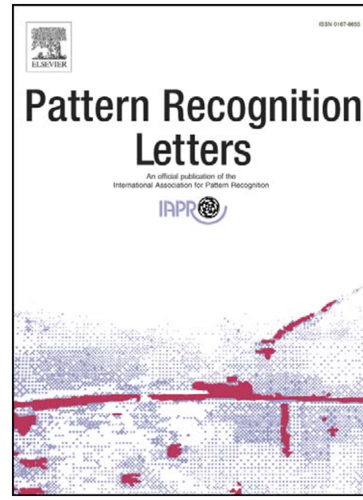
Accepted Manuscript



Video Sequence Matching based on Temporal Ordinal Measurement

Li Chen, F.W.M. Stentiford

PII: S0167-8655(08)00169-4
DOI: 10.1016/j.patrec.2008.05.015
Reference: PATREC 4432

To appear in: *Pattern Recognition Letters*

Received Date: 13 May 2007
Revised Date: 22 March 2008
Accepted Date: 12 May 2008

Please cite this article as: Chen, L., Stentiford, F.W.M., Video Sequence Matching based on Temporal Ordinal Measurement, *Pattern Recognition Letters* (2008), doi: 10.1016/j.patrec.2008.05.015

Video Sequence Matching based on Temporal Ordinal Measurement

Li Chen[a1] and F. W. M. Stentiford[b]
[a] University College London, Adastral Park Campus, Martlesham Heath, Ipswich, Suffolk, IP5 3RE, UK, E-mail: li.chen00@gmail.com
[b] University College London, Adastral Park Campus, Martlesham Heath, Ipswich, Suffolk, IP5 3RE, UK. E-mail: f.stentiford@adastral.ucl.ac.uk

[a] Corresponding author. Tel.: +44 1223 488529

Abstract

This paper proposes a novel video sequence matching method based on temporal ordinal measurements. Each frame is divided into a grid and corresponding grids along a time series are sorted in an ordinal ranking sequence, which gives a global and local description of temporal variation. A video sequence matching means not only finding which video a query belongs to, but also a precise temporal localization. *Robustness* and *discriminability* are two important issues of video sequence matching. A quantitative method is also presented to measure the robustness and discriminability attributes of the matching methods. Experiments are conducted on a BBC open news archive with a comparison of several methods.

Keywords: Video sequence matching; Ordinal measure; Video copy detection

[1]This work within this paper was done when Li Chen worked in University College London, Adastral Park Campus, Martlesham Heath, Ipswich, Suffolk, IP5 3RE, UK. She now works with Virage Security and Surveillance, Autonomy Systems Ltd. Cambridge Business Park, Cambridge, CB4 0WZ.

1. Introduction

Digital videos are increasingly broadcasted to millions of homes and delivered on Internet. This has encouraged much work on video browsing, video retrieval, video editing and video monitors in commercial as well as research areas [1-5, 9, 11, 14]. Video sequence matching is a basic requirement for the effective and efficient delivery of media data and the protection of intellectual property right. One example is monitoring commercial TV channels to check whether or not there is illegal usage of unauthorised video records and this equally applies to material placed on the Internet [1, 2]. Another example is tracking interesting video clips in a database to find its origins for different purposes such as browsing or editing. The problem of video sequence matching is defined in this paper as that of determining the presence of a video clip in the database and also the precise localization of a query video in a target video. Temporal localization has not been explored in most video sequence matching except [3, 6, 9]. For example, Cheung and Zakhor [8, 12] used a small set of frames as seeds, and the similarity between two video sequences is estimated by comparing their frames to seeds respectively. This approach cannot obtain the temporal location of a short video sequence in a long one. We argue that in many applications precise localization is important in addition to finding its existence in a target video. This is relevant for example, in monitoring a TV channel in which copied video sequences are located in a long time stream, and in analysing statistics of particular broadcast advertisements. In this context, precise localization is essential to the applications.

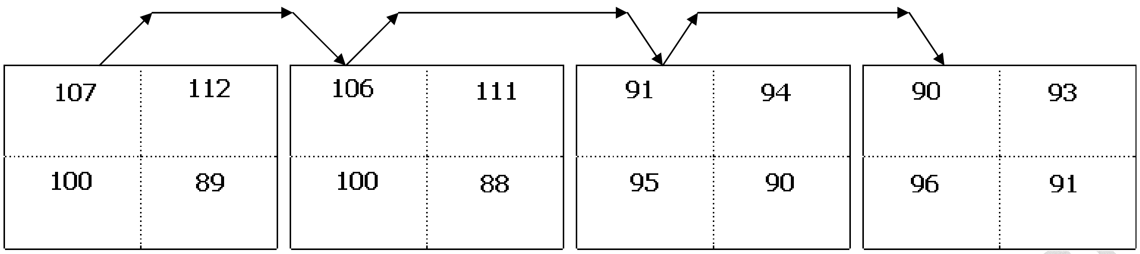
There are two approaches to the protection of copyright: watermarking and content based copy detection (CBCD). Watermarking inserts information into the media prior to distribution, which can be later extracted to establish ownership. However, sufficient fidelity and robustness of effective watermarking algorithms are not yet available. The primary theme of CBCD is "the media itself is the watermark", i.e. the media (video, audio, image) contains enough unique information that can be used for detecting copies [5]. In this regard, considerable effort has been devoted to effective representation of video signatures and similarity matching.

Recently Yang and Tian [2] gave a general survey on several content based video identification methods including video signature, matching methods, identification levels and application modes. Kim and Park [11] presented a modified Hausdorff distance calculation method for video sequence matching based on key frames extracted by directed divergence of histograms between successive frames. Oostveen, Kalker and Haitsma. [9] presented the concept of video *fingerprinting* as a tool for video identification and developed a spatial-temporal fingerprinting based on the differential of luminance of partitioned grids in spatial and temporal regions. Mohan [3] first introduced an ordinal measure [13] for video sequence matching and demonstrated its efficiency and effectiveness for matching videos based on similarity of "actions". Hua, Chen and Zhang [6] applied an ordinal measure to spatial arrangement while a different distance calculation method, called sequence shape similarity, was applied to video sequence matching. Furthermore dynamic programming was employed to locate a short query video clip in a long target. These ideas contrasted with traditional image (frame-wise) similarity measures. The distance between two video clips is calculated by adding and averaging distances of corresponding frame pairs [3, 5, 14], the statistics of matched frames [6], or keyframe-to-keyframe comparison [11]. Hampapur and Bolle [15] gives a comparison of up to 8 frame-wise based sequence matching for video copy detection.

Hampapur, Bolle and Hyun [5] later introduced motion based signature for copy detection and compared intensity and colour-based signatures: they found that region based ordinal matching is superior to global colour features; also intuitive global features were found to be affected by different distortions due to its weak representation of image signature that ignored the spatial distribution of colours. Kim and Vasudev [4] argued that ordinal matrix in terms of 2x2 partitioned frames is robust to different display format conversions including “letter-box” and “pillar-box” transformations. More importantly, a temporal signature is introduced: simply assign -1/0/1 to record the shift of intensity values between corresponding regions in the two adjacent frames in a video. Their experiments demonstrated improved performance compared to the ordinal measure based spatial signature matching in [5]. We argue that such a temporal signature is too inaccurate to obtain precise location in videos, and it does not give a global description of temporal variation. To support this, we propose a temporal ordinal measurement method for video similarity matching.

There are two issues concerning the devising of video sequence matching: *robustness* and *discriminability*. Robustness determines the tolerance of the system to different digitization and encoding processing that give rise to several distortions and transforms including additive Gaussian noise, illumination shift, contrast changes and different display formats, etc. *Discriminability* enables the system to disregard irrelevant videos and reduce the false alarm rate. The contribution of this paper is two fold: firstly we propose a video sequence matching method based on temporal ordinal measurement and secondly it provides precise temporal localisation. An evaluation method is presented to measure the robustness and discriminability of the matching methods in a quantitative fashion. The remainder of this paper is arranged as follows. The proposed video sequence matching method is presented in Section II. Experiments are conducted on the BBC open news archives that compare several methods in Section III, followed by the evaluation of robustness and discriminability in Section IV. Further discussion and conclusions are addressed in Section V and Section VI respectively.

(a)



(b)

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \\ 3 & 4 & 2 & 1 \end{bmatrix} \begin{array}{l} - \quad top \quad left \quad region \\ - \quad top \quad right \quad region \\ - \quad bottom \quad left \quad region \\ - \quad bottom \quad right \quad region \end{array}$$

(c)

Fig. 1. An example of a temporal ordinal measure (a) A video clip with four frames partitioned into 2x2 (b) Average grey values of each region overlapped on corresponding grids (c) Temporal ordinal ranking matrix based on grey values with time series

2. Temporal Ordinal Measurement of Video Sequence

Previous work in describing temporal information normally focuses on two successive frames; for example [4, 5, 9] extract intensity change or motion between adjacent frames. This approximate temporal estimation of a video sequence does not carry a global temporal description only local changes; it is only by chance that there are important variations within two adjacent frames. We observed that ordinal measures along the time series can describe local temporal variations as well as global ones.

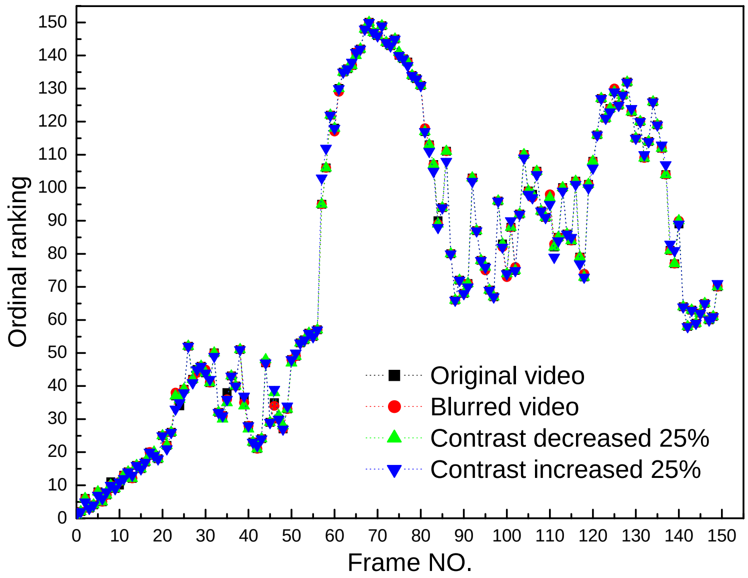


Fig. 2. An example of temporal ordinal ranking of three transformed copies of a video

Given a query video $V_q = [v_q^1, v_q^2, ..., v_q^M]$ and arbitrary target video $V_t = [v_t^1, v_t^2, ..., v_t^N]$ (M<<N) in a database, a video sequence match happens if the dissimilarity between V_q and $V_t^i = [v_t^i, v_t^{i+1}, ..., v_t^{i+M-1}]$ is less than a threshold $\varepsilon \in [0,1]$ where *i* is the *temporal position* of the matching. Assume that each frame is partitioned into K regions. The ranking matrix of region *k* in a query video, denoted by $V_q(k) = [v_q^1(k), v_q^2(k), ..., v_q^M(k)]$, is described by λ_q^k, and for a target sub-video $V_t^p = [v_t^p, v_t^{p+1}, ..., v_t^{p+M-1}]$, the ordinal measure of region *k* is written as $\lambda_t^{p,k}$. Fig. 1 illustrates an example of a temporal signature based on an ordinal measure on a 2x2 partition, where arrows show the temporal ordinal ranking for the top left region.

The distance is calculated as:

$$D\,(V_q, V_t^p) = \frac{1}{K}\sum_{k=1}^{K} d(\lambda_q^k, \lambda_t^{p,k})) \qquad (1)$$

where

$$d(\lambda_q^k, \lambda_t^{p,k}) = \frac{1}{C_M}\sum_{i=1}^{M}\left|\lambda_q^k(i) - \lambda_t^{p,k}(p+i-1)\right| \qquad (2)$$

C_M is a normalising factor, which is the maximum distance between two ranking matrices λ_q^k and $\lambda_t^{p,k}$. C_M is obtained when the two permutations of ordinal ranking matrices are the reverse of each other; that is, it is calculated as $C_M = \sum_{i=1}^{M}\left|M+1-2*i\right|$. For example, if M = 9, then C_M = 40.

Fig. 2 illustrates an example of temporal ordinal ranking of three transformed copies for one region from the same source with 150 frames. The three transformations include Gaussian radius-2 blur, contrast increased and decreased by 25% of the original video. It demonstrates that the temporal ordinal ranking matrices are nearly identical.

In order to process videos with different frame rates, or videos with frame drops due to compression, a resample technique can be applied as a pre-processing step to solve this issue. Alternatively keyframe extraction techniques can be used to approach the problem. To simplify the research reported in this paper we consider videos with the same frame rates.

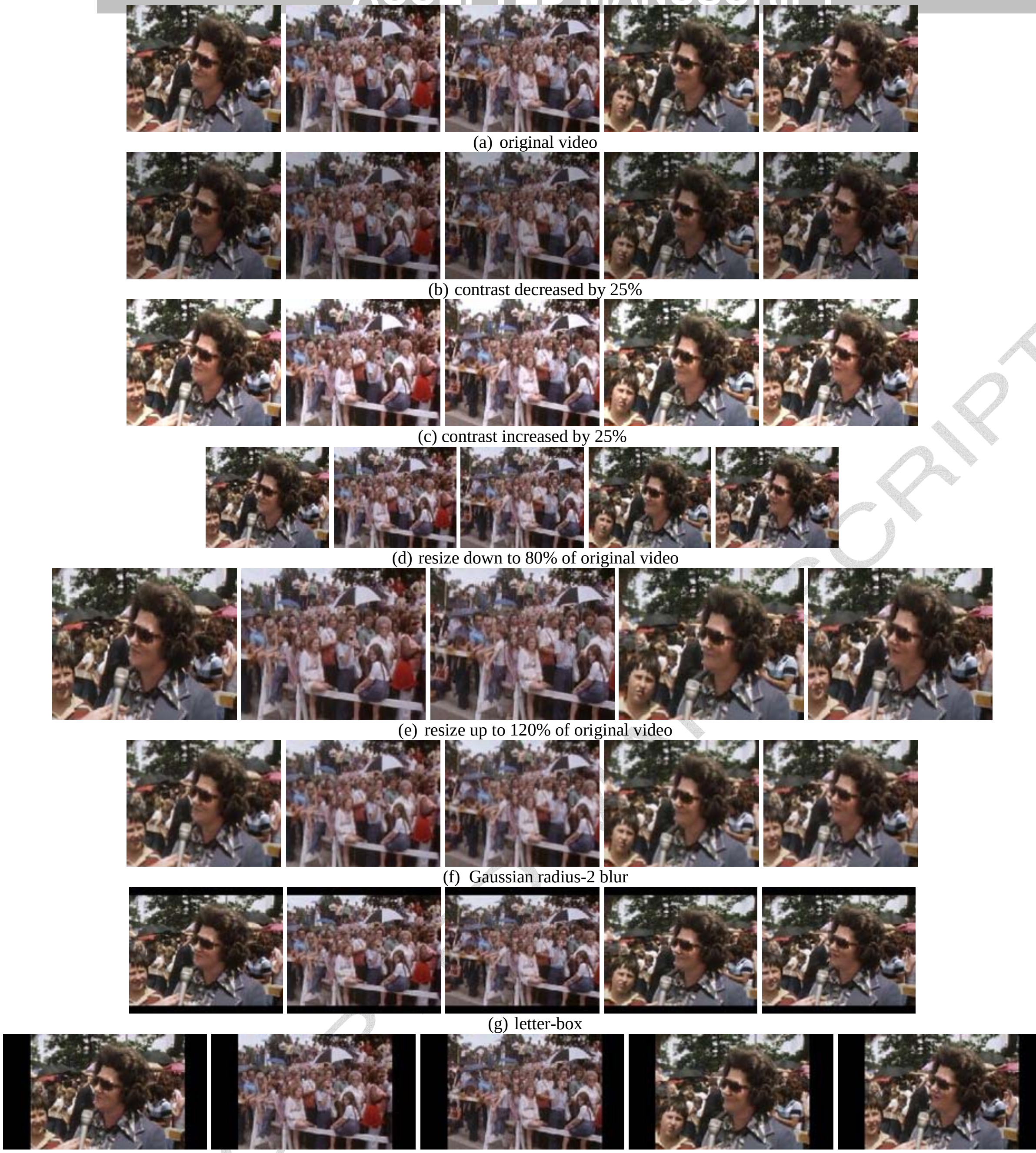
3. Experiments

3.1 Experiment setup

Experiments were carried out on the BBC open news archives [7]. 79 videos (285 535 frames, about 3.1 hours) cover different topics including conflicts and wars, disasters, personalities and leaders, politics, science & technology, and sports (shown in Table I). 8 query videos were randomly extracted from the database with 150 frames and 7 transformations were applied to each video. These transforms (see Fig. 3) were

- contrast increased by 25%
- contrast decreased by 25%
- resize down to 80% of original video
- resize up to 120% of original video
- Gaussian radius-2 blur
- letter-box
- pillar-box (4:3 to 16:9)

In addition the query videos were cut to 50 and 100 frames to test the performance of the proposed method for different lengths. Thus a total of 192 (=8x8x3) video clips were input as query videos. A query video $V_q = [v_q^1, v_q^2, ..., v_q^M]$ went through the following pseudo-code to find a match in the database.



(a) original video

(b) contrast decreased by 25%

(c) contrast increased by 25%

(d) resize down to 80% of original video

(e) resize up to 120% of original video

(f) Gaussian radius-2 blur

(g) letter-box

(h) pillar-box

FIG. 3. EXAMPLES OF EVERY 30TH FRAME FROM ORIGINAL QUERY VIDEO AND ITS TRANSFORMED COPIES

TABLE 1 INDEX OF VIDEO DATABASE BY LOCATION AND TOPICS (ADOPTED FROM [7])

	Africa	Americas	Asia Pacific	Europe	Middle East	South Asia
Conflicts and wars	2	2	9	17	3	2
Personalities and leaders	4	3	4	16	0	1
Disasters	3	4	4	10	0	1
Science&technology	0	3	2	5	0	2
Sport	0	0	1	4	1	0
Politics	4	0	4	2	0	0

```
for each video $V_t = [v_t^1, v_t^2, ..., v_t^N]$ in the database {
    p=1;
    while (p<N-M)    {
        calculate dissimilarity $D(V_q, V_t^p)$
        if $D(V_q, V_t^p) \prec \varepsilon$
            $(V_q, V_t^p)$ matches
        p=p+1;
    }
}
```

In our experiments, each frame is evenly divided into 4 regions, and this suggests that the proposed signature can manage different resolutions and screen display formats. Also due to the temporal attributes of the ordinal measure, this signature is naturally free from luminance or contrast shifts between different copies.

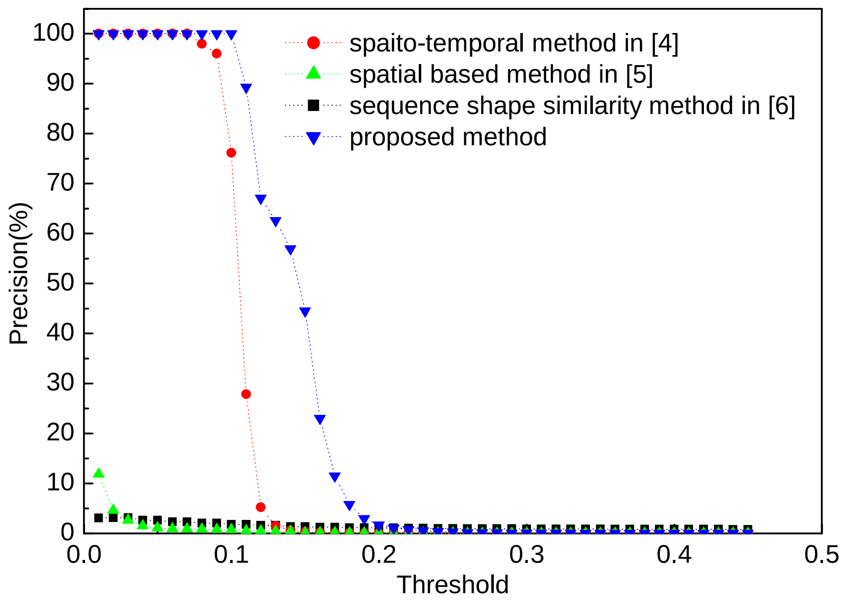
3.2 Precision and recall

Precision and recall [4] are used to evaluate the system defined as below:

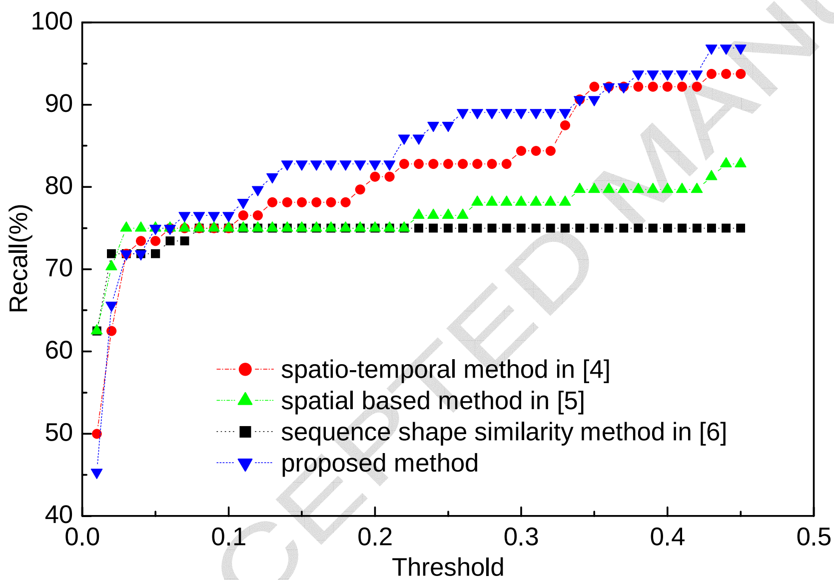
$$precision = \frac{number\ \ of\ \ correct\ \ dectection\ of\ \ matched\ \ videos}{number\ \ of\ \ matched\ \ videos} \tag{3}$$

$$recall = \frac{number\ \ of\ \ correct\ \ dectection\ \ of\ \ matched\ \ videos}{total\ \ number\ \ of\ \ copied\ \ videos} \tag{4}$$

It is worth noting that a correct detection of a matched video means not only identifying the target video, but also the precise temporal position. In order to compare other methods in the literature, the methods developed in [4, 5, 6] were also implemented. Precision and recall against thresholds are plotted in Fig. 4(a) and 4(b) respectively using query video lengths of 50 frames. It can be seen that the proposed video sequence matching obtains better performance than other three methods. The sequence shape similarity method [6] was applied to query videos of more than 20 seconds in length in their original experiment with high precision and recall up to 100%; however our results (see Fig. 4) show that this method achieves very low precision. This very possibly indicates that this method is not suitable for the situation with short length video as query; for example in the experiment, a query video is with 50 frames. The frame-wise similarity method in [5] also produced a poor performance. As [4] argued that the partitioning scheme becomes more critical for different formats of copies, the asymmetrical changes introduced in query video leads to poor precision of the method [5].



(a)



(b)

Fig. 4. Precision and recall against threshold at length = 50 (a) precision against threshold (b) recall against threshold

Indeed Hampapur, Bolle and Hyun. [5] observed that a spatial signature based on ordinal ranking is sensitive to different lengths of query videos, i.e., the matching performance improves with increasing video lengths. The same effect of sensibility to query length happens in the method [6] (see Fig. 4). Thus we introduce different lengths of query videos to further test the proposed method compared to the competing ones considering sensibility to different video lengths. Fig. 5 plots the precision-recall curves with query length changes (50 frames, 100 frames and 150 frames). It demonstrates that the proposed method achieves a better performance and also is less sensitive to query length changes. Our recent work on a comparative study of video copy detection

[16] further demonstrates that with the introduction of more transforms such as zoon in/out, crop and insert logos, the competing methods are sensitive to these "strong" distortion.

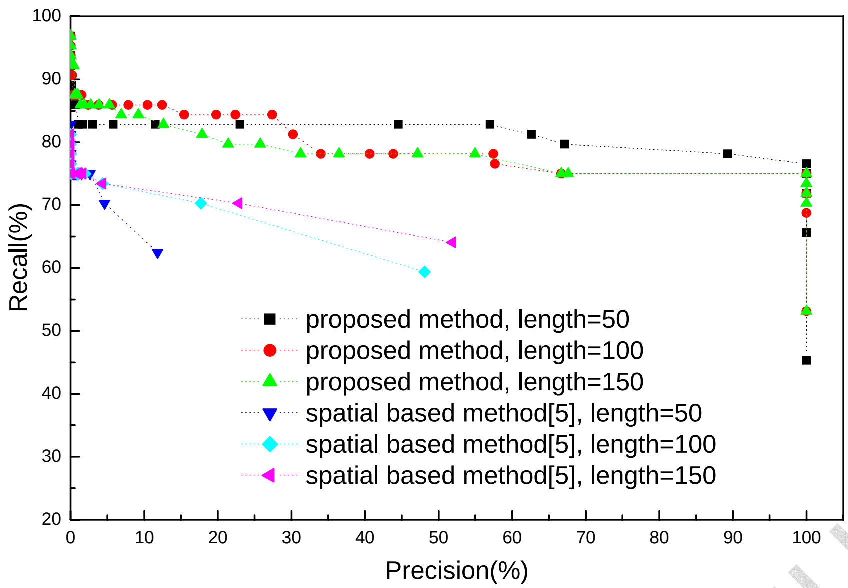


Fig. 5. Precision and recall with different lengths of query videos

4 Estimation of Robustness and Discriminability

If a matching method is robust, the dissimilarity score between original (V) and forged (V') copy should be small enough, i.e., $D(V,V') \to 0$ while sufficient discrimination is obtained between different videos. To given an estimation of robustness, the numbers of R videos are randomly sampled from a database and 7 different transforms applied to these videos (see Fig. 3). Each pair of original and forged copies is considered, so for each group of original and 7 transformed copies, there are 28 ($={}_8C_2$) dissimilarity values and in total there are 28R pairs. According to Function (1), the following probability distribution of dissimilar scores between copied videos is calculated as,

$$P(D(V,V') \le \varepsilon) \tag{5}$$

On the other hand, an effective method also requires discriminability, i.e.,

$$D(V_{M1}, V_{M2}) \to 1$$

where $|V_{M1}| = |V_{M2}|$, and V_{M1} and V_{M2} are irrelevant videos. The probability distribution of dissimilar scores between different videos (V_{M1}, V_{M2}) is calculated as:

$$P(D(V_{M1},V_{M2}) \geq \varepsilon) \quad (6)$$

Assuming the length of a signature is *L*, the number of possible pairs of (V_{M1},V_{M2}) is $L!$. 64 random query videos (8 original and 56 transformed copies) each with 150 frames are used to test robustness and discriminability. Thus 224 (28x8) pairs of copied videos and 1792 ($={}_{64}C_2 - 224$) pairs of different videos are applied to Functions (5) and (6) respectively, and Fig. 6 plots the corresponding probability distributions. A robust method should produce higher values of Function (5) with smaller ε; a discriminating method should produce higher values of Function (6) with higher values of ε. Fig. 6 shows that the proposed method has better robustness and discriminability compared to the methods in [4] and [5].

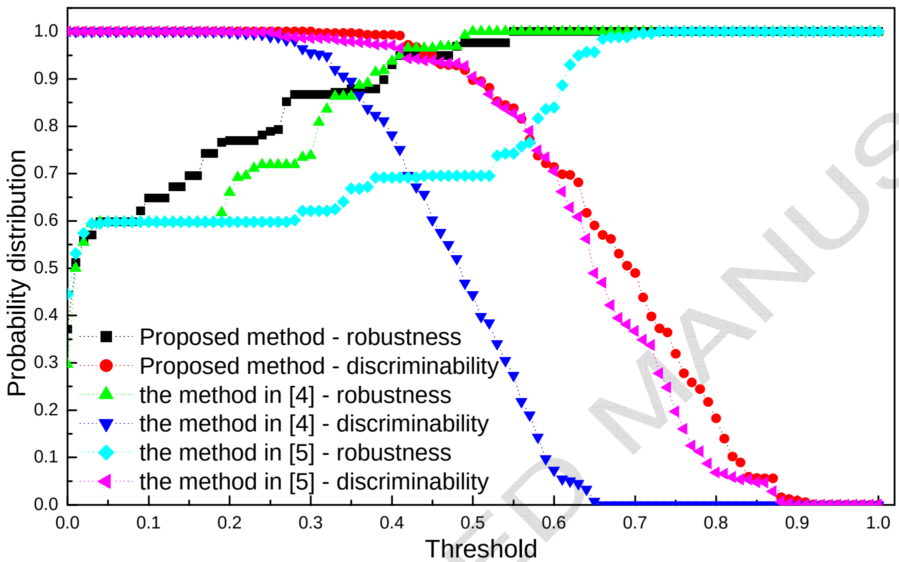


Fig. 6. Probability distributions of dissimilar scores

5 Discussion

Using the proposed approach, a video can be expressed compactly. For a video $V = [v^1, v^2, ..., v^N]$ in the database where each frame is partitioned into K regions, the ranking of each frame is expressed using $K \log_2 N$ bits. The proposed scheme is efficient in video sequence matching: matching a 50 frame query video to the database (285535 frames) takes about 9 seconds in a single thread on a Pentium 4 computer with 2.0GHZ CPU. Furthermore, multiple thread programming architectures or parallel calculations will dramatically reduce retrieval times.

Another aspect of research in CBCD is the difficulty of comparing methods in the absence of a benchmark. The BBC open news archive provides an open video database covering comparatively diverse topics, and offers an excellent benchmark as a reference in the video copy detection field.

To test the relative performance of frame-wise signatures in video sequence matching, the spatial signatures as described in [4, 5] were integrated with the temporal ordinal measure, where each contributes equally to the video sequence matching. More formally, let the [1xK] ordinal ranking matrices of a query video $V_q = [v_q^1, v_q^2, ..., v_q^M]$ be

$$\pi_q = [\pi_{q,1}, \pi_{q,2}, ..., \pi_{q,M}],$$

where $\pi_{q,i} =< \pi_{q,i}^1, \pi_{q,i}^2, ..., \pi_{q,i}^K >$ is a spatial ordinal ranking of regions in a frame v_q^i. Similarly, a spatial signature of a frame in target sub-video $V_t^p = [v_t^p, v_t^{p+1}, ..., v_t^{p+M-1}]$ is written as $\pi_{t,p+i}$ $i = 0, 1, ..., M-1$.

The spatial distance between two frames is defined as [4]:

$$d(\pi_{q,i}, \pi_{t,p+i}) = \frac{1}{C_K} \sum_{k=1}^{K} \left| \pi_{q,i}^k - \pi_{t,p+i}^k \right| \tag{7}$$

where $d(\pi_{q,i}, \pi_{t,p+i})$ is the normalised distance between two ordinal ranking matrices and C_K is the maximum distance between two ranking matrices $\pi_{q,i}$ and $\pi_{t,p+i}$. C_K is obtained when the two permutations of ordinal ranking matrices are the reverse of each other. For example, if K = 4, then C_K = 8.

The spatial dissimilarity between two sequences is computed by averaging over M pairs of frame-wise dissimilarities in Function (7), i.e.,

$$D_s(V_q, V_t^p) = \frac{1}{M} \sum_{m=1}^{M} d(\pi_{q,m}, \pi_{t,p+m-1})) \tag{8}$$

Combining the proposed temporal ordinal measure in Function (1), the final dissimilarity score is calculated as

$$D_{ST}(V_q, V_t^p) = \alpha D_s(V_q, V_t^p) + (1-\alpha) D\ (V_q, V_t^p) \tag{9}$$

As in [4] α is set as 0.5 to balance the contribution of the spatial and temporal ordinal measures. Fig. 7 illustrates the performance of the temporal ordinal measure based matching and the one with spatial signature added. The extra spatial information did not improve the performance but in fact reduced it a little. This indicates that the spatial data does not add useful information and might indicate that it is not independent of the temporal data as it is gathered in this approach.

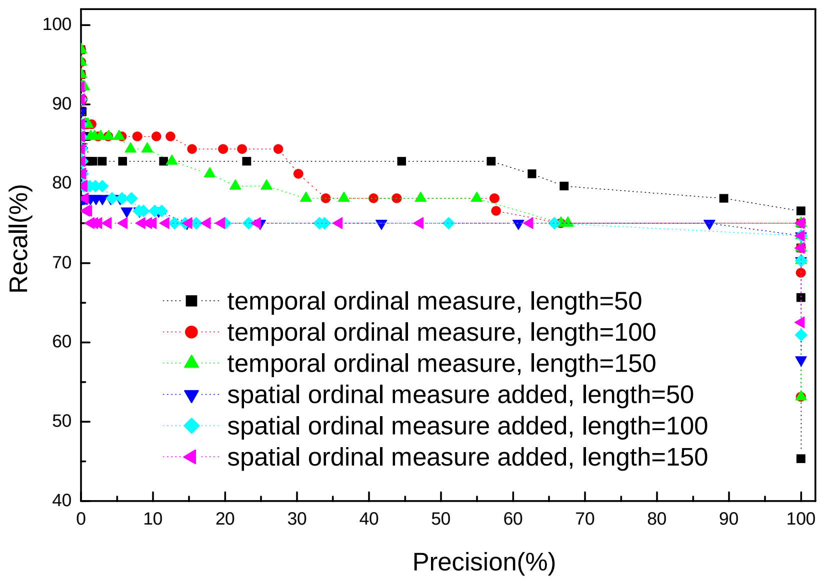


Fig. 7. Performance comparison between the proposed method and the method with spatial signatures added

6 Conclusions

This paper has presented a new video sequence matching method based on a temporal ordinal measure to obtain a promising solution to video copy detection. Such a temporal ordinal measure is different from the previous methods which simply assigned -1/0/+1 to record the variation in two successive frames without considering a global order of temporal variety. Experiments were conducted on the BBC open news archives which include 79 videos (285535 frames, about 3.1 hours) covering different topics including conflicts and wars, disasters, personalities and leaders, politics, science & technology, and sports. Three recent proposed methods [4, 5, 6] by different research institutes and companies were compared to the proposed method. The proposed method yielded the best performance, followed by the spatio-temporal signature matching [4], the frame-wise ordinal measure matching method [5] and the sequence shape similarity matching based on an ordinal measure [6]. Also the proposed method was less affected by the query clip length.

Apart from the classical evaluation criteria of precision and recall, a statistical method has been introduced to give a quantitative evaluation of robustness and discriminability, which are two important aspects to judge sequence matching methods. Again, the proposed method performed with greater robustness and discriminability compared to the other methods.

Acknowledgment

This research has been conducted within the framework of the European Commission funded Network of Excellence “Multimedia Understanding through Semantics, Computation and Learning” (MUSCLE) [10].

References

[1] J. Law-To, V. Gouet-Brunet, O. Buisson and N. Boujemaa, "Local behaviour labelling for content-based video copy detection," in *Proc. of the 18th International Conference on Pattern Recognition* (ICPR'06), August 2006, pp. 232-235.

[2] X. F. Yang, Q. Sun, and Q. Tian, "Content-based video identification: a survey," in *Proc. of Int. Conf. Information Technology: Research and Education*, 11-13 Aug. 2003 Page(s):50-54.

[3] R. Mohan, "Video sequence matching," in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, ICASSP 1998, Volume 6, 12-15 May 1998 Page(s):3697-3700.

[4] C. Kim and B. Vasudev, "Spatiotemporal sequence matching for efficient video copy detection," *IEEE Trans on Circuits and Systems for Video Technology*, 15(1), Jan. 2005, pp. 127-132.

[5] A. Hampapur, Rudolf M. Bolle and Ki-Ho Hyun, "Comparison of sequence matching techniques for video copy detection," *Proc of SPIE: Storage and Retrieval for Media Databases*, 2001.

[6] X-S Hua, X. Chen and H-J Zhang, "Robust video signature based on ordinal measure," in *Proc. of IEEE Int. Conf. on Image Processing*, ICIP 2004, vol. 1, 24-27 Oct. 2004, pp. 685-688.

[7] http://www.bbc.co.uk/calc/news/

[8] S. C. Cheung and A. Zakhor, "Estimation of web video multiplicity," in *Proc. of the SPIE - Internet Imaging*, San Jose, California. January 22-28, 2000, vol. 3964, pp. 34-46.

[9] J. Oostveen, T. Kalker and J. Haitsma, "Feature extraction and a database strategy for video fingerprinting source," in *Proc. of the 5th International Conference on Recent Advances in Visual Information Systems*, London, UK. 2002. Lecture Notes In Computer Science; Vol. 2314, ISBN:3-540-43358-9, Springer-Verlag, pp. 117-128.

[10] Multimedia Understanding through Semantics, Computation and Learning, EC 6th Framework Programme. FP6-507752. (2005) http://www.muscle-noe.org/

[11] S. H. Kim and R-H Park, "An efficient algorithm for video sequence matching using the modified Hausdorff distance and the directed divergence," *IEEE Transactions on Circuits and Systems for Video Technology*, 12(7), July 2002, pp. 592 – 596.

[12] S. C. Cheung and A. Zakhor, "Efficient video similarity measurement and search," in *Proc. of Int. Conf. on Image Processing*, ICIP 2000. 10-13 Sept. 2000, vol. 1, pp. 85-88.

[13] D. N Bhat and S. K. Nayar, "Ordinal measures for image correspondence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4), April 1998, pp. 415-423.

[14] Y. J. Li, J. S. Jin, and X. F. Zhou, "Video matching using binary signature," in *Proc. of Int. Symposium on Intelligent Signal Processing and Communication Systems, ISPACS 2005*, 13-16 Dec. 2005, pp. 317-320.

[15] A. Hampapur and R. M. Bolle, “Comparison of distance measures for video copy detection,” in *Proc. of Int. Conf. on Multimedia and Expo,* ICME 2001, 22-25 Aug. 2001, pp. 737-740.

[16] Julien Law-To, Li Chen, A. Joly et al. Video Copy Detection: a Comparative Study, *ACM International Conference on Image and Video Retrieval (CIVR'07)*, July 9-11 2007, Amsterdam, the Netherlands, pp. 371-378.

Vitae

L. Chen received her B. E degree from Xi'an Jiao Tong University, China, in 1998, M. E degree from University of Science and Technology of China in 2001, Ph.D. degree from Surrey University, UK in 2005, respectively. This work was done when she was with University College London, Adastral park campus, UK. Her research interests include machine learning, image analysis and retrieval, video analysis and retrieval.

F. W. M. Stentiford won a scholarship to study mathematics at St Catharine's College, Cambridge, and obtained a Ph.D. in Pattern Recognition at Southampton University. He first joined the Plessey Company to work on various applications including the recognition of fingerprints and patterns in time varying magnetic fields. He then joined British Telecom and carried out research on optical character recognition and speech recognition. During this time he led a team developing systems employing pattern recognition methods for the machine translation of text and speech. This work led to the world's first demonstration of automatic translation of speech between different languages. He is now a professor of Telecommunications in the Electronic & Electrical Engineering department of University College London where he leads a team researching multimedia content understanding. He has published over 70 papers and filed over 20 patents. He is a member of the British Computer Society and the Institute of Engineering and Technology. He is a member of the IEEE Multimedia Signal Processing technical committee.