

Discovering Meaningful Multimedia Patterns with Audio-Visual Concepts and Associated Text

L. Xie, L. Kennedy, S.-F. Chang, A. Divakaran, H. Sun, and C.-Y. Lin

TR2004-128 May 2005

Abstract

This paper presents algorithms for finding the meanings of the audio-visual video patterns obtained in the unsupervised discovery process. This problem is interesting in domains where neither perceptual patterns nor semantic concepts have simple structures. The patterns in the video are modeled with hierarchical hidden Markov models, with efficient algorithms to jointly learn the model parameters, the optimal model complexity, as well as the relevant feature subsets. The meanings are contained in words of the speech transcript of the video. The pattern-word association is obtained via co-occurrence analysis and machine translation models. Promising results are obtained on TRECVID news videos: video patterns that associate with distinct topics such as el-nino and politics are identified; a temporal structure model compares favorably to a non-temporal clustering algorithm.

Proc. IEEE Intl Conf. Image Processing

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

DISCOVERING MEANINGFUL MULTIMEDIA PATTERNS WITH AUDIO-VISUAL CONCEPTS AND ASSOCIATED TEXT

L. Xie[†], L. Kennedy[†], S.-F. Chang[†], A. Divakaran[§], H. Sun[§], C.-Y. Lin[‡]

[†]Dept. of Electrical Engineering, Columbia University, New York, NY

[‡]IBM T.J. Watson Research Center, Hawthorne, NY

[§]Mitsubishi Electric Research Labs, Cambridge, MA

ABSTRACT

This paper presents algorithms for finding the meanings of the audio-visual video patterns obtained in the unsupervised discovery process. This problem is interesting in domains where neither perceptual patterns nor semantic concepts have simple structures. The patterns in the video are modeled with hierarchical hidden Markov models, with efficient algorithms to jointly learn the model parameters, the optimal model complexity, as well as the relevant feature subsets. The meanings are contained in words of the speech transcript of the video. The pattern-word association is obtained via co-occurrence analysis and machine translation models. Promising results are obtained on TRECVID news videos: video patterns that associate with distinct topics such as *el-nino* and *politics* are identified; a temporal structure model compares favorably to a non-temporal clustering algorithm.

1. INTRODUCTION

This paper presents solutions towards discovering meaningful structures in video in multiple modalities. Structures, or patterns, in temporal sequences refer to the repetitive segments that bear consistent characteristics in the observations and the dynamics. Automatic identification of structures from video is an interesting topic for both theoretical problems on learning in multi-modality and applications in multimedia content organization.

Supervised learning techniques are capable of learning the target structure once the domain knowledge is encoded in the choice of the feature set, the construction of the statistical model, and the design of the learning algorithms. Unsupervised structure discovery, on the other hand, tries to find statistical descriptions of the structures. It using has been shown [1] that in interesting game states in sports videos can be revealed without supervision using a few low-level features. Temporal structures identified by pure computational criteria awaits the association and evaluation with *meanings* before they become usable. This association can be manually performed when meanings in the original content are unambiguous and few, such as in sports programs. However, in domains with more general and diverse content, such as news videos, tagging structures with meanings is no longer a straightforward task, and the difficulty comes from the large number of topics present and the inherent uncertainty in the large hierarchy of *meanings*. In produced video content, the prevalence of metadata, such as closed captions, speech transcript, gamestats or screenplays, provides a complimentary channel of semantic information for bridging this gap.

In this work, we aim to discover meaningful structures in audio-visual concept space with the additional information provided in

the metadata. An audio-visual concept space is a collection of elementary concepts such as “people”, “building”, “monologue” each of which was learned from low-level features in a separate supervised training process [2]. We believe that such mid-level concepts offer a promising direction to revealing the semantic meanings in patterns, since grouping and post-processing beyond the signal level is a vital part for the understanding of sensory inputs [3], and multi-modal perception is no less complicated than perception in individual senses. Temporal structures in the audio-visual sequence, characterized by the strength in each concept, the mutual dependency among concepts and their temporal evolutions, are captured by a hierarchical hidden Markov model (HHMM), learnable with statistical inference techniques such as Expectation-Maximization (EM) and Monte Carlo method. Once a description of temporal structure is in place, the first step towards understanding its meaning is to examine the co-occurrence statistics between the structure labels and words obtained from the speech transcript. The co-occurrence statistics are further refined by estimating an underlying generative probability between the labels and the words with machine translation models. These techniques were first proposed by Brown et. al. [4], and later used to associate images blobs with words [5, 6]. The former [5] was set in a context with clean text labels that can be treated as concepts in themselves; while the latter [6] operates on the keyframes in video shots without taking into account the temporal structure. We use news videos as the test domain and find promising associations from the video patterns to distinct topics such as *el-nino* or *politics*, we have also demonstrated the advantage of using a dynamic structure model to a plain clustering alternative.

The rest of this paper is organized as follows, Section 2 discusses the unsupervised discovery of video patterns using HHMM, Section 3 presents algorithms for associating the patterns with the speech transcript; Section 4 includes the experiment results on news videos; Section 5 summarizes this work and discusses open issues.

2. UNSUPERVISED LEARNING OF VIDEO PATTERNS

Solutions to unsupervised structure discovery address two objectives in one pass: finding a statistical description of the common properties and locating the corresponding segments in the sequence. We are interested in models that describe the properties of each individual video unit (frame or shot) as well as the temporal transitions among these units. Distinct appearance and transition patterns exist in produced video contents such as TV programs and feature films, and a two-level HHMM is a model with an efficient inference algorithm suitable for this purpose. We use the algo-

gorithms described in an earlier work [1] to learn the HHMM; a summary is presented in Section 2.1 for completeness.

Our feature set consists of the confidence values from twelve concept detectors obtained from the IBM concept detection system in TRECVID evaluations [7, 2]. The confidence scores are obtained by fusing the results of multiple support vector machine (SVM) classifiers applied to the key frame of each shot¹ in the video. We uniformly quantize the scores into three levels before learning the HHMM. The concepts are $\{\textit{weather, people, sports, non-studio, nature-vegetation, outdoors, news-subject-face, female speech, airplane, vehicle, building, road}\}$, selected from the 16 TRECVID-evaluated concepts that have a reported average precision greater than 50%. Using shots as the basic analysis units is advantageous for the news domain, because the production syntax usually produces clear cuts, and the content within a shot is usually consistent. We use this concept space mainly for its availability and performance assurance, while the choice of an optimal concept space or a proper concept lexicon is still an open question.

2.1. Hierarchical hidden Markov models

HHMM is a generalization of HMM with hierarchical control structure in the hidden states; it is also a specialization of Dynamic Bayesian network (DBN). The parameter set Θ of a two-level HHMM consists of within-level across-level transition probabilities and emission parameters that specifies the distribution of observations conditioned on the state configuration. Parameter estimation on this model can be carried out with an EM algorithm similar to that of the HMM, taking into account additional transition and control constraints. The complexity of the algorithm is $O(T)$, where T is the sequence length.

The size of the HHMM state-space represents the number of interesting structures in the data, and it is often desirable to determine a suitable size automatically rather than manually supply a fixed value. Reverse-jump Markov chain Monte Carlo (MCMC) provides an effective computational framework for achieving this goal. The algorithm involves (1) proposals towards changing the size of the HHMM state-space (*jump*) or staying put, which are stochastically (*Monte Carlo*) related the current configuration (*Markov chain*); (2) an assessment of the jump with the Bayesian information Criteria (BIC), followed by probabilistic decision of acceptance or rejection (*reversible*). The proposals are repeated until an *optimal* size-parameter configuration declared or a maximum number of rounds reached. Details of this approach is in [1].

2.2. Maintaining a group of models

The pattern discovery method above can be used on any feature set – leading to the question of relevant feature grouping. In an unsupervised learning context, the criteria for *relevance*, i.e., agreement among a group of features in representing a concept becomes relative to the numerous concepts simultaneously present but not explicitly pinpointed in the content. Therefore, we adopt a wrapper method to maintain a pool of models learned over different feature subsets for further exploration into the meaning of the clusters.

Using all 2^D possible feature subsets among the original D -dimensional features is obviously undesirable in both computation and representation. We select L ($L \ll 2^D$) feature subsets using mutual information criteria wrapped around the HHMM model. An HHMM model is first learned over each one-dimensional feature sequence, generating D different hidden state labels for each observation; next, we measure the mutual information $I(Q^{d_1}, Q^{d_2})$

(where Q is the random variable for the hidden states) between any pair of labels using features d_1 and d_2 , respectively; we then perform agglomerative clustering on the D features using their pair-wise mutual information as the similarity measure; the feature clusters are traversed breadth-first from the top of the agglomerative grouping, until there are L feature subsets; finally, we re-learn K HHMMs from the identified feature groups, and keep M label sequences $q_{1:S_k}^{(m)}$, $k = 1, \dots, L$, $m = 1, \dots, M$ for each video clip of length S_m . Grouping single features using pair-wise mutual information neglects the higher-order interactions among multiple features, but this yields more stable results than the greedy growing strategy for partitioning the feature pool [1].

3. ASSOCIATING STRUCTURE WITH METADATA

The quest for meanings in the temporal structures begins by associating HHMM *labels* (taking values from a *pattern lexicon*) to *tokens* (from a *word lexicon*) in the metadata stream. The objective of this association is two-fold: tagging structures with meanings and assessing the goodness of them.

3.1. Text processing from speech transcript

The Automatic Speech Recognition (ASR) transcript of the TREC videos are in the form of time-stamped words (t, \hat{w}_t) . The discourse style of a news program is usually concise and direct, hence it suffices to stay at the level of individual words rather than going to higher-level concepts via linguistic or semantic analysis, to represent the subject of discussion. We choose to focus on a lexicon of frequent and meaningful words, freeing ourselves from the noise introduced by stop words and statistically insignificant ones. The lexicon is obtained from the TRECVID development corpus after the following shallow parsing operations: (1) Stem the words from ASR output, resulting in about 12,500 distinct *tokens*; (2) Prune tokens that appear on average less than once in each half-hour news video, 647 tokens survived; (3) Perform rule-based part-of-speech tagging on the tokens², and only retain the 502 nouns, verbs, adjectives or adverbs; (4) Further prune away a few frequent tokens with no concrete meaning such as the verbs “be”, “do”, “have”, the adverbs “more”, “even”, “just”, “still”, and words too common in news programs such as “new”, “today”. Denote the set of pruned ASR tokens from the original transcript as (t, w_t) , taking values from the final lexicon $W = \{w_1, w_2, \dots, w_N\}$.

3.2. Co-occurrence analysis

We obtain the co-occurrence statistic $C(q, w)$ for a HHMM *label* q and a *token* w by counting the number of times that the state label q and the word w both appear in the same temporal segment among all video clips.

Denote the set of K videos as $\mathcal{S} = \{S_1, \dots, S_K\}$, let each of the video S_k be partitioned into a set of closed, non-overlapping segments $\mathcal{S}_k = \{s_i, i = 1, \dots, S_k\}$. Denote the maximum-likelihood state *labels* obtained with the HHMMs as $q_{\bar{s}}^m \in Q^m$, $m = 1, \dots, M$, $\bar{s} = 1, \dots, \bar{S}_k$, $k = 1, \dots, K$, where m indexes the M HHMM models, Q^m is the state-space of the m -th HHMM, and \bar{s} indexes the shots in each of the K clips in the current set of videos. The co-occurrence statistic $C(q^m, w)$, defined as the number of segments that label q^m and token w appear together, is accumulated across all the video segments as follows, where “ \vee ” denotes logical OR, and $1()$ is the indicator function.

²Hugo Liu, MontyTagger: Commonsense-Informed Part-of-Speech Tagging, <http://web.media.mit.edu/~hugo/montytagger/>

¹A *shot* refers to a continuous camera take in both time and space.

$$C(q^m, w) = \sum_{k=1}^K \sum_{s \in S_k} \bigvee_{\bar{s} \cap s \neq \phi} 1(q_s^m = q^m) \cdot \bigvee_{t \in s} 1(w_t = w) \quad (1)$$

$$\forall q^m \in Q^m, w \in W$$

There are two natural temporal divisions in news videos on which we can compute the co-occurrence statistics: shots¹ and stories. The latter is defined as “a segment of a news broadcast with a coherent news focus which contains at least two independent, declarative clauses” by TRECVID. Despite the convenience of directly using shots as the temporal division on which the HHMM labels are generated, we find it beneficial to use stories in establishing the *label-token* correspondence. This is because: (1) Meanings in news are conveyed at the story level with a sequence of shots in the visual channel and several sentences in the speech; the transition in topics happen at story boundaries. (2) Within a shot, words being said are often not in sync with the visual content. Evaluations show that co-occurrence on shots yields precisions only about one-tenth of that of the stories while producing comparable recalls. Without overloading the notations, the set of all stories in video S_k is denoted as $S_k = \{s_i, i = 1, \dots, S_k\}$ hereafter.

Once the co-occurrence statistics are in place, we normalize the co-occurrence counts to obtain the empirical conditional probabilities of tokens and labels, and use these quantities as a basis for “predicting words” in a new video clip. Since the normalization is done within the same HHMM model m , we omit the superscript m in the following sections when there is no confusion.

$$c(w|q^m) = \frac{C(q^m, w)}{\sum_w C(q^m, w)}, \quad c(q^m|w) = \frac{C(q^m, w)}{\sum_{q^m} C(q^m, w)} \quad (2)$$

3.3. Refining association with machine translation

In co-occurrence analysis, we associated *all* the labels with *all* the tokens present in the same story, this actually causes many entries $C(q, w)$ to receive more counts than they “deserve”. Take an ideal example, if label q_1 and token w_1, q_2 and w_2 always occur and only occur at exactly the same time, respectively; then for each story that contains both q_1 and q_2 , the entries $C(q_1, w_2)$ and $C(q_2, w_1)$ will receive one extra count. In other words, we would like to reduce the *smoothing* effects of token and label correlation resulting from the imprecise association within each story.

It turns out that we cannot just undo the correlation in both dimensions of C simultaneously, however the conditional co-occurrence $c(\cdot|\cdot)$ can be sharpened assuming independence of the variable being conditioned on. A mathematical model for this type of *un-smoothing* from co-occurrences has been explored in the machine translation (MT) literature [4], where the correspondence from English word e to a French word f are recovered from aligned sentences in both languages, by estimating the *translation* probabilities $t(f|e)$ of f given e , independent of the position of the words. In our context, it is appropriate to estimate both conditionals $t(w|q^m)$ and $t(q^m|w)$, as there is no obvious independence in either labels or tokens. Moreover, we do not want to loose the association to either labels or tokens as many entries in $t(\cdot|\cdot)$ diminishes to zero in the estimation process. The translation parameters are estimated with the EM algorithm, which we present from model 1 by Brown et. al. [4] for completeness:

$$\mathbf{E} : \quad \bar{t}(q|w) = \frac{t(q|w)}{\sum_q t(q|w)}, \quad \bar{t}(w|q) = \frac{t(w|q)}{\sum_w t(w|q)} \quad (3)$$

$$\mathbf{M} : \quad t(w|q) \leftarrow \frac{C(q, w)\bar{t}(q|w)}{\sum_w C(q, w)\bar{t}(q|w)}, \quad t(q|w) \leftarrow \frac{C(q, w)\bar{t}(w|q)}{\sum_q C(q, w)\bar{t}(w|q)}$$

4. EXPERIMENTS

In this section, we discuss results of predicting the correspondence using the co-occurrence statistic and refined probabilities. Our dataset are ABC and CNN news broadcasts taken from the TRECVID 2003 corpus, consisting of 22 half-hour clips from each channel. Each video comes with the audio-visual stream, the ASR words, and the ground-truth for story boundaries. We divide the data into four 11-broadcast sets from either station and rotate their roles as the training set and the test set. The cross-channel testing results are notably worse than that of the same channel, therefore we only report the latter due to space limit.

In the experiments, HHMMs were learned on one of the 11-video sets (the training set), generating $K = 10$ (manually set for traversing into considerable depth of the hierarchical mutual information clusters) different models using different subset of the 12 concepts. The resulting HHMMs typically have $5 \sim 10$ distinct states. Aside from one singular state that only corresponds to one shot in the corpus, most of the states take up an appropriate portion of the video sequence. The correspondence of the state labels in all models to the word-stems in the ASR transcript (in the training set) is then estimated according to Equations (2)(3) to produce conditional confidence values $c(w|q), c(q|w)$ and $t(w|q), t(q|w)$, respectively. These probabilities can be interpreted in two complementary contexts. One is *auto-annotation*, i.e., predicting words upon seeing an HHMM label, $c(w|q)$ is the *precision* value of this token-prediction process on the testing set by the counting processing in Equation (1); the other is *retrieval*, i.e., producing possible labels upon seeing a word, and $c(q|w)$ is *recall* of this label-retrieval process. It is easy to see, however, that the precision is biased towards frequent tokens or infrequent labels, while the recall tends to be large with infrequent tokens or frequent labels.

Hence we examine an alternative measure to trade-off this bias: the *likelihood ratio* L of the estimated conditional to the prior probability. i.e.,

$$L_w^c = c(w|q)/P(w), \quad L_q^c = c(q|w)/P(q);$$

$$L_w^t = t(w|q)/P(w), \quad L_q^t = t(q|w)/P(q);$$

Where $P(q)$ and $P(w)$ are the empirical fraction of stories that has label q or token w , respectively. Note $L_w^c = L_q^c$ from the normalizing relationship in Equation 2, and we denote both as L^c hereafter. This likelihood ratio measure is essentially the widely used *tf · idf* measure in text processing [8], since we are penalizing frequent *labels* or *tokens* by scaling with their *inverse story frequency*. Intuitively, $L \gg 1$ implies a strong association, and $L \rightarrow 0$ implies a strong exclusion, while values close to 1 implies no better than random guess.

We sort all the (*label, token*) pairs based on L^c, L_w^t and L_q^t , respectively, we then examine the *salient* pairs that lies in the top 5% of each L values. Table 1 shows details of an interesting label identified. This is the first label (among a total of seven) in the ninth model learnt over the concepts $\{\textit{outdoors, news-subject-face, building}\}$. The HHMM emission probabilities for this label shows low confidence for the concept *outdoors* and high confidence for the other two. The list of words intuitively hint the topic of politics and legal affairs. Further examining the actual stories that contain this label (42 out of a total 216), we find that these 42 stories cover all the Iraqi weapon inspection topic with 25.5% recall and 15.7% precision, and simultaneously contain the clinton-jones lawsuits with 44.3% recall and 14.3% precision at the shot-level. Another

$L^c L_w^t L_q^t$	token	$c(w q)$	$c(q w)$	$t(w q)$	$t(q w)$
●●●	murder	0.095	0.571	0.028	1.000
●●●	lewinski	0.238	0.556	0.074	0.999
●●●	congress	0.119	0.556	0.032	0.985
●○●	alleg	0.143	0.545	0.026	0.990
●●●	juri	0.167	0.500	0.055	1.000
●○●	judg	0.048	0.500	0.003	0.573
●○●	clinton	0.310	0.500	0.059	0.994
●●●	presid	0.452	0.475	0.179	0.999
●○●	polit	0.167	0.467	0.031	0.972
●●●	saddam	0.143	0.462	0.063	1.000
●●●	lawyer	0.143	0.462	0.039	0.999
●○●	independ	0.190	0.444	0.031	0.980
●○●	accus	0.095	0.444	0.000	0.010
●○●	monica	0.167	0.438	0.024	0.915
●○●	white	0.381	0.432	0.061	0.876
○●●	charg	0.190	0.381	0.055	0.998
○●●	investig	0.167	0.412	0.055	1.000
○●●	offic	0.190	0.364	0.006	0.581
○●●	public	0.143	0.300	0.046	0.901
○●●	secretari	0.190	0.364	0.019	0.689
○●●	washington	0.262	0.355	0.092	0.995

Table 1. Statistics of words associated with label $q^0 = 1$ before and after MT. Note “●” denotes that the corresponding L value lies in the overall top 5%, and “○” is for the complement. Using set A of channel ABC.

interesting label is $q^6 = 3$, indicating high confidence in both of its raw concepts $\{people, non-studio-setting\}$. With a word-list of $\{storm, rain, forecast, flood, coast, el, nino, administr, water, cost, weather, protect, starr, north, plane, northern, attorney, california, defens, feder, gulf\}$, it clearly indicates the topic *weather*. In fact, it covers the el-nino and storms that prevailed the US continent in the spring of 1998 with 80% and 78% recall on the training and testing set, respectively. Note this weather cluster is found without either the original “weather” concept or any dedicated weather sections in ABC News.

Figure 1 compares the likelihood ratio L^c of the HHMMs and the K-means clustering, the latter uses the feature set and cluster size chosen by the HHMM (Section 2.1- 2.2). From the graph we can see much more strong associations (bright peaks) and exclusions (dark valleys) in the labels obtained with HHMM than that of the K-means, and this shows that temporal modeling is indeed effective for the news domain. Look further into the K-means clusters, take the afore-mentioned $\{people, non-studio-setting\}$ for example, each of the six cluster labels is more spread out across all kinds of stories (appears in at least 2/3 of all stories), which makes its association with topic less distinctive.

It’s worth to note that: Words \neq meanings. While the words associated with a few labels are easy to decipher, most labels are associated with diverse words from which distinct topics are difficult to find. Natural language processing techniques such as latent semantic analysis can be employed to unveil the inherent structure in text (e.g., “white” and “house” often appear together) and to embody the ambiguity of going from words to semantics (e.g., “rise” can refer to the stock index, the temperature, a higher elevation, or even an angry political reaction). On the other hand, the concepts in the audio-visual stream may not be those present in the speech transcript, unlike the sentence-wise aligned bitext between two languages [4].

5. CONCLUSION

In this paper, we propose a method for discovering meaningful structures in news video through unsupervised learning of temporal clusters and associating them to words in the ASR transcript using co-occurrence analysis and machine translation techniques. We are able to find a few convincing translations between state la-

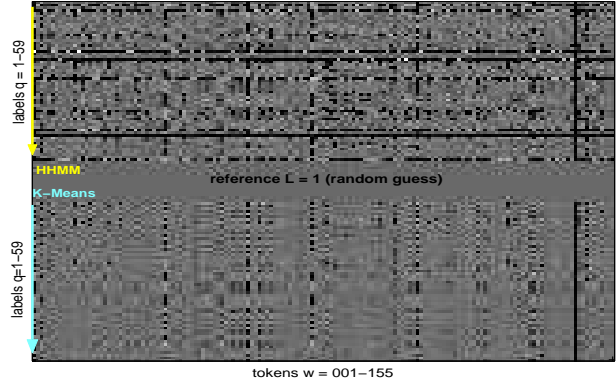


Fig. 1. Likelihood ratios $\log L^c(q, w)$ for HHMM (top) and K-means (bottom). Using set A of channel ABC.

bels and words. We have also observed that temporal models are indeed better in capturing the semantics than non-temporal clustering.

A few interesting issues remain: (1) Taking into account that the audio-visual concepts may or may not actually correspond to the words being said (such as in commercials); (2) Using text processing techniques to extract higher-level semantic concepts from raw textual tokens; (3) Jointly learning of the temporal model and the semantic association so as to get more meaningful labels.

ACKNOWLEDGEMENT

We would like to thank the TRECVID team [2] at IBM T. J. Watson Research Center for the concept detection results, and M. Naphade, A. Natsev, and B. Tseng for providing additional visual concept detection results. We are grateful to W. Hsu and D. Zhang for the tokenized speech transcript and helpful discussions.

6. REFERENCES

- [1] L. Xie, S.-F. Chang, A. Divakaran, and H. Sun, “Unsupervised mining of statistical temporal structures in video,” in *Video Mining*, Kluwer Academic Publishers, 2003.
- [2] A. Amir, G. Iyengar, C.-Y. Lin, C. Dorai, M. Naphade, A. Natsev, C. Neti, H. Nock, I. Sachdev, J. Smith, Y. Wu, B. Tseng, and D. Zhang, “The IBM semantic concept detection framework,” in *TREC Video Retrieval Evaluation Workshop*, 2003.
- [3] D. Marr, *Vision. A Computational Investigation into the Human Representation of Visual Information*. New York: Freeman, 1982.
- [4] P. F. Brown, S. A. D. Pietra, V. J. D. Pietra, and R. L. Mercer, “The mathematics of statistical machine translation: Parameter estimation,” *Computational Linguistics*, vol. 19, pp. 263–311, June 1993.
- [5] P. Duygulu, K. Barnard, N. de Freitas, and D. A. Forsyth, “Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary,” in *ECCV*, May–June 2002.
- [6] P. Duygulu and H. Wactlar, “Associating video frames with text,” in *Multimedia Information Retrieval Workshop, in conjunction with SIGIR 2003*, (Toronto, Canada), August 2003.
- [7] The National Institute of Standards and Technology (NIST), “TREC video retrieval evaluation,” 2001–2004. <http://www-nlpir.nist.gov/projects/trecvid/>.
- [8] D. Jurafsky and J. H. Martin, *Speech and Language Processing*. Prentice Hall, 2000.