Video copy detection in large databases: A local signatures probabilistic similarity search approach

A. Joly, C. Frélicot and O. Buisson

INRIA Rocquencourt, BP 105 Rocquencourt, 78153 Le Chesnay Cedex France

Abstract

Content-based copy detection (CBCD) is one of the emerging multimedia applications for which there is a need of a concerted effort from the database community and the computer vision community. In this work, we put forward a new approximate search paradigm dedicated to CBCD in large databases and we evaluate it in a complete video CBCD framework based on local signatures. The search of similar signatures in the database is not based on classical range or KNN queries but on probabilistic distortion-based queries.

Copies and distorted signatures

A copy is not only a similar document but also a transformed document:

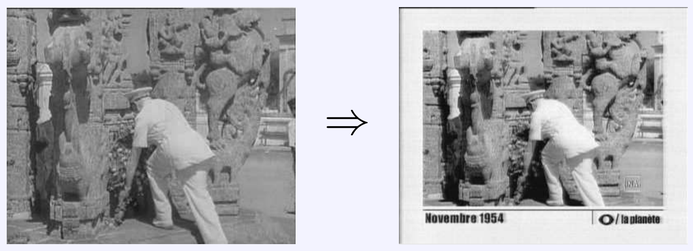


Fig. 1 Two copies and their original

A signature $\boldsymbol{S}(t(M))$, extracted in a transformed document $t(M)$, can thus be considered as a distorted version of the original signature $\boldsymbol{S}(M)$, extracted in the original document M. We define the distortion as the following variable:

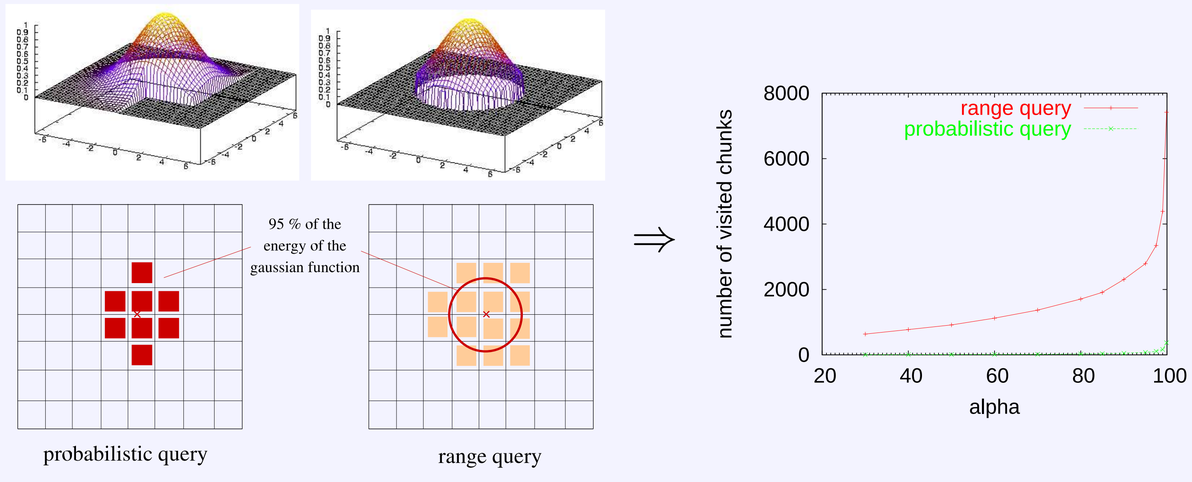
$$\boldsymbol{\Delta S} = \boldsymbol{S}(M) - \boldsymbol{S}(t(M))$$

Distortion-based Probabilistic Queries

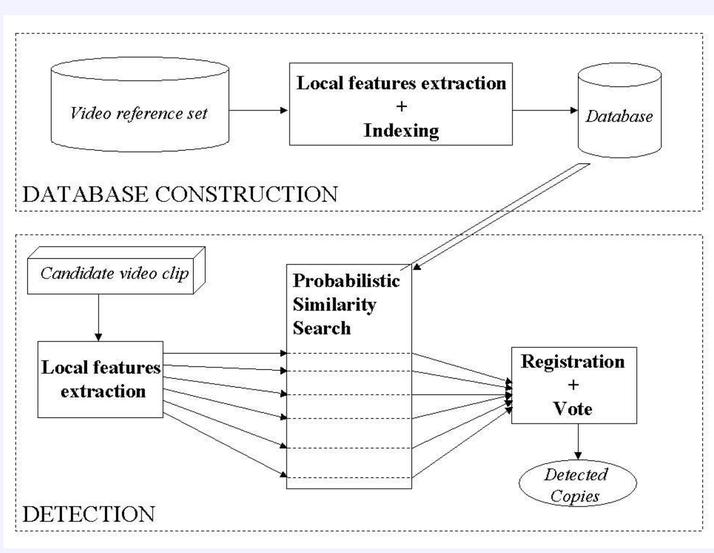
We define a *distortion-based probabilistic query*, associated to a probability equal to α, as the search of all the signatures contained in a region V_α of the feature space satisfying:

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

where $\boldsymbol{Q}$ is the query (i.e. the candidate signature) and $p_{\boldsymbol{\Delta S}}(.)$ is the probability density function of the distortion. Intuitively, the probabilistic query selects a region of the feature space such as the probability of finding signatures that could belong to a copy is equal to α.



Video Copy Detection Framework



Local signatures = differential invariants extracted around Harris interest points in key images ($\approx 1\,billion\,signatures$ for $10,000$ hours of video). Geometric consistency is post-computed from all local results by a spatio-temporal registration on the points positions P_i:

$$\left(\widehat{\boldsymbol{A}}(V_h), \widehat{\boldsymbol{B}}(V_h)\right) = \arg\min_{\boldsymbol{A},\boldsymbol{B}} \sum_{i=1}^{n_c} \min_{\substack{k \in K_i \\ V_{ik}=V_h}} \rho\left(\|\boldsymbol{P_{ik}} - (\boldsymbol{A P_i} + \boldsymbol{B})\|\right) \qquad (2)$$

Experimental results

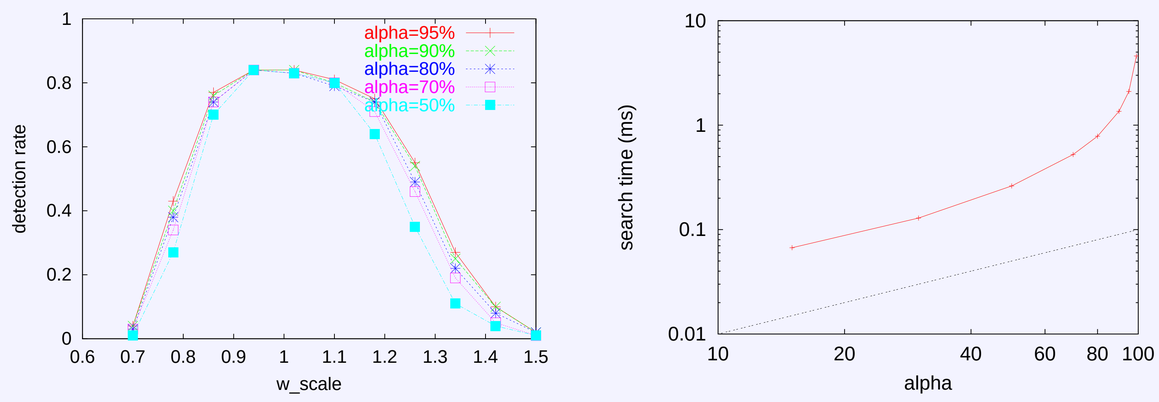


Fig. 3 Influence of α on the recall after an image resizing (left) and the speed (right)

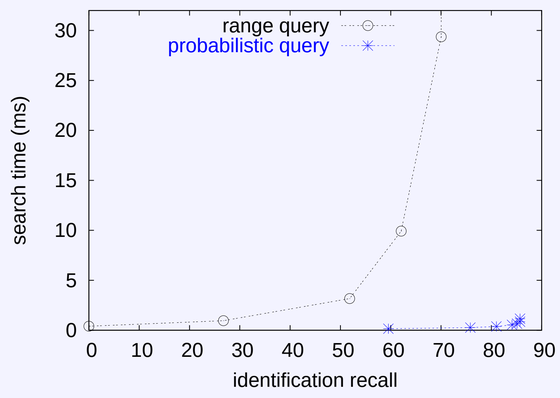


Fig. 2 Comparison of distortion-based probabilistic queries and exact range queries - Search time with respect to recognition recall, at constant precision ($p_r = 90\%$)

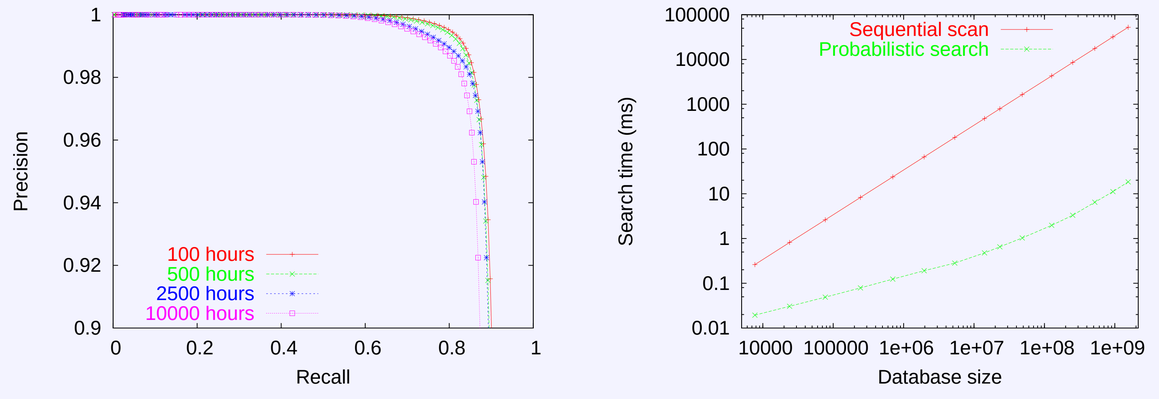


Fig. 3 Influence of the database size on ROC (left) and speed (right)