An application of Gaussian regression and SURF method for Mexican history of painting

Computational Statistics MA-6973

Liam Carpenter (lc3919), Mariana Martinez (mgm9745)

Courant Institute, New York University

December 7, 2021

Sections

Mexican Art History

We worked with paintings from the following artists:

- Santiago Rebull (1829-1902)
- José María Velasco (1840-1912)
- Alfredo Ramos Martínez (1871-1946)
- Gerardo Murillo, Dr Atl (1875-1964)
- José Clemente Orozco (1883-1949)
- Diego Rivera (1886-1957)
- Rufino Tamayo (1899-1991)
- Frida Kahlo (1907-1954)
- Remedios Varo (1908-1963)
- Javier Marin (1962)

Examples

Objective of this project

Investigate the possibility of using methods such as Gaussian regression to help solve the problem of lack of dates for paintings.

SURF Method

Two Main Components

- Interest Point Detector (Approximate Hessian)
- Interest Point Descriptor (Haar Wavelets)

SURF achieves faster performance by using integral images.

1.

31	2	4	33	5	36
12	26	9	10	29	25
13	17	21	22	20	18
24	23	15	16	14	19
30	8	28	27	11	7
1	35	34	3	32	6

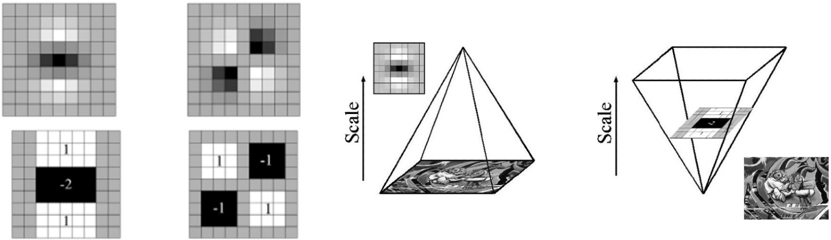
2.

31	33	37	70	75	111
43	71	84	127	161	222
56	101	135	200	254	333
80	148	197	278	346	444
110	186	263	371	450	555
111	222	333	444	555	666

$$\boxed{15 + 16 + 14 + 28 + 27 + 11} =$$
$$101 + 450 - 254 - 186 = 111$$

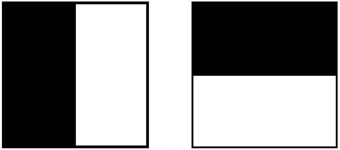
Detector

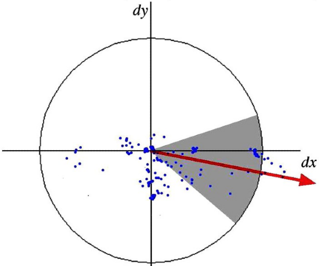
- Approximate Gaussian Derivatives With Box Filters
- Fast Computation Across Scales With Integral Images

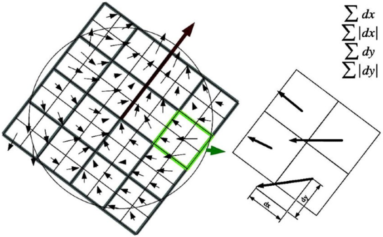


Orientation and Description

- Haar Wavelets for Orientation Assignment
- Haar Wavelets for Descriptor Components



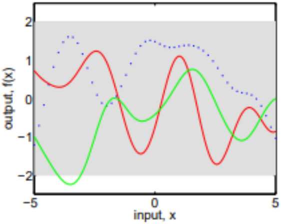




Set Up

Supervised Setting with n observed data points $\mathcal{D} = \{x_i, y_i\}_{i=1}^{n}, x_i \in \mathbb{R}^k$. Test points $\mathcal{T} = \{x_j^*\}_{j=1}^{n^*}, x_j^* \in \mathbb{R}^k$. Given mean and co-variance functions we assume a multivariate Gaussian prior and compute the joint distribution from this.

$$y \sim \mathcal{N}(\bar{0}, k(X,X)+\sigma^2 I), \begin{pmatrix} y \\ y^* \end{pmatrix} \sim \mathcal{N}(\bar{0}, \begin{pmatrix} K(X,X)+\sigma^2 I & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{pmatrix})$$

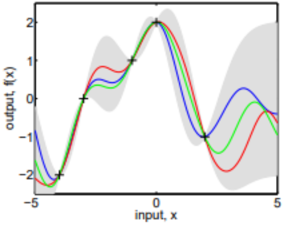


(a), prior

Predictions

The predictive distribution is obtained by marginalizing over the observed y. We use the mean of this distribution as our predictor.

$$y^* \sim \mathcal{N}(K(X^*, X)(K(X, X) + \sigma^2 I)^{-1} y, K(X^*, X^*) - K(X^*, X)(K(X, X) + \sigma^2 I)$$



(b), posterior

Data Base used

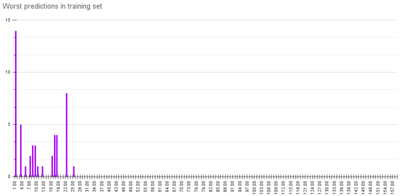
Artist	Number of paintings used	Years covered
Santiago Rebull	2	1851-1875
José María Velasco	11	1840-1898
Alfredo Ramos Martínez	16	1929-1945
Gerardo Murillo	17	1908-1962
José Clemente Orozco	16	1925-1949
Diego Rivera	30	1907-1957
Rufino Tamayo	26	1934-1991
Frida Kahlo	19	1925-1954
Remedios Varo	16	1935-1963
Javier Marín	7	1987-2002

Table 1: Paintings and years per artist in the training set

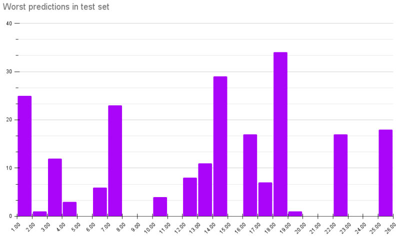
Methodology

1. Transform images into vectors for each color channel (SURF points, SURF points + PCA, SURF features, SURF features + PCA)
2. Fit different Gaussian regressions with different kernels and different optimization methods
3. Results analysis

Results and worst classifications per model



(a) Training set



(b) Test set

Example of analysis

Figure 2: Bailarinas, Diego Rivera, 1939, gouache on rice paper 38.6 × 28 cm

Other paintings done in 1939

Conclusions

- Working with more data (and more reliable) would be better
- Maybe in the future, try to fit Gaussian regressions for different art schools or use them for style evolution (for example, the development of impressionism)