A Measure of Perceptual Aliasing in Image Descriptors

Jiemin Wang, Hong Zhang
Department of Computing Science
University of Alberta
Edmonton, Canada
Email: {*Jiemin2, hzhang*}*@ualberta.ca*

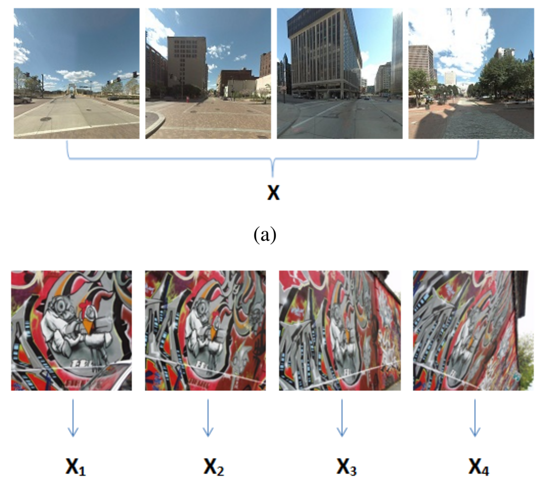
***Abstract*—Researchers exploring problems in image matching tasks face the curse of perceptual aliasing that is originally used in characterizing a sensing process. Perceptual aliasing occurs when the one-to-one mapping relations between world states (objects) and their representations (descriptors) are not maintained. In this paper, we introduce a novel method for quantifying perceptual aliasing. Our method measures the discriminating power of an image descriptor in terms of its ability to distinguish between images of different objects and to match images of the same object. To illustrate our method, we apply it in the evaluation of popular global image descriptors that do not require local feature or keypoint detection. Specifically, our method runs spectral clustering on the similarity matrix computed with descriptors of known image clusters and measures the performance of an image descriptor by its ability to maintain the original clusters, using two indices, MRI-1 and MRI-2, that are based on Rand index.**

***Keywords*-Perceptual aliasing; performance metric; image descriptors; discriminating power.**

I. Introduction

Perceptual aliasing used in characterizing a sensing process was first defined in 1991 [1]. The concept has been used in reinforcement learning, network sensor, psychology, etc. In image matching tasks, perceptual aliasing occurs when the one-to-one mapping between the world states (objects) and the internal representations (descriptors) is not maintained. Two types of perceptual aliasing can occur [1]. First, multiple world states share the same internal representation; e.g., images of different objects are perceived the same according to the image descriptor. Second, one world state has more than one internal representation; e.g., images of the same object are described by different image descriptors. Simple examples of the two types of perceptual aliasing are illustrated in Figure 1.

Extensive literature exists on comparison of keypoint descriptors [2]–[8] which include SIFT [9], SURF [10] and the more recent BRIEF [11], and the focus of these studies is often on the invariance properties in regard to scale, rotation, illumination, etc. Performance is measured in terms of repeatability of the keypoints, and the precision and the recall of matching detected keypoints. Other evaluation techniques are used in the context of an application where each competing alternative is substituted in turn and an appropriate performance metric is defined – such as how often



(b)

Figure 1. Examples of two types of perceptual aliasing. (a) Type 1, images of multiple world states are perceived the same with descriptor vector X. (b) Type 2, one world state has multiple internal representation, i.e., the descriptor vectors X_i, where $X_i \neq X_j$ if $i \neq j$.

a query image finds a correct match among the top x-percent of the retrieved images in an image retrieval application [12] – as the basis for choosing the optimal image descriptor for this application. However, these performance metrics are not defined to capture the two types of perceptual aliasing and they may ignore the similarities and differences between each image pair but only focus on the similarities between the query image and the top retrieved ones.

In this paper, we offer a novel measure of perceptual aliasing in an image descriptor. Our goal is a simple method for quantifying perceptual aliasing so as to compare image descriptors and assist in the process of descriptor selection. Given the dataset within a specific application, the performance can be evaluated if the similarity can be computed between image pairs. Specifically, an optimal image descriptor should minimize perceptual aliasing when it is used to compare images, i.e., similar images remain similar in the descriptor space and vice versa. To quantify our comparative analysis, we borrow the technique in clustering analysis of using Rand index [13] to evaluate the clustering results and define two perceptual aliasing corresponding to the two types of perceptual aliasing, referred to as modified Rand index 1 and 2 (MRI-1 and MRI-2).

The rest of this paper is organized as follows. Section II reviews the related performance comparison methods for image descriptors people have used to measure perceptual aliasing. Section III gives the detailed description of the proposed method. Experimental results and explanations are shown in Section IV. The paper is concluded with a discussion in Section V.

II. Related Work

When an image is described, perceptual aliasing introduced by the description step can be indirectly determined by the precision and the recall computed from matching the image descriptors. The seminal work of Mikolajczyk and Schmid [2] presented perhaps the most popular framework for comparing the performance of keypoint descriptors. They carefully constructed image datasets under various transformations including affine, compression, blur, scale, rotation and illumination, and measured the performance of keypoint detectors in terms of repeatability and that of keypoint descriptors in terms of precision and recall. Unfortunately, it may not be appropriate to quantify perceptual aliasing using precision and recall because they are designed to measure the discriminating power of image descriptors on distinguishing images of different objects and matching images of the same object.

Rather than precision and recall, other metrics have been developed for evaluating the performance of an image descriptor, and they are almost exclusively always in the context of specific applications such as content-based image retrieval or place recognition. These metrics have their own focuses in terms of what is important to measure. To measure the quality of the image retrieval result, average normalized rank of relevant images, with values ranging between 0 and 1, was proposed in [14]. This metric was later used in the video Google work [15] by Sivic and Zisserman to analyze the performance of their visual BoW image representation. The query images have to be substituted each time and the retrieval has to be run repeatedly to evaluate the performance of image descriptors. Another possible performance metric for image retrieval was based on top relevant images [12], motivated by the intuition that the correct images have to be at the top of the matched list for the image descriptor and the retrieval algorithm in general to be effective. In this case, the performance is defined in terms of the percentage of ground truth images that are among the top x percent of the returned images. Generalizing precision and recall, [16] proposed mean average precision (mAP) and recall@R to compare the performance of the proposed compact global image descriptor with the state-of-the-art image descriptors in large scale image retrieval. Intuitively, mAP computes the average area under the precision-recall curve for a set of queries, thereby eliminating the need for choosing recall levels. Recall@R, on the other hand, computes the recall for the first R returned images, in order to overcome the difficulty in setting different recall levels for the high number of returned images in large scale image retrieval.

In contrast to all the previous studies, our performance indices focus on perceptual aliasing and quantify it directly. Since image description is a process of transforming an image from intensity space to a descriptor space, perceptual aliasing happens when dissimilar images are mapped to similar descriptors, or when images of the same objects or places under various transformations such as view point, scale, or illumination changes are mapped to different descriptors. This leads to the idea of constructing a dataset with similar images within individual clusters but with dissimilar images between clusters. We can then quantify perceptual aliasing by clustering analysis and computing the ratio of image pairs belonging to different objects falling into the same image cluster and the ratio of image pairs belonging to the same object dropping into different image clusters. We choose to use spectral clustering due to the general superior performance of spectral clustering over, for example, k-means, and in order to handle any image description techniques in which one can calculate the similarity between two images even when a descriptor space is of no fixed dimension nor explicitly available. The results of our method can help choose the optimal image descriptor within a specific application when the dataset used in calculating the indices consists of images from that application.

III. Method

In this section, we will describe the proposed method in detail. To investigate the extent of perceptual aliasing caused by a particular image descriptor, we first construct a similarity matrix with respect to many clusters of images such that images within a cluster are similar and images between clusters are dissimilar. This can be done easily by grouping consecutive images of a video into clusters – assuming the camera captures the images continuously. The similarity matrix can be computed by matching local feature/keypoint descriptors like SIFT and SURF or global image descriptors like GIST and HOG. We will then explain the two perceptual aliasing indices.

A. Similarity Matrix Construction

Each element s_{ij} in similarity matrix S indicates the similarity between two data points. To construct a similarity matrix, we collect images to create a dataset such as those shown in Figure 2 that includes street views of the city of Pittsburgh. We include c adjacent images of the video in a cluster and select m clusters in total. Images in the same cluster contain similar scenes and images in different clusters do not overlap by requiring each cluster every $n \gg c$ images.

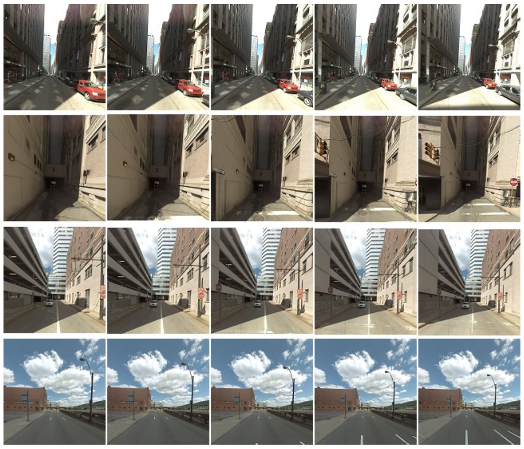


Figure 2. Four image clusters from the dataset. Each cluster contains 5 adjacent images taken from the back perspective. Images in the same cluster contain the same scene and there is no overlapping between two clusters.

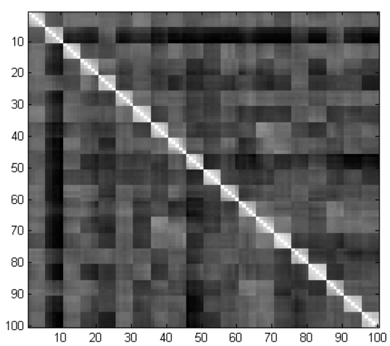


Figure 3. Similarity matrix computed with GIST descriptors. $m = 20$ and $c = 5$. Similarity matrix is of size $mc \times mc = 100 \times 100$.

The similarity can be computed by matching the local feature/keypoint descriptors or calculating the distance between two global image descriptors if the image descriptors are explicit and of the same length such as GIST. As a result, the similarity matrix S is of size $mc \times mc$. An example similarity matrix, computed from the GIST image descriptors with 20 clusters and five images per cluster, is plotted in Figure 3. Since images of the same cluster contain similar scenes, elements near the diagonal form $c \times c$ bright blocks indicating much higher similarity within each block than between blocks. Due to perceptual aliasing, however, many off-diagonal elements of the similarity matrix have high values. This common weakness is shared by all image descriptors to various extents.

B. Perceptual Aliasing Measurement

With the similarity matrix, we are then able to group images according to their descriptors using spectral clustering and perform an analysis of perceptual aliasing based on the clustering results. Spectral clustering works on the spectrum of the similarity matrix and is superior in solving general clustering problems even when the clusters are not convex sets [17], [18]. With the clustering results, since the ground truth clusters are known in our dataset, we are then able to measure the performance an image descriptor by examining how well the clusters have been retained.

We resort to Rand index to quantify perceptual aliasing [13]. Rand index was originally defined to compare the similarity between two clustering results. Specifically, given two partitions of a set S, $X = \{X_1, \ldots, X_r\}$ and $Y = \{Y_1, \ldots, Y_l\}$, Rand index is defined as,

$$R = \frac{a+b}{a+b+c+d} \tag{1}$$

where a refers to the number of element pairs in S that are in the same subset of X and in the same subset of Y, b refers to the number of element pairs in S that are in different subsets of X and in different subsets of Y, c indicates the number of element pairs in S that are in the same subset of X but in different subsets of Y, d indicates the number of element pairs in S that are in different subsets of X but in the same subset of Y. Rand Index measures the similarity between two partitions of a set. In our case, the ground truth image clusters $X = \{X_1, \ldots, X_r\}$ and the result image clusters $Y = \{Y_1, \ldots, Y_l\}$ from spectral clustering are two partitions of the whole dataset. The first type of perceptual aliasing denotes images belonging to different objects but have similar image descriptors, as a result of which, the result image cluster $Y_i \in Y$ contains images of multiple objects. We defined the modified Rand index (MRI) for the first type of perceptual aliasing as,

$$\text{MRI-1} = \frac{1}{l}\Sigma_{i=1}^{l}\frac{d_i}{\text{C}(|Y_i|, 2)} \tag{2}$$

where $\text{C}(|Y_i|, 2)$ indicates the number of 2-combinations of the set Y_i, $|Y_i|$ denotes the cardinality of the set Y_i and d_i is the number of image pairs that are in different subsets of X but in the same subset of Y. MRI-1 takes the average over the ratios computed on each subset $Y_i \in Y$ in the range $[0, 1]$. The smaller the MRI-1 value of an image descriptor, the superior the ability of this descriptor against the first type of perceptual aliasing.

The second type of perceptual aliasing occurs when images of the same object are mapped to dissimilar image descriptors, i.e., images that belong to X_i are clustered into different subsets of Y. MRI-2 is defined to characterize this type of perceptual aliasing as,

$$\text{MRI-2} = \frac{1}{r}\Sigma_{i=1}^{r}\frac{c_i}{\text{C}(|X_i|, 2)} \tag{3}$$

where c_i is the number of image pairs that are in the same subset of X but different subsets of Y. Similar to MRI-1, MRI-2 then takes the average over the ratios computed on each subset $X_i \in X$ and lies in the range $[0, 1]$. The smaller the MRI-2, the less severe the second type of perceptual aliasing in the image descriptor.

TThe overall process of computing MRI-1 and MRI-2 is summarized in Algorithm 1. The number of image clusters k is set from 2 to K, where $K \leq m$, so that we are able to observe the trend of perceptual aliasing with more image clusters. For each k, the corresponding similarity matrix $S^{kc \times kc}$ is extracted from the original similarity matrix $S^{mc \times mc}$. C_X and C_Y are vectors of the result cluster labels and the ground truth cluster labels of the images. PA is the method to compute MRI-1 and MRI-2 according to Equations (2, 3) given the cluster labels. The experimental steps are then repeated with each k for max_iter times and the final results are the average over the results of all iterations.

Algorithm 1 A method to measure perceptual aliasing

Require: $S \leftarrow$ similarity_matrix, $K \leftarrow$ #image_cluseters, $t \leftarrow 1$

MRI-1$_table = zeros(max_iter,\ k)$;
MRI-2$_table = zeros(max_iter,\ k)$;
for $k = 2 \rightarrow K$ **do**
  while $t \leq max\_iter$ **do**
    $S^{kc \times kc}$ = sub similarity matrix corresponds to randomly selected k image clusters;
    $C_Y = Spectral\_clustering(S^{kc \times kc}, k)$;
    $[\text{MRI-1}, \text{MRI-2}] = PA(C_X, C_Y)$;
    MRI-1$_table(t,\ k) =$ MRI-1;
    MRI-2$_table(t,\ k) =$ MRI-2;
  end while
end for
MRI-1$_result = mean($MRI-1$_table, 1)$;
MRI-2$_result = mean($MRI-2$_table, 1)$

IV. Experimental Results

In this section, we evaluate and compare the performance of the common global image descriptors in terms of MRI-1 and MRI-2 defined in the previous section. We will first give a brief introduction to the image descriptors and the parameters used in creating the descriptors and constructing the similarity matrices. One evaluation is concerned with different versions of the global BRIEF descriptor and our result will demonstrate that the conclusion from our method is consistent with the literature. To illustrate the utility of our method in image descriptor comparison, we will run experiments on two image datasets, the Pittsburgh Street View dataset and the Benchmark dataset[1], to examine image descriptors within location recognition and image retrieval applications.

[1]Benchmark dataset: http://www.vis.uky.edu/ stewe/ukbench/

A. Image Descriptors

We evaluated perceptual aliasing that exists in several common global image descriptors, which do not require local feature or keypoint detection, including: BRIEF-Gist, GIST, WI-SIFT and WI-SURF. We also ran experiments with the local SIFT descriptor [9], [19] since it shows outstanding performance and serves as a baseline to observe the performance of global image descriptors intuitively and objectively. The similarity matrix for SIFT descriptors is created by matching the local feature/keypoint descriptors.

BRIEF-Gist [20] is a global image descriptor based on the idea of the binary image descriptor BRIEF [11]. Rather than extracting local binary image descriptors by comparing the intensity of randomly chosen local keypoint pairs, BRIEF-Gist downsamples the original image to proper descriptor patch size and builds the binary image descriptor for the whole image. To preserve information of the original image and create more distinctive image descriptor, BRIEF_$n \times n$ divides an image into $n \times n$ tiles, builds a descriptor for each tile and concatenates the $n \times n$ short descriptors into one image descriptor. In our experiments, we set the size of each tile to be 60×60 pixels after the whole image is resized to $60 \cdot 7 \times 60 \cdot 7$ pixels as in [20]. The similarity is computed with hamming distance.

WI-SIFT and WI-SURF [21] are the global versions of the corresponding local normalized keypoint descriptors SIFT and SURF, where ”WI” stands for ”whole image”. In this case, we downsample the images to 128×128 pixels, take the center point of the image as a keypoint and create a SIFT or SURF image descriptor describing the characteristics of its neighbourhood. The similarity between WI-SIFT or WI-SURF image descriptors are calculated with Euclidean distance in our experiments, although others could be used as well. The similarity is normalized with ‘1’indicating perfectly matched and ‘0’indicating least matched.

GIST [22] is another global image descriptor that describes the characteristic of an image in terms of its response to a Gabor filter bank. Again, Euclidean distance is used to compute the similarity and normalization is applied in the same way with WI-SIFT and WI-SURF.

B. Comparison Results

Given the similarity matrices computed with the common global image descriptors in our study, we set the maximum number of image clusters K to 50 so that k is from 2 to 50. Practically, it is difficult to determine the range of k. However, since we are interested in measuring the relative performance against perceptual aliasing of different image descriptors, we can just set the k from 2 to 50 and observe the comparison results with different numbers of image clusters. The maximum number of iterations max_iter has been set to 100 since the perceptual aliasing can be different within different combinations of image clusters and the results can be averaged to show the overall performance.

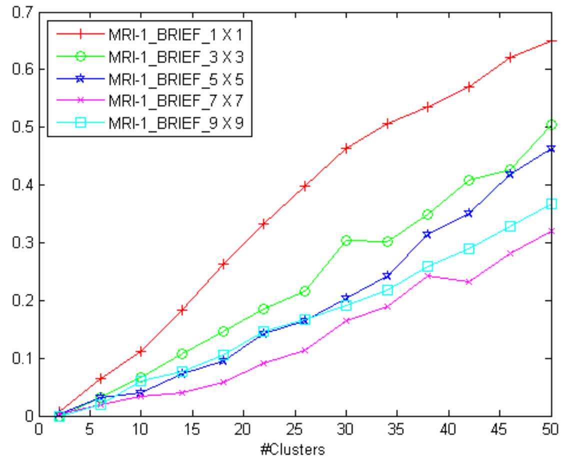


Figure 4. Comparison of the performance against perceptual aliasing of global BRIEF descriptors with MRI-1 on Pittsburgh Street View dataset.

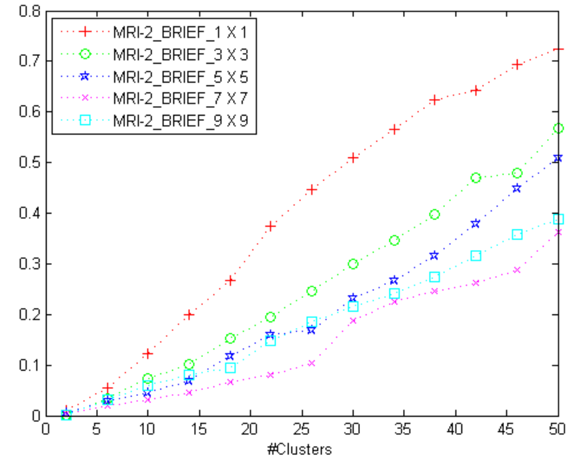


Figure 5. Comparison of the performance against perceptual aliasing of global BRIEF descriptors with MRI-2 on Pittsburgh Street View dataset.

The number of images per cluster c is 5, and $m = 100$ image clusters were selected from the Pittsburgh Street View dataset.

To examine our method, we ran the experiments to evaluate perceptual aliasing using different versions of global BRIEF descriptors. [20] proves that BRIEF with 7×7 tiles achieves the better performance than a smaller number of tiles. With a larger number of tiles, the performance will also not improve. Our results of the relative performance of different versions of BRIEF are shown in Figure 4 and Figure 5 where one can observe that they are consistent with that reported in [20]. However, our method provides specific performance evaluation in regard to the two types of perceptual aliasing.

We ran the experiments on Pittsburgh Street View dataset and Benchmark dataset. Figure 6 and Figure 7 show the experimental results on Pittsburgh dataset. Local SIFT shows the best performance among the tested image descriptors, which is not unexpected since a local keypoint descriptor is capable of preserving more information in general. BRIEF-Gist was recently designed and applied in SLAM problem where the dataset is similar to the dataset we used in our experiments. Figure 6 and Figure 7 show that BRIEF achieves better performance than the other global image descriptors. Following BRIEF-Gist are WI-SIFT and WI-SURF, which are quite similar in performance in terms of perceptual aliasing, perhaps due to the similar way in which they describe an image with SURF aiming to speed up the computation. GIST was originally proposed to represent the global structure of the scene and was used to describe the naturalness of the image content, such as buildings, mountains, trees, etc. In this case, the Street View dataset contains images of an urban city environment, which a similar degree of naturalness that GIST captures. Therefore, it is difficult for GIST to distinguish and match images in this dataset.

Figure 9 and Figure 10 show the experimental results on Benchmark dataset that contains image clusters of different objects, such as shoes, CD covers, clocks, etc. This dataset is different from the Pittsburgh Street View dataset and each cluster has 4 images as shown in Figure 8. We selected 100 image clusters. The other parameters are the same as those of the previous experiments. The local feature descriptor again achieves the best performance as expected. Among the global image descriptors, WI-SURF gains the best performance, followed by WI-SIFT, GIST and BRIEF_7×7. It is interesting to observe that the rank of image descriptors is different from that on the Street View dataset, and there is a simple explanation. Since this dataset contains many indoor images with varying amounts of blur and under different lighting conditions, WI-SURF is expected to be robust to these variations. BRIEF_7×7 however is not invariant to rotation changes and unable to handle many of the images in the dataset that experience rotational changes. GIST also seems have difficulty with this dataset. It performs better than BRIEF in MRI-1 but worse with large number of image clusters in MRI-2.

The results in this section show that our performance indices, MRI-1 and MRI-2, can be used to select image descriptors in different applications. For example, in SLAM, the robot will easily lose itself if images of multiple locations look similar to each other in the descriptor space, i.e., if MRI-1 is very serious with the description of an image descriptor, the localization mission will fail. Better ability of an image descriptor against MRI-1 is preferred in this situation. On the other hand, in image retrieval, it is important for

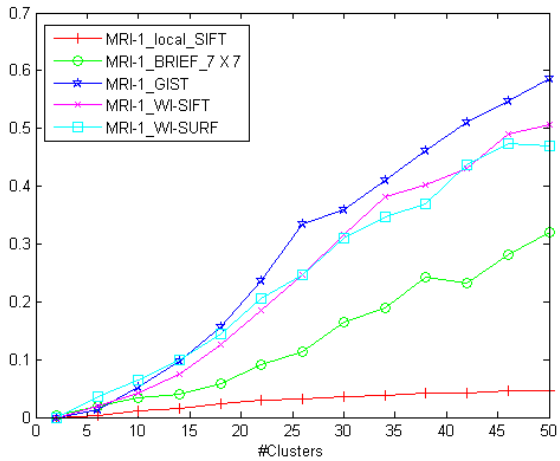


Figure 6. Comparison of the performance against perceptual aliasing of the common image descriptors with MRI-1 on Pittsburgh Street View dataset.

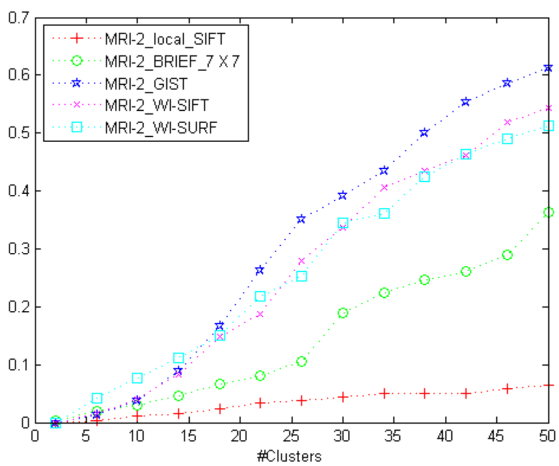


Figure 7. Comparison of the performance against perceptual aliasing of the common image descriptors with MRI-2 on Pittsburgh Street View dataset.

the matching images to be at or near the top in the ranked retrieved results, i.e., the robustness against the second type of perceptual aliasing should instead be favoured.

V. Conclusion

In this paper, we have presented a pair of novel performance metrics for comparing and measuring perceptual aliasing in image descriptors. The metrics are formulated directly from the definition of perceptual aliasing and calculated in a similar way to Rand index, used popularly in comparing clustering results. We have also introduced a procedure in which images in the intended application are selected and organized into clusters and a similarity matrix constructed using descriptors of these images. With the help

Figure 8. Four image clusters from Benchmark dataset. Each cluster contains 4 images of different changes in scale, rotation, illumination, blur, etc.

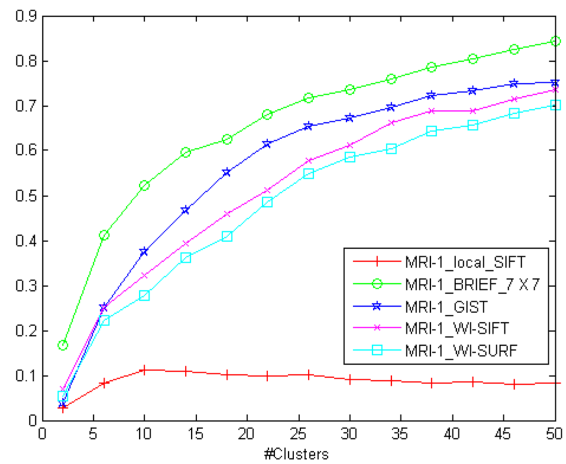


Figure 9. Comparison of the performance against perceptual aliasing of the common image descriptors with MRI-1 on Benchmark dataset.

of spectral clustering and comparison with the ground truth clusters, we are thus able to evaluate the performance against perceptual aliasing of any descriptors. We should add that our method is applicable even when image descriptor is of no fixed length or explicitly available.

We have demonstrated the reliability and usefulness of our method by using different common global image descriptors. The comparison results from our method are consistent with the literature but provide more detailed conclusion concerning perceptual aliasing. Furthermore, the proposed method is efficient as it does not involve an application. The results can be taken as an important reference when choosing the appropriate image descriptor for a specific application.

Our future work includes the comparison of our MRI metrics with other existing performance metrics for image descriptor evaluation. In particular, we are working on demonstrating that the MRI metrics are able to capture

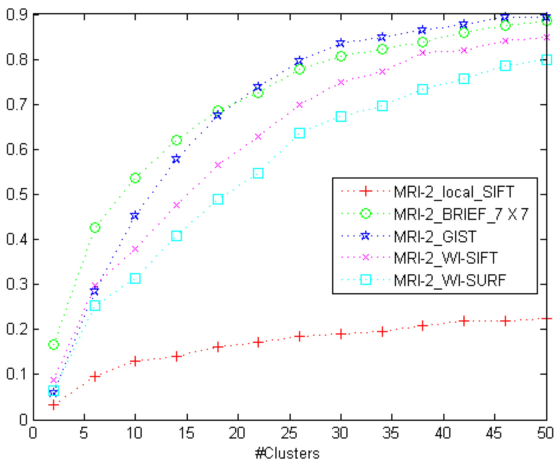


Figure 10. Comparison of the performance against perceptual aliasing of the common image descriptors with MRI-2 on Benchmark dataset.

important descriptor properties that other common metrics such as precision and recall fail to capture.

REFERENCES

[1] S. D. Whitehead and D. H. Ballard, "Learning to perceive and act by trial and error," *Machine Learning*, vol. 7, no. 1, pp. 45–83, Jul. 1991.

[2] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[3] S. Gauglitz, T. Höllerer, and M. Turk, "Evaluation of interest point detectors and feature descriptors for visual tracking," *International Journal of Computer Vision*, vol. 94, no. 3, pp. 335–360, Sep. 2011.

[4] T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: an experimental comparison," *Information Retrieval*, vol. 11, no. 2, pp. 77–107, Apr. 2008.

[5] E. Hörster, T. Greif, R. Lienhart, and M. Slaney, "Comparing local feature descriptors in plsa-based image models," in *Proceedings of the 30th DAGM symposium on Pattern Recognition*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 446–455.

[6] F. Viksten, P.-E. Forssén, B. Johansson, and A. Moe, "Comparison of local image descriptors for full 6 degree-of-freedom pose estimation," in *Proceedings of the 2009 IEEE international conference on Robotics and Automation*, 2009, pp. 1139–1146.

[7] J. Li and N. M. Allinson, "A comprehensive review of current local features for computer vision," *Neurocomputing*, vol. 71, no. 10-12, pp. 1771–1787, Jun. 2008.

[8] V. Chandrasekhar, D. M. Chen, A. Lin, G. Takacs, S. S. Tsai, N.-M. Cheung, Y. A. Reznik, R. Grzeszczuk, and B. Girod, "Comparison of local feature descriptors for mobile visual search." in *ICIP*. IEEE, 2010, pp. 3885–3888.

[9] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision*, 1999, pp. 1150–1157.

[10] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[11] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: binary robust independent elementary features," in *ECCV*, 2010, pp. 778–792.

[12] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *CVPR*, 2006, pp. 2161–2168.

[13] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, 1971.

[14] H. Müller, S. Marchand-Maillet, and T. Pun, "The truth about corel - evaluation in image retrieval," in *Proceedings of the International Conference on Image and Video Retrieval*, 2002, pp. 38–49.

[15] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proceedings of the International Conference on Computer Vision*, vol. 2, Oct. 2003, pp. 1470–1477.

[16] H. Jégou, M. Douze, C. Schmid, and P. Pérez, "Aggregating local descriptors into a compact image representation," in *CVPR*, Jun 2010, pp. 3304–3311.

[17] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 888–905, 1997.

[18] U. Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[20] N. Sunderhauf and P. Protzel, "Brief-gist - closing the loop by simple means." in *IROS*. IEEE, 2011, pp. 1234–1241.

[21] H. Badino, D. F. Huber, and T. Kanade, "Real-time topometric localization." in *ICRA*, 2012, pp. 1635–1642.

[22] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, pp. 145–175, 2001.