Photo Uncrop (ECCV'14)

Qi Shan, Brian Curless, Yasutaka Furukawa

Content

The presentation is divided into following parts:

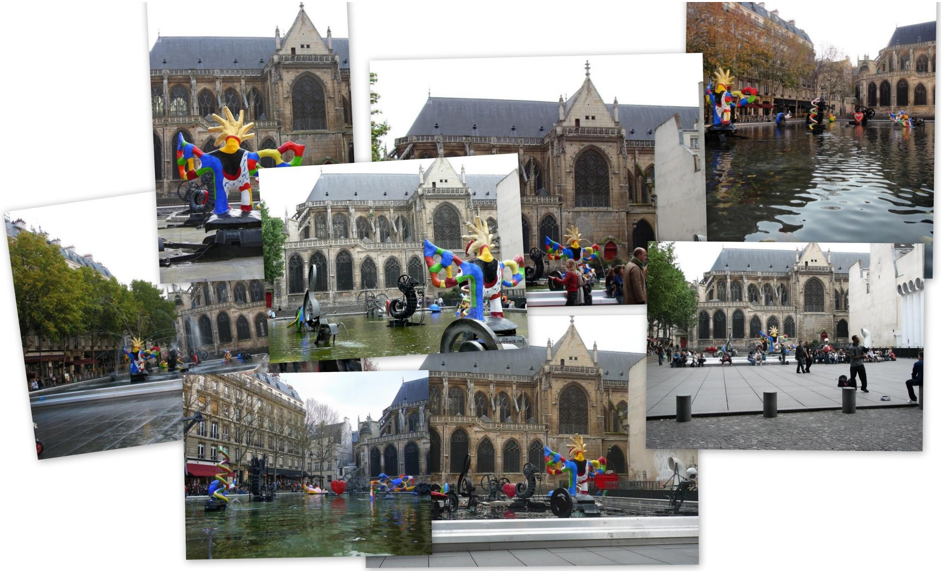
Existing Challenge

Existing methods for extending image like texture synthesize fail in a sense, that they are not able incorporate new information from *true scene*.

Abstract

This paper addresses the problem of extending the field of view of a photo - *Uncrop*. Given a reference photograph to be uncropped, the approach selects, reprojects, and composites a subset of Internet imagery taken near the reference into a larger image round of the reference using the underlying scene geometry. The proposed MRF based approach is capable of handling large photo collection with arbitrary viewpoints, dramatic appearance variation, and complicated scene layout.

+



=

Concepts Involved

- Structure from Motion (SfM)
- Markov Random Field
- Poisson Blending

Markov Random Field(MRF) Cost Formulation

$$E(l) = \sum_{p} E_{\text{unary}}(p, l(p)) + \sum_{\{p,q\} \in \mathcal{N}(p,q)} E_{\text{binary}}(p, l(p), q, l(q)) + E_{\text{label}}(l).$$

$\mathcal{N}(p, q)$ denotes pairs of neighboring pixels

l here denotes the set of all the labels in the image.

Markov Random Field(MRF) Cost Formulation

Unary Cost:

$$E_{\text{unary}}(p,l) = E_{\text{geometry}}(p,l) + \alpha_1 E_{\text{appearance}}(l) + \alpha_2 E_{\text{contrast}}(p,l)$$

$$+ \alpha_3 E_{\text{reference}}(p,l)$$

where $\alpha_1 = 10$, $\alpha_2 = 5$, $\alpha_3 = 1$ are used

Markov Random Field(MRF) Cost Formulation

$$E_{\text{geometry}}(p, l) = \max(|p - p_{\text{near}}(p, l)|, |p - p_{\text{far}}(p, l)|)$$

Determined by two factors, accuracy of the original depth value, and the baseline between the reference view and source view.

Let u denote a source pixel in image I, and U to be the corresponding 3D point on the depthmap, which is re-projected to p in the reference. We look at a local neighbourhood of 11x11 pixels, centered at u, and compute the minimum and maximum depth values in the window. We take 3D point U and shift location to the minimum depth locations, and project it to the reference image. The two projected locations thus, are Pnear and Pfar.

Markov Random Field(MRF) Cost Formulation

$$E_{\text{appearance}}(l) = k_l / N$$

It is important to encourage the use of images with similar appearance.

To do this, we assign an appearance cost to each source image. Specifically, we take the color histogram of each image, and score it by its KL divergence from the histogram of the reference image. Then the images are stored in ascending order. Let k_l be the index of image l in this sorted list.

Markov Random Field(MRF) Cost Formulation

$$E_{\text{contrast}}(p,l) = \frac{1}{|\Omega|}\sum_{v\in\Omega}\sqrt{(1-|G_x^l(v)|)^2 + (1-|G_y^l(v)|)^2}$$

Undesirable appearance variations such as shadows and over-saturation can be penalized based on the the contrast. This is addressed by defining a local contrast cost.

Let (G^l_x,G^l_y) be the finite difference gradient of image l after mapping l to grayscale (intensity values ∈ [0,1]. 11x11 window Ω centered at u in image l, which corresponds to p after the warping.

If multiple pixels from source image l map to p after warping, we again simply take the average of their scores.

Markov Random Field(MRF) Cost Formulation

$$E_{\text{reference}}(p, l) = \begin{cases} 0, & l = l_{\text{ref}} \\ 10000, & l = -1 \\ 100, & p \notin \Omega_{\text{core}} \\ \infty, & p \in \Omega_{\text{core}} \end{cases}$$

It is important to respect the reference image. Core region of the image (Ω_{core}) is defined to be the set of pixels inside the reference image and more than 11 pixels in distance from its boundary.

Here, l_{ref} is the label of the reference image. It is possible that some of the pixels in the target image are not covered by any of the images, thus the label l = -1 is assigned the highest cost.

Markov Random Field(MRF) Cost Formulation

Binary Cost:

$$E_{\text{binary}} = E_{\text{edge}} + \beta E_{\text{compatibility}}$$

We set $\beta = 10$ in all of our experiments.

Markov Random Field(MRF) Cost Formulation

$$E_S(u,l) = \left(6 - \frac{||S(u,l)||_1}{4}\right)^2$$

We first define a Sobel filter cost for a single pixel u and in (unwarped) source image I.
S(u,I) is the concatenation of the Sobel filter responses in the x and y directions for each of the r, g and b color channels, where we take the L_1 norm of this 6-dimensional vector.

Markov Random Field(MRF) Cost Formulation

$$E_{\text{edge}}(p, l, q, m) = \begin{cases} 0, & l = m \\ E_S(u, l) + E_S(u, m), & l \neq m \end{cases}$$

Now, for neighbouring target pixels p and q with labels l and m, respectively, the binary edge cost is defined as above.

If multiple pixels correspond to p after warping, we take their average over u.

Markov Random Field(MRF) Cost Formulation

$$E_{\text{compatibility}}(p, l, q, m) = 1 - \max_{n} \text{NCC}\left[\frac{1}{2}\left(W_{p,q}(l) + W_{p,q}(m)\right), W_{p,q}(n)\right]$$

*To encourage regions in the target image to resemble regions in the source image, we introduce a novel label compatibility term

Consider a pixel p and one of its neighbors q in the target image, and an image l. We define an 11 x 11 window around the two pixels and collect the pixels of $C_l(p)$ (corresponding to the warped version of image l) in the overlap into a vector $W_{p,q}(l)$. If there is a transition between labels l and m in going from p to q, respectively, then the resulting window in the final result will likely resemble to average of the windows $W_{p,q}(l)$ and $W_{p,q}(m)$. This average in turn should resemble at least one of the (warped) source images.

Since, this is a multi-label MRF problem, therefore alpha expansion is used.

Poisson Blending

Input image

Label map

MRF composite (without Poisson blending)

Final blend composite

(a) $E_{geometry} = 0$ (b) $E_{compatibility} = 0$ (c) With both terms effective

Implementation Details

- Depth Map Reconstruction: We use publicly available multi-view stereo software (Reference 8 in the paper) to reconstruct per-view depth maps, then apply cross bilateral often cause artifacts during image warping. Local window radius is 50, and the regularization parameter is 0.16. Note: We use corresponding color image as the reference for the bilateral filtering. Finally, we compute per-pixel based depths.

- Image Selection & Warping: Given a reference photograph and the SfM reconstruction, we first move each source image with an optical center that is more than a distance τ_{cop} from the reference. We set $\tau_{cop} = 50$

- Next we forward-wrap the remaining source images into the target image using splatting and a soft Z-buffer algorithm

OUR PROGRESS(Mid Evaluations)

- Reading & Understanding Research Paper
- Data Collection
- Extracted SfM reconstruction and relative optical centres.
- Implemented Max-Flow for graph-cut.

OUR PROGRESS(Final Evaluations)

- Depth Map Reconstruction
- Graph Construction.
- Poisson Blending.

- **Results are available in the project report.**

Manan Sharma (20161013)
Aakash KT (20161202)
Mohd Sharique (20161081)