On Communicating Computational Research

Dan Ellis

Laboratory for Recognition and Organization of Speech and Audio Dept. Electrical Eng., Columbia Univ., NY USA

dpwe@ee.columbia.edu

http://labrosa.ee.columbia.edu/



Computational Research
 Five Ways to Present Results
 Sharing Code
 Conclusions



Communicating Computational Results - Dan Ellis

I. Computational Research

• 50 years of computational research (in SP):

1963





1988



2013

Accessible	Within lab	Researchers	Anyone
Coding time (FFT)	I month	I day	<1 minute

The Paradigm

- Computational approaches can be very complex
- Proposed techniques are often elude theoretical analysis
- Empirical results are "the proof of the pudding"



Example: Soundtrack Classification

Ellis, Zheng, McDermott 'I I

• Trained models using "texture" features



Results on 9,317 videos (210 hours)



Communicating Computational Results - Dan Ellis

2013-04-04 4 /13

resenting Results

^g Correlogram time slice

• Compute times:

aundonal

Subband VQ

Subband VQ

Subband VQ

Subband VQ

Table 1. Comparison of feature properties. Calculation timesare over the 210 h CCV data set on a single CPU.

Histogram

paper:

	MFCC	SAI (reduced)	SBPCA
Feature extraction	5.6 h	1087 h	310 h
Feature/patch dims	60	48	60
<pre># patches/codebooks</pre>	1	24 (8)	4
Codebook size	3000	1000	1000
Histogram size	3000	24000 (8000)	4000

SUBBAND AUTOCORRELATION FEATURES FOR VIDEO SOUNDTRACK CLASSIFICATION

Courtenay V. Cotton, Daniel P. W. Ellis

LabROSA, Dept. of Electrical Engineering Columbia University {cvcotton,dpwe}@ee.columbia.edu

ABSTRACT

Inspired by prior work on stabilized auditory image features, we have developed novel auditory-model-based features that preserve the fine time structure lost in conventional frame-based features. While the original auditory model is computationally intense, we present a simpler system that runs about ten times faster but achieves equivalent performance. We use these features for video soundtrack classification with the Columbia Consumer Video dataset, showing that the new features alone are roughly comparable to traditional MFCCs, but combining classifiers based on both features achieves a 15% improvement in mean Average Precision over the MFCC baseline.

Index Terms – Acoustic signal processing, Multimedia databases, Video indexing, Auditory models 1. INTRODUCTION particularly useful for the identification of sounds in mixtures. Since we are working with broadly similar problems of classifying unconstrained environmental audio, we attempting to replicate their system as closely as possible to test it on a consumer video soundtrack retrieval task.

The next sections introduce our data/domain, and then describe our results using an available implementation of the auditory model front-end, and our modified, simplified features aiming to capture the same information. Sections 5 and 6 describe other experimentation with the original system, experimenting with replacing the original PAMIR retrieval model and with more common Support Vector Machine (SVM) classifiers, and with reduce the dimensionality of the representation. Section 7 describes the further improvements we obtained by fusing these novel features with the existing baseline MFCCs.

2. DATASET AND TASK

As the means to collect and share video and audio become increasingly ubiquitous and cheap, automatic taggine retrieval of multimedia content become portant. Although much recome



Communicating Computational Results - Dan Ellis

2013-04-04 5 /13

Presenting Results

Code & Data release ~5000 lines of Matlab ~5 GB of data

● ● ●	genVidClassif	- Generic Video	Soundtrack Cl	assification fram	iework		M ²¹
🔺 🕨 🙆 😰 🔘 🖾 🌀 www.ee.columbia.edu/~dpwe/resources/matlab/gen 🖒 Reader							
↔ ∰ Ⅲ D	elici.dpwe ->delic	RAPS Site elmp	AladdinShPt	babelswordfish	CVNFacCtr	GooDocs	>>
Panel to	Former	www.sta	Columbi	macboo	genVid	c	+ Imi

Dan Ellis : Resources: Matlab:

genVidClassif - Generic Video Soundtrack Classification framework

This package provides a set of routines for training and applying classifiers for video soundtracks.

Because I was working with multiple different sets of videos with different label definitions, I tried to come up with a common convention they could all use. This is based around a set of text files with fixed names. I'm calling this framework "gen" (for generic video soundtrack classification), so the directory

/u/drspeech/data/aladdin/data/gen/

.. contains several data sets arranged in this format.

For this example, I'm going to use the "small" MED2011 DEV set, which is the 2062 example videos covering the 15 event categories, plus the 4292 distractor videos from LDC2011E06. I call this set DEVT1+Ev, so the files describing it are in:

/u/drspeech/data/aladdin/data/gen/DEVT1+Ev/

which contains:

categories.txt



Columbia Consumer Video (CCV) Database

--- A Benchmark for Consumer Video Analysis

Summary

Communicating Computational Results - Dan Ellis

2013-04-04 6 /13

2. Five Ways to Present Results

- Researchers want to present their results
 - "publish or perish"
 - paper & citation counts
 - impact & fame
- Math/Humanities model
 the paper is the product
- Science/Engineering model
 the paper describes the product

Five Ways...

I. Traditional Publications

pro: Present your "spin" con: Not the whole story

2. Talks / Demos / videos o pro: Quick hook

• con: Distorting

SUBBAND AUTOCORRELATION FEATURES FOR VIDEO SOUNDTRACK CLASSIFICATION

Courtenay V. Cotton, Daniel P. W. Ellis

LabROSA, Dept. of Electrical Engineering Columbia University {cvcotton,dpwe}@ee.columbia.edu

ABSTRACT

Inspired by prior work on stabilized auditory image features, we have developed novel auditory-model-based features that preserve the fine time structure lost in conventional fram-based features. While the original auditory model is computationally intense, we present a simpler system that runs about ten times faster but achieves equivalent performance. We use these features for video soundtrake classification with the Columbia Consumer Video dataset, showing that the new features alone are roughly comparable to traditional MFCCs, but combining classifiers based on both features achieves a 15% improvement in mean Average Precision over the MFCC baseline. particularly useful for the identification of sounds in mixtures. Since we are working with broadly similar problems of classifying unconstrained environmental audio, we attempting to replicate their system as closely as possible to test it on a consumer video soundtrack retrieval task.

The next sections introduce our data/domain, and then describe our results using an available implementation of the auditory model front-end, and our modified, simplified features aiming to capture the same information. Sections 5 and 6 describe other experimentation with the original system, experimenting with replacing the original PAMIR retrieval model and with more common Support Vector Machine (SVM) classifiers, and with reduce the dimensionality of the representation. Section 7 describes the further improvements we obtained by fusing these novel features with the existing baseline MFCCS.

databases, Video indexing, Auditory models

2. DATASET AND TASK

I. INTRODUCTION

As the means to collect and share video and audio become
increasingly ubiquitous and cheap, automatic tagging
retrieval of multimedia content become
portant. Although much

Index Terms- Acoustic signal processing, Multimedia



Communicating Computational Results - Dan Ellis

2013-04-04 8 /13

Five Ways...

3. Interactive demos

pro: Let people ask their own questions con: Substantial additional development



4. Libraries / APIs

 pro: Promotes uptake
 con: Development and support liability

<pre>#echonest (</pre>	Develope	er Accou	nt Down	loads Forums Blog	
Table Of Contents	Track AF	PI Meth	ods		
Echo Nest API Overview Overview Keys Ground Rules Encoding Authentication Rate Limits Identifiers	Methods for analyzing or getting info about tracks. Analysis - The track analysis includes summary information about a track including ten danceability, loudness, liveness, speechinesss and energy along with detailed informa structure (bars, beats tatums) and detailed info about timbre, pitch and loudness envelo interpret the analyzer output see the Analyze Documentation				
Project Rosetta Stone Standard Parameters Response Codes	profile				
Resources Discussion Release History	Get info about to	acks given an	id or md5. Th	e md5 parameter is the file md5.	
biographies	ani key	ves	no		
blogs familiarity hotttnesss	id	one of id or md5	no	TRTLKZV12E5AC92E11	
images list control	md5	one of id or	no	881f4e47e88e8b570e34a3b49c8262ac	

The Fifth Way: Code Sharing

Complete description of what you did

• "share the research equipment"

Pros

- every detail, regardless of your spin
- allows replication & reuse
- the best way to uncover bugs

Cons

time to preparedirty laundrycompetitive edge



3. Sharing Code

Code Sharing & the Scientific Mission

- Scientific fields traditionally struggle to develop reproducible protocols
- Commodity computers & software support unprecedented reproducibility

• Barriers [Stodden 2010]

Code		Data
77%	Time to document and clean up	54%
52%	Dealing with questions from users	34%
44%	Not receiving attribution	42%
40%	Possibility of patents	-
34%	Legal Barriers (ie. copyright)	41%
-	Time to verify release with admin	38%
30%	Potential loss of future publications	35%
30%	Competitors may get an advantage	33%
20%	Web/disk space limitations	29%

Communicating Computational Results - Dan Ellis

2013-04-04 11/13

The Future of Sharing Code

- The opportunity to share code is novel
 Better, more consistent, high-level platforms
 Open Source movement
- There are drawbacks

Time to prepareFear of humiliation

There are huge advantages
 Scientific mission: reproduction, verification, debugging
 Impact



Computational research is qualitatively different
 o and great

• but hard to comprehend

- Traditional publications describe superficially
 not a good match
 editorial choices about "what matters"
- Cheap & powerful computers support code sharing
 O ''if I cap't fix it I dop't own it''

• ''if I can't fix it, I don't own it''

• but: airing dirty laundry

• Waiting for a generational change...