iCEST 2011 Serbia, Niš, June 29 - July 1, 2011

Image Stitching – Basic Problems and Approaches for Their Solutions

Yulka Petkova[1], Todorka Georgieva[2]

Abstract –**Combining images in a greater image is widely used in computer vision. The paper is a brief survey of published methods and techniques for image stitching. Basic steps of image stitching procedure are described. Different approaches for their solutions are presented and analysed.**

Keywords –**Image Stitching, Template Matching, Image Processing, Representative Points, Control Points, Panoramas**

I. INTRODUCTION

The goal of image stitching is to assemble together different smaller images into a single high-resolution seamless image. Image stitching has become very important nowadays. It is due to the limited abilities of conventional capturing devices to produce great images. A wide variety of applications need this process. Some of them are:

- Combining satellite images, producing a greater image of an area [20];
- Combining aviation and astronomical images[11];
- Producing large medical images for the purposes of diagnostic [7, 10];
- Combining microscope images for so called Virtual Microscopy [17];
- Videoconferences [1, 5,8, 9, 19];
- Architectural walk-through [4];
- Making panorama high-resolution photo-images, using a set of pictures, taking even by handheld cameras or mobile devices [3, 20], etc.

Image stitching process proceeds after images acquisition and preprocessing. After the images have been acquired, some processing have to be applied basically to remove undesired noises inserted during the first step. The image stitching it self consists of two basic steps – Image registration and alignment and Image assembling, as it is shown in fig. 1.

Image registration is focused on establishing correspondence between objects from one image with the objects from the other image. During the step of image alignment a proper mathematical model, which connects pixel coordinates from one image with the pixel coordinates from another image has to be formulated. After that the right alignment between images must be established. During the second step aligned images are blended, removing the seam between them.

The common problem of all the steps in image stitching process is the great computational complexity. This is the reason a lot of investigations and researches to be made in order to accelerate computations, saving the quality of input images.

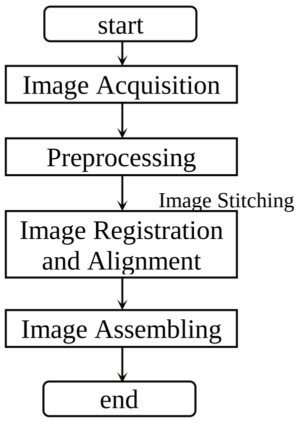

Fig. 1.Image stitching process

II. METHODS FOR IMAGE REGISTRATION

The key challenge in image stitching is the displacement of the objects in two different views of the same scene (parallax); moreover the displacement is different for objects at different depth levels (for cameras that do not have the same optical center). Thus, the objective of image registration stage is to find all matching (i. e. overlapping) images. Connected sets of image matches will be stitched later in a greater image. The problem is quadratic in the number of images, since each image could potentially match every other.

A method for image registration has to deal with a lot of problems, due to the methods of image acquisition. These problems can be:

- Differences between the intensities of the stitched images. They can be a result of changes in lightening, varying of angles between camera and lightening source, changing of the contrast between images, etc.;
- Presence of super illuminated areas, due to reflective objects in the scene;
- Presence of noise, due to the blurred lenses, dust, etc.
- Object moving during the process of images acquisition.

A method for image registration usually consists of four components. They are: a set of parameters which have to be compared; similarity measure, searching area and searching strategy.

Set of parameters for comparisons

Methods for automatic image registration can be divided in two major groups: direct or pixel-to-pixel comparison and feature-based comparison.

[1]Yulka Petkova is with the Department of Computer Sciences and Engineering at Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: jppet@abv.bg

[2]Todorka Georgieva is with the Communication Department at Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: todorka.georgieva@mail.bg

The advantage of direct methods [9, 17, 21], is that they use all possible information about the images, because they measure the contribution of every pixel in the image. Thus the final result is a precise registration. The basic disadvantage is that they are computationally very expensive and computational complexity is strongly dependent on the resolution, especially in the case of high-definition images. This is the reason they are not very suitable for real-time applications and they are not so widely used in practice. But direct methods can be used to refine the results, obtained by feature-based methods.

Feature-based methods use limited set of features, which are involved in comparisons. These features can be contours, edges, texture, colors, etc. [12, 13, 14, 17, 22, 24]. Set of features has to be properly chosen for any application. The basic advantage of these methods is the reduced computational complexity. As a rule these methods give precise enough results for the most of applications they are used in. It determines the great variety of algorithms, based on these methods [2].

For matching sequential frames in a video, the direct approach always works. For matching partially overlapped images for the purposes of photo-panoramas making it is not so useful.

Sometimes combinations of both feature-based and direct methods are used. These methods compare intensities of the selected control points (corners, edges, contours, etc.) [2, 18]. In other applications a feature-based methods are used for coarse registration and after that a more accurate direct comparison is used to refine the results [4].

Zoghlami et al. [24] use line segments together with control points to estimate homographies between compared images. Brown and Lowe [2] propose an approach, which basic advantage is that using invariant SIFT (Scale Invariant Feature Transform)features allows set of input images to be compared despite of rotations, scaling and different illuminations in them. Kumar et al. [10] propose to match the histograms of the component images in parts and to find correct correspondence between them (correct pixel coordinates of the relating pixels).

To summarize it can be noted that direct methods are more accurate but slower than feature-based. A lot of approaches use advantages of both direct and feature-based methods.

Similarity measures

Similarity measures are functions which return a value, corresponding to the similarity between comparing features. Similarities can be evaluated between orientation, size, color, intensities, etc. Values, obtained by computing the similarity measures are used to determine transformations, necessary for image alignment.

A suitable similarity measure should be chosen depending on type of comparison. The most frequently used similarity measures are:

- Sum of Squared Differences – SSD:

$$E_{SSD}(u)=\sum_i[I_1(x_i+u)-I_0(x_i)]^2=\sum_i e_i^2 \qquad (1)$$

where $I_0(x_i)$ is the template image, sampled in discrete pixel location x_i, $I_1(x_i+u)$ is the corresponding part of a searched image, where the similarity measure is computed, u is the displacement; e_i is the residual error.

- Robust error metrics – SRD (Sum of Robust Differences):

$$E_{SRD}(u)=\sum_i\rho\big(I_1(x_i+u)-I_0(x_i)\big)=\sum_i\rho(e_i) \qquad (2)$$

$\rho(e_i)$ is a robust function that grows less quickly than the quadratic penalty associated with least squares.

- Sum of Absolute Differences – SAD:

$$E_{SAD}(u)=\sum_i\left|I_1(x_i+u)-I_0(x_i)\right|=\sum_i|e_i| \qquad (3)$$

This function is preferred in motion estimation for video coding because of its speed. It is not proper for gradient descent approaches, because it is not differentiable at the origin.

- German-McClure function:

$$\rho_{GM}(x)=\frac{x^2}{1+x^2/a^2} \qquad (4)$$

This is an example of a smoothly varying function that is quadratic for small values but grows more slowly away from the origin. Here a is a constant that can be thought of as an outlier threshold, x is the intensity.

- Normalized Cross-Correlation:

$$E_{NCC}(u)=\frac{\sum_i\left[I_0(x_i)-\bar{I}_0\right]\left[I_1(x_i+u)-\bar{I}_1\right]}{\sqrt{\sum_i\left[I_0(x_i)-\bar{I}_0\right]^2.\left[I_1(x_i+u)-\bar{I}_1\right]^2}} \qquad (5)$$

where $\bar{I}_o=\frac{1}{N}\sum_i I_0(x_i)$ and $\bar{I}_1=\frac{1}{N}\sum_i I_1(x_i+u)$ are the mean images (usually intensities) of the corresponding parts of compared images; N is the number of pixels in the part. E_{NCC} is in the range [-1, 1], it guaranties high reliability of the results and this makes it suitable for some higher-level applications.

Normalized Cross-Correlation has its interpretation in frequency domain, where the convolution in the spatial domain corresponds to multiplication in Fourier domain. Applying Fast Fourier Transform algorithm significantly decreases computational complexity.

Phase Correlation is also used in some applications [2] where motion estimation has to be computed.

Kumar et al. [4] propose mutual information to be used for establishing the best matching.

$$MI(I_0,I_1)=\frac{H(I_0)+H(I_1)}{H(I_0,I_1)} \qquad (6)$$

where $H(I_0),H(I_1)$ are entropies of images; $H(I_0,I_1)$ is the joint entropy of two images; $MI(I_0,I_1)$ is the mutual information.

Patrik Nyman [16] proposes using of SURF (Speeded Up Robust Features) for image registration and alignment. SURF are compared according to the Euclidean distance and the minimum distance between them is found.

Brawn et al. [2]use SIFT features which are located at scale space maxima/minima of a difference of Gaussian function. Scale and orientation establishment at each feature location gives a similarity invariant frame in which to make measurements. SIFT features are invariant under rotation and scale. This is the reason their method can handle images with arbitrary orientations and zoom.

To summarize it can be noted that the most precise similarity measure is NCC, but it is computationally very expensive. A trade-off between precision and computational complexity is usually made depending on the requirements to the particular application.

III. Searching Strategy

Searching strategy is an algorithm, which decides how to choose next transformations from the searching set. Searching methods usually use pattern matching for image alignment [18].

The simplest searching algorithm is an exhaustive comparison with the template which calculates the similarity measure for each position and transformation in the searched set (so called brute-force algorithm). In this way the optimal similarity measure is guaranteed to be the globally optimal measure. But this algorithm has got a great computational complexity, thus it works slowly which makes it unworkable for real-time applications.

Acceleration of algorithms, keeping high reliability and precision, is another provocation to researchers. One of the possible decisions is to restrict the searched positions, for example – to search around the most probable position which is known in advance [2].Other techniques use different feature sets. In this method binary images, generated from the selected features(usually corner or edge points) are compared.

Another technique is so called coarse-to-fine search. It is an iteration process which uses a coarse determination of the most probable position. After it has been determined a fine search around it is performed in order to refine the result.

Registration with step search strategy is proposed by Tzi in [23]. In this strategy only five positions are evaluated for their respective similarity measures in each iteration of the search. The five positions include a center point and four points respectively in the north, east, south and west of the center point. The distances between the center point and the four other points are the step sizes. Initially, the vertical step is half the distance between the center point and the top of border, and the horizontal step size is half the distance between the center point and the left border. During the processing step sizes are corrected. When both step sizes reach 1, the similarity measures of eight positions around the center point are evaluated to determine the position with the best similarity measure. Since the step size is reduced by half in each iteration, the search algorithm converges a solution very quickly.

Some algorithms use combinations of methods. One such method is registration with binary edge image and restricted search set. With this combination, the chance of misalignment can be reduced by limiting the search to a defined neighborhood, within the optimal overlapping position is guaranteed to occur.

Another example of combined algorithms is a combination of restricted search set and step search. Chia-Yen Chen in [4]proposes the feature set to be the averaged intensity or the binary edge image. The similarity measure is either the sum of absolute differences or the standard deviation of the intensity differences. A restricted search set is used to decrease the chance of the algorithm converging towards the local optimum away from the position with the globally optimal similarity measure.

Other techniques use image pyramids to find the optimal match and transformation between two images at successively higher resolutions. Once the pyramids have been built, the registration is fast, but the process of building images is very time consuming.

To summarize it can be said that brute-force searching strategy is the most reliable but computationally most expensive. The other strategies are faster, but there is a real chance to skip the right correspondence between images.

IV. Methods for Image Assembling

After the steps of registration and alignment images have to be assembled or blended in a common image. Image assembling is a process of adjusting the values of pixels in the registered and aligned images, such that when the images are joined, the transition from one image to the other is invisible. With other words - the basic problem of this process is how to merge the images so that the seam between them to be visually undetectable. A seam is the artificial edge generated by the intensity differences of pixels immediately next to where the images are joined.

Methods for image blending can be separated in two categories: transition smoothing methods and finding the optimal seam.

Transition smoothing methods try to minimize the seam between images smoothing the edges of the image. The basic disadvantage of these methods is that blurry areas are created. Recently some methods, using multi-resolution blending, wavelets and gradient-domain blending are published [21].Gradient blending calculates a smooth weighted blend from one side of the overlapping parts to the other. The effect reduces issues like varying background intensities and provides the smooth edge transition between adjacent images. These methods need finding a least square solution of a Poison equation which is computationally very expensive [6].

Methods with optimal seam [21] try to put the seam where differences between images are as small as possible. Patrik Nyman [16] proposes so called watershed segmentation to be used for seam position establishment. In watershed segmentation one regards the image as a topological map, where watershed barriers define the different segments. The source and sink of the problem are defined as the non-overlapping regions from the first and second image respectively. The segments from the watershed algorithm are set as the nodes in the graph. The total sum in the difference image of the boundary pixels between two segments is set as the weight between the segments. Using a max-flow algorithm the minimum cut is found. It is then the optimal seam between both images.

One approach to remove the seam is to perform the intensity adjustment locally, within a defined neighborhood of the seam [15]. Another approach is to perform a global intensity adjustment on the images to be merged, so that apart from the intensity values outside the overlapping regions may also need to be adjusted [15].

Chia-Yen Chen [4] investigates four different intensity adjustments: linear distribution of intensity differences; linear distribution of median intensity differences; intensity adjustment with respect to median filtered regions and intensity adjustment with respect to corresponding pixels in overlapping region. The second method used by him gave the best results. The reason is that the adjustment to the original intensity levels is kept to a necessary minimum. The amount of adjustment is also proportional to the intensity differences between the joined images. Therefore, large intensity differences between the images indicate that the merged images may not retain the quality of the original images as well as when the intensity differences are small.

For obtaining a seamless stitching Kumar et al. [10] propose triangulation averaging to be applied on the overlapping area of the assembled images. The overlapped area of the left image is multiplied with averaging image whose intensity starts with 0 and changes to 1. The overlapped area of the right image is multiplied with averaging image whose intensity starts with 1 and changes to 0. The algorithm does not change the original quality of the images except in the overlapped area, but it produces a seamless image.

No quantitative evaluations are used for the quality of image blending. Visual evaluation is always used.

To summarize it can be noted that better methods for seamless image blending are based on finding an optimal seam between assembled images.

V. Conclusions

In this paper different methods and approaches for image stitching are briefly considered in order to be helpful for better understanding of different stages involved in generation of panoramic images. According to the published results a comparison between them is made.

Set of parameters, taking part in comparisons, similarity measures and different searching strategies are described. Methods for seamless image assembling (or blending) are also briefly presented.

Acknowledgements

The work presented in this paper was partially supported within the project BG 051PO001-3.3.04/13 of the HR Development OP of the European Social Fund 2007-2013.

References

[1] Adam M., C. Jung, S. Roth, G. Brunnett, Real-time Stereo-image Stitching using GPU-based Belief Propagation, VMV 2009, pp. 215-224

[2] Brown M., D. G. Lowe, Automatic Panoramic Image Stitching using Invariant Features, International Journal of Computer Vision, Volume 74 Issue 1, August 2007

[3] Brunton A., C. Shu, Belief Propagation for Panorama Generation, 3D Data Processing Visualization and Transmission, International Symposium on, pp. 885-892, 2006

[4] Chia-Yen Chen, Image Stitching – Comparisons and New Techniques, CITR-TR-30, Computer Science Department of the University of Auckland, (http://www.tcs.auckland.ac.nz)

[5] Coleshill E., Alexander Ferworn, Panoramic Spherical Video – The Space Ball, ICCSA, 2003

[6] Eden A., M. Uyttendaele, R. Szeliski, Seamless Image Stitching of Scenes with Large Motions and Exposure Differences, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'2006), pp. 2498-2505

[7] Goossen A., T. Pralow, R. Grigat, Automatic Stitching of Digital Radiographies using Image Interpretation, 2009

[8] Heinrich G., M. Leitner, C. Jung, F. Logemann V. Hahn, A Platform for Audiovisual Telepresence using Model- and Data-based Wave-field Synthesys, 125th AES Convention San Francisco, 2008

[9] Irani M., P. Anandan, About Direct Methods, Vision Algorithms: Theory and Practice, No 1883 in LNCs, pp. 267-277, Springer-Verlag, 1999

[10] Kumar A., R. S. Bandaru, at al., Automatic Image Alignment and Stitching of Medical Images with Seam Blending, World Academy of Science, Engineering and Technology, 65, 2010, pp. 110 – 115

[11] Majumdar J., S. Vinay, S. Selvi, Registration and Mosaicing for Images Obtained from UAV, International Conference on Image Processing and Communications, pp. 198-203, 2004

[12] McLauchlan P., A. Jaenicke, Image Mosaicing Using Sequential Bundle Adjustment, Image and Vision Computing, 20 (9-10):751-759, 2002

[13] Mikolajczyk K., C. Schmid, Scale& Affine Invariant Interest Point Detectors, International Journal on Computer Vision 60 (1), 63-86, 2004, Kluwer Academic Publishers

[14] Mikolajczyk K., T. Tuytelaars, C. Schmid, A Comparison of Affine Region Detectors, International Journal on Computer Vision Springer Science + Business Media, Inc., DOI: 10.1007/sl 1263-005-3848-x

[15] Milgram D., Adaptive Techniques for Photomosaicing, IEEE Transactions on Computers, vol. C-26, pp. 1175-1180, 1977

[16] Nyman P., Image Stitching using Watersheds and Graph Cuts, Centre of Math. Sciences, Lund University, Sweden, 2007

[17] Romero E., F. Gomez, M. Iregui, Virtual Microscopy in Medical Images: a Survey, *Modern Research and Educational Topics in Microscopy*, Formatex 2007, pp. 996-1006

[18] Ritter G., J. Wilson, Handbook of Computer Vision Algorithms in Image Algebra, CRC Press, 1996, pp. 225-239

[19] Steedly D., et al., Efficiently registering video into panoramic mosaics, In Tenth International Conference on Computer Vision, Beijing, China, 2005, pp. 1300 - 1307

[20] Szeliski R., H. Shum, Creating full Vew Panoramic Image Mosaics and Environment Maps, SIGGRPH’97, Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, pp. 251-258, 1997

[21] Szeliski R., S. Kang, Direct Methods for Visual Scene Reconstruction, In IEEE Workshop on Representation of Visual Scenes, pp. 26-33, Cambridge, 1995

[22] Tuytelaars T., L. V. Gool, Matching Widely Separated Views Based on Affine Invariant Regions, International Journal on Computer Vision 59 (1), 61-85, 2004

[23] Tziritas G., C. Labit, Advances in Image Communication, vol. 4, Motion Analysis for Image Sequence Coding, Elcevier, 1994

[24] Zoghlami I., O. Faugeras, R. Deriche, Using Geometric Corners to build a 2D Mosaic from a Set of Images, In Proceedings of the International Conference on Computer Vision and Pattern Recognition, IEEE, June 1997