Expert Systems With Applications 91 (2018) 322–346

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Ultra-fast basic geometrical transformations on linear image data structure

Vladan Vučković[a,*], Boban Arizanović[a], Simon Le Blond[b]

[a] Faculty of Electronic Engineering, Computer Department, P.O. Box 73, 18000 Niš, Serbia
[b] University of Bath, Department of Electronic & Electrical Engineering, United Kingdom

ARTICLE INFO

Article history:
Received 1 May 2017
Revised 8 September 2017
Accepted 9 September 2017
Available online 9 September 2017

Keywords:
Linear transformations
Affine transformations
Spatial transformations
LUT
Rotation
Machine optimization

ABSTRACT

This paper presents a general, ultra-fast approach for geometrical image transformations, based on the usage of linear lookup hash tables. The new method is developed to fix distortions on document images as part of a real-time optical character recognition (OCR) system. The approach is generalized and uses linear image representation combined with pre-computed lookup tables. Backward mapping is used for generation of lookup tables, while forward mapping is presented as an alternative and more efficient mapping model for specific cases. Also, a theoretical space and time complexity analysis of the proposed method is provided. To achieve maximal computational performance, pointer arithmetic and highly-optimized low-level machine code implementations are provided, including the specialized implementations for horizontal mirror, vertical mirror, and 90° rotation. Also, a modified variant of the approach, based on auto-generated machine code is presented. Very high computational performances are achieved at the expense of memory usage. The performances from the perspective of time complexity are analyzed and compared with classical implementation, FPGA implementation, and other implementations of the image rotation. Numerical results are given for a set of different PC specifications to provide full insight into the implementation performances. The processing time for very large images are below 200 ms for backward mapping and below 100 ms for forward mapping for most machines, which is 30–60 times faster than the classical implementation, 5–20 times faster than the FPGA implementation, and up to 6 times faster than other implementations of image rotation. Original documents belonging to Nikola Tesla are used for visual demonstration of performance.

1. Introduction

The performances of OCR systems are usually highly dependent on the input document image which is being processed. Distortions, which can appear in a document image, have to be fixed from the outset using geometrical image transformations. Due to their importance in OCR, such transformations and their implementations have thus long been the subject of research. Although the transformations can be applied individually, generally, their use makes more sense in complex systems (Eldon, 1988; Giulieri, Nolibe, & Richon, 1988; Hagege & Francos, 2005; Liu, Yin, Liu, & Wei, 2013; Schmalz, 1993; Younes, 2006), where they are usually focused on fixing distortions (Chang & Fitzpatrick, 1990). The emphasis of this new method is linear transformations, but recent research also includes perspective and nonlinear transformations (Devich & Weinhaus, 1980; Evemy, Allerton, & Zaluska, 1989; Weeks, Myler, & Emery, 1994). Geometrical image transformations can be realized using quantum algorithms based on n-qubit normal arbitrary superposition state (NASS) (Fan, Zhou, Jing, & Li, 2016). Handwritten documents can be classified using affine transformation and 2D projection transformation (Yamashita & Wakahara, 2016). A fast FPGA based implementation for 3D affine transform was proposed by Mondal, Biswal, and Banerjee (2016). Evaluation of the linear geometrical image transformations is usually performed by analyzing the implementation time complexity and the interpolation method, which determines the image quality after the transformation is applied. Estimation of the linear transformations by analyzing the periodic properties of interpolation using the second-derivative of the transformed image, was proposed by Ryu and Lee (2014). Affine transformations can be used in order to perform a fractal based image compression (Raittinen & Kaski, 1993). Representation of most classes of 3D objects can be achieved using feed-forward neural networks and affine transformations can be applied to objects in order to prove the

* Corresponding author.
E-mail addresses: vladan.vuckovic@elfak.ni.ac.rs, vladanvuckovic24@gmail.com (V. Vučković), bobanarizanovic@hotmail.com (B. Arizanović), S.P.leBlond@bath.ac.uk (S.L. Blond).

http://dx.doi.org/10.1016/j.eswa.2017.09.011
0957-4174/

possibility of this modeling (Piperakis & Kumazawa, 2001). Affine transformation can also find application in cryptography for the simultaneous encryption-decryption of two images (Kovalchuk, Peleshko, Navytka, & Sviridova, 2011). Some works deal with applications of affine transformations for motion compensation (Lopes & Ghanbari, 2002; Nakaya & Harashima, 1994; Rodrigues, da Silva, & de Faria, 2001). Pham and Nakamura (2015) proposed a new algorithm based on affine transformations for deformation of robot trajectories. Affine transformations can also be performed in the frequency domain (Lucchese, 2001). Pei and Hsiao (2015) proposed a method for spatial affine transformations using a fractional shift Fourier transform. An FPGA implementation of affine transformations was proposed by Bensaali, Amira, Uzun, and Ahmedsaid (2003). Decomposition of rotation into a sequence of one-dimensional translations, which represents a fast convolution-based interpolation and preserves the high quality of the rotated image, was described by Unser, Thévenaz, and Yaroslavsky (1995). High precision rotation angle estimation for rotated images was proposed by Qian, Li, and Yu (1993). A method which exploits the hidden periodicities in the rotated image in the frequency domain using the 2D spectrum was demonstrated by Chen, Ni, and Shen (2014). The distortion problem arising from repeated image rotation applied on the compressed image was addressed by Yi, Joo, and Kim (2015). An algorithm for image rotation and correction based on local feature was proposed by Li and Dan (2013). Image rotation can be also used in video processing, for example, image rotation algorithm for a traffic monitoring system (Tan, Zhang, & Song, 2012). Image scaling has been accomplished using a real-time FPGA-based hardware architecture for implementation of bicubic interpolation (HABI) (Nuno-Maganda & Aries-Estrada, 2005). Another FPGA implementation of image rotation in video was proposed by Berthaud, Bourennane, Paindavoine, and Milan (1998), which exploits B-spline interpolation.

As mentioned previously, geometrical image transformations normally form part of a complex system, such as an OCR algorithm. Geometrical image transformations are applied in a very early stage of the character segmentation process, since further processing is impossible if distortions are present. The most frequent use of geometrical image transformation is for document image skew correction (Vučković & Arizanović, 2017b). Document skew correction is applied in combination with document skew estimation, and image rotation is used for skew correction process (Yu & Jain, 1996). To achieve real-time character recognition of Chinese documents for a reading robot, Yu, Dong, Wei, and Shen (2006) proposed a fast image rotation algorithm. Another image rotation algorithm used for document skew correction avoids multiplications in order to reduce the computational cost (Cao, Wang, & Li, 2003). Kapoor, Bagai, and Kamal (2004) proposed a document skew correction algorithm which exploits the Radon transform. In order to ensure distortion-free rotation, document skew correction can also be achieved using multi-rate signal processing principles (Mahata & Ramakrishnan, 2000).

One design objective of the new method was to achieve very high computational performances, since it is intended for real-time OCR systems (Vučković & Arizanović, 2017a). The approach is based on pre-computed lookup tables for mapping offsets to reduce the transformation processing time. To provide the highly-optimized implementation, linear image representation is used. Implementations using pointer arithmetic and highly-optimized machine code are used for optimal efficiency. Implementation time complexity is evaluated on several different PC machines, to provide a range of results for a fuller understanding of performance. Results show that the processing time even for very large images are below 200 ms in the case of backward mapping and below 100 ms in the case of forward mapping, which proved to be the more efficient mapping model but can only be used in specific cases. Pointer arithmetic and highly optimized machine code implementations proved to be almost 50 times faster than the classical procedure for image rotation, 5–20 times faster than the FPGA implementation, and up to 6 times faster than diverse algorithms for forward and inverse image rotation. Also, the image rotation and scaling quality using the proposed approach is quantitatively compared with equivalent Photoshop® transformations. Beside the optimized implementations of the new method, specialized procedures for horizontal mirror, vertical mirror, and 90° rotation are provided along with performance results. Since the new technique requires three memory accesses which limits the processing time, a modified implementation based on auto-generated machine code is also presented. The visual performances of these new procedures are demonstrated using original Nikola Tesla documents from the "Nikola Tesla Museum" in Belgrade. The numerical results demonstrate the new method is ideal for implementation in real-time systems.

This paper is organized as follows: Section 2 provides description of the related works with the focus on image rotation, which are used later for comparison and evaluation of the new method. Section 3 provides the detailed mathematical background of linear transformations and gives general definition of transformation matrices used for spatial image transformations. In Section 4 the complete approach is presented, including the crucial optimization steps which enable fast implementation, as well as theoretical space and time complexity analysis. Section 5 offers more details about the fast implementations using pointer arithmetic and highly-optimized low-level machine code. In Section 6, experimental results for time complexity and transformation quality are presented using large images with different dimensions. In the concluding Section 7, the pros and cons of the new method are discussed, in addition to plans for future work.

2. Related works

Within the field of linear image transformations, image rotation and scaling are frequent research topics. Related work is usually focused on efficient implementations without compromising image quality.

Zhu, Deng, and Gao (2016) proposed a learning-to-rank approach for estimation of the image scaling factor. This uses the normalized energy density features and moment features and is based on training the parameters which represent the difference of previously mentioned features for ordered image pairs.

Ashtari, Nordin, and Kahaki (2015) proposed a fast image rotation algorithm which preserves quality of the image. The method determines base-line equation on the target image and uses floating-point multiplications for this task. Other lines are determined using the base-line pixel coordinates.

Cheng and Wan (2015) proposed an image rotation which uses a radial basis function (RBF). The advantages are related to the improvement of the interpolation mechanism, giving better results than all classical interpolation methods, including the method based on Hermite basis expansion.

The previously described image rotation approaches mainly use image processing techniques in the spatial domain. Fu and Wan (2015) proposed image rotation based in the frequency domain using a discrete cosine transform (DCT). This involves performing a two-dimensional DCT on image blocks for obtaining the frequency domain information, followed by a two-dimensional IDCT before the final interpolation.

In order to evaluate the computational performances of the new method, it is necessary to give more details about the approaches which are used for comparison. Bourennane, Milan, Paindavoine, and Bouchoux (2002) proposed a real-time image rotation implementation using FPGAs. For this task, static and dynamic