Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal 2022, 16(3), 269–281
https://doi.org/10.12913/22998624/150320
ISSN 2299–8624, License CC-BY 4.0

Received: 2022.04.04
Accepted: 2022.05.14
Published: 2022.06.01

Wild Image Retrieval with HAAR Features and Hybrid DBSCAN Clustering for 3D Cultural Artefact Landmarks Reconstruction

Perumal Pitchandi[1]

[1] Department of Computer Science and Engineering, Sri Ramakrishna Engineering College, Coimbatore, India

[*] Corresponding author's e-mail: perumalp@srec.ac.in

ABSTRACT

In this digital age large amounts of information, images and videos can be found in the web repositories which accumulate this information. These repositories include personal, historic, cultural and business event images. Image mining is a limited field in research where most techniques look at processing images instead of mining. Very limited tools are found for mining these images, specifically 3D images. Open source image datasets are not structured making it difficult for query based retrievals. Techniques extracting visual features from these datasets result in low precision values as images lack proper descriptions or numerous samples exist for the same image or images are in 3D. This work proposes an extraction scheme for retrieving cultural artefact based on voxel descriptors. Image anomalies are eliminated with a new clustering technique and the 3D images are used for reconstructing cultural artefact objects. Corresponding cultural 3D images are grouped for a 3D reconstruction engine's optimized performance. Spatial clustering techniques based on density like Particle Varied Density Based Spatial Clustering of Applications with Noise (PVDBSCAN) eliminate image outliers. Hence, PVDBSCAN is selected in this work for its capability to handle a variety of outliers. Clustering based on Information theory is also used in this work to identify cultural object's image views which are then reconstructed using 3D motions. The proposed scheme is benchmarked with Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to prove the proposed scheme's efficiency. Evaluation on a dataset of about 31,000 cultural heritage images being retrieved from internet collections with many outliers indicate the robustness and cost effectiveness of the proposed method towards a reliable and just-in-time 3D reconstruction than existing state-of-the-art techniques

Keywords: outliers removal, culturalartefact objects, 3D reconstruction, particle swarm optimization, density based spatial clustering of applications with noise

INTRODUCTION

Awareness on cultural institutions like museums is very low making them conclude that their digital presence needs improvements. The web is increasingly used for collecting, storing and referring events or images of the past and present. This is mainly due to the availability of cheap digital hardware. Images range from simple photographs to 3D scans. For example Google's Art Project [1] has National Gallery (London), Palace of Versailles [2] or Van Gogh Museum as the project's contributors for 3D presentations. Khufu Pyramid [3] is fully 3D and can be viewed from any direction. In the creation of documents on relics or eternal subjects only multimedia presentations do not help as they need accuracy of resolution in devices that measure them [4, 5].

Increasing interest in techniques for 3D digitization has resulted in the proposals of various digital modelling techniques for 3D objects and used in visualizations or documentations. Examples are cultural artefact objects [6], healthcare screening and evaluations [7], entertainment gadgets [8]. 3D visualizations are have industrial or commercial applications in reverse engineering, fashion technology, crime investigations and industrial quality control to name a few. In spite of a variety of applications of 3D imaging they are limited by complexity and duration while measuring or using metrics. They typically use bi-directional measurements which require skilled personnel. 3D scanning is a common approach used in e-documentations of cultural artefact s [9]. The study in [10] proposed an automated 3D measurement approach for preserving cultural artefacts while [11] developed software to handle voluminous 3D scanned data. High resolution volumetric maps created for 3D images were used for documenting cultural assets in [12].

Specially acquired 3D images using specialized equipment (wild image collections) can be exploited for creating cultural e-documentations using the medium of internet. The difficulty however in using such collections is their lack structure in storage and calls for content-based filtering techniques to retrieve them based on subjects. Automatic generation of geo-location tags has helped in improving content visualizations, but performances of techniques retrieving them have been found to be low mainly due to mismatches is tagging. Thus, content filtering techniques are imperative to for efficient 3D reconstructions and for while maximizing exploitation of image collections on the internet.

3D images are acquired with a planning phase where strategies for measurements are defined. In the scanning phase, multi-directional measurements catering to each object's digitization are considered. These measurements are then integrated to obtain a complete 3D model. In the final phase techniques like filtering and normalization using triangle meshes are used. Most of these processes are done manually. Scanner placements, defining measurement strategies and integrating obtained views are done with human intervention and by skilled operators. The absence of automation in these phases eliminates in candidature in creating professional digital documents. The main issue in automatic implementations of accurate 3D/4D reconstructions lies in its lack of a proper structure as they happen to be reconstructions evoked due to personal interest. While being displayed on the internet, they manage to form outliers which again recursively affects performances and cost of retrieving algorithms. Robust algorithms for 3D reconstructions in spite of noises do exist [13, 14] but their complexity in computations increases drastically. Hence, quick and efficient matching of images is very critical to 3D reconstructions of images.

This research work proposes automation of content filtering for cultural artefacts in images using Internet based image repositories. Outliers are eliminated by 3D reconstruction algorithms using structures of objects in motion and thus preparing them for digital documentations. LVC (Localized Visual Content) representations are clubbed with textual extractions and geo-tagged information in this work. The main objective of using LVCs is to identify images with same objects in the background but different foregrounds and use them in 3D reconstructions. LVCs identification is done using ORB (Oriented FAST and Rotated BRIEF) as the local descriptor. This work also implements voxel descriptors to extract Haar-like features from Cultural image landmarks. These features are selected based on a voxel's neighbourhood and thus examining its local appearance. A similarity matrix is constructed from the extractions for indicating the similarity of a visual content in images. This localized pair matches yield different models of a cultural artefact's views in a 2D image space. Outliers are identified by considering each image as a point of a multi-dimensional hyperspace where its coordinates depict its position in the hyperspace. As they project the closeness of images in the hyperspace spatially applicable PVDBSCAN can be used to remove outlier points efficiently. DBSCAN is mainly used due to its capability to eliminate outliers with robustness. The use of Information theory clustering in this research work helps quicken image matching in image views of a cultural artefact.

RELATED WORK

Images can be identified in searches and their features can also be extracted by CBIR (Content-Based Image Retrieval) techniques. CBIRs extract an image's shape, color and texture statistically or through recognitions of pattern recognitions, computer vision and processing signals. These techniques have been applied in the areas of healthcare, crime recognition, publishing and fashion technology.

In [15] this article, we propose an automatic scheme for 3D modeling/reconstruction of objects of interest by collecting pools of short duration videos that have been captured mainly for touristic purposes. Initially a video summarization algorithm is introduced using a discriminant Principal Component Analysis (d-PCA). The goal of this innovative scheme is to extract the frames so that bunches within each video cluster that contains videos of content referring to the same object present the maximum coherency of image data while content across bunches the minimum one. Experimental results on cultural objects indicate the efficiency of the proposed method to 3D reconstruct assets of interest using an unstructured image content information.

In [16] introduce a new approach for large-scale scene image retrieval to solve the problems of massive image processing using traditional image retrieval methods. First, we improved traditional -Means clustering algorithm, which optimized the selection of the initial cluster centers and iteration procedure. Second, we presented a parallel design and realization method for improved -Means algorithm applied it to feature clustering of scene images. Finally, a storage and retrieval scheme for large-scale scene images was put forward using the large storage capacity and powerful parallel computing ability of the Hadoop distributed platform. The experimental results demonstrated that the proposed method achieved good performance.

In [17] the author aims to examine the possibility to extract metric information of historic buildings from historical film footage for their 3D virtual reconstruction. In order to make automatic the research of a specific monument to document, in the first part of the study an algorithm for the detection of architectural heritage in historical film footage was developed using Machine Learning. This algorithm allowed the identification of the frames in which the monument appeared and their processing with photogrammetry. In the second part, with the implementation of open source Structure-from-Motion algorithms, the 3D virtual reconstruction of the monument and its metric information were obtained. The results were compared with a benchmark for evaluate the metric quality of the model, according to specific camera motion. This research, analysing the metric potentialities of historical film footage, provides fundamental support to documentation of Cultural Heritage, creating tools useful for both geomatics and historians.

It can be said that main limitation of the above listed techniques is in their use of global visual information to identify required images amongst noisy images. CBIR approaches have been found to be useful in retrieving images based on user query images, but have their own pitfalls when applied to 3D image reconstructions.

In [18], we proposed a deep learning-based, soft-edge-enhanced depth estimation method and applied it to the 3D reconstruction of Borobudur reliefs. We introduced an edge guidance layer to the depth estimation network to improve the reconstruction accuracy of the relief details. In [19] a new content-based image filtering is proposed to discard image outliers that either confuse or significantly delay the followed e-documentation tools, such as 3D reconstruction of a cultural heritage object. The presented approach exploits and fuses two unsupervised clustering techniques: DBSCAN and spectral clustering. DBSCAN algorithm is used to remove outliers from the initially retrieved dataset and spectral clustering discriminate the noise free image dataset into different categories each representing characteristic geometric views of cultural heritage objects. To discard the image outliers, we consider images as points onto a multi-dimensional manifold and the multi-dimensional scaling algorithm is adopted to relate the space of the image distances with the space of Gram matrices through which we are able to compute the image coordinates. Finally, structure from motion is utilized for 3D reconstruction of cultural heritage landmarks.

The study by Kekre et al [20] used signatures of images and created clusters from its colour features and stored it in a database as codebooks. These stored codebooks were referred with input query images for assessing pertinent visual matches. Visual clustering was focused by Simon et al [21]. Their scheme optimized image selections by creating a scene summary with canonical image views. One major limitation of these methods is in their use of image global features for encoding visual information which is not suitable for 3D image processing. Reconstructing 3D images have to select multiple views of the object from the background's spherical coordinates instead of matching similarity of information. Global visual representations fail to describe multiple object view instances as the foreground and background is two dimensional. User's geometric intuitions were used in [22]. The technique learnt about the feature space based on these intuitions. RF (random Forest) algorithms can generically handle two classes of images namely relevant or irrelevant. RF can also be viewed as relevant or irrelevant groups. For example, while classifying cars, irrespective of the car colour they can be categorized as relevant while other as irrelevant in feature space representations. The study used multi-level relevance scores and integrated relative degrees of relevance on images based on the query of the user. Geo-clustering figured in the work of [23] where landmark images were retrieved. The study's engines combined geo information and hierarchical agglomerative visual clustering for getting dense groups. In the work's phase 1, image information with similar geo-tags was extracted. After this phase, outliers in images were eliminated using visual clustering.

METHODOLOGY

The main purpose of this proposed work is to use visual clustering in CBIRs for 3D reconstructions of images and their effective use in e-documentation of cultural artifacts. The foremost research challenge lies in reconstructing remote unstructured images using the heterogeneous internet medium. Very huge volumes noise/outlier in images exist when they are placed on the web. The present work's emphasis lies in identifying cultural artifacts using CBIRs and depicted as dotted lines in Figure 1. Geo-tags text is used by the searching part of the proposal where internet multimedia databases like Picasa, Flickr and Photosynth are searched for identifying cultural objects which can effectively depicts cultural artifacts in these data stores. Most of these images accumulate noise in retrievals due to multiple reasons: user generated monumental photos get overlaid by personal content hiding objects; annotations of images are complex and ad-hoc like a Parthenon temple is the same as Acropolis hill in annotations though they are different; people may refer to important parts of an artefact in their own unique way (Athenian view from the hill is mostly annotated as Acropolis) and various objects share the same name making it a complex issue. Hence, this work attempts to overcome these hurdles using the proposed framework.

The proposed methodology eliminates outliers and then important characteristics are discriminated for an objects multiple geometric views which then become the input for a 3D engine. PVDBSCAN clustering eliminates image outliers. This technique is used as it is more robust than other techniques which fail to separate outliers from relevant images. The selected and identified images are then processed for selection of M elements that can relate to accuracy of builds in 3D reconstructions. This is accomplished by selecting more "uncorrelated" information (canonical geometric views) from image data. Spectral information theoretic clustering is used mainly because of its efficiency in defining sub-graphs for depicting maximum coherence in intra cluster and the minimum coherence is used in inter-clusters.

PROBLEM FORMULATION

Assuming X is the set of images retrieved from with text/geo tags from multiple datasets on a user query for identifying cultural artefact objects, then X can be viewed as a combination of two mutually exclusive sets C (relevant images) and O (Outliers) such that $X = C \cup O$, and $C \cap O = \emptyset$. If Z is the set of retrieved images then C and O have to be computed. The set of representative images M that can enhance 3D reconstructions accurately exists in C. 3D reconstructions accuracy raises as M increases while increasing the complexity. Thus, the problem turns to selecting M from C for improving accuracy of 3D reconstructions. M can be selected

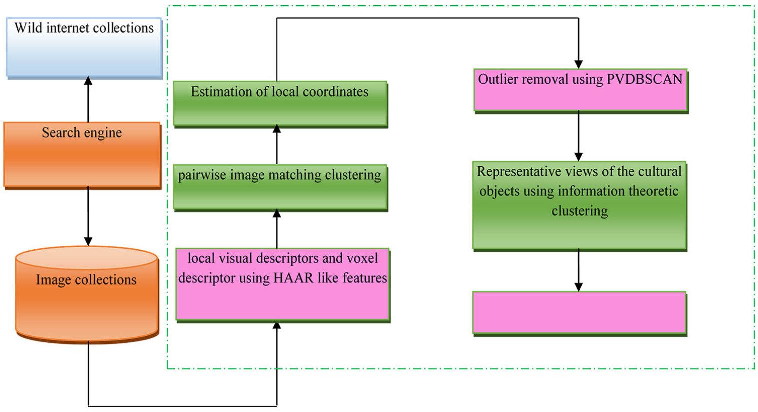
Wild internet collections
Search engine
Image collections
Estimation of local coordinates
pairwise image matching clustering
local visual descriptors and voxel descriptor using HAAR like features
Outlier removal using PVDBSCAN
Representative views of the cultural objects using information theoretic clustering

Fig. 1. Proposed methodology for 3D cultural artefacts image reconstructions

by dividing C into mutually exclusive groups/clusters where C=C$_r$r=1, 2… M. Similarity of canonical forms do not yield accurate 3D reconstructions and hence M consists of "uncorrelated" cluster samples belonging to different clusters for maximizing 3D reconstructions accuracy. Also M should not have redundant information. The problem is depicted as Equation (1)

$$\begin{aligned}&\hat{C}_r\text{: min}\sum_{i=1}^{M} P_r = \sum_{i\in C_r, j\notin C_r} d_{i,j}\\ &and \text{ max}\sum_{i=1}^{M} Q_r = \sum_{i\in C_r, j\in C_r} d_{i,j}\end{aligned} \tag{1}$$

where: $\hat{C}_r$ – C's optimal rth partition in M and $d_{i,j}$- distance between images $i,j \in C$.

LHS of (1) reduces correlations between clusters in M for obtaining "uncorrelated" images. RHS in (1) maximizes class coherences. The issue in this formulation is computing $d_{i,j}$ along with C. Hence, this work formulates multi-dimensional images as data points in a feature space. The distances between points cab be computed using cMDS (classical Multi-Dimensional Scaling) algorithm [24] based on connection between points and visual matches of images (Theorem 1 [25]). Visual coherences between images are treated as Euclidean distances between points to formulate the required set C. The set of outlier data points are also identified. Any center-based technique would fail in identifying C and O as their divisions lead to more number of outliers in partitioning. Densitybased clusters overcome this issue and frame C as it is formulated from denser areas while also identifying O. This work uses PVDBSCAN base with strong constraints which ensures most confident images get aligned with their corresponding sets. The focus of this work thus is in creating a compact set with multiple geometric views of an object for 3D reconstructions and exclude outlier images in processing for reduced computational complexity.

GEOMETRIC INVARIANT DESCRIPTOR AND HAAR LIKE FEATURE EXTRACTION

Assuming N images retrieved from web datasets like Picasa are then extracted to form the set X. N is formed using geo-location/textual tags or they have textual and geo-location information similarity. The images are filtered using Visual features for maximizing 3D reconstructions from N while maintaining computational complexity to a minimum.

ORB based visual content representation

This work uses Visual descriptors for capturing an object's multiple views with geometric perspectives which are used for 3D reconstructions. The descriptors can find visual similarities in images being invariant to affine transformations and convert them into a distance form between two images. The descriptor based on ORB, finds key points in image cornersand associates key points to a descriptive vector.

Assuming FAST detects corner pixels (p_c – pixel) they are processed with a series of binary tests $T = \{\tau_1, \tau_2, \ldots, \tau_n\}$ and n is pre-defined scalar. If, $\ell\,(p_c)$ is an Image patch surrounding then $p_c\tau_i(\ell\,(p_c); q, r)$= 1 = 1 when $I(q) < I(r)$ or otherwise 0. q,r are pixels in $\ell\,(p_c)$ while *I(q), I(r)*, are intensities of the corresponding pixels. The constructed features from T is expressed as Equation (2)

$$f_n^{(I)}(\ell\,(p_c)) = \sum_{1\le i\le n} 2^{i-1}\,\tau_i(\ell\,(p_c); q, r) \tag{2}$$

The patch around a pixel found using intensity's centroid orientation in a corner using Equation (3)

$$\theta(\ell\,(p_c)) = \text{arctan}\left(m_{01}(\ell\,(p_c)), m_{10}(\ell\,(p_c))\right) \tag{3}$$

where: $m_{01}(\ell\,(p_c)), m_{10}(\ell\,(p_c))$ – unprocessed moments of the patch $\ell\,(p_c)$. Feature vector $f_n^{(I)}(\ell\,(p_c))$ projection in angle $\theta(\ell\,(p_c))$ outputs a rotation-invariant binaryrepresentation vector, $f_n^{(I)}(\ell\,(p_c))$, of patch $\ell\,(p_c)$. Visual content's matrix $F^{(I)} \in \{0,1\}^{K\times n}$can be equated to Equation (4)

$$\begin{aligned}F^{(I)} &= \Big[f_n^{(I)}(\ell\,(p_1)),\\ f_n^{(I)}(\ell\,(p_2)) &\ldots f_n^{(I)}(\ell\,(p_k))\Big]^T\end{aligned} \tag{4}$$

Voxel descriptor

3D voxel descriptors model Haar-like features [26] $Hr_F(I)$ which can be obtained a sample voxel's neighbourhood. Haar-like features are extracted in the study using ten filter templates as shown in Figure 2 using the scales of 22–8 voxelsand 11 translations along each axis. Each feature is computed as the difference between the sum of intensities inside the grey region and the sum of intensities inside the white region. Extractions are constrained to 31×31 voxel blocks which can almost 91,594 Haar-like features per image voxel. These features do not use a complete Haar wavelet band in computations.

Degree of Visual Resemblance: this work calculates voxel wise similarities between two images in their corresponding points. This helps in identifying cultural artefacts from different images with geometrical perspectives. An image corner's closest neighbours are matched in images as ORB's key points are binary patterns. The matches are executed using multi-probe Locality Sensitive Hashing which identify nearest neighboursusing Hamming distance, DH. If ORB extracts an image A then $k_i^{(A)}$ is the ith image corner and described as $f_n^{(A)}(\ell\ (p_c))$ (vector), then the image B's most relevant corners $k_i^{(B)}$ w.r.t $k_i^{(A)}$, can be obtained using minimization achieved by Equation (5)

$$j_i \ = \ \underset{j=1,..K}{\arg\min}(D_H(f_n^{(A)}\left(\ell\ (p_i)\right), f_n^{(B)}\ell\ (p_i))) \quad (5)$$

where: $k_i^{(A)}$ and $k_i^{(B)}$ are corresponding key points of the images. Once all corresponding points are detected, the set $M^{(A\to B)}$ contains key-point pairs $k_i^{(A)}$, i=1,2,…,K with corresponding B's $k_i^{(B)}$ the set can be defined as Equation (6)

$$M^{(A\to B)} = \left\{\left(k_i^{(A)}, k_{j_i}^{(B)}\right)\middle| i = 1,..K\right\} \quad (6)$$

A Bi-way matching is executed between A and B for obtaining the set M(A,B) which has corresponding matches of the sets. The intersection of M(A→B) and M(B→A)sets can be defined as Equation (7)

$$M^{(A,B)} = M^{(A\to B)} \cap M^{(B\to A)} \quad (7)$$

The chosen matches can be justified by their nearest neighbour. In case of differences in matches, they are compensated by Bi-way matching. K, the count of extracted matches is used in assessing visual similarity between images i = A and j = B using Equation (8)

$$S_{i=A,j=B} = \frac{\left|M^{(A,B)}\right|}{K} \quad (8)$$

where: $\left|M^{(A,B)}\right|$ is the cardinality of the set M (A, B).

The output of (8) is aN×N symmetrical matrix S for N images where $s_{ij} \in [0,1], i,j = 1,2,\dots,N. s_{ij}$=0 when there is absence of relation in the visual contents of images i and j. $s_{i,j} = 1$ implies the images are similar. If D is S's logarithmic version similar images near 0 while dissimilar images show a high value. D is a symmetric square N×N matrix with positive values where the main diagonal has zero values. $D = [d_{ij}] = -\log(S)$ where

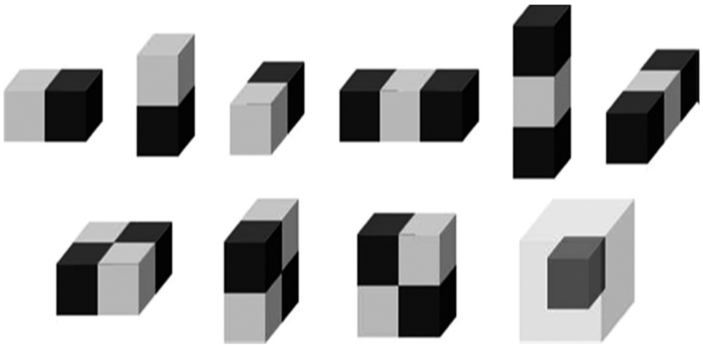

Fig. 2. Filter templates used for extracting 3D Haar-like feature

$d_{i,j}$ is an element of D. The treatment of images as multi-dimensional points helps in identifying and eliminating outliers in the retrieved images as they have low or very low densities when compared to relevant images.

PROPOSED SOLUTION

Conventional center based partitioning schemes like k-means clustering fail to bifurcate relevant images C, from outliers O, properly. Such precise partitions require complicated processes like Information theory clustering. The creation of relevant set of images in this work is done using density variations in a feature space, instead of their positions in the feature space for image partitions.

Estimating spatial densities in images

An image's density can be defined as the count of points u that exist within r the specified in a hyperspace. In an image A, assuming the existence of a relationship that is non-linear then the function $g(A)(\cdot)$ relates r to u as $r = g(A)(u)$. The function also depicts the required distance of r to be a part of the space for u points within r in an image. A line segment that connects these points is defined in the plane (u,r) where c1 = $(u = 1, r = g(A)(1))$ and $cN = (u = N, r = g(A)(N))$ The curve's point at maximum distance is defined by the function which can be detected where a unit vector can defined as Equation (9)

$$u = \frac{C_N - C_1}{\left||C_N - C_1|\right|_2} \tag{9}$$

In the above equation, DBSCAN finds clusters (C_1, C_2) and marks its neighbourhood points as noise. The algorithm has problems with high-dimensional data as it is difficult for the algorithm to define density in high dimensional data. This work focuses on solving this deficiency of DBSCAN while clustering data points with density variations. Real times applications have great utilizations for density based clustering during analysis. Also, partial densities help in revealing clusters that could be formed in a data space's regions. This work uses an improved version of DBSCAN called PVDBSCAN which overcomes hurdles of DBSCAN's single global parameter for clustering data points with variations in density during analysis.

The proposed PVDBSCAN (1) stores its computed k-dist of pixel points for partitioning k-dist plots. (2) The formulated k-dist plots reveal intuitively the densities. (3) The parameter $Epsi$ for each density is chosen automatically. (4) $Epsi$ is used to scan the dataset and obtain different density clusters. (5) Show valid density clusters. Since, everything narrows down to the parameter $Epsi$ its choice is a significant factor and a very complex process. This is overcome in this work by the use of CSP (Core Sample Partitioning) for the selection which is then optimized using PSO (Particle Swarm Optimization) algorithm.

Step 1: Selecting parameter Epsi

Selecting this is parameter is a key step as K-dist plotting are done with this parameter along with an analysis of density levels, though wide variations in density cluster densities may vary, but in the case of equal density, the variations are marginal. This leads to smooth curves getting connected to curves with variations. A dataset has n density levels when there are n smooth curves in the k-dist plot. Varying density dataset is has several densities or n variances in density. Single density datasets yield only one smooth curve in its k-dist plot.

Assuming the neighbourhood of a point p is Nr (p) where $p \in \mathbb{R}^{\mu}$i in a multidimensional hyperspace, then the neighbourhood has points $q \in \Theta$, where its distance w. r. t p <= r and the radius of the clusters is computed using Equation (10):

$$N_r(P) = \left\{q \in \Theta \middle| \mathrm{d_{p,q}} < r\right\} \tag{10}$$

The output of (10) is optimized by PSO which is a nature inspired algorithm modeled on the dynamic movements and social behavior of fishes, birds and insects [30–31]. Velocity changes are also computed for cluster data points in a dimension d, and velocity changes are updated using Equations (11) and (12):

$$V_{id} = \omega_{id} V_{id} + C_1 r_1 (p_{id} - X_{id}) + \\ + C_2 r_2 (p_{gd} - X_{id}) \tag{11}$$

$$X_{id} = X_{id} + V_{id} \tag{12}$$

where: r_1, r_2 , – random numbers in the interval [0, 1] based on global/local radius ranges in clusters,
V_{id} – momentum,
ω_{id} – inertia,
C_1 – cognitive learning parameter and
C_2– social collaboration parameter,
$X_{id} = (x_{i1}, x_{i2}, \dots x_{id})$ – energy of the i th particle,
$P_i = (p_{i1}, p_{i2}, \dots p_{id})$– best prior positions based on fitness energy values.

Inertia weight balances exploitation and exploration processes. It finds the rate of particles between current and previous velocities in the current time step. Multiple types of inertia weights like constant weights or random weights have been used. PSO was modified for updating velocity of the *i* th particle in a dimension *d* using Equations (13) and (14):

$$V_{id} = \lambda\left[\omega_{id} V_{id} + C_r r_1 (p_{id} - X_{id}) + + C_2 r_2 \left(p_{gd} - X_{id}\right)\right] \tag{13}$$

$$X_{id} = X_{id} + (\omega V_{id}) \tag{14}$$

where: λ – factor of convergence and computed using Equation (15)

$$\lambda = \frac{2}{\left|2 - C - \sqrt{C^2 - 4C}\right|} \tag{15}$$

where: $C = C_1 + C_2$,

The proposed algorithm ω_{id} is computed using Equation (16)

$$\omega_{id} = 0.9 - \frac{t}{T_{max}} * 0.5 \tag{16}$$

where: t – total iterations and
T_{max} – max no of iterations. An increasing t, decreases ω linearly from 0.9 to 0.4.

Step 2: Varied-density clustering

This study modifies DBSCAN for the parameter Epsi where i=1, 2, 3… n, n being density levels. The parameter is ordered as k-dist line curves implying $Epsi < Epsi + 1 (i < n)$ DBSCAN, marks Epsi+1, points in clusters based on $Epsi$ value. $Ci - t$ depicts the points belonging to a cluster t with i density. The marking are done to avoid re-processing of points in iterations. Unmarked points are treated as outliers and only $Ci - t$ are displayed as results. The point p with neighbourhood cardinality $\|N_r(p)\| \geq u$ is the core sample. When the point $q \in N_r(p)$ with cardinality $\|N_r(p)\| \geq u$ then p and q are directly density-reachable or when exists a chain of points $p_1, \dots, p_n$ with $p_1 = p$ and pn=qsuch that pi is directly density-reachable from pi+1. Though CSP partitioning with stringent norms increases FPR (False Positive Rates) values when outliers are removed, it also reduces false negatives. Thus, outlier eliminations may create confusions for algorithms in selecting the most suitable views for 3D reconstructions, but they generate a compact subset of visually similar images in a lesser time frame. The generated subsets also need to be identified as groups that contain visual representation of the cultural artefact which can be regenerated into 3D view.

Thus, the next phase of this research work is to identify and separate the processed image data into regions of relevance. Image regions with spherical representations are input into a 3D reconstruction engine which uses these inputs to construct a 3D image, a complex process. In this process the point similarities in images can be depicted as a graph G = (V, E) where each vertex of a graph i is a data point x_i. Assuming i and j are vertices similarity between them w_{ij} namely x_i and x_j is a non-zero value and weighed by w_{ij} which highlights the distance between the vertices. Clustering these data points can now be reassessed using similarities in the graph and narrows down to find partitions where the edges in the graph are identified by their weights i.e. lower weights form a cluster and higher weights forma different cluster. The degree of weights, a matrix $DM = \{d_1 \dots d_n\}$ can be defined as a diagonal matrix. DM can be normalized into a Laplacian Matrix L where $L = 1 - I - D - 1W$. The term Graph Laplacian for such matrices somehow seem to rare or non-existent in literature [29]. Spectral Clustering variants are mostly use

Laplacian matrix's eigenvectors to symbolize abstract data points as real points in a Euclidean space. Clustering algorithms like k-means then use these vectors to obtain clusters from the given inputs data. A transitional matrix with non-negative values whose sum of elements equals 1 is defined as $P(n\ x\ n) = D - 1\ W = I - L$. P is then used to define a Marko chain corresponding to the graph's random walk. The n valued stationary Markov chain for $X = \{X_t\}$ is defined in Equation (17):

$$P_{ij} = (D^{-1}W)_{ij} = p(X_2 = j|X_1 = i) = \frac{w_{ij}}{\sum_k w_{ik}} \quad (17)$$

The probability P_{ij} in formulating from i to j (by transformation) in a singular step is directly proportional to the weight of the edge w_{ij}. If it is assumed that $\pi = (\pi_1, \ldots, \pi_n)$, where $\pi i = \frac{d_i}{\sum_j d_j}$, then $P^T\pi = \pi$ can also be verified, thus it can be said that the graph is a non-bipartite connected graph where π signifies unique and stationary distribution of P's Markov chain [27]. Thejoint stationary probability of X_1 and X_2 is defined in Equation (18):

$$p(X_1 = i, X_2 = j) = \frac{w_{ij}}{\sum_k w_{kl}} \quad (18)$$

and the combined distribution of (Y_1, Y_2), random variables in the clusters can be defined as Equation (19):

$$p(Y_1 = i, Y_2 = j) = p\big(X_1 \in A_i, X_2 \in A_j\big) =$$

$$= \frac{1}{vol\ (V)} \sum_{k \in A_i, l \in A_j} w_{kl} \quad (19)$$

where: $vol(V) = \sum_{ij} w_{ij}$.

This work additionally applies information theory to obtain minimal losses in random walks of the clusters. The information shared while clustering C = {A1, ..., Am}, defined as Equation (20):

$$MI(A_1, \ldots A_m) = I(Y_1; Y_2) =$$

$$= \sum_{ij} p(Y_1 = i, Y_2 = j) \log \frac{p(Y_1 = i, Y_2 = j)}{p(Y_1 = i)p(Y_2 = j)} \quad (20)$$

The above equation results are rewritten in formulating a matrix from MI vectors. If $a_r = [\cdots a_r^u \cdots]^T$ is the index vector where the uth entry equals one, the uth image is appended to Cr 'srth partition as depicted in Equation (21).

$$a_r^u = \begin{cases} 1\ iftheu^{th}iamgeisassignedtor^{th}partition \\ 0\ otherwise \end{cases} \quad (21)$$

It can be comprehended from the above discussions that when an image is chosen from retrieved images, updating and re-estimations are needed in the cluster from where it is chosen. Same representations are maintained for reduction of computational complexities. Hence, the selections are sorted based on their representations and retrieved for maximizing dissimilarity and successful regenerations of 3D images.

EXPERIMENTAL RESULTS

The proposed technique's implementation results are displayed in this section. The design, analysis and validation results for 4D modelling of cultural artefacts using 3D reconstructions have been presented. The dataset images selected from the wild (being captured and stored in image repositories for non-professional 3D reconstruction use) and support the aim of digital libraries European and UNESCO Memory of the World to build a sense of a shard European cultural history and identity.

Figure 3 is considered as a input image, which is initially retrieved image set from Flickr by using as query the keyword "Porta Nigra". Though a 3D reconstruction engine, such as the structure from motion algorithm, can exclude image outliers, the respective computational cost is quartically increasing with respect to the number of input images. This makes 3D reconstruction process practically impossible to be implemented for real-time application scenarios. To solve this problem, a content-based filtering is required to "sort" the retrieved data according to "their contribution" to the 3D reconstruction. Thus, we first need to discriminate the relevant/irrelevant image data as shown in Figure 4. Figure 5 shows the relevant image set to localize those images that represent as much as possible the different canonical views (geometric perspectives) of the cultural object.

Evaluation results of the proposed scheme

The proposed modified PVDBSCAN and clustering algorithms use Precision, Recall and F1 Score in evaluations. These values have been represented in Equation (22), (23) and (24):

$$p_r = \frac{\left|\left|C_{vs} \cap C_{ret}\right|\right|}{\left|\left|C_{ret}\right|\right|} \quad (22)$$

$$r_e = \frac{\left|\left|C_{vs} \cap C_{ret}\right|\right|}{\left|\left|C_{vs}\right|\right|} \quad (23)$$

$$f_1 = 2.\frac{p_r.r_e}{p_r + r_e} \quad (24)$$

Where, Cvs denotes visually similar images while Cret are this work's identified similar images. Annotated images amounting to 31,000 images were split into two categories (sets), one for relevance and other for outliers. PVDBSCAN density-based clustering discriminated outliers from relevant visuals in images, thus helping the construction the two set Cvs. Cret was created using Proposed density-based partitioning which is compared to various existing approaches like K-Means algorithm, Density-based partitioning based on core samples partitioning (DBSCAN-CSP), Mean shift algorithm, Density-based partitioning algorithm (DBSCAN).

The proposed PVDBSCAN was benchmarked with other similar clustering approaches. The performance analysis of DBSCAN and PVDBSCAN for elimination of outliers is depicted in Figure 6. The results show an enhanced performance in PVDBSCAN outlier removal operations when compared to DBSCAN. This enhancement is caused as PVDBSCAN yields more false negatives when compared to DBSCAN's increased false positives. This implies center based

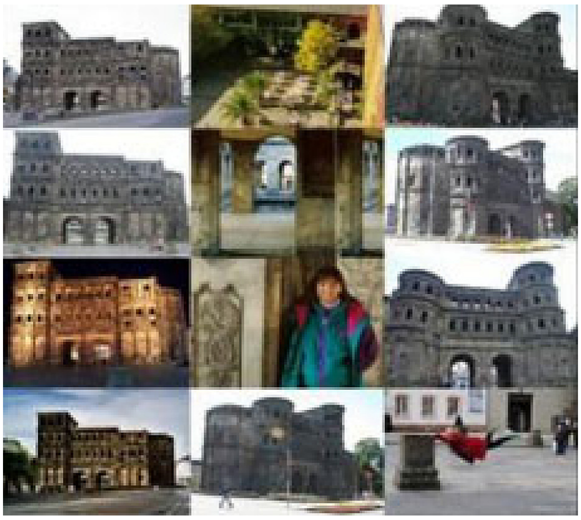

Fig. 3. Input image

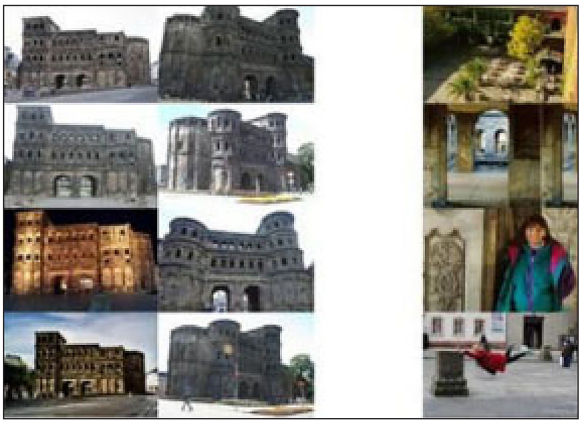

Fig. 4. Outlier removed image using PVDBSCAN

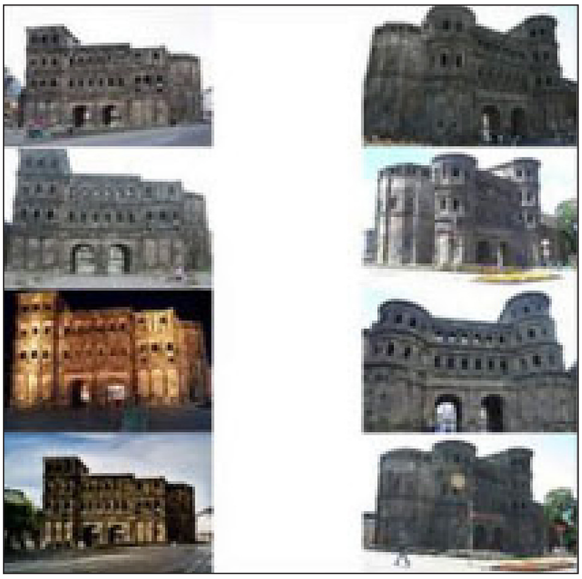

Fig. 5. Spectral clustering to discriminate images depicting the rear and the front view of the monument

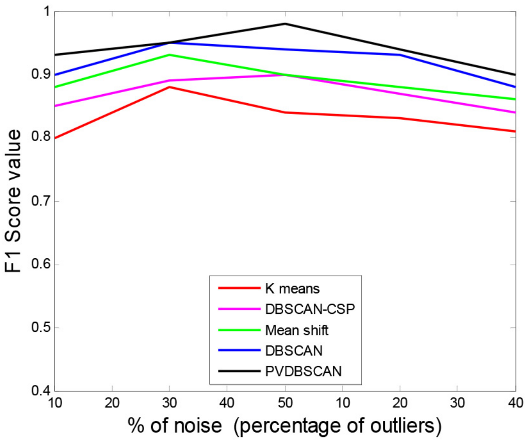

Fig. 6. F1 Score regarding partitioning performance using the different clustering methods

clustering is overwhelmed by density based clustering. PVDBSCAN manages to outperform k-means and Mean-Shift in evaluations. CSPs show better performances than DBSCANs and specifically in terms of increased image outliers. These results on cultural artefact have been obtained by averaging F1 scores in a "wild" image dataset.

Comparative precision and recall values of the benchmarked techniques, DBSCAN, PVDBSCAN and of CSP for noise identification are shown in Figure 7 which is based on average precision, recall values on different cultural artefact objects retrieved from the wild. It can be verified that CSP partitioning yields more false positives and less false negatives. Noise relates to the count of image outliers in web retrieved image dataset. Noisy images are identified based on their visual content and as noisy images in the initial retrieved bring won any proposed system's performance. Spectral clustering removes noises/outliers in a multi-dimensional space. In the proposed work N images were selected for accuracy in reconstructions irrespective of the time taken or computational cost incurred. This work defines reconstruction accuracy using cardinality as:

$$A = \frac{|C_n \cap C_e, i|}{|C_n|}.$$

It is clear from Figure 7 that when n images were extracted in cases of n/5, 2n/5, 3n/5, 4n/5, the corresponding accuracies were 20, 40, 60, 80and 100 in percentages. The proposed work's scheme also performs better in these evaluations when compared with min cut and normalized cut spectral clustering algorithm. Figure 8 displays reconstruction accuracy results.

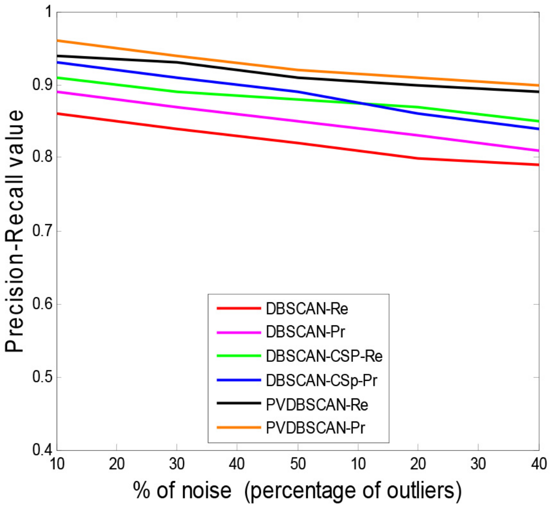

Fig. 7. Precision and recall using the different clustering methods for removing outliers

It is evident from Figure 9 that algorithms perform better with increasing count of samples. This also implies that as the clusters count increases, selection of true positives increases. Min cut spectral clustering performs the worst in reconstructions. Spectral clustering approaches outperform spectral clustering with min cut and k-Means in all cases.

Finally, Figure 10, presents a 3D reconstruction for the Porta Nigra monument. For this reconstruction 30 images dataset was used that contained 20% of outliers.

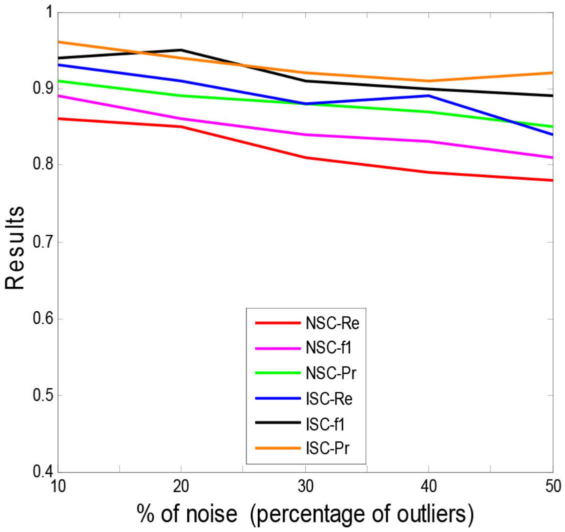

Fig. 8. Precision, recall and F1 score diagram for spectral clustering and Information theory clustering algorithm versus the noise of the initially retrieved image collections

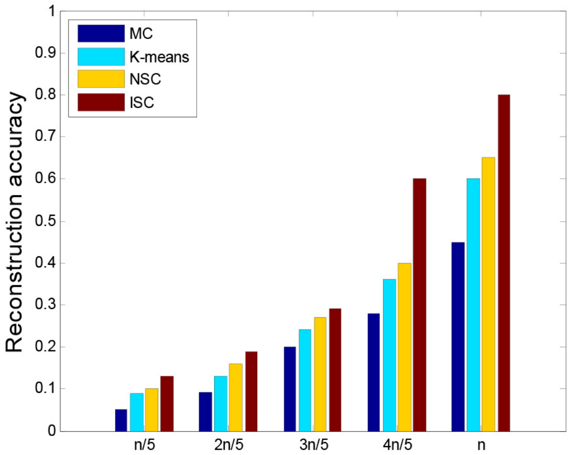

Fig. 9. Reconstruction accuracy in regard to the number of selected representatives

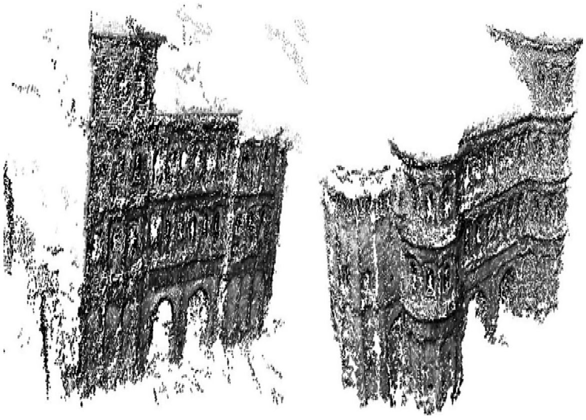

Fig. 10. 3D reconstruction of rear and front view sides of Porta Nigra. For this reconstruction 30 images were used that contained 20% of outliers

CONCLUSIONS

The main goal of the proposed work is to discard image outliers that are often retrieved from such internet collections selecting few but appropriate images needed to be fed as input in the 3D reconstruction process. Initially, local visual descriptors are extracted to capture different geometric properties and perspectives of an object. This work introduces a voxel descriptors are proposed to extract Haar-like features from Cultural artefact landmarks. The Haar-like features are computed from the neighborhood of each sample voxel under consideration, thus representing the local appearance of the voxel. This way, construct a similarity matrix that indicates how close the visual content of two images. In order to unsupervised remove the image outliers from the retrieved image set (that is, without any knowledge), each image is considered as a point onto a multi-dimensional hyperspace manifold. In this work proposed a PVDBSCAN based clustering algorithm is to set more strict criteria for partitioning the dataset so that only the most confident image inliers will be included in the relevant dataset. In this case the focus is to create a compact relevant set that contains all different geometric views of an object. Then, partition this "relevant" dataset to find images that contain the most representative geometric perspectives of an object. These representative views are used for a computational efficient 3D reconstruction without spoiling its performance. Experimental results showed that the system is capable to eliminate outliers from the initial retrieved datasets, even if they contain a large percentage of noise. In addition, the information theoretic clustering algorithm can define different geometric views of an object reducing the cost of 3D reconstruction without spoiling its performance. Information theoretic clustering algorithm based on applying a random-walk associated with the affinity matrix of the data points and computing the mutual information between visited clusters. The constructed ground truth data are used to evaluate the efficiency of the image clustering algorithm applied to estimate the most representative geometric views of an object. Images are placed onto a multidimensional manifold and the spatial density is exploited for partitioning the space into two disjoint subspaces containing the visually similar and non-similar (outliers) images. Obviously, the percentage of outliers in the initial retrieved dataset determines the spatial density of the multi-dimensional space and thus it affects the performance of the algorithm.

REFERENCES

1. Google Art Project, http://www.googleartproject.com/.
2. Grand Versailles Numérique, http://www.gvn.chateauversailles.fr.
3. The Khufu Pyramid, http://www.3dvia.com/3d_experiences/view_experience.php?experienceId=1.
4. Bunsch E., Sitnik R. and Michoński J. Art documentation quality in function of 3D scanning resolution and precision. In: Proc. SPIE 7869, 2011.
5. Bunsch E. and Sitnik R. Documentation instead of visualization – applications of 3D scanning in works of art analysis. In: Proc. SPIE 7531, 2010.

6. Pavlidisa G., Koutsoudisa A., Arnaoutogloua F., Tsioukasb V. and Chamzas Ch.. Methods for 3D digitization of cultural artefact. Journal of Cultural artefact, 2007; 8(1): pp. 93–98.
7. Glinkowski W., Michonski J., Sitnik R. and Witkowski M. 3D diagnostic system for anatomical structures detection based on a parameterized method of body surface analysis. Information Technologies in Biomedicine, Advances in Intelligent and Soft Computing, 2010; 69, 153–164.
8. Zhao Lei and Duan Qing Xu. Immersive display and interactive techniques for the modeling and rendering of virtual heritage environments. In: International Conference on Information Management and Engineering, 2009, 601–606.
9. Barone S., Paoli A. and Razionale A.V. 3D virtual reconstructions of artworks by a multiview scanning process. In: 18th International Conference on Virtual Systems and Multimedia (VSMM). 2012, 259–265.
10. Karaszewski M., Sitnik R., Bunsch E. On-line, collision-free positioning of a scanner during fulIyautomated three-dimensional measurement of cultural artefact objects. Robotics and Autonomous Systems, 2012; 60(9):1205–1219.
11. Sitnik R, and Karaszewski M. Automated Processing of Data from 3D Scanning of Cultural artefact Objects. In: Proceedings of the Third international conference on Digital heritage, 2010.
12. Ioannides M., Fellner D., Georgopoulos A., Hadjimitsis D.G. (Eds). Digital Heritage. In: 3rd International Conference dedicated on Digital Heritage. 2010: 28–41.
13. Bay H., Tuytelaars T. and Gool L.V. SURF: Speeded up Robust Features. In: Computer Vision–ECCV 2006. 2006, 404–417.
14. Wu C., Agarwal S., Curless B. and Seitz S.M. Multicore bundle adjustment. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2011,3057–3064.
15. Wu Doulamis, A. Automatic 3D Reconstruction From Unstructured Videos Combining Video Summarization and Structure From Motion. Frontiers in ICT, 2018, 5, 29.
16. Cao J., Wang M., Shi H., Hu G., & Tian Y. A new approach for large-scale scene image retrieval based on improved parallel-means algorithm in mapreduce environment. Mathematical Problems in Engineering, 2016.
17. Condorelli F., Rinaudo F., Processing historical film footage with Photogrammetry and Machine Learning for Cultural Heritage documentation. In: ACM Multimedia Conference Proceedings, October 2019, Nice, France. ACM, NY, USA. 8 pages. https://doi.org/10.1145/3347317.3357248
18. Pan J., Li L., Yamaguchi H., Hasegawa K., Thufail F.I. & Tanaka S. 3D reconstruction of Borobudur reliefs from 2D monocular photographs based on soft-edge enhanced deep learning. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 183, 439–450.
19. Makantasis K., Doulamis A., Doulamis N. & Ioannides M. In the wild image retrieval and clustering for 3D cultural heritage landmarks reconstruction. Multimedia Tools and Applications, 2016, 75(7), 3593–3629.
20. Kekre D.H.B., Sarode T.K., Thepade S.D. and Vaishali V. Improved texture feature based image retrieval using Kekre's fast codebook generation algorithm. In: Pise S.J. (Ed.) Thinkquest~2010, Springer, India, 2011.
21. Simon I., Snavely N. and Seitz S.M. Scene Summarization for Online Image Collections. In: IEEE 11th International Conference on Computer Vision, ICCV 2007, 1–8.
22. He X., W.-Y. Ma, and H.-J. Zhang. Learning an Image Manifold for Retrieval. In: Proc. ACM Multimedia, 2004.
23. Zheng Y-T, Zhao M, Song Y, Adam H, Buddemeier U, Bissacco A, Brucher F, Chua T-S and Neven H. Tour the world: Building a web-scale landmark recognition engine. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009, 1085–1092.
24. Cox T., Cox M., Multidimensional Scaling, Second Edition, Chapman & Hall/CRC, 2000.
25. Cayton L. Algorithms for manifold learning, University of California, San Diego, Tech. Rep. CS2008–0923, 2005.
26. Tu Z. and Bai X. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2010; 32(10):1744–1757.
27. von Luxburg U. A tutorial on spectral clustering. Statistics and Computing, 2007; 17(4):395– 416.