Introduction to video hashing

Mu Li

iPAL Group Meeting

October 8, 2010

Outline

- Definition and requirements of video hashing
- Applications of video hashing
- Recent methods of video hashing
- Four typical methods of video hashing
- Open problems and conclusions

What's video hashing?

- Hashing:

$$K \to h(K) \tag{1}$$

- If K is video, then $h(K)$ is video hashing.
- Specific requirements of video hashing:
  - Perceptual uniqueness: Hashing values of "visually" different videos should be different.
  - Robustness: Hashing values extracted from a video clip subjected to content-preserving distortions should be similar to the ones extracted from the original video clip.
  - Security: Randomization is needed in order to prevent adversarial attack.
  - Computational efficiency: Due to the high dimension essential of video, time complexity of video hash function should be low.
- Usually, there exists a tradeoff between robustness and security: The more secure a hash function is, the less robust it is.

Applications of video hashing

- Automatic video clip identification in a video database or in broadcasting.
- Online search in a streaming video.
- Authentication of the video content.
- Content-based watermarking.

Recent methods of video hashing

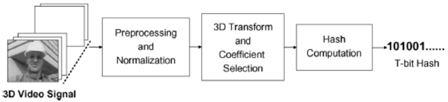
- S. C. Cheung and A. Zakhor, Efficient video similarity measurement with video signature, IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 1, pp. 5974, Jan. 2003.
- K. Kashino, T. Kurozumi, and H. Murase, A quick search method for audio and video signal based on histogram pruning, IEEE Trans. Multimedia, vol. 5, no. 3, pp. 348357, Sep. 2003.
- X.-S. Hua, X. Chen, and H.-J. Zhang, Robust video signature based on ordinal measure, in Proc. Int. Conf. Image Process. (ICIP 04), vol. 1. Singapore, Oct. 2004, pp. 685688.
- C. D. Roover, C. D. Vleeschouwer, F. Lefebvre, and B. Macq, Robust video hashing based on radial projection of key frames, IEEE Trans. Signal Process., vol. 53, no. 10, pp. 40204037, Oct. 2005.
- C. Kim and B. Vasudev, Spatiotemporal sequence matching for efficient video copy detection, IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 1, pp. 127132, Jan. 2005.

Recent methods of video hashing

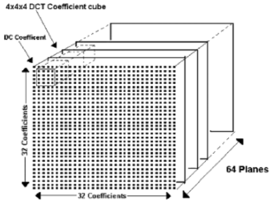
- S. Lee and C. D. Yoo, Video fingerprinting based on centroids of gradient orientations, in Proc. ICASSP 06, vol. 2. Toulouse, France, May 2006, pp. 401404.
- B. Coskun, B. Sankur, and N. Memon, Spatio-temporal transform based video hashing, IEEE Trans. Multimedia, vol. 8, no. 6, pp. 11901208, Dec. 2006.
- S. Lee and C. D. Yoo, Robust video fingerprinting based on affine covariant regions, in Proc. ICASSP 08, Las Vegas, NV, Apr. 2008, pp. 12371240.
- S. Lee and C. D. Yoo, Robust video fingerprinting for content-based video identification, IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 7, pp. 983988, Jul. 2008.
- Sunil Lee, Chang D. Yoo, and Ton Kalker, "Robust Video Fingerprinting Based on Symmetric Pairwise Boosting," IEEE Trans. Circuits and Systems for Video Technology, vol. 19, no. 9, pp. 1379-1388, Sept. 2009.

Typical algorithm1: Hashing based on 3-D DCT[1]

1. Core idea: Quantize the low frequency conponent coefficients of 3-D DCT as hashing values.
2. flowchart:



3. coefficient selection:



4. RBT: In order to increase security, we can introduce randomness to the cosine frequency coefficients.

[1] Baris etc, Spatio-Temporal Transform based video hashing, Trans. multimedia, 2006

Typical algorithm1: Hashing based on 3-D DCT[2]

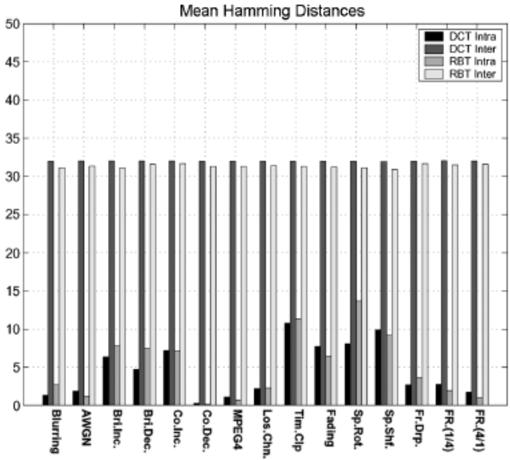
1. Performance experiment:

	Mean **Intra**-Hamming Distances		Mean **Inter**-Hamming Distances	
Modif. Type	**DCT**	**RBT**	**DCT**	**RBT**
Blurring	1.33	2.69	32.01	31.12
AWGN	1.86	1.13	32.02	31.33
Brightness Increase	6.35	7.81	32.02	31.1
Brightness Decrease	4.72	7.51	32.02	31.59
Contrast Increase	7.19	7.12	32.04	31.64
Contrast Decrease	0.29	0.18	32.01	31.29
MPEG4 Comp.	1.12	0.7	32.01	31.31
Lossy Channel	2.2	2.3	32.00	31.43
Clipping in Time	10.79	11.29	31.98	31.30
Fadeover	7.71	6.45	32.00	31.23
Frame rotation (3^o)	8.1	13.7	32.00	31.09
Frame shift (3%)	9.9	9.2	31.95	30.90
Frame drop (70%)	2.7	3.7	32.00	31.66
Framerate chge (1/4)	2.75	1.95	32.05	31.53
Framerate chge (4/1)	1.75	1.05	32.04	31.59

[2] Baris etc, Spatio-Temporal Transform based video hashing, Trans. multimedia, 2006

Typical algorithm1: Hashing based on 3-D DCT[3]

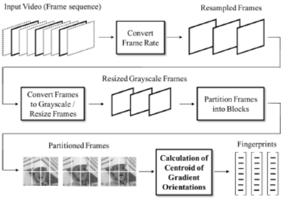
1. Performance experiment:



[3] Baris etc, Spatio-Temporal Transform based video hashing, Trans. multimedia, 2006

Typical algorithm2: Hashing based on centroid of gradient orientations[4]

1. flowchart:



2. calculate centroid of gradient orientations:

$$\nabla \mathbf{f} = [G_x \quad G_y] = \left[\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y}\right].$$

$$G_x = f[x+1, y, k] - f[x-1, y, k]$$
$$G_y = f[x, y+1, k] - f[x, y-1, k].$$

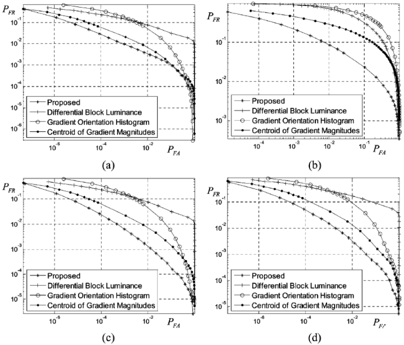
$$r[x, y, k] = \sqrt{G_x^2 + G_y^2}$$
$$\theta[x, y, k] = \tan^{-1}\left(\frac{G_y}{G_x}\right).$$

[4] Sunil Lee, Chang D. Yoo, Ton Kalker, Robust Video Fingerprinting for Content-Based Video Identification, IEEE Trans. circuits and systems for video technology, 2008

Typical algorithm2: Hashing based on centroid of gradient orientations

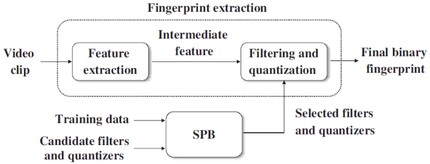
$$c[n,m,k] = \frac{\sum_{(x,y)\in B_{n,m,k}} r[x,y,k]\theta[x,y,k]}{\sum_{(x,y)\in B_{n,m,k}} r[x,y,k]}$$

1. performance comparisons:



Typical algorithm3: Hashing based on Symmetric Pairwise Boosting[5]

1. Core idea: Use some learning algorithm(symmetric pairwise boosting), instead of conventional heuristically and manually derived extraction methods(whose threshold is determined as mean or median value of the elements), to quantize the extracted intermediate features, so that the robustness and discriminability of the hashing method is increased simultaneously.
2. flowchart:



3. Step1: Extraction of Intermediate Features
block mean luminance (BML):

$$x_t(r, c) = \frac{1}{|B_{r,c,t}|} \sum_{(i,j) \in B_{r,c,t}} l_t(i, j)$$

[5] Sunil Lee, Chang D. Yoo, Ton Kalker, Robust video fingerprinting based on symmetric pairwise bo systems for video technology, 2009

Typical algorithm3: Hashing based on Symmetric Pairwise Boosting[6]

1. Step1(Cont.): or extract centroid of gradient orientations

$$x_t(r,c) = \frac{\sum_{(i,j)\in B_{r,c,t}} m_t(i,j)\theta_t(i,j)}{\sum_{(i,j)\in B_{r,c,t}} m_t(i,j)}$$

2. Step2: Use symmetric pairwise boosting to train a set of classifiers.

Input

N pairs of sequences of intermediate features $\{(X_{1:T}^{(1,n)}, X_{1:T}^{(2,n)}, y_n) | n = 1, \ldots, N\}$ with label $y_n \in \{-1, +1\}$.

Initialization

Distribution $d_n^{(1)} = \frac{1}{N}, n = 1, \ldots, N.$

Do for $m = 1, \ldots, M$

1) Find the classifier $h_m \in \mathcal{H}$ that minimizes the weighted error

$$\epsilon_m = \sum_{n=1}^{N} d_n^{(m)} \cdot \mathbf{I}\left[h_m(X_{1:T}^{(1,n)}, X_{1:T}^{(2,n)}) \neq y_n\right]$$

where $\mathcal{H}$ is a class of classifiers, and $\mathbf{I}[\varepsilon] = 1$, if the event ε occurs; $\mathbf{I}[\varepsilon] = 0$, otherwise.

2) Compute weight (confidence) of the chosen classifier

$$c_m = \log\{(1 - \epsilon_m)/\epsilon_m\}.$$

3) Update distribution

$$d_n^{(m+1)} = d_n^{(m)} \cdot \exp\left(-c_m y_n h_m\left(X_{1:T}^{(1,n)}, X_{1:T}^{(2,n)}\right)\right)/Z_m$$

where Z_m is a normalization factor.

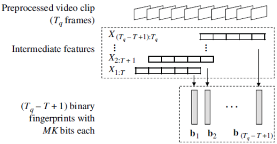
Output

M pairs of filter and quantizer $\{(f_m, Q_m) \mid m = 1, \ldots, M\}$ which parameterize the chosen M classifiers $\{h_1, \ldots, h_M\}$.

[6] Sunil Lee, Chang D. Yoo, Ton Kalker, Robust video fingerprinting based on symmetric pairwise bc[...] systems for video technology, 2009

Typical algorithm3: Hashing based on Symmetric Pairwise Boosting[7]

1. Step3: Use the classifiers to select the filters and quantizers that apply to the extractd intermediate features to get the final hashing.



$$\mathbf{b}_t = [Q_1(f_1(X_{t:(t+T-1)})) \cdots Q_M(f_M(X_{t:(t+T-1)}))]$$

2. Step4: Database Search and Fingerprint Matching

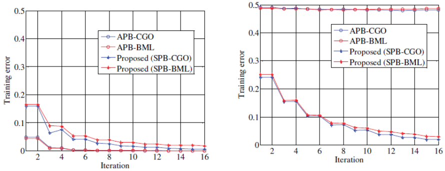
$$\mathbf{b}_t = [Q_1(f_1(X_{t:(t+T-1)})) \cdots Q_M(f_M(X_{t:(t+T-1)}))]$$

$$D(v_q, v_c) = \sum_{i=1}^{T_q-T+1} d_H(\mathbf{b}_i^q, \mathbf{b}_i^c)$$

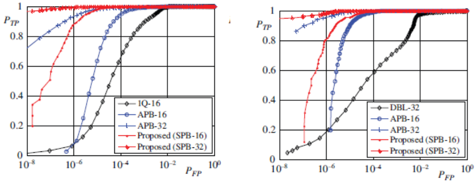
[7] Sunil Lee, Chang D. Yoo, Ton Kalker, Robust video fingerprinting based on symmetric pairwise boosting, IEEE Trans. circuits and systems for video technology, 2009

Typical algorithm3: Hashing based on Symmetric Pairwise Boosting[8]

1. Training error for (a) matching pairs and (b) non-matching pairs



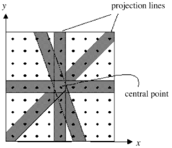
2. performance comparisons, with intermediate features as CGO, BML respectively.



[8] Sunil Lee, Chang D. Yoo, Ton Kalker, Robust video fingerprinting based on symmetric pairwise boosting, IEEE Trans. circuits and systems for video technology, 2009

Typical algorithm4: Hashing based on Radial Projections of Key Frames[9]

1. Step1: Radial hASH



$$-\frac{1}{2} \leq (x - x').\cos\phi + (y - y') \cdot \sin\phi \leq \frac{1}{2}.$$

$$P(\phi) = \frac{\sum_{(x,y)\in\Gamma(\phi)} I^2(x,y)}{\#\Gamma(\phi)} - \left(\frac{\sum_{(x,y)\in\Gamma(\phi)} I(x,y)}{\#\Gamma(\phi)}\right)^2.$$

$$D(n) = \sqrt{\frac{2}{N}} \cdot \sum_{\phi=0}^{N-1} \left(R(\phi) \cdot \cos\frac{\pi \cdot (2\phi + 1) \cdot n}{2N}\right)$$

[9] C. D. Roover, C. D. Vleeschouwer, F. Lefebvre, and B. Macq, Robust video hashing based on rad
Trans. Signal Process., 2005.

Typical algorithm4: Hashing based on Radial Projections of Key Frames[10]

1. Step2: Key Frame Selection
2. Boundary Frame Selection

- First, extract a feature from each frame of the video sequence.
- Second, use a metric $d(k, k')$ to measure the distance between the features extracted at time indices k and k'. The distance $d(k, k')$ is expected to measure the disparity between the kth and the k'th frames.
- Third, compare the distance values $d(k, k')$ to a threshold T. If $d(k, k') > T$, the kth frame is marked as being a boundary frame. In general, $k' = k - 1$.

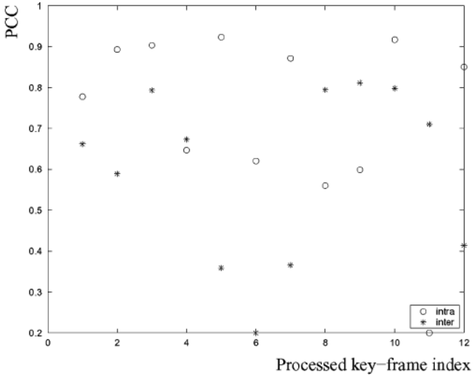
3. Key Frame Selection

$$r = \arg \min_{k_1 < k \leq k_2} d(k, k-1).$$

[10] C. D. Roover, C. D. Vleeschouwer, F. Lefebvre, and B. Macq, Robust video hashing based on radial projection of key frames, IEEE Trans. Signal Process., 2005.

Typical algorithm4: Hashing based on Radial Projections of Key Frames[11]

1. Performance experiment:



[11] C. D. Roover, C. D. Vleeschouwer, F. Lefebvre, and B. Macq, Robust video hashing based on radial projection of key frames, IEEE Trans. Signal Process., 2005.

Open problems

1. Whenever the brightness manipulation is taken to the extreme of saturation (too dark, clipped to 0 or too bright, saturated to 255), the hash function based on 3-D DCT suffers.[12]
2. Quantitatively analyse the robust and secure performances of video hashing algorithms.

[12] Baris etc, Spatio-Temporal Transform based video hashing, Trans. multimedia, 2006

Conclusions

1. Video hashing defines a feature vector that characterizes the video content, independently of "nonsignificant" distortions.
2. A good video hash function should be perceptual unique, robust, secure and computational efficient.
3. Video hashing technique has wide applications in video authentication and verification.