Problem statement:

Be more specific!

Offer an optimal photo taking strategy for those who want to take a satisfying picture with someone in it. Given any photo with landscape or artificial element as a background, output a position where a human can stand to form the most aesthetic photo.

Map human as a X:X rectangle.

Approach:

Pre-pro: where to find new data,

How to measure scale: if something daily things appear in the photo, it can be a reference of human size

How to measure success:

Do we output a normal photo?

1. Depth and scale of human
2. Standing point of human (Human does not stand in the sky or in the water)

Crowdsource our output?

Evaluation system

1. Learn AVA
2. Learn other dataset
3. Define our own evaluation system.
4. Algorithms from other papers/ Github repos

To do:

Figure out how AVA evaluates scores. How to learn evaluation criterion from dataset? Hy

(from AVA)

(just binary, either good or bad) Linear SVMs with Stochastic Gradient Descent(SGD) on Fisher Vector(FV) signatures computed from color and SIFT(scale invariant feature transform) descriptors

Large set of training data

Large delta (take out 5-delta ~ 5+delta)

<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6126246>

Extraction of advanced features

Sophisticated learning algorithm

<http://delivery.acm.org/10.1145/1880000/1873990/p271-bhattacahrya.pdf?ip=128.237.146.245&id=1873990&acc=ACTIVE%20SERVICE&key=A792924B58C015C1%2E5A12BE0369099858%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&__acm__=1517710897_03c4c583e7503d09c39b0fc9f6447dd3>

Visual saliency technique to estimate spatial location of foreground object

supervised learning-based scene classification method proposed by Hoiem et al → identify sky and support

Morphological processing tool to mask the foreground

Relative foreground position is defined as the normalized Euclidean distance between the foreground’s center of mass, also called the visual attention center, to each of four symmetric stress points in the image frame

a four dimensional feature vector

Visual weight ratio can be described as the ratio of approximate number of pixels in the sky region, to that in the support region (ground or sea).

The ratios between the areas of these rectangles should be close to the golden ratio [15] for a better appeal

SVM to learn the score

→ solve optimization problem, and find best location

→ rescale the object

Breakthrough points? :

It is only good for 1 foreground object

assumption that the photographic frame is approximately aligned with the horizon

(only 2 quantities are used for judging composition: relative foreground position and visual weight, CAN WE DO MORE?)

Find datasets with human and landscape. WY Z

Human detection state-of-the-art (half-body or even larger)? Xueqiang Wang

https://www.cs.umd.edu/sites/default/files/scholarly\_papers/neetiPaper\_1.pdf%20

How to analyse depth and scale of a photo. YH

Depth from Focus (DfF) is a traditional way to measure relative depth from a scene *by using the information acquired through the change of the focus of a camera*  Behind this method, focus measure (FM) is the key to accuracy. Noise Robust Depth from Focus using a Ring Difference Filter proposed a new FM and a processing pipeline to get accurate relative depth from a image.

-> We can implement it and get relative depths in a image.

How to identify floor?

Do we have to consider color/brightness information when we train model from AVA dataset? Yes, but the influence is small since same pic with different person position varies little. (3.1,3.2,3.3) We still have a highest score 3.3.

Pending:

How many people in our photo? 1 2 or 3?