Automatic Image Annotation Based on Low-Level Features and Classification of the Statistical Classes*

Andrey Bronevich and Alexandra Melnichenko

Mathematics Department, Technological Institute of Southern Federal University,
Taganrog, Nekrasovskij Street 44, Taganrog 347928, Russia
brone@mail.ru, alexandramelnichenko@gmail.com

Abstract. This work is devoted to the problem of automatic image annotation. This problem consists in assigning words of a natural language to an arbitrary image by analyzing textural characteristics (low-level features) of images without any other additional information. It can help to extract intellectual information from images and to organize searching procedures in a huge image base according to a textual query. We propose the general annotation scheme based on the statistical classes and their classification. This scheme consists in the following. First we derive the low-level features of images that can be presented by histograms. After that we represent these histograms by statistical classes and compute secondary features based on introduced inclusion measures of statistical classes. The automatic annotation is produced by aggregating secondary features using linear decision functions.

Keywords: automatic image annotation, image retrieval, low-level features, statistical classes, inclusion measures.

1 Introduction

Nowadays there are many visual data bases accessible in Internet, but it is hard to find such information because the usual textual query cannot be processed properly because, in general, images do not have additional description, that can be matched with textual representation. To solve this problem, it is necessary to provide effective methods for automatic image annotation enabling to obtain image annotations consisting of words and describing the image content.

There are many methods proposed for this problem which differ in some aspects such as image features representation and type of classifier used [1,2]. In the recent investigations two main features representations are used equally often: *regional representation*, when each image region is described separately with own feature vector, and *global representation*, where whole image is described by the one vector of low-level features. While regional representation could be more

* This work is supported by RFBR, projects #10-07-00135, #10-07-00478, #11-07-00591.

S.O. Kuznetsov et al. (Eds.): RSFDGrC 2011, LNAI 6743, pp. 314–321, 2011.

discriminative, the accurate image segmentation, required for it, is hard problem. As regards annotation model, in the early years various probabilistic methods dominate here (such as Bayesian classifiers and particularly Cross-Media Relevance Models [3,4])along with some machine learning algorithms [2,5]).Recent years Theory of Rough Sets proposes promising approaches for building classifiers and dicision rules for solving this sort of classification tasks [13].

The automatic image annotation can be considered as a classification problem, in which we should choose words from a given vocabulary that describe the image relevantly. In this investigation we don't consider the semantic textual descriptions and our output result should be the list of words ordered by their significance or relevance to a given image. Let us notice that the annotations based on textural features of images can reflect general image characteristics, that represented by words like "day", "night", "sea", "tree", "city", that describe the image in the whole. And we apply for this low-level features based on evaluation of gradient, colors distribution and various textural characteristics. As a rule, low-level image features can be some numerical characteristics or samples that are often represented by histograms. This way of representation can be used for describing colors distributions, the distribution of gradient directions and so on. These characteristics should be stable to scene illumination and scaling. Because it is hard to find an explicit connection between low-level features and words, it is reliable to use methods from pattern recognition theory: according to the problem statement we have a learning sample of annotated images and we have to build decision functions allowing us to classify an arbitrary image using words of a given vocabulary. The main characteristics of this pattern recognition problem are the following:

1. A huge number of classes to which a given image can belong (the number of classes is equal to the cardinality of a chosen vocabulary).
2. Classes are not disjoint in general.
3. It is impossible precisely to define boundaries between classes.
4. As a rule, low-level features can be represented as independent samples of a random variable that characterizes the image.
5. A very high dimension of feature space in which the classification problem should be solved.

These characteristics of the classification problem can be easily derived by analyzing its nature. For example, a city landscape can include houses, trees, and some times a part of the picture can include sea outlook. If a picture contains a palm, it is not possible to judge whether the photo was made within the house or outside. If we classify images using words "morning", "day", "evening", and "night", then it is hard to define exact boundaries between classes "day" and "evening", "night" and "morning". Trying to increase classification quality, we should increase the number of the used low-level features, and this also forced the increasing of the feature space dimension. These characteristics lead to the following classification scheme, depicted on Fig. 1.

According to this scheme, images should be processed first for extracting low-level features, then the secondary features are derived, and the annotation is

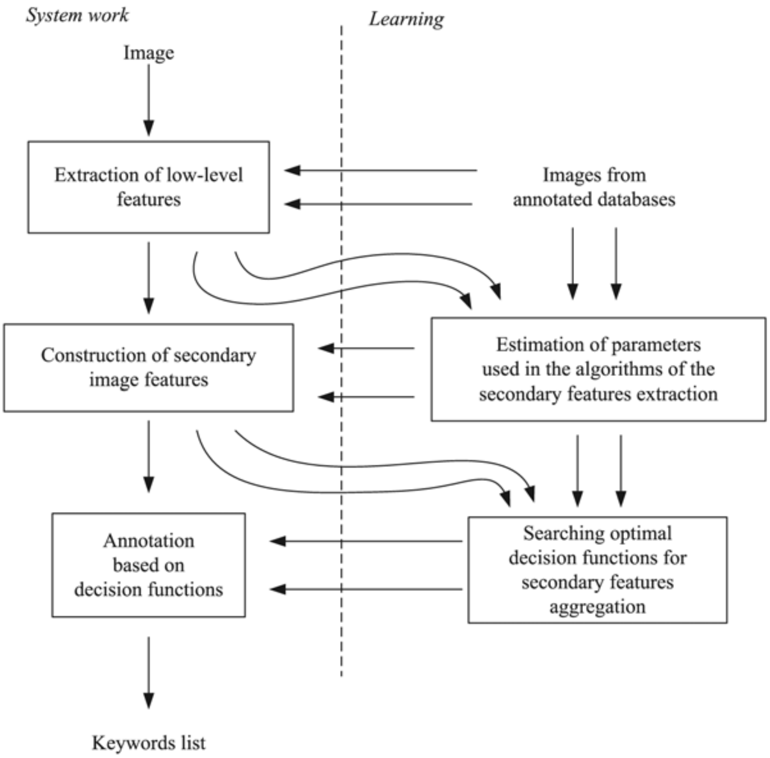
System work
Learning
Image
Extraction of low-level features
Images from annotated databases
Construction of secondary image features
Estimation of parameters used in the algorithms of the secondary features extraction
Annotation based on decision functions
Searching optimal decision functions for secondary features aggregation
Keywords list

Fig. 1. General scheme of the automatic image annotation procedure

constructed by aggregating secondary features. Sometimes the secondary features extraction can be understood as a problem of decreasing feature space dimension.

2 Low-Level Image Features

Low-level feature extraction is a very important stage of automatic annotation algorithm because it provides the base for building the image representation. Low-level features producing image annotation should meet humans perception of image similarity and satisfy certain conditions allowing to consider these features as patterns for classification. The main of such requirements are:

- invariance with respect to image scaling and to lighting conditions of image capturing;
- small correlation of different features;
- the dimension of obtained patterns should be adequate to the size of keywords used.

We briefly describe here four low-level feature types which are the most appropriate in terms of required properties. We have successfully used these features for the construction of automatic image annotation system.

Histogram of Oriented Gradient (HoG). allows to determine appearance of the local objects and to recognize their shape [5]. HoG-descriptors calculation is performed using rectangular grid. The grid consists of cells small spatial regions which are combined to the larger intersected areas called "blocks". For an image the matrices of gradient magnitudes in horizontal and vertical directions are calculated for each color channel in RGB color space. As the orientation and magnitude for each pixel the corresponded values are taken from the color channel with the maximal magnitude. For each cell one calculates histograms of gradient orientations and joins histograms which compose a block. Histograms belonging to the one block are normalized to achieve invariance to the local illumination changes.

Measure of the background homogeneity. can be useful, for example, to distinguish such scene types as landscape, portrait or macro [6]. Image background is relatively large areas of connected pixels with the similar color characteristics. We calculate the measure of background homogeneity using Shannon entropy of every color channel as described in [6]. To obtain more informative image representation, we calculate the measure of background homogeneity for particular rectangles from some regular grid.

Color histograms. evaluate an image property that is very important for human visual perception — the color distribution. The key issue here is the choice of the appropriate color space. We use for this purpose *CIELab* color space, which has been designed as the space with linear color changes with respect to the human perception. To achieve illumination invariance we discard L ("lightness") component of the (L, a, b) pixel component and build two-dimensional histogram from two remaining chrominance color components.

Texture image features. Texture determines surfaces features which are helpful for objects recognition. One of the most informative texture features is one introduced in [7], which includes such characteristics as coarseness, contrast and directionality of the texture. The coarseness characterizes the size of the structural units forming the texture. The texture contrast value indicates how much gray levels vary within an image and in what degree their distribution is biased to white or black. Feature of texture directionality is calculated in the each pixel from the magnitudes of histogram peaks. For detailed description of calculation procedure see [7].

3 Statistical Classes Classification

3.1 Notion of Statistical Class

Here we use the **notion of statistical class** [8], that is introduced for the finite case as follows. Let $X = \{x_1, ..., x_n\}$ be a finite universal set and let $\mathcal{U} = 2^X$ be the powerset of X. Assume also that the space X is equipped with an additive measure V, called the volume measure. Then any statistical class F can be defined by a probability measure P on $\mathcal{U}$, that has to be absolutely continuous w.r.t. the volume measure V. Because we assume that all low-level features can be described by histograms, we postulate that any such feature is a histogram, which can be considered as an evaluation of a probability distribution. Secondary features are computed by using inclusion measures of statistical classes. Let us remind that the absolute continuity for a finite case means that $x_i \in X$ and $P(\{x_i\}) > 0$ implies that $V(\{x_i\}) > 0$. Therefore, it is possible to define a probability density by formula

$$h(x) = \begin{cases} P(\{x\})/V(\{x\}) & \text{if V}(\{\text{x}\}) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Using the probability density we can compute the probability of any event $A \in \mathcal{U}$ by $P(A) = \sum\limits_{x \in A} h(x)V(\{x\})$. The last sum can be considered as an integral sum for Lebesgue integral, therefore, we can write: $P(A) = \int\limits_A h(x)dV$. It is clear that the density function can be considered as another way for defining the statistical class. In real applications, the volume measure has to be chosen such that it can discriminate statistical classes in the best way. If we have no sufficient prior information, we can assume $V(\{x\}) = c > 0$ for all $x \in X$, in particular, $c = 1$ or $c = 1/n$. Obviously, in the last case, V is a probability measure on $\mathcal{U}$.

Theoretically the **inclusion relation** of statistical classes is introduced with the so-called minimal events. In this paper we drop this theoretical construction (see for details [8]). For practical applications it is sufficient to know of how this relation is defined by using membership functions. Given a statistical class F, defined by a probability measure P_F with a density $h_F(x)$. Then functions

$$\underline{\mu}_F(x) = \sum_{y \in X | h_F(y) < h_F(x)} P_F(\{y\}) \text{ and } \bar{\mu}_F(x) = \sum_{y \in X | h_F(y) \leq h_F(x)} P_F(\{y\})$$

are called a lower and an upper membership functions of the statistical class F respectively. By definition, the statistical class F_1 is included to the statistical class F_2, i.e. $F_1 \subseteq F_2$, if $\underline{\mu}_{F_1}(x) \leq \underline{\mu}_{F_2}(x)$ and $\bar{\mu}_{F_1}(x) \leq \bar{\mu}_{F_2}(x)$ for all $x \in X$. It is possible to prove that membership functions define each statistical class uniquely. In the next, we consider set-theoretical operations on statistical classes, which are produced with the help of min and max operations:

1. $\underline{\mu}_{F_1 \cap F_2}(x) = \min\left(\underline{\mu}_{F_1}(x), \underline{\mu}_{F_2}(x)\right)$, $\bar{\mu}_{F_1 \cap F_2}(x) = \min\left(\bar{\mu}_{F_1}(x), \bar{\mu}_{F_2}(x)\right)$ are membership functions of the statistical class $F_1 \cap F_2$;
2. $\underline{\mu}_{F_1 \cup F_2}(x) = \max\left(\underline{\mu}_{F_1}(x), \underline{\mu}_{F_2}(x)\right)$, $\bar{\mu}_{F_1 \cup F_2}(x) = \max\left(\bar{\mu}_{F_1}(x), \bar{\mu}_{F_2}(x)\right)$ are membership functions of the statistical class $F_1 \cup F_2$.
   It is possible that a statistical class $F_1 \cap F_2$ or $F_1 \cup F_2$ can not be generated by a probability measure. The sense of this can be explained while considering classification problems.

An **inclusion measure of statistical classes** $\mu(F_1 \subseteq F_2)$ is introduced for evaluating an inclusion degree of statistical class F_1 into statistical class F_2. By definition, $\mu(F_1 \subseteq F_2) \in [0, 1]$ and $\mu(F_1 \subseteq F_2) = 1$ iff $F_1 \subseteq F_2$. This functional can be introduced in various ways but we don't describe all of them, see [8,9,10] for details. Here we use the inclusion measure axiomatically defined in [9]. Let us introduce an auxiliary function

$$\psi\left(F_1 \subseteq F_2\right) = 0.5 \int_X \left(\underline{\mu}_{F_2}\left(x\right) + \bar{\mu}_{F_2}\left(x\right)\right) dP_1 = 0.5 \sum_{x \in X} \left(\underline{\mu}_{F_2}\left(x\right) + \bar{\mu}_{F_2}\left(x\right)\right) P_1\left(x\right).$$

Then an inclusion measure is defined as

$$\mu\left(F_1 \subseteq F_2\right) = \psi\left(F_1, F_1 \cap F_2\right) + 1 - \psi\left(F_2, F_1 \cup F_2\right).$$

Let us notice that in the last formula it is necessary to compute membership functions of statistical classes $F_1 \cap F_2$ and $F_1 \cup F_2$.

3.2 Secondary Features Construction Based on Inclusion Measures

Let we have a set $\{S_1, S_2, ..., S_n\}$ of etalon statistical classes. Then the classification of any statistical class F consists in computing the following classifying vector: $(\mu\left(F \subseteq S_1\right), ..., \mu\left(F \subseteq S_n\right))$.

Let us show how to get secondary features using inclusion measures. Since secondary features can be represented by means of histograms, we can assume that any low-level feature is a histogram being an estimate of probability distribution. For example, features derived with the help of the gradient directions are the set of histograms that correspond to different positions of the scanning window. Features of the texture coarseness and background homogeneity can be represented as a histogram if we calculate these features for different positions of the scanning window.

Let us consider how to build etalon classes. Assume that images are annotated such that a keyword corresponds to the part or the whole image. For example, if an image is annotated by a word "building", then the annotation procedure would be more precise if to compute the statistical characteristics for the part of the image, where the building is situated. Assume that we are going to extract secondary features for buildings. For this aim, we should first construct etalon classes corresponding to the word "building". The simplest way to do so is to compute the histogram for the all images from the learning sample that are annotated by the word "building". Let us assume that the etalon class S has been constructed that corresponds to the word w ("building") and a to given low-level feature b. Then for image classification, we has to compute histograms, corresponding to b. Let these histograms are statistical classes $F_1, ..., F_l$. Then the secondary feature is $p(w|b) = \max\left\{\mu\left(F_1 \subseteq S\right), ..., \mu\left(F_l \subseteq S\right)\right\}$. Notice that in the last formula the maximum is used, because we choose in this case the part of the image that is the most relevant to the keyword w. Notice also that the inclusion measure has a probabilistic interpretation: it is a mean value of conditional probability of minimal events that correspond to the etalon class S provided that we observe the statistical class F. Hence, the greater value $p(w|b)$ is, the greater probability is that the image is annotated by the keyword w.

3.3 The Aggregation of Secondary Features

Assume that on this stage we have a vocabulary $W = \{w_1, ..., w_m\}$ and a set of features $B = \{b_1, ..., b_n\}$. The secondary features are presented by $p(w_i|b_j)$. The next procedure is to construct aggregation functions $\varphi_i : [0,1]^n \rightarrow [0,1]$, allowing us to compute the global features: $p(w_i) = \varphi_i\left(p(w_i|b_1), ..., p(w_i|b_n)\right)$, $i = 1, ..., n$, which have to give us the global evaluation that the keyword w_i is relevant to the analyzed image.

Let $\varphi : [0,1]^n \rightarrow [0,1]$ be a aggregation function. Then it has the following properties [11]: 1) $\varphi(\mathbf{0}) = 0$ and $\varphi(\mathbf{1}) = 1$, where $\mathbf{0} = (0, ..., 0)$ $\mathbf{1} = (1, ..., 1)$; 2) $\varphi(\mathbf{x}) \leq \varphi(\mathbf{y})$ for $\mathbf{x} \leq \mathbf{y}$, where $\mathbf{x} = (x_1, ..., x_n)$, $\mathbf{y} = (y_1, ..., y_n)$, and $\mathbf{x} \leq \mathbf{y}$ if $x_i \leq y_i$ for all $i \in \{1, ..., n\}$.

If we assume that the features are independent, then it is rational to use linear aggregation functions of the type $\varphi(\mathbf{x}) = \sum\limits_{i=1}^{n} a_i x_i$, where $a_i \geq 0$, $i = 1, ..., n$, and $\sum\limits_{i=1}^{n} a_i = 1$. Let we annotate images by the rule: an image is annotated by a keyword w_i if $p(w_i) > \varepsilon_i$. In this scheme parameters of aggregation functions φ_i and non-negative numbers ε_i have to be estimated using the learning sample. Suppose that the learning sample consists of N images. In this case any image with a number $k \in \{1, ..., N\}$ is described by a vector of secondary features $\mathbf{p}_k = (p_k(w_i|b_1), ..., p_k(w_i|b_n))$, that characterizes the relevance of the keyword w_i. Assume further that we code with a number $\delta_k \in \{-1, 1\}$ the information whether or not the image with the number k is annotated by the keyword w_i, assuming that $\delta_k = 1$ if w_i is in the image annotation, and $\delta_k = -1$, otherwise. Then we have a learning problem of searching a vector $\mathbf{a} = (a_1, ..., a_n)^T$ and a threshold value ε so that the number of false classifications would be minimal. In other words, the number of true inequalities $\delta_k(\mathbf{p}_k\mathbf{a} - \varepsilon) > 0$ $k = 1, ..., N$, would be maximal. Such optimization problem of finding a linear classifier is classical in pattern recognition theory and can be solved by any well-known algorithm, in particular, perceptron algorithm [12].

4 Conclusion

In this paper a problem of automatic image annotation is considered and a general scheme for this problem is presented based on low-level image features extraction. The key properties of low-level features are discussed and several feature types with desired properties are briefly described. The further annotation procedure is based on extracting secondary features from the low-level features and on classifying the obtained patterns. For this purpose, the notion of statistical class and the inclusion measure of statistical classes are introduced. In our problem, we propose to use statistical classes for representing probability distributions of low-level features. A scheme of classifying statistical classes into etalon classes, which correspond to keywords, is given. The generation of annotations is produced by the aggregation of secondary features using linear decision functions constructed by the learning procedure based on perceptron algorithm. The presented annotation scheme is implemented practically and has shown its effectiveness provided by the proposed algorithms.

References

1. Tsai, C., Hung, C.: Automatically Annotating Images with Keywords: A Review of Image Annotation Systems. Recent Patents on Computer Science 1, 55–68 (2008)
2. Hanbury, A.: A Survey of Methods for Image Annotation. Journal of Visual Languages & Computing 19(5), 617–627 (2008)
3. Duygulu, P., Barnard, K., de Freitas, J.F.G., Forsyth, D.: Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2353, pp. 97–112. Springer, Heidelberg (2002)
4. Jeon, J., Lavrenko, V., Manmatha, R.: Automatic Image Annotation and Retrieval Using Cross-Media Relevance Models. In: Proc. of the ACM SIGIR Conference, vol. 1, pp. 119–126 (2003)
5. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 428–441. Springer, Heidelberg (2006)
6. Abramov, S.K., Lukin, V.V., Ponomarenko, N.N.: Entropy Based Background Measure Calculation for Images Searching and Sorting in the Large Collections. Electronics and Computer Systems 2(21), 24–28 (2007)
7. Tamura, H., Mori, S., Yamawaki, T.: Texture Features Corresponding to Visual Perception. IEEE Trans. On Sys. Man, and Cyb. 8(6), 460–473 (1978)
8. Bronevich, A.G., Karkishchenko, A.N.: Statistical Classes and Fuzzy Set Theoretical Classification of Possibility Distributions. In: Bertoluzza, C., Gil, M.A., Ralescu, D.A. (eds.) Statistical Modeling, Analysis and Management of Fuzzy Data, pp. 173–198. Physica-Verl., Heidelberg (2002)
9. Bronevich, A.G., Karkishchenko, A.N.: Application of Possibility Theory for Ranking Probability Distributions. In: Proc. of the European Congress on Intelligent Techniques and Soft Computing, pp. 310–314 (1997)
10. Bronevich, A.G., Karkishchenko, A.N.: Fuzzy Classification of Probability Distributions. In: Proc. of the Fourth European Congress on Intelligent Techniques and Soft Computing, vol. 1, pp. 120–124 (1996)
11. Grabisch, M., Pap, E., Mesiar, R., Marichal, J.-L.: Aggregation Functions. Cambridge University Press, Cambridge (2009)
12. Tsypkin, Y.Z.: Adaptation and Learning in Automatic Systems. Academic Press, Inc., Orlando (1971)
13. Skowron, A., Swiniarski, R.W.: Information Granulation and Pattern Recognition. Rough-Neural Computing. In: Techniques for Computing with Words, pp. 599–636. Springer, Heidelberg (2004)