PERFORMANCE ANALYSIS OF BAG-OF-FEATURES BASED CONTENT IDENTIFICATION SYSTEMS

S. Voloshynovskiy, M. Diephuis, T. Holotyak

University of Geneva, Stochastic Information Processing (SIP) Group
7, route de Drize, 1227 Geneva, Switzerland

ABSTRACT

Many state-of-the-art methods in image retrieval, classification and copy detection are based on the Bag-of-Features (BOF) framework. However, the performance of these systems is mostly experimentally evaluated and little results are reported on theoretical performance. In this paper, we present a statistical framework that makes it possible to analyse the performance of a simple BOF-system and to better understand the impact of different design elements such as the robustness of descriptors, the accuracy of encoding/assignment, information preserving pooling and finally decision making. The proposed framework can be also of interest for a security and privacy analysis of BOF systems.

Index Terms— Bag-of-features, content identification

1. INTRODUCTION

In recent years BOF based recognition, classification and retrieval systems have become the state-of-the-art in many applications ranging from multimedia management to security. In addition, in many applications facing strict memory-complexity restrictions, like in the on-line mobile phone visual or audio search systems, the BOF systems with carefully designed descriptors probably remain the only suitable technology in comparison to emerging yet complex deep learning frameworks [1, 2, 3, 4]. There are also more complex aggregation frameworks [3], but here we consider a simple yet tractable BOF system.

However, besides of its remarkable experimental performance, the theoretical analysis of BOF systems remains largely unexplored. They are often considered as black boxes where performance is estimated based on a public database, for some type of descriptors (e.g., SIFT [5], SURF [6], CHOG [1], ORB [7], BRIEF [8] , etc.) or a certain type of aggregation method [3], with little theoretical insight. In addition, it is not completely clear which descriptors in which particular application framework contribute to successful identification. Further more, it is not obvious which factors influence security and which elements of the BOF systems should be properly protected. Finally, the optimality of different encoding/assignment and pooling methods is based on empirical evidence rather than on strictly proven theoretical results.

This work was partially supported by the SNF project No. 200020-146379

2. STATE-OF-THE-ART BOF-BASED CONTENT IDENTIFICATION

Currently, most BOF systems are used for CBIR, object recognition and copy detection. We will consider *content identification* where M items are enrolled and given a probe, the system should determine the corresponding item or issue a rejection. When it is not possible to return a single index item, the system should retrieve a list of indices whilst ensuring that the true item index is on the list. The CBIR counterpart of content identification retrieves a list of indices of items similar to the probe.

The performance of BOF-systems is generally evaluated by simulation, and existing theoretical works [9, 10, 11, 12] mostly consider content identification based on content fingerprinting where a sufficiently long fingerprint is deduced to represent the content. In the most theoretical works, perfect synchronization between the enrolled fingerprint and the probe fingerprint is assumed with one notable exception [9] where fingerprint de-synchronization was modeled by a random shift parameter. However, in practice it is not feasible to design one single fingerprint or descriptor that would be invariant to all types of distortions, hence multiple local descriptors per image are used. Despite popularity, SIFT descriptors are characterized by the high computational complexity and relatively long length. In the recent years, a number of short binary descriptors have been proposed (BRIEF, ORB, etc) [8, 7]. However, in this case the length of the deployed descriptors does not satisfy the asymptotic assumptions considered in the theoretical works [9, 10, 11, 12] which makes the analysis of practical BOF-systems intractable.

This work has focused on the theoretic analysis of BOF based content identification systems. To our knowledge there is little work on the theoretical analysis of BOF-systems' performance besides [13] and none on BOF based content identification. Therefore, the goal of this paper is to provide a simple and tractable model allowing to analyze, optimize and guide the design of BOF systems. In this paper, we will con-

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

sider the case of non-compressed features to reveal the theoretical limits of BOF based identification systems, analyze the impact of descriptor compression and encoding/assignment as well as discovering the impact of geometrical consistency between the descriptors on overall system performance. Such a formulation was not considered in earlier studies.

The paper is organized as follows. The BOF based content identification problem is formulated in Section 3. Section 4 introduces the statistical model and Section 5 summarises the performance of the systems. Section 6 and Section 7 present the results and the conclusions.

3. BOF-BASED CONTENT IDENTIFICATION: PROBLEM STATEMENT

A content database is defined as a collection of M items $\mathbf{x}(m)$ represented by their features/descriptors $\mathbf{x}(m) = \{\mathbf{x}^1(m), \cdots, \mathbf{x}^{J_\mathrm{x}(m)}(m)\}$, $1 \leq m \leq M$, with each descriptor $\mathbf{x}^i(m) \in \mathcal{X}^L$, $1 \leq i \leq J_\mathrm{x}(m)$.

The problem is to decide if a query $\mathbf{y} = \{\mathbf{y}^1, \cdots, \mathbf{y}^{J_\mathrm{y}}\}$ is related to some elements or not. In the general case, $J_\mathrm{y} \neq J_\mathrm{x}(m)$.

The system should produce a list of indices $\mathcal{L}(\mathbf{y})$ whilst ensuring that the correct index m is always on this list and an empty set, if the probe $\mathbf{y}$ is not related to any item in the database.

The system performance is evaluated by the probability of missing a correct item and probability of falsely accepting an unrelated item $\mathbf{y}$ as related to some item in the database leading to the average list of accepted items. Other parameters include memory storage, search complexity, security and privacy [**?**, 12]. In this paper, we will focus on the performance of the identification system for a given database size M, parameters of descriptors, their numbers $J_\mathrm{x}(m)$ and J_y as well as targeting efficient search complexity based on inverted files.

The core idea behind the BOF systems consists in a local feature based representation of each image aggregated into a fixed dimensional vector.

Such a representation should ensure fast search of ϵ-NN or k-NN. It is also closely related to the approximations or compression of the descriptors in the visual words space which can be roughly classified in three groups:

- *VQ (hard)-assignment* [14]: the descriptor $\mathbf{x}^i$ is quantized by a coarse vector quantizer (VQ) $\mathbb{Q}_c(.)$ to its nearest visual word resulting in the approximate $\hat{\mathbf{x}}^i = \mathbb{Q}_c(\mathbf{x}^i)$;
- *source coding with refinement* [15]: the descriptor $\mathbf{x}^i$ is quantized using the VQ $\mathbb{Q}_c(.)$ to its nearest visual word with simultaneous storage of the quantized refinement information about the descriptor with respect to its quantized version given by fine quantization $\mathbb{Q}_f(.)$ of the residual vector: $\hat{\mathbf{x}}^i = \mathbb{Q}_c(\mathbf{x}^i) + \mathbb{Q}_f(\mathbf{x}^i - \mathbb{Q}_c(\mathbf{x}^i))$;

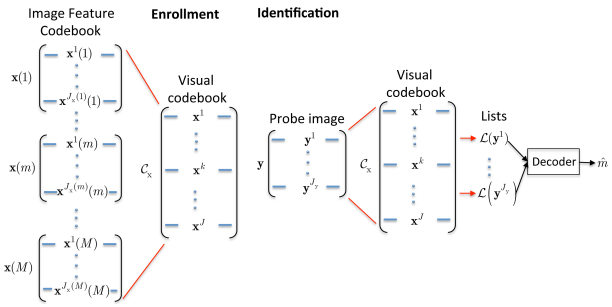

Fig. 1. Enrolment and identification via the visual codebook.

- *soft-assignment* [16, 17]: the descriptor $\mathbf{x}^i$ is approximated in the space of visual words $\mathbf{c}^j$ by the linear approximation: $\hat{\mathbf{x}}^i = \sum_{j=1}^{J} w_j \mathbf{c}^j$ with weights $\{w_j\}$. The different strategies such as sparse coding and LLC include different selection strategies of weight coefficients $\{w_j\}$.

The design of a visual codebook in terms of the best approximation of the descriptors is vital for the trade-off between memory storage, search complexity and accuracy. That is why the descriptors are stored in a compressed or approximated form $\mathbf{x}^i \rightarrow \hat{\mathbf{x}}^i$ with special indexing using hierarchical structures. However, to reveal the theoretical limits of the identification systems based on BOF, we will assume that the descriptors are uncompressed corresponding to a very fine approximation targeted by many state-of-the-art techniques [2].

For these reasons, we will consider the equivalent model shown in Figure 1 consisting of enrollment and identification via the equivalent codebook $\mathcal{C}_\mathrm{x} = (\mathbf{x}^1, \cdots, \mathbf{x}^J)^T \in \mathbb{R}^{J \times L}$. This equivalent codebook contains all unique descriptors, the composition of which gives the representation of a particular image $\mathbf{x}(m)$. Note that the representation of each image in terms of the visual codebook with an appropriate indexing structure in the form of the inverted files makes it possible to obtain an efficient search [15].

4. STATISTICAL MODEL OF BOF CONTENT IDENTIFICATION

The statistical model of BOF content identification includes the definition of: (a) statistics of descriptors $\mathbf{x}^i$, (b) statistical observation model $p(\mathbf{y}^k|\mathbf{x}^i)$, (c) model of encoding/assignment, (d) model of decision making about the descriptor presence/absence and geometric consistency verification [1], and finally (e) model of global decision making or decoding.

Database of descriptors In this paper, we will assume that the local sparse descriptors $\mathbf{x}^i \in \mathcal{X}^L$ are i.i.d. like ORB following some distribution $\mathbf{X}^i \sim p(\mathbf{x}^i) = \prod_{n=1}^{L} p(x_n^i)$ [2].

[1] We will also investigate the upper theoretic limit assuming the perfect synchronization.

[2] One can consider different descriptors: local or global, sparse or dense, multilevel and binary. The i.i.d. assumption is not valid for SIFT descriptors

Statistical observation model The statistical observation model for the entire image is expressed in terms of statistical model for the local descriptors:

$$p(\mathbf{y}|\mathbf{x}(m)) \equiv \prod_{k=1}^{J_\mathrm{y}} \prod_{i=1}^{J_\mathrm{x}(m)} p(\mathbf{y}^k|\mathbf{x}^i(m)), \quad (1)$$

which reduces to $p(\mathbf{y}|\mathbf{x}(m)) \equiv \prod_{k=1}^{J_e} p(\mathbf{y}^k|\mathbf{x}^k(m))$ in the *synchronized* case, i.e., the exact correspondence between the descriptors is known with $J_e = \min\{J_\mathrm{x}(m), J_\mathrm{y}\}$.

The above probabilistic model can be also mapped into some metric space via $p(\mathbf{y}^k|\mathbf{x}^i(m)) \propto e^{-d(\mathbf{y}^k,\mathbf{x}^i(m))}$, where $d(\mathbf{y}^k, \mathbf{x}^i(m))$ represents a distance between two descriptors.

The performance of the descriptors is measured in terms of their ROCs defined by the probabilities of miss $P_M^D = \Pr\{d(\mathbf{x}^k, \mathbf{Y}^k) \geq \epsilon L\}$ and probability of false acceptance $P_F^D = \Pr\{d(\mathbf{x}^i, \mathbf{Y}^k) < \epsilon L\}$ where ϵ is the threshold. In this paper, we assume that the descriptors of non enrolled items under hypothesis $\mathcal{H}_{m'}$ follow the same statistical distribution as under hypothesis $\mathcal{H}_m$.

Model of encoding/assignment In this paper, we will consider hard assignment to investigate the system performance under the minimum requested memory storage requirements [15, 18] [3]. The encoding matrix can be generally constructed as $\mathbf{C}_\mathrm{x}(m) = (\mathbf{c}_\mathrm{x}^1(m), \cdots, \mathbf{c}_\mathrm{x}^{J_\mathrm{x}(m)}(m)) \in \mathbb{R}^{J \times J_\mathrm{x}(m)}$, where each column $\mathbf{c}_\mathrm{x}^i(m)$ stands for the code representing encoding of the descriptor $\mathbf{x}^i(m)$, $1 \leq i \leq J_\mathrm{x}(m)$ with respect to the visual codebook $\mathcal{C}_\mathrm{x}$. In the case of hard assignment, $\mathbf{C}_\mathrm{x}(m) \in \{0,1\}^{J \times J_\mathrm{x}(m)}$ with the elements $c_{\mathrm{x}j}^i(m) = 1$ for $j : \mathbf{x}^j = \mathbf{x}^i(m)$ or zero-distance between the descriptor and codebook codeword, i.e., $j \in \mathcal{L}(\mathbf{x}^i(m))$ with the list $\mathcal{L}(\mathbf{x}^i(m)) = \{j \in \{1, \cdots, J\} : d(\mathbf{x}^j, \mathbf{x}^i(m)) = 0\}$.

Given the case that the descriptors are matched without geometrical consistency, i.e., they are desynchronized, and there are generally a different number of descriptors in the enrolled image $J_\mathrm{x}(m)$ and probe J_y, *pooling* is used. To address this, there are two common types of pooling *average-* and *max-* pooling. In the case of hard assignment at the enrollment stage they are equivalent. The enrolled fixed-length sparse code for the image m is $\mathbf{d}_\mathrm{x}(m) = (d_\mathrm{x}^1(m), \cdots, d_\mathrm{x}^J(m))^T \in \{0,1\}^J$ whose elements are $d_\mathrm{x}^{av_j}(m) = \sum_{i=1}^{J_\mathrm{x}(m)} c_{\mathrm{x}_j}^i(m)$ in the average-pooling and $d_\mathrm{x}^{max_j}(m) = \max_{1 \leq i \leq J_\mathrm{x}(m)} c_{\mathrm{x}_j}^i(m)$ in the max-pooling.

Model of decision making about the descriptor presence/absence and geometric consistency verification Given a probe $\mathbf{y}$ consisting of J_y descriptors, the encoding matrix for the probe is defined as $\mathbf{C}_\mathrm{y} = (\mathbf{c}_\mathrm{y}^1, \cdots, \mathbf{c}_\mathrm{y}^{J_\mathrm{y}}) \in \{0,1\}^{J \times J_\mathrm{y}}$, with $c_{\mathrm{y}j}^k = 1$ for $j \in \mathcal{L}(\mathbf{y}^k)$ with the list $\mathcal{L}(\mathbf{y}^k) = \{j \in \{1, \cdots, J\} : d(\mathbf{x}^j, \mathbf{y}^k) \leq \epsilon L\}$. This decoder corresponds to the BDD or ϵ-NN decoder which seeks all $\{\mathbf{x}^j\}$ NNs in the radius ϵL from the query descriptor $\mathbf{y}^k$.

The corresponding average- and max-pooled fixed-length vectors are defined as $d_\mathrm{y}^{av_j} = \sum_{k=1}^{J_\mathrm{y}} c_{\mathrm{y}_j}^k$ in the average-pooling and $d_\mathrm{y}^{max_j} = \max_{1 \leq k \leq J_\mathrm{y}} c_{\mathrm{y}_j}^k$ in the max-pooling. In following, we will only consider the max-pooling to its reported superior performance [16].

The statistics of matrix $\mathbf{C}_\mathrm{y}$ are completely defined by the probabilities of descriptor miss P_M^D and false acceptance P_F^D defined above.

Model of global decision making or decoding The final decision is based on the list decoder that produces a list of possible candidates:

$$\mathcal{L}(\mathbf{y}) = \{m \in \{1, \cdots, M\} : t(m) \geq \tau J\}, \quad (2)$$

where $t(m) = \mathbf{d}_\mathrm{x}^T(m)\mathbf{d}_\mathrm{y}$ stands for the similarity score between two vectors that can also represent a cosine distance, that is often used in the BOF systems, if the vectors are normalized by their norms $||\mathbf{d}_\mathrm{x}(m)||$ and $||\mathbf{d}_\mathrm{y}||$.

Remark: In the synchronized case, when the correspondence between the descriptors from two images is established, one can estimate the upper bound on the system performance by evaluating the similarity between two matrices as $t(m) = \mathbf{C}_\mathrm{x}(m) \odot \mathbf{C}_\mathrm{y}$, where $\odot$ denotes the Frobenius inner product [4].

5. PERFORMANCE UNDER MAX-POOLING AND PERFECT SYNCHRONIZATION

The overall system performance is evaluated by the probability of miss P_M, i.e., the correct m does not appear on the decoder's list under the hypothesis $\mathcal{H}_m$, $P_M = \Pr\{T(m) \leq \tau J_e|\mathcal{H}_m\}$ and by the probability of false acceptance P_F, i.e., an incorrect m' appears on the decoder's list under the hypothesis $\mathcal{H}_{m'}$, $P_F = \Pr\{T(m) > \tau J_e|\mathcal{H}_{m'}\}$, where $\tau \in (0,1)$ is the threshold and J_e stands for the equivalent length under different pooling strategies. The average list size can be estimated as $\mathbb{E}\{|\mathcal{L}(\mathbf{y})|\} = MP_F$. In the case of unique identification, the list size is 1.

Without loss of generality, we will assume that the same number of descriptors is enrolled for all images, i.e., $J_\mathrm{x}(m) = J_\mathrm{x}$, which is a reasonable assumption for most of the identification systems where the enrolment is under the control.

The sufficient statistics in the case of max-pooling and perfect synchronization are:

$$T(m) \sim \begin{cases} \mathcal{B}(J_e, \theta(m)), & \text{for } \mathcal{H}_m, \\ \mathcal{B}(J_e, \theta(m')), & \text{for } \mathcal{H}_{m'}, \end{cases} \quad (3)$$

where $\mathcal{B}$ denotes the Bernoulli distribution and for the max-pooling: $J_e = \min\{J_\mathrm{x}, J_\mathrm{y}\}$, $\theta(m) = 1 - (1 - P_D^D)(1 - P_F^D)^{J_\mathrm{y}-1}$ and $\theta(m') = 1 - (1 - P_F^D)^{J_\mathrm{y}}$ for the perfectly synchronized case: $\theta(m) = P_D^D$ and $\theta(m') = P_F^D$.

which manifest high correlation between elements.

[3] The hard/soft assignments represent a trade-off between the memory storage and decoding complexity.

[4] In the synchronized case, the matrices are of the same size.

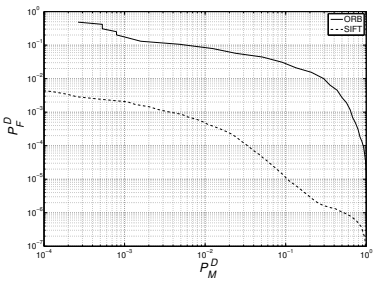

Fig. 2. Typical ROCs for matched and non-matched SIFT and ORB descriptors under scaling 0.5, rotation 10^0 and JPEG 75.

The performance of the content identification system is estimated based on the list decoder which is characterized by the probability of miss:

$$\begin{aligned} P_M &= Pr\{T(m) \leq \tau J_e | \mathcal{H}_m\} \\ &= \sum_{d=0}^{\tau J_e} \binom{J_e}{d} \theta^d(m)(1-\theta(m))^{J_e-d} \\ &\leq 2^{-J_e \mathcal{D}(\tau \| \theta(m))}, \end{aligned} \quad (4)$$

where $\mathcal{D}(\tau\|\theta(m))$ denotes the divergence, the threshold should satisfy $0 \leq \theta(m') < \tau < \theta(m) \leq 1$ and the average list of candidates is:

$$\begin{aligned} \mathbb{E}\{|\mathcal{L}(\mathbf{Y})|\} &= MP_F \\ &= M\ Pr\{T(m) > \tau J_e | \mathcal{H}_{m'}\} \\ &= M \sum_{d=\tau J_e}^{J_e} \binom{J_e}{d} \theta^d(m')(1-\theta(m'))^{J_e-d} \\ &\leq M 2^{-J_e \mathcal{D}(\tau \| \theta(m'))}. \end{aligned} \quad (5)$$

To keep the non-exponential size of the list of candidates, M should be chosen accordingly. Using the notion of the identification rate as $R = 1/J_e \log_2 M$, one can target the condition $R \leq \mathcal{D}(\tau\|\theta(m'))$ to keep this list small.

In some applications, it is interesting to keep both probabilities of errors small, for which one can minimize the maximum probability of error under optimal τ and ϵ as follows: $(\hat{\tau}, \hat{\epsilon}) = \text{argmin}_{\tau,\epsilon}(\max\{P_M(\tau,\epsilon), P_F(\tau,\epsilon)\})$.

6. RESULTS OF COMPUTER SIMULATION

Since the overall system performance is determined by the ROC curves of the descriptors, we first investigated the typical ROC curves for SIFT and ORB descriptors shown in Figure 2 for the *copydays* database [19] containing 157 images generating approximately 100'000 descriptors and SIFT produces better results at a computational cost.

The experimental distributions of parameter $T(m)$ and their theoretical counterparts (3) under the hypothesis $\mathcal{H}_m$ and $\mathcal{H}_{m'}$ for the max-pooling and synchronized case for a pair $P_M^D = 0.3$, $P_F^D = 0.001$ and $J_x = J_y = 50$ are shown in

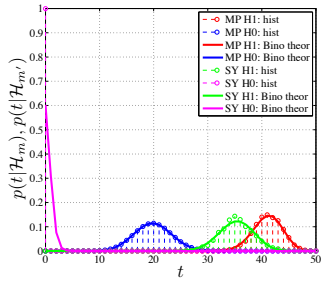

Fig. 3. Distributions of similarity score for the ORB descriptor with max-pooling (MP) and perfect synchronization (SY) under correct and incorrect hypothesis.

Figure 3. The theoretically predicted curves are in accordance with the experimental data.

Finally, the overall performance of system is summarized in terms of $\max\{P_M(\tau,\epsilon), P_F(\tau,\epsilon)\}$ as a function of ϵ and τ for the theoretical distributions [5]. Note that the system has a global minimum for optimal pair of the thresholds τ, ϵ. Therefore, if one specifies the descriptor, our model suggests a set of optimal parameters under max-pooling to optimize overall performance. Additionally, the theoretically attainable performance for the perfectly synchronized descriptors indicates the gap with max-pooling and significance of further re-ranking steps.

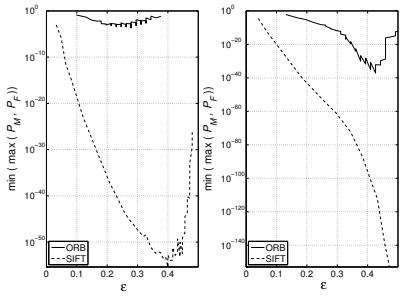

Fig. 4. Performance of BOF-based identification system for the max-pooling (left) and synchronized (right) system with SIFT and ORB descriptors.

7. CONCLUSION

In this paper, we introduced a simple and tractable model of BOF content identification systems. We plan to extend this model to (a) compare the max- and average-pooling strategies, (b) find an optimal regime for the descriptors encoding for different pooling methods, (c) find the number of items M leading to a non-exponential size of list of candidates and finally (d) investigate the impact of compression and soft assignment on overall system performance.

[5] The results are obtained based on the analytical pmfs to highlight the order of expected exponents which can not be attained by limited computer simulations.

8. REFERENCES

[1] B. Girod, V. Chandrasekhar, D. Chen, N.-M. Cheung, R. Grzeszczuk, Y. Reznik, G. Takacs, S. Tsai, and R. Vedantham, "Mobile visual search," *Signal Processing Magazine, IEEE*, vol. 28, no. 4, pp. 61–76, 2011.

[2] Christian Wengert, Matthijs Douze, and Hervé Jégou, "Bag-of-colors for improved image search," in *ACM Multimedia*, Scottsdale, United States, Oct. 2011.

[3] Hervé Jégou, Florent Perronnin, Matthijs Douze, Jorge Sánchez, Patrick Pérez, and Cordelia Schmid, "Aggregating local image descriptors into compact codes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1704–1716, Sept. 2012, QUAERO.

[4] Jorge Sanchez, Florent Perronnin, Thomas Mensink, and Jakob Verbeek, "Image Classification with the Fisher Vector: Theory and Practice," *International Journal of Computer Vision*, vol. 105, no. 3, pp. 222–245, Dec. 2013.

[5] D. Lowe, "Distinctive image features from scale-invariant keypoints," in *International Journal of Computer Vision*, 2003, vol. 20, pp. 91–110.

[6] H. Bay, T .Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *In ECCV*, 2006, pp. 404–417.

[7] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: an efficient alternative to sift or surf," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2564–2571.

[8] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *Computer Vision–ECCV 2010*, pp. 778–792. Springer, 2010.

[9] P. Moulin, "Statistical modeling and analysis of content identification," in *Information Theory and Applications Workshop (ITA), 2010*. IEEE, 2010, pp. 1–5.

[10] A. Varna and M. Wu, "Modeling and analysis of correlated binary fingerprints for content identification," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 3, pp. 1146–1159, 2011.

[11] F. Farhadzadeh, S. Voloshynovskiy, and O. Koval, "Performance analysis of content-based identification using constrained list-based decoding," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1652–1667, 2012.

[12] S. Voloshynovskiy, O. Koval, F. Beekhof, F. Farhadzadeh, and T. Holotyak, "Information-theoretical analysis of private content identification," in *IEEE Information Theory Workshop, ITW2010*, Dublin, Ireland, Aug.30-Sep.3 2010.

[13] Y. Zhang, R. Jin, and Z.-H. Zhou, "Understanding bag-of-words model: a statistical framework," *International Journal of Machine Learning and Cybernetics*, vol. 1, no. 1-4, pp. 43–52, 2010.

[14] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, 2003, pp. 1470–1477.

[15] H. Jegou, M. Douze, and C. Schmid, "Product quantization for nearest neighbor search," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 1, pp. 117–128, 2011.

[16] L. Liu, L. Wang, and X. Liu, "In defense of soft-assignment coding," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2486–2493.

[17] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 3360–3367.

[18] F. Farhadzadeh, F. M.J. Willems, and S. Voloshynovskiy, "Fundamental limits of identification: Identification rate, search and memory complexity trade–off," in *IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, July 7–12 2013.

[19] H. Jégou, M. Douze, and . Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *European Conference on Computer Vision*, Andrew Zisserman David Forsyth, Philip Torr, Ed. oct 2008, vol. I of *LNCS*, pp. 304–317, Springer.