ICIAP 2013
CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

An Early Framework for Determining Artistic Influence

Kanako Abe and Babak Saleh
Ahmed Elgammal
kanaabe,babaks@rutgers.edu
elgammal@cs.rutgers.edu

Computer Science Department
Rutgers University

Abstract. Considering the huge amount of art pieces that exist, there is valuable information to be discovered. Focusing on paintings as one kind of artistic creature that is printed on a surface, artists can determine its genre and time period that paintings can belong to. In this work we are proposing the interesting problem of influence determination between painters which has not been explored well. We answer the question "Who influenced this artist?" by looking at his masterpieces and comparing them to others. We pose this interesting question as a knowledge discovery problem. We presented a novel dataset of paintings for the interdisciplinary field of computer science and art and showed interesting results for the task of influence finding.

1 Introduction

How do artists describe their paintings? They talk about their works using several different concepts. The elements of art are the basic ways in which artists talk about their works. Some of the elements of art include space, texture, form, shape, color, tone and line [7]. Each work of art can, in the most general sense, be described using these seven concepts. Another important descriptive set is the principles of art. These include movement, unity, harmony, variety, balance, contrast, proportion, and pattern. Other topics may include subject matter, brush stroke, meaning, and historical context. As seen, there are many descriptive attributes in which works of art can be talked about.

One important task of art historians is to find influences and connections between artists. By doing so, the conversation of art continues and new intuitions about art can be made. An artist might be inspired by one painting, a body of work, or even an entire genre of art.

So how is this influence determined? Which paintings influence each other? Which artists influence each other? Art historians are able to find which artists influence each other by examining the same descriptive attributes of art which were mentioned above. Similarities are noted and inferences are suggested.

It must be mentioned that determining influence is always a subjective decision. We will not know if an artist was ever truly inspired by a work unless he or she has said so. However, for the sake of finding connections and progressing through movements of art, a general consensus is agreed upon if the argument is convincing enough. Figure 1 represents a commonly cited comparison for studying influence.

Fig. 1. An example of an often cited comparison in the context of influence. Diego Velázquez's Portrait of Pope Innocent X (left) and Francis Bacon's Study After Velázquez's Portrait of Pope Innocent X (right). Similar composition, pose, and subject matter but a different view of the work.

Is influence a task that a computer can measure? In Computer Vision, there has been considerable research on the object-recognition in images [1], similarity between images, and some research on automated classification of paintings [2,3,9,8]. However, there is very little research done on measuring and determining influence between artists [9]. Measuring influence is a very difficult task because of the broad criteria for what influence between artists can mean. As mentioned earlier, there are many different ways in which paintings can be described. Some of these descriptions can be translated to a computer. Some research includes brushwork analysis [9] and color analysis to determine a painting style. For the purpose of this project, we do not focus on a specific element of art or principle of art but instead we focus on finding new comparisons by experimenting with different similarity measures.

Although the meaning of a painting is unique to each artist and is completely subjective, it can somewhat be measured by the symbols and objects in the painting. Symbols are visual words that often express something about the meaning of a work as well. For example, the works of Renaissance artists such as Giovanni Bellini and Jan Van-Eyck use religious symbols such as a cross, wings, and animals to tell stories in the bible. This shows the need for an object-based representation of images. We should be able to describe the painting from a list of many different object classes. By having an object-based representation, the image is described in a high-level semantic as opposed to low-level semantics such as color and texture. By using the Classemes [11] feature, we are able to capture both high-level and low-level semantics. For example, Figure 1 may not look like similar images, but when considering the objects placed in each

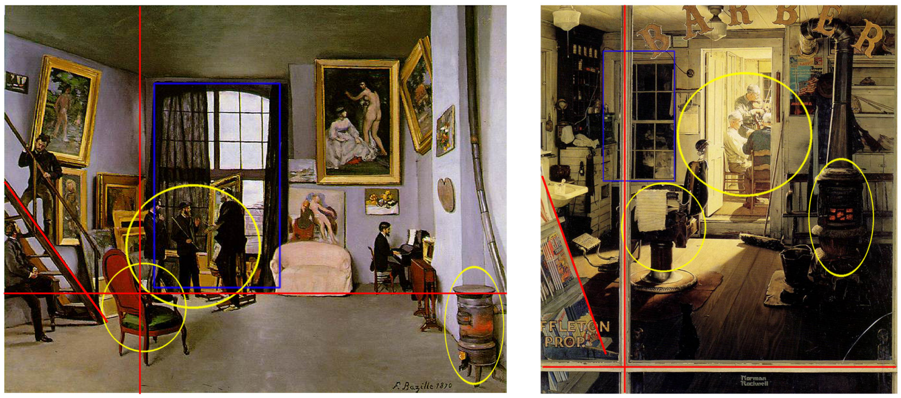

Fig. 2. Frédéric Bazille's *Studio 9 Rue de la Condamine* (left) and Norman Rockwell's *Shuffleton's Barber Shop* (right). The composition of both paintings is divided in a similar way. Yellow circles indicate similar objects, red circles indicate composition, and the blue square represents similar structural element. The objects seen – a fire stove, three men clustered, chairs, and window are seen in both paintings along with a similar position in the paintings. After browsing through many publications and websites, we conclude that this comparison has not been made by an art historian before.

of the paintings, similarity becomes clear. This comparison is a result from our experiments which we describe later.

One important factor of finding influence is therefore having a good measure of similarity. Paintings do not necessarily have to look alike but if they do or have reoccurring objects (high-level semantics), then they will be considered similar. If influence is found by looking at similar characteristics of paintings, the importance of finding a good similarity measure becomes prominent. Time is also a necessary factor in determining influence. An artist cannot influence another artist in the past. Therefore the linearity of paintings cuts down the possibilities of influence.

By including a computer's intuition about which artists and paintings may have similarities, it not only finds new knowledge about which paintings are connected in a mathematical criteria but also keeps the conversation going for artists. It challenges people to consider possible connections in the timeline of art history that may have never been seen before. We are not asserting truths but instead suggesting a possible path towards a difficult task of measuring influence. The main contribution of this paper is proposing the interesting task of determining influence between artist as a knowledge discovery problem and proposing a new dataset. Best of our knowledge Carneiro et al[4] recently published the "PRINTART" on paintings along with primarily experiments on image retrieval and genre classification. However this dataset contains only monochromatic artistic images. Our dataset have chromatic images and its size is about double the "PRINTART" dataset.

2 Influence Inference Framework

Consider a set of artists, X. For each artist, X_i, we have a ground truth time period t_i that artist X_i has completed works for. Also consider a set of images I, for each artist X_i. We extract Classeme features [11] as visual features for each image and show it by a vector called $C = (c_1, ..., c_N)$ where N represents dimension of the feature space.

We represent the problem of influence as similarity following time. For the statement $X_i \Rightarrow X_j$ to be true where the arrow indicates the left side influencing the right side, two requirements must be met. In the time constraint, artist t_i must be either come before or overlap t_j since t is represented as a time period. Thus we not only allow artists to be influenced by their past but also by people from an overlapping time period since this is generally true for influence within a genre.

For the second requirement, two artists in the the feature space should be similar $X_i \sim X_j$. For determining similarity, averaging of each artist X_i's image set I_i will not work. Doing so results in a loss of information. If an artist X_i has significantly less images in I_i than an artist X_j does in X_j, then X_j will have a larger variation of images. This may result in skewed information about the similarity between artists since it reflects each I's number of images. Therefore, a method that measures distance between sets is important. This way, artists can be represented by their entire work yet still keep information about each individual painting. The goal is to avoid the risk of losing information about a painting while providing meaningful set distances.

We consider potential influence if it reflects ground-truth or artists are of similar genre. Our ground-truth is a collection of known influences and general consensus of influences.

Once a good similarity measurement is found, we can map artists into a space and here, knowledge discovery becomes prominent. Which artists are similar but have not been talked about before? How will different distance metrics lead to different conclusions about artists? This portion of the study is also important for contribution to the art world.

3 Dataset

Our dataset contains a total of 1710 works by 66 artists chosen from Mark Harden's Artchive database of fine-art. Each image is annotated with the artist's first name, last name, title of work, year made, and genre. The majority of the images are of the full work while a few are details of the work. We are primarily dealing with paintings but we have included very few images of sculptures as well. The artist with the most images is Paul Cézanne with 140 images and the artist with the least number of works is Hans Hoffmann with 1 image.

The artists themselves ranged 13 different genres throughout art history. These include Expressionism (10 artists), Impressionism (10), Renaissance (12), Romanticism (5), Cubism (4), Baroque (5), Pop (4), Abstract Contemporary (7), Surrealism (2), American Modernism (2), Post-Impressionism (3), Symbolism (1), and Neoclassical (1). The number in the parenthesis refers to the number of artists in each genre. Some genres were condensed such as *Abstract Contemporary* which includes works in the *Abstract Expressionism, Contemporary,* and *De Stijl* periods. The *Renaissance* period has the most images (336 images) while *American Modernism* has the least (23 images). The average number of images per genre is 132. The earliest work is a piece by Donatello in 1412 while the

Fig. 3. Examples of paintings from thirteen genres: Renaissance, Baroque, Neoclassical, Romanticism, Impressionism, Post-Impressionism, Expressionism, Cubism, Surrealism, Symbolism, American Modernism, Pop, and Abstract Contemporary.

most recent work is a self portrait by Gerhard Richter done in 1996. The earliest genre is the *Renaissance* period with artists like Titian and Michelangelo during the 14th to 17th century. As for the most recent genre, art movements tend to overlap more in recent years. Richter's painting from 1996 is in the *Abstract Contemporary* genre.

4 Experiments

4.1 Visual features

We extracted the Classeme feature vector [11] as the visual feature for our experiments. Classeme features are output of a set of classifiers corresponding to a set of C category labels, which are drawn from an appropriate term list (defined in [11] and not related to our textual features). For each category $c \in \{1 \cdots C\}$, a set of training images is gathered by issuing a query on the category label to an image search engine.
After a set of coarse feature descriptors (Pyramid HOG, GIST) is extracted, a subset of feature dimensions was selected [11]. Considering this reduced dimension feature a one-versus-all classifier ϕ_c is trained for each category. The classifier output is real-valued, and is such that $\phi_c(x) > \phi_c(y)$ implies that x is

more similar to class c than y is. Given an image x, the feature vector (descriptor) used to represent it is the classeme vector $[\phi_1(x), \cdots, \phi_C(x)]$. The Classeme feature is of dimensionality $N = 2569$.

4.2 Similar artists

To judge about similarity between artists we computed the Euclidean distance between Classeme features corresponding to each image in the dataset. The results showed some interesting cases, several of which have not been studied by art historians as a potential comparison before. Figure 1 is an example of this, as well as Figure 4 and Figure 5.

Fig. 4. Vincent van Gogh's *Old Vineyard with Peasant Woman* 1890 (left) and Joan Miro's *The Farm* 1922 (Right). Similar objects and scenery but different moods and style.

We researched known influences between artists within our dataset from multiple resources such as The Art Story Foundation and The Metropolitan Museum of Art. For example, there is a general consensus among art historians that Paul Cézanne's use of fragmented spaces had a large impact on Pablo Picasso's work.

We say there is a good artist-to-artist similarity if 1) it reflects the ground-truth artist influence list or 2) they are of similar genres. We include the genre as another indication of influence because the works of artists in the same genre tend to be influenced by the same people and also by each other. After computing distances, an affinity matrix of similar artists is made. To measure accuracy, we focus on the top 5 artists similar to each artist X and considered how many of them are correct based on annotation.

We tried several different methods for measuring the distance between two sets of artists. First, we used the Hausdorff distance to measure the distance between sets of artists. We computed the distance between each artist set in a Euclidean metric space. Our result had an accuracy of 22.73%.

In another variation, we modified the Hausdorff distance to consider a subset of distances. Our modification only considers 90 percent of the paintings for every artist set. This would presumably eliminate 10 percent of the least similar images. The results were not very different from the previous case. Those artists

Fig. 5. Georges Braque's *Man with a Violin* 1912 (Left) and Pablo Picasso's *Spanish Still Life: Sun and Shadow* 1912 (Right).

which were affected had slightly better results than regular Hausdorff distance. This results held an accuracy of 23.03%

Next we tried modified Hausdorff distance (MHD) proposed by Dubuisson et al [5]. This adjusted version of Hausdorff distance is shown to work better for object matching. This result had an accuracy of 30%, which was our best result.

Previously, we had tried using only symmetric measurements. If artists are influenced by each other (meaning they have overlapping t time periods), it may be important to describe which artist influences the other more. Further experiments were done on asymmetric and symmetric affinity matrices and those results are seen in Figure 8.

We also tried to see if we can reduce the feature dimensionality before computing the similarity or not. We applied various methods including PCA, MDS, LLE, Isomap, etc but got a worse results emphasizing classemes feature are optimized in terms of dimension.

4.3 Locally Linear Embedding and Mapping

In order to visualize artists in a new space based on similarity, we used a non-linear approach, namely Locally Linear Embedding (LLE) [10]. With LLE, we map our high-dimensional data to a 2D and 3D space. The embedding provides relationships between artists in relation to the artist space as a whole.

First, we applied LLE to the affinity matrix computed previously down to one dimension. In this reduction, Modern and Abstract artists seemed to be grouping together while other artists were too clustered to determine groupings. In a different mapping, we use LLE to reduce the affinity matrix to two dimensions. Figure 6 visualizes this two dimension varying mapping in both the x and y axis. We color each artist name according to its genre (13 colors for 13 genres) to get a better sense of groupings.

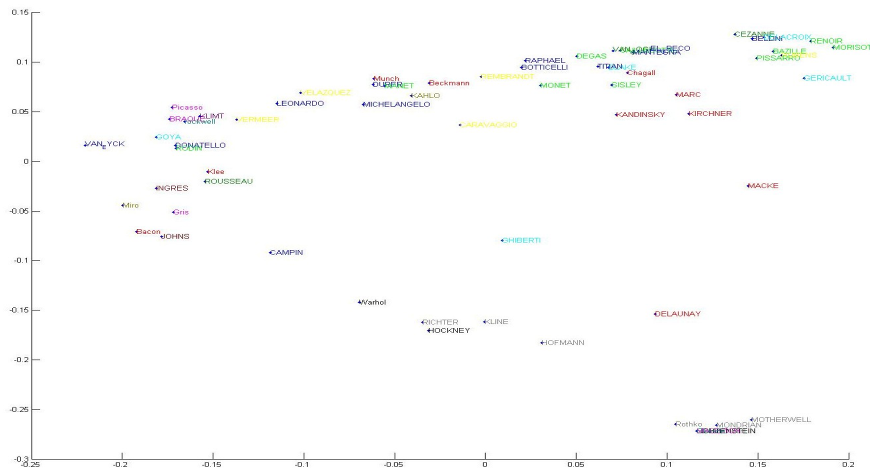

Fig. 6. Similar artists in two dimensions

In Figure 6 we can see a few *Expressionist* artists (in red) clustered together as well as *Abstract Contemporary* artists (in grey-blue). As seen, the artists at the bottom of the mapping are Lichtenstein, Hepworth, Malevich, Mondrian, Motherwell, O'Keffe, and Rothko who are all Modern and Abstract artists. Their genres differ slightly but all share some stylistic approaches and time period. . However, this time we can see that the *Impressionists* and *Renaissance* artists seem to have similar values in one dimension but not the other. Other genres such as *Romanticism* seem to have a broader range of values.

Some artists in this mapping seem to cluster according to their genre, but in the context of influence, it is also important to think about the similarities between artists instead of the classification of genre. This is especially true as style varies in certain genres. This is yet another complication of the task of measuring influence.

Therefore, another way to analyze this graph is to disregard genre all together. We can wonder whether Richter and Hockney share a connection because they lie close to each other. Or we can wonder if Klimt was influenced by Picasso or Braque. In fact, both Picasso and Braque were listed as influences of Klimt in our ground-truth list. When comparing these close mappings to the ground truth influence, some are reasonable while others seem less coherent. In another example, Bazille lies close to Renoir and Delacroix who were both influences of Bazille. Renoir was also influenced by Delacroix according to our research and it is reflected in our mapping. Other successful mappings include Munch's influence on Beckmann, Pissarro's influence on Cezanne, Degas's influence on Caillebotte, Velazquez's influence on Manet, and so on. Caillebotte and Van Gogh are nearly mapped on top of each other. Although it is not reflected in our sparse ground-truth list, it is known that both Caillebotte and Van Gogh were influenced by an outside source, Japanese art and composition [6]

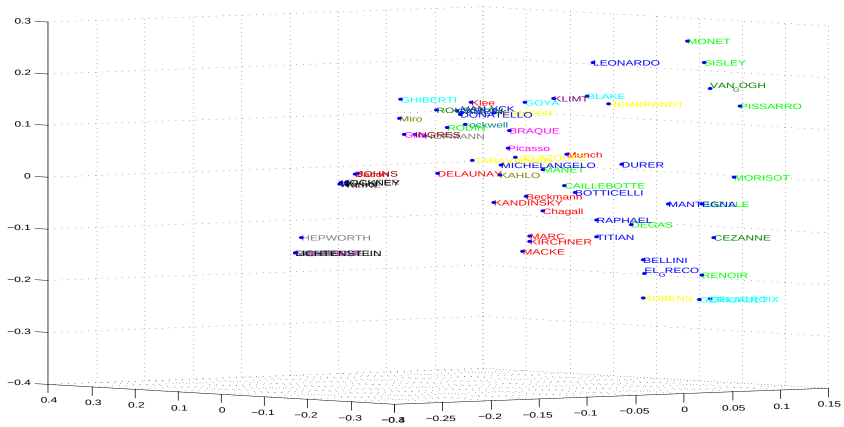

Fig. 7. Embedded mapping of artists in three dimensions using an asymmetric Hausdorff distance

We try another mapping with a 3-dimensional reduced space. Figure 7 shows a 3-dimensional view of artists mapped using an asymmetric affinity matrix. Here, more of the genres seem to be clustering together. There is a clear band of Impressionists (green), Renaissance (blue), and Expressionist (red) artists. Other interesting pairs include Ruben's influence on Delacroix, Pissarro and Monet's influence on Van Gogh, Manet's influence on Munch, and the congestion of Modern and Abstract artists. This asymmetric 3 dimensional LLE mapping shows a better categorization of genres then any other mapping we tried.

Other reduction methods besides LLE were tried as seen in the table in Figure 9. Those methods are PCA, MDS, Linear Discriminant Analysis (LDA), and Isomap. The table shows that LLE dimension reduction is the best method our of five for accuracy in the context of influence.

Variation of Hausdorff distance Accuracy		
	Symmetric	Asymmetric
Hausdorff	22.73%	20.00%
Percentile Hausdorff	23.03%	20.30%
Modified Hausdorff	30.00%	26.36%

Fig. 8. Accuracy rates of asymmetric vs. symmetric affinity matrices for similarity measuring.

Dimension Reduction Accuracy	
LLE	28.18%
PCA	21.52%
MDS	21.52%
LDA	18.79%
Isomap	12.54%
LLE (Symmetric Affinity Matrix)	22.42%

Fig. 9. Accuracy rates of five dimension reduction methods.

Through our experiments we have found that Modified Hausdorff Distance (MHD) works well on image similarity. We also found that combining this with Locally Linear Embedding is the best technique for mapping a new space. Mapping results in lower accuracy than without the mapping but the trade-off is that we are able to visualize artists in relation to each other. MHD worked best for symmetric affinity matrices at 30% accuracy. However, applying LLE to an asymmetric matrix shows better accuracy than a symmetric one and it can be explained by the fact that LLE is graph based dimension reduction and asymmetric MHD will make more accurate directed graph to apply LLE on it.

5 Conclusion and Future Works

In this paper we presented a new dataset with diverse set of artists and wide range of paintings. This dataset will be publicly available and can be used for interdisciplinary tasks of Art and Computer Science. We also posed the interesting question of finding influence between painters as a knowledge discovery problem and showed interesting results for both of the qualitative and quantitative measurements.

References

1. Ravneet Singh Arora and Ahmed M. Elgammal. Towards automated classification of fine-art painting style: A comparative study. In *ICPR*, 2012.
2. Ricardo S. Cabral, Joo P. Costeira, Fernando De la Torre, Alexandre Bernardino, and Gustavo Carneiro. Time and order estimation of paintings based on visual features and expert priors. In *SPIE Electronic Imaging, Computer Vision and Image Analysis of Art II*, 2011.
3. Gustavo Carneiro. Graph-based methods for the automatic annotation and retrieval of art prints. In *ICMR*, 2011.
4. Gustavo Carneiro, Nuno Pinho da Silva, Alessio Del Bue, and João Paulo Costeira. Artistic image classification: An analysis on the printart database. In *ECCV*, 2012.
5. M-P Dubuisson and Anil K Jain. A modified hausdorff distance for object matching. In *Pattern Recognition*, 1994.
6. Derek Fell. *Van Gogh's Garden.* Simon Schuster, 2001.
7. Lois Fichner-Rathus. *Foundations of Art and Design.* Clark Baxter.
8. Friedenberg J. Rockmore D. Graham, D. Mapping the similarity space of paintings: image statistics and visual perception. *Visual Cognition*, 2010.
9. Jia Li, Lei Yao, Ella Hendriks, and James Z. Wang. Rhythmic brushstrokes distinguish van gogh from his contemporaries: Findings via automated brushstroke extraction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012.
10. Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000.
11. Lorenzo Torresani, Martin Szummer, and Andrew Fitzgibbon. Efficient object category recognition using classemes. In *ECCV*, 2010.