2010 IEEE International Symposium on Multimedia

Ring Fingerprint Based on Interest Points for Video Copy Detection

Guiguang Ding
School of Software
Tsinghua University
Beijing, China
dinggg@tsinghua.edu.cn

Rongxian Nie
School of Software
Tsinghua University
Beijing, China
nierx07@mails.tsinghua.edu.cn

***Abstract*—In the field of information security, assurance, copyright protection etc., content-based copy detection is more and more important, which consists of two technologies - fingerprint extraction and fingerprint matching. Fingerprints extracted from videos are mainly described as global video fingerprints and local video fingerprints. To make best use of the advantages and bypass the disadvantages of different video fingerprints, we propose the ring fingerprint based on interest points and ordinal measure in this paper. A last we examine the proposed method with lots of experiments and discuss the performance of approaches. The experimental results also demonstrate the effectiveness of the proposed method for video copy detection.**

Keywords- Interest Points; Ring Fingerprint; Video Copy Detection

I. INTRODUCTION

With the fast growth in the information technology, communication networks and multimedia broadcasting, masses of internet video sites are appearing, for example video on demand, network podcasts, streaming media, peer-to-peer system, and so on. Sharing and broadcasting the video with convenience, significantly increases the risk of unauthorized use of video content. Content-Based Copy Detection (CBCD) is a new effective approach for the digital copyright protection.

CBCD is different from digital rights management (DRM), digital fingerprint (DF) and digital watermark (DWM) in protection of digital content. DRM is a generic term that refers to access control technologies that can be used by hardware manufacturers, publishers, copyright holders and individuals to try to impose limitations on the usage of digital content and devices [1]. DRM technologies used are dependent on three crucial elements: technology, law and business model [2]. DF uses mathematical algorithms to generate a string of digits and numbers that represents the contents of an electronically-transmitted message [3]. The DWM needs to embed the additional information [4], which may reduce the content's quality. And the embedded information may be lost partly or totally when it is suffered from malicious attack, so that the embedded fingerprint may be not detected rightly. Now CBCD is being studied by more and more scholars. CBCD extracts the unique feature information (content fingerprint) from video content itself, and then the illegal distribution can be detected through its exclusive feature. The fingerprint of CBCD is fundamentally different from the fingerprint of DF and DWM. The fingerprint of DF and DWM is the additional information embedded into the video. And the fingerprint of CBCD is extracted from the video's content, which is unique information used to identify the video. So the fingerprint of CBCD is often referred to "content fingerprint" or "digital DNA".

There are both similarities and differences between CBCD and content-based video retrieval (CBVR). They all need to extract fingerprint and match fingerprint of video content. The difference is that the CBCD is to detect the copied videos and the CBVR is to search the similar videos [5]. A copy is not an identical or a near replicated video sequence but rather a transformed video sequence. These transformations can strongly change the signal (gamma and contrast transformations, overlay, shift etc...) therefore a copy can be visually less similar than other kinds of similar videos. And a similar video is only visually similar among videos in despite of their origins.

Based on the differences between CBCD and CBVR, there are two issues in CBCD: the robustness issue and the discriminability issue [6]. The robustness is that the illegal video is not visually similar with the original video after transformation, but this relationship should be detected by CBCD technology. The discriminability issue is that CBCD technology should not take two similar videos which are not copies as copies. So CBCD is to compare video segments to check if they have same video content.

Fingerprint extraction and fingerprint matching are two key technologies of CBCD. In fingerprint extraction, the global descriptors (color histogram, ordinal measure and etc.) are robust to transformations of video but poor in the discriminability issue for the insensitivity to local changes. But the local descriptors (based on interest points, based on time-space information, and etc.) have a good performance in the discriminability issue but less robust for the insensitivity to global changes. To reach the equilibrium of advantages and disadvantages of global descriptors and local descriptors, the ring fingerprint based on interest points and ordinal measure is proposed in this paper.

II. RELATED WORKS

A. Application of CBCD

There are several promising applications of CBCD, e.g., the copyright control, the media tracking, the video search systems and etc.

978-0-7695-4217-1/10 $26.00 © 2010 IEEE
DOI 10.1109/ISM.2010.59

In copyright control, CBCD has many advantages compared with the traditional methods. CBCD doesn't need to embed additional information into original video content. It only extracts fingerprint from the video content itself and matches finger printers of different videos [7].

Media tracking is the problem of keeping track of when and where a particular known piece of media has been used. Examples include, tracking when (at what time) and where (which channel) a particular TV commercial was aired. Analogously, on the web, when (date and time) and where (URL) was particular piece of video content available [8]. Thus, CBCD is useful for media tracking.

CBCD can also be used for video search engine. The results of a video search include many same videos or copied videos. CBCD can filter the search results for users. And CBCD is important for business intelligence and advertisement tracking, law enforcement investigations, etc. It offers an alternative to watermarking.

B. *Workflow of CBCD*

CBCD system generally consists of two parts [5]: off-line indexing and on-line retrieval. The off-line indexing component extracts the video signatures from the reference content database (containing the original videos), stores them in the reference fingerprint database and builds a specific index of this database. In the application context, the original content database increases at a slow rate over time, so indexing can be done off line.

The on-line retrieval component checks whether a query video is a copy of an original video stored in the reference content database. Fingerprints are first extracted from the candidate video, then the fingerprint is used as a query to retrieve similar fingerprint from the reference fingerprint database.

So the fingerprint extracting is critical for off-line indexing and fingerprint matching is important for on-line retrieve. The fingerprint extracting and matching are the two main steps in CBCD. Most of the existing proposals for video CBCD follow this generic workflow but employ different content extracting schemes and different solutions for matching-based retrieval.

C. *Methods of CBCD*

To extract fingerprint, the video is firstly segmented into shots and each shot is represented by one or more key-frames [9]. The key frames are then represented by certain high dimensional feature vector. Finally the sequence of the key frames' features is taken as the fingerprint of the whole video. Video fingerprint includes two classes: global descriptors and local descriptors.

The global descriptors of video based on the motion, ordinal intensity and color histogram are compared by Hampapur and Bolle in [10]. A representative global descriptor – the ordinal measure was originally proposed by Bhat and Nayar [11-12] for computing image correspondences. In ordinal measure, the image is partitioned into M×N equal-sized blocks. After sorting the average intensity values of blocks, each block can get an ordinal number. The vector in m×n dimensions with ordinal numbers of blocks is the fingerprint of the image. It is not complex to acquire the ordinal measure, so ordinal measure was adapted by Mohan in [13] for video purpose and the corresponding matching approach was proposed. Therefore the ordinal measure is used to other fields [14-18]. Kim and Vasudev [19] combined spatial matching of ordinal measure obtained from the partitions of each key frame and temporal matching of temporal signatures from the temporal trails of the partitions. The paper [20] examined the image distance measures of smoothed image difference, histogram intersection (RGB, HSV and Gradient Direction), Hausdorff distance [21], local edge representation and invariant moments in context of video copy detection and compared their performances.

Besides the global descriptors, local features or interest points provide compact and abstract representations of patterns. Joly, Buisson and Frelicot [22] described a local descriptor - an improved version of the Harris interest point detector and a differential description of the local region around each interest point. Laptev and Lindeberg in [23] build on the idea of the Harris and Forstner interest point operators and detect local structures in space-time where the image values have significant local variations in both space and time, to detect spatio-temporal events. Law-To [24] presented a comprehensive comparative study of different state-of-state techniques for video copy detection: global video descriptors based on temporal, ordinal measurement and temporal ordinal measurement; local descriptors based on Harris interest points and space time interest points (STIP). And Law-To also evaluated two existing CBCD systems (developed by INRIA and INA) in paper [25].

III. RING FINGERPRINT EXTRACTING BASED ON INTEREST POINTS

As a global descriptor, ordinal measure reflects the relative intensity distribution [12]. The video frame or key frame is partitioned into N=m×n equal-size blocks and the average gray level in each block is computed. Then the set of average intensities is sorted in ascending order and the rank is assigned to each block. The ranked m×n dimensional vector is the fingerprint (ordinal measure) of the frame. The sequence of all frames' fingerprint is the fingerprint of the video. So the number of block is an important influence factor to robustness of ordinal measure. For the modification of "letter-box" or "pillar-box", if the frame is partitioned into 2×2 blocks, the two videos with the original content and the modified content have identical ordinal measures. If the frame is partitioned into 3×3 blocks, the ordinal measures of two videos are distinct.

As reflecting the intensity distribution, ordinal measure is more robust and less discriminable, but fingerprint based on interest points, as a local descriptor, is more discriminable and less robust. Neither ordinal measure nor interest points can be both robust and discriminable. To meet the two essential properties (robustness and discriminability) required by CBCD, this paper proposes the ring fingerprint based on interest points and ordinal measure.

To compute the ring fingerprint, interest points are extracted from each key frame of the video firstly. Then each key frame is divided into equal-area circle rings based on the boundaries of key frame and the center point which can be found from the interest points. The equal-area circle rings are further divided into equal-area sectors. Based on the intensity of equal-area sectors, we can get the ordinal measure of the key frame, and the sequence of the ordinal measure vectors is taken as the ring fingerprint of the whole video. The ring fingerprint extraction method is illustrated as figure 1.

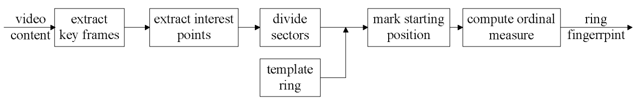


Figure 1. The ring fingerprint extraction

The steps of the ring fingerprint extraction can be described as following in details.

(1) Extract key frames

For the given video V, we can use the key frame extraction algorithm to extract V's key frames $(P_1, P_2, ..., P_n)$. Here we extract a frame per second as the key frame, because the shot divided by the algorithm may be distinct when the video is suffered from malicious attacks. This paper doesn't select the frame from the shot as the key frame but take the frame per second as the key frame.

(2) Extract interest points

For each key frame P_i, we extract its interest points $p_1(x_1, y_1)$, $p_2(x_2, y_2)$, ... , $p_n(x_n, y_n)$ (where x_i is the horizontal axis value of P_i and y_i is vertical axis value of P_i). This paper takes the surf feature points as the interest points.

(3) Compute the centre and radius of the circle

The centre point $p_0(x_0, y_0)$ can be got from the equation (1) and (2).

$$x_0 = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$y_0 = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{2}$$

If w and h represent the wide and high of the key frame P_i, radius of the circle can be computed from equation (3).

$$R = \min(x_0, y_0, |w - x_0|, |h - y_0|, d_p) \tag{3}$$

where $x_0, y_0, |w - x_0|, |h - y_0|$ are distances from the centre $p_0(x_0, y_0)$ to the boundaries of the key frame, and d_p is the longest distance from the center $p_0(x_0, y_0)$ to each interest point as equation (4).

$$d_p = \max(\sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2}) \text{ (i=1, 2, ... , n)} \tag{4}$$

(4) Divide equal-area rings and sectors

The circle with the centre point $p_0(x_0, y_0)$ and R as radius is divided into m equal-area rings, so the radius of each ring is $r_i = \sqrt{i \cdot R}$ (i=1,2,...,m). Here we set m = 4 as illustrated in Figure 2.

Then the rings are divided into equal-area sectors. Here the paper uses four lines to divide each ring into eight equal-area sectors n0, n1, n2, ... , n7 as illustrated in Figure 2.

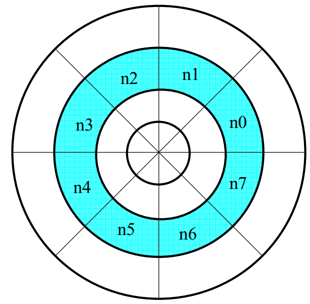


Figure 2. Divide equal-area rings and sectors

(5) Compute the ring's values

If the sectors' average gray values of the k th ring are t0, t1, t2, ... , t7, the k th ring's value is :

$$v_k = \sum_{i=0}^{n-1} g(t_i) * h(k) \tag{5}$$

where $g(t_i) = \begin{cases} 1 & t_i > t_{i-1} \\ 0 & t_i \le t_{i-1} \end{cases}$, i=1,2,...,7, when i = 0, i-1 = 7; and $h(k) = 2^k$, k = i. which is illustrated as Figure 3.

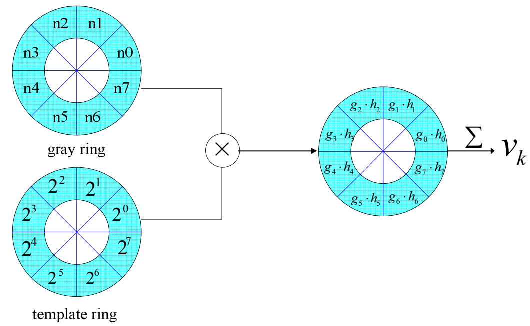


Figure 3. Compute the ring value

(6) Mark starting position

Computing the k th ring's value v_k, the template ring can be rotated through 45 degrees every time, which means in equation (5), $h(k) = 2^k$, where k = i, i+1, i+2, ... , i+7, we can get $v_k^0, v_k^1, ..., v_k^7$ respectively. If v_k^l is the minimum value in $v_k^0, v_k^1, ..., v_k^7$,

$$v_k = \min(v_k^0, v_k^1, ..., v_k^7) = v_k^l \quad 0 \le v_k^l \le 7 \tag{6}$$

Then we take the l as the starting position.

(7) The ring ordinal measure

Taking the l as the starting position, the sectors' average gray value of the k th ring $t' = (t_0', t_1', ..., t_7')$ are sorted to $\bar{t} = (\bar{t}_0, \bar{t}_1, ..., \bar{t}_7)$. Based on the t' and $\bar{t}$, we can compute the kth ring's ordinal measure $p_k = (p_k^0, p_k^1, ..., p_k^7)$ from equation (7).

$$p_i = k \quad \text{if } t_i' = \bar{t}_k \quad (i = 0,1,...,7) \tag{7}$$

(8) The ring fingerprint of the video

From all concentric rings' ordinal measure vectors of the key frame P_i, the ring fingerprint of the key frame is $P_k = (p_0, p_1, ..., p_m)$, where p_i is the ith ring's ordinal measure. The sequence of the ring fingerprint of the key frame can be taken as the whole video's ring fingerprint $P = (P_0, P_1, ..., P_n)$, where $P_0, P_1, ..., P$ are the key frames' ring fingerprint.

IV. Experimental Results

In the experiment, our approaches are evaluated based on the database MUSCLE-VCD-2007 from INRIA-Roquencour, which contains about 100 hours of video materials from different source: web video clips, TV archives, movies and others. Each video in database is in MPEG-1 format at PAL frame rate of 30 fps. The ground truth is set up by copying and transferring ten videos selected from the database as original videos. The transformations of original videos are brightness increased doubly, brightness decreased by half, contrast increased by 200%, contrast decreased by 50%, Gaussian radius-2 blurriness, quick blurriness 200%, letterbox, pillar-box, crop 10% with black window and insertion of a logo as illustrated in Figure 4.

The 100 transformation videos are added into the database, so the database consists of 200 hours of video materials which is extracted about 720000 key frames at equal distance of one second. The query sequences are the sequences and sub-sequences of the ten original videos. The results of videos in database matched with query sequences can estimate the approaches.

Figure 4. Transformations

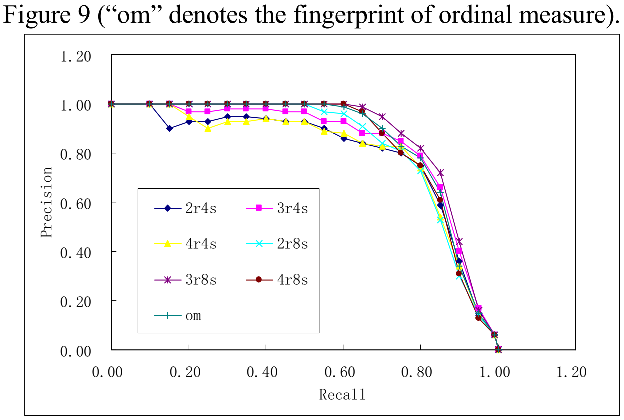
In order to comprehensively evaluate the ring fingerprint and the matching method based on bipartite graph (as explained in section 3 and 4), we take the traditional ordinal measure and sliding window matching as the compared fingerprint and matching method. The performance of approaches is evaluated by the PR (Precision-Recall) curves and accuracy. If tp, tn, fp and fn represent number of true-positives, true-negatives, false-positives and false-negatives respectively. Then the precision, recall and accuracy can be got as following formulas:

$$\Pr ecision = \frac{tp}{tp + fp} \tag{9}$$

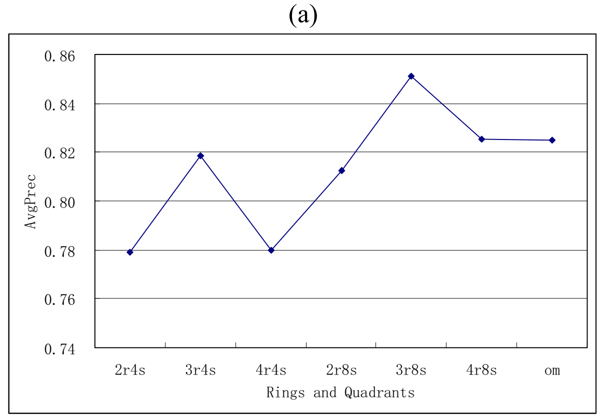
$$\text{Re}\,call = \frac{tp}{tp + fn} \tag{10}$$

$$Avg\,\Pr ec = \frac{\sum_{i=1}^{N} (\Pr ecision)_i}{N} \tag{11}$$

Firstly, the traditional ordinal measure and the ring fingerprint with different parameters are compared using the same matching method of slide window matching. The result is illustrated as Figure 5. We extract the ring fingerprint of all videos by dividing the circle into 2, 3 and 4 (m=2, 3, 4) equal-area rings and then the rings are divided into 4 and 8 equal-area sectors in the 4th step in section 3, which is denoted by 2r4q, 3r4q, 4r4q, 2r8q, 3r8q, 4r8q respectively in Figure 9 ("om" denotes the fingerprint of ordinal measure).



(a)



(b)

Figure 5. The PR-curves (a) and AvgPrec curve (b) of different rings and quadrants

From the PR-curves in Figure 5 (a), the performance of ring fingerprint with 3 rings and 8 sectors is the best. If the numbers of rings and sectors (e.g. the parameter is 2r4s) is too little, it can bring more "dirty" videos in the result, which can reduce the discriminability of the algorithm. At the same time, if the dimensionality is too high (e.g. the parameter is 4r8s) in the ring fingerprint, many "clean" videos are estimated wrongly with poor robust. The AvgPrec curve in Figure 5 (b) can show the performance with different parameter more distinctly.

To evaluate the ring fingerprint's sensitivity to different transfers of video, we compare the ranks of the best ring fingerprint (3r4s) with the ordinal measure in Figure 6. When the transfers are brightness 200%, brightness 50%, contrast 200%, pillar box and logo, the ranks of ring fingerprint are smaller than ordinal measure, which means the ring fingerprint is more sensitive than ordinal measure. And for the same reason, the ordinal measure is more sensitive than ring fingerprint for the transfers of contrast 50%, Gaussian blurriness 200% and quick blurriness 200%. They almost have the same sensitivity to the transfers of crop 10% and letterbox.

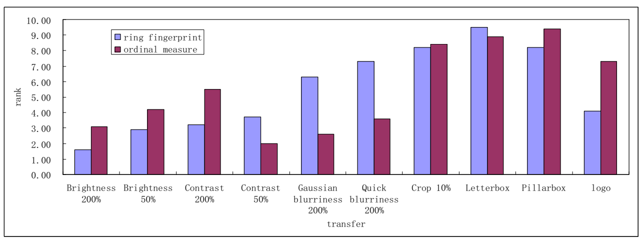


Figure 6. The fingerprints' ranks to different transfers

In order to show the sensitivity to different transfers more clearly, the value of $Rank_{\text{global/local}}$ is defined as equation (12).

$$Rank_{\text{global/local}} = \frac{Avg-Rank_{\text{global transfer}}}{Avg-Rank_{\text{local transfer}}} \tag{12}$$

In equation (12) the value $Avg-Rank_{\text{global transfer}}$ is the average rank of the global transfers, which include brightness 200%, brightness 50%, contrast 200%, contrast 50%, Gaussian radius-2 blurriness and quick blurriness 200%. And similarly, the value $Avg-Rank_{\text{local transfer}}$ is the average rank of the local transfers, which include crop 10%, letterbox, pillar box and insert logo.

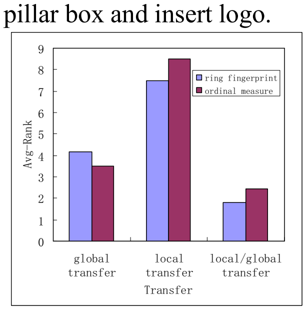


Figure 7. The fingerprint's sensitivity to different transfers

As shown in Figure 7, for the global transfer, the ordinal measure is more sensitive than the ring fingerprint, and the ring fingerprint is more sensitive than ordinal measure for the local transfer. But in view of the overall situation, $Rank_{\text{global/local}}$ of the ring fingerprint is more close to the value 1 than ordinal measure, so the copy video detecting performance of the ring fingerprint is better when the video's transfer is unknown in the general situation.

From the Figures above, we can see that the ring fingerprint is better than the traditional ordinal measure with the same matching method of slide window matching illustrated as Figure 5, Figure 6 and Figure 7.

V. Conclusions

In this paper, we propose the ring fingerprint based on interest points and ordinal measure. The experiment results demonstrate their effects through comparing with the traditional fingerprint algorithms. And the performances of fingerprint algorithms with different parameter or threshold are discussed at the same time.

Acknowledgments. The authors acknowledge the support received from the National Natural Science Foundation of China (Grant No. 60972096) and the National 863 Plans Projects (Grant No. 2009AA01Z410).

References

[1] "Digital Rights Management," Wikipedia; http://en.wikipedia.org/wiki/ Digital_rights_management.

[2] N. Rump.: Definition, Aspects, and Overview. Digital Rights Management, Springer Berlin / Heidelberg, pp.3-15(2003)

[3] Jen-Hao H., Cheng-Hung L.: Efficient digital fingerprint production and evaluation for secure communication systems based on genetic algorithms. Lecture Notes in Computer Science, Springer Berlin / Heidelberg, pp.628-633(2004)

[4] Frank Y., Shih and Yi-Ta Wu: Information hiding by digital watermarking", Studies in Computational Intelligence, Springer / Heidelberg, pp. 205-223(2009)

[5] L. Law-To, O. Buisson, V. Gouet-Brunet, N. Boujemaa: Robust voting algorithm based on labels of behavior for video copy detection", Proceedings of 14th annual ACM international conference, pp.835-844(2006)

[6] J. Law-to, A. Joly, L. Joyeux: Video and image copy detection demo. Proceedings of the 6th ACM international conference, pp.97-100(2007)

[7] De Roover, C.; De Vleeschouwer, C.; Lefebvre, F.; Macq, B.: Robust video hashing based on radial projections of key frames. IEEE Transactions on Signal Processing, Volume 53, Issue 10, Part 2, pp. 4020 – 4037(2005)

[8] A. Hampapur and R.M.Bolle.: Feature based indexing for media tracking. IEEE Int. Conf. Multimedia and Expo (ICME), pp.67-70(2000)

[9] Zhonghua S., Kebin J., Hexin C.: Video key frame extraction based on spatial temporal color distribution. Proceedings of the 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Precessing, pp.196-199(2008)

[10] Hampapur, A., Bolle, R. M., and Hyun, K.-H.: Comparison of sequence matching techniques for video copy detection. Proceedings of SPIE: Storage and Retrieval for Media Database, pp.194-201(2002)

[11] Bhat, D.N. and Nayar, S.K.: Ordinal measures for visual correspondence. Columbia University, Computer Science, Technology reports, CUCS-009-96, pp.351-357(1996)

[12] Bhat, D.N. and Nayar, S.K.: Ordinal measures for image correspondence. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 4, pp.415-423(1998)

[13] R. Mohan: Video Sequence Matching. IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 6, pp. 3697-3700(1998)

[14] Aurnhammer, M. and Mayoral, R.: Improving Seismic Horizon Matching by Ordinal Measure. Proceedings of the 17th International Conference on Pattern Recognition, Volume 3, pp. 642 – 645(2004)

[15] Adrian N. Evans: Cloud Tracking Using Ordinal Measure and Relaxation Labeling IGARSS '99 Proceedings. IEEE 1999 International, vol.2, pp. 1259-1261(1999)

[16] Z. Sun and T. Tan: Ordinal Measures for Iris Recognition. IEEE Transaction on Pattern Analysis and Machine Intelligence. pp. 2211-2226(2009)

[17] X.-S. Hua, X. Chen, and H.-J. Zhang.: Robust video signature based on ordinal measure. In International Conference on Image Processing, pp. 685-688(2004)

[18] Cheikh, F.A., Cramariuc, B., Partio, M., Reijonen, P. and Gabbouj, M.: Ordinal measure based shape correspondence. EURASIP Journal on Applied Signal Processing, Vol. 5, pp. 362-371(2002)

[19] Chang Kim and Vasudev B.: Spatiotemporal Sequence Matching for Efficient Video Copy Detection. IEEE Transaction on Circuits and Systems for Video Technology, pp. 127-132(2005)

[20] A. Hampapur and R. M. Bolle: Comparison of Distance Measure for Video Copy Detection. in Proc. IEEE Int. Conf. Multimedia and Expo (ICME), pp. 188-192(2001)

[21] D.P. Huttenlocher, D. Klanderman, W.J. Rucklidge: Comparing Images Using the Hausdorff distance. International Conference on Computer Vision and Pattern Recognition, pp. 705-706(1993)

[22] A. Joly, O. buisson, and C. Frelicot.: Content-based Copy Detection Using Distortion-based Probabilistic Similarity Search. IEEE Transactions on Multimedia, pp. 293-306(2007)

[23] I. Laptev and T. Lindeberg.: Space-time interest points. In ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision, Washington, DC, USA, IEEE Computer Society, pp. 432-429(2003)

[24] J. Law-To, L. Chen, A. Joly, I. Laptev, O. Buisson, V. Gouet-Brunet, N. Boujemaa and F. Stentiford: Video Copy Detection: a Comparative Study. In ACM International Conference on Image and Video Retrieval, Amsterdam, The Netherlands,pp.371-378 (2007)

[25] C.-Y. Chiu, C.-H. Li, H.-A. Wang, C.-S. Chen and L.-F. Chien: A Time Warping Based Approach for Video Copy Detection. In ICPR'06, pp. 228-231(2006)