

# UNSUPERVISED OBJECT MATCHING AND CATEGORIZATION VIA AGGLOMERATIVE CORRESPONDENCE CLUSTERING

Md. Shafayat Hossain, Ahmedullah Aziz and Mohammad Wahidur Rahman

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh  
rumi3.1416@gmail.com

## **ABSTRACT**

*This paper presents computationally efficient object detection, matching and categorization via Agglomerative Correspondence Clustering (ACC). We implement ACC for feature correspondence and object-based image matching exploiting both photometric similarity and geometric consistency from local invariant features. Object-based image matching is formulated here as an unsupervised multi-class clustering problem on a set of candidate feature matches linking maximally stable external regions features and scale invariant features in the framework of hierarchical agglomerative clustering. The algorithm is capable to handle significant amount of outliers and deformations such as scaling and rotation as well as multiple clusters, thus powering simultaneous feature matching and clustering from real-world image pairs with significant clutter and multiple objects. The experimental assessment on feature correspondence, object recognition, and object-based image matching demonstrates that, this method is capable of rigorously handling scaling, rotation, and deformation and can be applied to a wide range of image matching and object recognition and categorization related real-world problems.*

## **KEYWORDS**

*Agglomerative Correspondence Clustering, MSER, Object matching, SIFT & UNN*

## **1. INTRODUCTION**

Object matching between unsupervised real-world images is examined in this paper. Direct image matching in an unsupervised setup has been a perplexing and thought-provoking open problem in present computer vision research. In such cases, establishing reliable feature-level correspondences is hindered by many outliers and abstruse matches, thereby requiring an approach interleaved with discovering high-level correspondences. A simultaneous feature matching and clustering method is adopted to address the interconnected problems of feature matching, object matching, and outlier elimination in an effective and integrated way for our purpose. The problem of object-based image matching is formulated as an unsupervised multi-class clustering problem on a set of candidate feature correspondences. A robust dissimilarity measure based on local invariant features is developed, and a novel matching algorithm is applied in the framework of hierarchical agglomerative clustering. The algorithm controls both photometric and geometric feature-consistency. It can also handle a significant amount of outliers and deformation, as well as multiple clusters. The experimental evaluation of feature matching, object recognition and categorization, and object-based image matching concludes that the applied method is a powerful and efficient tool for a wide range of image matching problems.

In this paper two standard datasets are used for object matching and categorization using ACC. Then some real world images are matched by means of object-matching.

## 2. METHODOLOGY

The object matching problem can be formulated as unsupervised multi-class clustering of candidate feature correspondences as follows. We consider two sets of features  $P$  and  $Q$ , each obtained from an image, and a set of candidate correspondences  $L \subset P \times Q$ . For each candidate correspondence  $c_i = (p, q) \in L$ , a unary dissimilarity score determines the dissimilarity between  $p \in P$  and  $q \in Q$ . For each pair of correspondences  $(c_i, c_j)$  where  $c_i, c_j \in L$ , the geometric deformation between  $c_i$  and  $c_j$  is evaluated by the pair-wise dissimilarity score. The target is to formulate clusters of mutually coherent feature correspondences eliminating outliers from  $L$ . The solution can be denoted by a mutually disjoint set of correspondence clusters  $\theta = \{C_1, C_2, \dots, C_n\}$  where  $\forall C_k \in \theta, C_k \subset L$ . Mapping constraints used in other methods appeared in literature [1], [2], [3] can also be employed here, such as: one-to-one constraint permitting one feature from  $P$  to match at most one feature from  $Q$ , or one-to-many constraint permitting one feature from  $P$  to match more than one features from  $Q$ . This method allows arbitrary cross- mapping.

Dissimilarity between two candidate feature matches can be regarded as the internal force driven feature correspondence and clustering. In this work, we use MSER detectors [4], [5] for features, and SIFT [6] for descriptors, which are widely used in the literature. Thus, we represent a candidate feature correspondence  $c_i = (p, q) \in L$  as a candidate match  $m_i = (x_i, x'_i, H_i)$ , interchangeably, where  $x_i, x'_i, H_i$  denotes center of  $p$ , center of  $q$ , and homography from  $p$  to  $q$ , respectively [5].

Photometric dissimilarity  $d_{app}(m_i)$  of a feature correspondence  $c_i = (p, q)$  can be defined by the Euclidean distance between corresponding SIFT descriptors of feature  $p$  and  $q$ . We can assume that two features of a true correspondence will have similar local descriptors. Therefore, a set of candidate correspondences  $L$  is usually formulated using this photometric dissimilarity. Geometric dissimilarity  $d_{geo}(m_i, m_j)$  between two feature correspondences  $m_i = (x_i, x'_i, H_i)$  and  $m_j = (x_j, x'_j, H_j)$  is defined by

$$d_{geo}(m_i, m_j) = \frac{1}{2} (d_{geo}(m_j | m_i) + d_{geo}(m_i | m_j)) \quad (1)$$

$$d_{geo}(m_i, m_j) = \frac{1}{2} (|x'_j - H_i x_j| + |x_j - H_i^{-1} x'_j|) \quad (2)$$

$$d_{geo}(m_i, m_j) = \frac{1}{2} (|x'_i - H_j x_i| + |x_i - H_j^{-1} x'_i|) \quad (3)$$

where  $|\cdot|$  denotes the Euclidean distance function. It corresponds to a projection error, which will be small if  $H_i$  and  $H_j$  are similar to each other. Exploiting the homographies of two correspondences, this geometric dissimilarity provide a discriminative measure for this agglomerative algorithm.

Combining the photometric and geometric dissimilarity, we define an overall pairwise dissimilarity function as

$$d(m_i, m_j) = d_{geo}(m_i, m_j) + \alpha \max(d_{app}(m_i), d_{app}(m_j)) \quad (4)$$

where  $\alpha$  is a weighting factor for expressing photometric dissimilarity. The max operator means that both of two correspondences in the pair should have low dissimilarity for a confident correspondence pair.

## 2.1. Hierarchical Agglomerative Clustering

Hierarchical Agglomerative clustering is a widely used algorithm [7], [8]. The single-link algorithm defines the cluster dissimilarity function  $D$  as the minimum among all pairwise dissimilarities between elements of two clusters. Elongated or connected clusters can be recovered using this method. The single-link dissimilarity between two correspondence clusters  $C_a$  and  $C_b$  is represented according to our problem by

$$D_{sl}(C_a, C_b) = \min_{i \in C_a, j \in C_b} d(m_i, m_j) \quad (5)$$

The complete-link dissimilarity via complete-link algorithm [8]

$$D_{cl}(C_a, C_b) = \max_{i \in C_a, j \in C_b} d(m_i, m_j) \quad (6)$$

The average-link dissimilarity through average-link algorithm is [8]

$$D_{av}(C_a, C_b) = \frac{1}{|C_a||C_b|} \sum_{i \in C_a} \sum_{j \in C_b} d(m_i, m_j) \quad (7)$$

The complete-link and average-link are capable of handling chaining effects but ineffective against decentralized clusters.

## 2.2. Adaptive Partial Linkage Model

The kNN linkage model to adapt the connectedness of the elements

$$D_{kNN}(k, C_a, C_b) = \frac{1}{|\Gamma|} \min \sum_{(i,j) \in \Gamma} d(m_i, m_j) \quad (8)$$

$$\text{Such that } \Gamma \subset C_a \times C_b \quad |\Gamma| = \min(k, |C_a||C_b|)$$

$|C_a|$  and  $|C_b|$  are number of elements in the corresponding clusters.  $|C_a||C_b|$  represents number of possible pairs between two clusters.  $k$  minimum dissimilarity among all dissimilarities is used in this model which is robust against chaining effect.

To avoid asymptotic chaining effect, adaptive partial linkage model is used [9].

$$D_{AP}(C_a, C_b) = D_{kNN}(k, C_a, C_b) \quad (9)$$

$$\begin{cases} k = k_{AP} \text{ if } |C_a||C_b| \leq \frac{k_{AP}}{\tau_{AP}} \\ k = \tau_{AP} |C_a||C_b| \text{ if } |C_a||C_b| > \frac{k_{AP}}{\tau_{AP}} \end{cases} \quad (10)$$

$k_{AP}$  and  $\tau_{AP}$  are the control parameters.

### 2.3. Agglomerative Correspondence Clustering

The algorithm is adopted from Minsu Cho et al. [9].

Algorithm 1. Agglomerative Correspondence Clustering process

<p>1: (Initialization)</p> <p>2: Establish a set of candidate correspondences</p> <p><math>L = \{c_1, c_2, \dots, c_n\}</math></p> <p>3: Set <math>i</math>th cluster <math>C_i = \{c_i\}</math> (<math>i = 1, \dots, n</math>)</p> <p>4: <math>\theta_0 = \{C_1, C_2, \dots, C_n\}</math></p> <p>5: (Agglomerative Clustering)</p> <p>6: <b>repeat</b></p> <p>7:   <math>t = t + 1</math></p> <p>8:   Find <math>C_a, C_b</math></p> <p>      Such that. <math>D_{AP}(C_a, C_b) = \min_{r,s} D_{AP}(C_r, C_s)</math></p> <p>9:   Merge <math>C_a</math> and <math>C_b</math> into a single cluster <math>C_q</math></p> <p>10:   Remove from <math>\theta_{t-1}</math> all correspondences in conflict with those in <math>C_q</math></p> <p>      (given a mapping constraint)</p> <p>11:   <math>\theta_t = (\theta_{t-1} - \{C_a, C_b\}) \cup \{C_q\}</math></p> <p>12:   <b>until</b> <math> \theta_t  == 1</math> or <math>\min_{r,s} D_{AP}(C_r, C_s) &gt; \delta_D</math></p> <p>13: (Cluster Selection)</p> <p>14: Eliminate unreliable clusters from <math>\theta</math></p>
--

### 3. MATCHING TESTS

We demonstrate unsupervised object-based image matching experiments performed on Caltech 101 object categories [10], ETHZ shape database [11]-[14] and then on challenging real images. As shown in the Figures below, the dataset contains objects of different categories. The MSER [5] detector and the SIFT descriptor [6] are used to generate candidate correspondence for each image pair. According to the distance of 128-dimensional SIFT descriptor, the best 1000 candidate matches are considered. Each feature can have multiple correspondences. Then, the ground truths are labeled manually for all the candidates considered. To measure the dissimilarity between two candidate region correspondences, we adopted the affine transfer error function. For evaluation, the result of ACC is compared with UNN [6] as a baseline. UNN shows less number of matches between two similar objects. As the results suggest, ACC accomplishes successful

many-to-many object matching under intra-cluster one-to-one constraint. ACC yields higher recall rate than the other methods. Fig. 1-25 show test on different object categories on Caltech 101 set. Fig. 26-30 show test results on ETHZ shape dataset. Fig. 31-33 show tests on real-life objects on which only one sample gives false result.

### 3.1. Test on Caltech 101 Object Category

#### 3.1.1. Garfield

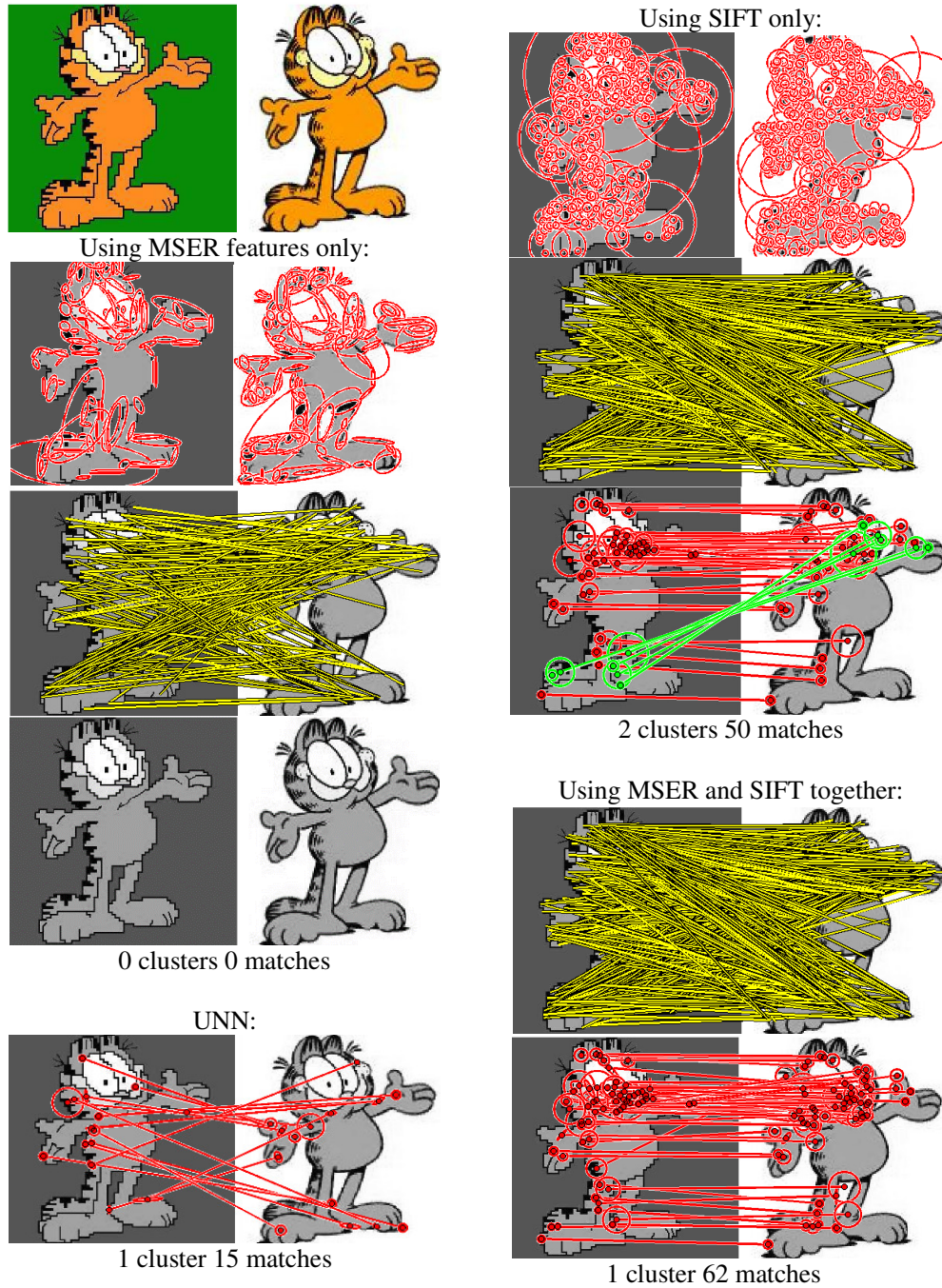


Figure 1. Test results on Garfield

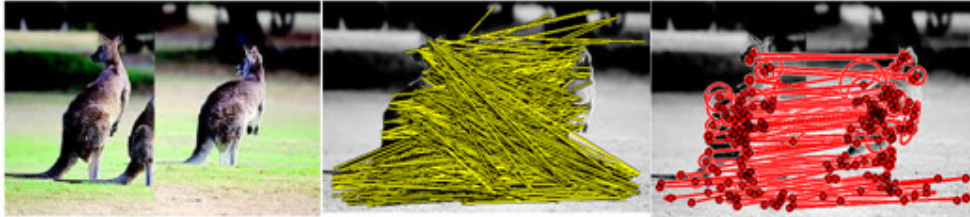
### 3.1.2. Faces



2 clusters 85 matches

Figure 2. Test results on Human Face

### 3.1.3. Kangaroo



1 cluster 98 matches

Figure 3. Test results on Kangaroo

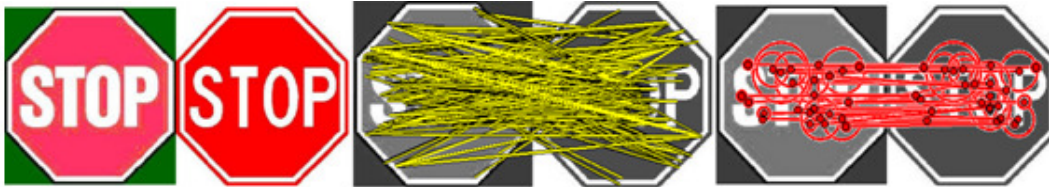
### 3.1.4. Cup



1 cluster 77 matches

Figure 4. Test results on Cup

### 3.1.5. Stop Sign



1 cluster 25 matches

Figure 5. Test results on Stop Sign

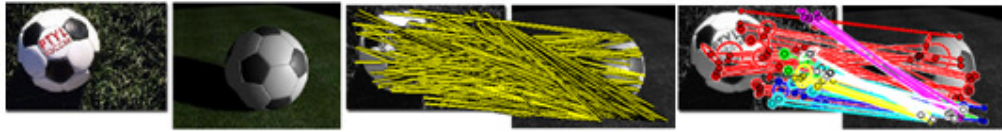
### 3.1.6. Watch



7 clusters 138 matches

Figure 6. Test results on Watch

### 3.1.7. Football



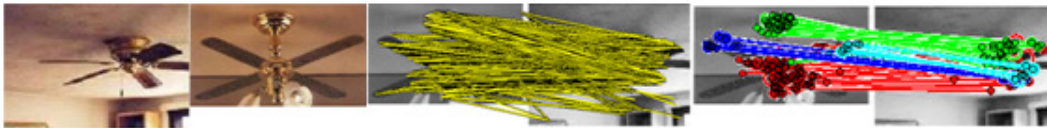
7 clusters 71 matches  
Figure 7. Test results on Football

### 3.1.8. Motorbikes



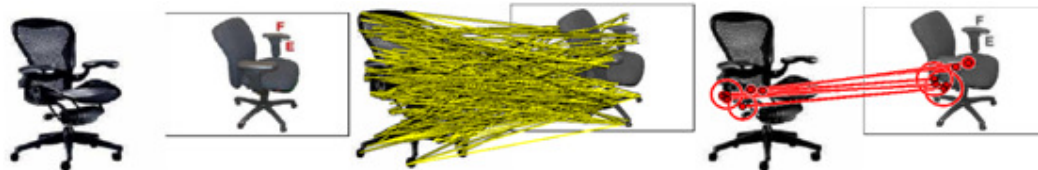
1 cluster 72 matches  
Figure 8. Test results on Motorbikes

### 3.1.9. Ceiling Fan



4 cluster 120 matches  
Figure 9. Test results on Ceiling Fan

### 3.1.10. Chair



1 cluster 6 matches  
Figure 10. Test results on Chair

### 3.1.11. Airplane



1 cluster 60 matches  
Figure 11. Test results on Airplane

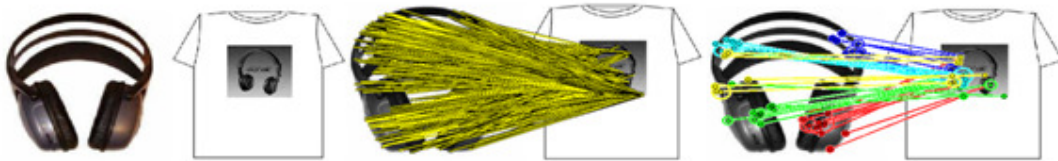
**3.1.12. Camera**



4 clusters 52 matches

Figure 12. Test results on Camera

**3.1.13. Headphone**



5 clusters 45 matches

Figure 13. Test results on Headphone

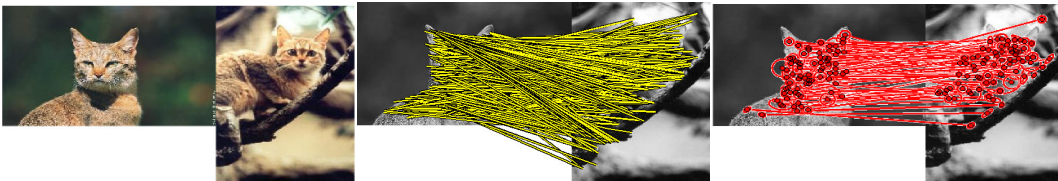
**3.1.14. Laptop**



4 clusters 103 matches

Figure 14. Test results on Laptop

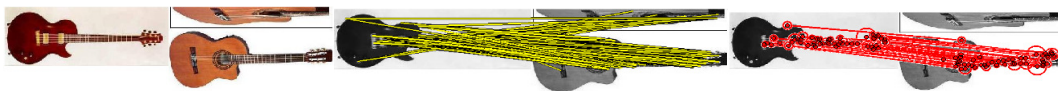
**3.1.15. Wildcat**



1 cluster 76 matches

Figure 15. Test results on Wildcat

**3.1.16. Guitar**

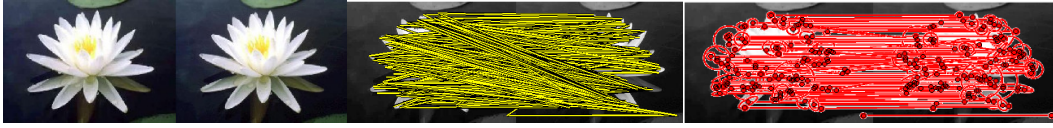


1 cluster 48 matches

Figure 16. Test results on Guitar



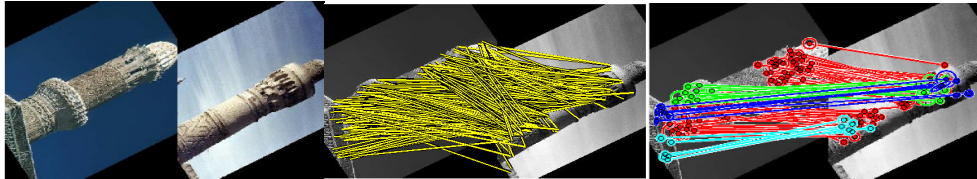
**3.1.17. Water Lilly**



1 cluster 119 matches

Figure 17. Test results on Water Lilly

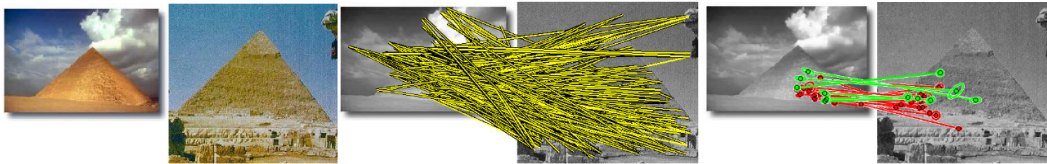
**3.1.18. Minaret**



4 clusters 85 matches

Figure 18. Test results on Minaret

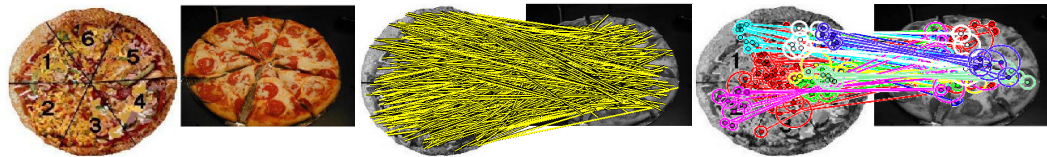
**3.1.19. Pyramid**



2 clusters 21 matches

Figure 19. Test results on Pyramid

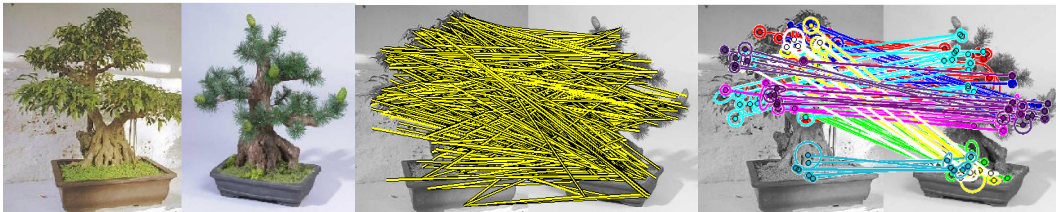
**3.1.20. Pizza**



9 clusters 122 matches

Figure 20. Test results on Pizza

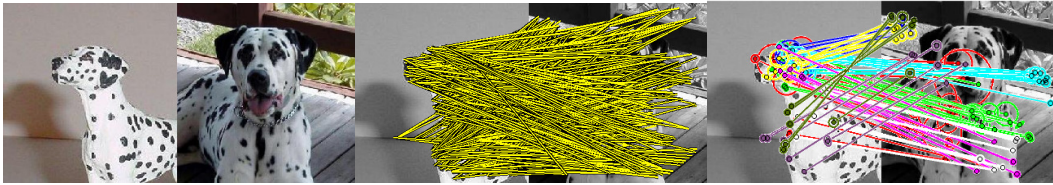
**3.1.21. Bonsai**



10 clusters 117 matches

Figure 21. Test results on Bonsai

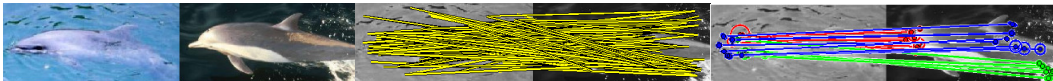
### 3.1.22. Dalmatian



9 clusters 91 matches

Figure 22. Test results on Dalmatian

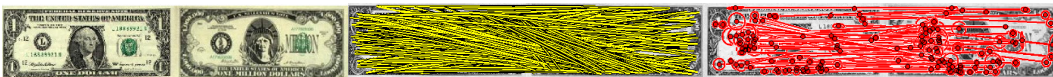
### 3.1.23. Dolphin



3 clusters 22 matches

Figure 23. Test results on Dolphin

### 3.1.24. Dollar bill



1 cluster 70 matches

Figure 24. Test results on Dollar bill

### 3.1.25. Helicopter

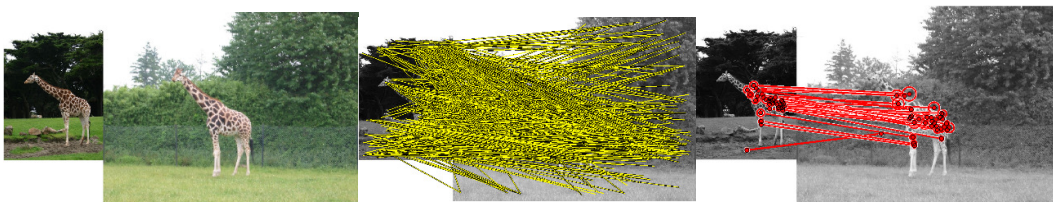


2 clusters 61 matches

Figure 25. Test results on Helicopter

## 3.2. Test on ETHZ shape database

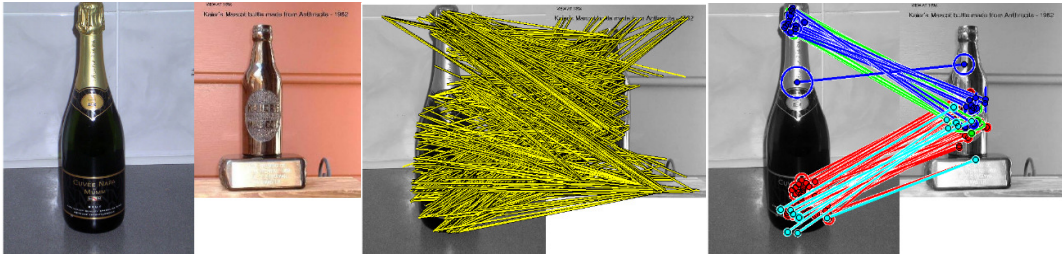
### 3.2.1. Giraffes



1 cluster 34 matches

Figure 26. Test results on Giraffes

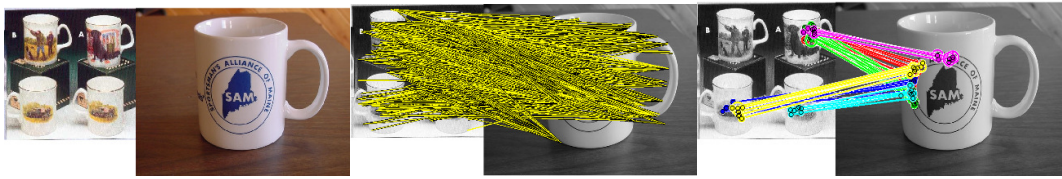
### 3.2.2. Bottle



4 clusters 74 matches

Figure 27. Test results on Bottle

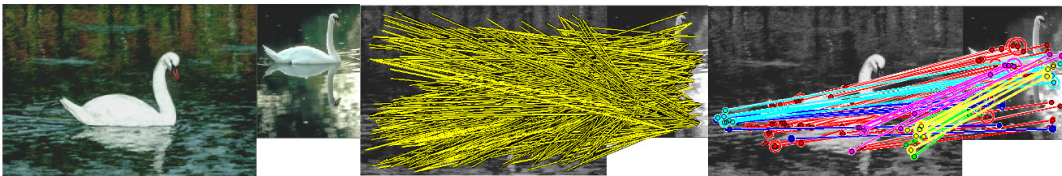
### 3.2.3. Mugs



6 clusters 64 matches

Figure 28. Test results on Mugs

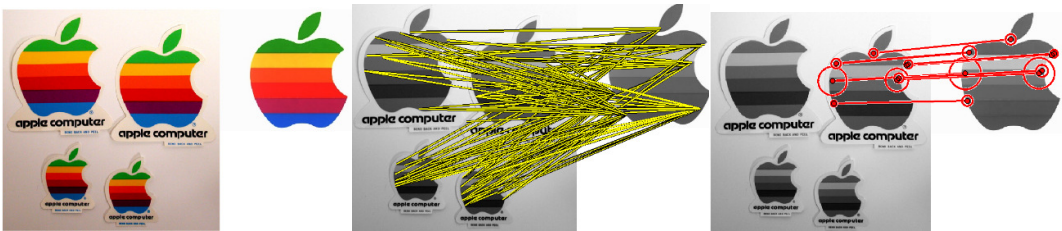
### 3.2.4. Swan



6 clusters 72 matches

Figure 29. Test results on Swan

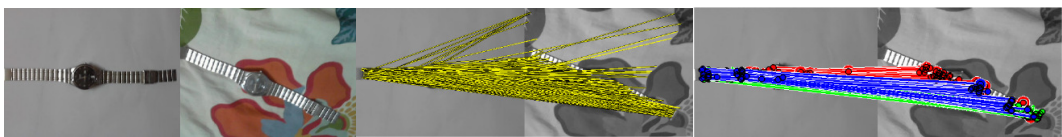
### 3.2.5. Apple logo



1 clusters 8 matches

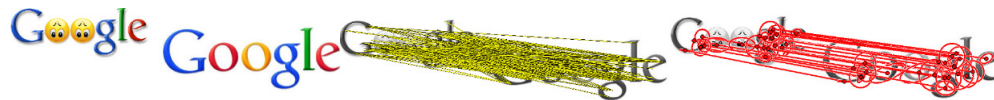
Figure 30. Test results on Apple logo

### 3.3. Test on Random Real life Images



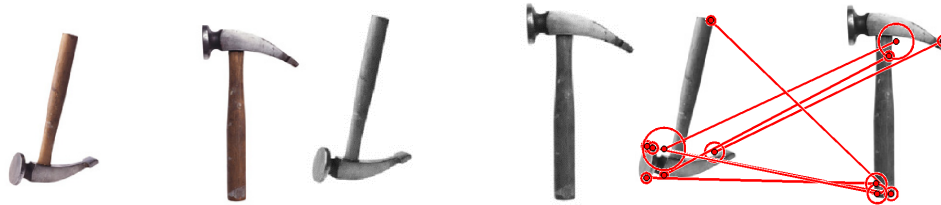
3 clusters 105 matches

Figure 31. Test results on random Watch



1 cluster 23 matches

Figure 32. Test results on Google logo



No match found via ACC; 1 cluster 7 matches via UNN

Figure 33. Test results on an instrument. UNN is successful to detect matches. But ACC cannot find any match in this case.

#### 4. CONCLUSIONS

MSER features and SIFT descriptors are hierarchically combined in a framework called Agglomerative Correspondence Clustering for object matching and categorization. This technique is proved quite accurate in matching object of same category judged by its performance in object-matching using standard datasets. ACC is proven satisfactory in case of matching challenging real-world images with very small failure.

#### REFERENCES

- [1] A. C. Berg, T. L. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences", CVPR, pp. 26–33, 2005.
- [2] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints", ICCV, pp. 1482–1489, 2005.
- [3] L. Torresani, V. Kolmogorov, and C. Rother, "Feature correspondence via graph matching: Models and global optimization", ECCV, pp. 596–609, 2008.
- [4] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable external regions", BMVC, 2002.
- [5] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors", IJCV, 60(1) pp.63–86, Oct. 2004.
- [6] D. G. Lowe, "Object recognition from local scale-invariant features", ICCV, pp. 1150–57, 1999.
- [7] A.K.Jain and R.C.Dubes, "Algorithms for Clustering Data", Prentice Hall, 1998.
- [8] S. Theodoridis and K. Koutroumbas. "Pattern Recognition", 3<sup>rd</sup> Edition. Academic Press, 2006.
- [9] Minsu Cho, Jungmin Lee, Kyoung Mu Lee. "Feature Correspondence and Deformable Object Matching via Agglomerative Correspondence Clustering", IEEE International Conference on Computer Vision (ICCV), 2009.
- [10] L. Fei-Fei, R. Fergus and P. Perona. "Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories", IEEE. CVPR 2004, Workshop on Generative-Model Based Vision. 2004.
- [11] V. Ferrari, T. Tuytelaars and L. V. Gool, "Object Detection by Contour Segment Networks", ECCV 2006, Graz, Austria, 2006.

- [12] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid, "Groups of Adjacent Segments for Object Detection", PAMI, January 2008.
- [13] V. Ferrari, F. Jurie, and C. Schmid, "Accurate Object Detection with Deformable Shape Models Learnt from Images", CVPR 2007, Minneapolis, USA, 2007.
- [14] V. Ferrari, F. Jurie, and C. Schmid, "From Images to Shape Models for Object Detection", IJCV 2009.

**Authors**

Md. Shafayat Hossain was born in Dhaka, Bangladesh. He is currently an undergraduate student of Bangladesh University of Engineering and Technology (BUET) in department of Electrical and Electronic Engineering. His research interests include analytical modeling & simulation of Nano-scale electron device, robotics, computer vision and embedded system design.



Ahmedullah Aziz was born in Dhaka, Bangladesh. He is currently an undergraduate student of Bangladesh University of Engineering and Technology (BUET) in department of Electrical and Electronic Engineering. His current research interests include Robotics, Novel semiconductor based thin film characterization, embedded system design etc.



Mohammad Wahidur Rahman is currently an undergraduate student of Bangladesh University of Engineering and Technology (BUET) in department of Electrical and Electronic Engineering. His current research interests include Novel semiconductor based thin film characterization, Computer Vision technology etc.

