Visual Group Binary Signature for Video Copy Detection

Savaş Özkan†‡, Ersin Esen† and Gözde Bozdağı Akar‡

†: Image Processing Group, TUBITAK UZAY, Turkey
‡: Dept. of Electrical-Electronics Engineering, Middle East Technical University, Turkey
Outline

• Motivation

• Proposed Method
  • Local Descriptor
  • Visual Group Binary Signature
  • Visual Indexing
  • Ranking

• Experiments

• Conclusion
Motivation

• Proliferation of copyright infringements have become an emerging problem.

• Several crucial reasons;
  • development of effective video compressions,
  • increasing bandwidth availability,
  • increasing popularity of video sharing websites,
Motivation

• Is it real hard to obtain copy video in an archive?

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

• Is it real hard to obtain copy video in an archive?
  • Actually, It is a bit more complicated.

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

• Therefore, in this work, we propose a complete content-based video copy detection system.
Proposed Method

• Local Descriptor


• Flip-Invariant SIFT preserves original structure of classical SIFT.

• Flip-Invariant SIFT generates a parameter which makes a decision about local region need to be flipped or not.
Proposed Method

• Visual Group Binary Signature
  • Inspired from [Dai 2013]

• Approach Motivation:
  • Uses existence or non-existence of interest points in a circular region.
  • Encodes as a binary string.
  • Scale, orientation and flip invariant.
Proposed Method

• Visual Group Binary Signature

• Approach Detail:
  • Define a circular region around a center point \( k \).
  • Partition the region into patches \( G^k \) in scale (1 \( \ldots \) \( N_{\delta_s} \)) and angular (1 \( \ldots \) \( N_{\delta_\theta} \)) domains.

\[
G^k = \{ G_{1,1}^k, \ldots, G_{N_{\delta_\theta},N_{\delta_s}}^k \}
\]
Proposed Method

• Visual Group Binary Signature

• Approach Detail:
  - According to existence or non-existence of a interest point, assign bit value as 1 or 0 for each.
  - Concatenate these bits $b_{vg}^k$ in clock-wise (flip case) or counter clock-wise manner (normal case).

$$b_{vg}^k = \{b_p(G_{1,1}^k), ..., b_p(G_{N_{\theta,\theta}^k,N_{\delta,\delta}^k})\}$$

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

• Visual Group Binary Signature

• Approach Detail:

scale level 1: 00110000
scale level 2: 00010000
scale level 3: 00001000
scale level 4: 00000010
Proposed Method

• Visual Group Binary Signature

• Approach Detail:
  • Similarity score of two binary signatures $k$ and $l$;

$$s_{vg}(b_{vg}^k, b_{vg}^l) = \frac{1}{N_{\text{norm}}} \sum_{1 \leq i \leq N_{\delta}} b_p(G_{i,j}^k) \times b_p(G_{i,j}^l)$$

- $N_{\text{norm}}$: maximum of number of filled patches in $G^k$ and $G^l$. 


Proposed Method

- Visual Indexing
  - *Bag-of-Word* [Sivic 2003]
    - Quantize feature vector $v$ to closest cluster center $q_c(v)$.

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

• Visual Indexing

• *Hamming Embedding* [Jegou 2008]
  
  • Encode in-class location of \( v \) with binary string \( b(v) \).

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

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

• Visual Indexing
  
  • Product Quantization [Jegou 2010]
    
    • Quantize sub-vectors $v_m$ of $v$ separately.

    \[ q_p(v) = (q_1(v_1), q_2(v_2), ..., q_M(v_m)) \]

    • Encode residual error $\sigma$ of $v$ to closest cluster center $q_c(v)$

    \[ \sigma = v - q_c(v) \]
Proposed Method

• Visual Indexing
  • Product Quantization

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

• \( q_m(.) \): quantizer of sub-residual error.
• \( NN_m(.) \): nearest neighbor order of two sub-residues.
• \( t_p \): upper limit for nearest neighbor order.
Proposed Method

• Ranking
  • Content similarity

\[ s_{c_{t_i,t_j}} := w_{t_{tfidf}}^{t_i,m} \times s_{model}(v^{t_i,m}, v^{t_j,n}) \times s_{vg}(b^{t_i,m}_{vg}, b^{t_j,n}_{vg}) \]

• Geometric consistency
  • Simplified and improved version of E-WGC [Zhao 2010].
Proposed Method

• Ranking
  • Geometric consistency.

Normal Case:

\[
\begin{bmatrix}
  x^q \\
y^q
\end{bmatrix} = \tilde{s} \times \begin{bmatrix}
  x^r \\
y^r
\end{bmatrix} + \begin{bmatrix}
t_x \\
t_y
\end{bmatrix}
\]

\[
\tau = |x^q - \tilde{s} \times x^r| + |y^q - \tilde{s} \times y^r|
\]

Final Score:

\[
s_{final} = \max(h_+^{\tau}, h_-^{\tau})
\]

Flip Case:

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

\[
\tau = |x^q + \tilde{s} \times x^r| + |y^q - \tilde{s} \times y^r|
\]
Proposed Method

• Ranking
  • Temporal Alignment.
    • Initialize a score vector.
    • Slide frames by $t_i - t_j$.
    • Find maximum score bin.

• Constrain: first highest score must be at least twice of second.
Experiments

- TRECVID 2009 CBCD dataset.
  - 400 hours of ref. videos.
  - 1407 query videos.
  - 7 different visual attacks.
    - picture-in-picture (T2), insertion of pattern (T3), strong re-encoding (T4), change-of-gamma (T5), decrease-in-quality (T6), post processing (T8) and combination of 5 different attacks (T10)
Experiments

- Recall score with average comparison time.

<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>
<th>Time(second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoW+WGC</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>
<td>0.0402</td>
</tr>
<tr>
<td>BoW+WGC+VGBP</td>
<td>0.4850</td>
<td>0.8805</td>
<td>0.6940</td>
<td>0.8731</td>
<td>0.9447</td>
<td>0.5298</td>
<td>0.4925</td>
<td>0.7004</td>
<td>0.0629</td>
</tr>
<tr>
<td>BoW+EWGC</td>
<td>0.6044</td>
<td>0.8955</td>
<td>0.7910</td>
<td>0.9253</td>
<td>0.9701</td>
<td>0.7238</td>
<td>0.6119</td>
<td>0.7889</td>
<td>0.955</td>
</tr>
<tr>
<td>BoW+EWGC+VGBP</td>
<td>0.6044</td>
<td>0.9253</td>
<td>0.7985</td>
<td>0.9402</td>
<td>0.9776</td>
<td>0.7238</td>
<td>0.6194</td>
<td>0.7985</td>
<td>0.1202</td>
</tr>
<tr>
<td>HE+WGC</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>
<td>0.0531</td>
</tr>
<tr>
<td>HE+WGC+VGBP</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>
<td>0.0702</td>
</tr>
<tr>
<td>HE+EWGC</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>
<td>0.1095</td>
</tr>
<tr>
<td>HE+EWGC+VGBP</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>0.6856</td>
<td>0.8516</td>
<td>0.1276</td>
</tr>
<tr>
<td>PQ+WGC</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>
<td>0.0657</td>
</tr>
<tr>
<td>PQ+WGC+VGBP</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>
<td>0.0862</td>
</tr>
<tr>
<td>PQ+EWGC</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>
<td>0.1140</td>
</tr>
<tr>
<td>PQ+EWGC+VGBP</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>
<td>0.1352</td>
</tr>
</tbody>
</table>
Conclusion

• **Visual Group Binary Signature** improves recall accuracy.
  - Combination with WGC gives approximately similar results to E-WGC.

• **Hamming Embedding and Product Quantization** yield better results.
  - Modeling in-class variation is sensible rather than selecting higher k value.
Conclusion

- F-SIFT is invariant for flip case but it adds some noise for other cases.

<table>
<thead>
<tr>
<th></th>
<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>F-SIFT</td>
<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></td>
<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></td>
<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></td>
<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>SIFT</td>
<td>HE+SWGC</td>
<td>0.6417</td>
<td>0.8955</td>
<td>0.7611</td>
<td>0.9179</td>
<td>0.9701</td>
<td>0.3880</td>
<td>0.4701</td>
<td>0.7206</td>
</tr>
<tr>
<td></td>
<td>HE+TWGC</td>
<td>0.7313</td>
<td>0.9402</td>
<td>0.8731</td>
<td>0.9626</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5522</td>
<td>0.7942</td>
</tr>
<tr>
<td></td>
<td>HE+SWGC+VGBS</td>
<td>0.6865</td>
<td>0.9328</td>
<td>0.8358</td>
<td>0.9477</td>
<td>0.9925</td>
<td>0.4402</td>
<td>0.5223</td>
<td>0.7654</td>
</tr>
<tr>
<td></td>
<td>HE+TWGC+VGBS</td>
<td>0.7462</td>
<td>0.9701</td>
<td>0.8955</td>
<td>0.9552</td>
<td>1.0</td>
<td>0.4850</td>
<td>0.5522</td>
<td>0.8006</td>
</tr>
</tbody>
</table>
Conclusion

• Our Visual Group Binary Signature gives compatible results with less memory and complexity compare to Visual Group [Dai 2013]

• Particularly, Visual Group fails on T4 (strong re-encoding).
Question

• Thanks for your attention...