Modelling Approaches and Results of the FHDO Biomedical Computer Science Group at ImageCLEF 2015 Medical Classification Task

Obioma Pelka Christoph M. Friedrich

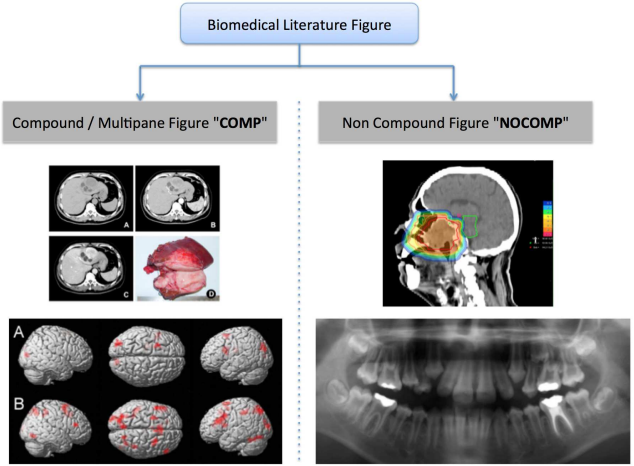
University of Applied Sciences and Arts Dortmund (FHDO)
Department of Computer Science

09. September 2015

- First participation in ImageCLEF 2015 Medical Classification Task
- Biomedical Computer Science Group (BCSG)
  1. Task1 - Compound Figure Detection Task
     - 1 Visual Run
     - 1 Textual Run
     - 4 Mixed (Visual + Textual) Runs
  2. Task4 - Subfigure Classification Task
     - 1 Visual Run
     - 1 Textual Run
     - 6 Mixed (Visual + Textual) Runs

Task Definition

Discriminate compound from non-compound figures



[DISCLAIMER: Figures used for illustration are from the ImageCLEF 2015 Medical Classification Task Training Set]

State of the art

Task non-specific

Bag-of-Keypoints[4]

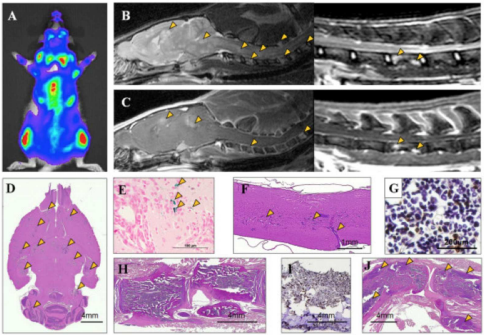
- Dense Scale Invariant Feature Transform
- Approximated nearest neighbour $k-$means Clustering

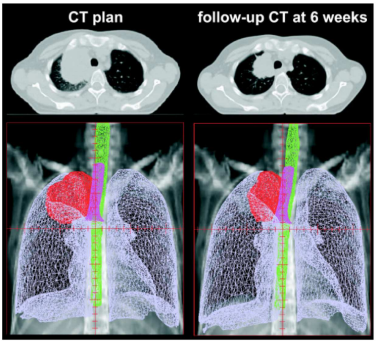
Bag-of-Words[7]

- Porter stemming
- Removal of stop-words
- Frequency of words
- Occurrence class

Border Profile

Task specific

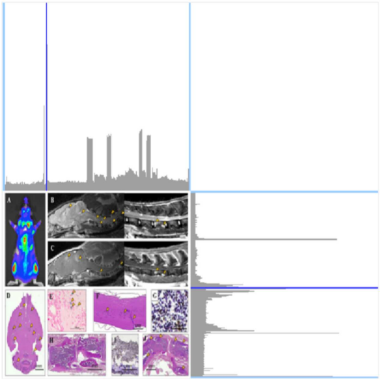


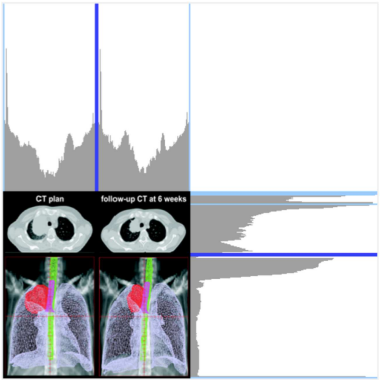


[DISCLAIMER: Figures used for illustration are from the ImageCLEF 2015 Medical Classification Task Training Set]

Border Profile

Task specific





[DISCLAIMER: Figures used for illustration are from the ImageCLEF 2015 Medical Classification Task Training Set]

Characteristic Delimiter

Task specific

- Compound Figure = 2 or more subfigures
- Subfigures are mentioned in the caption

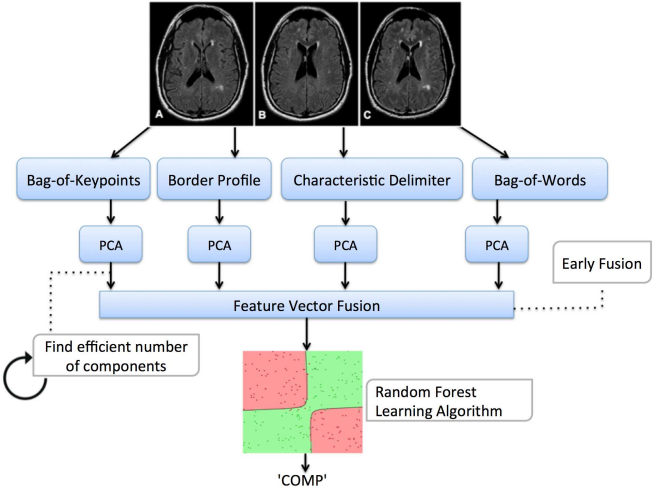
```
<figures>
    <figure iri="1479-5876-7-88-4">
        <caption> Central Nervous System and skeletal involvement by breast cancer .  (A) Bioluminescence
            image of rat at week 3 shows high photon flux activity from the brain, spine, and joints. (B)
            Sagittal T2w MRI and (C) contrast enhanced T1w MRI show hyperintense lesions on the brain, spinal
            cord, and vertebral bodies (arrowheads). Histological section of brain (D) and spinal cord (F)
            with hematoxylin and eosin (HE) staining from group 2 rat euthanized at week 4 reveals numerous
            metastases (arrowheads). E) Prussian blue staining of the consecutive brain section from D shows
            few isolated iron positive cells near the tumor (arrows). H) Thoracic spine with tumor
            infiltration on HE stain. Cytokeratin immuno-histochemical staining of the bone marrow aspirates
            (G) and Spine (I) is positive for tumor. J) Knee joint metastasis with extraskeletal involvement
            is seen (arrowheads) on HE stain.
        </caption>
    </figure>
```

- Detect delimiter pairs

	1	2	...
1	A	B	
2	(A.	(B.	
3	(A)	(B)	
4	i	ii	
5	upper	lower	
6	1)	2)	
...			

[DISCLAIMER: Captions used for illustration are from the ImageCLEF 2015 Medical Classification Task Training Set]

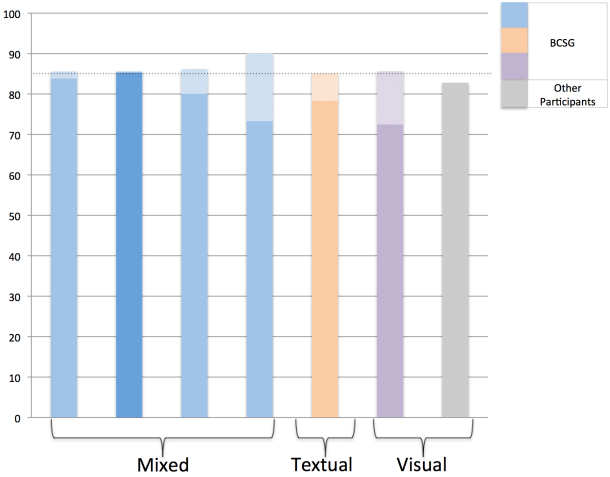
Classifier Setup



Descriptor	Original Vector Size	Reduced Vector Size
BoK	12000	20
BoW	3906	20
Border Profile	4	4
Charact. Deli	2	2

[DISCLAIMER: Figures used for illustration are from the ImageCLEF 2015 Medical Classification Task Training Set]

Results of submitted runs



- Discrepancy between evaluation and development results
- Supplementing visual with textual features improves prediction accuracy

Task Definition

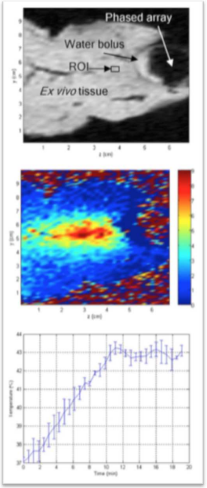


Fig-X-100

Fig-X-100-1

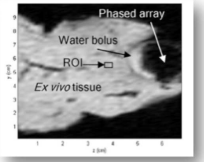


Fig-X-100-2

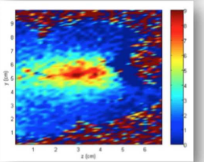
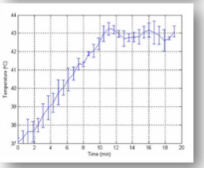


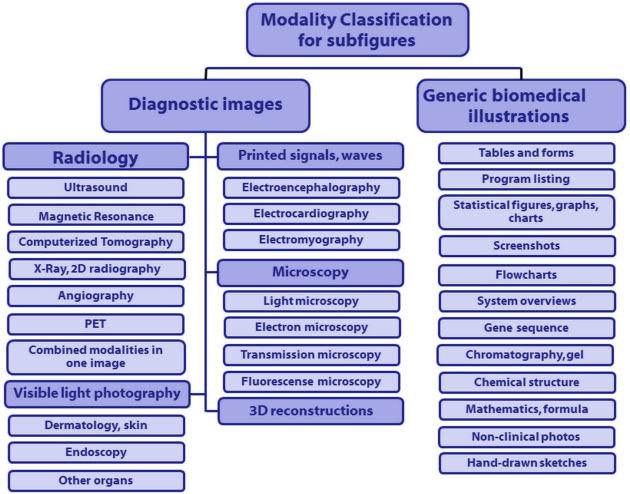
Fig-X-100-3



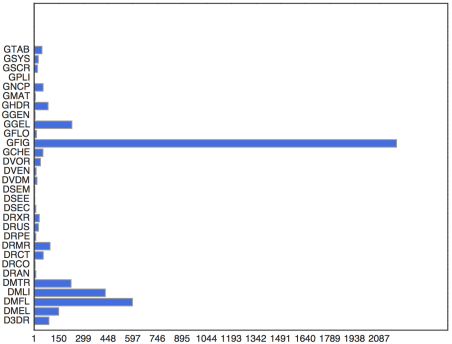
[DISCLAIMER: Figures used for illustration are from the ImageCLEF 2015 Medical Classification Task Training Set]

Task Definition

Flat modality hierarchy
70% learning data and 30% validation data



Task Definition



Extended Learning Collection

1. **DataSet$_1$:** ImageCLEF 2015 Medical Classification
2. **DataSet$_2$:** DataSet$_1$ + ImageCLEF 2013 AMIA Medical Task
3. **DataSet$_3$:** DataSet$_1$ + ImageCLEF 2013 AMIA Modality Classification
4. **DataSet$_4$:** ImageCLEF 2013 Modality Classification

Visual Features

Color

- Fuzzy Color Histogram (FCH)[5]
- Color and Edge Directivity Descriptor (CEDD)[2]
- Joint Composite Descriptor (JCD)[2]

Texture

- Tamura[8]
- Gabor[9]

Local

- Bag-of-Keypoints (Dense Scale Invariant Feature Transform)[4]
- Pyramid Histogram of Oriented Gradients (PHOG)[1]

Global

- Basic Features (BAF)[6]

Textual Features

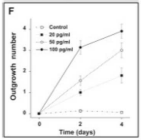
Bag-of-Words[7]

- Codebook built with DataSet_4
- χ^2-test to compute attribute importance[3]
- Caption trimmed to relevance

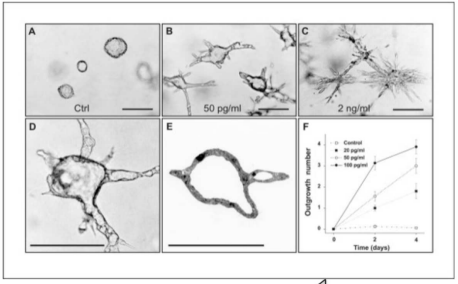
[DISCLAIMER: Figures and captions used for illustration are from the ImageCLEF 2015 Medical Classification Task Training Set]

Textual Features

1471-213X-7-7-2-6



1471-213X-7-7-2

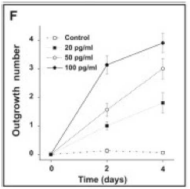


Low concentrations of exogenous TGF- β 1 induce morphogenesis of branching tubules. **(A)** J3B1A cells grown in a collagen gel in defined medium for a total of 10 days. **(B)** Parallel culture in which J3B1A cells were grown in a collagen gel for 6 days to allow cyst formation and were subsequently treated with 50 pg/ml TGF- β 1 for an additional 4 days. TGF- β 1 has induced the outgrowth of tube-like structures from the wall of existing cysts. **(C)** Treatment with 2 ng/ml TGF- β 1 has resulted in the formation of numerous thin cell cords extending out into the surrounding collagen matrix. Notably, at this relatively high concentration, TGF- β 1 also disrupts the organization of preformed cysts, resulting in lumen obliteration. **(D)** Higher magnification view of a multicellular structure formed in a culture treated with 20 pg/ml TGF- β 1 for 4 days. The outgrowths enclose a patent lumen, which at least in some tubes is continuous with the cavity of the cyst. **(E)** Semi-thin section of a collagen gel culture of J3B1A cells treated with 50 pg/ml TGF- β 1 for 4 days. Bars (A-E), 200 μ m. **(F)** Quantitative analysis of TGF- β 1-induced tubulogenesis. J3B1A cells were grown in a collagen gel for 6 days to allow cyst formation and were subsequently treated with different concentrations of TGF- β 1.

[DISCLAIMER: Figures and captions used for illustration are from the ImageCLEF 2015 Medical Classification Task Training Set]

Textual Features

1471-213X-7-7-2-6



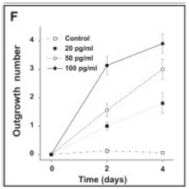
	1	2	3	4	5	6	7	...
1	A	B	C	D	E	F	G	
2	(A.	(B.	(C.	(D.	(E.	(F.	(G.	
3	(A)	(B)	(C)	(D)	(E)	(F)	(G)	
...								

Low concentrations of exogenous TGF- β 1 induce morphogenesis of branching tubules. **(A)** J3B1A cells grown in a collagen gel in defined medium for a total of 10 days. **(B)** Parallel culture in which J3B1A cells were grown in a collagen gel for 6 days to allow cyst formation and were subsequently treated with 50 pg/ml TGF- β 1 for an additional 4 days. TGF- β 1 has induced the outgrowth of tube-like structures from the wall of existing cysts. **(C)** Treatment with 2 ng/ml TGF- β 1 has resulted in the formation of numerous thin cell cords extending out into the surrounding collagen matrix. Notably, at this relatively high concentration, TGF- β 1 also disrupts the organization of preformed cysts, resulting in lumen obliteration. **(D)** Higher magnification view of a multicellular structure formed in a culture treated with 20 pg/ml TGF- β 1 for 4 days. The outgrowths enclose a patent lumen, which at least in some tubes is continuous with the cavity of the cyst. **(E)** Semi-thin section of a collagen gel culture of J3B1A cells treated with 50 pg/ml TGF- β 1 for 4 days. Bars (A-E), 200 μ m. **(F)** Quantitative analysis of TGF- β 1-induced tubulogenesis. J3B1A cells were grown in a collagen gel for 6 days to allow cyst formation and were subsequently treated with different concentrations of TGF- β 1.

[DISCLAIMER: Figures and captions used for illustration are from the ImageCLEF 2015 Medical Classification Task Training Set]

Textual Features

1471-213X-7-7-2-6



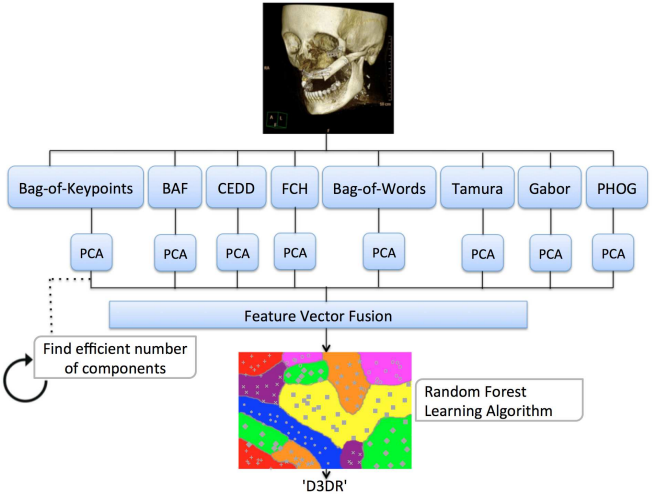
	1	2	3	4	5	6	7	...
1	A	B	C	D	E	F	G	
2	(A.	(B.	(C.	(D.	(E.	(F.	(G.	
3	(A)	(B)	(C)	(D)	(E)	(F)	(G)	
...								

Low concentrations of exogenous TGF- β 1 induce morphogenesis of branching tubules. **(A)** J3B1A cells grown in a collagen gel in defined medium for a total of 10 days. **(B)** Parallel culture in which J3B1A cells were grown in a collagen gel for 6 days to allow cyst formation and were subsequently treated with 50 pg/ml TGF- β 1 for an additional 4 days. TGF- β 1 has induced the outgrowth of tube-like structures from the wall of existing cysts. **(C)** Treatment with 2 ng/ml TGF- β 1 has resulted in the formation of numerous thin cell cords extending out into the surrounding collagen matrix. Notably, at this relatively high concentration, TGF- β 1 also disrupts the organization of preformed cysts, resulting in lumen obliteration. **(D)** Higher magnification view of a multicellular structure formed in a culture treated with 20 pg/ml TGF- β 1 for 4 days. The outgrowths enclose a patent lumen, which at least in some tubes is continuous with the cavity of the cyst. **(E)** Semi-thin section of a collagen gel culture of J3B1A cells treated with 50 pg/ml TGF- β 1 for 4 days. Bars (A-E), 200 μ m. **(F)** Quantitative analysis of TGF- β 1-induced tubulogenesis. J3B1A cells were grown in a collagen gel for 6 days to allow cyst formation and were subsequently treated with different concentrations of TGF- β 1.

[DISCLAIMER: Figures and captions used for illustration are from the ImageCLEF 2015 Medical Classification Task Training Set]

Classifier Setup

Early Fusion

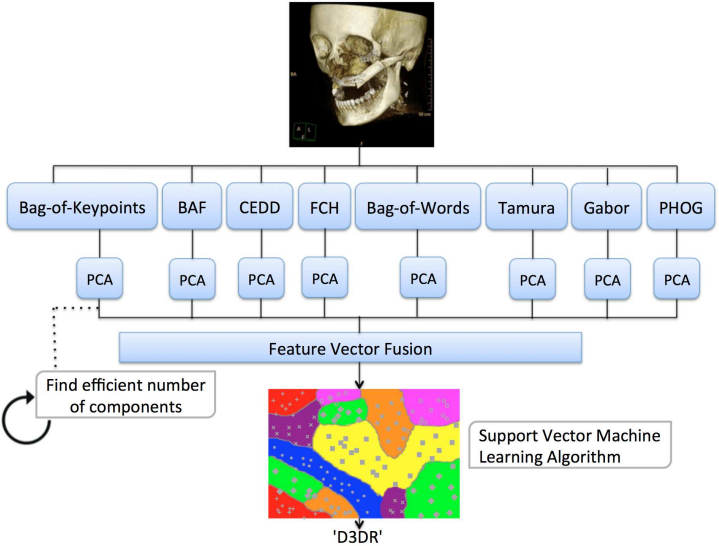


Descriptor	Original Vector Size	Reduced Vector Size
BoK	12000	25
BoW	438	40
BAF	8	8
CEDD	144	5
FCH	10	3
Gabor	60	0
JCD	168	5
Tamura	18	2
PHOG	630	2

[DISCLAIMER: Figures used for illustration are from the ImageCLEF 2015 Medical Classification Task Training Set]

Classifier Setup

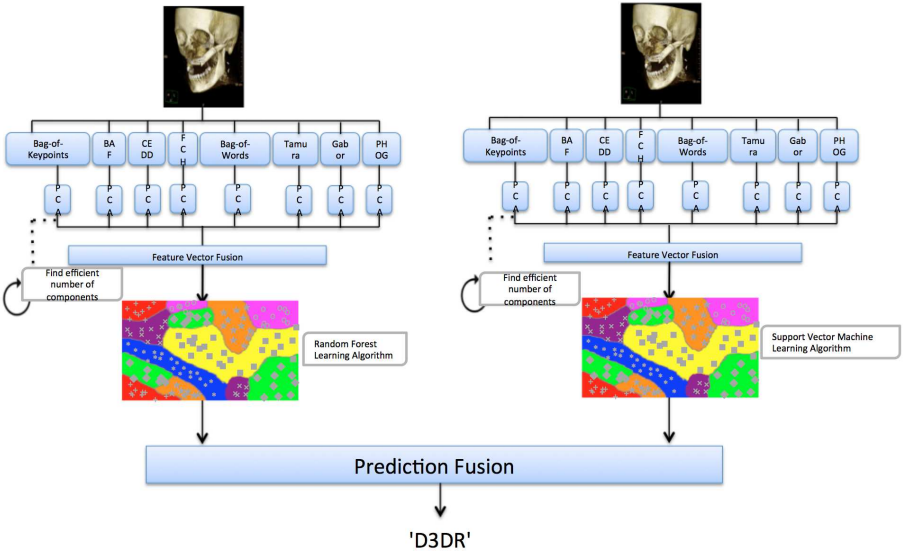
Early Fusion



[DISCLAIMER: Figures used for illustration are from the ImageCLEF 2015 Medical Classification Task Training Set]

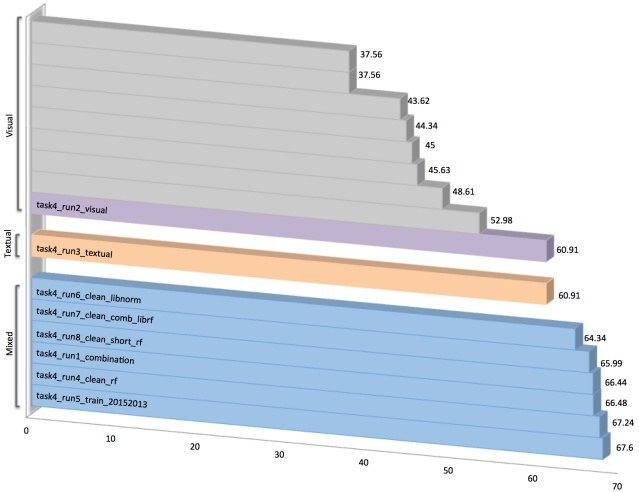
Classifier Setup

Late Fusion



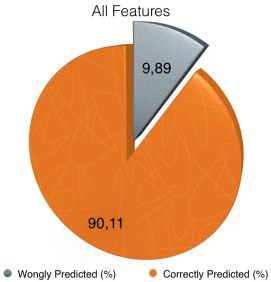
[DISCLAIMER: Figures used for illustration are from the

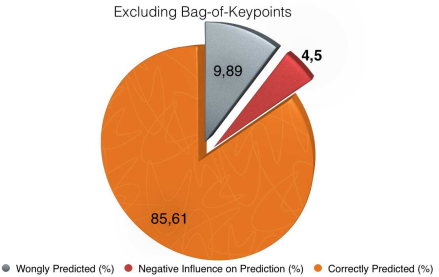
Results of submitted runs



- Colored bars: Results of Biomedical Computer Science Group
- Gray bars: Results of other participants
- Runs to all submission categories: Visual, Textual, Mixed(Visual + Textual)

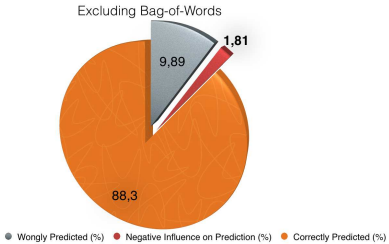
Feature Contribution - Compound Figure Detection

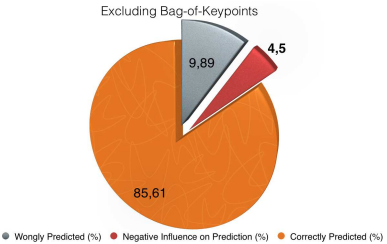


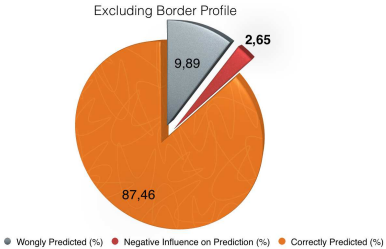


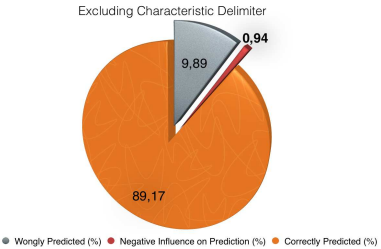
Analysis computed by applying classifier model Run4 using sampled training and validation set

Feature Contribution - Compound Figure Detection

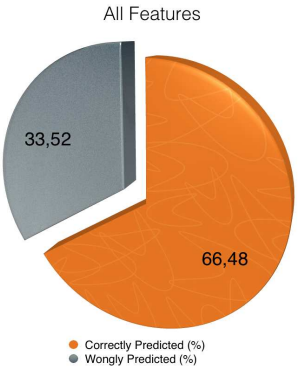


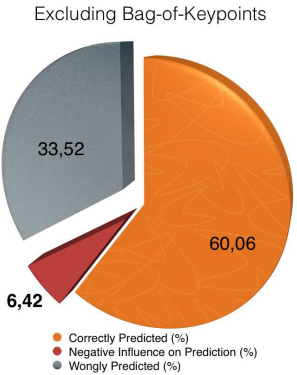






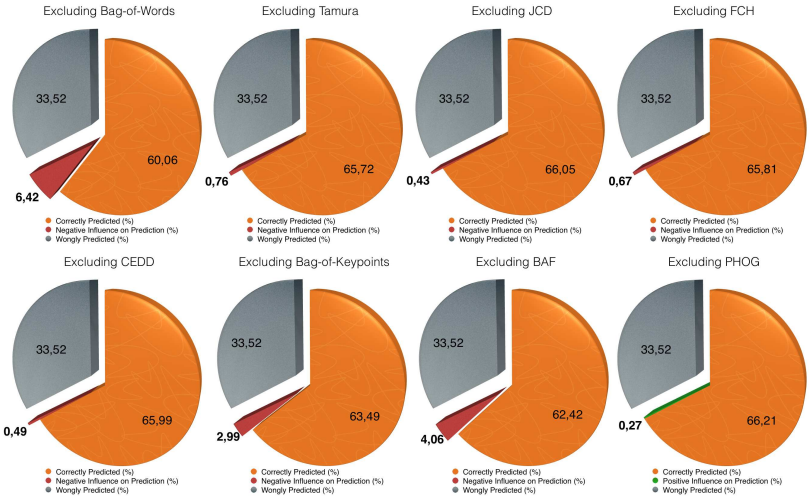
Feature Contribution - Subfigure Classification





Analysis computed by applying classifier model Run1 using the original evaluation set

Feature Contribution - Subfigure Classification



Findings

1. Better results with dense SIFT than Lowe SIFT
2. Precision Format Double vs Single
3. Supplementing visual with textual features increases prediction accuracy
4. Border Profile can be further extended with additional colors
5. Image to single before feature extraction

Thank you for listening!

References

[1] Anna Bosch, Andrew Zisserman, and Xavier Munoz.
Representing shape with a spatial pyramid kernel.
In Proceedings of the 6th ACM International Conference on Image and Video Retrieval, CIVR '07, pages 401–408, New York, NY, USA, 2007. ACM.

[2] Savvas A. Chatzichristofis and Yiannis S. Boutalis.
Compact Composite Descriptors for Content Based Image Retrieval: Basics, Concepts, Tools.
VDM Verlag, Saarbrücken, Germany, 2011.

[3] William G. Cochran.
The χ^2 test of goodness of fit.
Ann. Math. Statist., 23(3):315–345, 1952.

[4] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and CÃ©dric Bray.
Visual categorization with bags of keypoints.
In In Workshop on Statistical Learning in Computer Vision, ECCV, pages 1–22, 2004.

[5] Ju Han and Kai Kuang Ma.
Fuzzy color histogram and its use in color image retrieval.
IEEE Transactions on Image Processing, 11(8):944–952, 2002.

[6] Mathias Lux and Savvas A. Chatzichristofis.
Lire: Lucene image retrieval an extensible java cbir library.
In Abdulmotaleb El-Saddik, Son Vuong, Carsten Griwodz, Alberto Del Bimbo, K. SelÃ§uk Candan, and Alejandro Jaimes, editors, ACM Multimedia, pages 1085–1088. ACM, 2008.

[7] Gerard Salton and Michael J. McGill.
Introduction to modern information retrieval.
McGraw-Hill computer science series. McGraw-Hill, New York, 1983.

[8] Hideyuki Tamura, Shunji Mori, and Takashi Yamawaki.
Texture features corresponding to visual perception.
IEEE Transactions on System, Man and Cybernatic, 6, 1978.

[9] M R Turner.
Texture discrimination by gabor functions.
Biol. Cybern., 55(2-3):71–82, 1986.