Inferring semantics from textual information in multimedia retrieval

Mats Sjoberg, Jorma Laaksonen, Timo Honkela, Matti Polla

Sistemi Cognitivi 13/14

Perlo Daniele

Obbiettivo

Low Level Features -> GAP -> Semantic Concepts

Symbol Grounding Problem
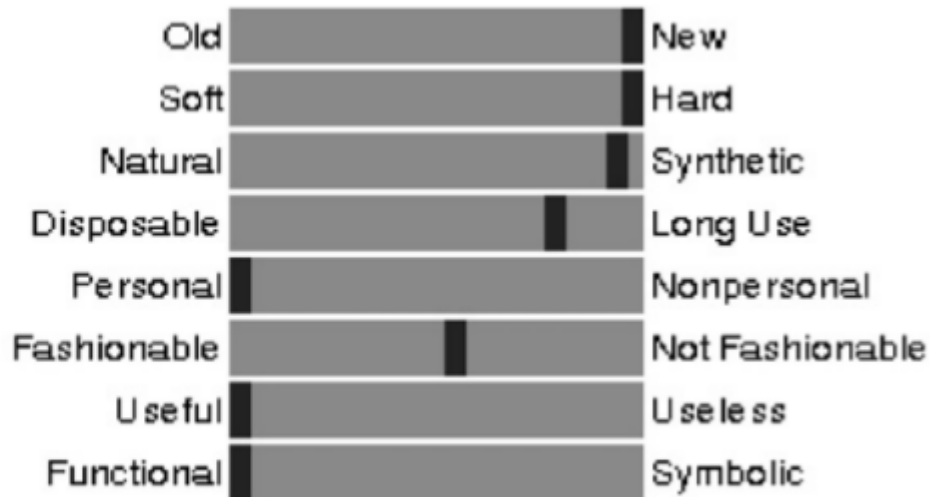
Recupero delle informazioni testuali

Fig. 3. An "attackalarm" and the property values given by its owner.

Semantic classes from the "Pockets full of memories" data set

Semantic class description	A priori (%)	Training set	Test set
Old	27.1	501	400
New	60.6	938	1077
Soft	32.8	541	551
Hard	53.8	894	895
Natural	23.2	343	429
Synthetic	70.4	1207	1134
Disposable	24.0	422	375
Long-use	67.3	1083	1155
Personal	70.7	1135	1217
Nonpersonal	21.3	373	336
Fashionable	54.8	860	963
Not-fashionable	24.5	426	389
Useful	74.6	1212	1270
Useless	17.9	332	264
Functional	58.5	903	1044
Symbolic	31.5	571	476
All objects	100	1678	1649

Semantic classes from the TRECVID 2005 data set

Semantic class description	A priori (%)	Training set	Test set
An explosion or a fire	1.08	109	265
Regional territory graphic (map)	1.90	376	282
A US flag	0.79	123	151
An exterior of a building	7.28	1578	943
Waterscape or waterfront	2.30	375	420
A captive person	0.16	23	32
Any sport in action	2.59	460	437
A car	7.27	1279	1239
All objects	100	17 230	17 407

Features
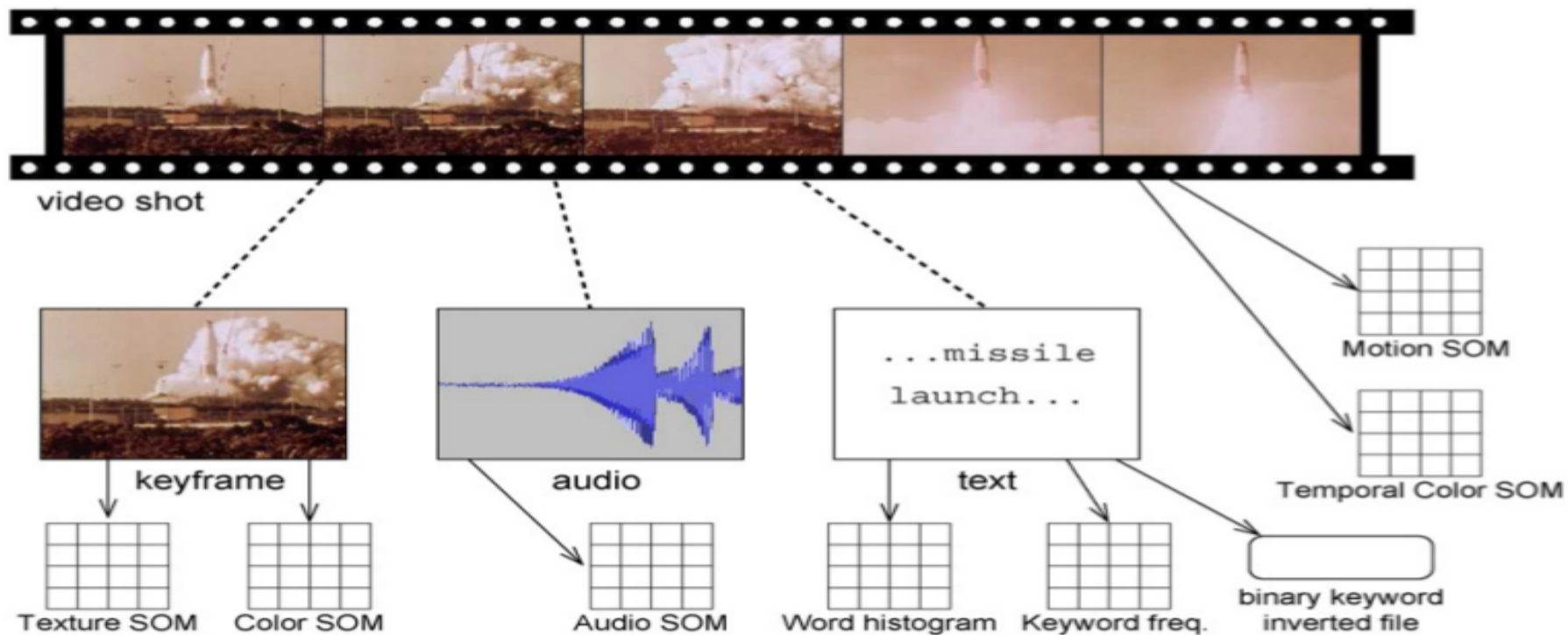
Fig. 2. The hierarchy of videos and examples of multi-modal SOMs.

Training: SOMs

Self-Organizing Map

$\forall$ semantic contest c
$\quad \forall$ feature f
$\qquad$ create a SOM_cf
$\qquad \forall$ object o about c
$\qquad\quad$ add f_o at SOM_cf
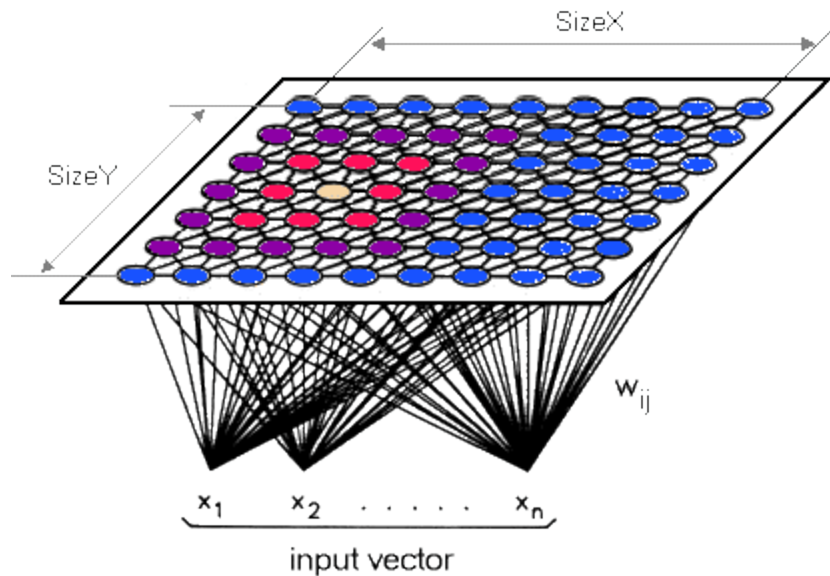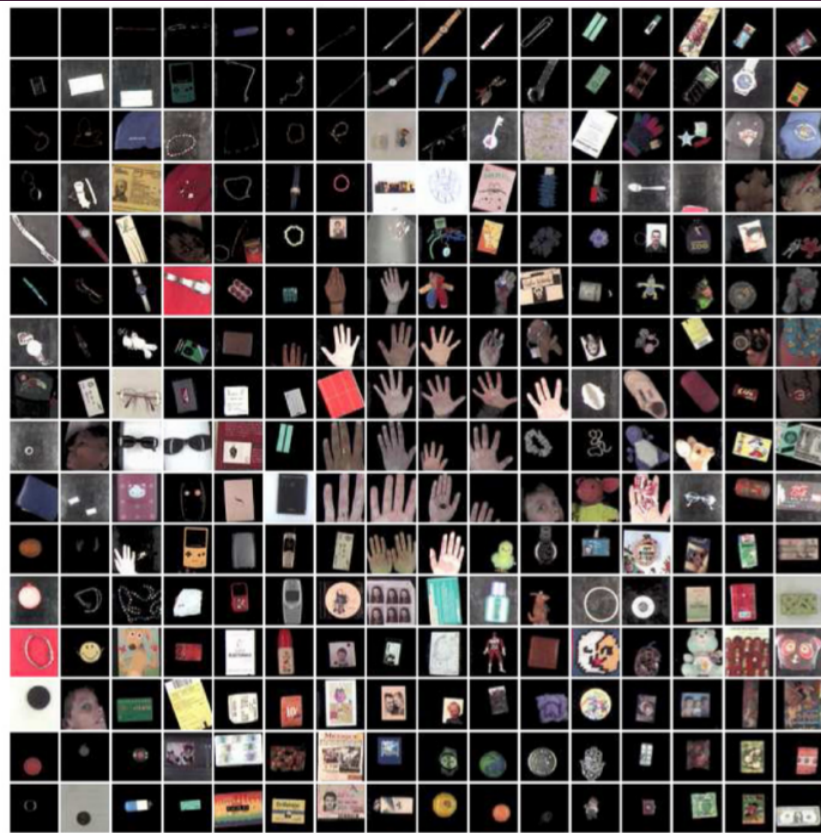
Edge Histogram
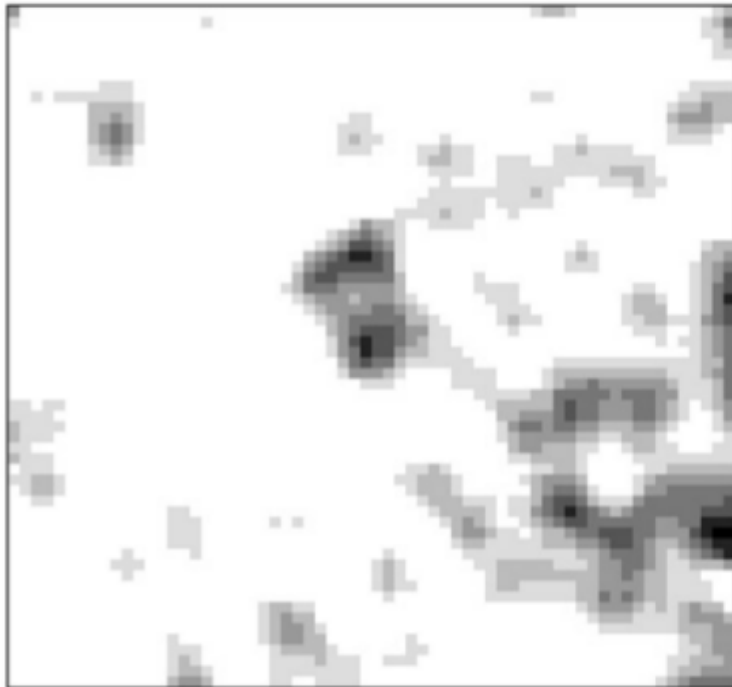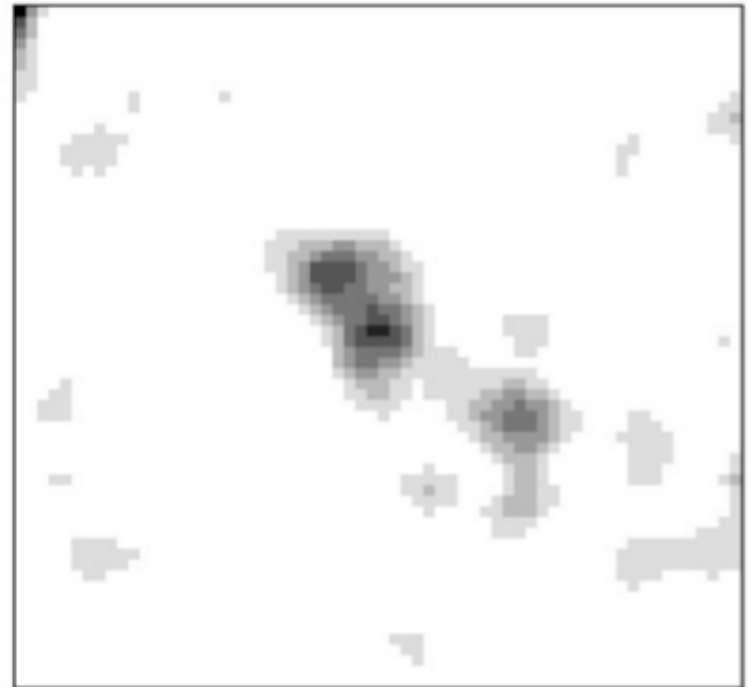
soft   natural

Textual features

Word Histogram

$$tf * \log(idf)$$

Textual features

Keyword Frequency

$\forall$ semantic contest c

keyword list from all obj in c from frequency

keyword ranks ->

score(keyword list, keyword reference list)

Textual features

Binary Keywords

$\forall$ semantic contest c

$$S_c(k) = \frac{N_{c,k}}{N_c} - \frac{N_k}{N} \qquad S_{i,c} = \sum_k \frac{\delta_{i,k}}{N_k}, \quad \text{where } \delta_{i,k} = \begin{cases} 1 & \text{if } k \text{ exists in } i, \\ 0 & \text{otherwise.} \end{cases}$$

based on Inverse List

Retrivial: PicSOM
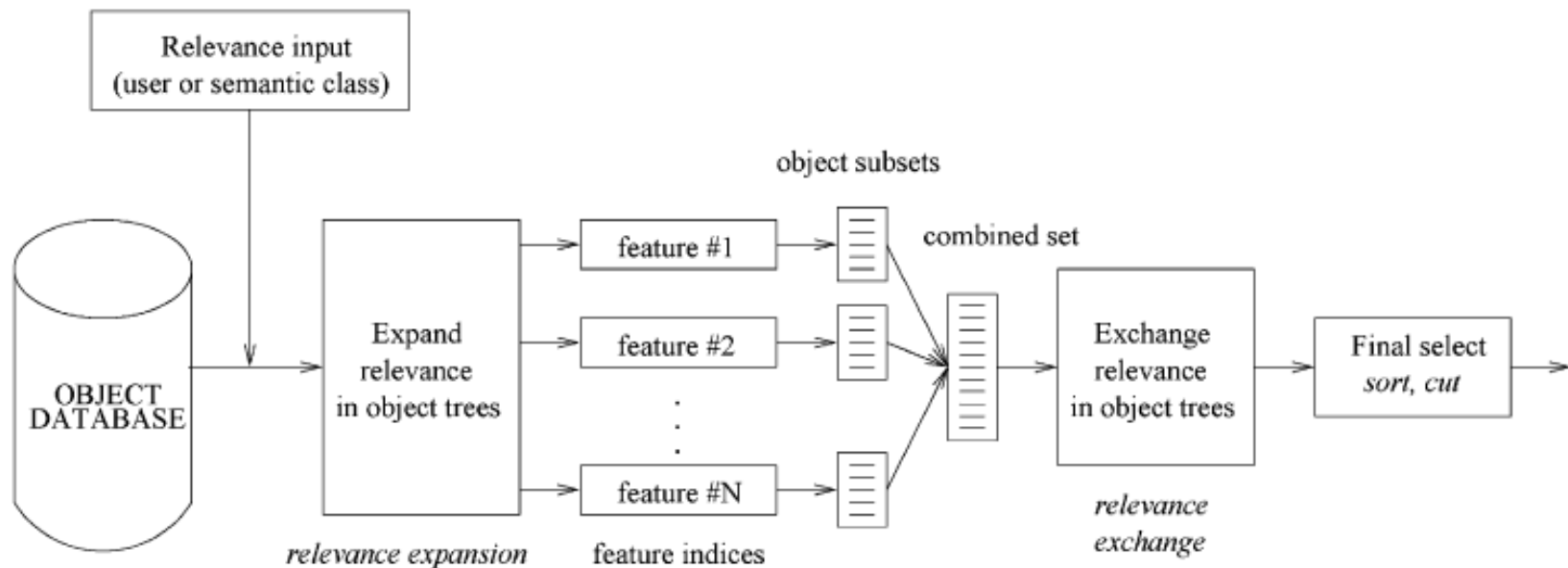
Fig. 1. Processing stages in PicSOM.
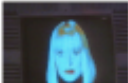
Retrivial: PicSOM

Input: Test Set & Semantic Contenst c

Relevant obj: $w>0$ Not Relevant obj: $w>0$

$\forall$ object o
  $\forall$ frature f
    add w at BMU in SOM_cf

-> QUALIFICATION VALUE _ ocf

Retrivial: PicSOM

obj Rank:

Sum of qualification values

Results

Precision & Recall

Features:

- only non-textual features (*nt*)
- word histogram (*wh*)
- keyword frequency (*kwf*)
- binary keywords (*bkw*)

Average precision results for "Pockets full of memories" experiments

Semantic class	nt	wh	bkw	wh + nt	bkw + nt
Old	0.262	0.302	0.344	0.309	**0.350**
New	0.594	0.630	**0.706**	0.636	0.699
Soft	0.337	0.398	0.514	0.405	**0.522**
Hard	0.546	0.562	0.704	0.573	**0.706**
Natural	0.260	0.353	**0.462**	0.343	0.458
Synthetic	0.677	0.723	**0.852**	0.727	**0.852**
Disposable	0.237	0.302	**0.394**	0.297	0.393
Long-use	0.641	0.698	**0.821**	0.695	0.820
Personal	0.652	0.729	**0.806**	0.717	0.804
Nonpersonal	0.248	0.255	0.399	0.270	**0.400**
Fashionable	0.502	0.586	**0.650**	0.578	0.644
Not-fashionable	0.253	0.269	0.348	0.264	**0.354**
Useful	0.692	0.758	**0.842**	0.753	0.839
Useless	0.188	0.211	0.327	0.229	**0.331**
Functional	0.557	0.639	**0.740**	0.631	0.737
Symbolic	0.322	0.354	0.496	0.368	**0.503**
Average	0.436	0.486	**0.588**	0.487	**0.588**

Average precision results for TRECVID experiments

Semantic class	nt	wh	kwf	bkw	wh + nt	kwf + nt	bkw + nt
Explosion or fire	0.0567	0.0061	0.0104	0.0285	0.0595	0.0583	**0.0779**
Map	0.3396	0.0061	0.0411	0.0049	0.3402	0.3402	**0.3433**
Depicting US flag	0.0713	0.0023	0.0052	0.0059	0.0763	0.0825	**0.0848**
Exterior of building	0.0988	0.0108	0.0042	0.0068	0.0985	**0.0996**	0.0990
Waterscape or front	0.2524	0.0056	0.0090	0.0053	0.2482	0.2484	**0.2515**
Captive person	0.0054	0.0043	0.0025	0.0000	**0.0161**	**0.0161**	0.0158
Sport in action	0.2240	0.0108	0.0338	0.0970	0.2207	0.2328	**0.2666**
A car	0.2818	0.0177	0.0065	0.0114	**0.2824**	0.2820	0.2800
Average	0.1662	0.0080	0.0141	0.0200	0.1677	0.1700	**0.1774**