International Journal of Combinatorial Optimization Problems and Informatics, 13(3), Sep-Dec 2022, 43-55. ISSN: 2007-1558

www.editada.org

A Reduced Spherical Model for Optimization of Image Recognition through 3D Colour Histograms

Felipe Trujillo-Romero[1], Santiago-Omar Caballero-Morales[2], Karen-Lizbeth Flores-Rodriguez[3], Carlos Garcia-Capulin[1], Raúl Sanchez-Yanez[1]*

[1]División de Ingenierías, Campus Irapuato-Salamanca - Universidad de Guanajuato, Salamanca, Guanajuato (México)
[2]Facultad de Ingeniería Industrial y Logística, Universidad Popular Autónoma del Estado de Puebla A.C., Puebla, Puebla (México)
[3]Instituto Politécnico Nacional, CICATA-Unidad Querétaro, Querétaro (México)
*Correspondence: fdj.trujillo@ugto.mx

Abstract. A 3D colour histogram is an image processing technique used to visualize the distribution of colours (Red-Blue-Green) in a picture. Because colour distribution does not significantly change if a pictured object is translated or rotated, a 3D colour histogram can be used as a descriptor for automatic object recognition. However, this task requires cubes with high dimensionality. Within this context, the present work contributes with an approach to reduce the high dimensionality of the 3D colour histogram and improve it as a descriptor for object recognition. Tests performed with three databases (COIL-100, an own database, and CO3D) and three recognition systems corroborated its suitability for efficient object recognition, achieving overall recognition rates of 97.0% for objects with complex geometry and reflectance features. These results are more competitive when compared with other colour descriptors as C-SIFT, RGB-SIFT, Colour moments and RGB histograms.
Keywords: image recognition, 3d colour histogram, dimension reduction, neural networks

Article Info
Received Jul 13, 2022
Accepted Aug 18, 2022

1 Introduction

In computer vision, object recognition is a challenging task as real objects exhibit different visual features such as colour, texture and shape. Also, these features may be influenced by external factors such as background, perspective, pose, and lighting conditions. In this context, computer vision researchers have developed approaches to extract and characterize the main features of objects and improve their classification. Table 1 presents a review of the features considered by object recognition works.

As reviewed, colour is frequently used as the main feature for object or image recognition. This is understandable as the human perception of objects is influenced by their colours, and all objects have distinctive colour patterns [1], [2]. This is independent of the nature of the object, and colour has been a suitable descriptor to recognize fruits and vegetables [3]–[5], human emotions and faces [6], [7], people [8] and buildings [9]. Also, research has corroborated the influence of colour data in the accuracy and speed of object recognition systems [10], [11].

Note that, for object recognition purposes, colour data must be described through appropriate extraction and processing techniques. In this context, 3D colour histograms have been proposed as suitable descriptors as they are invariant to translation and rotation about the viewing axis, and get minimally affected to scale variation and occlusion [12]–[14].

However, depending of the object and required recognition rate, 3D colour histograms can be prone to high dimensionality [12].
Within this context, the present work contributes with a novel approach to reduce the high dimensionality of 3D colour histograms and improve them as descriptors for object recognition tasks. This reduction was achieved by modeling the 3D colour histogram's bins as spheres. With this approach, which led to 90.0% reduction in dimensionality when compared to standard 3D colour histograms, recognition rates of 97.0% were consistently obtained throughout different object databases, including an own database. Thus, robust descriptors can be obtained for object recognition. These results also were more competitive when compared to other colour descriptors such as C-SIFT, RGB-SIFT, Colour moments and RGB histograms.

The present work is structured as follows: in Section 2 the details on the databases used for training and testing are presented, including the characteristics of the new database created by us. Then, in Section 3 the details of the reduction approach for improvement of the 3D colour histogram descriptor are presented. The details of three pattern recognition systems considered for assessment of the descriptor are presented in Section 4. The results are presented and analyzed in Section 5. Finally, in Section 6 our conclusions and future work are presented.

Table1. Review of Features for Object Recognition.

Work	Year	Feature Descriptor
[15]	2019	Colour and Texture
[16]	2019	Colour
[17]	2019	Colour and Depth Data: RGB-D (Red, Green, Blue and Depth)
[18]	2018	Colour, Depth Data and DWT: RGB-D (Red, Green, Blue and Depth) and Discrete Wavelet Transform
[19]	2018	Colour
[20]	2018	Colour
[21]	2017	Colour and Depth Data: RGB-D (Red, Green, Blue and Depth)
[22]	2016	Colour and Depth Data: RGB-D (Red, Green, Blue and Depth)
[23]	2013	Colour
[24]	2012	Colour and Zernike Moments
[25]	2012	Colour and Depth Data: RGB-D (Red, Green, Blue and Depth)
[26]	2011	Luminance Contours
[27]	2011	Colour and Depth Data: RGB-D (Red, Green, Blue and Depth)
[28]	2010	Colour
[29]	2009	Colour and Gray-Values
[30]	2006	Colour and SIFT (Scale Invariant Feature Transformation)

2 Object Databases

Databases are important resources to train, test and validate pattern recognition approaches. In this work we considered two well-known databases within the field of image processing. We also developed a new database with more natural environmental features for robustness assessment. These databases are described in the following sections.

2.1 COIL-100

The Columbia Object Image Library (COIL-100) [31] is a widely used database for testing object recognition algorithms. This database consists of 100 objects. Each object was placed in a motorized turntable which varied its pose through 360 degrees with respect to a fixed colour camera. Colour images of each object were taken at pose intervals of 5 degrees, leading to 72 images for each object, and 7200 images for all objects in the database. This enables the learning processes

the object from a set of views and evaluates the recognition system using the remaining images. Figure 1 presents examples of the objects included in the COIL-100 database.

Fig. 1. Objects selected from the COIL-100 database.

2.2 CO3D

The Common Objects in 3D (CO3D) [32] is a recently created database to test different algorithms for 3D object reconstruction. The CO3D is composed by 46 object categories, and each category has several subsets of the same class. Figure 2 presents examples of the objects included in the CO3D database.

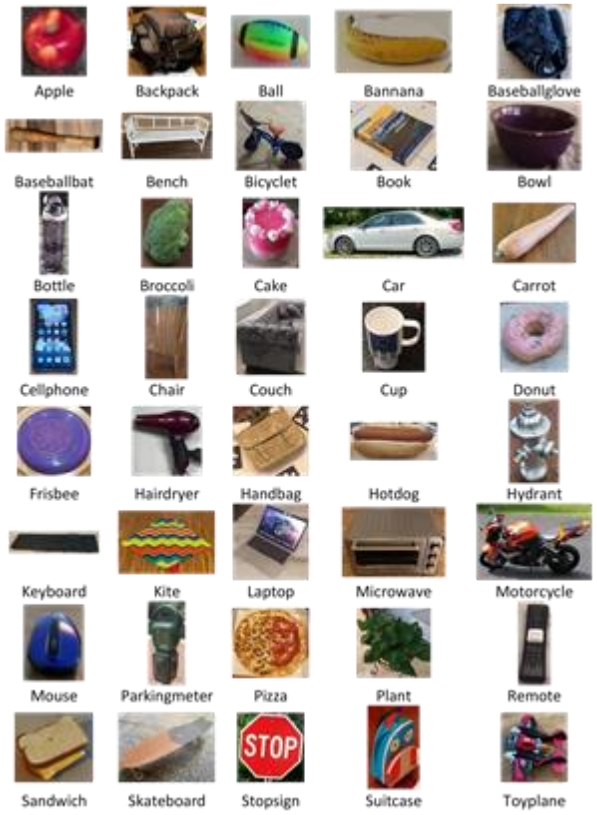

Fig. 2. Objects selected from the CO3D database.

This database is meaningful for the present work due to the following features: (a) it is composed of everyday or usual objects in a natural human environment, (b) the images were not taken in particular scenarios, cameras, or structured lights, (c) there are objects of different sizes within the same category, and (d) some images have different noise elements such as blurring, shining, shadows, transparency, and occlusions (see Figure 3). The noise elements represent challenges for training and testing of the recognition system as they affect perception of colour and object segmentation.

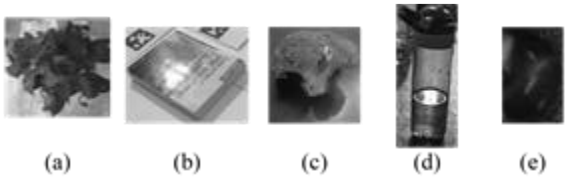

Fig. 3. Objects selected from the CO3D database with noise elements: (a) blurring, (b) shining, (c) shadows, (d) translucent, and (e) occlusion (note: gray-scale conversion was performed to enhance the presence of these elements).

2.3 OLib

An Own Library (OLib) is a database created by us with the purpose of validating our approach. This database consists of 20 common objects. Figure 4 presents examples of the objects included in this database. In contrast to the COIL-100 database, the OLib database has the following features: (a) no uniform pattern was considered while taking the images (e.g., the images were taken from different positions of the object without a particular sequence), (b) the images were taken without using any kind of structured light (just daylight was considered), (c) the background is white, thus more shadows are generated, and (d) all objects are placed at different distances from the camera, leading to all objects having different scales within the fixed image frame. As presented in Figure 5 these features also led to the presence of noise elements such as in the CO3D database.

Fig. 4. Objects within the OLib database.

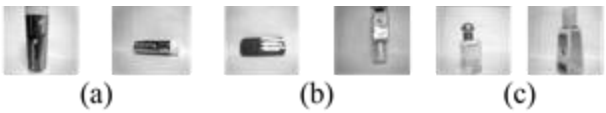

Fig. 5. Objects selected from the OLib database with noise elements: (a) scale change, (b) shining, and (c) translucent (note: gray-scale conversion was performed to enhance the presence of these elements).

3 Approach to 3D Colour Histogram Enhancement

3.1 Background Removal

Background removal is important to reduce or eliminate the interference that the background can introduce to the generation of the 3D colour histogram. For this purpose, the snake algorithm reported in [33] was considered. This algorithm detects the contour of the object, enabling its segmentation and consequent elimination of the background. This particular version of the snake algorithm uses local information to generate a segmentation, thus contours can be detected in objects with different feature profiles. Figure 6 presents four stages of the segmentation process for an object within the COIL-100 database. While it is possible that a little bit of background still remains, its influence in the generation of the 3D colour histogram can be considered as not significant.

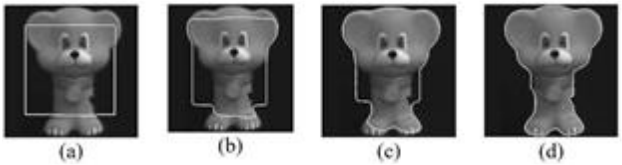

Fig. 6. Stages of the segmentation process for background removal.

3.2 Reduction of the 3D Colour Histogram Descriptor

In the digital world, colours are usually defined by three primary components: red, green, and blue *(r, g, b).* Based on this, a three-dimensional space can be created by considering the *r*, *g*, and *b* components as the ordinary Cartesian coordinates of a Euclidean system. In this three-dimensional model, non-negative values within the range [0, 255] are associated with each *(r, g, b)* triplet, which leads to a specific colour. This approach allows computations of the colour similarity of two objects by merely calculating the distance between their *(r, g, b)* patterns: the shorter the distance, the higher the similarity. Out-of-range computations can also be performed in this way.

Although this representation cannot be used directly as a descriptor, if each colour in the image is mapped into the RGB cube and its frequency is computed, it is possible to create a colour description from it. This leads to the 3D colour histogram, which is a RGB cube of $n \times n \times n$ dimension, where n is the number of layers of the histogram, and each bin (or cell within the cube) represents the frequency distribution of binned *(r, g, b)* triplets within an interval.

Note that, while the colour distribution within the RGB space of the 3D histogram can be obtained using different tools (i.e., ImageJ [34]), the high dimensionality may restrict their use for object recognition. The present work addresses this aspect through the use of a spherical model for the colour bins of the 3D colour histogram which is built as follows:

(a) Define the parameter n as the size of one dimension of the cube, where $n > 1$ and n^3 is the total dimension of the cube. Figure 7 presents examples of 3D colour histograms with different values for n.
(b) Estimate the maximum radius of the spherical bins (rs) of the 3D histogram as rs = 256/n. This defines the maximum length of the sphere's radius, the radius of the spherical bins for *(r, g, b)* triplets will be proportional to their colour frequencies. Note that all of these elements is directly dependent of the n parameter.
(c) Fill the spherical bins of the 3D colour histogram as determined by the colour frequencies.

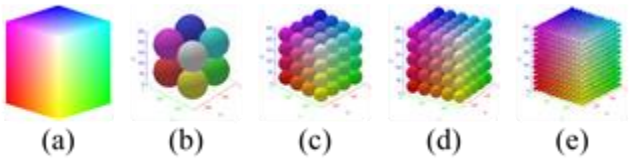

Fig. 7. 3D colour histograms with dimension (a) n=1, (b) n=2, (c) n=4, (d) n=5, and (e) n = 12.

Figure 8 presents some examples of 3D colour histograms with frequency-sized spherical bins for different images of the COIL- 100 database. As observed, the colour distribution patterns show significant differences between objects. Note

that n = 6 (as considered for Figure 8) is not an arbitrary value. It is consequence of different tests done for different values of n which led to a maximum number of bins (dimension) of $6^3 = 216$.

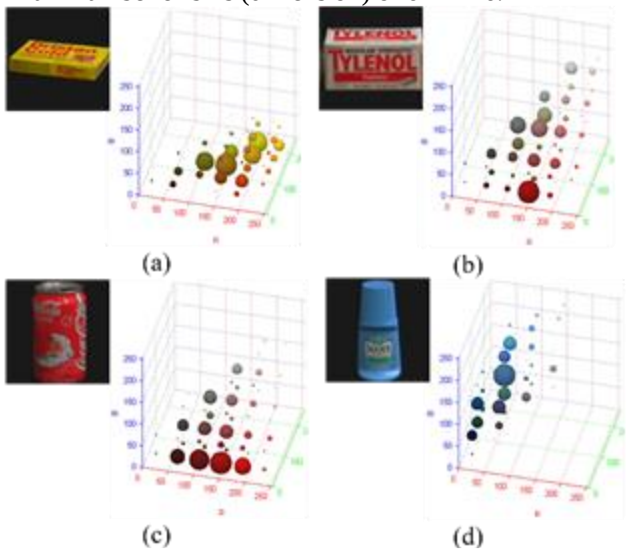

Fig. 8. 3D colour histograms (n=6) from selected objects of the COIL-100 database: (a) Object 1, (b) Object 31, (c) Object 62, and (d) Object 88.

4 Recognition Systems

Once the object descriptor is finished, the next step consists on defining the structure of the recognition system. Figure 9 presents the general structure of the system which includes the modules of data modeling (learning), testing (recognition), pre-processing and data descriptor (segmentation, 3D colour histogram), and histogram reference data (histogram database).

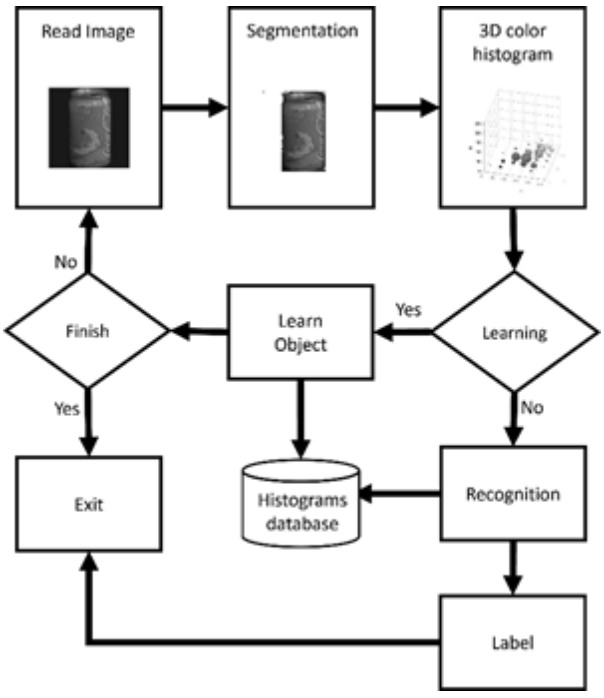

Fig. 9. Modules of the recognition system for assessment of the colour descriptor.

As presented, the input consists of the images of the objects which are processed to be described as 3D colour histograms. This process includes segmentation and background removal as described in Section 3. After the 3D colour histogram are created, there are two possible stages to follow:

- Training of the recognition system, where the system "learns" the information provided by the descriptors (i.e., the 3D colour histograms of the considered objects). The learning stage is repeated until all the considered objects are learned, and it leads to a reference database of 3D histograms.
- Testing of the recognition system, where the trained system executes recognition of input data (i.e., the 3D colour histograms of the considered objects). This data must be different from the data considered to train the recognition system. This stage is executed only one time.

The learning / recognition module can be implemented through different techniques. In this work, three techniques were considered and these are described in the following sections.

4.1 Normalized Euclidean Distance (NED)

The Euclidean distance is a positive number that indicates the separation between two points in a space where the axioms and theorems of Euclid's geometry are fulfilled. Mathematically, the one-dimensional Euclidean distance *d(A, B)* between points *A* and *B* is defined as the square root of the square of the differences in their *X* coordinates. This can be extended to high-order dimensions, such as a third-order dimension.

An inconvenient of the Euclidean distance is its sensitivity to the measurement units of the variables: the differences between the values of variables measured with high values will contribute much more than the differences between the importance of the variables with low weights. Consequently, changes in scale will also determine changes in the distance between points. A possible way of solving this problem is the previous typification of the variables or the normalized Euclidean distance.

Therefore, the Normalized Euclidean Distance gives the squared distance between two vectors whose lengths have been scaled to have a unit norm. This is helpful with the changing of scale in the objects. Thus, the Normalized Euclidean Distance (NED) would be considered the error (*e*) between the two analyzed vectors. In our case, the NED or error *e* between two histograms is computed as:

$$e = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}({}^{2}f_{ijk} - {}^{1}f_{ijk})^{2}}{n^{3}}}.$$

4.2 Self-Organizing Map (SOM)

The Self-Organizing Map (SOM) [35] technique belongs to the category of neural networks of unsupervised learning. The main idea of this algorithm is that, given a set *A* of input vectors, the Kohonen network should generate a partition of the set *A* into m disjoint regions $a_1, a_2, \ldots, a_m$. Then, the objective is that the Kohonen network covers the set *A* so that one and only one neuron is activated for each input vector. If set *A* is divided into *m* regions, then the Kohonen network must have at least m neurons, and each neuron will specialize in one and only one region.

For implementing the recognition system, the architecture of the self-associative map neural network was composed as follows: a relation of 60 neurons in an array of 6×10; it is to say, 60 cells in the map. The maximum iteration epochs used to train the neural network was 9000, the α was set to 0.2, and the Manhattan distance was considered.

4.3 Growing Cell Structure (GCS)

Growing Cell Structure (GCS) [36] is a derivation of the Kohonen's self-organizing neural network. Usually, the network starts with a certain number of neurons and depending of an error which is estimated from the input patterns, it adapts and inserts new neurons or eliminates those that are no longer useful in order to reduce this error. This allows the recognition module to discriminate the objects learned from the characteristics extracted by a descriptor during the learning stage.

The parameters used for learning in GCS are neurons and epoch or adaptation step values. These epochs are selected according to the maximal number of the images remaining in the databases used. For example, in our database, the epoch parameter was chosen in a range between 250-300. If it is assumed that a neuron is needed for every class, then at least 45 neurons are necessary. Therefore, in a similar way that the SOM algorithm, the maximal neurons were set to 90. Other required parameters for the GCS algorithm are the constants of adaptation for the winner neuron ε_b = 0.05, and for the neighborhood of the winner neuron ε_n = 0.005.

5 Results

For the learning and recognition processes of the Euclidean Distance, SOM and GCS systems, the following database sets of objects were considered:

- COIL-100: 25 objects (shown in Figure 1);
- CO3D: 45 objects;
- OLib: 20 objects.

Learning was performed with only four views of each object. Thus, 100 different images from COIL-100, 180 from CO3D, and 80 from our database OLib were used for training of the recognition systems.

The four images of each object were processed as described in Section 3 to obtain their 3D colour histogram descriptors. To optimize searching time during the recognition stage, the four descriptors corresponding to these images were unified into a single 3D colour histogram. This was performed by considering the strategy described in [37]. Then, the unified descriptor is stored within the system's database (NED, SOM, and GCS) for future use in the recognition stage.

Recognition was performed with the remaining images of each database and object (images not considered for training/learning).

5.1 Recognition Performance on the COIL-100 Database

Table 2 presents the results of the recognition systems with data from the COIL-100 database. Note that average recognition rates are higher than 97.00% for all systems and these results are consistent through all objects as the standard deviation is minimal as measured by the coefficient of variability which is less than 5.0%. These results are significant as test data consists of ((72- 4)/72) = 94.0% of the whole database and training was performed with only (4/72) = 6.0% of the database. Also, test data is likely to have images with view angles / poses which may be highly different from those used for training.

Table 2. Recognition rates (%) of the NED, SOM and GCS systems with the test data of the COIL-100 database.

Object	NED	SOM	GCS	Object	NED	SOM	GCS
1	98.00	100.00	100.00	55	99.00	98.15	100.00
7	99.00	98.15	96.67	59	100.00	94.44	100.00
12	99.00	100.00	96.67	62	100.00	98.15	100.00
18	80.00	92.59	83.33	68	99.00	96.30	100.00
24	98.00	96.30	90.00	75	100.00	100.00	93.00
25	98.00	96.30	100.00	78	99.00	98.15	96.67
30	98.00	100.00	100.00	82	100.00	98.15	100.00
31	100.00	100.00	96.67	83	100.00	96.30	96.67
36	99.00	92.59	96.67	86	100.00	92.59	94.52
39	98.00	98.33	100.00	88	99.00	100.00	100.00
43	100.00	100.00	100.00	96	100.00	94.44	100.00
46	100.00	100.00	100.00	μ	**97.76**	**97.10**	**97.23**
50	82.00	86.66	93.33	σ	**5.01**	**3.28**	**3.94**
52	99.00	100.00	96.67	cv	**0.05**	**0.03**	**0.04**

5.2 Recognition Performance on the CO3D Database

Table 3 presents the results of the recognition systems with data from the CO3D database. In general, the NED system achieves the lowest recognition rate with an average of 94.57%, reporting rates of 85.00% - 89.00% for 20.00% of the objects within the database (banana, bicycle, cellphone, chair, couch, handbag, hotdog, microwave, skateboard, and wineglass).

Nevertheless, it is important to highlight that the CO3D database consists of objects with noise elements such as shadows and blurring, and the images were taken in non-controlled environments. Also, testing is performed with ((50-4)/50) = 92.00% of each object's set of images, while training was performed with only (4/50) = 8.00% of the object's set.

Regarding the SOM and GCS systems, both achieve the highest recognition rates with an average of approximately 96.50%. As in the case of the COIL-100 database, these results are consistent through all objects as standard deviation is minimal as measured by the coefficient of variability which is less than 5.0%.

Particularly, SOM had some difficulties recognizing four objects: couch, hotdog, microwave, and wineglass. On the other hand, GCS had some difficulties recognizing five objects: couch, hotdog, microwave, skateboard, and wineglass.

Table 3. Recognition rates (%) of the NED, SOM and GCS systems with the test data of the CO3D database.

Object	NED	SOM	GCS	Object	NED	SOM	GCS
Apple	99.50	100.00	100.00	Hydra	98.50	100.00	100.00
BackP	95.00	95.00	96.60	Keyb	98.80	100.00	100.00
Ball	94.00	97.50	98.30	Kite	97.60	98.80	98.80
Banana	89.00	98.30	98.50	Laptop	95.00	95.40	95.40
Bglove	94.00	98.00	98.80	Mwave	85.60	88.90	88.60
Bbat	94.65	98.70	98.30	Mcycle	97.60	98.00	98.10
Bench	95.50	99.10	99.20	Mouse	98.50	100.00	100.00
Bicy	88.30	90.80	90.00	Pmeter	97.20	97.40	97.40
Book	91.10	93.60	93.70	Pizza	98.75	100.00	100.00
Bowl	98.50	100.00	100.00	Plant	93.70	95.75	96.00
Bottle	95.40	97.50	97.20	Remot	97.70	99.00	99.00
Broc	96.90	95.70	96.40	Sandw	91.90	94.40	94.60
Cake	98.00	100.00	100.00	Skateb	88.00	90.60	89.75
Car	93.00	96.70	96.30	Stops	98.15	99.00	99.20
Carrot	91.30	92.20	92.00	Suitcas	96.10	98.00	98.00
Phone	88.00	90.00	90.50	Tplane	96.10	97.80	97.70
Chair	86.75	90.00	89.75	Ttrain	98.10	100.00	99.75
Couch	87.60	88.00	87.90	Ttruck	98.75	100.00	100.00
Cup	98.40	100.00	100.00	Umbrel	96.00	97.75	97.70
Donut	99.00	100.00	100.00	Vase	99.40	100.00	100.00
Frisb	100.00	100.00	100.00	Wglass	85.10	88.50	87.90
Hdryer	98.30	99.00	99.10	μ	**94.57**	**96.48**	**96.51**
Hbag	88.80	92.60	92.90	σ	**4.41**	**3.85**	**3.96**
Hotdog	88.00	89.80	89.50	cv	**0.04**	**0.04**	**0.04**

5.3 Recognition Performance on the OLib Database

Table 4 presents the results of the recognition systems with data from the CO3D database. Testing is performed with ((20- 4)/20) = 80.00% of each object's set of images, while training was performed with only (4/20) = 20.00% of the object's set.

While the NED recognizer reports some issues with six objects (9, 10, 15, 16, 18, and 20), a general average recognition rate of 96.50% corroborates its high accuracy. This is important as the objects in our database include noise elements

such as shadows, scale variation, shining, and translucency, which makes the task difficult. Regarding the SOM system, there are specific difficulties with objects 18 and 19. This led to an average recognition rate of 95.96% which is lower than the performance of the NED system. Finally, the highest recognition rate of 96.88% is obtained with the GCS system. In contrast to the experiments with the COIL-100 and CO3D databases, performance reports a minimal increase in variability of 1.0%.

Table 4. Recognition rates (%) of the NED, SOM and GCS systems with the test data of the OLib database.

Object	NED	SOM	GCS	Object	NED	SOM	GCS
1	100.00	100.00	100.00	13	100.00	95.00	100.00
2	98.66	97.00	100.00	14	100.00	95.00	92.00
3	97.00	98.30	100.00	15	90.00	100.00	95.00
4	100.00	100.00	100.00	16	85.00	94.40	96.67
5	97.00	100.00	100.00	17	95.00	100.00	100.00
6	99.33	100.00	98.32	18	90.00	80.00	89.66
7	100.00	100.00	100.00	19	100.00	85.00	93.00
8	100.00	100.00	100.00	20	95.00	97.00	95.00
9	90.00	90.00	90.45	μ	**96.50**	**95.96**	**96.88**
10	93.00	93.30	95.00	σ	**4.58**	**5.44**	**3.70**
11	100.00	97.50	100.00	cv	**0.04**	**0.05**	**0.04**
12	100.00	96.67	92.55				

5.4 Recognition Performance vs. Alternative Colour Descriptors

To conclude the assessment for the improved 3D colour histogram descriptor, a comparison with other colour descriptors reported in the literature was performed. The considered descriptors were:

- C-SIFT [30]: a coloured SIFT which integrates the information of colour variations as well as geometrical data.
- RGB-SIFT [38]: a SIFT descriptor is computed for each colour channel.
- Colour moments [39]: it is based on the assumption that the distribution of colours can be modeled as a probability distribution. Therefore, an image can be represented by the central moments (mean, variance and skewness) of the distribution of the RGB colours. Thus, moments are computed for these colour channels and consequently the image is described by nine moments (three for each colour).
- RGB histogram: this is a 1-D histogram which combines the information of the RGB channels.

Table 5 presents the results of the three recognition systems on the test data of the COIL-100 database. These results compare the performance of the improved 3D colour histogram with the previously reviewed colour descriptors. With the COIL-100 database, it is clear the advantage of the proposed 3D colour histogram descriptor independently of the recognition system.

Table 6 and Table 7 present the results of the three recognition system on the test data of the CO3D and OLib databases respectively. Although for these databases the RGB-SIFT descriptor increases the performance of the SOM system only, the proposed 3D colour histogram descriptor achieves very similar performance, outperforming the C-SIFT, Colour moments, and RGB histogram descriptors. In all other databases and recognition systems, the proposed 3D colour histogram achieves the highest recognition rates.

Table 5. Comparison of recognition rates (%) obtained with alternative colour descriptors: COIL-100 database

Descriptor	NED	SOM	GCS
3D Colour Histogram	**97.80**	**97.10**	**97.20**
C-SIFT	95.30	90.10	91.50
RGB-SIFT	95.00	94.25	94.00
Colour moments	92.35	91.60	92.20
RGB histogram	89.50	90.00	90.50

Table 6. Comparison of recognition rates (%) obtained with alternative colour descriptors: CO3D database

Descriptor	NED	SOM	GCS
3D Colour Histogram	**94.57**	96.48	**96.51**
C-SIFT	92.93	95.09	96.31
RGB-SIFT	93.10	**97.00**	96.05
Colour moments	93.66	94.25	95.35
RGB histogram	92.83	93.80	94.00

Table 7. Comparison of recognition rates (%) obtained with alternative colour descriptors: OLib database

Descriptor	NED	SOM	GCS
3D Colour Histogram	**96.50**	96.00	**96.80**
C-SIFT	94.40	93.50	93.30
RGB-SIFT	95.30	**96.45**	96.00
Colour moments	93.10	93.45	94.00
RGB histogram	93.00	92.80	92.60

6 Conclusions

In this work a colour-based object descriptor was defined. Such descriptor is a 216-bin 3D colour histogram. To evaluate its performance, three different recognition systems were implemented using NED, SOM, and GCS, to discriminate objects from three databases: COIL-100, CO3D, and OLib (own database). Recognition tests corroborated that the proposed descriptor is able to discriminate objects based in colour, reaching a recognition rate of around 97.0% despite the reduced number of samples used for training. When compared to other colour descriptors, the proposed 3D colour histogram descriptor generally outperformed C-SIFT, RGB-SIFT, Colour moments, and RGB histogram throughout all databases.

In summary, from the results obtained from the different evaluations carried out with the 3D colour histogram descriptor, we can conclude the following:

a) The descriptor is easy to implement.
b) The performance obtained is even better than other object colour-based descriptors.
c) It could be used as an object recognition system on a mobile or industrial robot.

Nevertheless, the proposed descriptor, as any colour-based descriptor, is prone to further improvements. As mentioned in [40] colour histograms can potentially be identical for two images with different object content. Also, similar objects of different colours may be indistinguishable based solely on colour histogram comparisons. Thus, integration with other descriptors is important to enhance feature extraction and reduce the likelihood of these potential risks.

References

1. F. Shahbaz-Khan, J. van-de Weijer, M. Vanrell, “Top-down colour attention for object recognition,” in 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 2009, pp. 979–986.
2. I. Bramao, L. Faísca, K. Petersson, A. Reis, “The contribution of colour to object recognition,” in Advances in Object Recognition Systems, I. Kypraios Ed., Rijeka: IntechOpen, 2012, pp. 73–111.
3. H. Gan, W. Lee, V. Alchanatis, R. Ehsani, J. Schueller, “Immature green citrus fruit detection using colour and thermal images,” Computers and Electronics in Agriculture, vol. 152, pp. 117–125, 2018, doi: 10.1016/j.compag.2018.07.011.
4. J. Feng, L. Zeng, L. He, “Apple fruit recognition algorithm based on multi-spectral dynamic image analysis,” Sensors, vol. 19, no. 4, 2019.
5. G. Liu, S. Mao, H. Jin, J. Kim, “A robust mature tomato detection in greenhouse scenes using machine learning and colour analysis,” in 11th International Conference on Machine Learning and Computing (ICMLC ’19), Zhuhai, China, 2019, pp. 17–21.
6. S. Wang, W. Yan, X. Li, G. Zhao, C. Zhou, X. Fu, M. Yang, J. Tao, “Micro-expression recognition using colour spaces,” IEEE Transactions on Image Processing, vol. 24, no. 12, pp. 6034–6047, 2015, doi: 10.1109/TIP.2015.2496314.

7. S. Bao, X. Song, G. Hu, X. Yang, C. Wang, “Colour face recognition using fuzzy quaternion-based discriminant analysis,” International Journal of Machine Learning and Cybernetics, vol. 10, p. 385–395, 2019, doi: 10.1007/s13042-017-0722-4.
8. Y. Wang, T. Wang, J. Yen, F. Wang, “Dynamic human object recognition by combining colour and depth information with a clothing image histogram,” International Journal of Advanced Robotic Systems, vol. 16, no. 1, 2019, doi: 10.1177/1729881419828105.
9. A. Ghandour, A. Jezzini, “Autonomous building detection using edge properties and image colour invariants,” Buildings, vol. 8, no. 5, 2018.
10. J. Herman, A. Bogadhi, R. Krauzlis, “Novel colour stimuli for studying spatial attention,” Journal of Vision, vol. 13, no. 9, 2013, doi: 10.1167/13.9.1124.
11. S. Hagen, Q. Vuong, L. Scott, T. Curran, J. Tanaka, “The role of colour in expert object recognition,” Journal of Vision, vol. 14, no. 9, 2014, doi: 10.1167/14.9.9
12. M. Swain, D. Ballard, “Colour indexing,” International Journal of Computer Vision, vol. 7, no. 1, pp. 11–32, 1991.
13. J. Ning, L. Zhang, D. Zhang, C. Wu, “Robust object tracking using joint colour-texture histogram,” International Journal of Pattern Recognition and Artificial Intelligence, vol. 23, no. 7, pp. 1245–1263, 2009.
14. K. Berker-Logoglu, S. Kalkan, A. Temizel, “Cospair: Coloured histograms of spatial concentric surflet-pairs for 3d object recognition,” Robotics and Autonomous Systems, vol. 75, no. B, pp. 558–570, 2016, doi: 10.1016/j.robot.2015.09.027.
15. M. Rafi, S. Mukhopadhyay, “Salient object detection employing regional principal colour and texture cues,” Multimedia Tools and Applications, 2019, doi: 10.1007/s11042-019-7153-z.
16. G. Liu, J. Yang, “Exploiting colour volume and colour difference for salient region detection,” IEEE Transactions on Image Processing, vol. 28, no. 1, pp. 6–16, 2019, doi: 10.1109/TIP.2018.2847422.
17. M. Loghmani, M. Planamente, B. Caputo, M. Vincze, “Recurrent convolutional fusion for rgb-d object recognition,” IEEE Robotics and Automation Letters, vol. 4, no. 3, pp. 2878–2885, 2019, doi: 10.1109/LRA.2019.2921506.
18. R. Ashraf, M. Ahmed, S. Jabbar, S. Khalid, A. Ahmad, S. Din, G. Jeon, “Content based image retrieval by using colour descriptor and discrete wavelet transform,” Journal of Medical Systems, vol. 42, no. 3, pp. 2878–2885, 2018, doi: 10.1007/s10916-017-0880-7.
19. C. Reta, J. Cantoral-Ceballos, I. Solis-Moreno, J. Gonzalez, R. Alvarez-Vargas, N. Delgadillo-Checa, “Colour uniformity descriptor: An efficient contextual colour representation for image indexing and retrieval,” Journal of Visual Communication and Image Representation, vol. 54, pp. 39–50, 2018, doi: 10.1016/j.jvcir.2018.04.009.
20. R. Boukezzoula, D. Coquin, T. Nguyen, S. Perrin, “Multi- sensor information fusion: Combination of fuzzy systems and evidence theory approaches in colour recognition for the nao humanoid robot,” Robotics and Autonomous Systems, vol. 100, pp. 302–316, 2018, doi: 10.1016/j.robot.2017.12.002.
21. X. Li, M. Fang, J. Zhang, J. Wu, “Learning coupled classifiers with rgb images for rgb-d object recognition,” Pattern Recognition, vol. 61, pp. 433–446, 2017, doi: 10.1016/j.patcog.2016.08.016.
22. Y. Huang, F. Zhu, L. Shao, A. Frangi, “Colour object recognition via cross-domain learning on rgb-d images,” in 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 2016, pp. 1672–1677.
23. P. Caleiro, A. Neves, A. Pinho, “Colour-spaces and colour segmentation for real-time object recognition in robotic applications,” Electrónica e Telecomunicacões, vol. 4, no. 8, pp. 940–945, 2013.
24. B. Chen, H. Shu, H. Zhang, G. Chen, C. Toumoulin, J. Dillenseger, L. Luo, “Quaternion zernike moments and their invariants for colour image analysis and object recognition,” Signal Processing, vol. 92, no. 2, pp. 308–318, 2012.
25. M. Blum, J. Springenberg, J. Wulfing, M. Riedmiller, “A learned feature descriptor for object recognition in rgb-d data,” in 2012 IEEE International Conference on Robotics and Automation (ICRA), Saint Paul, USA, 2012, pp. 1298–1303.
26. J. Martinovic, J. Mordal, S. Wuerger, “Event-related potentials reveal an early advantage for luminance contours in the processing of objects,” Journal of Vision, vol. 11, no. 7, 2011, doi: 10.1167/11.7.1.
27. K. Lai, L. Bo, X. Ren, D. Fox, “Sparse distance learning for object recognition combining rgb and depth information,” in 2011 IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, 2011, pp. 4007–4013.
28. K. V. D. Sande, T. Gevers, C. Snoek, “Evaluating colour descriptors for object and scene recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1582–1596, 2010.
29. G. Burghouts, J. Geusebroek, “Performance evaluation of local colour invariants,” Computer Vision and Image Understanding, vol. 113, no. 1, pp. 48–62, 2009.
30. A. Abdel-Hakim, A. Farag, “Csift: A sift descriptor with colour invariant characteristics,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1978–1983, 2006.
31. S. Nene, S. Nayar, H. Murase, “Columbia object image library (coil-100): Technical report cucs- 006-96,” Columbia Imaging and Vision Laboratory (CAVE) - Columbia University, New York, USA, 1996. Accessed: Feb. 15, 2022, [Online]. Available: https://www.cs.columbia.edu/CAVE/software/softlib/coil- 100.php, doi: 10.2779/97902.
32. J. Reizenstein, R. Shapovalov, P. Henzler, L. Sbordone, P. Labatut, D. Novotny, “Common objects in 3d: Large- scale learning and evaluation of real-life 3d category reconstruction,” in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, p. 10901–10911.
33. S. Lankton, A. Tannenbaum, “Localizing region-based active contours,” IEEE Transactions on Image Processing, vol. 17, no. 11, pp. 2029–2039, 2008, doi: 10.1109/TIP.2008.2004611.
34. “3d colour inspector/colour histogram,” Berlin, Germany, 2019. Accessed: Feb. 15, 2022, [Online]. Available: http://rsb.info.nih.gov/ij/plugins/colour-inspector.html.

35. T. Kohonen, Self-Organizing Maps. Springer, U.S.A, 2001.
36. B. Fritzke, “Growing cell structures, a self-organizing network for unsupervised and supervised learning,” Neural Network, vol. 7, no. 9, pp. 1441–1460, 1994.
37. K. Flores-Rodriguez, F. Trujillo-Romero, W. Suleiman, “Object recognition modular system implementation in a service robotics context,” in 2017 International Conference on Electronics, Communications and Computers (CONIELECOMP), 2017, pp. 1–6.
38. A. Verma, C. Liu, “Sift features in multiple colour spaces for improved image classification,” in Recent Advances in Intelligent Image Search and Video Retrieval, C. Liu Ed., Springer Cham, 2017, pp. 145–166.
39. N. Keen, “Colour moments,” The University of Edinburgh - School of Informatics, Scotland, UK, 2005. Accessed: Feb. 15, 2022.
40. X. Wang, J. Wu, H. Yang, “Robust image retrieval based on colour histogram of local feature regions,” Multimedia Tools and Applications, vol. 49, no. 2, pp. 323–345, 2009.