Event Recognition in Photo Albums Using Probabilistic Graphical Model and Feature Relevance

Siham Bacha
University of Blida 1,
Department of Computer Science
Blida, Algeria
Email: *siham.bacha@gmail.com*

Mohand Said Allili
University of Quebec in Outaouais,
Department of Computer Science
Gatineau, Canada
Email: *mohandsaid.allili@uqo.ca*

Nadjia Benblidia
University of Blida 1
Department of Computer Science
Blida, Algeria
Email: *benblidia@gmail.com*

***Abstract*—The exponential use of digital cameras has raised a new problem: how to store/retrieve images/albums in very large photo databases that correspond to special events. In this paper, we propose a new probabilistic graphical model (PGM) to recognize events in photo albums stored by users. The PGM combines high-level image features consisting of scenes and objects detected in images. To consider the discriminative power of features, our model integrates the object/scene relevance for more precise prediction of semantic events in photo albums. Experimental results carried out on the challenging PEC dataset with 807 photo albums are presented.**

***Index Terms*—event recognition, feature relevance, probabilistic graphical models.**

I. Introduction

In recent years, digital cameras have become widespread and very accessible to users. Consequently, the fast-growing quantity of personal photos raises the need for efficient photo management and retrieval systems. Most of methods in the past dealing with this issue try to describe the visual image content at semantic level; thus attempting to bridge the so called semantic gap [22]. Several approaches have used different semantic cues such as faces [18], [6] and person identification [19] to facilitate management of personal photo collections. Furthermore, contextual information (e.g., date of acquisition, scene types) have also been exploited [13], [26].

In real-world scenarios, people often organize their collections according to events (e.g., *birthday* and *wedding*, etc.) [32]. Therefore, several approaches have been proposed for event recognition in personal photo collections based on visual features [11], [28]. In addition, contextual information (e.g., time-stamps, GPS, etc.) has been successfully incorporated to facilitate event recognition in photo albums [21], [29]. However, methods based solely on low-level image features lack semantically meaningful patterns, which makes them less discriminative for event recognition. Coming one step closer to describe images as perceived by humans, researchers have shifted their focus to process images at a finer level of granularity, including details of objects [24], [25] and scenes [8], in order to provide richer descriptive semantics. The combination of object and scene information has emerged, therefore, as a promising approach for further improving event recognition efficiency [17], [27].

In [17], a statistical model integrating scene and object information for event recognition on single images is proposed. Recently, [27] propose a promising avenue for event recognition in images by using Convolutional Neural Networks (CNNs) [14] for scene and object recognition. In this work, the output of CNNs for scenes and objects are combined for boosting event recognition. It remains, however, that these methods detect events only in single images which provide less rich and complete information about events compared to entire albums. On the one hand, photos constituting albums are usually taken at timestamps reflecting important moments of the events [5] and, therefore, can be more informative about the events. On the other hand, different photos can give a more exhaustive set of objects/scenes involved in the events.

In this paper, a probabilistic graphical model (PGM) combining high-level object and scene features for event recognition in personal photo collections is proposed. Given a photo album, CNNs are used to extract scene and object information from images, which represent the high-level features for event recognition. Moreover, our model incorporates scene/object relevance which allows to boost event prediction efficiency. For example, a 'Christmas tree' object can help to determine the event *Christmas* and the same applies for scene 'Mountain with snow' and event *skiing*. To infer event categories of new albums, we combine features of all album images and use maximum a posteriori probability (MAP) to estimate the event category. Contrary to previous work for event recognition in sets of images, our approach relies only on visual information. Experiments on a challenging dataset have demonstrated the performance of our approach by comparison with recent state-of-the-art methods.

The remainder of this paper is organized as follows. Section II presents the proposed model. Section III presents our model's parameter learning steps. Section IV presents how to infer events for new albums. Finally, Section V presents some experimental results for validation.

II. Proposed Model

A. Image representation

Images belonging to albums are related to events such as *birthday*, *wedding*, etc. The high-level visual features that can be exploited to recognize the events are objects and scenes, which represent semantically meaningful patterns. Recently, Convolutional Neural Networks (CNNs) have shown a great success in object detection and recognition in images [12], [16], [30] and scene categorization [31], even on most challenging datasets such as Places205 [31] and ImageNet [10]. Motivated by the promising results of the very deep CNN architecture, we adopt the GoogLeNet architecture [23] for both object detection and scene recognition.

B. Graphical model for event recognition

Our objective is to propose a model that associates an album of images with one of the event semantic classes. For this purpose, we use a probabilistic graphical model (PGM) combining high-level features to infer event categories for sets of photos. Fig. (1) shows a representation of the proposed model, where gray nodes represent observed variables in both training and testing phases. Nodes that have no shading are not observed in the testing phase. In the outer boxes, nodes with dotted lines are calculated variables. Finally, the green nodes and the rounded green boxes are the parameters and hyper-parameters of the model, respectively.

We consider the generative process of an album $\mathcal{A}$ of a particular event containing a set of N images denoted by $\mathcal{A} = \{I_1, .., I_N\}$, where I_i is the i-th image of the album. Let $e \in \{1, 2, ..., N_e\}$ be a discrete random variable representing the event category and N_e is the number of event categories. We suppose that the album $\mathcal{A}$ belongs to one event category e. An event category label e is chosen according to a Multinoulli distribution $e \sim p(e|\eta)$, where η is its N_e-dimensional parameter vector.

An event category e can exhibit multiple object and scene latent topics that represent semantic contexts in which the object and the scene may appear in the event, respectively. For example, in a *Round-trip* event, we can consider 'scholar road trip' and 'non-scholar road trip' object topics, and 'littoral road trip' and 'mountain road trip' as scene topics. Selecting a 'scholar road trip' object topic will privilege objects that occur frequently in this theme (e.g., scholar bus, etc.). To generate an image for a particular event category, we first generate scene and object topics from mixtures of available topics. Then, given the selected topics, we generate the object and scene instances of an image. First, we go through the generative process for each scene of $I_i \in \mathcal{A}$. Let s be a random variable representing a scene. An image can contain L_i scenes denoted by $\mathbf{s}_i = \{s_{il}\}$, where $l \in \{1, \ldots, L_i\}$, $L_i \in \{1, ..., L_{max}\}$ and s_{il} is the l-th scene of the i-th image of the album. For each scene s_{il}:

- Choose a scene topic t_{il}, where t_{il} is governed by a Multinoulli distribution $Mut(\psi^{(e)})$ with N_t-dimensional vector, where N_t is the number of topics in the scene latent space, and $t_{il}^{(v)} = 1$ indicates that the v-th topic is selected. ξ is the parameter vector of the Dirichlet prior distribution for $\psi^{(e)}$.
- Given the scene topic $v \in \{1, ..., N_t\}$, generate a scene instance s_{il} according to a Multinoulli distribution $Mut(\gamma^{(v)})$, with N_s-dimensional vector, where N_s is the number of scenes. δ is the parameter of the Dirichlet prior for $\gamma^{(v)}$.

Second, we go through the generative process of object instances in the album. The entity object o follows a Multinoulli distribution with N_o-dimensional vector, where N_o is the number of object classes. An image I_i can contain M_i objects denoted by $\mathbf{o}_i = \{o_{im}\}$, where $m \in \{1, \ldots, M_i\}$, $M_i \in \{1, ..., M_{max}\}$, and o_{im} is the mth object of the image I_i. For each object o_{im} of the album:

- Choose an object topic z_{im}, that indicates from which topic the object is generated, where z_{im} follows a Multinoulli distribution $Mut(\phi^{(e)})$ with N_z-dimensional parameter vector $\phi^{(e)}$. Let $z_{im}^{(u)} = 1$ indicates that the u-th topic is selected, and N_z is the number of topics in the object latent space. κ denotes the hyper-parameter of the Dirichlet prior distribution for $\phi^{(e)}$.
- Given the object topic $u \in \{1, ..., N_z\}$, the object o_{im} is generated according to a Multinoulli distribution $Mut(\pi^u)$ with N_s-dimensional parameter vector $\pi^{(u)}$. The vector $\pi^{(u)}$ has a Dirichlet prior with μ.

Finally, we introduce new variables $r^{(o,e)}$ and $r^{(s,e)}$ to encode the discrimination power of each scene and object, respectively, for the prediction of the event categories. Once all scenes and objects of the events are generated, we estimate their relevance $r^{(o,e)}$ and $r^{(s,e)}$ as follows:

- Let $r^{(o,e)}$ be binary discrete random variable, where $r^{(o,e)} = 1$ if the object o is relevant to the event category e and $r^{(o,e)} = 0$, otherwise. Let $p(r^{(o,e)} = 1|e, o) = \theta_{o,e}$ be a Bernoulli distribution $Ber(\theta_{o,e})$ with parameter $\theta_{o,e}$. The parameter $\theta_{o,e}$ has a Beta prior with hyper-parameters α and β.
- Let $r^{(s,e)}$ be binary discrete random variable, where $r^{(s,e)} = 1$ if the scene s is relevant to the event category e and $r^{(s,e)} = 0$, otherwise. Let $p(r^{(s,e)} = 1|e, s) = \omega_{s,e}$ be a Bernoulli distribution $Ber(\omega_{s,e})$ with parameter $\omega_{s,e}$. The parameter $\omega_{s,e}$ has a Beta prior with hyper-parameters α' and β'.

III. Learning parameters

A. Estimation of PGM parameters

In this section, we describe the learning of the model parameters $\{\phi, \pi, \psi, \gamma\}$ as depicted in Fig. (1). We assume that event categories are governed by a uniform distribution, where $p(e) = 1/N_e$. Furthermore, we suppose that objects and scenes are independent given the event. Thus, the object parameters $\{\phi, \pi\}$ and the scene parameters $\{\psi, \gamma\}$ can be learned separately. Without loss of generality, we present the learning steps for object parameters. Then, the same steps can be followed for learning scene parameters.

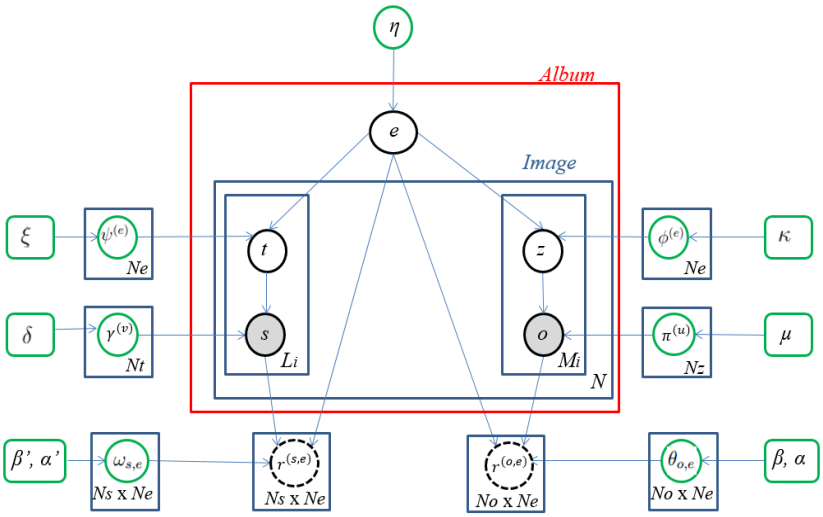

Fig. 1. Graphical representation of our model. Boxes denote replication of the corresponding random variables: there are N images in an album $\mathcal{A}$, with M_i observed objects and L_i observed scenes in image I_i. The variables e, o and s represent the event, object and scene classes, respectively. The latent variables for generating object and scene are z and t, respectively.

The N_z-dimensional parameter vector $\phi^{(e)}$ has Dirichlet distribution prior with the hyper-parameter κ. It governs the distribution of topics z_{im} given an event category e. The Bayesian estimation of the entries of the vector $\phi^{(e)}$ is given as follows:

$$\phi_u^{(e)} = p(z^{(u)} = 1|e) = \frac{n_{u,e} + \kappa_u}{\sum_{j=1}^{N_z} n_{j,e} + N_z \times \kappa_u}, \quad (1)$$

where $n_{u,e}$ is the number of occurrences of object topic $z^{(u)}$ in the event category e and κ_u is the u-th entry of κ.

The parameter $\pi^{(u)}$ governs the distribution of an object category o given that an object theme $z^{(u)} = 1$. We suppose a Dirichlet prior with hyper-parameter μ for the parameter vector $\pi^{(u)}$. The Bayesian estimation for $\pi^{(u)}$ is given by:

$$\pi_o^{(u)} = p(o|z^{(u)} = 1) = \frac{n'_{o,u} + \mu_u}{\sum_{k=1}^{N_o} n'_{k,u} + N_o \times \mu_u}, \quad (2)$$

where $n'_{q,u}$ is the number of occurrences of object category o given the topic $z^{(u)}$, and μ_u is the u-th entry of the vector μ. Finally, we choose a uniform prior for the distribution of hyper-parameters $\{\kappa, \mu, \xi, \delta\}$, respectively [20].

B. Estimation of relevance parameters

In this section, we estimate the relevance parameters $\theta_{o,e}$ and $\omega_{s,e}$. We recall that $p(r^{(o,e)} = 1|\theta_{o,e})$ is a Bernoulli distribution that represents an object relevance (resp. $p(r^{(s,e)} = 1|\omega_{s,e})$ for a scene relevance) for recognizing event e. Without loss of generality, we presents the steps for estimating object relevance parameters. The same steps can be followed for estimating scene relevance parameters. The Bernoulli distribution for the variable $r^{(o,e)}$ is formulated as follows:

$$p(r^{(o,e)}) = \theta_{o,e}^{r^{(o,e)}} (1 - \theta_{o,e})^{(1-r^{(o,e)})}. \quad (3)$$

To estimate $r^{(o,e)}$, we generate T samples from several albums. Let $\mathcal{D} = \{r_1^{(o,e)}, \ldots, r_T^{(o,e)}\}$ be T (calculated) observations of the variable $r^{(o,e)}$. Each observation is calculated from a given sample as: $r^{(o,e)} \approx 1 - \exp[-MI(o,e))]$, where $MI(o,e)$ is the mutual information between object o and event e. Finally, the maximum likelihood estimation for $\theta_{o,e}$ can be given as follows:

$$\hat{\theta}_{o,e} = \arg\max_{\{\theta_{o,e}\}} p(\mathcal{D}|\theta_{o,e}) \quad (4)$$

where $p(\mathcal{D}|\theta_{o,e})$ is the likelihood of the parameters given by the following factorization:

$$p(\mathcal{D}|\theta_{o,e}) = \theta_{o,e}^{N_1}(1 - \theta_{o,e})^{N_2}, \quad (5)$$

where $N_1 = \sum_i^T r_i^{(o,e)}$ and $N_2 = T - N_1$ gives the number of samples where the object is found relevant for the event category. We suppose a Beta distribution prior as a conjugate prior for $\theta_{o,e}$ with hype-parameter α and β. The MAP estimation for $\theta_{o,e}$ is then given as follows:

$$\begin{aligned} \hat{\theta}_{o,e} &= \arg\max_{\{\theta_{o,e}\}} (Beta(\alpha + N_1 - 1, \beta + N_2 - 1)) \\ &= \frac{\alpha + N_1 - 1}{\alpha + \beta + T - 2} \end{aligned} \quad (6)$$

IV. Inferring an event category for a new album

We consider the full graphical model of Fig. (1). Let $\Theta = \{\eta, \{\phi^{(e)}, \pi^{(z)}, \psi^{(e)}, \gamma^{(t)}\}, \omega, \theta\}$ denote the set of parameters used in our PGM, $\mathbf{r_o} = \{r^{(o,e)}|o = 1 : N_o, e = 1 : N_e\}$, and $\mathbf{r_s} = \{r^{(s,e)}|s = 1 : N_s, e = 1 : N_e\}$ are the estimated relevance for object and scene respectively. Given a new album $\mathcal{A}$, the goal is to infer its event category e. This is achieved by calculating the maximum likelihood function for the album given each event. We assume that the images in $\mathcal{A} = \{I_1, I_2, ..., I_N\}$ are independent and that the probability

of an event category depends only on the parameters the event class. Writing the album in term of images gives,

$$\begin{aligned} p(e|\mathcal{A}, \mathbf{r_s}, \mathbf{r_o}, \Theta) &\propto \prod_{i=1}^{N} p(e|I_i, r^{(s,e)}, r^{(o,e)}, \Theta) \\ &= \prod_{i=1}^{N} p(e|\mathbf{o}_i, \mathbf{s}_i, r^{(s,e)}, r^{(o,e)}, \Theta) \\ &= \prod_{i=1}^{N} \Bigg\{ \prod_{m=1}^{M_i} p(e|o_{im}, r^{(o,e)}, \Theta) \\ &\quad \prod_{l=1}^{L_i} p(e|s_{il}, r^{(s,e)}, \Theta) \Bigg\}, \end{aligned} \tag{7}$$

where the second equality comes from the assumption that building blocks of an image I_i are constituted of scenes and objects: $I_i = \{\mathbf{o}_i, \mathbf{s}_i\}$. The third equality comes by assuming that the objects and scenes of each image are detected separately in independent phases. The likelihood of the album images at the object level, which is expressed in the first term of Eq. (7), can be marginalized on the set of object topics as follows:

$$\begin{aligned} p(e|o_{im}, r^{(o,e)}, \Theta) &\propto p(r^{(o,e)}|o_{im}, \Theta) \\ &\quad \sum_{z_{im}} p(o_{im}|z_{im}, \Theta) p(z_{im}|e, \Theta) p(e|\eta) \end{aligned} \tag{8}$$

The expansion of individual terms in Eq. (8) depends on the full structure graphical of Fig. (1). Similarly to objects, the second term of Eq. (7) expresses the likelihood of the album images at the scene level. this term can be marginalized on the set of scene topics as follows:

$$\begin{aligned} p(e|s_{il}, r^{(s,e)}, \Theta) &\propto p(r^{(s,e)}|s_{il}, \Theta) \\ &\quad \sum_{t_{il}} p(s_{il}|t_{il}, \Theta) p(t_{il}|e, \Theta) p(e|\eta) \end{aligned} \tag{9}$$

Finally, we compute the maximum posterior probability (MAP) according to each event category $e \in \{1, 2, ..., N_e\}$ using the following maximization:

$$\hat{e} = \arg\max_{e} p(e|\mathcal{A}, \mathbf{r_s}, \mathbf{r_o}, \Theta) \tag{10}$$

where $\hat{e}$ represents the most likely event label that will be assigned to the album $\mathcal{A}$.

V. Experimental Results

A. Experimental Dataset

The experiments are carried out on Personal Event Collect (PEC) dataset, which is photo collections of provided by [5]. Namely, the PEC dataset consists of 807 collections, which are collected from Flicker with total of 61,000 images belonging to 14 event classes such as birthday, wedding and cruise, etc. We use the same experimental protocol suggested by the dataset creators. Particularly, we kept the 140 albums using for the test and selected randomly 84 albums for the training the model.

Methods \ Measure	Average accuracy	Average F_1 measure
Bossard *et al.* [5]	55.71%	56.16%
Wu et *al.* [28]	73.43%	57.68%
Our model	**74.28**%	**74.82**%

TABLE I
Comparison of our method with state-of-art methods using the PEC dataset.

B. Comparative experiments and discussion

We compare our proposed model against the most recent work in the literature provided by Wu et al. [28] and PEC dataset creators Bossard et al.[5]. The results of the comparison are shown in Table I. We recall that [28] have presented multiple results using different configurations of their method. We picked the best results obtained by their configurations.

In terms of classification accuracy, we can note that our method achieves an average of 74.29%, exceeding the best average accuracies obtained by [28] and [5] by 0.85% and 18.57%, respectively. More specifically, our method outperforms the others in the events *Birthday*, *Children's birthday*, *Easter*, *Graduation* and *Wedding*. For other event categories, we have achieved a close performance to the compared methods. In terms of the F_1 score, we have obtained an average score of 74.82%, exceeding the best F_1 scores obtained by [28] and [5] by 17.17% and 18.66%, respectively. This can be explained by the fact that the F_1 score is a compromise measure between the *precision* and *recall*, whereas the accuracy is calculated only from *recall* information. In other words, the obtained *precision* for our method is generally higher than those of the compared ones.

These results confirm, among other things, that the combination of scene and object cues and the integration of their relevance is a very suitable method for event recognition in photo albums. Not only our method outperforms recent state-of-the art methods, but feature relevance can be useful for more interpretable results which can be of prominent importance for applications such as album summarization and retrieval.

VI. Conclusion

We have introduced a probabilistic graphical model for album classification in social event categories. Our method takes advantage of recent developments on object/scene recognition in images to build high-level features for event recognition in photo albums. Moreover, we introduced a probabilistic feature relevance scheme that tunes the contribution of features according to their power of event discrimination. This allows to boost event recognition accuracy and obtain more interpretable results.

VII. Acknowledgements

This work has been achieved thanks to the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] M.S. Allili. Wavelet Modeling Using Finite Mixtures of Generalized Gaussian Distributions: Application to Texture Discrimination and Retrieval. *IEEE Trans. on Image Processing*, 21(4): 1452-1464, 2012.

[2] M.S. Allili, N. Baaziz and M. Mejri. Texture Modeling Using Contourlets and Finite Mixtures of Generalized Gaussian Distributions and Applications. *IEEE Trans. on Multimedia*, 16(3): 772-784 (2014)

[3] M.S. Allili, D. Ziou. Object of Interest Segmentation and Tracking Using Feature Selection and Active Contours. *IEEE Conf. on Computer Vision and Pattern Recognition*, 1-8, 2007.

[4] M.S. Allili, D. Ziou. Automatic Colour-Texture Image Segmentation using Active Contours. *Int'l J. of Computer Mathematics*, 84(9):1325-1338, 2007.

[5] L. Bossard, M. Guillaumin and L. Van Gool. Event Recognition in Photo Collections with a Stopwatch HMM. *IEEE Int'l Conference on Computer Vision*, 1193-1200, 2013.

[6] M. Brenner and E. Izquierdo. Joint People Recognition Across Photo Collections Using Sparse Markov Random Fields. *MultiMedia Modeling, Lecture Notes in Computer Science* 8325, 340-352. Springer, 2014.

[7] L. Cao, J. Luo, and T-S. Huang. Annotating Photo Collections by Label Propagation According to Multiple Similarity Cues. *ACM Int'l Conference on Multimedia*, 121-130, 2008.

[8] L. Cao, J. Luo, H. Kautz and T.S. Huang. Image Annotation Within the Context of Personal Photo Collections Using Hierarchical Event and Scene Models. *IEEE Trans. on Multimedia*, 11(2):208-219, 2009.

[9] M. Das and A.C. Loui. Event Classification in Personal Image Collections. *IEEE Int'l Conference on Multimedia and Expo*, 1660-1663, 2009.

[10] J. Deng, W. Dong, R. Socher, L-J. Li, K. Li and L. Fei-Fei. Imagenet: A Large-Scale Hierarchical Image Database. IEEE Conference on Computer Vision and Pattern Recognition, 248-255, 2009.

[11] L. Duan, D. Xu, S-F. Chang. Exploiting Web Images for Event Recognition in Consumer Videos: A Multiple Source Domain Adaptation Approach. *IEEE Conference on Computer Vision and Pattern Recognition*, 1338-1345, 2012.

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 38(1):142-158, 2016.

[13] S. Kisilevich, D. Keim, N. Andrienko and G. Andrienko. Towards Acquisition of Semantics of Places and Events by Multi-Perspective Analysis of Geo-Tagged Photo Collections. *Geospatial Visualisation, Lecture Notes in Geoinformation and Cartography*, 211-233. Springer, 2013.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25*, 11061114, 2012.

[15] G. Larivière and M.S.Allili. A Learning Probabilistic Approach for Object Segmentation. *IEEE Canadian Conf. on COmputer and Robot Vision*, 86-93, 2012.

[16] M. Liang and X. Hu. Recurrent Convolutional Neural Network for Object Recognition. *IEEE Conf. on Computer Vision and Pattern Recognition*, 3367-3375, 2015.

[17] L-J. Li and L. Fei-Fei. What, Where and Who? Classifying Events by Scene and Object Recognition. *IEEE Int'l Conference on Computer Vision*, 1-8, 2007.

[18] W.W.Y. Ng, T-M. Zheng, P.P.K. Chan and D.S. Yeung. Social Relationship Discovery and Face Annotation in Personal Photo Collection. *Int'l Conference on Machine Learning and Cybernetics*, 631-637, 2011.

[19] N. O'Hare and A.F. Smeaton. Context-Aware Person Identification in Personal Photo Collections. *IEEE Transactions on Multimedia*, 11(2):220-228, 2009.

[20] C. Robert. *The Bayesian Choice*. Springer, 2001.

[21] A. Salvador, M. Zeppelzauer, D. Manchon-Vizuete, A. Calafell and X.G. Nieto. Cultural Event Recognition with Visual Convnets and Temporal Models. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 36-44, 2015.

[22] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(12):1349-1380, 2000.

[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich. Going Deeper With Convolutions. CoRR, abs/1409.4842, 2014.

[24] S.-F. Tsai, T.S. Huang, and F. Tang. Album-Based Object-Centric Event Recognition. *IEEE Int'l Conference on Multimedia and Expo*, 1-6, 2011.

[25] S-F. Tsai, L. Cao, F. Tang and T-S. Huang. Compositional Object Pattern: A New Model for Album Event Recognition. *ACM Int'l Conference on Multimedia*, 1361-1364, 2011.

[26] A. Ulges, M. Worring, and T. Breuel. Learning visual contexts for image annotation from flickr groups. *IEEE Trans. on Multimedia*, 13(2):330-341, 2011.

[27] L. Wang, Z. Wang, W. Du, and Y. Qiao. Object-Scene Convolutional Neural Networks for Event Recognition in Images. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 30-35, 2015.

[28] Z. Wu, Y. Huang, and L. Wang. Learning Representative Deep Features for Image Set Analysis. *IEEE Trans. on Multimedia*, 17(11):1960-1968, 2015.

[29] J. Yuan, J. Luo and Y. Wu. Mining Compositional Features From GPS and Visual Cues for Event Recognition In Photo Collections. *IEEE Trans. on Multimedia*, 12(7):705-716, 2010.

[30] Y. Zhang, K. Sohn, R. Villegas, G. Pan and H. Lee. Improving Object Detection With Deep Convolutional Networks via Bayesian Optimization and Structured Prediction. *IEEE Conf. on Computer Vision and Pattern Recognition*, 249-258, 2015.

[31] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba and A. Oliva. Learning Deep Features for Scene Recognition Using Places Database. *Neural Information Processing Systems* (NIPS), 487-495, 2014.

[32] C. Zigkolis, S. Papadopoulos, G. Filippou, Y. Kompatsiaris and A. Vakali. Collaborative Event Annotation in Tagged Photo Collections. *Multimedia Tools and Applications*, 70(1):89-118, 2014.