Exploiting Class-Specific Features in Multi-feature Dissimilarity Space for Efficient Querying of Images

Turgay Yilmaz, Adnan Yazici, and Yakup Yildirim*

Dept. of Computer Engineering, Middle East Technical University, Ankara, Turkey
{turgay,yazici}@ceng.metu.edu.tr

Abstract. Combining multiple features is an empirically validated approach in the literature, which increases the accuracy in querying. However, it entails processing intrinsic high-dimensionality of features and complicates realizing an efficient system. Two primary problems can be discussed for efficient querying: representation of images and selection of features. In this paper, a class-specific feature selection approach with a dissimilarity based representation method is proposed. The class-specific features are determined by using the representativeness and discriminativeness of features for each image class. The calculations are based on the statistics on the dissimilarity values of training images.

1 Introduction

CBIR systems aim to retrieve pictures from large image repositories according to the needs of the users [6]. In CBIR systems, images are usually modelled with a set of low level features, such as color, texture or shape, from which underlying similarity functions are used to perform queries [1]. The ultimate goal of designing CBIR systems is to achieve the best possible retrieval accuracy. To achieve high accuracy on a retrieval task, traditional approaches prefer creating superior low level features than the currently available ones, or optimization of them [5, 14]. However, the noise in sensed data, non-universality of any single low level feature and performance upper bounds prevent relying on a single feature [22]. In the information fusion literature, fusing multiple features is an empirically validated approach for increasing the retrieval performance [8, 14, 17, 27].

Dealing with multiple features entails processing intrinsic high-dimensionality of each feature and handling heterogeneous dimensions / scales of different features. Modelling the CBIR system to operate in feature space (storing image features in the database) makes the system struggle with the heterogeneity of different features and prevents it from being fast and flexible [3]. Such a system is not fast since similarity calculation is done at query-time. Also, it cannot be

* This work is supported in part by a research grant from TÜBİTAK EEEAG with grant number 109E014.

H. Christiansen et al. (Eds.): FQAS 2011, LNAI 7022, pp. 149–161, 2011.

flexible either, considering that handling a new feature requires renewing the system for processing the dimensionality and scale of the new feature. Therefore, an alternative approach, that regards the fastness and the flexibility issues, is modelling the system in dissimilarity space. In accordance with the ideas of [20, 21] for representing images with dissimilarities, Bruno et al. [4] present fusing multiple features in dissimilarity space. In dissimilarity space, the images in the database are represented with the dissimilarity values to prototype objects of the particular image categories. Thus, both the retrieval operation is faster and adding new features to the system is easier as long as the distance function is available at once for processing the images in the database.

Beyond the representation problem of images, another crucial issue is to find out the features that are more beneficial for fusion. This problem, namely the feature selection problem, tries to determine which subset of features yield to an optimal result. In [12], Jain et al. group widely-used techniques with a general aspect of view: exhaustive search, branch-and-bound search, best individual features, sequential forward/backward selection, sequential forward/backward floating search. The methods except the exhaustive search provide computationally better ways of finding an optimal set, however exhaustive search guarantees to find the optimal solution. For each of these methods, selection criteria during forward/backward selection operations can differ; information gain, previously-defined quality metrics or the complexity can be considerations. With a more specific view on the problem, some of the recent approaches in the information fusion literature can be listed as: Finding principal/independent components [16, 26], selecting the most coherent and less complex features according to the heterogeneity issue [15], calculating the information gain obtained [2, 13] and defining quality and reliability metrics on features [22, 23].

Although there are many different approaches for the selection of features, all of them have a common preference: The selection process is independent of the category (semantic meaning) of the images. However, considering the idea that different features can be more effective, representative and discriminative for different image categories, using a category dependent feature selection approach can be more beneficial.

In this study, we propose a class-specific feature selection approach for the fusion of multiple features. In order to eliminate the high-dimensionality of multiple features and provide efficient querying over the images, we prefer a dissimilarity based approach. To learn the class-specific features, we carry out a training phase. During the training, the class-specific features are determined by using the representativeness and discriminativeness of features for each image class. The calculations of representativeness and discriminativeness are based on the statistics on the dissimilarity values of training images.

The remainder of this paper is organized as follows: First, the multi-feature modelling in dissimilarity space is introduced in Section 2. Then, the class-specific feature selection approach is given in detail, in Section 3. In Section 4, the empirical results and the evaluations are presented. Lastly, in Section 6, some conclusions are drawn and further study is discussed.

2 Multi-feature Modelling in Dissimilarity Space

The literature of information fusion agrees on the idea that combining multiple features enhances the efficiency. However, how to combine such information is still studied. One of the discussed issues is the representation of images. In feature based representation, an image is usually represented with a multi-dimensional feature vector and having multiple features causes dealing with multiple of such multi-dimensional feature vectors, each having different dimensions and scales. Handling the complexity of different dimensions and scales of different features makes the CBIR system more dependent on the currently available features and less flexible to new features. In [3], Bruno et al. discuss these issues in detail. Still, a more crucial flaw for feature-based representation is the inefficiency of the fast querying capabilities. Having features in the database requires calculating the similarities of related images for every query task.

A more convenient way is the dissimilarity based representation [3,7,20,21]. In dissimilarity based representation, feature values are not stored in the database; instead the dissimilarity values of images are stored. Thus, the CBIR system does not need to deal with the intrinsic dimensionality of features to combine them. In addition, a query task is simpler; it does not require similarity calculations for each query. The dissimilarity values of images are calculated once, before including the image into the CBIR system. To calculate the dissimilarity values, the dissimilarity functions of each feature are utilized. Hence dissimilarity-based representation is a more flexible and fast way of representing the images in a CBIR system employing multiple features.

In dissimilarity based representation, the dissimilarities between each image couple is not necessary. Instead, the dissimilarities of the images in the image database with prototype images of the system are enough. The number of prototype images is quite smaller than the size of the image database. Usually, the prototype images are grouped according to their image classes (semantic meanings of images) in order to meet semantic query requirements. In a multi-feature CBIR system, such distance values between the images in the image database and the prototype images should be stored separately for each feature.

More formally, assuming that $F = \{f_1, f_2, ..f_k\}$ is the set of features available for the CBIR system having k number of features, $C = \{c_1, c_2, .., c_m\}$ is the image database having m number of images, $P = \{P_1, P_2, ..P_n\}$ is the set of prototype image classes containing n number of image classes, each prototype image class is $P_i = \{p_1^i, p_2^i, ..p_t^i\}$ where number of prototype images is t and t is not necessarily the same in all prototype image classes; the multi-feature CBIR system owns following distance-based representation for each image class i and feature f:

$$D_f^i = \begin{pmatrix} d_f(c_1, p_1^i) & d_f(c_1, p_2^i) & \cdots & d_f(c_1, p_t^i) \\ d_f(c_2, p_1^i) & d_f(c_2, p_2^i) & \cdots & d_f(c_2, p_t^i) \\ \vdots & \vdots & \vdots & \vdots \\ d_f(c_m, p_1^i) & d_f(c_m, p_2^i) & \cdots & d_f(c_m, p_t^i) \end{pmatrix} \quad (1)$$

where $d_f(x, y)$ is the dissimilarity between the database image x and the prototype image y for feature f.

A semantic query (for instance "Find pictures of cars") executed in this kind of CBIR system is handled as follows: The distance matrices of D_f^is are evaluated, where i is the class of 'car' images and $f \in F$. First, for each matrix, prototype aggregation with a predefined algorithm is performed and an aggregated distance vector that represents the distances of all images in the image database to the 'car' semantic image class is obtained. Then k number of distance vectors, each representing a different feature, are combined with a feature selection algorithm. The combination of k number of distance vectors results with a single distance vector which shows the distances of all database images to the 'car' class.

In this study, we propose a class-specific feature selection approach for the feature selection problem stated above. The prototype aggregation problem is beyond the scope of this paper. However two different basic aggregation methods (minimum and average) are utilized during the empirical study in order to see the effect of using different aggregation techniques.

3 Exploiting Class-Specific Features

In CBIR systems, as mentioned in Section 1, a particular feature or a common set of features is usually used to compare the query image with the database images. In these systems, the features are selected to represent the problem domain. However, if the size of the database and/or the diversity of image collection is increased, these methods fail to give satisfactory results. Specifically, using the same features for different domains and types of objects yields unsatisfactory results. Finding a solution to the problem is quite simple: using different features for different object types. For example, shape features are more important than color features for a 'car' object whereas a 'sea' object can be defined with color and texture features.

To describe the approach more formally, assume an image database having images from 2 semantic classes. It is assumed that class C_1 contains n_1 number of images and C_2 contains n_2 number of images in the database. Also, it is assumed that the images of class C_1 can be defined better with color features and the images of C_2 can be defined better with shape features. If this database is used in a CBIR system that compares images according to only color features or shape features, the performance of the system is nearly 50% in terms of accuracy. If color features are used, the performance of the system is satisfactory for C_1, but not for C_2. To obtain a satisfactory performance for the whole system, different features should be used for different classes.

By considering this idea, in [25], Uysal et al. utilized an approach identifying the Best Representative Feature (BRF) for each object class, which maximizes the correct match in a training set. Similarly, in [24] Swets et al. propose to use Most Expressive Features and Most Discriminating Features. However, these approaches lacks the advantages of fusing multiple features since they select only one feature for each class.

Besides, Jain et al. [10] apply the idea in biometrics domain. They propose combining multiple traits by selecting person-specific traits for recognition. However, they do not propose a feature selection methodology. They obtain the person-specific traits after an exhaustive search process on the training data.

In this study, we propose a class-specific feature selection mechanism by finding out the representative and discriminative features for each image class. Representative characteristics of features are calculated according to the dissimilarities of images within the same class, and discriminative characteristics are calculated according to the ability of features to distinguish between different image classes. Using these characteristics, the importance values of features for each image class are calculated as detailed below. The importance values of features for each category is also called the Class-Specific Features (CSF) index. The mechanism is based on statistical calculations over the dissimilarity values of all prototype images. Providing such prototype images can be considered as the training phase of the CBIR system. The CSF indices are used as the weights of the features during feature combination process.

3.1 Calculation of CSF Indices

To calculate the CSF indices, firstly the dissimilarity values of prototype images to each other is calculated and a dissimilarity matrix is obtained as $D_f^i(P)$ for each f, similar to the one given in Section 2. Differently, $D_f^i(P)$ includes dissimilarities of prototype images in image class i to all prototype images of all image classes. $D_f^i(P)$ contains $n \cdot t$ rows and t columns.

$$D_f^i(P) = \begin{pmatrix} d_f(p_1^1, p_1^i) & d_f(p_1^1, p_2^i) & \cdots & d_f(p_1^1, p_t^i) \\ \vdots & \vdots & \vdots & \vdots \\ d_f(p_t^1, p_1^i) & d_f(p_t^1, p_2^i) & \cdots & d_f(p_t^1, p_t^i) \\ \vdots & \vdots & \vdots & \vdots \\ d_f(p_1^n, p_1^i) & d_f(p_1^n, p_2^i) & \cdots & d_f(p_1^n, p_t^i) \\ \vdots & \vdots & \vdots & \vdots \\ d_f(p_t^n, p_1^i) & d_f(p_t^n, p_2^i) & \cdots & d_f(p_t^n, p_t^i) \end{pmatrix} \tag{2}$$

After obtaining the dissimilarity matrices $D_f^i(P)$ for each feature and image class, dissimilarity values of each image category in each matrix are aggregated both column-wise and row-wise. Thus, the mean and standard deviation vectors are obtained as follows;

$$\mu(D_f^i(P)) = \left[\mu_f^{i,1} \mu_f^{i,2} \cdots \mu_f^{i,n} \right]^T \tag{3}$$

$$\sigma(D_f^i(P)) = \left[\sigma_f^{i,1}\ \sigma_f^{i,2} \cdots \sigma_f^{i,n} \right]^T \tag{4}$$

Here, $\mu_f^{i,j}$ denotes the mean of dissimilarities from all images in class i to all images in class j for feature f. Also, $\sigma_f^{i,j}$ denotes the corresponding standard deviation.

To obtain the CSF indices, four important parameters are extracted from the above given vectors of $\mu(D_f^i(P))$ and $\sigma(D_f^i(P))$:

- Mean of Class ($\mu_f^{i,i}$): $\mu_f^{i,i}$ is the average dissimilarity value of a class to itself, for a particular feature f. Mean of Class is a representative characteristic for features. For a selected class, the features with lower dissimilarity values represent the image class better. Thus, the CSF index is inversely proportional to the mean of the category.
- Standard Deviation of Class ($\sigma_f^{i,i}$): $\sigma_f^{i,i}$ is an another important representative property. For any class, a feature with small standard deviation entails close image-to-image dissimilarity values within the class. Such a feature can be considered as a better feature. Thus, the CSF index is inversely proportional to the standard deviation of an image class.
- Standard Mean Distance to Other Classes (δ_f^i): Standard mean distance to other classes is a discriminative feature which is calculated by using the dissimilarities of a class to other classes. It is calculated as follows:

$$\delta_f^i = \sqrt{\frac{\sum_{j=1}^{n} (\mu_f^{i,i} - \mu_f^{i,j})^2}{n}} \tag{5}$$

  where n is the number of image classes. This calculation gives us the average dissimilarity of an image class i to all other classes. Thus, having a greater dissimilarity means better discrimination among all categories, which means that the CSF index is directly proportional to δ_f^i.
- Correctness Ratio (ω_f^i): Although the three parameters given above are important and provide good representation and discrimination, the issue of correctness of the feature is not considered. It is important for a feature to give the lowest dissimilarity values for the images in a class which is the same with the class of the query images. Correctness ratio of a particular feature f can be defined as what percentage of the means in a $\mu(D_f^i(P))$ vector are larger than the mean value of the class i ($\mu_f^{i,i}$). As the correctness ratio decreases, the representation ability decreases, which means that the CSF index is directly proportional with the correctness ratio.

Considering the effects of the above parameters, the CSF index of a particular feature f on a particular image class i is calculated using the formula below:

$$CSF_f^i = \frac{(1 - \mu_f^{i,i}) \cdot \delta_f^i \cdot \omega_f^i}{\sigma_f^{i,i}} \tag{6}$$

3.2 Normalization on Dissimilarities

As mentioned before, CBIR system having dissimilarity-based representation does not need to deal with the intrinsic dimensionality of features to combine them. However, different scales of different features are still a problem to be solved. Different scales of the values contained in the features causes dissimilarity values to be in different scales.

In the literature, there are several normalization methods to handle the different scales of multiple features [11]: Min-max, decimal scaling, z-score, median, double sigmoid, tanh estimators, biweight estimators. In [11], Jain et al. empirically show that min-max, z-score and tanh estimators methods are superior. Also they note that the simplest method (min-max) would suffice when the minimum and maximum values are known. Min-max normalization transforms values from a known (or estimated) range $[min, max]$ into $[0, 1]$ range with the following basic formulation: $x' = (x - min)/(max - min)$. Considering that we have the prototype images and dissimilarity values of prototype images to themselves, it is easy to find the minimum and maximum dissimilarity values for each feature. Thus the min-max normalization approach is preferred in this study.

4 Evaluation

To demonstrate the validity of the proposed approach, a number of experiments are carried out. For the experiments, the CalTech 101 image dataset [9] is used. It contains pictures of objects belonging to 101 categories. During the tests, all of the 101 classes in the dataset are used. Randomly selected 10 images for each class, hence a total of 1010 images, are treated as the prototype images. For the query purposes, randomly selected 20 images for each class and a total of 2020 images are employed the image database. In addition, as the features to be combined, 8 visual features of MPEG-7 [18] in three types are utilized: Color descriptors of Color Layout(CL), Color Structure(CS), Dominant Color(DC), Scalable Color(SC); Shape descriptors of Contour Shape(CSh), Region Shape(RS); Texture descriptors of Edge Histogram(EH), Homogeneous Texture(HT). The dissimilarities of the images for these features are calculated by using the MPEG-7 reference software (eXperimentation Model, XM) [19].

The tests are mainly performed on semantic retrieval of images; the semantic classes are queried over the image database. The images are fetched and sorted according to the dissimilarity values. To measure the retrieval accuracy, *Precision*, *Recall*, *Average Precision(AP)* and *Mean Average Precision(MAP)* metrics are used. *Precision* is the fraction of retrieved images that are relevant to the search, whereas *Recall* is the ratio of the number of relevant images retrieved to the total number of relevant images in the collection. The AP is the sum of the precision at each relevant hit in the retrieved list, divided by the minimum between the number of relevant documents in the collection and the length of the list. Considering that image collection in our test contains 2020 images, AP is measured at 2020. MAP is the AP averaged over several image classes.

As the primary test, the accuracy of the proposed method on semantic retrieval is measured. In order to perform a detailed comparison, this test is executed in four steps. As the first step, the retrieval accuracies of each single feature is calculated. For the second step, following simple combination approaches are tested: Minimum Distance(MD), Average Distance(AD), Euclidian Distance(ED). The combined dissimilarity is obtained by selecting the minimum dissimilarity (distance) in MD, averaging all available dissimilarities in AD and

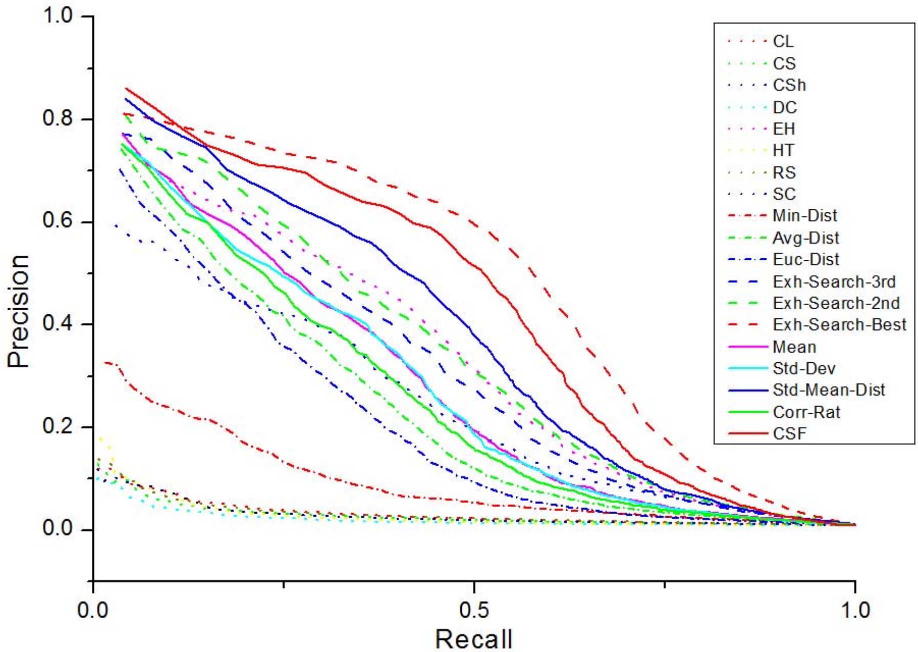

Fig. 1. Precision-Recall Graph for Semantic Retrieval

calculating an Euclidian distance on the available dissimilarities in ED. For the third step, feature selection by an exhaustive search approach is applied and the combined dissimilarity is calculated by averaging the dissimilarities of resultant features from the feature selection. An exhaustive search for feature selection requires calculating all combinations of available features, 2^8 cases in total for our test. Considering that performing an exhaustive search during each query is not applicable due to the time cost, the selection process is executed once on the prototype images. Then, 10 best selections (ES[1-10]) are found and semantic retrieval test is performed for each of these 10 feature selections. As the last step, the approach proposed in this study is performed for feature selection. Calculated CSF indices are used to combine the dissimilarity values with a weighted-sum approach. Not only the CSF index, but also the four parameters of the CSF are tested separately in order to see which one is more influential. In Figure 1, the Precision-Recall graphs of these methods are given. In addition, the AP of some sample categories, MAP of Best 10, 20, 50 and all 101 categories are presented in Table 1. Also, how many times each method has the best score and mean ranks of each method are included in the table.

Considering the test results, it is observed that obtaining an increase in the accuracy requires a good selection on the features. Simple methods like MD, AD and ED are not enough for selection. MD lacks the advantages of combining multiple features whereas AD and ED always combine all of the features and are affected by the unfavorable features. Besides, the exhaustive search

Table 1. Semantic Query Results

		electric guitar	saxophone	inline skate	stop sign	revolver	Best-10	Best-20	Best-50	All-101	Number of Best Scores	Mean Rank
Single Features	CL	0.013	0.035	0.127	0.376	0.023	0.406	0.259	0.129	0.073	0	22.6
	CS	0.028	0.135	0.037	0.388	0.021	0.361	0.241	0.119	0.066	0	23.0
	CSh	0.865	0.766	0.725	0.253	0.361	0.841	0.743	0.542	0.339	3	13.9
	DC	0.020	0.033	0.055	0.427	0.019	0.258	0.171	0.088	0.050	0	23.6
	EH	0.895	0.874	0.633	0.827	0.928	0.924	0.855	0.667	0.424	6	10.1
	HT	0.006	0.063	0.159	0.235	0.029	0.304	0.210	0.112	0.063	0	23.0
	RS	0.097	0.120	0.153	0.624	0.114	0.354	0.233	0.121	0.070	0	22.4
	SC	0.016	0.065	0.057	0.654	0.064	0.352	0.229	0.116	0.066	1	22.8
Simple	MD	0.401	0.253	0.190	0.253	0.733	0.703	0.550	0.318	0.176	0	19.1
	AD	0.029	0.614	0.734	0.863	0.615	0.813	0.716	0.493	0.310	1	14.2
	ED	0.014	0.563	0.704	0.792	0.542	0.766	0.677	0.457	0.284	0	15.9
Exh. Search	ES1	**0.958**	0.964	0.870	0.856	0.970	0.963	0.927	**0.806**	**0.563**	**36**	**5.0**
	ES2	0.841	0.919	0.763	0.917	0.830	0.895	0.820	0.630	0.418	3	9.6
	ES3	0.565	0.951	0.831	0.985	0.898	0.923	0.828	0.616	0.400	1	11.1
	ES4	0.928	0.960	0.806	0.880	0.797	0.923	0.862	0.693	0.459	5	8.0
	ES5	0.934	0.916	0.844	0.910	**0.973**	0.927	0.872	0.704	0.484	7	6.9
	ES6	0.815	0.885	0.720	0.811	0.794	0.867	0.780	0.609	0.405	4	10.5
	ES7	0.641	0.968	**0.911**	0.981	0.959	0.932	0.855	0.663	0.441	6	8.7
	ES8	0.587	0.916	0.785	0.935	0.854	0.896	0.797	0.594	0.387	2	11.3
	ES9	0.578	0.972	0.844	0.979	0.852	0.927	0.845	0.634	0.409	3	10.6
	ES10	0.746	0.942	0.886	0.981	0.841	0.926	0.864	0.714	0.482	8	7.1
Proposed	μ	0.583	0.700	0.762	0.893	0.653	0.834	0.747	0.556	0.359	2	12.3
	σ	0.174	0.878	0.786	0.959	0.803	0.876	0.787	0.567	0.362	1	11.5
	δ	0.867	0.887	0.835	0.961	0.959	0.942	0.866	0.683	0.458	4	7.4
	ω	0.315	0.617	0.734	0.862	0.640	0.817	0.722	0.518	0.333	1	13.1
	CSF	0.955	**0.981**	0.889	**0.987**	0.957	**0.970**	**0.928**	0.769	0.521	24	5.5

guarantees to find the optimal feature selection by evaluating all possible combinations. Therefore ES1 outperforms the other methods. However the ES[1-10] ranking obtained at the training phase is not the same during the querying. For instance, ES5 performs better than ES2, ES3 and ES4. Such situation is caused by difference between training and query images. Although it is not observed in this test conditions, it could be possible that the best combination obtained during the training phase do not give best results during querying. It is possible to handle such incompliance by executing the exhaustive search during each query, but it causes a time inefficiency.

On the other hand, our proposed method of CSF gives successful accuracy results that are very close to the best selection in total and even better for one fourth of the image classes. Regarding that the results of the best selection in ES can be considered as the upper-bound for the retrieval task, the CSF method can be qualified as a robust and successful approach. In addition, our claim of exploting class-specific features can be supported by the results of ES method. Different feature combinations in ES selections perform better in different image classes, which results different classes requires the use of different features.

Another observation on the results is the superiority of δ parameter of CSF approach among other parameters. Therefore, it can be stated that the discriminativeness characteristics of features are more effective than the representativeness.

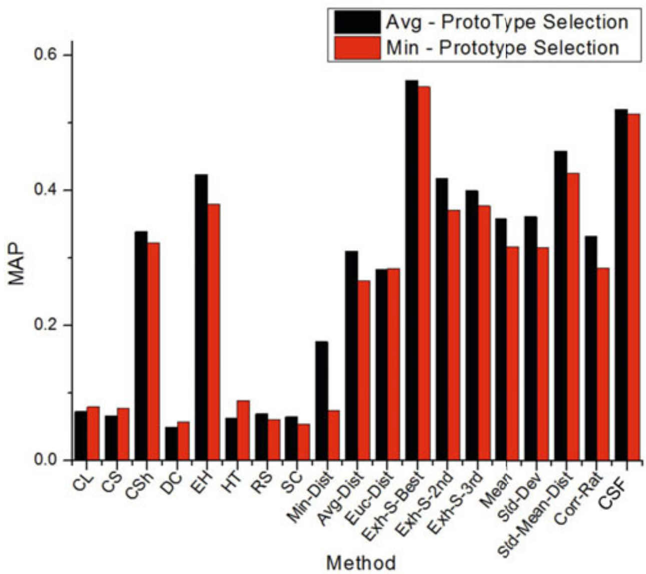

Fig. 2. Comparison of Average and Minimum Aggregation Methods

An important discussion for combining multiple features is the independency of features. Using complementary features with the methods requiring independent inputs can cause a decrease in the accuracies. Therefore, many studies exist in the information fusion literature that performs an independence analysis [16]. In this empirical study, the features utilized are not fully independent. It is previously stated that simple methods like MD, AD and ED are not successful enough for the selection task. One important reason in their inefficiency is the fact that they can not eliminate complementary information and the violation of independence assumption decreases their performance. However, the ES and CSF approaches enables selecting different combinations and eliminates complementary features.

As mentioned in Section 2, a prototype aggregation is necessary to combine the dissimilarities of multiple prototypes. Although prototype aggregation is beyond our scope, a secondary test is performed to show the effect of prototype aggregation. During the first test, *averaging* is used for aggregation. In this test, the previous test is repeated with a *minimum* aggregation method. The comparison of two methods are given in Figure 2. It is clearly shown that *averaging* is superior than *minimum*. However, these two are very simplistic methods and there are better ways of exploiting the information included in the prototypes.

As the last test, the time complexities of our proposed method and exhaustive search are compared. The query execution times of these two approaches are quite the same since querying includes only a weighted/unweighted summation of several features. However, the execution times for the training phases, which are carried out in order to find out the optimal set of features, differ much. Time complexity of exhaustive search is $O(m^2 \cdot 2^n)$ where m is the total number of prototype images and n is the number of features. Whereas, time complexity of our proposed method is $O(m^2 \cdot n)$. Time-measurements obtained in this test validated these theoretical definitions. Results are given in Table 2. The results show us that CSF approach is 50 times better than the ES approach, in our case. If the number of features increases, execution time for ES could be worse.

Table 2. Execution Times for Training Phases

	Total Execution Time
Exhaustive Search	1,049,652 msec
CSF Calculation	19,802 msec

5 Conclusion

In this paper, a class-specific feature selection approach for the fusion of multiple features is presented. In order to eliminate the high-dimensionality of multiple features and provide efficient querying over the images, a dissimilarity based approach is utilized. The class-specific features are determined by using the representativeness and discriminativeness of features for each image class. The calculations of representativeness and discriminativeness are based on the statistics on the dissimilarity values of training images. The approach is tested on CalTech 101 dataset by using 8 MPEG-7 features and compared with the single features, simple combination approaches and exhaustive search approach. Test results showed that proposed class-specific feature selection approach is a timely-efficient, accurate and robust way of feature selection.

Some further research direction can be as follows: Employing prototype selection and aggregation methods within the proposed approach, utilizing proposed approach with a dissimilarity based classification mechanism and performing multi-modal feature selection obtained from video data.

References

1. Arevalillo-Herráez, M., Domingo, J., Ferri, F.J.: Combining similarity measures in content-based image retrieval. Pattern Recogn. Lett. 29, 2174–2181 (2008)
2. Atrey, P.K., Kankanhalli, M.S., Oommen, J.B.: Goal-oriented optimal subset selection of correlated multimedia streams. ACM Trans. Multimedia Comput. Commun. Appl. 3 (February 2007)
3. Bruno, E., Marchand-Maillet, S.: Multimodal preference aggregation for multimedia information retrieval. Journal of Multimedia 4 (5), 321–329 (2009)
4. Bruno, E., Moënne-Loccoz, N., Marchand-Maillet, S.: Design of multimodal dissimilarity spaces for retrieval of multimedia documents. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(9), 1520–1533 (2008)
5. Chibelushi, C., Deravi, F., Mason, J.: Adaptive classifier integration for robust pattern recognition. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 29(6), 902–907 (1999)

6. Datta, R., Li, J., Wang, J.Z.: Content-based image retrieval: approaches and trends of the new age. In: MIR 2005: Proc. of the 7th ACM SIGMM Workshop on Multimedia Information Retrieval, pp. 253–262. ACM Press, New York (2005)
7. Duin, R.P.W., Loog, M., Pękalska, E., Tax, D.M.J.: Feature-based dissimilarity space classification. In: Ünay, D., Çataltepe, Z., Aksoy, S. (eds.) ICPR 2010. LNCS, vol. 6388, pp. 46–55. Springer, Heidelberg (2010)
8. Duin, R.P.W., Tax, D.M.J.: Experiments with classifier combining rules. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 16–29. Springer, Heidelberg (2000)
9. Fei-Fei, L., Fergus, R., Perona, P.: One-shot learning of object categories. IEEE Trans. Pattern Anal. Mach. Intell. 28(4), 594 (2006)
10. Jain, A., Ross, A.: Learning user-specific parameters in a multibiometric system. In: Proceedings of 2002 International Conference on Image Processing (2002)
11. Jain, A., Nandakumar, K., Ross, A.: Score normalization in multimodal biometric systems. Pattern Recognition 38(12), 2270–2285 (2005)
12. Jain, A.K., Duin, R.P., Mao, J.: Statistical pattern recognition: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 4–37 (2000)
13. Kankanhalli, M., Wang, J.: Experiential sampling on multiple data streams. IEEE Transactions on Multimedia 8(5), 947–955 (2006)
14. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Trans. Pattern Anal. Mach. Intell. 20, 226–239 (1998)
15. Kludas, J., Bruno, E., Marchand-Maillet, S.: Information fusion in multimedia information retrieval. In: Boujemaa, N., Detyniecki, M., Nürnberger, A. (eds.) AMR 2007. LNCS, vol. 4918, pp. 147–159. Springer, Heidelberg (2008)
16. Kludas, J., Bruno, E., Marchand-Maillet, S.: Can feature information interaction help for information fusion in multimedia problems? Multimedia Tools Appl. 42, 57–71 (2009)
17. Kuncheva, L.I.: Switching between selection and fusion in combining classifiers: an experiment. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 32(2), 146–156 (2002)
18. Martínez, J.: Mpeg-7 overview (version 10). Requirements ISO/IEC JTC1 /SC29 /WG11 N6828, International Organisation For Standardisation (October 2003)
19. MPEG: Mpeg-7 reference software experimentation model (2003), `http://standards.iso.org/ittf/PubliclyAvailableStandards/c035364_ISO_IEC_15938-6(E)_Reference_Software.zip` (Online; accessed February 01, 2011)
20. Nguyen, G.P., Worring, M., Smeulders, A.W.M.: Similarity learning via dissimilarity space in cbir. In: Proc. of the 8th ACM International Workshop on Multimedia Information Retrieval, MIR 2006, pp. 107–116. ACM, New York (2006)
21. Pekalska, E., Paclík, P., Duin, R.P.W.: A generalized kernel approach to dissimilarity-based classification. Journal of Machine Learning Research 2, 175–211 (2001)
22. Poh, N., Kittler, J.: Multimodal Information Fusion: Theory and Applications for Human-Computer Interaction. 8, pp. 153–169. Academic Press, London (2010)
23. Snidaro, L., Niu, R., Foresti, G., Varshney, P.: Quality-based fusion of multiple video sensors for video surveillance. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 37(4), 1044–1051 (2007)
24. Swets, D.L., Weng, J.J.: Shoslif-o: Shoslif for object recognition and image retrieval (phase ii). Tech. Rep. CPS 95-39, Michigan State University, Department of Computer Science (1995)

25. Uysal, M., Yarman-Vural, F.T.: Selection of the best representative feature and membership assignment for content-based fuzzy image database. In: CIVR, pp. 141–151 (2003)
26. Wu, Y., Chang, E.Y., Chang, K.C.C., Smith, J.R.: Optimal multimodal fusion for multimedia data analysis. In: Proc. of the 12th Annual ACM International Conf. on Multimedia, MULTIMEDIA 2004, pp. 572–579. ACM, New York (2004)
27. Xu, L., Krzyzak, A., Suen, C.: Methods of combining multiple classifiers and their applications to handwriting recognition. IEEE Transactions on Systems, Man and Cybernetics 22(3), 418–435 (1992)