AUTOMATIC MOSAICKING OF PLANAR SURFACES: A PROGRESS REPORT

G.M. Cortelazzo, L. Lucchese, R. Frezza † A. Vettore ‡

† Department of Electronics and Informatics
‡ Department of Topography
UNIVERSITY OF PADOVA, ITALY

Commission VI, Working group 3

ABSTRACT

Mosaicking of planar surfaces is a classical photogrammetric problem for which a wealth of well established non-automatic solutions are available. Solving this task in a fully automatic way is a hard problem still open to investigations.This work reports on an hybrid technique operating both in the spatial and in the frequency domain which can be fully automated.

1. Introduction

Taking pictures in single shots produces images whose resolution depends on the size of the imaged scene. Images of a scene at a predetermined spatial resolution must be taken in pieces (subimages) to be subsequently composed, or mosaicked, into full images. The subimages can be registered from the common overlapping areas. Image registration can be performed non-automatically in many ways: the taking can be directly performed by orthophotographic arrangements or the images can be subsequently corrected so that they correspond to such an arrangement; contiguous images can be registered with or without markers. In the latter case the position of the left image with respect to the right one can be determined from the mutual correlation between the overlapping portions of the two images.

In this work we propose a procedure for automatically mosaicking subimages of a planar scene into a full image; the subimages can be taken from any position and their registration is completely automatic. Let us stress that the proposed method addresses only the case of planar objects, as it is based on projective transformations. It is well know that images of a planar surface, taken from any distance and view angle, are related by perspective transformations. The use of projective transformations waves from camera calibration since the intrinsic orientation parameters are implicitly included into the coefficients of the projective transformation. Therefore the procedure is suited to be used with commercial inexpensive camera equipment. Some calibration, aimed to correct for geometric distortion, may in principle be necessary if the photographic lenses are of poor quality or cover wide angles. Geometric distortions of medium range lenses of average commercial quality were found to be neglectable for this application.

The first part of this work presents a fully automatic procedure for determining the coefficients of the projective transformations among pairs of images, from which their mosaick can be straightforwardly obtained. Unfortunately this procedure is plagued by high computational complexity, essentially because it blindly matches features over the whole overlap area of contiguous images.

The second part of the work presents a procedure for automatically estimating affine transformations between pairs of images, which can cure the computational problems of the method for automatically estimating the coefficients of the projective transformations, since it can limit its search area for matching features with dramatic reductions of the computational costs.

The determination of the affine transformation between pairs of images is based on their Fourier transform. In this way it can be performed automatically and efficiently because it uses FFT algorithms. Furthermore, since it operates with the whole information of the images and not just with subsets of it (as any set of features is, no matter how meaningful may be the information associated with it) this method is very robust against noise and it works with highly subsampled images.

At the time of this writing the experimental work is still in progress, and final results assessing the performance of the proposed techniques are not available; nevertheless we thought worth reporting our procedure because of its originality and potential. This procedure was originally conceived as a tool for the automatic authoring of virtual visits of historical monuments, however it may be of interest also in other areas. Two of the authors are currently considering its possible application to videophotogrammetry and robotic navigation.

2. Automatic determination of projective transformations between images pairs

The automatic determination of the projective transformation between two images of the same area of a planar surface can be obtained in the spatial domain by the following method.

The features of the two images can be first automatically extracted by standard methods *[6]*. A quintuplet of features of the left $\mathcal{C}_l$ image forming a "regular" pentagon - the sense of the word "regular" will be subsequently explained - is matched against all possible quintuplets of the right image $\mathcal{C}_r$. Each correspondence of pentagons pairs is associated to a projective transformation, denoted as $\mathcal{P}$. The projective transformation $\mathcal{P}$ is applied to all the features of the left image, denoted as x, in order to evaluate the error measure defined as

$$e = \sum_{\substack{x \in \mathcal{C}_l \\ x' \in \mathcal{C}_r}} \|\mathcal{P}(x) - x'\| \qquad (1)$$

where x' is the feature of the right image spatially closest to $\mathcal{P}(x)$. If is less than a predetermined threshold, the projective transformation $\mathcal{P}$ is taken as estimate of the transformation between $\mathcal{C}_l$ and $\mathcal{C}_r$. In this way feature matching is included in the estimate of the projective transformation.

The use of quintuplets is due to the following reasons. Since the number of n-plets from F features-points is

$$\binom{F}{n} = \frac{\prod_{k=F-n+1}^{F} k}{n!} > \frac{(F-n+1)^n}{n!} \xrightarrow{F \to \infty} \frac{F^\infty}{n!} \qquad (2)$$

it grows more than exponentially with n. The minimum number of corresponding points necessary to identify a projective transformation is four and five is the smallest number next to it. Quintuplets have been preferred to quadruplets because of the pruning conditions associated to them. Two quintuplets are considered a valid match only if the pentagons associated to them satisfy to the following conditions: convexity; sum of their internal angles suitably bounded; smallest side not smaller than a predetermined value; the cross-ratios test satisfied *[4]*. The previously mentioned regularity characteristic has to be intended in this sense.

Unfortunately the computational complexity associated to the pentagons matching procedure is too high, even with pruning provision. As an example the number of quintuplets one can extract from a set of $F = 30$ features is $\binom{30}{5} = 142{,}506$; after pruning by means of the above conditions, simulations give $46{,}574$ possible pentagons, leading to $46{,}574^2$ matches! Therefore exhaustive searching for pentagon matches until the error measure is less than a predetermined value is in general not practicable. One way to make this search practically feasible is to keep F small, i.e. to use small size search areas in order to limit the number of features.

3. Reduction of the search area by means of affine transformation

The search area can be reduced by a first rough guess of the projective transformation given by an affine transformation. If the projective transformation $\mathcal{P}$ between the two images is

$$x' = \frac{Ax + b}{c^T x + 1} \qquad A \in R^{2\times 2} \qquad b, c \in R^2 \qquad (3)$$

then if relationship *(3)* can be approximated by an affine transformation $\mathcal{A}$

$$x' = Ax + b \qquad (4)$$

Condition $c^T x << 1$ corresponds to taking situations with little tilt. The estimate of A and b can be automatically performed in the frequency domain by the algorithm of the next section. A feature of the left image in principle may match any feature of the right image; however if $\mathcal{A}$ approximates $\mathcal{P}$ the matches for the features of a subregion $\mathcal{S}_l$ of the left image can be sought only in the subregion $\mathcal{S}'_r \doteq \mathcal{A}(x), x \in \mathcal{S}_l$. For safety we use a search area $\mathcal{S}_r$ slightly greater than $\mathcal{S}_l$ given by the union of $\mathcal{S}_r$ with a suitable tolerance border around it.

The control of the search area in the right image can be extremely effective with respect to computational complexity. Indeed if there are F features in $\mathcal{S}_l$ without any assumption on $\mathcal{P}$, one should search for their matches in the whole right image $\mathcal{C}_r$ formed by $N \times N$ pixels. If $\mathcal{S}_l$ has $M \times M$ pixels then also $\mathcal{S}'_r$ and its inclusion with a tolerance border $\mathcal{S}_r$ will approximately have $M \times M$ pixels only. If the features of $\mathcal{C}_l$ are F, the features associated to $\mathcal{S}_l$ and $\mathcal{S}_r$ are approximately $F \cdot \left(\frac{M}{N}\right)^2$.

Therefore by these geometrical considerations the number of features with respect to the blind search among $\mathcal{C}_l$ and $\mathcal{C}_r$ or $\mathcal{S}_l$ and $\mathcal{C}_r$, has been reduced by a factor $\left(\frac{M}{N}\right)^2$.

This complexity reduction easily amounts to various orders of magnitude; for instance for $N = 500$ and $M = 20$, it is $\left(\frac{M}{N}\right)^2 = 1.6 \cdot 10^{-3}$.

4. Frequency domain analysis of affine transformations

This section derives some relationships between the Fourier transforms of a pair of images supposedly related by an affine transformation. Let's assume that two images $I_1(x)$ and $I_2(x)$, $x \in R^2$, are related as

$$I_2(x) = I_1\left(A^{-1}x - b\right) = I_1\left(A^{-1}(x - Ab)\right) \qquad (5)$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in R^{2\times 2},\ det(A) > 0,\ b \in R^2 \qquad (6)$$

Equation *(3)* implies that image $l_2(x)$ is obtained from image $l_1(x)$ by first translating each point of $l_1(x)$ by b and then by applying the affine matrix A to the result of this translation. This is one of the two models for representing affine transformations in R^2. The other possible model is

$$l_2(x) = l_1\left(A^{-1}(x-c)\right), \qquad c \in R^2 \qquad (7)$$

which implies that $l_2(x)$ is obtained by inverting the order of the two operations on $l_1(x)$ with respect to the previous model. The two affine transformation models become equivalent for $c = Ab$. Throughout the text we will refer to *(3)*.

It is worth observing that constraint $det(A) > 0$ preserves the relative orientation of the images, i.e., $l_2(x)$ could be superimposed to $l_1(x)$ and viceversa without any reflection with respect to the axes.

Let $L_n(k) \doteq \mathcal{F}[l_n(x)|k]$, $k \in R^2$, denote the Fourier transform of image $l_n(x)$. Then the Fourier transform of equation *(3)* yields *[1]*

$$L_2(k) = det(A)\cdot L_1\left(A^T k\right)\cdot e^{-j2\pi k^T Ab} \qquad (8)$$

Since the translational vector b affects only the phases of the Fourier transforms in *(8)*, one may first estimate A from the magnitudes of *(8)*

$$\left|L_2(k)\right| = det(A)\cdot\left|L_1\left(A^T k\right)\right| \qquad (9)$$

An estimate Δ of $|det(A)|$ can be computed as $\Delta = L_2(0)/L_1(0)$. By defining $\tilde{L}_2(k) \doteq L_2(k)/\Delta$ equation *(9)* can be rewritten in normalised form as

$$\left|\tilde{L}_2(k)\right| = \left|L_1\left(A^T k\right)\right| \qquad (10)$$

The following relationships between areas of slices of $|L_1(k)|$ and $|\tilde{L}_2(k)|$ will be used in the proposed algorithm. For simplicity let's define $\mathcal{L}_1(k) \doteq |L_1(k)|$, $\mathcal{L}_2(k) \doteq |L_2(k)|$ and write *(10)* as $\mathcal{L}_2(k) = \mathcal{L}_1\left(A^T k\right)$ or, more explicitly, as

$$\mathcal{L}_2\left(k_x,k_y\right) = \mathcal{L}_1\left(a_{11}k_x + a_{21}k_y, a_{12}k_x + a_{22}k_y\right) \qquad (11)$$

It can be easily proved that the slice of along the line $k_y = k_x \tan\theta_1$ is mapped into a companion slice of along the line $k_y = k_x \tan\theta_2$, with θ_1 and θ_2 related as

$$\theta_1 = \phi_1(\theta_2) = \arctan\left(\frac{a_{12} + a_{22}\tan\theta_2}{a_{11} + a_{21}\tan\theta_2}\right) \qquad (12)$$

Furthermore, the slice of $\mathcal{L}_2(k)$ is a version of the corresponding slice $\mathcal{L}_1(k)$ of stretched by

$$\lambda_1(\theta_2) = \sqrt{\frac{\left(a_{11} + a_{21}\tan\theta_2\right)^2 + \left(a_{12} + a_{22}\tan\theta_2\right)^2}{1 + \tan^2\theta_2}} \qquad (13)$$

This concept can be precisely formalised by equation

$$\bar{\mathcal{L}}_2(\rho,\theta_2) = \bar{\mathcal{L}}_1(\lambda_1\rho,\theta_1) \qquad (14)$$

where $\bar{\mathcal{L}}_n(\rho,\theta) \doteq \mathcal{L}_n(\rho\cos\theta, \rho\sin\theta)$, $n = 1,2$, are the expressions of $\mathcal{L}_n\left(k_x,k_y\right)$ according to the polar coordinates $k_x = \rho\cos\theta$, $k_y = \rho\sin\theta$, $\rho \geq 0$, $0 \leq \theta < \pi$. The angle θ is limited in this range owing to the hermitian symmetry of the Fourier transform.

Of course equation *(14)* can also be rewritten as $\bar{\mathcal{L}}_1(\rho,\theta_1) = \bar{\mathcal{L}}_2(\lambda_2\rho,\theta_2)$ where $\lambda_2 = \lambda_1^{-1}$. Equation *(14)* reduces the 2-D problem of simultaneously estimating the four parameters a_{ij} of A into two 1-D problems. In each 1-D problem one has to find triplet $(\theta_1,\theta_2,\lambda_1)$ of *(14)*. Parameters a_{ij} can be subsequently determined via *(12)* and *(13)* from a pair of these triplets. Call

$$\beta_1 \doteq a_{12} + a_{22}\tan\theta_2 \quad \text{and} \quad \beta_2 \doteq a_{11} + a_{21}\tan\theta_2 \qquad (15)$$

and rewrite *(12)* and *(13)* respectively as

$$\theta_1 = \arctan\frac{\beta_1}{\beta_2} \quad \text{and} \quad \lambda_1 = \sqrt{\frac{\beta_1^2 + \beta_2^2}{1 + \tan^2\theta_2}} \qquad (16)$$

It is straightforward to show that β_1, and β_2, of *(16)* are

$$\beta_1 = \pm\lambda_1 \tan\theta_1 \sqrt{\frac{1 + \tan^2\theta_2}{1 + \tan^2\theta_1}} \qquad (17)$$

and

$$\beta_2 = \pm\lambda_1 \sqrt{\frac{1 + \tan^2\theta_2}{1 + \tan^2\theta_1}} \qquad (18)$$

A pair of triplets $\left(\theta_1^{(1)},\theta_2^{(1)},\lambda_1^{(1)}\right)$ and $\left(\theta_1^{(2)},\theta_2^{(2)},\lambda_1^{(2)}\right)$

gives, through *(15)*, *(17)* and *(18)*, the following linear system in the four unknowns a_{ij}

$$\begin{bmatrix} 0 & 1 & 0 & \tan\theta_2^{(1)} \\ 1 & 0 & \tan\theta_2^{(1)} & 0 \\ 0 & 1 & 0 & \tan\theta_2^{(2)} \\ 1 & 0 & \tan\theta_2^{(2)} & 0 \end{bmatrix} \cdot \begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{bmatrix} = \begin{bmatrix} \beta_1^{(1)} \\ \beta_2^{(1)} \\ \beta_1^{(2)} \\ \beta_2^{(2)} \end{bmatrix} \qquad (19)$$

Define the radial projections of magnitudes $\bar{\mathcal{L}}_n(\rho,\theta)$, $n=1,2$, as

$$\mathcal{P}_n(\theta)=\int_0^\infty \bar{\mathcal{L}}_n(\rho,\theta)d\rho,\ \ 0\le\theta<\pi,\ \ n=1,2 \qquad (20)$$

From the use of *(12)*, *(13)* and *(14)* in *(20)* it can be easily proved that

$$\mathcal{P}_2(\theta)=\frac{1}{\lambda_1(\theta)}\mathcal{P}_1(\phi_1(\theta)) \qquad (21)$$

5. Estimation of the affine approximation in the frequency domain

Expression *(21)* shows that $\mathcal{P}_2(\theta)$ is a version of $\mathcal{P}_1(\theta)$ modulated with respect to magnitude and phase. Let us observe that if there was no denominator at the right hand side of *(21)*, $\mathcal{P}_2(\theta)$ would simply be a scale-distorted replica of $\mathcal{P}_1(\theta)$. As $\lambda_1(\theta)$ is a smoothly varying function, the behaviour of $\mathcal{P}_2(\theta)$ roughly looks like that of $\mathcal{P}_1(\theta)$ with the scale distorsion[1]. That suggest that the local maxima of $\mathcal{P}_1(\theta)$ are mapped into the local maxima of $\mathcal{P}_2(\theta)$.

This idea motivates the following algorithm for determining the two triplets $\left(\theta_1^{(1)},\theta_2^{(1)},\lambda_1^{(1)}\right)$ and $\left(\theta_1^{(2)},\theta_2^{(2)},\lambda_1^{(2)}\right)$ in order to solve *(19)*.

Algorithm

1. find the two angles $\theta_1^{(1)}$ and $\theta_2^{(1)}$ where $\mathcal{P}_1(\theta)$ exhibits the two highest local maxima;
2. find the N angles $\theta_2^{(n)}$, $n=1,\ldots,N$, where $\mathcal{P}_2(\theta)$ exhibits the N highest local maxima;
3. for $i=1$ to $i=2$

    $\hat{\theta}_1^{(i)}=\theta_1^{(i)}$

    for $j=1$ to $j=N$

        compute $\lambda_1=P_1\left(\theta_1^{(i)}\right)/P_2\left(\theta_2^{(j)}\right)$

        if $\lambda_1>1$

            then stretch $\bar{\mathcal{L}}_2\left(\rho,\theta_2^{(j)}\right)$

                $\bar{\mathcal{L}}_{2s}\left(\rho,\theta_2^{(j)}\right)=\bar{\mathcal{L}}_2\left(\lambda_1\rho,\theta_2^{(j)}\right)$

                $e_j(\rho)=\bar{\mathcal{L}}_1\left(\rho,\theta_1^{(i)}\right)-\bar{\mathcal{L}}_{2s}\left(\rho,\theta_2^{(j)}\right)$

            else stretch $\bar{\mathcal{L}}_1\left(\rho,\theta_1^{(i)}\right)$

                $\bar{\mathcal{L}}_1\left(\rho,\theta_1^{(i)}\right)=\bar{\mathcal{L}}_1\left(\rho/\lambda_1,\theta_1^{(i)}\right)$

                $e_j(\rho)=\bar{\mathcal{L}}_{1s}\left(\rho,\theta_1^{(i)}\right)-\bar{\mathcal{L}}_2\left(\rho,\theta_2^{(j)}\right)$

        end

    end

    $\left(\hat{\theta}_2^{(i)},\hat{\lambda}_1^{(i)}\right)=\arg\min_j\int_0^\infty\left|e_j(\rho)\right|d\rho$

    end
4. solve system *(19)* with respect to a_{ij} by means of $\left(\hat{\theta}_1^{(1)},\hat{\theta}_2^{(1)},\hat{\lambda}_1^{(1)}\right)$ and $\left(\hat{\theta}_1^{(2)},\hat{\theta}_2^{(2)},\hat{\lambda}_1^{(2)}\right)$

[1]The details of the complete procedure, which considers a preliminary rough estimation of the function $\lambda_1(\theta)$ by means a non-linear least-square algorithm applied to quotient $q(\theta)\doteq\mathcal{P}_1(\theta)/\mathcal{P}_2(\theta)$, is reported in *[5]*.

Uncertainty, inherent the quadratic nature of relationships *(12)* and *(13)*, gives four consistent solutions for system *(19)*. Two of them can be ruled out since they do not satisfy constraint $det(A)>0$. Therefore step *4* of the algorithm returns two estimates for A:

$$\hat{A}_1 \quad \text{and} \quad \hat{A}_2=-\hat{A}_1 \qquad (22)$$

which can be disambiguated as follows. Consider the ideal case in which $\hat{A}_1=A$ and $\hat{A}_2=-A$ and compute

$$d'(x)=l_2(Ax)=l_1(x-b) \qquad (23)$$

and

$$d''(x)=l_2(-Ax)=l_1(-x-b)=d'(-x) \qquad (24)$$

Image $d'(x)$ is a version $l_1(x)$ translated by b. Image $d''(x)$ is symmetric to $d'(x)$ with respect to the origin. This fact suggests the use of phase correlation algorithm *[2, 3]* not only for estimating shift b, but also for solving disambiguation. Compute

$$\Phi'(K)\doteq\frac{L_1^*(k)\cdot D_1'(k)}{\left|L_1(k)\cdot D_1'(k)\right|}=e^{-j2\pi k^Tb} \qquad (25)$$

and

$$\Phi''(K)\doteq\frac{L_1^*(k)\cdot D_1'(k)}{\left|L_1(k)\cdot D_1'(k)\right|}=e^{-j2\theta(k)+j2\pi k^Tb} \qquad (26)$$

where

$$D_1^i(k) \doteq \mathcal{F}\left[d_1^i(k)|k\right] \quad \text{and} \quad \theta(k) \doteq arg\left(L_1(k)\right) \qquad (27)$$

The inverse Fourier transform of $\Phi'(k)$ is

$$\phi'(x) \doteq \mathcal{F}^{-1}\left[\Phi'(k)|x\right] = \delta(x-b) \qquad (28)$$

i.e. an impulsive function located at b. The inverse Fourier transform of $\Phi''(k)$, denoted as $\phi''(x)$, is not an impulsive one. Therefore comparison between the peaks of $\phi'(x)$ and $\phi''(x)$ allows one to rule out solution $\hat{A}_2$, related to $\phi''(x)$, and vector b can be estimated from the position of the peak of $\phi'(x)$. The complete procedure we devise for automatically estimating projective transformations has the following steps:

1. apply the algorithm of section *4* to two corresponding subregions $\mathcal{C}_l$ and $\mathcal{C}_r$ of the left and right images and obtain affine parameters A, b approximating the projective transformation;

2. automatically extract a limited number of features in $\mathcal{C}_l$ by *[6]* by means of suitable threshold values;

3. apply the algorithm of section *2*, using as search areas all pairs $\{\mathcal{S}_l, \mathcal{S}_r\}$.

6. Conclusions

The above procedure is a new frequency domain method for estimating affine transformations between images pairs. Our approach exploits the separation of the contributions of affine deformation and translation operated by the Fourier transform and the stretching properties of the Fourier transforms magnitudes.
This technique is used in order to reduce the computational complexity of a technique for automatically estimating projective transformations images of planar scenes and to obtain a method for automatically composing images of planar structures by mosaick.

Acknowledgements

This work was partially supported by *Progetto CNR "Realizzazione di immagini a colori colorimetricamente e geometricamente calibrate tramite mosaico di riprese".*

REFERENCES

[1] A. Papoulis, *The Fourier Integral and Its Applications,* New York, McGraw-Hill, 1962.

[2] C.D. Kuglin and D.C. Hines, *The phase-correlation image alignement method,* in Proc. IEEE 1975 Int. Conf. Cybern. Soc., 1975, pp. 163-165.

[3] S. Alliney and C. Morandi, *Digital* image *registration using projections,* IEEE Trans. Patt. Anal. Machine Intell., vol. PAMI-8, pp. 222-233, 1986.

[4] P.R. Wolf, *Elements of Photogrammetry,* New York: McGraw-Hill, 1983.

[5] L. Lucchese, *Frequency Domain Analysis and Estimate of Motion,* Ph.D. Thesis, University of Padova, February 1997.

[6] B.D. Lukas and T. Kanade, *An iterative image registration technique with an application to stereo vision,* Proc. of the 7th International Joint Conference of Artificial Intelligence, 1981.