Automatic Image Annotation using Possibilistic Clustering Algorithm

Mohamed Maher Ben Ismail, Sara N. Alfaraj, and Ouiem Bchir

Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

ljfis

Abstract

In this paper, the proposed PCMRM (possibilistic based cross-media relevance model) annotates images based on their visual contents. PCMRM framework relies on unsupervised learning to group the visually similar image regions into homogeneous clusters, along with the cross-media relevance model (CMRM) that is used to estimate the joint distribution of textual keywords and images. Besides, the unsupervised learning task exploits the robustness to noise of a possibilistic clustering algorithm, and generates membership degrees that represent the typicality of image regions with respect to the obtained clusters. To validate and assess the proposed system, we used the standard Corel dataset. PCMRM produced promising results. The reported performance measures proved that the proposed automatic image annotation approach outperforms similar state of the art solutions. This attainment is mainly attributed to the exploitation of the possibilistic membership produced by the clustering algorithm which allowed accurate learning of the association between annotating labels and the visual content of the image regions.

1. Introduction

Lately, the amounts of images including visual information have witnessed a continuous increase boosted by the spread of smart multimedia devices. Consequently, retrieval tools able to mine images upon user demand have become a compulsory need. Traditionally, image retrieval frameworks were designed with the assumption that images are associated with textual labels to make databases indexing and querying more useful and straightforward [1]. However, the amount of labor needed to manually annotate and/or correct the labels of large images' collections along with the subjectivity of the annotators affected the usability of such text-based retrieval solutions [2]. The Content-Based Image Retrieval (CBIR) systems [3] emerged as an alternative to relaxed the assumption that the image retrieval requires the association of labels with the stored images. In fact, typical CBIR systems use some low-level features such as color, texture, and shape to represent the images' content and estimate their pairwise similarity [4]. However, such CBIR systems failed bridge the semantic gap between the visual characteristics of the images and their high-level meanings [5].

Thus, the most accurate state-of-the-art image retrieval solutions rely on the assumption that the images are annotated/labeled. Therefore, automatic image annotation (AIA) that can be defined as the automatic assignment of textual keywords to images by a computer can be perceived as a natural solution to support the image retrieval systems. In other words, the

Received: Feb. 15, 2019
 Revised : Nov. 22, 2019
 Accepted: Dec. 1, 2019

Correspondence to: Mohamed Maher Ben Ismail
 (maher.benismail@gamil.com)
 ©The Korean Institute of Intelligent Systems

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Figure 1. Overview of a typical AIA system.

image retrieval gets more convenient if the images are automatically, accurately and objectively annotated. AIA aims to learn the associations between the images and text keywords. Specifically, it intends to assign textual labels to unannotated images in an unsupervised manner. Typically, machine learning techniques are used to learn the associations between the low-level visual features of the images and the high-level semantics of the user. Thus, accurate AIA yields better image management and retrieval using typical text based approaches. Figure 1 provides an overview of typical AIA systems.

Initially, the system is fed with a training set including labeled images. In general, each training image is then segmented into regions, and local features are extracted and used to encode the regions content. Next, the resulting feature vectors are categorized in an unsupervised manner into homogenous clusters. Then, a machine learning algorithm is used to learn the associations between the cluster representatives (blobs) and the keywords used to annotate the training images. For testing, the model built during the training phase is used to assign keywords to the image based on its visual properties.

Despite the proposal of promising CBIR systems, their performance remains constrained by the semantic gap between the visual characteristics of the images and their high-level meanings. Nevertheless, more accurate AIA systems would narrow this gap and yield better exploitation of the image databases.

In this research, we propose an AIA system which learns the associations between the image visual features and the textual labels (i.e., keywords) using machine-learning techniques. More precisely, we extend the cross-media relevance model (CMRM) [6] through the use of possibilistic membership degrees generated by a clustering algorithm to annotate the unlabeled images. In particular, the clustering algorithm generates membership degrees that represent the typicality of each data instance with respect to every cluster. Those memberships are

then used to estimate the probabilities of the CMRM. The low time complexity and the robustness to noise of the possibilistic clustering algorithm would improve the overall performance of the proposed AIA system.

2. Related Work

Recently, numerous researchers have attempted to bring more attention to bridging the semantic gap in image retrieval systems by proposing AIA systems. In [7], the researchers proposed an approach to annotate images based on grouping adjacent regions to solve the homogeneity problem. They grouped the image regions in order to have a compact object that could be annotated with semantic and appropriate keywords. The k-means algorithm was used to segment images, while texture and GIST descriptor [8] was used as low-level features to represent image content. Additionally, a Bayesian network was used as classifier to find and assign the appropriate keywords to the visual content. In [9], a quantum particle swarm optimization (IQPSO) was proposed. The AIA scheme bridges the gap between low-level features and high-level semantics by adopting the visual features' selection (VFS) [10]. To extract the visual features, a binary encoding was performed for each particle to simplify the image representation. Additionally, the population diversity measurements were proposed to maintain the population diversity and control conditions of the visual features selection process. The annotation approach used an ensemble stratagem based on boosting to fuse several classifiers' results—LDA [11], KNN [12], and SVM [13]—in order to improve the annotation performance. A diffusion-based image annotation framework that exploits images disseminated within online social networks was proposed in [14]. It assumed that the images in social webs are highly related to their annotations as well as to the user preferences. Therefore, a common-interest model was introduced

to analyze and learn social diffusion records to generate subgraphs. Then, the system extracted several features to represent the subgraphs and the corresponding annotations. Similarly, the hybrid probabilistic model (HPM) in [15] assumes that web images have a few existing semantic labels. By exploiting those labels, the system can automatically and collaboratively label a given image, and therefore perform ranking based on the label occurrence in the entire dataset. HPM integrates low-level features of the image and the high-level labels provided by the users to exploit the benefit of both content-based and collaborative approaches to predict the label under the probabilistic framework.

Other studies were devoted to learn statistical generative models to annotate and retrieve images. For instance, the automatic linguistic indexing of pictures real-time (ALIPR) [16] relies on statistical modelling and optimization to learn associations between the images and the keywords, and therefore to assign annotations for online images. ALIPR improves an algorithm. Namely, they adopted a D2-clustering algorithm to represent the objects using discrete distributions. In [17], the authors introduced a statistical generative model using multiple Bernoulli relevance models to predict keywords. Whereas researchers in [18] proposed a discriminative model called TagProp for annotation propagation. It combines a weighted nearest-neighbor approach with metric learning by directly maximizing the likelihood of the annotation predictions in the training set. The authors in [19] proposed a system that integrates two complementary of support vector machines (SVMs) for multiple instance learning (MIL) [20]. The optimum number of SVMs hyperplanes was obtained based on the extracted bag-features. The research in [21] focused on selecting the appropriate features to improve the annotation performance instead of modelling the keywords association. A robust group sparsity-based approach was introduced to cluster the extracted features. Different approaches constructed their annotation models using the nearest neighbors approach to constitute semantic neighborhoods. For instance, the 2-pass K-nearest neighbor model [22] relies on two steps of the k-nearest neighbor algorithm. The first step estimates the image-to-label similarities while the second one computes the image-to-image similarities. In [23], an AIA approach, named a NSIDML (neighborhood set based on image distance metric learning) that learns the distance metric in an unsupervised manner for a more accurate similarity calculation.

An image annotation approach based on a semi-supervised learning method was proposed in [24]. The proposed CGSSL (compact graph-based semi-supervised learning) builds a com-

pact local reconstruction graph with normalization to approximate a sample with its neighborhoods and to calculate the graph weights. More specifically, CGSSL aims to reconstruct the sample by using the neighbors of its adjacent samples and then preserve the graph weights corresponding to the minimum error. In this way, a more compact graph can be constructed to better capture the manifold structure of the dataset. Also, symmetrizing and normalizing this compact graph will establish the connection to the normalized graph. A co-occurrence approaches has considered the AIA problem as the problem of assigning probabilities to every word and image region, then annotating the image with words with the highest probability [23]. Based on this approach, a framework for automatic annotation of image regions based on segmentation using semantic analysis and discriminative classification is proposed in [25]. First, the researchers proposed a texture-enhanced JSEG algorithm [26] in order to segment the test image. These regions were then represented using a bag-of-words model to encode the image region content and study the semantic relationship between the visual words. Finally, the region labels were predicted using a maximal figure-of-merit algorithm [27]. These models were trained using image regions with multiple associations between regions and concepts. Motivated by the fact that multiple concepts that frequently co-occur over images form patterns that could provide contextual cues for proper concept inference, the researchers in [28] outlined an AIA approach that exploited the co-occurrence patterns as hierarchical communities in the pre-labelled training set. In addition, the model called automatic linguistic indexing of pictures (ALIP) [29] was used to compute the likelihood of the occurrence of images based on a describing stochastic process.

Particularly, AIA methods based on relevance models perform image annotation by maximizing the joint probability of images and words, which is calculated by the expectation over training images. In [30], a dual CMRM between image and keyword performed the image annotation task that involved two relations: word-to-image and word-to-word. These relations were estimated by training data and used search techniques on web data. Then, in [31], a cross-model approach to learn a representation that maximized the correlation between visual features and tags in a common semantic subspace is outlined. This approach relied on a learning procedure based on kernel canonical correlation analysis [32] which maps visual content to textual words by projecting them into a latent meaning space. Then, the learned mapping was used to annotate unlabeled images using advanced nearest-neighbor voting methods. In [33],

the researchers presented an AIA approach that integrated the fuzzy logic to the CMRMs. First, it performed identification of homogenous image regions, which share the same semantics using a fuzzy clustering algorithm, and then it learned the association between keywords and image regions using a membership-based CMRM. This work used the fuzzy logic as a knowledge representation of objects and scenes. Moreover, a novel inference-based algorithm was used to reduce the error propagation through the inference of scenes. In [34], the researchers proposed an AIA model based on fuzzy association rules and decision tree. First, the association rules that represent the correlations between image features and the high-level semantic concepts were mined. Then, the decision tree was added to discard the irrelevant rules [34]. In [35], an annotation framework based on the semantic analysis of images using fuzzy sets was introduced. It used a region-growing segmentation processes and it used a contextual representation approach combining fuzzy algebra with fuzzy theory to assign membership degrees of the labels. Most of these fuzzy clustering approaches are derived from the fuzzy C-means (FCM) algorithm [36] that considers the probabilistic constraint that the membership of a data point with respect to all clusters sums to 1. However,

FCM does not always correspond to the intuitive concept of the degree of belonging or compatibility. Thus, the performance of such FCM-based algorithms proved to be considerably affected in noisy environments [26]. To avoid this drawback, the possibilistic C-means (PCM) [37] algorithm was proposed as an alternative approach to make the FCM robust to noise and outliers by relaxing the constraint of the membership degree.

In summary, AIA aims to predict a set of semantic labels for a given image in an unsupervised manner. Several studies reported researchers' efforts which focus on system development rather than the feature extraction or machine learning tasks. Whereas the contributions in other studies are relevant to image processing and visual descriptors aspects. One should note that feature extraction and image segmentation, along with the learning task represent the main components of typical automatic annotation systems.

3. Proposed Approach

The proposed system consists of two main stages as depicted in Figure 2. The offline stage is intended to build the automatic annotation model, while the label assignment to the unannotated

Figure 2. Proposed AIA system architecture (PCMRM).

Figure 3. Example of two clusters dataset, with FCM membership degree.

images is held during the online stage.

The offline stage is fed with manually annotated images to conduct the model training. Initially, these images are partitioned using a simple segmentation algorithm to obtain image regions which inherit the image level labels. Then, the low-level feature extraction from the image regions is performed to encode the visual content of the images regions. PCMRM uses a possibilistic clustering algorithm to categorize all the obtained numeric vectors into homogenous clusters (i.e., blobs). The annotation model is built through the computation of the probabilities that a given word ω will appear in each blob b as well as the probability that each word ω occurs in given blob b . The online stage employs then this annotation model to assign annotations to the unlabeled images. Note that each test image is segmented to obtain its regions, and the regions visual features are used to assign them to the relevant blob. More specifically, each feature vector is assigned to the closest cluster to form a combination of blobs $\{b_1 \dots b_m\}$. Finally, the annotation model yields the computation of the probability of each word ω occurring in the blob b and returns the top N keywords with the highest probability.

3.1 Possibilistic Clustering

FCM algorithm [38] proved to be sensitive to noise and outliers because of the probabilistic constraint which forces the membership degrees of a given instance with respect to all clusters to sum to 1. More specifically, the memberships generated by the FCM algorithm fail to reflect the intuitive concept of the fuzzy degree of belonging for noise points. An illustrative example is given in Figure 3 to show how the fuzzy constraint of the FCM

algorithm breaks down with such outliers/noisy points.

Let C_1 and C_2 be two clusters, and $\{x_1, x_2, \dots, x_n\}$ be data points having membership degrees with respect to both clusters. As shown in Figure 3(a), the FCM generated different membership values, denoted by u for the points x_1 and x_2 with respect to C_1 , even though they are equidistant from the cluster centroid. This is due to the fuzzy constraint. Similarly, in Figure 3(b), the point x_2 and the point x_3 had equal membership values (≈ 0.5) assigned with respect to C_1 and C_2 , even though x_2 is more typical than point x_3 . Consequently, x_3 which can be a noise point was handled as any other typical point, and would have affected the overall clustering results. In other words, the membership degree of a point to a cluster is a relative number, and depends on the membership degree of the point with respect all other classes.

On the other hand, PCM [37] addresses the sensitivity of FCM to noise by relaxing the constraint to sum the membership degrees to 1. The objective function of the PCM algorithm is formulated as follows [37]:

$$J = \sum_{i=1}^k \sum_{j=1}^n u_{ij}^m d_{ij}^2(x_i, c_j) + \sum_{i=1}^k \eta_i \sum_{j=1}^n (1 - u_{ij}), \quad (1)$$

where $d_{ij}^2(x_j, c_i)$ represents the distance from the feature vector x_i to the cluster c_j , u_{ij} is the membership degree of x_i belonging to cluster j , η_i are positive numbers that relate to the overall size and shape of the cluster, m is the control number, and k, n refers to the number of clusters and data point, respectively.

The term $\sum_{i=1}^k \sum_{j=1}^n u_{ij}^m d_{ij}^2(x_i, c_j)$ in (1) is formulated to minimize

the sum of intra-cluster distances, while the term $\sum_{i=1}^k \eta_i \sum_{j=1}^n (1 - u_{ij})$ is intended to induce u_{ij} to be larger and avoid the case where all u_{ij} are null. The constraints on the possibilistic membership functions become

$$\begin{cases} u_{ij} \in [1.0] \quad \forall i, j, \\ 0 < \sum_{j=1}^n u_{ij} < n \quad \forall i. \end{cases} \quad (2)$$

Minimizing Eq. (1) with respect to $U = [u_{ij}]$, which is the $k \times n$ matrix of possibilistic membership values, leads to:

$$u_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^2}{\eta_i} \right)^{\frac{1}{m-1}}}. \quad (3)$$

The update equation of the centroids is:

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m \cdot x_j}{\sum_{j=1}^n u_{ij}^m}. \quad (4)$$

The update equation of η_i which determines the relative importance of the second term of in (1) becomes [37]:

$$\eta_i = \frac{\sum_{j=1}^n u_{ij}^m d_{ij}^2}{\sum_{j=1}^n u_{ij}^m}. \quad (5)$$

As illustrated in Figure 4, PCM assigns low possibilistic membership to noise points with respect to all clusters. In

Figure 4. Example of two clusters dataset, with PCM membership degree.

particular, Figure 4(a) shows that low possibilistic membership degrees with respect to both clusters were assigned to the point x_3 , while x_2 had larger membership degrees. This reflects a better typicality of the data points.

Substantially, we intend to exploit the robustness to noise of the possibilistic clustering algorithm to enhance the annotation accuracy. That is, if the instance is typical for a given cluster, it should be assigned a high membership degree regardless of other clusters, while noise instances will be detected easily because they exhibit low membership degrees with respect to all clusters.

3.2 Possibilistic Cross Media Relevance Model

PCMRM can be perceived as a machine translation model for AIA. Specifically, a segmentation algorithm is first deployed to partition images into object-shaped regions. Next, a visual vocabulary called “blobs” is learned by grouping visually similar image regions into homogeneous clusters. Namely, PCM algorithm [37] is used to discover those clusters. Finally, the relationship between blobs and keywords is estimated using a probabilistic model. In other words, PCMRM approximates the probability of observing a set of blobs and words in a given image.

Typically, relevance models [30] are used to estimate the joint distribution of words and images based on the training images set. The CMRM in [6] used the k -means [39] clustering algorithm to generate crisp membership values. Thus, for a given unannotated image I , the blob representation is $I = \{b_1 \dots b_m\}$ where m is the number of blobs in that image. The probability distribution of all possible words, related to the blobs in image I is calculated as:

$$P(\omega \mid I) \approx P(\omega \mid b_1 \dots b_m). \quad (6)$$

In fact, CMRM uses a training set T of annotated images J_n to estimate the probability of observing the word ω and blobs, the joint distribution can be computed as:

$$P(\omega, b_1 \dots b_m) = \sum_{J \in T} P(J) P(\omega, b_1 \dots b_m \mid J), \quad (7)$$

where $P(J)$ denotes the prior probabilities overall images in T . Assuming the independence of ω and $b_1 \dots b_m$, we get:

$$P(\omega, b_1 \dots b_m) = \sum_{J \in T} P(J) P(\omega \mid J) \prod_{i=1}^m P(b_i \mid J). \quad (8)$$

The annotations for unlabeled images is obtained by maximizing the expectation as follows [6]:

$$P = \max \left\{ \sum_{J \in T} P(J) P(\omega | J) P(b | J) \right\}, \quad (9)$$

where $P(\omega | J)$ refers to the likelihood of word ω given the training image J , $P(b | J)$ refers to the probability of b given J , and $P(J)$ is kept uniform for all training images.

Since each image $J \in T$ contains words and blobs, the maximum likelihood interpolated to estimate $P(\omega | J)$ is [6]:

$$P(\omega | J) = (1 - \alpha_J) \frac{\#(\omega.J)}{|J|} + \alpha_J \frac{\#(\omega.T)}{|T|}, \quad (10)$$

where $\#(\omega.J)$ denotes the occurrence of words in image J , which is either 1 or 0. Additionally, $\#(\omega.T)$ is the total numbers of words and blobs occurring in training set T . $|J|$ represents the overall count of all words and blobs occurring in image J , whereas $|T|$ represents the total size of the training set.

On the other hand, the probability of b given J is obtained using:

$$P(b | J) = (1 - \beta_J) \frac{f}{|J|} + \beta_J \frac{\#(b.T)}{|T|} \quad (11)$$

with

$$f = \sum_{r \in J} u_{rb}, \quad (12)$$

where u_{rb} is the possibilistic membership degree generated by PCM for the region r (in image J) with respect to the blob b , and $\#(b.T)$ represents the number of images' blobs in which the word appears. Besides α_J and β_J determine the degree of interpolation between the maximum likelihood estimates and the probabilities for the words and the blobs.

4. Experiments

In this research, the experiments were conducted using a subset of the Corel Stock Photo library, which is a standard annotated dataset and typical benchmark for keyword-based image retrieval systems [40]. The subset dataset (namely Corel 5k) consists of 5,000 images from 50 classes that originated from the library collection, with different classes including 100 images within a specific concept. The Corel 5k dataset consists of various image categories such as "nature scenes", "buses", "flowers", "dinosaur", etc. In addition, each image is associated

Figure 5. The frequencies of the first 100 keywords used as training images' annotation.

with a 3–5 manually textual keyword that depicts the primary objects appearing in the image. Overall there are 374 keywords in the dataset. The Corel 5k dataset includes a training set of 4,500 images and a test set of 500 images.

A grid segmentation of the images was conducted using a 3×3 kernel. This results in 9 regions per image. The obtained 40,500 regions were then clustered into blobs using possibilistic clustering. One should also note that the main scope of this research is the investigation of the possibilistic clustering and its effect to enhance the AIA. One should mention that the experiments have been deployed on a laptop with Mac OS Sierra, 2.9 GHz, Intel Core i7 processor, and 16 GB of RAM.

The vocabulary we used in our experiments exhibits high variation of the keyword frequency. Figure 5 displays the occurrences of the first 100 keywords used to annotate the training images. In fact, some keywords such as 'tree' and 'water' are highly frequent, and their frequencies exceed 700 times in the image collection. On the other hands, a keyword such as 'city' appears less than 30 times in the dataset. Similarly, some keywords are very rare. For instance, if we set a threshold to 30, only 216 words from the 374 original keywords vocabulary are considered as frequent.

In these experiments, we conducted an empirical optimization of the different parameters. In particular, during the training phase, we investigated various combinations of the smoothing parameters α and β in order to find the optimum values. Rather than using an exhaustive search, we naively set the first parameter to an initial value and vary the second one to seek local optima. Thus, the optimal value for the smoothing parameter α was 0.6. On the other hand the optimal smoothing parameter β was set to 0.9.

Table 1. Performance measures obtained using different number of clusters

Number of clusters	Precision (%)	Recall (%)	F-score (%)
100	25.34	69.74	37.17
300	25.11	68.90	36.80
600	24.79	68.35	36.38

Table 2. Sample performance measures obtained using different number labeling keywords

Number of labeling keywords	Precision (%)	Recall (%)	F-score (%)
3	35.69	64.53	45.96
5	24.82	69.44	36.57
7	18.44	70.02	29.20

Table 3. Sample performance measures obtained using different values of the fuzzyfier m

Fuzzyfier value	Precision (%)	Recall (%)	F-score (%)
1.1	21.06	69.70	32.35
1.3	25.34	69.74	37.17
1.5	24.82	69.44	36.57

Moreover, to determine the appropriate number of clusters needed to partition the image regions into homogeneous groups, we varied the number of clusters, and we compared the obtained performance measures for each value. As it can be seen in Table 1, setting the number of clusters to 100 yields the best performance in terms of precision, recall and f-score.

On the other hand, Table 2 shows the precision, recall and f-score obtained using different number of labelling keywords assigned to the test images. To maintain a trade-off between precision and recall along with an acceptable f-score, one can claim that using 5 keywords to annotate the test images is the best choice to assess the proposed PCMRM.

Similarly, in Table 3 we report sample performance measures obtained by varying the value of the fuzzyfier m used by the clustering algorithm. The value of the fuzzyfier m that yielded the best results corresponds to 1.3. Besides, the obtained results proved the robustness of PCM to the variation of m compared to FCM. In fact, the possibilistic membership reflects the “typicality” of each image region with respect to a given cluster and

does not take into consideration its membership value in the other clusters as expressed by the closed form in Eq. (3).

Table 4 shows sample labels automatically generated by the proposed method, as well as relevant state of the art methods, to annotate test images.

Specifically, the fuzzy CMRM attains a mean precision of 11.44%, while the typical CMRM yields 7.58%. On the other hand, the proposed PCMRM outperforms both methods and gives a precision of 25.34%. Similarly, for the mean recall, the proposed model overtakes the state-of-the-art solutions by achieving 69.74% versus 20.31% and 13.90% for the fuzzy CMRM [33] and the original CMRM [6], respectively. Moreover, for the f-score metric, the 37.17% attained by our method exceeds the attainment of both the fuzzy CMRM [33] and CMRM [6].

This performance can be attributed to the quality of the clustering process we adopted in our approach. In fact, the memberships associated with the image regions by PCM are independent of each other and do not have to sum to one. Our AIA approach exploited this valuable property of PCM. Specifically, low possibilistic memberships with respect to all clusters are assigned to outlier image regions. Note that the outliers and noise instances consist in the heterogeneous image regions which include parts from different objects. Thus, the PCM robustness to noise reduced the contribution of the noise regions to the learned CMRM probabilities, compared to the fuzzy CMRM and the original CMRM [6]. Moreover, if PCM [37] clustering is conducted using an initial over-specified number of clusters, the partitions/clusters will be constructed independently and some of them will converge to the same relatively dense regions of the feature space. This also represents a considerable advantage compared to FCM [36], and allows PCMRM to reduce the impact of the clustering quality on the AIA task.

On the other hand, the sensitivity to the smoothing parameters β is reduced with the use of possibilistic memberships compared to the fuzzy CMRM. In fact, β determines the contribution of the regions of a given test image to estimate the underlying image model. In other words, it defines the interpolation level between the background probability and the maximum likelihood of the image regions. In particular, larger β as in our experiments, gets the model biased towards the regions having the highest possibilistic memberships.

Further analysis of the obtained results showed that many test images were annotated using the most frequent keywords. Namely, the words “water”, “sky” and “tree” were wrongly assigned to some test images. If we give up those keywords when

Table 4. Sample keywords automatically generated using the original CMRM [6], the fuzzy CMRM [33] and the proposed method

Test images	Ground truth	CMRM [6]	Fuzzy CMRM [33]	Proposed PCMRM
"Cars, tracks, grass"	"Water, tree, sky, people, grass"	"Cars, tracks, turn, prototype, water"	"Cars, tracks, snow, grass, water"
"Sky, tree, castle"	"People, building, oahu, water, tree"	"People, waves, oahu, water, tree"	"Sky, mountain, tree, building, grass"
"Horser, tree, grass"	"Water, tree, sky, people, snow"	"Tree, horses, foals, mare, garden"	"Tree, horses, bear, grass, garden"
"Flowers, petals, leaf"	"Sky, water, people, tree, grass"	"Flowers, petals, water, grass, tree"	"Flowers, petals, grass, snow, leaf"
"Sunset, tree, mountain"	"Sky, flowers, water, grass, tree"	"People, sunset, sky, water, beach"	"Water, sunset, beach, mountain, sky"
"Flowers, tree, sky"	"Flowers, tree, grass, lawn, sky"	"Flowers, tree, grass, lawn, sky"	"Sky, flowers, tree, grass, mountain"
"Sky, plane, runway"	"Plane, jet, sky, car, tracks"	"Plane, jet, bird, sky, tracks"	"Plane, jet, sky, runway, tracks"

calculating the performance measures of the proposed system, the precision, recall, and f-score reaches 26.59%, 71.11%, and 38.70%, respectively. This proves that the value of the smoothing parameter α we used in our experiments did not cause a considerable bias towards the most frequent words.

In order to assess how PCMRM is sensitive to the low-level features used to encode the visual content of the image regions, we used the 5-diemsional edge histogram descriptor (EHD) as a texture feature along with the color moments to encode the content of the image regions. After concatenating both

low-level features (color and texture), each image region was represented using a 20-dimensional feature vector. This resulted in 2.75% as average improvement for the precision, recall, and f-score. Specifically, 30.03%, 72.53%, and 42.47% were attained as precision, recall, and f-score, respectively. This demonstrates that more visual descriptors would yield better representation of the image region content and better learning of the association between the labeling keywords and visual properties of the image regions.

In Figure 6, we show the average accuracy of some labeling keyword. For the display purpose, we selected these keywords from the vocabulary to illustrate the comparison. Basically, it is the number of occurrences of a given keyword in the top five labels automatically assigned to the test images. For the most frequent keywords in this collection such as ‘sky’, ‘grass’, ‘water’ and ‘tree’, all methods give comparable acceptable accuracy. On the other hand, the proposed approach (blue curve) and the fuzzy based CMRM [33] (red curve) overtake the original CMRM in [6]. The resulting average accuracies are 34%, 23%, and 14% for the proposed method, the fuzzy CMRM in [33], and the original CMRM in [6], respectively. These results confirm the precision, recall, and f-score attainments reported above.

5. Conclusion

In this paper, we proposed an AIA framework based on a possibilistic unsupervised learning algorithm to partition the image regions into homogeneous clusters. Moreover, it relies on an extension of the CMRM which uses the possibilistic membership degrees generated by the clustering algorithm to estimate the joint distribution of the textual keywords and the images. To validate and assess the proposed PCMRM, we used the standard Corel dataset. The reported performance measures proved that the proposed AIA approach outperforms similar state of the art solutions. This attainment is mainly attributed to the exploitation of the possibilistic membership produced by the clustering algorithm which allowed accurate learning of the association between annotating labels and the visual content of the image regions.

Despite the promising results the performance of proposed approach can be increased through the following future works:

- The use of a more advanced image segmentation technique to partition the images.
- The fusion of more low-level features to represent the image regions in an efficient manner and avoid the curse of dimension-

Figure 6. Comparison of the accuracy of the most frequent words using different annotation methods.

ality.

- The introduction of some supervision information to guide the clustering algorithm and ensure its convergence to the optimal solution.

- The use of kernel based distance metric rather than the Euclidean distance which suitable only if the clusters exhibit spherical shape.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgements

The authors are grateful for the support by the Research Center of the College of Computer and Information Sciences, King Saud University.

References

- [1] Y. Rui, T. S. Huang, and S. F. Chang, "Image retrieval: current techniques, promising directions, and open issues," *Journal of Visual Communication and Image Representation*, vol. 10, no. 1, pp. 39-62, 1999. <https://doi.org/10.1006/jvci.1999.0413>
- [2] Y. Liu, D. Zhang, G. Lu, and W. Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, no. 1, pp. 262-282, 2007. <https://doi.org/10.1016/j.patcog.2006.04.045>
- [3] V. N. Gudivada and V. V. Raghavan, "Content based image retrieval systems," *Computer*, vol. 28, no. 9, pp. 18-22, 1995. <https://doi.org/10.1109/2.410145>
- [4] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognition*, vol. 45, no. 1, pp. 346-362, 2012. <https://doi.org/10.1016/j.patcog.2011.05.013>
- [5] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 22, no. 12, pp. 1349-1380, 2000. <https://doi.org/10.1109/34.895972>
- [6] J. Jeon, V. Lavrenko, and R. Manmatha, "Automatic image annotation and retrieval using cross-media relevance models," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, Toronto, Canada, 2003, pp. 119-126. <https://doi.org/10.1145/860435.860459>
- [7] M. Oujaoura, R. El Ayachi, B. Minaoui, M. Fakir, and O. Bencharef, "Grouping K-means adjacent regions for semantic image annotation using Bayesian networks," in *Proceedings of 2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*, Beni Mellal, Morocco, 2016, pp. 243-248. <https://doi.org/10.1109/CGiV.2016.54>
- [8] M. Douze, H. Jegou, H. Sandhawalia, L. Amsaleg, and C. Schmid, "Evaluation of gist descriptors for web-scale image search," in *Proceedings of the ACM International Conference on Image and Video Retrieval*, Santorini, Greece 2009. <https://doi.org/10.1145/1646396.1646421>
- [9] C. Jin and S. W. Jin, "Automatic image annotation using feature selection based on improving quantum particle swarm optimization," *Signal Processing*, vol. 109, pp. 172-181, 2015. <https://doi.org/10.1016/j.sigpro.2014.10.031>
- [10] M. Wang, X. S. Hua, J. Tang, and R. Hong, "Beyond distance measurement: constructing neighborhood similarity for video annotation," *IEEE Transactions on Multimedia*, vol. 11, no. 3, pp. 465-476, 2009. <https://doi.org/10.1109/TMM.2009.2012919>
- [11] S. Balakrishnama and A. Ganapathiraju, "Linear discriminant analysis-a brief tutorial," *Institute for Signal and Information Processing*, vol. 18, pp. 1-8, 1998.
- [12] M. Y. Su, "Using clustering to improve the KNN-based classifiers for online anomaly network traffic identification," *Journal of Network and Computer Applications*, vol. 34, no. 2, pp. 722-730, 2011. <https://doi.org/10.1016/j.jnca.2010.10.009>
- [13] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293-300, 1999. <https://doi.org/10.1023/A:1018628609742>
- [14] C. Lei, D. Liu, and W. Li, "Social diffusion analysis with common-interest model for image annotation," *IEEE Transactions on Multimedia*, vol. 18, no. 4, pp. 687-701, 2015. <https://doi.org/10.1109/TMM.2015.2477277>

- [15] N. Zhou, W. K. Cheung, G. Qiu, and X. Xue, "A hybrid probabilistic model for unified collaborative and content-based image tagging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 7, pp. 1281-1294, 2011. <https://doi.org/10.1109/TPAMI.2010.204>
- [16] J. Li and J. Z. Wang, "Real-time computerized annotation of pictures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 985-1002, 2008. <https://doi.org/10.1109/TPAMI.2007.70847>
- [17] S. L. Feng, R. Manmatha, and V. Lavrenko, "Multiple Bernoulli relevance models for image and video annotation," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, 2004, pp. 1002-1009. <https://doi.org/10.1109/CVPR.2004.1315274>
- [18] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "Tagprop: discriminative metric learning in nearest neighbor models for image auto-annotation," in *Proceedings of 2009 IEEE 12th International Conference on Computer Vision*, Kyoto, Japan, 2009, pp. 309-316. <https://doi.org/10.1109/ICCV.2009.5459266>
- [19] X. Qi and Y. Han, "Incorporating multiple SVMs for automatic image annotation," *Pattern Recognition*, vol. 40, no. 2, pp. 728-741, 2007. <https://doi.org/10.1016/j.patcog.2006.04.042>
- [20] S. Andrews, I. Tsochantaridis, and T. Hofmann, "Support vector machines for multiple-instance learning," *Advances in Neural Information Processing Systems*, vol. 16, pp. 577-584, 2003.
- [21] S. Zhang, J. Huang, Y. Huang, Y. Yu, H. Li, and D. N. Metaxas, "Automatic image annotation using group sparsity," in *Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, 2010, pp. 3312-3319. <https://doi.org/10.1109/CVPR.2010.5540036>
- [22] Y. Verma and C. V. Jawahar, "Image annotation by propagating labels from semantic neighbourhoods," *International Journal of Computer Vision*, vol. 121, no. 1, pp. 126-148, 2017. <https://doi.org/10.1007/s11263-016-0927-0>
- [23] M. Zhao, T. W. Chow, Z. Zhang, and B. Li, "Automatic image annotation via compact graph based semi-supervised learning," *Knowledge-Based Systems*, vol. 76, pp. 148-165, 2015. <https://doi.org/10.1016/j.knosys.2014.12.014>
- [24] Y. Mori, H. Takahashi, and R. Oka, "Image-to-word transformation based on dividing and vector quantizing images with words," in *Proceedings of the 1st International Workshop on Multimedia Intelligent Storage and Retrieval Management*, Orlando, FL, 1999, pp. 1-9.
- [25] J. Zhang, Y. Gao, S. Feng, Y. Yuan, and C. H. Lee, "Automatic image region annotation through segmentation based visual semantic analysis and discriminative classification," in *Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016, pp. 1956-1960. <https://doi.org/10.1109/ICASSP.2016.7472018>
- [26] Y. G. Wang, J. Yang, and Y. C. Chang, "Color-texture image segmentation by integrating directional operators into JSEG method," *Pattern Recognition Letters*, vol. 27, no. 16, pp. 1983-1990, 2006. <https://doi.org/10.1016/j.patrec.2006.05.010>
- [27] C. Jin and S. W. Jin, "Image distance metric learning based on neighborhood sets for automatic image annotation," *Journal of Visual Communication and Image Representation*, vol. 34, pp. 167-175, 2016. <https://doi.org/10.1016/j.jvcir.2015.10.017>
- [28] S. Gao, W. Wu, C. H. Lee, and T. S. Chua, "A maximal figure-of-merit learning approach to text categorization," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, Toronto, Canada, 2003, pp. 174-181. <https://doi.org/10.1145/860435.860469>
- [29] L. Feng and B. Bhanu, "Semantic concept co-occurrence patterns for image annotation and retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 4, pp. 785-799, 2016. <https://doi.org/10.1109/TPAMI.2015.2469281>
- [30] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075-1088, 2003.
- [31] J. Liu, B. Wang, M. Li, Z. Li, W. Ma, H. Lu, and S. Ma, "Dual cross-media relevance model for image annotation," in *Proceedings of the 15th ACM International Conference*

on *Multimedia*, Augsburg, Germany, 2007, pp. 605-614. <https://doi.org/10.1145/1291233.1291380>

- [32] L. Ballan, T. Uricchio, L. Seidenari, and A. Del Bimbo, "A cross-media model for automatic image annotation," in *Proceedings of International Conference on Multimedia Retrieval*, Glasgow, UK, 2014. <https://doi.org/10.1145/2578726.2578728>
- [33] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: an overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639-2664, 2004. <https://doi.org/10.1162/0899766042321814>
- [34] M. Alkaoud, I. AshShohail, and M. M. B. Ismail, "Automatic image annotation using fuzzy cross-media relevance models," *Journal of Image and Graphics*, vol. 2, no. 1, pp. 59-63, 2014.
- [35] Z. Li, L. Li, K. Yan, and C. Zhang, "Automatic image annotation based on fuzzy association rule and decision tree," in *Proceedings of the 7th International Conference on Internet Multimedia Computing and Service*, Zhangjiajie, China, 2015. <https://doi.org/10.1145/2808492.2808504>
- [36] T. Athanasiadis, P. Mylonas, Y. Avrithis, and S. Kollias, "Semantic image segmentation and object labeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 3, pp. 298-312, 2007. <https://doi.org/10.1109/TCSVT.2007.890636>
- [37] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191-203, 1984. [https://doi.org/10.1016/0098-3004\(84\)90020-7](https://doi.org/10.1016/0098-3004(84)90020-7)
- [38] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98-110, 1993. <https://doi.org/10.1109/91.227387>
- [39] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, NY: Springer Science & Business Media, 2013.
- [40] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: a k-means clustering algorithm," *Journal of the Royal Statistical Society Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100-108, 1979. <https://doi.org/10.2307/2346830>

[41] The COREL database for content based image retrieval, Available <https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval>

[42] C. Jin and S. W. Jin, "Image distance metric learning based on neighborhood sets for automatic image annotation," *Journal of Visual Communication and Image Representation*, vol. 34, pp. 167-175, 2016. <https://doi.org/10.1016/j.jvcir.2015.10.017>

Mohamed Maher Ben Ismail is an associate professor at the computer science department of the College of Computer and Information Sciences at King Saud University. He received his Ph.D. degree in Computer Science from the University of Louisville in 2011. His research interests include pattern recognition, machine learning, data mining and image processing.

E-mail: maher.benismail@gmail.com

Sara N. Alfaraj got her master's degree in computer science from King Saud University, Riyadh, Saudi Arabia. The photo is not included according to the author's request.

E-mail: lathamax@gmail.com

Ouïem Bchir is an associate professor at the Computer Science Department, College of Computer and Information Sciences (CCIS), King Saud University, Riyadh, Saudi Arabia. She obtained her Ph.D. from the University of Louisville, KY, USA. Her research interests are Spectral and kernel clustering, pattern recognition, hyperspectral image analysis, local distance measure learning, and Unsupervised and Semi-supervised machine learning techniques. She received the University of Louisville Dean's Citation, the University of Louisville CSE Doctoral Award, and the Tunisian presidential award for the electrical engineering diploma.

E-mail: deenaiee@yahoo.com