ASSIGNMENT #4 (EC720 A1)
"Digital Video Processing"
Date: October 31, 2017
Due date: November 13, 2017

1. *Motion estimation* (20 points).
This problem is about 2-D motion estimation based on phase correlation, briefly mentioned in class. The method is applied to a *real-valued* image sequence $u[\boldsymbol{x}, t]$, $\boldsymbol{x} \in \Lambda \subset R^2$, $t = kT$, where Λ is a 2-D orthonormal lattice. Note that each frame of the image sequence is considered to have infinite extent spatially to simplify the derivation.

   (a) Express the following correlation function:
   $$R_{uu}(\boldsymbol{y}) = \sum_{\boldsymbol{x}\in\Lambda} u[\boldsymbol{x}, t_1]u[\boldsymbol{x}+\boldsymbol{y}, t_2], \qquad \boldsymbol{y} \in \Lambda$$
   in terms of a spatial convolution.

   (b) The phase correlation function $\tilde{R}_{uu}$ is computed in 3 steps as follows:
      i. 2-D Fourier transform of R_{uu}: $\Psi_{t_1,t_2}(\boldsymbol{f}) = \mathcal{F}\{R_{uu}(\boldsymbol{y})\}$,
      ii. normalization of the result: $\tilde{\Psi}_{t_1,t_2}(\boldsymbol{f}) = \frac{\Psi_{t_1,t_2}(\boldsymbol{f})}{|\Psi_{t_1,t_2}(\boldsymbol{f})|}$,
      iii. inverse 2-D Fourier transform of the normalized result: $\tilde{R}_{uu}(\boldsymbol{y}) = \mathcal{F}^{-1}\{\tilde{\Psi}_{t_1,t_2}(\boldsymbol{f})\}$.

   Completing steps i. and ii., express $\tilde{\Psi}_{t_1,t_2}(\boldsymbol{f})$ as a function of $U_{t_1}(\boldsymbol{f}) = \mathcal{F}\{u[\boldsymbol{x}, t_1]\}$ and $U_{t_2}(\boldsymbol{f}) = \mathcal{F}\{u[\boldsymbol{x}, t_2]\}$. What is the physical meaning of $\tilde{\Psi}_{t_1,t_2}(\boldsymbol{f})$?

   (c) Consider the ideal case of translational, global motion between two frames:
   $$u[\boldsymbol{x}, t_2] = u[\boldsymbol{x} - \boldsymbol{d}, t_1], \qquad t_2 > t_1.$$
   Calculate $\tilde{R}_{uu}(\boldsymbol{y})$ in this case. What does it look like? Comment.

   (d) Explain the importance of step ii., i.e., what would happen had it been removed?

   (e) How can one apply this method in practice when motion in a sequence is not uniformly translational? Describe step-by-step how you would implement such an approach. Provide either a block diagram or a sequence of steps to take.

2. *Motion estimation comparison – Matlab* (50 points)
Based on your developments in the first problem, implement 2-D motion estimation *via* phase correlation, and test it on 3 image pairs: *missa_80* and *missa_84*, *coastguard_90* and *coastguard_95*, *container_1* and *container_30* in the following scenarios:

   (a) apply your method to the whole image, finding 3 most likely displacement estimates $\boldsymbol{d}$ (3 largest local maxima of $\tilde{R}_{uu}(\boldsymbol{y})$),

   (b) apply your method to 32×32 blocks, finding 3 most likely displacement estimates $\boldsymbol{d}$ for *each block*; then, in each block find the *optimal displacement* be evaluating the *mean absolute error* (MAE): $\sum_{\boldsymbol{x}\in B} |u[\boldsymbol{x}, t_2] - u[\boldsymbol{x} - \boldsymbol{d}, t_1]|$, where B denotes a 32×32 block and pixels outside the image are assumed zero-valued,

   (c) repeat the previous step for 16×16 blocks,

(d) for the case of 32×32 and 16×16 blocks also run the exhaustive-search disjoint-block matching with absolute-value criterion that you developed in assignment #3,

(e) compare the phase-correlation and block-matching results visually as well as numerically (entropy and MAE), and also compare the computational complexity (CPU time of *Matlab* code execution),

(f) with your solutions include the following plots for each image tested:

i. full-page correlation surface for the full-image case, with 3 most likely displacements in the title,

ii. 3×3 array of plots for 32-pixel blocks, in the following order (*subplot* notation):

- original image (3,3,1),
- prediction error for phase correlation, with entropy/MAE/CPU in title (3,3,2),
- prediction error for block matching, with entropy/MAE/CPU in title (3,3,3),
- vector field of optimal displacements for phase correlation (3,3,5),
- vector field for block matching (3,3,6),
- vector field of phase correlation displacements for largest maxima (3,3,7),
- vector field of phase correlation displacements for 2nd-largest maxima (3,3,8),
- vector field of phase correlation displacements for 3rd-largest maxima (3,3,9),

iii. 3×3 array of plots for 16-pixel blocks (same order as above)

(g) draw conclusions, especially concerning the optimal displacement field for phase correlation, performance of each algorithm and computational complexity.

3. *Parametric motion models - small rotations* (10 points).
Suppose a camera viewing distant scene (large Z) undergoes small 3-D rotation (small angles θ_X, θ_Y, and θ_Z) between times t and t', for example due to vibrations experienced by camera mounting hardware. First, derive the rotation matrix R for this case. Then, use perspective projection under large Z to derive an expression, in terms of the three angles, for 2-D displacement $\boldsymbol{d}[\boldsymbol{x}]$ of a point at location $\boldsymbol{x}$ on image sensor at time t, i.e., such that $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{d}[\boldsymbol{x}]$ where $\boldsymbol{x}'$ is this point's location at time t'.

4. *Parametric motion models - planar object* (20 points).
Recall the general case of 3-D affine motion of a rigid body, i.e., 3-D rotation plus 3-D translation, as described in class. This 3-D affine mapping described by equation:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$$

results in the following 2-D mapping:

$$x' = f\frac{(r_1x + r_2y + r_3f)Z + T_xf}{(r_7x + r_8y + r_9f)Z + T_zf}, \qquad y' = f\frac{(r_4x + r_5y + r_6f)Z + T_yf}{(r_7x + r_8y + r_9f)Z + T_zf}.$$

Show, that in the particular case of a planar surface $\alpha X + \beta Y + \gamma Z = 1$, the above general mapping simplifies to the eight-parameter projective mapping:

$$x' = \frac{a_0 + a_1x + a_2y}{1 + c_1x + c_2y}, \qquad y' = \frac{b_0 + b_1x + b_2y}{1 + c_1x + c_2y}.$$

Express, each of the coefficients $(a_0, a_1, a_2, b_0, b_1, b_2, c_1, c_2)$ in terms of rotation parameters r_*, translation T_*, surface parameters α, β, γ, and focal length f.