The dimensionality of episodic images

Vishnu Sreekumar (sreekumar.1@buckeyemail.osu.edu)
Department of Psychology, Ohio State University
Columbus, OH 43201 USA

Yuwen Zhuang (zhuang.14@buckeyemail.osu.edu)
Department of Computer Science and Engineering, Ohio State University
Columbus, OH 43201 USA

Simon J. Dennis (simon.dennis@gmail.com)
Department of Psychology, Ohio State University
Columbus, OH 43201 USA

Mikhail Belkin (mbelkin@cse.ohio-state.edu)
Department of Computer Science and Engineering, Ohio State University
Columbus, OH 43201 USA

Abstract

Previous studies (Doxas, Dennis, & Oliver, 2010) show that natural language discourse exhibits a two-scale structure with a lower dimension at short distances and larger dimension at long distances. We attempt to search for the source of this constraint in the visual input that goes into forming episodic experiences in human beings. This information is assumed to be approximated well by images captured by a Microsoft^{TM}Research SenseCam that our subjects used. The hypothesis is that if the same two scale structure is observed here, the constraint is possibly not one that is imposed by the cognitive system. We use and contrast two methods by which images can be represented: the traditional color histogram and a more recently developed color correlogram method. The color correlogram is established to work better for our current purposes. We observe hints of a two scale structure in the correlation dimension plots but these are not conclusive.

Keywords: Episodic Memory; Correlation Dimension; Networks; Graphs.

Introduction

The existing models of episodic memory assume a representation of context. Retrieval of episodes involves reinstatement of context. The current literature does not address the nature of representation of context and the question of how the representation was formed in the first place. Our ultimate goal is to model contextual reinstatement as a search over episodic networks. We begin by looking at the images that people encounter everyday. In a parallel study, graphs of these images are constructed and the structure of the graphs is investigated. People are extremely fast at isolating episodes from memory. Such a search has to be fast and efficient. The graph has to satisfy certain properties for it to be efficiently searchable (Steyvers & Tenenbaum, 2005). We attempt to test the idea that contextual reinstatement can be modeled as a network search. One prerequisite for this model to be feasible is that the episodic network must be quickly searchable.

We encode events into our memory as we encounter and experience them. What kinds of constraints are inherent to this input information? Such a question is motivated by previous studies on natural language discourse where paragraph spaces of corpora of different languages exhibited a two-scale structure (Doxas et al., 2010). Doxas et al. did a correlation dimension analysis on the paragraph spaces of text corpora taken from five different languages and genres. The correlation dimension is a measure of how points within a given distance r scales with that distance. The paragraph spaces were found to exhibit a low dimensional structure at short distances and a higher dimensional structure at larger distances. This is similar to a "weave" structure. For example, if we zoom in to look at a thread that is part of a shirt, the observed dimensionality is one. If we zoom out to intermediate length scales, we would start observing a two dimensional structure. Further zooming out will further increase the dimensionality. The finding of this "weave" structure in natural language discourse raises an important question regarding the origin of this constraint. Is this constraint one that is imposed by the cognitive system or is it a property of the input the system receives that is being mirrored by the cognitive system? We attempt to address this question in the current study. To investigate this, we used a Microsoft^{TM}Research SenseCam to capture images that can be thought of as representative of a person's (visual) episodic experience. A dimensionality analysis was then done on these images.

The paper is organized as follows. The next section outlines the method used to capture and represent the images on which the dimensionality analysis is done. The Microsoft Research SenseCam device is described briefly. Two different image representation schemes and their corresponding distance measures are discussed. The two methods are then contrasted using a definition of a ratio that is based on the requirement that these methods must, among other things, successfully identify images that belong to the same contexts. The subsequent section describes the correlation dimension. The results section discusses the correlation dimension plots for the image sets obtained from different individuals. The paper concludes with a discussion of the structure that is observed in the correlation dimension plots of the image data.

Image Data Collection, Representation and Distance Measures

Microsoft Research SenseCam

To capture a sufficient number of images that can sufficiently represent an individual's visual episodic experience for a period of about a week, we used a Microsoft Research SenseCam. Subjects hung the camera around their necks for about a week each. The SenseCam contains sensors which can detect changes in color, light-intensity and temperature. Changes in these sensor readings can be set to automatically trigger the SenseCam to take pictures. The camera can also be set to a timer mode where pictures can be captured periodically. Our camera captures an image once every eight to ten seconds. The camera has wide-angle (fish-eye) lens that maximizes its field-of-view. The resulting images are particularly useful for studying episodic experience because these images are fragmentary, time compressed, temporally ordered, and have a 'field perspective' (Berry et al., 2006).

HSV Space

The HSV (hue, saturation, value) color space is very different from the better known RGB (red, green, blue) color space.The problem with using the RGB color space is that it is not perceptually uniform. To get a satisfactory representation of the image in the RGB space, the quantization step sizes should be fine such that distinct colors are not assigned to the same bin. This increase in the number of bins affects performance in terms of computation time. The oversampling also produces a larger set of colors than are necessary and this is not an accurate representation of human visual discrimination of colors.

A three dimensional representation of the HSV color space is a hexacone (Stockman & Shapiro, 2001). The central axis represents the intensity. Hue is defined as an angle in the range [0,2π] relative to the red axis such that red is at angle 0, green is at $2\pi/3$, blue at $4\pi/3$ and red again at 2π. Saturation takes values between 0 and 1. Saturation is the depth or purity of the color. It is measured as a radial distance from the central axis. The saturation value is 0 at the central axis and is 1 at the outer surface. As saturation varies from 0 to 1, the corresponding hues vary from unsaturated (shades of gray) to fully saturated (no white component, pure form of the color represented by its hue). In other words, for a low value of saturation, a color can be approximated by a gray value specified by the intensity value and for a high value of saturation, the color can be approximated by its hue. HSV separates out the light-intensity information (luminance) from the color information (chromaticity).

Color Histogram Representation

A color histogram for an image is generated by concatenating 'N' higher order bits for the Red, Green and Blue values in the RGB space (Swain & Ballard, 1991). The histogram is generated by counting the number of pixels with the same color and accumulating it in 2^{3N} bins . We generate such a histogram from the representation of each image in the HSV space. Quantizing the hue component more precisely than the value and saturation components makes the HSV histogram more sensitive to color differences and less sensitive to brightness and depth differences. We found it sufficient to use a (h=30 levels, s=10 levels, v=3 levels) quantization to generate the histograms based on the fact that the human eye is more sensitive to variations in hue and intensity than variations in saturation.

Several distance measures can be used to calculate distance between images (Jeong, Won, & Gray, 2004). These include the histogram euclidean (HE) distance and the histogram intersection (HI) distance (Smith & Chang, 1995, 1996). A Kullback-Liebler divergence (Greenspan, Goldberger, & Ridel, 2001) measure is also discussed which has been established to work better than the HE and HI measures in information retrieval tasks (Goldberger, Gordon, & Greenspan, 2006).

Histogram Euclidean Distance If **h** and **g** represent two color histograms, the euclidean distance between them is given by

$$d^2(h,g) = \sum_A \sum_B \sum_C (h(a,b,c) - g(a,b,c))^2 \qquad (1)$$

A,B and C are the three colors (RGB or HSV). In this formula, all bins contribute equally to the distance and only identical bins in the respective histograms are compared.

Histogram Intersection Distance The histogram intersection (HI) distance (Swain & Ballard, 1991) between **h** and **g** is given by

$$d(h,g) = \frac{\sum_A \sum_B \sum_C \min(h(a,b,c), g(a,b,c))}{\min(|h|, |g|)} \qquad (2)$$

$|\mathbf{h}|$ and $|\mathbf{g}|$ are the number of samples in the respective histograms. The sum is normalized by the histogram with the lesser number of samples. We used the histogram intersection distance for our initial analysis. The distance tends to 1 if the images are highly similar and 0 if they are highly dissimilar.

Square root of the Jensen-Shannon divergence: a proper metric A better measure to calculate similarity between images is the information theoretic Kullback-Liebler (KL) divergence (Greenspan et al., 2001). This is a non-symmetric measure of the difference between two probability distributions. It has been shown to perform better than HI in image search and retrieval tasks (Goldberger et al., 2006). Though the intuition is to use the KL divergence directly as a distance measure, it is not a true metric. A symmetrical version of the KL divergence is the Jensen-Shannon (JS) divergence, the square root of which is a metric. Using a proper metric is important since we intend to study the dimensionality of the space of these image representations. Our color histogram results here are based on the distance measure that is the square

root of the JS divergence. Figure 1 shows a query image and the retrieved images that are similar to the query image based on the JS distance. The distance is printed on top of each of the retrieved images.

The Kullback-Liebler divergence of Q from P is defined as

$$D_{KL}(P||Q) = \sum_i log \frac{P(i)}{Q(i)} \tag{3}$$

where P and Q are probability distributions of a discrete random variable. The symmetric Jensen-Shannon divergence is given by

$$D_{JS} = \frac{1}{2} D_{KL}(P||M) + \frac{1}{2} D_{KL}(Q||M) \tag{4}$$

where $M = \frac{1}{2}(P+Q)$

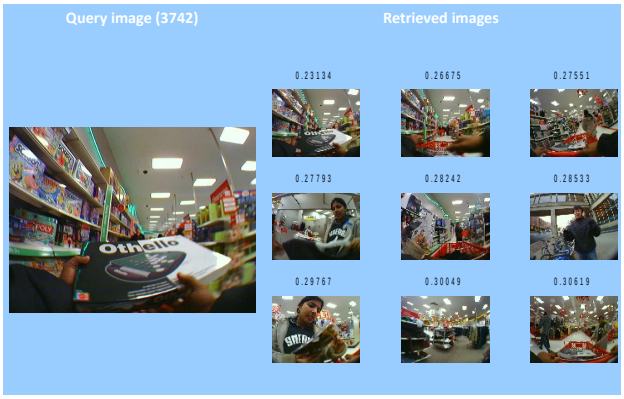

Figure 1: Query image and retrieved images (JS divergence distance method).

Color Correlogram Representation

The color histogram has the drawback of being a purely global description of the color content in an image. It does not include any spatial information. Purely local properties when used can be extremely sensitive to appearance changes due to slight changes in angle, zoom, etc. Purely global properties (like those used in the color histograms) can give false positives as it can classify images from widely separated scenes as belonging to the same scene if they have similar color content. An example of this can be found in figure 1. The third image in the second row of the retrieved images is a false positive because that image belongs to an entirely different event.

A color correlogram (Huang, Kumar, Mitra, Zhu, & Zabih, 1997) describes global distributions of local spatial color correlations. In other words, a correlogram of an image is a three dimensional matrix whose k-th entry for $\langle i, j \rangle$ is the probability of finding a pixel of color j at a distance k from a pixel of color i. This makes the correlogram robust to changes in appearance caused by occlusions, zoom, viewing angles, etc. We use a special case of the correlogram for ease of computation: the banded correlogram (Huang, 1998). Figure 2 shows the same query image as earlier and the retrieved images that are based on the relative L_1 distances between images represented as banded correlograms. The distance is printed on top of each of the retrieved images. There are no false positives in these retrieved images.

Let I be an $n \times m$ image. The colors in I are quantized into k colors $c_1, c_2, \cdots c_k$. For a pixel $p = (x, y) \in I$, let $I(p)$ denote its color. $I_c \triangleq \{p | I(p) = c\}$ where $c \in \{c_1, c_2, \cdots c_k\}$. For pixels $p_1 = (x_1, y_1), p_2 = (x_2, y_2)$, we define L_∞ norm to measure the distance between them, such that $|p_1 - p_2| \triangleq max\{|x_1 - x_2|, |y_1 - y_2|\}$.

The correlogram of I is defined for $i, j \in \{1, 2, 3, \cdots, k\}, d \in \{1, 2, 3, \cdots, l\}$ where distance d is fixed a priori, such that

$$\gamma^{(d)}_{c_i,c_j}(I) \triangleq \Pr_{p_1 \in I_{c_i}, p_2 \in I}[p_2 \in I_{c_j} \, | \, |p_2 - p_1| = d] \triangleq \frac{|I_{c_j} \cap I^d_{c_i}|}{|I^d_{c_i}|} \tag{5}$$

where $I^d_c \triangleq \{p_2 | p_1 \in I_c \wedge |p_2 - p_1| = d\}$, where $d \in \{1, 2, 3, \cdots, l\}$ is a distance between two given pixels in the image. Given any pixel of color c_i in the image, $\gamma^{(d)}_{c_i,c_j}(I)$ gives the probability that a pixel at distance d away from the given pixel is of color c_j. Hence, the color correlogram is a three-dimensional table indexed by color and distance between pixels and the size of the correlogram is $O(k^2 l)$.

The banded correlogram (Huang, 1998) is for storage trimming. Given b, for $1 \leqslant d \leqslant l/b$,

$$\bar{\gamma}^{(d)}_{c_i,c_j}(I) \triangleq \sum_{d'=(d-1)b+1}^{db} \gamma^{(d')}_{c_i,c_j}(I) \tag{6}$$

For each color pair (c_i, c_j), the probability values for the distances in the selected distance set whose cardinality is b are summed as a single number. Hence, a banded color correlogram is a restricted version of the color correlogram.

Distance Measure We use a relatively weighted L_1 distance measure for computing the distance between images I and I' as follows:

$$|I - I'|_{\gamma, L_1} \triangleq \sum_{i,j,d} \frac{|\gamma^{(d)}_{c_i,c_j}(I) - \gamma^{(d)}_{c_i,c_j}(I')|}{1 + \gamma^{(d)}_{c_i,c_j}(I) + \gamma^{(d)}_{c_i,c_j}(I')} \tag{7}$$

where $i, j \in \{1, 2, 3, \cdots, k\}$, and $d \in \{1, 2, 3, \cdots, l\}$.

The L_1 distance is also known as the manhattan distance. The manhattan distance between two points in an n-dimensional vector space with a fixed cartesian coordinate system is just the sum of the lengths of the projections of the line segment between the two points onto the coordinate axes. The normalization is such that non-uniform weights are assigned to the contribution of different colors to the dissimilarity between the two images. This is in keeping with the intuition that a difference in the number of pixels in any given

color bucket has a more significant contribution to the perceived dissimilarity if the content of that color in the image is low to start with. The same difference but when the color content is extremely high shouldn't contribute too much to the perceived dissimilarity between two images.

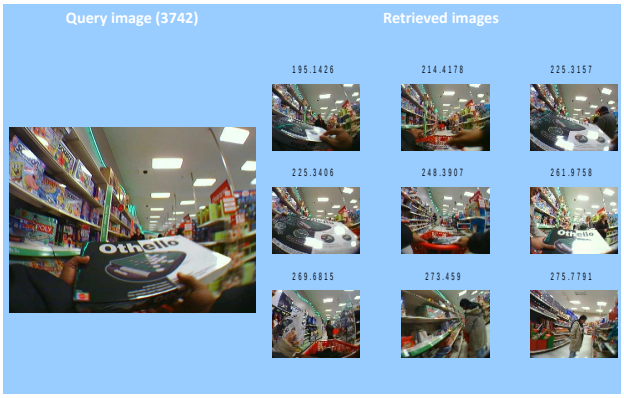

Figure 2: Query image and retrieved images (the color correlogram method).

Comparison: Common Neighbor Ratio

We now need to compare the performance of the two methods for our current purpose: to check if the distance measure on the respective representation does a good job of identifying as neighbors images that really are closely spaced in the time sequence. Events within a context are closely spaced in time and one of the major tasks for our method is to be able to accurately retrieve images that are from the same context. The idea here is that most of the closely spaced images as characterized by the distance measure ought to be closely spaced in time. Periodic events are exceptions where people might return to the same place after a certain duration. The images from those two episodes will be closely spaced but might be far apart in time. With this in mind, we define the common neighbor ratio. Given a positive integer k, for each image I, we find its k nearest neighbors both in the distance domain and in the time domain. Suppose $D_I = \{I_{d1}, I_{d2}, \cdots, I_{dk}\}$ are image I's k nearest neighbors in the distance space and $T_I = \{I_{t1}, I_{t2}, \cdots, I_{tk}\}$ are image I's k nearest neighbors in the time space (the images come with timestamps on them which are used in this calculation), then

$$common\ neighbor\ ratio = \frac{\sum_{I=1}^{n} |D_I \cap T_I|}{n \times k} \qquad (8)$$

where n is the total number of images. If k equals to n, then the ratio is 1. The method that has a higher common neighbor ratio is the better one. Figure 3 shows clearly that the correlogram representation and its corresponding distance measure outperforms the traditional histogram representation

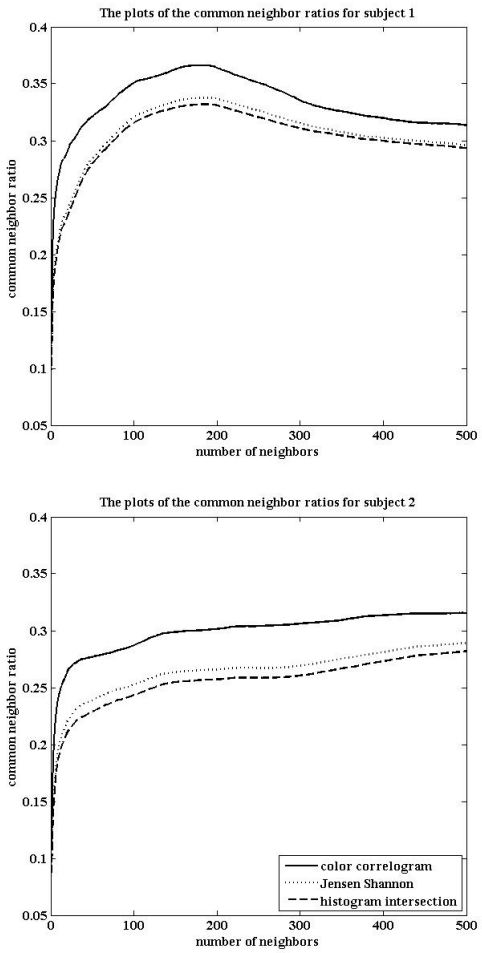

Figure 3: The common neighbor ratio as a function of number of nearest neighbors for image data from two subjects. The correlogram-relative L_1 distance method gives a higher ratio than the histogram-JS distance and the histogram intersection distance methods.

and the associated JS divergence distance measure. It can also be seen that the JS divergence measure works slightly better than the histogram intersection distance.

Correlation Dimension Analysis

Dimension measures are used to quantify the space filling properties of a set. A fractal dimension is a more informative measurement than a topological dimension which can take only integer values. For example, the topological dimension of a point is 0, of a line is 1 and of a surface is 2. A wiggly line is more space filling than a straight line but has a topological dimension of 1. The wiggly line is said to be a fractal if its fractal dimension is greater than its topological dimension (Mandelbrot, 1967). Fractal dimension measurements have been widely used in nonlinear dynamics time series analysis.

If a time series is from a nonlinear dynamical system or from a random process, the time series is irregular in both

time and frequency domains. Methods of time series analysis based on phase space reconstructions can reveal structure in time series from nonlinear dynamical systems as opposed to little structure in time series from random processes. Many popular methods of analysis involve correlation dimension estimates. There are several dimension measurements that are possible (Camastra, 2003). The correlation dimension is one of the simplest to calculate and is the most widely used dimension measurement in time series analysis. The correlation dimension is also related to the minimum number of variables needed to model the system's behavior in phase space.

The correlation dimension is a measure of the dimensionality of the space occupied by a set of points and is a type of fractal dimension because it allows non-integer values. Grassberger and Procaccia (1983a, 1983b) introduced the correlation dimension to characterize phase space filling properties of attractors. The set is covered by spheres of a given size r and the correlation dimension v is defined by:

$$v = \lim_{r \to 0} \sum_{i} \frac{\log\left(\sum_i p_i(r)^2\right)}{\log r} \tag{9}$$

where $\sum_i p_i(r)^2$ is the probability of finding a pair of points in a sphere of size r. For small values of r, this probability is the same as the probability of finding a pair of points separated by less than r. This probability, for large data sets, is given by the correlation sum. For N points in an M-dimensional space, the correlation sum is given by

$$C(r) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1 j \neq i}^{N} H\left(r - \left|\vec{X}_i - \vec{X}_j\right|\right) \tag{10}$$

H is the heaviside function. Here, it counts the number of pairs of points which are separated by less than r. For sufficiently small r and large number of points N,

$$C(r) \propto r^v \tag{11}$$

Taking logarithms of each side, we get:

$$v \sim \frac{\log\left(C\left(r\right)\right)}{\log\left(r\right)} \tag{12}$$

v is calculated from the slope of the straight line scaling region of a $log(C(r)$ versus $log(r)$ plot.

Results

The color histogram method was used to represent the images and the square root of the Jensen-Shannon divergence was used to calculate the similarity between pairs of images. $log(C(r))$ was then recorded in a series of 1000 bins. The correlation dimension(s) v is the slope $\frac{d\log(C(r))}{d\log(r)}$ of the linear portion(s) of the $log(C(r))$ versus $log(r)$ plot. The same procedure was repeated for the color correlogram representations using the relative L_1 distances to calculate similarity between images. Figure 4 shows the correlation dimension plots for image data taken from 2 subjects. The left panel contains the results for the correlation dimension using the color histogram representation and the associated square root of the JS Divergence. The right panel contains the results using the color correlogram representation and the associated relative L_1 distance measure. Points close to zero have been discarded in the correlogram correlation dimension plots due to insufficient pairs of points in that region.

There are hints of a two scale structure in the histogram based correlation dimension plots but the correlogram based correlation dimension plots do not show this structure. More discussion follows in the next section.

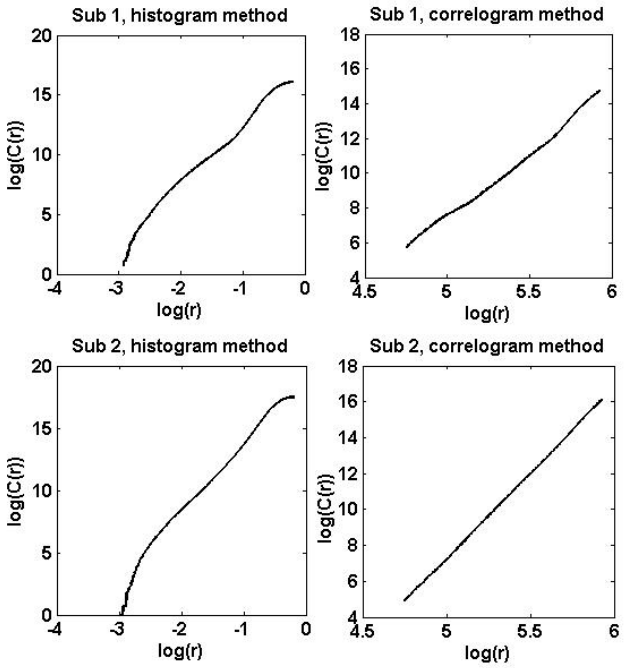

Figure 4: The correlation dimension plots for 2 subjects: The left panel is with the color histogram-JS div distance method and the right panel is with the correlogram-L1 distance method.

Conclusion and Discussion

Images were captured by subjects using a Microsoft^{TM}Research SenseCam. A correlation dimension analysis was done on images that were obtained from each subject. These images can be considered as representative of the visual input that goes into an individual's episodic memory. Distances between pairs of images represented by color histograms were calculated using the square root of the Jensen-Shannon divergence. Color histograms do not include spatial information. HSV autocorrelograms have been found to work better in image retrieval studies (Ojala, Rautiainen, Matinmikko, & Aittola, 2001). Spatial information in the images may be relevant here. For example, how do people recognize that two very different images in terms of color

content belong to the same episode? The distances calculated from the HSV histogram have given us sufficiently accurate nearest neighbour pairs as demonstrated in Figure 1 but the correlogram method and the associated L_1 distance measure was found to work better for our current purposes based on our definition of the common neighbor ratio. We conclude that the better method is the one that correctly identifies images that are close in time (within context) by classifying them as close in space based on the distance measure employed by the respective method.

A two scale structure was found in earlier studies on corpora of different languages (Doxas et al., 2010). The trajectory through a semantic space as one transitions from paragraph to paragraph in written discourse was shown to display a low dimensionality at short distances and higher at larger distances. This structure was observed in five corpora of written text in English, French, Modern Greek, Homeric Greek, and German respectively. The lower scale dimension of eight was observed to be approximately the same across languages. These structures suggest that there are strong constraints on the topology of the space through which authors move as they write and through which readers move as they read. The question now is if this is a constraint imposed by the cognitive system. This study is aimed at addressing this question. The images used represent the visual input that goes into a person's episodic experience, i.e., of the everyday events that one encounters (visually). The correlation dimension plots however don't reliably show a two scale structure here. Further exploration is necessary, however, to determine if the image representation meets all of the assumptions of the correlation dimension analysis as it has been used in this study. One such assumption is that the space has orthonormal basis vectors.

Acknowledgments

This work was supported by the Air Force Office of Scientific Research (AFOSR) under grant number FA9550-09-1-0614. We thank Microsoft for providing us the Microsoft Research SenseCams.

Références

Berry, E., Conway, M., Moulin, C., Williams, H., Hodges, S., Williams, L., et al. (2006). Stimulating episodic memory: Initial explorations using sensecam. In *Abstracts of the psychonomic society. 47th annual meeting* (Vol. 11, p. 56-57). Oxford University Press.

Camastra, F. (2003). Data dimensionality estimation methods: A survey. *Pattern Recognition*, *36*, 635–652.

Chiu, G. S. (2002). *Bent-cable regression for assessing abruptness of change*. Doctoral dissertation, Department of Statistics and Actuarial Science, Simon Fraser University.

Doxas, I., Dennis, S., & Oliver, W. L. (2010). Dimensionality of discourse. *Proceedings of the National Academy of Sciences*, *107*.

Goldberger, J., Gordon, S., & Greenspan, H. (2006). Unsupervised image-set clustering using an information theoretic framework. *IEEE Trans Image Process*, *15*.

Grassberger, P., & Procaccia, I. (1983a). Characterization of strange attractors. *Physical Review Letters*, *50*, 346–349.

Grassberger, P., & Procaccia, I. (1983b). Measuring the strangeness of strange attractors. *Physica D*, *9*, 189–208.

Greenspan, H., Goldberger, J., & Ridel, L. (2001). A continuous probabilistic framework for image matching. *Journal of Computer Vision and Image Understanding*, *84*, 384–406.

Huang, J. (1998). *Color-spatial image indexing and applications*. Doctoral dissertation, Department of Computer Science, Cornell University.

Huang, J., Kumar, S. R., Mitra, M., Zhu, W. J., & Zabih, R. (1997). Image indexing using color correlograms. In *Proceedings of the 1997 conference on computer vision and pattern recognition.*

Jeong, S., Won, C. S., & Gray, R. M. (2004). Image retrieval using color histograms generated by gauss mixture. *Computer Vision and Image Understanding: Special Issue on Color for Image Indexing and Retrieval*, *94*, 44–66.

Mandelbrot, B. (1967). How long is the coast of britain? statistical self-similarity and fractional dimension. *Science*, *156*, 636–638.

Ojala, T., Rautiainen, M., Matinmikko, E., & Aittola, M. (2001). Semantic image retrieval with hsv correlograms. In *Proc. 12th scandinavian conference on image analysis.* Bergen, Norway.

Smith, J. R., & Chang, S. F. (1995). *Automated image retrieval using color and texture* (Rapport technique N^{o} CU/CTR 408-95-14). Columbia University.

Smith, J. R., & Chang, S. F. (1996). Tools and techniques for color image retrieval. In *Symposium on electronic imaging: Science and technology - storage retrieval for image and video databases iv* (Vol. 2670). San Jose, CA.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*, 41–78.

Stockman, G., & Shapiro, L. (2001). *Computer vision*. Prentice-Hall.

Swain, M., & Ballard, D. (1991). Color indexing. *International Journal of Computer Vision*, *7*, 11–32.