HIERARCHICAL ONLINE IMAGE REPRESENTATION BASED ON 3D CAMERA GEOMETRY

Sang Min Yoon and Holger Graf
GRIS, TU Darmstadt, Rundeturmstrasse 10, Darmstadt, Germany
ZGDV, Computer Graphics Center, Rundeturmstrasse 10, Darmstadt, Germany

Keywords: Hierarchical image clustering.

Abstract: Within this paper, we present a hierarchical online image representation method with 3D camera position to efficiently summarize and classify the images on the web. The framework of our proposed hierarchical online image representation methodology is composed of multiple layers: at the lowest layer in the hierarchical structure, relationship between multiple images is represented by their recovered 3D camera parameters by automatic feature detection and matching. At the upper layers, images are classified using constrained agglomerative hierarchical image clustering techniques, in which the feature space established at the lowest layer consists of the camera's 3D position. Constrained agglomerative hierarchical online image clustering method is efficient to balance the hierarchical layers whether images in the cluster are many or not. Our proposed hierarchical online image representation method can be used to classify online images within large image repositories by their camera view position and orientation. It provides a convenient way to image browsing, navigating and categorizing of the online images that have various view points, illumination, and partial occlusion.

1 INTRODUCTION

As the use of digital cameras, cell phones, or PDAs with embedded cameras is increasing, managing, browsing, querying and summarizing photos from personal image libraries or on the web is becoming more critical. Online image retrieval and browsing applications (Dent et al, 2001, Jhanwar et al, 2002, Krishnamachari et al, 1999 , Qian et al, 2000) were developed encouraging people to freely explore any place in the world and discover interesting locations and photographs.

In various web applications, geographical online map services such as Google Earth or Virtual Earth have become very popular with web-users. These applications allow the users to view and navigate their way through high resolution satellite images from within their home environment resp. their desk. It also offers local information and photographs of specific places, as well as the ability to view different geographic levels of detail. If a specific place is very popular and interesting, many photographs are uploaded and geo-tagged (Jaffe et al, 2006) with this location by many users.

Nevertheless, online images which are uploaded by numerous users are so various that they have illumination variation, view changes, resolution, partial occlusion, and noise. That is one of the bottlenecks in extracting and matching features from online images. The locations of images uploaded by anonymous users to an online satellite map like Google Earth are not exact and many photographs that contain no tags or titles which represent their location, or are incorrectly geo-tagged (Jaffe et al, 2006). To collect the exact 3D position of the online images, we need to recover the camera's extrinsic parameters of previously uploaded images. Without an appropriate clustering of images, we are distracted in navigation and view of the map.

Our objective of this paper is on how to select the representative images of an interesting site and how to summarize the online images hierarchically taking into account the difference level of the map's zoom in/out. For this, the camera geometric information based image classification is used. Retrieving the representative images of a specific site with its camera's orientation and translation information needs robust feature detection, matching, and image classification (Brown et al, 2005).

Figure 1 shows the structure of our proposed approach

Min Yoon S. and Graf H. (2009).
HIERARCHICAL ONLINE IMAGE REPRESENTATION BASED ON 3D CAMERA GEOMETRY.
In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pages 54-59
DOI: 10.5220/0001790600540059

(a) The concept of our proposed hierarchical online image representation method (b) Left 7 images are the representative images in each layer of the category

Figure 1: The concept of our proposed hierarchial representation method of online images and some representative images in the layers which are used in our experiments.

and representative images of each category and cluster. This hierarchical structure is composed with multiple layers: the highest layer is called "Category", middle layers are defined as "Cluster", and the lowest level shows the relative 3D camera position of the images. Our representative images of each category and cluster is the closest image through clustering optimization and by hierarchical online image clustering. Figure 1 also shows each images in the lowest layer and upper layers which are constructed by online image clustering from a collection of online images.

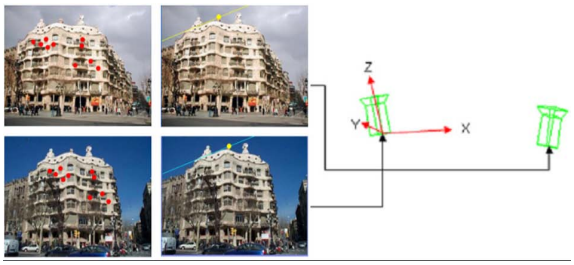
2 PREVIOUS RESEARCH

Research about geographic location based image retrieval (Naaman et al, 2004) or browsing increase during the last few years. The organization of image collections has been accomplished by several classification criteria such as detecting significant events, geographical characteristics within a specific location, or tags in titles of a photographs (Cai et la, 2004, Gao et al, 2005, Jaffe et al, 2006). However, current research efforts (Ahern et al, 2007) for image retrieval based on common context and visual features within online image repositories try to summarize the collection of images. Hierarchical online image representation tasks are composed of various technologies of Image Based Modelling for extracting the 3D camera geometry out of multiple images, feature classification for online image clustering, and image similarity measures (Simon et al, 2007, Snavely et al, 2006). There are also similar approaches about online image representation methods, interactive browsing and exploration of a collection of photographs and online image summarization to represent visual content of a given set (Wang et al, 2006).

Clustering is the unsupervised classification of patterns into groups. Geographical tagging or title of photographs have been used for online image retrieval or browsing applications (Choubassi et al, 2007). Nevertheless, the goal of feature classification and clustering in image processing and computer vision is how to deal with images for classification, in order to separate the images by low-level features such as color, texture, shape, or by high level semantics, or a combination of those (Cai et al, 2004, Chen et al, 1999, Deng et al, 2001, Gao et al, 2005, Jhanwar et al, 2002, Rege et al, 2007). A similarity measure using these features between online images is one of the critical issues because it is still weak in partial occlusion, view translation, orientation and noise (Fergus et al, 2005). Most classification approaches into three main categories; partition, Division, and Aggregation (Cormark et al, 1971). One of the most popular methods in partition methodology is the *k-Means* method (Bradley et al, 1998), the division follows *kd*-tree method. When we compare to previous research related to online image summarization (Simon et al, 2007) and browsing, our proposed agglomerative clustering has the advantage in showing the representative images according to a geographical zoom in/out. As we zoom out of the geographical hotspot, we only show the representative images of the site or building. Otherwise, we zoom in the map, we browse the online images according to the 3D camera's parameters. Another advantage of our proposed algorithm is to provide an efficient hierarchical structure of the online image data set.

3 HIERARCHICAL STRUCTURE OF ONLINE IMAGES

In the following sections, we will explain how we extract the 3D extrinsic camera parameters from multiple images, establish relationship between 3D positions from multiple images and adequate classifica-



(a) Epipolar geometry and matching points with automatic feature detection and matching (b) Recovered 3D camera position with SIFT and RANSAC

Figure 2: Recovering the 3D camera parameters in the world coordinate system by automatic feature detection and matching.

tion methods of images at the upper layers based on the similarity measure derived from the camera's 3D extrinsic parameters.

3.1 Recovery of 3D Camera Parameter

Given N images in the database, the extrinsic camera parameters of each image, $E_i(r,t), (i = 1,...,n)$, where r is 3**x**3 rotation matrix, and t is 1**x**3 translation vectors, are recovered by adequate feature detection mechanism in each camera, feature matching between multiple images, the calculation of the epipolar geometry, and the 3D position estimation within the world coordinate system (Hartley et al, 2004). Figure 2 highlights the process of recovering the $E_i(r,t)$ from multiple images (Chaman et al, 1993). Figure 2-(a) shows the example images within a collection of online images. We have no prior knowledge such as image resolution or tags or title (Jaffe et al, 2006). It also shows matching features after the Scale Invariant Feature Transform (SIFT) (Lowe, 2004) and Random Sample Consensus (RANSAC). Figure 2-(b) displays the relative 3D position and rotation of the camera within the world coordinate system.

From the epipolar geometry and matching points, we extracted the rotation and orientation of the cameras. By calculating the epipolar geometry and extracting the extrinsic camera parameters of multiple images, we can sketch the relationship of the images. With SIFT and RANSAC, the epipolar geometry and 3D camera position of a set of online images is estimated as shown in Figure 3. In this figure, the recovered 3D camera position within the world coordinate system with the online images of Casa Mila, Barcelona, Spain. The lowest layer of our hierarchical structure is constructed using extrinsic camera parameters of the images. In a next step, we describe the clustering of images based on the distance of each 3D camera position within the world coordinate system.

3.2 Online Image Clustering

In online image applications, unsupervised image clustering can be separated with non-hierarchical and hierarchical clustering algorithms (Krishnamachari et al, 1999). In numerous non-hierarchical clustering methods (Goldberger et al, 2006) which are extensively used in data classification or data mining in various areas, *k-Means* clustering is an algorithm to cluster n images based on attributes into k partitions, where k is less than n, to form a k-block set partition of data and to find good local minimum and have linear complexity $O(k_{min})$ with respect to the number of instances. However, the algorithm is sensitive to initial starting conditions and hence must be randomly repeated many times (Davidson et al, 2005). Conversely, hierarchical clustering algorithms are run once and create a *dendrogram* which is a tree structure containing a k-block set partition for each value of k between 1 and n, where n is the number of online images at the lowest level to cluster allowing the user to choose a particular clustering granularity.

Let

$$S = \{E_1, E_2, ..., E_{n-1}, E_n\} \quad (1)$$

be the set of 3D extrinsic parameters, E_i, to be clustered. At the initial status, the number of clusters is same to the number of images, n, and each cluster C_i is represented as E_i for every i. Then we progressively join the closest clusters through the equation shown in equation-(2) until$k$= 1.

$$s(i,j) = D(C_i, C_j), \forall i, j; l, m = argmin_{a,b} d(a,b), \quad (2)$$

$$C_l = Join(C_l, C_m); Remove(C_m) \quad (3)$$

where $s(i,j)$ is the similarity measure between cluster C_i and C_j. In this paper, the similarity measure between clusters are calculated by the Euclidean distance, D, of the camera's 3D extrinsic parameters.

The objective of our hierarchical clustering algorithm is to extract a multi-level partitioning of images based on 3D camera parameters, i.e. a partitioning which groups images into a set of clusters and then, recursively, partitions them into smaller sub-clusters, until some stop criteria are satisfied. Agglomerative hierarchical clustering algorithms start with several clusters containing only one object, and iteratively two clusters are chosen and merged to form one larger cluster. This process is repeated until only one large cluster is left, that contains all objects. Divisive algorithms work in the symmetrical way. Figure 4 shows the original agglomerative hierarchical online image clustering. A Similarity measure between multiple

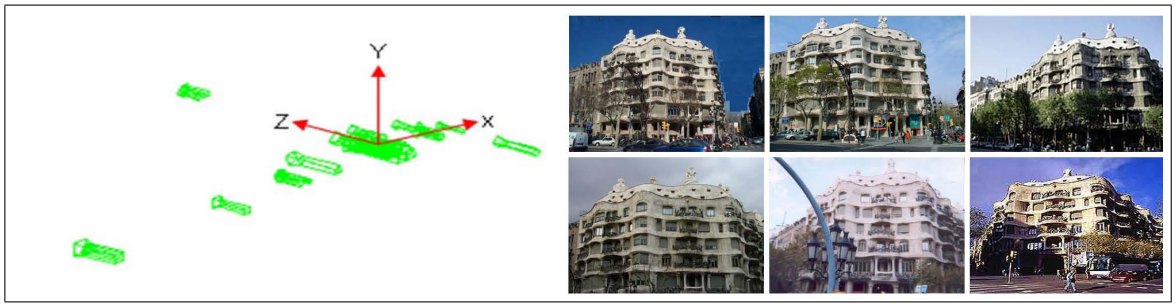


Figure 3: 3D Camera position and orientation of the multiple images which are extracted with SIFT and RANSAC within category and some example images in the cluster.

images is computed by the Euclidean distance between 3D camera's position. From multiple images of the Casa Mila, layers are automatically clustered into 12 layers. This number of layers is different from site to site or change of view points.

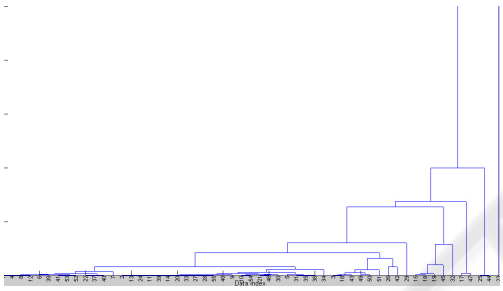


Figure 4: Unconstrained agglomerative hierarchical online image clustering with 3D extrinsic camera parameters.

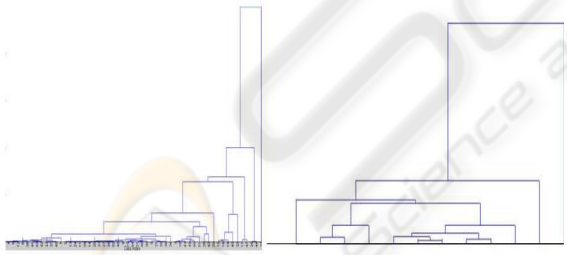


Figure 5: Comparison of unconstrained and constrained agglomerative hierarchical online image clustering methods from recovered 3D camera parameters

The unconstrained version of agglomerative hierarchical online image clustering builds a *dendrogram* for all values of k. If there are many online images at a public site, the *dendrogram* will be high. However, there are many places that have only few uploaded online photographs in the internet. To balance the hierarchical layers of each sites, we impose some constraints on the hierarchical clustering (Zho et al, 2005). When building the Dendrogram, we need the constraint of the numbers of dendrogram, by W-constraint, and B-constraint algorithm. We can prune the dendrogram by starting building clusters at k_{max} and stop building the tree after k_{min} clusters are reached. B-constraint is defined as the distance between any pair of images in two different clusters to be at least B_{min}, and W-constraint requires that for each point x in C_i, there must be another point y in C_i such that the distance between x and y is at most W_{max} (Davidson et al, 2005). At the initial of unconstrained hierarchical clustering approach, the number of clusters was equal to the number of images, n. However, we construct an initial cluster by B-constraints and W-constraints. This constrained agglomerative hierarchical clustering algorithm procedure is shown below:

$$k_{max}, k_{min} = calculateBound(W_{max}, B_{min}) \quad (4)$$

$$s(C_i, C_j) \geq B_{min} \forall i, j, s(x, y) \geq W_{max} \forall x, y \in C_i \quad (5)$$

where eq-(5) is the distance bound for the distance between clusters and within cluster. Within this boundary, we join the closed cluster until the dendrogram is k_{min}. Figure 5 shows the constrained agglomerative hierarchical online image clustering method with the constrained number of k, euclidean distance constraint based on the recovered 3D camera's position.

4 EXPERIMENTS

We lead experiments with various online images which are downloaded from internet. 90 images are used for our experiments with multiple online images of Casa Mila, Barcelona, on the web. Images in Casa Mila are roughly separated with 3 categories and the number of images in each category were 55 images, 21 images, and 14 images. We show the result of our automatic hierarchical online image clustering from the front and near view images. In the previous section, we already represented the recovered camera's

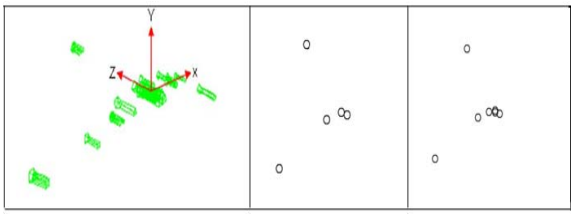
3D position and unconstrained and constrained online image representation methodology.
Our next experiment was comparison of the non-hierarchical clustering algorithm such as *k-Means* (Simon et al, 2007) and Mean-Shift clustering algorithms [Comaniciu et al, 2002, Xu et al, 2005) to check the efficiency in a hierarchical structure as shown in Figure 6. Figure 67-(a) is the result of the *k-Means* clustering when k is 5 and 7. Figure 6-(b) shows the Mean-Shift cluster method, that automatically separates 6 clusters. The representative image in the cluster shown also in figure 6-(b) is the center of gravity of the cluster. Constrained hierarchical online image clustering method is better than non-hierarchical online image clustering in automatically constructing a structure and balancing the hierarchy of the areas that have many online photographs or not. In the end of this paper, we also showed extracted 3D camera position, unconstrained, and constrained hierarchical online image clustering method in 28 images of the Blue Mosque in Instanbul, Turkey in figure 7 .
Table 1 shows the number of hierarchical layers of the category when we tested the unconstrained and constrained hierarchical image clustering method. As shown in table 1, we can see that the constrained hierarchical image clustering method is efficient in balancing the hierarchical layers of the categories and also useful in browsing the representative images of the site.

Table 1: Comparison of the number of hierarchical layers between unconstrained and constrained hierarchical clustering to show the balance of the hierarchical layers in various sites.

Site	Nr. of Image Set	Unconstrained Layer's Nr.	Constrained Layer's Nr.
Site1	55	12	8
Site2	21	7	4
Site3	14	6	3
Site4	28	13	7

5 CONCLUSIONS AND FUTURE WORKS

In this paper, we have presented the hierarchical online image representation method for the efficient browsing and navigation within a geographic online map. We also presented a new approach in order to estimate the relationship between a collection of online images, how to select a representative image using the 3D camera position and orientation, and how to construct a hierarchy of online images with a constrained agglomerative clustering methodology. The



(a) *k-Means* clustering of online images when k=5 and k=7

(b) *Mean-Shift* is separated with 6 clusters and its size of the cluster is proportional to the number of images in the cluster

Figure 6: Non-hierarchical online image clustering like *k-Means* and *Mean-Shift* with camera's 3D extrinsic parameters

hierarchical tree which we presented in this paper can be useful to many applications involving large collections of digital photographs. We are able to sort and view the images that are geographically close to an 3D camera position that users want to watch. It gives convenience and immersion related to applications involving large data on web. Our future work improve this system for industrial applications. The processing time to estimate the 3D position of billion of images is the critical problem in the works of online images. We will focus on the advanced interaction with user and our hierarchical structure is needed for immersive navigation or viewing.

REFERENCES

Bradley, P., Fayyad, U., and Reina, C., 1998. Scaling Clustering Algorithms to Large Databases. *In Proceeding of ACM 4th DKK Conference.*

Brown, M., Lowe, D, G., 2005 Unsupervised 3D Object Recognition and Reconstruction in Unordered Datasets. *5th International Conference on 3D Imaging and Modelling.*

Cai, D., He, X., Li, Z., Ma, W. Y., and Wen, J. R., 2004 Hierarchical Clustering of WWW Image Search Results Using Visual Textual and Link Analysis. *In Proceeding of 12th ACM Multimedia*

Chaman, L., and Sabharwal, 1993. Recovering 3D image parameters from corresponding two 2 images. *In Proceeding of SIGGRAPP Symposium on Applied Computing.*

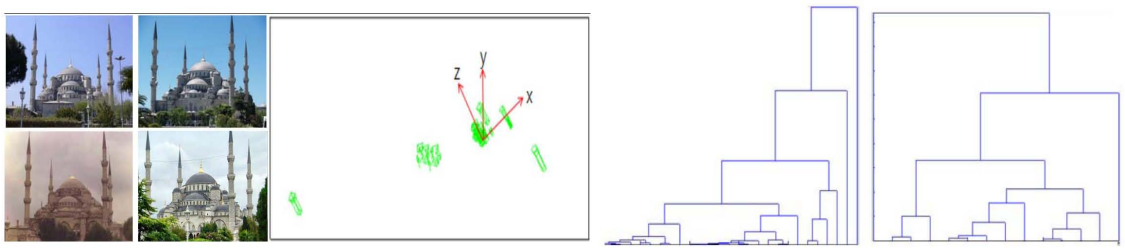


Figure 7: (a) Some representative images in the hierarchical structure of online images and recovery of 3D camera's extrinsic parameters at Blue Mosque, Istanbul, Turkey (b)Unconstrained and constrained agglomerative hierarchical clustering method Hierarchical online image representation method in Blue Mosque, Istanbul, Turkey.

Chen, Y., and Wong, E., 1999. Augmented Image Histogram for Image and Video Similarity Search. *In Proceeding of SPIE Storage and Retrieval for Image and Video Database.*

Comaniciu, D., and Meer, P., 2002. Mean Shift: A robust approach toward feature space analysis. *IEEE Transaction on PAMI.*

Cormark, R., 1971. A review of classification. *Journal of the Royal Statistical Society Series A 134.*

Deng, Y., Manjunath, B. S., Kenney, C., Moore, M. S., and Shin, H., 2001 An efficient color representation for image retrieval. *IEEE Transaction on Image Processing.*

Duda, R. D., Har, P. E., and Stork, D. G., 2001. Pattern Classification. *Wiley second edition*

El Choubassi, M., Nefian, A. V., Kozintsev, I., Bouguet, J.-Y, and YiWu., 2007. Web Image Clustering. *In Proceeding of IEEE ICASSP.*

Fergus, R., Fei-Fei, L., Pernona, P., Zisserman,A., 2005. Learning object categories from google's image search. *In Proceeding of CVPR.*

Gao,B., Lie, T., Qin, T., Zheng, X., Cheng, Q., and Ma, W., 2005. Web image clustering by consistent utilization of visual features and surrounding texts. *In the Proceeding of ACM Multimedia.*

Goldberger, J., Gordon, S., and Greenspan, H., 2006. Unsupervised Image-Set Clustering Using an Information Theoretic Framework. *IEEE Transaction on Image Processing.*

Hartley, R., and Zisserman, A., 2004. Multiple View Geometry. *Cambridge University Press.*

Jaffe, A., Naaman,M., Tassa, T., and Davis, M., 2006. Generating summaries and visualization for large collection of geo-referenced photographs. *In the proceeding of ACM Workshop on Multimedia information Retrieval.*

Jhanwar, N., Chaudhuri, S., Seetharaman, G., Zavidovique, B., 2002. Content-based image retrieval using motif cooccurence matrix. *In Proceeding of Image Vision Computing.*

Krishnamachari, S., and Abdel-Mottaleb, M., 1999. Hierarchical Clustering Algorithm for fast Image Retrieval.

Naaman, M., Song, Y. J., Paepcke, A., and Garcia Molina, H., 2004. Automatic organization for digital photographs with geographical coordinates. *In the Proceeding of ACM/IEEE Joint Conference on Digital Library .*

Lowe, D., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *IJCV.*

Rege, M., Dong, M., and Hua, J., 2007. Clustering web image with multi-modal features. *In Proceeding of ACM Multimedia .*

Qian, R., van Beek, L. P., and Ibrahim Sezan, M., 2000. Image Retrieval Using Blob Histogram. *In the Proceeding of ICME*

Snavely, N., Seitz, S. M., and Syeliski, R., 2006. Photo Tourism: Exploring collection in 3D. *In the Proceeding of SIGGRAPH.*

Svoboda, T., Martinec, D., and Pajdla, T., 2005. A convenient multi-camera self-calibration for virtual environments. *PRESENCE: Teleoperators and Virtual Environments.*

Wang, J., Sun,J., Quan, L., Tang, X., and Shum, H. Y., 2006. Picture Collage. *In the Proceeding of CVPR.*

Xu, D., Wang, Y. and An, J., 2005. Applying a New Spatial Color Histogram in Mean Shift Based Tracking Algorithm. *In Proceeding of Image and Vision Computing.*

Zeng, H. J., He, Q. C., Chen, Z., Ma, W. Y., and M, J. W., 2004. Learing to cluster web search results. *In Proceeding of 27th International ACM SIGIR Conference.*

Zho, Y., and Karypis, G., 2005. Hierarchical Clustering Algorithms for Document Datasets. *In Proceeding of Data Mining and Knowledge Discovery.*