection accuracy.
  - Even if Opponent SIFT is more discriminative than any other descriptors, it gives the worse results due to the its higher dimensionality.
  - Flip-invariant SIFT descriptor gives worse result on re-encoding attacks compare to classic SIFT descriptors.
Experimental Results

• Discussion and Observations.

  • Elimination of static trajectories decreases the performance.

  • Spatio-temporal with motion content fails on re-encoding attacks.

  • Utilization of geometric consistency improves the accuracy for both spatial and spatio-temporal domains.

  • Comparison time for even the slowest combination is in acceptable range.

  • Purely utilization of visual group binary signature with scale distribution geometric consistency yields compatible results over complex results like weak geometric consistency with translation distribution.
Conclusion & Future Work

• Novelities and contributions of this study.

  • Within the scope of this thesis, we have proposed an overall content-based video copy detection that consists of three main stages, feature extraction, indexing and geometric consistency.

  • Firstly,
    • Densely sampled feature model are deployed on this problem.
    • Soft-assignments instead of hardcoded similarity score for product quantization.
    • A novel compact and effective local geometric signature is proposed.
    • A novel trajectory-based weak geometric consistency for spatio-temporal descriptors.
    • A novel scheme for combination of flip invariant and original weak geometric consistency for translation distribution.

  • Best of all, we have reached several important observations which will sight in future research.
Conclusion & Future Work

• Academic publications


Questions?

• Thank you for your attention.