

Content -Built Visual Copy Discovery in Huge Databanks: A Native Characteristics Statistical Resemblance Quest Method

A. Samayadevi^{1*} and A. Aarthi²

^{1*}Research Scholar, PREC, Anna University, India.

²Department of Communication System, PREC, Anna University, India.

www.ijcaonline.org

Received: Oct/26/2014

Revised: Nov/09/2014

Accepted: Nov/20/2014

Published: Nov/30/2014

Abstract— Recent methods based on interest points and local fingerprints have been proposed to perform robust CBVCD (content-built visual copy discovery) of images and video. They include two steps: the search for similar local fingerprints in the database (DB) and a voting strategy that merges all the local results in order to perform a global decision. In most image or video retrieval systems, the search for similar features in the DB is performed by a geometrical query in a multidimensional index structure. Recently, the paradigm of approximate k-nearest neighbors query has shown that trading quality for time can be widely profitable in that context. In this paper, we evaluate a new approximate search paradigm, called Statistical Similarity Search (S3) in a complete CBVCD scheme based on video local fingerprints. Experimental results show that these statistical queries allow high performance gains compared to classical -range queries and that trading quality for time during the search does not degrade seriously the global robustness of the system, even with very large DBs including more than 20,000 hours of video.

Keywords—Component; Formatting; Style; Styling; Insert

I. INTRODUCTION

Content-built visual copy discovery (CBVCD) schemes are an alternative to the watermarking approach for persistent identification of images and video clips [1, 2, 3, and 4]. It generally consists in extracting as few features as possible from the candidate objects and matching them with a DB. Since the features must be discriminant enough to identify an image or a video, the features are often called fingerprints or signatures [4]. Recently, robust CBVCD schemes based on local fingerprints have been proposed to deal with geometrical transformations, such as resizing, shifting or inserting [1, 3]. In these techniques, the detection includes two steps: the search for similar local fingerprints in the DB and a voting strategy that merges all the local results in order to decide which of these results some copies of the candidate object are. The method proposed in [1] is dedicated to static images and the voting strategy is only based on the image identifiers of the local fingerprints returned by the search. In [3], we proposed a method dedicated to video. The experimental results we present in this paper were obtained in that context.

As for many content based retrieval systems, one of the difficult task of a CBVCD scheme is the time cost of the similarity search in the fingerprints reference DB, which can be very large. To solve this problem, most multimedia retrieval systems use multidimensional index structures [5, 6] or improved sequential techniques, e.g. the V A-file [7]. For the last few years, researchers are interested in trading quality for time [8, 9, 10, and 11] and the paradigm of approximate similarity search has emerged [12]. Some of the

proposed solutions are simply early stopping approaches [9] and other techniques, e.g. [13], are based on geometrical approximations during the filtering rules of the search algorithm. More recent techniques are based on a probabilistic selection of the bounding regions used in the indexing structure [10, 11]. They allow to control directly the expected percentage of the real k-nearest neighbors of the query. Range queries are rarely used because the results of the similarity search are almost always directly linked to the results that are provided to the user, for whom it is useful to get always the same number of results. We think however that a k-nearest neighbor search is not appropriate to copy detection and especially for techniques that include a voting strategy after the search. The main reason is that the number of relevant fingerprints for a given query is highly variable. In a large TV archives DB, several video clips can be duplicated 600 times, whereas other video clips are unique. Furthermore, inside a single video sequence, points of the background are detected many times whereas others corresponding to moving objects are unique. The basic idea of the Statistical Similarity Search (S3) technique we proposed in [14] was to extend the approximate search paradigm to -range queries, leading to the statistical query paradigm (section 2). In this paper, we evaluate the S3 technique in a complete video CBVCD scheme, described in section 3. The evaluation description and the results are provided in section 4.

II. STATISTICAL QUERY

By excluding several regions of a hyper spherical query, having a too small intersection with the bounding regions of

Corresponding Author: A. Samayadevi

the index structure, it is possible to obtain high speed-up with very small losses in the results. However, it is not possible to take the volume percentage as an error measure because it would be equivalent to consider that the relevant similar fingerprints are uniformly distributed inside a hyper sphere. When the dimension increases, the fingerprints following such a distribution become closer and closer to the surface of the hyper sphere but this is not true in reality, as illustrated on Fig. 1. The solid curve (left) is the real distribution of the distance between referenced and distorted fingerprints issued from a transformed version of the referenced video sequences for the same interest points. In this example, the transformation was a resize of factor was = 0.8, but we observed the same kind of distributions for all studied transformations, including colorimetric distortions and noise addition. The two other dotted curves represent the estimated probability density function for two probabilistic models: a uniform spherical distribution (right) that would be obtained if we took the volume percentage as an error measure and a zero mean normal distribution (center) under components independence assumption. A simple independent normal distribution is much closer to the real distribution than the uniform one. The proposed statistical query paradigm relies on the distribution of the relevant similar fingerprints. Let the distortion vector ΔS be defined by:

$$\Delta S = S(m) - S(t(m)), \quad t \in T$$

Where $S(m)$ is the fingerprint of a referenced pattern m , $S(t(m))$ is the distorted fingerprint, i.e. the fingerprint of the transformed pattern $t(m)$ and T is the set of transformations that can be applied between a referenced sequence and a copy. The statistical query of expectation α is defined as the search of all the fingerprints contained in a region V_α of the feature space satisfying:

$$\int_{V_\alpha} p_{\Delta S}(X - Q) dX \geq \alpha \quad (1)$$

Where Q is a candidate fingerprint and $p_{\Delta S}(\cdot)$ is the probability density function of the distortion.

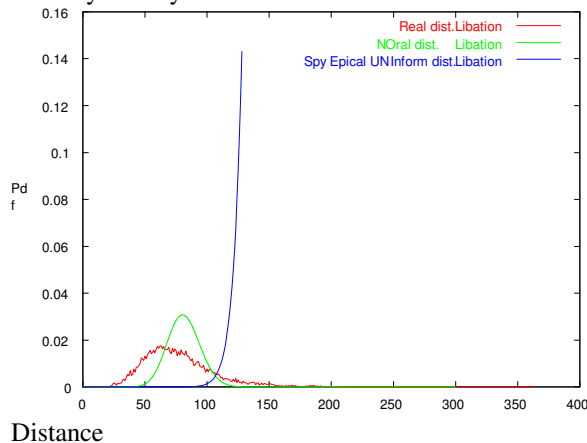


Fig. 1. Distribution of the distance between a fingerprint and its distorted version after resizing of a video sequence ($wsc = 0.8$)

In practice, the first step of a search in a multidimensional indexing structure is a set of geometric filtering rules that quickly exclude most of the bounding regions. To process the statistical queries in an indexing structure, we propose to replace the geometric rules by probabilistic rules, according to the distortion model. The main advantage is that a statistical query has no intrinsic shape constraint. Thus, the region V_α that makes equal the probability to find a relevant fingerprint to α is naturally adapted to the shape of the bounding regions used by an indexing structure. Details on the indexing structure and the search algorithms of the S3 technique are provided in [14]. The distortion model used in the rest of this paper is a zero mean normal distribution independent for each component:

$$p_{\Delta S_j}(x_j) = fN(0, \sigma)(x_j)$$

where the unique parameter σ is estimated on a set of representative transformations.

III. VIDEO LOCAL FINGER PRINTS AND VOTING

STRATEGY

- The local fingerprints extraction [3] includes three steps:
- A key-frame detection, based on the mean of the frames difference also called intensity of motion.
- An interest point detection in each key-frame, processed by an improved version of the Harris detector [15].
- a local characterization computed around each interest point, leading to a $D = 20$ -dimensional fingerprint S :

$$S = \left(\frac{s_1}{\|s_1\|}, \frac{s_2}{\|s_2\|}, \frac{s_3}{\|s_3\|}, \frac{s_4}{\|s_4\|} \right)$$

Where the six are 5-dimensional sub-fingerprints computed at four different patio-temporal positions distributed around the interest point. Each six is a differential decomposition of the gray level 2D signal $I(q, d)$ until the second order:

$$s_i = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, \frac{\partial I}{\partial x \partial y}, \frac{\partial^2 I}{\partial x^2}, \frac{\partial^2 I}{\partial y^2} \right)$$

In the indexing case, the fingerprints are simply inserted in the indexing structure with a video sequence identifier Id and a time-code tic . In the detection case, the statistical query returns for each candidate fingerprint so a set of K_i referenced fingerprints $\{S_{uk}\} k \in K_j$ with their identifiers $\{Id_{jk}\} k \in K_j$ and their time-codes $\{tic_{jk}\} k \in K_j$. These results are stored in a buffer for a fixed number of key-frames in order to estimate the best sequences. We note NC and the number of candidate fingerprints contained in the corresponding time interval ($j \in [1, NC]$ and).

The estimation is only based on the identifiers and the time-codes and not on the fingerprint itself. For each identifier id , the corresponding time-codes are used to

estimate the unique parameter b of the following temporal model:

$$tc' = tc + b$$

Where T_c represents a time-code of the candidate sequence and T_r a time-code of a referenced sequence. This simple estimation problem is solved by the following minimization equation:

$$\hat{b}(id) = \arg \min_b \left(\sum_{j=1}^{N_{candidate}} \min_{k \in \mathcal{K}_j} \rho(|tc'_j - (tc_{jk} + b)|) \right) \quad (2)$$

Where $\rho(u)$ is a non-decreasing cost function allowing to decrease the contribution of outliers. The Turkey's weight M-estimator was chosen for $\rho(u)$ (see [16] for details). Once $b(id)$ has been estimated, a similarity measure $snip$ is computed for each id represented in the results by a voting strategy. It simply consists in counting the number of candidate fingerprints (i.e. the number of interest points) that contribute to the solution $b(id)$ according to a small tolerance interval. By thresholding the value of $snip$, we finally decide which of the identifiers represented in the results correspond effectively to a copy of the candidate sequence.

IV. EXPERIEMENTS AND RESULTS

The DBs used in these experiments contain real video local fingerprints extracted by the method described in section 3. The referenced video sequences come from a French TV archives DB that contains all kinds of TV programs from the Forties at our days: news, sport, show, variety, films, reports, black white archives, advertisements, etc. The average number of local fingerprints per hour of video is about 50,000. Thus, a DB representing 10,000 hours of video contains about 500,000,000 fingerprints and the size of the corresponding DB file is about 13 Gb ($D = 20$ dimensional fingerprints + identifiers + time codes). Experiments were computed on a Pentium IV (CPU 2.5 GHz, cache size 512Kb, RAM 1.5 GB).

4.1. Statistical query compared to exact range query

In this first experiment, we compare the search time of a statistical query to those of a classical spherical range query of radius α (-range query). We do not aim at testing the relevance of the distortion model, but at showing the advantage of a statistical query compared to an exact range query when the distribution is perfectly known.

We randomly select 1000 real fingerprints S in the DB and construct 1000 queries $Q = S + \Delta S$, where the components of the distortion vector ΔS are independently generated according a zero mean normal distribution $p_{\Delta S_j}(ax) = \text{fan}(0, \sigma)(ax)$ with $\sigma = 18.0$. These queries are then searched in the DB using both a statistical query and an α -range query. For different values of the query expectation α , we measure, for both query types, the average time of a single search (Fig. 2) and the retrieval rate (Fig. 3), i.e. the percentage of queries for which the original fingerprint S

belongs to the results returned by the search. The radius of the range query was set in order to have the same expectation α than the statistical query. It is indeed easy to show that for the given distortion model, the L2 norm of the distortion has the following probability density function:

$$p_{\|\Delta S\|}(r) = \frac{f_{\mathcal{N}(0, \sigma)}(r)}{(2\pi\sigma)^{\frac{D-1}{2}} \Gamma\left(\frac{D}{2} + 1\right)} r^{D-1}$$

Where Γ is the gamma function and D is the dimension of the feature space. By tabulating the values of the corresponding cumulated density function, it is easy to choose the value of the radius r such as

$$\int_0^r p_{\|\Delta S\|}(r) dr = \alpha$$

The average search time curves of Fig. 2 (displayed in logarithmic coordinates) show that the statistical query approach outperforms the classical exact range query. Depending on α , it is from 17 to 132 times faster. The geometrical constraint of an exact α -range query degrades seriously the search time without improving the retrieval rate, as shown on Fig. 3. This result does not depend on the index structure. Whatever the shapes of the bounding regions are, it is indeed well known that the number of intersections with a hyper sphere becomes very high when the dimension increases. The main asset of the statistical query is that it does not impose any particular shape. It only uses the probability to find a relevant fingerprint inside the bounding regions.

4.2. Robustness of the video CBVCD system

The purpose of these experiments is to study the interaction between the S3 technique and the global robustness of the video CBVCD system. We extract randomly 100 video sequences of 10 seconds each from the reference DBs and apply to them five kinds of transformations: vertical shift of $wsh\%$ of the image, resize of factor wsc and gamma modification ($I'(x, y) = I(x, y)^{w\gamma}$).

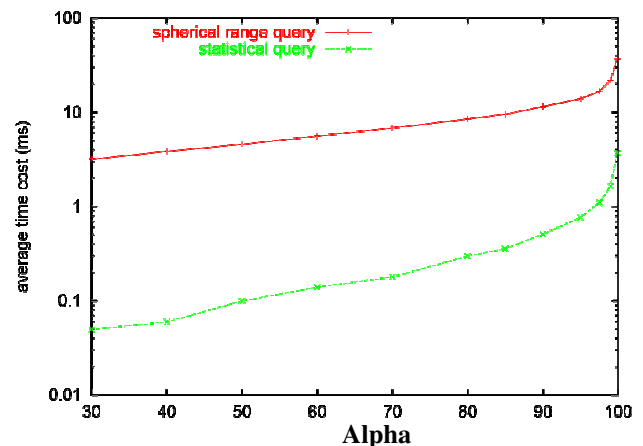


Fig. 2. Average search time (MS) vs. query expectation α

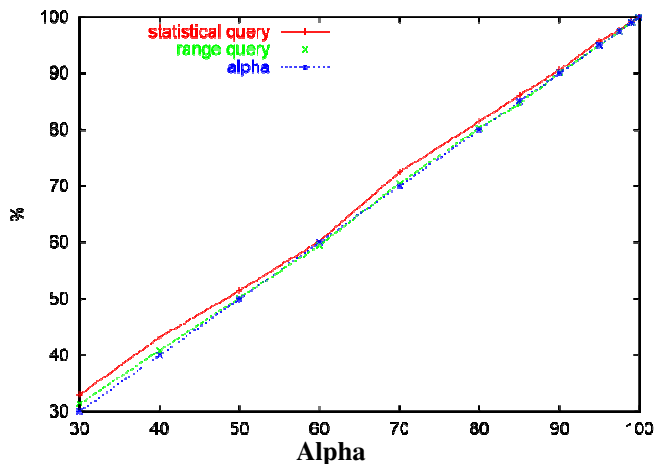


Fig. 3. Retrieval rate (%) vs. query expectation α

The 100 transformed sequences are submitted as candidates to the video CBVCD system and we measure the good detection rate with a tolerance of 2 frames. The threshold value of nsim was set so that in average less than 1 false alarm occurs per hour when the system is continuously monitoring a TV channel. The parameter σ of the distortion model was set to 20.0. Different DB sizes, different values of the expectation α and different values of transformations parameters are used in these experiments. Results are presented on Fig. 4 in the form of abacuses of

- α (left-hand side) using a DB containing about 3500 hours of video
- the DB size (right-hand side) with α set to 80%

Two tables presenting the average search time of one single fingerprint for the different values of α and DB size are given at the bottom of the figure. Right subplots show that the DB size does not affect so much the detection rate whatever the transformation is. The main reason is that the statistical query guarantees the same expectation for the similarity search whatever the DB size is. The increased number of false retrieved fingerprints does not degrade the final quality thanks to the voting strategy which is highly discriminant (the temporal coherence of many fingerprints is very rare).

It is important to note that it would not have not been the same if we had used a k-nearest neighbor search. When the DB size is multiplied by 100, the higher the density of fingerprints, the higher the chance to exclude relevant fingerprints from the results. Left subplots show that the detection rate remains almost invariant for all transformations as the expectation α decreases 95% down to 70% whereas the search is 4 times faster. For the most severe transformations, it begins to fall down when α equals 50%. The most important result of this experiment is that an approximate search is particularly profitable when a voting strategy is employed after the search. It is indeed useless to retrieve the less distortion-invariant fingerprints since they

seriously degrade the search time without really improving the robustness of the video CBVCD system.

video hours	time cost (MS)
110	0.69
875	1.81
3,500	3.45
10,000	9.11

A (%)	time cost (MS)
50	1.05
70	2.15
80	3.45
90	5.76
95	8.64

VII. CONCLUSION

In this paper, we evaluated a new approximate search paradigm based on statistical queries. We showed that the geometrical constraint of a classical exact -range query can degrade seriously the time cost of the search without systematically improve the results. We showed that the use of a statistical filtering instead of an exact geometrical filtering can lead to a search about 100 times faster. We then studied the influence of our statistical similarity search strategy on the global robustness of a video CBVCD system which is based on local fingerprints retrieval and a voting strategy. The experiments showed that in such a scheme, trading quality for time during the search is highly profitable, even when the size of the DB becomes very large.

The voting strategy will be our main future research topic for two reasons. Firstly, when the DBs becomes very large, the number of retrieved fingerprints during the search increases seriously and this step will probably become a new bottleneck of the total time cost of the CBVCD system. Secondly, we would like to extend the estimation step on the time codes to the spatial positions of the interest points to improve the discriminance. This will however complicate the simple parameter estimation problem of the S3 method and will require much more efficient algorithms.

REFERENCES

- [1] Linfang Dong ; Dept. of Comput. Sci. & Technol., Tianjin Univ. of Finance & Econ., Tianjin, China ; Shang Liu , "Adaptive Video Transmission Control Scheme in Wireless Networks", Published in: Distributed Computing and Applications to Business Engineering and Science (DCABES), 2010 Ninth

- International Symposium on Date of Conference: 10-12 Aug. 2010, Page(s): 557 – 560.
- [2] Kyeong-Jin Oh ; Dept. of Inf. Eng., Inha Univ., Incheon, South Korea ; Jin-Guk Jung ; Geun-Sik Jo, "A Context-Awareness for Mechanical Maintenance" Published in: Information Science and Applications (ICISA), 2011 International Conference on Page(s): 1 - 5
- [3] Feng Xue ; VCC Div., Hefei Univ. of Technol., Hefei, China ; Zengwei Jiang, "An improved mean shift algorithm for object tracking" Published in: Multimedia Technology (ICMT), 2011 International Conference on, Page(s): 4833 – 4836
- [4] Tankiz, S. ; Bilgisayar Muhendisligi Bolumu, Hacettepe Univ., Ankara, Turkey ; Ates, T.K. ; Can, A.B. "Content Based Video Copy Detection by shot segmentation" Published in: Signal Processing and Communications Applications (SIU), 2011 IEEE 19th Conference on Date of Conference:20-22 April 2011 Page(s):582 – 585
- [5] Xiao Bing Kang ; Key Lab. of Contemporary Design & Integrated Manuf. Technol., Northwestern Polytech. Univ., Xi'an, China ; Sheng Min Wei, "An efficient approach to still image copy detection based on SVD and block partition for digital forensics" Published in: Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on (Volume:3) Date of Conference:20-22 Nov. 2009 Page(s):457 – 461
- [6] Xiaorong Yang ; Coll. of Stat. & Math., Zhejiang Gongshang Univ., Hangzhou, China ; Ke-Ang Fu "Copy number detection using self-weighted least square regression" Published in: Systems Biology (ISB), 2011 IEEE International Conference on Date of Conference:2-4 Sept. 2011 Page(s):47 – 51
- [7] Ali Qureshi, M. ; Dept. of Electr. Eng., King Fahd Univ. of Pet. & Miner., Dhahran, Saudi Arabia ; Deriche, M." A review on copy move image forgery detection techniques" Published in: Multi-Conference on Systems, Signals & Devices (SSD), 2014 11th International, Date of Conference:11-14 Feb. 2014 Page(s):1 – 5
- [8] Muhammad, G. ; Coll. of Comput. & Inf. Sci., King Saud Univ., Riyadh, Saudi Arabia ; Hussain, M. ; Khawaji, K. ; Bebis, G., "Blind copy move image forgery detection using dyadic undecimated wavelet transform" Published in: Digital Signal Processing (DSP), 2011 17th International Conference on Date of Conference:6-8 July 2011 Page(s):1 – 6
- [9] Esmaeili, M.M. ; Dept. of Electr. & Comput. Eng., Univ. of British Columbia, Vancouver, BC, Canada ; Fatourech, M. ; Ward, R.K., "A Robust and Fast Video Copy Detection System Using Content-Based Fingerprinting" Published in: Information Forensics and Security, IEEE Transactions on (Volume:6 , Issue: 1) Date of Publication :20 December 2010, Page(s):213 – 226
- [10] Gao Jin-hua ; Shenzhen Inst. of Inf. Technol., Shenzhen, China ; Du Yi-hua ; Zhu Ying-ying ; Wen Zhen-kun, "Structure Information and Temporal Ordinal Measure Fused Video Copy Detection" Published in: Multimedia Information Networking and Security (MINES), 2011 Third International Conference on, Date of Conference:4-6 Nov. 2011 Page(s):51 – 54
- [11] El Adel, A. ; REGIM: Res. Group on Intell. Machines, Univ. of Sfax, Sfax, Tunisia ; Zaied, M. ; Ben Amar, C. "Learning wavelet networks based on Multiresolution analysis: Application to images copy detection" Published in: Communications, Computing and Control Applications (CCCA), 2011 International Conference on Date of Conference: 3-5 March 2011 Page(s):1 – 6
- [12] Barrios, J.M. ; Dept. of Comput. Sci., Univ. of Chile, Santiago, Chile ; Bustos, B." P-VCD: A pivot-based approach for Content-Based Video Copy Detection" Published in: Multimedia and Expo (ICME), 2011 IEEE International Conference on Date of Conference: 11-15 July 2011 Page(s): 1 – 6
- [13] Hung-Yi Lin ; Dept. of Logistics Eng. & Manage., Nat. Taichung Inst. of Technol., Taichung, Taiwan ; Shih-Ying Chen, "Indexing and Querying in Multimedia Databases" Published in: Intelligent Information Hiding and Multimedia Signal Processing, 2009. IHH-MSP '09. Fifth International Conference on Date of Conference: 12-14 Sept. 2009 Page(s): 475 – 478
- [14] Xiaorui Zhang ; Sch. of Manage., Changchun Inst. of Technol., Changchun, "A Linear Cost Function Model and its Application" Published in: Informatics in Control, Automation and Robotics, 2009. CAR '09. International Asia Conference on Date of Conference: 1-2 Feb. 2009 Page(s): 32 – 36
- [15] Bo Fu ; Coll. of Comput. & Inf. Technol., Liaoning Normal Univ., Dalian, China ; Chuan-Ming Song, "Object Enhancement and Recognition Based on Hough Forest for Underground Video" Published in: Parallel Architectures, Algorithms and Programming (PAAP), 2014 Sixth International Symposium on Date of Conference: 13-15 July 2014 Page(s): 75 - 80