CV Project Mid Evaluation

FrameBreak: Dramatic Image Extrapolation by Guided Shift-Maps

-Yinda Zhang, Jianxiong Xiao, James Hays, Ping Tan (CVPR 2013)

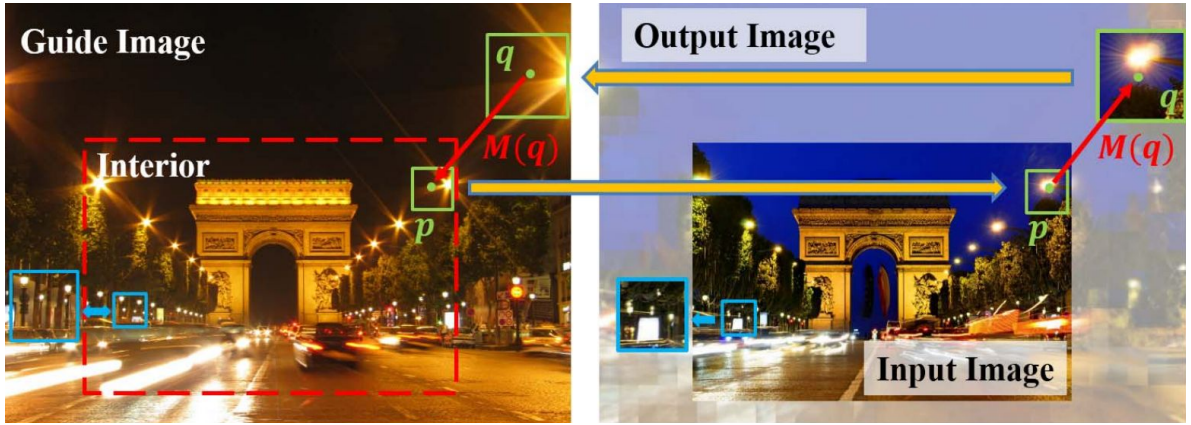
Team: Honest Liars

Parv Parkhiya (201430100)
Yogesh Sharma (201402083)
Akanksha Baranwal (201430015)

Problem Statement

Significantly extrapolate the field of view of a photograph by learning from a roughly aligned, wide-angle guide image of the same scene category keeping the layout similar as the guide image and texture similar to origin image.

Baseline Method



Assumption: Guide image I_g with desirable FOV is known.
Approach: For every exterior pixel **q** in I_g, transformation $M_{(q)}$ and interior pixel p are computed which takes patch around q to p in I_g^i with reasonable visual closeness.
Inverse of all computed transformations are applied to the respective pixels p locations in I_i to fully reconstruct I.

I_g : Guide image

I_i : Input image

I : Output image

I_g^i : The interior region of I_g where I_i is registered

M(q): the transformation to take pixel q (exterior) to pixel p (interior). $p = M(q) \circ q$

Limitation of Baseline Method

Guide Image

Input Image aligned with guide image

Baseline Method

Proposed Method

Generalised shift-map

$$E(M) = \sum_{q} E_d(M(q)) + \sum_{(p,q)\in N} E_s(M(p), M(q))$$

A graph optimisation is formulated to choose an optimal transformation (translation, scale, rotation, reflection) at each pixel of I from I_i.

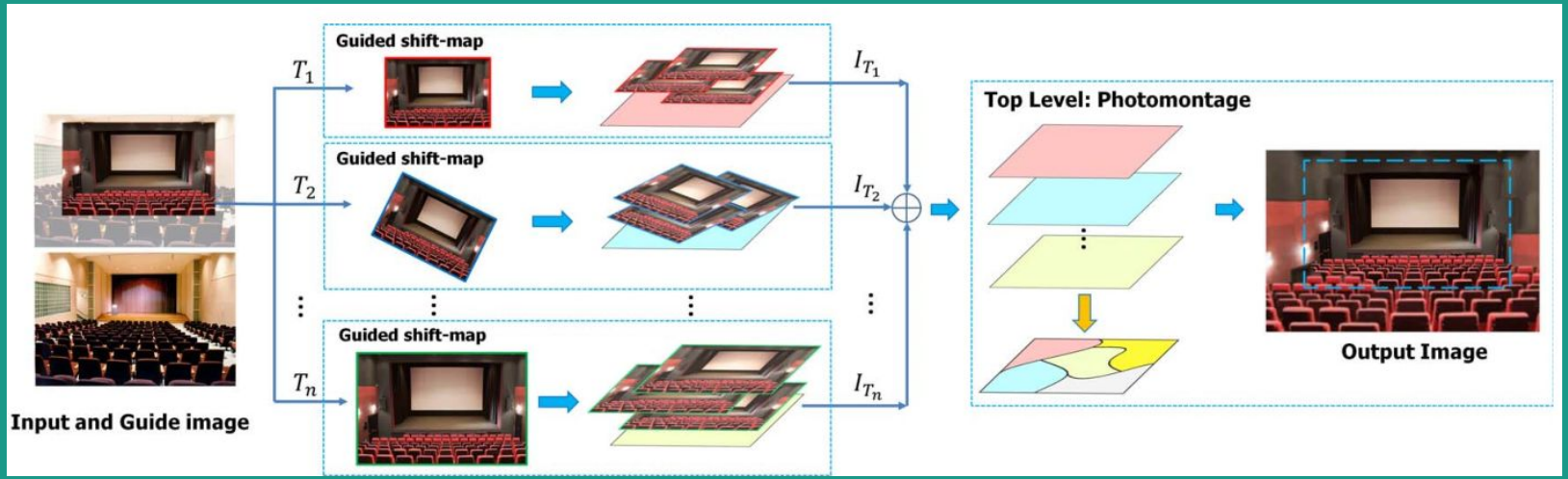
$$E_d(M(q)) = \|R(q, I_g) - R(q \circ M(q), I_g)\|_2$$

Ed is the data term to measure consistency of patch centred at q and q o M(q) in Ig
When Ed is small, q of Ig can be synthesised by copying the pixel at q o M(q).

$$\begin{aligned} E_s(M(p), M(q)) \quad &= \|I(q \circ M(q)) - I(q \circ M(p))\|_2 \\ &+ \|I(p \circ M(q)) - I(p \circ M(p))\|_2. \end{aligned}$$

Es is the smoothness term to measure compatibility of two neighboring pixels in the result image.
Smoothness cost penalises incoherent seams in the result image

Hierarchical Optimization



Step1: Fix the rotation, scaling and reflection parameters (T) and solve for optimal translation map.
Step2: Merge the results by solving for a optimal transformation (T) map to obtain the final output.

Step 1a: Guided shift-map at bottom level

- Transformation (T) (Rotation, Scale, Reflection) are uniformly sampled 11 rotation angles from the interval of [−45° ,45°], and 11 scales from [0.5,2.0]. Vertical reflection is indicated by a binary variable. (Total 11∗11∗2 = 242 T)
- For each T:
    - For each pixel q in the exterior of I_g, we search for its K nearest neighbors from the interior of I_g^i transformed by T, and choose only those whose distance is within a fixed threshold.
    - Each matched point p provides a shift vector.
    - We build a histogram of these shift vectors from all pixels in I_g.
    - The top 50 candidate translations form the set of representative translations M_T
- Based on number of patches covered in top 50 translation, we pick best 20-50 (T) transformation out of total 242.

Step 1b: Graph optimisation for each chosen T

- Translation vector at each pixel is chosen from the candidate set M_T by minimising the graph energy equation.
- For any translation M in M_T, I_i is first transformed by T and then shifted according to M.
- For pixels that cannot be covered by transformed I_i, cost is set to infinity.
- For pixels that voted for M, data cost is set to zero.
- For all other pixels data cost is set to a constant C. In implementation we use C=2.
- The graph equation is minimised using **alpha expansion** to find the optimal shift map for each chosen T transformation to give intermediate synthesis result I_T.

Step 2: Photomontage to pick optimum T at each pixel

$$\mathbb{E}(T) = \sum_{q} \mathbb{E}_d(T(q)) + \sum_{(p,q)\in N} \mathbb{E}_s(T(p), T(q))$$

We use one final alpha expansion to Minimize above energy function to get the final transformation for each pixel in image.

$$\mathbb{E}_d(T(q)) = E_d^T(M^T(q)) + \sum_{p\in N(q)} E_s^T(M^T(p), M^T(q))$$

$\mathbb{E}_d$ is the cost per pixel, which is same as the cost calculated for it in previous optimization.

$$\mathbb{E}_s(T(p), T(q)) = \|I_{T(p)}(q) - I_{T(q)}(q)\|_2 + \|I_{T(p)}(p) - I_{T(q)}(p)\|_2.$$

$\mathbb{E}_s$ is the smoothening cost for neighbourhood pixel. It minimizes the change in pixel values across transformations (T). (E_s is the smoothening cost across translations)

Limitations

This method focuses on simple filling and does not take care of semantic matching for extrapolation.

Technical details and requirements

Implementation platform: Matlab

Testing dataset: SUN360 panorama database

Thanks !