Image Similarities on the Basis of Visual Content – An Attempt to Bridge the Semantic Gap

Halina Kwasnicka[1], Mariusz Paradowski[1], Michal Stanek[1], Michal Spytkowski[1], and Andrzej Sluzek[2]

[1] Wroclaw University of Technology, Wroclaw, Poland
{halina.kwasnicka,mariusz.paradowski,michal.stanek}@pwr.wroc.pl
[2] School of Computer Engineering, Nanyang Technological University, Singapore

Abstract. Image similarities is a useful concept regarding to the image retrieval on the basis of visual content of the images (CBIR - **C**ontent **B**ased **I**mage **R**etrieval). Because an image can have far more interpretations than text, visual similarity can be totally different from semantic similarity. We have developed similar images searching tools using global approaches as well as local approaches to find near similar images. In this paper we propose a method of bridging local and global levels, what should solve the problem of limited, non-adaptable dictionary when we use automatic annotations in a similar images retrieving task. Our faraway goal is to face the difficult problem with all current approaches to CBIR systems, connected with visual similarity: the semantic gap between low-level content and higher-level concepts.

1 Introduction

The concept of similarity plays a key role in image analysis and, more specific, in image *retrieval*. Viable formulation of *image similarity* allows effectively recognizing and retrieving images with related content. The general concept of image similarity is vague and may be defined in multiple different ways. Man is able to determine the mutual similarity of two images shown to him. Also he is able to find the similar image to a given one, but this image is *similar in his view*. So, the term *images similarity* is not precise, it is very subjective when is considered by people. Let us see the formal definition of word 'similarity', defined in the American Heritage Dictionary [29]: ***similarity*** – is quality or condition of being similar; resemblance. Following words are the synonyms: *likeness*, *similarity*, *similitude*, *resemblance*, *analogy*, *affinity*. These words denote agreement or conformity, the *likeness* implies close agreement, *similarity* and *similitude* suggest agreement **only in some respects** or **to some degree**, while *resemblance* refers to similarity in external or superficial details. *Analogy* means similarity "as of properties or functions, between things that are otherwise not comparable". The last word, *affinity* is a "likeness deriving from kinship or from the possession of shared properties or sympathies" [29]. What we want is **to design a computer system which will be able to find similar images to a given one**. Taking into account the described meanings, this task is very difficult. One

N.T. Nguyen, C.-G. Kim, and A. Janiak (Eds.): ACIIDS 2011, LNAI 6591, pp. 14–26, 2011.

can expect that such computer system will be imprecise and, possibly, it will require be tuned for particular users.

Summing the above, different people consider different images as similar and would like to get different retrieval results. Images may be considered similar if they: have the same interpretation, share the same object(s), evoke the same emotions, have identical spatial arrangement, share the same colors or textures, have identical fragments, etc. Some interpretations of the mentioned similarity concept may be modeled using the object *recognition paradigm*, i.e. intelligent techniques. To make the situation even more complex, we also need to take into account the human perspective and expectations [24]. Thus, machine learning paradigm seems to be a reasonable solution to the problem of image similarity measurement.

Application of object recognition paradigm to image retrieval may be successfully implemented using the similarity of recognized concepts. Instead of low-level, pixel based queries, the user is able to formulate meaningful, concept based queries [14]. This image retrieval scheme is sometimes referred to as ***Annotation Based Image Retrieval*** [11] in contrast to classic ***Content Based Image Retrieval***. Despite its multiple advantages, researchers point out the key disadvantage of such approach: the number of concepts is *predefined* and *finite* [28]. This property makes the object recognition based paradigm inapplicable when faced with infinite diversity of the surrounding world [7]. Effective image retrieval may require continuous creation of new concepts which describe the environment in a precise way.

The paper is a continuation of our research on the mentioned problem. We seek how to automatically create new concepts without any a'priori information, purely on a visual basis [23] and seamlessly integrate them into the notion of image similarity. The presented idea consists of multiple subcomponents, solving various subproblems, but it may not be yet considered as complete or functional. Thus, in this paper we do not give clear answers, but rather we present our most recent ideas. Some methods which are regarded as components of the proposed idea were developed and studied [1,17,27].

2 Global Image Analysis

The first component of the presented solution is a global image analysis method where we are interested in extracting general, holistic image features. Such features are easy to generalize and efficient for processing by intelligent approaches. We may simply build image recognition methods based on global features and accompanying labels. These recognition methods may be image distance based which turns to be a quite effective approach. Having a set of labels, the image similarity retrieval becomes a text based retrieval. However, as mentioned above, we face the problem of finite, limited and non-adaptable dictionary.

2.1 Global Image Distances

Automatic methods of images analysis define *image similarity* as a distance measure between images, which is a sum of distances between global visual features of considered images. To obtain the similarity between two images one can measure the distance between visual vectors in metric or probabilistic space. Minkowski, Cosine, Correlation, Mahalanobis or EMD are commonly used measures to calculate distances between visual features. The other approaches use divergence between image probabilistic models calculated for the set of visual features. In that category commonly used measure is Kullback-Leibler divergence or its symmetric version Jehnsen-Shannon divergence.

Visual features of an image define its certain visual property. Global features capture some overall characteristics of an image, as color, texture and shape. An image can be divided into a number of sub-images; in such approach, the whole image is described by a vector of features calculated for each sub-image. The global approach has one important advantage: the high speed, both features extraction and similarity measure calculation [6]. However, the global features are usually too rigid to represent an image.

The second approach is extraction of local features, computed for every pixel using its neighborhood. Additional step, features summarization must be performed. Often data set based on a distribution for each pixel is calculated in summarization step.

Some features from MPEG-7 standard, as histogram-based descriptors, spatial color descriptors and texture descriptors seem to be well suited for natural images retrieval [4,5,6].

2.2 Automatic Image Annotation

Automated **I**mage **A**nnotation (AIA) is a process which describes previously unseen image $\boldsymbol{Q}$ by a set of concepts $\{w_1, w_2, \cdots, w_N\}$ from the semantic dictionary $\boldsymbol{D}$. Word assignment can be made by finding the correlation between visual features which characterize query image $\boldsymbol{Q}$ and high-level semantics (concepts). AIA is an integral part of modern CBIR systems. Image annotations can be seen as a bridge between textual queries and visual image content.

Machine learning techniques used to solve the AIA problem can be split into *classification* based methods and *probabilistic modeling* methods. Classification methods lie on training classifiers to recognize if a given word is present within the proper description of the image. Different classifiers can be used in this approach, good result and speed one can obtain with decision trees [13,22].

Probabilistic modeling methods, such as Hierarchical Probabilistic Mixture Model (HPMM) [10], Translation Model (TM) [8], Supervised Multi-Class Labeling (SML) [3], Continuous Relevance Model (CRM) [14] and Multiple Bernoulli Relevance Models (MBRM) [9], try to find the probability density function of visual features associated to concepts. Parametric or non parametric estimation can be used in this approach.

Results obtained by AIA methods can be further improved by using filter methods which take into account word co-occurrence models [15], words relations in Word-Net [12] or our GRWCO [13] method which reduces the difference between expected and resulted word count vectors to reranking the output annotations. Recently, Makadia et al. have proposed a new method based on the hypothesis that similar images are likely to share the same annotations [19]. In this approach, an image annotation is a process of transferring most frequent labels from nearest neighbors. The method **does not solve the fundamental problem of determining the number of annotations** that should be assigned to the target image, it assumes that the optimal annotation length is given.

In our recent research, we have extended this approach. We have proposed PATSI (**P**hoto **A**nnotation **T**hrough Finding **S**imilar **I**mages) annotator which introduces transfer function [27] as well as an optimization algorithm which can be used to find both, the optimal number of neighbors and the best transfer threshold according to the specified quality measure [17]. PATSI consists of two main phases: *preparation* and *query*. In the first phase, for each image repeat: (1) split the image into a number of regions (sub-images); (2 calculate statistical visual features for every region (sub-image); (3) create the model of the image. In the query phase, do: (1) split the query image into regions (sub-images); (2) build a model of the query image; (3) calculate distances between the query image and all images in the dataset; (4) Select k most similar images; (5) Transfer all words (annotations) with a weight dependent on a position of a considered image in a similarity ranking list (how much the image is similar to the query image); (6) Select words with sum of weights greater than the assumed threshold t. These words are the annotations of the query image. The more detailed description of the PATSI algorithm one can find in [17] and [27].

2.3 Image Retrieval Using Annotations

In the PATSI approach, concepts from the most similar images are transferred to the query image using transfer function. Finding the k most similar images is performed by calculating the distance measure between visual features of a query image and images in the training set. The resulting annotation consists of all the words whose transfer values are greater than a specified threshold value t. The threshold value t influences the resulting annotation length. Optimal threshold value t^* and number of neighbors k must be found using an optimization process [17]. Images retrieval using PATSI is embedded into the method. A query image is an image for which the similar images should be found. The third task in the query phase is calculation of distances between the query image and all other images in a dataset. The images from the dataset are ranked with increasing distances and are presented to a user with this ordering. Jehnsen-Shannon divergence is calculated between models of images built onto image visual features. Visual features are treated as a realization of multivariate random variable described by multivariate Gaussian distribution. The parameters of that distribution are calculated using the Expectation Maximization

Algorithm (EM) [27]. All images were split by 20-by-20 grid splitter and for every cell a mean color value as well as a color deviation in RGB color space were calculated. Additionally, for all segments their center points, and mean Eigen values calculated on color Hessians were stored. PATSI annotation results using F-measure for MGV2006 [22] dataset with different visual features as well as different distance measures are presented in Table 1. For all visual features as well as distance measures we used exactly 19 most similar images in transfer process. All words with transfer value greater than $t = 1.2$ were then treated as the final annotation. PATSI annotator run with using distances in metric space

Table 1. F-measure of AIA on MGV2006 dataset using PATSI annotator with different feature sets and distance measures in the metric space

Visual Feature	Distance measure					
	Cannbe-ra	Chebys-hev	City-block	Correla-tion	Cosine	Eucli-dean
Auto Color Collerogram	0.20	0.16	0.18	0.17	0.17	0.17
CEDD	0.25	0.18	0.25	0.27	0.27	0.27
FCTH	0.24	0.17	0.25	0.23	0.23	0.24
Fuzzy Color Histogram	0.12	0.13	0.13	0.16	0.16	0.13
Gabor	0.06	0.06	0.06	0.09	0.09	0.06
General Color Layout	0.14	0.09	0.14	0.09	0.08	0.11
JPEG Coefficient Histogram	0.20	0.18	0.21	0.21	0.22	0.21
Tamura	0.15	0.14	0.15	0.15	0.15	0.15
CoOccurance matrix	0.17	0.07	0.17	0.17	0.18	0.16
RGB	0.20	0.10	0.20	0.20	0.18	0.23
HSV	0.21	0.09	0.21	0.19	0.17	0.19
RGB + DEV.	0.23	0.09	0.21	0.21	0.18	0.20
HSV + DEV	0.22	0.09	0.22	0.18	0.19	0.19
RGB + DEV + HES	0.23	0.10	0.22	0.18	0.18	0.20
HSV + DEV + HES	0.22	0.09	0.22	0.19	0.19	0.19
RGB + DEV + XY + HES	0.22	0.10	0.22	0.22	0.22	0.20
HSV + DEV + XY + HES	0.23	0.09	0.22	0.20	0.20	0.19

achieved highest results for CEDD visual feature, and Euclidean measure. The best mean F measure was also achieved with Euclidean distance. Very interesting results can be achieved using PATSI annotator with distance measure calculated in probabilistic space, see Table 2. Using Jehnsen-Shannon divergence allows us to significantly improve annotation results in comparison to results presented in Table 1 as well as for the other state-of-art methods [27].

Examples of annotations generated by PATSI for images from ICPR2004 database are presented in Table 3. This table contains also images identified as the most similar images to the considered one. We use Jehnsen-Shannon divergence to calculate distances between images. The PATSI annotator performance in comparison to other state-of-the art method was improved by 20 percentage points [17], achieving F-Measure equal to 78% for the best 27% words in the

Table 2. F-measure of AIA on MGV2006 dataset using PATSI annotator using Jehnsen-Shanon divergence in comparison to other state-of-art methods

Method	Precision	Recall	F-measure
PATSI(HSV + DEV)	0.33	0.38	0.36
PATSI(RGB + Dev)	0.40	0.44	0.42
PATSI(RGB + DEV + HES + XY)	**0.42**	**0.43**	**0.43**
FastDIM	0.24	0.16	0.19
FastDIM + GRWCO	0.34	0.34	0.34
MCML	0.32	0.24	0.27
MCML + GRWCO	0.38	0.37	0.37
CRM	0.39	0.34	0.36

dictionary of MGV2006 database [22]. The results suggest that for a small number of concepts AIA can be now treated as the effective image retrieval tool.

During experiments we have noticed that some of the features as well as distance measures are more suitable to detect some groups of words, while showing a weak performance for others. By combining them together we can increase overall annotation performance. Current research is focused on combining many similarity measures and visual features in one annotation transfer process. We have extended the PATSI algorithm to the multi-PATSI method which performs annotation transfer process based onto many similarity matrices calculated using different feature sets and different similarity measures. The results are combined into the final annotation based on the quality of particular annotators for specific words.

3 Local Image Analysis

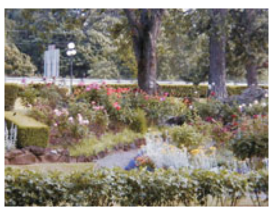
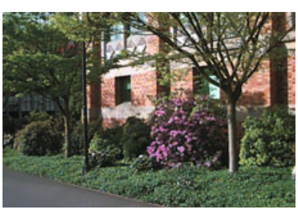
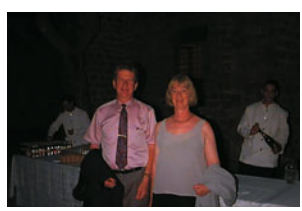
Local image analysis methods are built on the basis of local features, i.e., features calculated from very small image regions. Very popular and effective types of local features are *keypoints* [18,20]. Keypoints themselves are much harder to generalize (although such attempts exist, e.g., [21]) because they are much diversified along single objects. Yet, keypoints have a very nice property, they are able to capture the notion of sameness.

3.1 Image Matching

The goal of image matching is to detect whether two images share visually identical content. Image matching problem may be divided into many subproblems, such as: sub-image matching [16,30], image fragment matching [23], panorama recognition [2], etc. All these techniques provide high precision results.

Sub-image matching methods are able to determine if one image is a fragment of another image. Such approaches may be very useful for finding identical content in case where both images share only one common object. The key advantage of sub-image matching is the applicability of complex (even non-linear)

Table 3. PATSI annotation results for example images from ICPR2004 with their nearest neighbors

Original (query) image:	**Original (query) image:**	**Original (query) image:**
		
Original annotation: 'elk', 'greenery', 'ground', 'logs', 'tree', 'trunks'	**Original annotation:** 'river', 'trees'	**Original annotation:** 'man', 'people', 'table', 'woman'
Generated annotation: 'elk', 'greenery', 'ground', 'logs', 'tree', 'trunks'	**Generated annotation:** 'garden', 'grass', 'trees'	**Generated annotation:** 'man', '*microphone*', 'woman', 'people',
Similar images:	**Similar images:**	**Similar images:**
		
		
		
		

geometrical models for the matching process. This allows finding objects seen from different viewpoints or even deformed ones. These methods may be effectively used to capture large objects, such as monuments, buildings. However, they are ineffective when faced with a problem of finding multiple fragments on both images.

Image fragment matching utilizes simpler geometrical models, but is able to find multiple identical objects on scenes with cluttered background. The disadvantage of this approach is the relative simplicity of applied geometry. Deformed or strongly non-planar objects are harder to capture. These methods may be effectively used to capture small object, such as bottles, books, boxes, etc.

Panorama recognition techniques assume that there is only one object of interest. This object is however captured only partially, i.e., different images contain different fragments of the object of interest. These methods may be used to capture huge objects, such as e.g. landscapes, cityscapes.

3.2 Automatic Visual Object Formation

The last, and the most important, fragment of our solution in low level vision refers to the concept of visual objects [7]. We have proposed a grouping method which is able to automatically form *visual object* [24,25]. It is based on the image matching methods discussed in Section 3.1. Having a high precision matching routine we may expect that the created groups are free of errors. The method is able (in a very limited way) to find meaningful visual objects purely on a visual basis, without any training data or supporting information. In fact it is an attempt to bridge *the semantic gap* [6,26].

The *automatic visual object formation* method has four major steps: (1) *pre-retrieval* to make the process more efficient; (2) *image matching* to find similarities within the set; (3) formation of *prototypes*, which are an intermediate structure [7], and finally; (4) formation of *visual objects*. In the first step we measure similarities between all images in the database. For further processing we select only the most similar ones. In the second step we perform image matching for all pairs of similar images within the set. As a result we get a set of (nearly all) similar image fragments found within the input collection. Because each image is matched with multiple other images, some image regions on a single image may have multiple different matches with other images. In the third step we group all these regions found within a single image. Created groups are called *prototypes*. In the last step we group all prototypes according to matching information between images. Resulting groups are called *visual objects* and they represent frequently repeating, matched fragments from the input collection. Exemplary *visual objects* found in a database containing both indoor and outdoor scenes are presented in Fig. 1. Although, each *visual object* consists of images containing the manifestation of the same underlying, physical object, this information is very useful. It allows formulating very specific queries, we may seek for such specific objects as a road sign, a flu-remedy pack, a model of a ship or a car, a monument, a mountain or landscape, etc.

Fig. 1. Exemplary visual objects are outlined on images from a processed image collection

4 Bridging Local and Global Level Vision

Having described all necessary components, let us now present the main idea of our current research. We envision that both, global and local image analysis routines cooperate together. We would like to utilize global approaches to provide an effective retrieval tool, and we would like to enforce it by the local approach to solve the problem of limited, non-adaptable dictionary. Let us assume that the dictionary used in the global image processing is hierarchical, e.g., it is a fragment of some larger *ontology*. Some concepts in the hierarchy may be *contradictory* and cannot exist together.

Given a small set of hierarchically arranged concepts (e.g. inside, outside, mountain, ship, building, sky) and a collection of images containing multiple instances of identical objects (however seen in different scenes and contexts) we would like to make the hierarchy of concepts more specific and precise. This idea is illustrated in Fig. 2. First we detect all visual objects using the object formation routine discussed in Section 3.2. Having all identical objects captured, we would like to link them into our existing hierarchy of objects. To do this, we employ intelligent, global image analysis techniques, e.g. classification, automatic image annotation. If needed, we may use a different intelligent technique on each level of hierarchy. Recognition process should take into account shapes of regions creating a visual object. We divide the image into three separate segments, each having a different *meaning* for the processed visual object. These three segments are: *interior*, *context* and *environment*, they are illustrated in Fig. 3. Having recognized objects on all images belonging to a single visual object, we may decide where to attach it within the concept hierarchy. Usually, various recognition or annotation methods have one of three possible outputs: precise concept probability values, roughly estimated concept scores or just a subset of concepts from the dictionary. All these output types have to be processed in a different way. We have designed three decision rules, one for each type of output. Each decision rule outputs a single support value s_x^w for each concept w and each image x containing the visual object.

Final decision regarding of linking the new concept within the hierarchy is made on the basis of decision rule outputs. An averaged concept support values s_w is calculated and possible contradictions in the hierarchy are solved. Contradictory concepts in each level of hierarchy are modeled as a set of sets Z (multiple different rules on each level of hierarchy). Each set Z_i contains all contradictory concepts. In case there are two or more contradictory concepts, the ones with

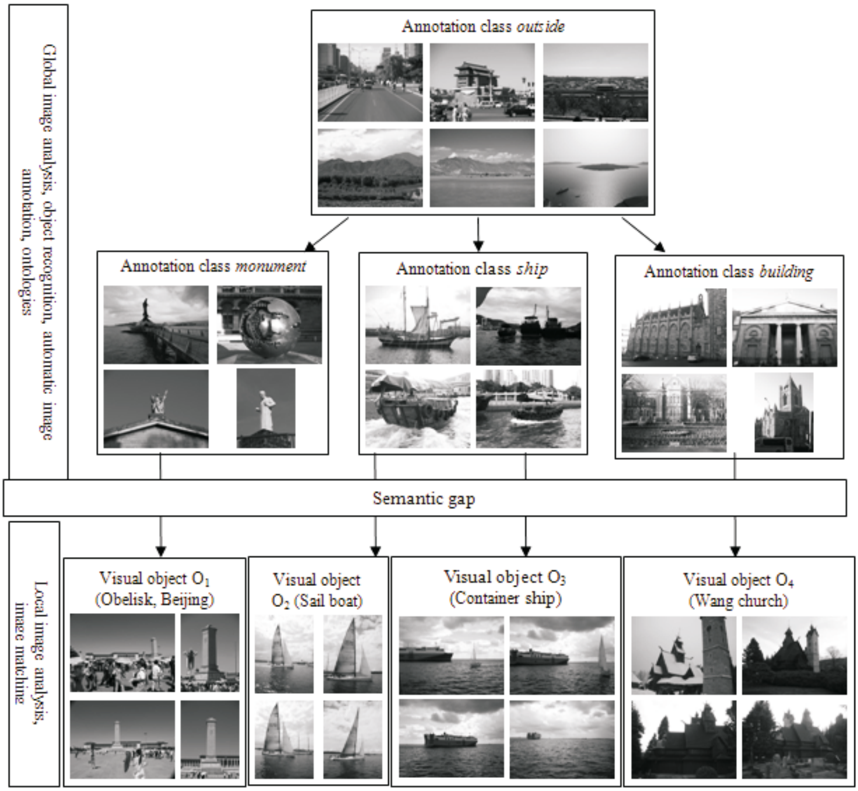


Fig. 2. The concept of bridging local and global level vision

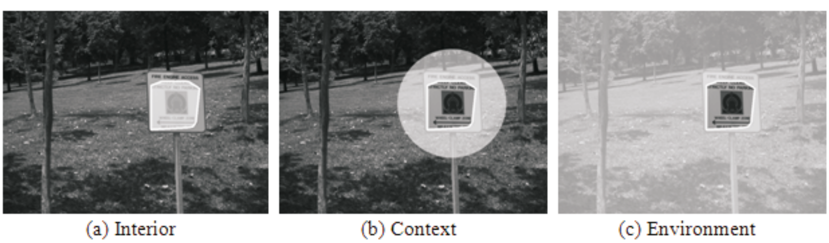


Fig. 3. Three different annotation regions for visual objects

the larger support are chosen by the decision rule. In case the decision rule d_w accepts the visual object it is processed deeper and deeper within the hierarchy.

5 Conclusion

The paper shows briefly the results of our methods concerning similar images retrieval using PATSI annotation algorithm (now we are testing multi-PATSI method) and the method of images matching – it detects whether two images share visually identical content. The important part of our research in low level vision refers to the concept of visual objects. We have proposed a grouping method which is able to automatically form visual object, this approach is based on the image matching methods. Our method is able to find meaningful visual objects purely on a visual basis.

Currently we try to join global and local image analysis routines. Global approaches should provide efficient retrieval tool, but it can work only on limited dictionary, with all words well represented in a training set. Such a dictionary can contain words from a given ontology, i.e., the dictionary consists of hierarchically arranged concepts. Captured in low level analysis identical objects can be linked into a hierarchy of concepts (objects) by global image analysis techniques, e.g., automatic image annotation method.

Our future plans concern with the above mentioned problem. Initial set of decision rules are proposed, but we do not have experimental results. Of course, all sub-methods in the proposed approach should work very well. Having weak one part of the method we are not able to obtain good final results. So, we plan to improve our global method (e.g., multi-PATSI method) as well as the automatic visual object formation methods. These two research topics will be conducted in parallel with studies on the 'bridge' method that should allow for filling up the semantic gap, perhaps even to a limited extent.

All the presented researches are dedicated to searching similar images, although we still have a problem with understanding the concept *images similarity*. Meaning of *similarity of images* still causes problem, however more of us can easily indicate the similar images within a not large collection of images. It is important that those images are usually similar in *the view of particular user*, and therefore the *term images similarity is not precise*. In our group we have developed computer program, called SIMILARIS, and a set of images used with this program. The main aim of that research is defining a kind of baseline – measures of images similarity when these images are evaluated by people. That data can be than used to find the efficient measure of image similarity. After finishing the testing phase and our preliminary study, the program SIMILARIS together with used collection of images will be published on the server with free access to researchers. Researchers on CBIR systems focus on building systems with the very high precision, but the fundamental question still remains without answer: is it possible to obtain CBIR systems with high precision and recall measures? The studies with SIMILARIS should help to find answer to the above question.

Acknowledgments. This work is partially financed from the Ministry of Science and Higher Education Republic of Poland resources in 2008-2010 years as a Poland-Singapore joint research project 65/N-SINGAPORE/2007/0.

References

1. Broda, B., Kwasnicka, H., Paradowski, M., Stanek, M.: MAGMA: efficient method for image annotation in low dimensional feature space based on Multivariate Gaussian Models. In: Ganzha, M., Paprzycki, M. (eds.) Proc. of the IMCSIT. Polish Information Processing Society 2009, pp. 131–138 (2009)
2. Brown, M., Lowe, D.G.: Recognising panoramas. In: Ninth IEEE International Conference on Computer Vision, vol. 2, pp. 12–18 (2003)
3. Carneiro, G., Chan, A., Moreno, P., Vasconcelos, N.: Supervised learning of semantic classes for image annotation and retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence 29(3), 394–410 (2007)
4. Chatzichristofis, S., Boutalis, Y.: Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. Computer Vision Systems, 312–322 (2008)
5. Chatzichristofis, S.A., Boutalis, Y.S.: Fcth: Fuzzy color and texture histogram – a low level feature for accurate image retrieval. In: Interactive Services, Intern. Workshop on Image Analysis for Multimedia, pp. 191–196 (2008)
6. Datta, R., Joshi, D., Li, J., Wang, J.: Image retrieval: Ideas, influences, and trends of the new age. ACM Transactions on Computing 40(2) (2008)
7. Dickinson, S.J., Leonardis, A., Schiele, B., Tarr, M.J. (eds.): Object Categorization: Computer and Human Vision Perspectives. Cambridge University Press, Cambridge (2009)
8. Duygulu, P., Barnard, K., de Freitas, J.F.G., Forsyth, D.A.: 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)
9. Feng, S.L., Manmatha, R., Lavrenko, V.: Multiple bernoulli relevance models for image and video annotation. In: Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1002–1009 (2004)
10. Hironobu, Y.M., Takahashi, H., Oka, R.: Image-to-word transformation based on dividing and vector quantizing images with words. Neural Networks in Boltzmann Machines 4 (1999)
11. Inoue, M.: On the need for annotation-based image retrieval. In: Proc. of the Information Retrieval in Context (IRiX), A Workshop at SIGIR 2004, pp. 44–46 (2004)
12. Jin, Y., Khan, L., Wang, L., Awad, M.: Image annotations by combining multiple evidence and wordnet. In: Proc. of the 13th Annual ACM Intern. Conf. on Multimedia (2005)
13. Kwasnicka, H., Paradowski, M.: Resulted word counts optimization-a new approach for better automatic image annotation. Pattern Recognition 41(12) (2008)
14. Lavrenko, V., Manmatha, R., Jeon, J.: A model for learning the semantics of pictures. In: Proc. of Neural Information Processing Systems (NIPS). MIT Press, Cambridge (2003)
15. Llorente, A., Motta, E., Roger, S.: Image annotation refinement using web-based keyword correlation. In: Proc. of the 4th International Conference on Semantic and Digital Media Technologies, Berlin, pp. 188–191 (2009)
16. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)

17. Maier, O., Stanek, M., Kwasnicka, H.: PATSI - photo annotation through similar images with annotation length optimization. In: Klopotek, M.A., et al. (eds.) Intelligent information systems 2010, pp. 219–232. Pub. House of Univ. of Podlasie (2010)
18. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In: Proc. of British Machine Vision Conference, pp. 384–393 (2002)
19. Makadia, A., Pavlovic, V., Kumar, S.: A new baseline for image annotation. In: Proc. of the 10th European Conference on Computer Vision, pp. 316–329 (2008)
20. Mikolajczyk, K., Schmid, C.: Scale and Affine Invariant Interest Point Detectors. International Journal of Computer Vision 60, 63–86 (2004)
21. Monay, F., Quelhas, P., Odobez, J.M., Gatica-Perez, D.: Contextual Classification of Image Patches with Latent Aspect Models. EURASIP Journal on Image and Video Processing, Article ID 602920 (2009), doi:10.1155/2009/602920
22. Paradowski, M.: Methods of automatic annotation as an efficient tool for images collections describing. PhD thesis, Wroclaw Univ. of Technology (2008) (in Polish)
23. Paradowski, M., Sluzek, A.: Automatic Visual Object Formation using Image Fragment Matching. In: Proc. of 2010 Intern. Multiconf. on Computer Science and Information Technology, vol. 5, pp. 45–51, IEEE Catalog CFP1064E-CDR (2010)
24. Russell, R., Sinha, P.: Perceptually-based comparison of image similarity metrics. Technical report, AIM-2001-014, CBCL-201 (2001)
25. Sluzek, A., Paradowski, M.: A Vision-based Technique for Assisting Visually Impaired People and Autonomous Agents. In: Proc. of 3th Intern. Conference on Human System Interaction, pp. 653–660 (2010)
26. Smeulders, A.W.M., Gupta, A.: Content-Based Image Retrieval at the End of the Early Years. Transactions on Pattern Analysis and Machine Intelligence 22(12), 1349–1380 (2000)
27. Stanek, M., Broda, B., Kwasnicka, H.: PATSI — Photo Annotation through Finding Similar Images with Multivariate Gaussian Models. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010. LNCS, vol. 6375, pp. 284–291. Springer, Heidelberg (2010)
28. Tadeusiewicz, R., Ogiela, M.R.: The New Concept In Computer Vision: Automatic Understanding of the Images. Proc. of Artificial Intelligence and Soft Computing, 133–144 (2004)
29. The American Heritage® Dictionary of the English Language. 4th edn. copyright ©2000, by Houghton Mifflin Company. Updated in 2009, Houghton Mifflin Company Pub. (2000)
30. Yang, D., Sluzek, A.: A low-dimensional local descriptor incorporating TPS warping for image matching. Image and Vision Computing 28(8) (2010)