Defence Research and Development Canada
Recherche et développement pour la défense Canada

Technology review:

Face recognition with poor quality images

Xue Dong Yang, Richard Dosselmann and Marian Moise

Defence R&D Canada – Ottawa

Contract Report
DRDC Ottawa CR 2012-012
May 2012

Canada

Technology review:

Face recognition with poor quality images

Xue Dong Yang
Richard Dosselmann
Marian Moise
University of Regina

Prepared By:
Xue Dong Yang
Department of Computer Science
University of Regina
Regina, SK S4S 0A2
Canada

Contract Project Manager: Qinghan Xiao, 613-998-1245
PWGSC Contract Number: W7714-4500823057
CSA: Qinghan Xiao, Defence Scientist, 613-998-1245

Defence R&D Canada – Ottawa

Contract Report
DRDC Ottawa CR 2012-012
May 2012

Scientific Authority

Original signed by Qinghan Xiao

Qinghan Xiao
Defence Scientist

Approved by

Original signed by Julie Lefebvre

Julie Lefebvre

Section Head

Approved for release by

Original signed by Chris McMillan

Chris McMillan

Chief Scientist

Abstract

Since the event of September 11th, 2001, research efforts in face recognition have intensified significantly, with much progress being made in the last decade in the essential areas of face detection and recognition algorithms. However, the development of face recognition systems that are truly useful and reliable in different and practical applications has been slow, mainly due to the limitations of existing algorithms. New challenges are raised in intelligence operations in which face images are captured in varying and unpredictable conditions. The resulting images usually contain many defects caused by poor or changing lighting, bad pose or camera angles that cause face occlusion and geometry changes. Worse yet, at a long distance, face images are sometimes of low resolution and contain high levels of noise. Before designing new application systems to deal with such situations, it is necessary to review and evaluate current face detection and facial recognition technologies. The objective of this technical report then is to review relevant research on facial recognition involving low resolution and/or poor quality images.

We review a large number of scientific publications that address some of the questions above, albeit at different depths. This review focuses on the major topics and representative examples to date, so as to provide the reader with an understanding of the primary issues and current status of this field. It is not intended to be an exhaustive survey nor an inclusive list of all published literature in the last decade. The contents are organized into four sections. Section 2 reviews preprocessing techniques that strive to improve the performance of face detection and recognition. Section 3 looks at new developments in face detection algorithms. Section 4 moves on to consider new face recognition algorithms developed in the past decade. Section 5 then takes a look at face tracking algorithms. Finally, Section 6 provides some concluding remarks.

Résumé

Depuis les attentats du 11 septembre 2001, la recherche en reconnaissance faciale s'est intensifiée et l'on a fait des pas de géant dans les domaines essentiels de la détection des visages et des algorithmes de reconnaissance. En revanche, la création de systèmes de reconnaissance faciale vraiment utiles et fiables dans différents contextes pratiques traîne, retardée principalement par les limites des algorithmes existants. Effectivement, les opérations de renseignement d'aujourd'hui présentent de nouvelles difficultés, car on doit souvent y prendre des images de visages dans des conditions variées et imprévisibles. Un éclairage mauvais ou inconstant, une pose ou un angle inadéquats encombrent de défauts la plupart de ces images, si bien que le visage s'en trouve caché ou subit des déformations géométriques. Pire encore, certaines images prises de loin ont une résolution trop basse et contiennent beaucoup de bruit. Avant de concevoir de nouveaux systèmes pour régler ce genre de problèmes, il importe de passer en revue les technologies de détection et de reconnaissance faciales. Le présent rapport technique examine donc la recherche actuelle sur le problème des images à basse résolution ou de mauvaise qualité en reconnaissance faciale.

Nous examinons (les uns plus en profondeur que les autres) quantité d'articles scientifiques qui abordent certains des thèmes ci-dessus. Nous nous concentrons sur les thèmes majeurs et sur les exemples représentatifs, pour que les lecteurs comprennent les grands enjeux et l'état actuel de ce domaine; il n'est pas question d'une revue exhaustive ni d'une liste complète de tous les articles publiés sur le sujet depuis dix ans. Le développement comporte quatre sections : la section 2 examine les techniques de prétraitement en usage pour rendre plus performantes la détection et la reconnaissance faciales; la section 3 donne les dernières nouvelles en matière d'algorithmes de détection faciale; la section 4 enchaîne en décrivant les algorithmes de reconnaissance faciale conçus dans les dix dernières années; la section 5 aborde les algorithmes de suivi des visages; finalement, la section 6 forme la conclusion.

Executive summary

Technology Review: Face Recognition with Poor Quality Images

Xue Dong Yang, Richard Dosselmann, Marian Moise; DRDC Ottawa CR 2012-012; Defence R&D Canada – Ottawa; May 2012.

Introduction or background: Much progress has been made in the last decade in the essential areas of face detection and recognition algorithms. However, the development of reliable face recognition systems has been slow. New challenges are raised in intelligence operations in which face images are captured in varying and unpredictable conditions. The resulting images usually contain many defects caused by poor or changing lighting, bad pose or camera angles that cause face occlusion and geometry changes. Before designing new application systems to deal with such situations, it is necessary to review and evaluate current face detection and facial recognition technologies. The objective of this technical report then is to review relevant research on facial recognition involving low resolution and/or poor quality images.

Results: The performance of face recognition algorithms has steadily improved in the last decade. However, these improvements, in our opinion, are evolutionary. True breakthroughs in this challenging area are rare. Great leaps in face representation and recognition are always possible, but cannot be expected to occur by any specified date. What can we do and what should we do before this happens? Based on experiences gained from other successful areas, we would make the following concluding remarks:

- The most promising direction is one that captures and stores face images in 3D format in a database, seeing that 3D models offer the best chance at solving the primary issues that affect face recognition, namely changes in pose and illumination conditions.
- If query images can also be acquired in 3D format, then high rate face recognition can be achieved.
- If the query images are 2D, but the database images are 3D, new images can be synthesized from the 3D model to match the pose and illumination conditions of those of the query image.
- When both the face image in the database and the query image are 2D and the query images are of low quality, we must rely on the best existing algorithms with fine tuning and optimization. In particular, preprocessing techniques should not be under-estimated. Finer face alignment at the sub-pixel level could lead to further performance improvements during the recognition stage. Face detection with very dark skin tones is also an under-studied area. More work in this area is needed.
- While most applications call for reliable positive identification, some applications can actually benefit from reliable negative rejection. If a face detection and recognition algorithm can reject all frames that either have no faces in them or that have faces that clearly do not match the query face, human operators need only look at the remaining frames containing possible matches.

Significance: Through this technology review, the readers will have a good understanding to the current status of face recognition technologies with low resolution and/or poor quality face

images. This is particularly valuable to the system developers for military applications because many of its operations are in the field with extreme varying and unpredictable conditions.

Future plans: Our research group, with a track record of research and development collaboration with DRDC on face recognition application system, we will be very enthusiastic to continue the participation in DRDC's future projects in this strategic area. Our previous research experience with 3D face matching and strong 3D computer graphics expertise make us an ideal partner to develop the next generation pure 3D and/or 3D/2D hybrid face recognition systems.

Sommaire

Technology review: Face recognition with poor quality images

Xue Dong Yang; Richard Dosselmann; Marian Moise

DRDC Ottawa CR 2012-012; R & D pour la défense Canada – Ottawa; mai 2012.

Introduction ou contexte: Depuis dix ans, les progrès n'ont pas manqué dans les secteurs essentiels des algorithmes de détection et de reconnaissance faciales. En revanche, la création de systèmes de reconnaissance faciale fiables traîne. Effectivement, les opérations de renseignement d'aujourd'hui présentent de nouvelles difficultés, car on doit souvent y prendre des images de visages dans des conditions variées et imprévisibles. Un éclairage mauvais ou inconstant, une pose ou un angle inadéquats encombrent de défauts la plupart de ces images, si bien que le visage s'en trouve caché ou subit des déformations géométriques. Pire encore, certaines images prises de loin ont une résolution trop basse et contiennent beaucoup de bruit. Avant de concevoir de nouveaux systèmes pour régler ce genre de problèmes, il importe de passer en revue les technologies de détection et de reconnaissance faciales. Le présent rapport technique examine donc la recherche actuelle sur le problème des images à basse résolution ou de mauvaise qualité en reconnaissance faciale.

Résultats: Depuis dix ans, la performance des algorithmes de reconnaissance faciale augmente progressivement. Nous qualifions les algorithmes d'évolutionnaires, au sens que dans ce domaine difficile, les percées sont rares. Bien sûr, une avancée majeure en reconnaissance ou en représentation des visages pourrait survenir, mais on ne saurait prévoir quand. D'ici là, que pouvons-nous faire, que devons-nous faire? L'expérience d'autres secteurs florissants nous inspire les observations finales suivantes :

- L'avenue la plus prometteuse consisterait à prendre des images de visages tridimensionnelles et à les stocker dans une base de données, car des modèles tridimensionnels seraient notre meilleure chance de lever les principaux obstacles à la reconnaissance faciale, à savoir les changements de poses et d'éclairage.
- En interrogeant la base de données à l'aide d'images elles-mêmes tridimensionnelles, nous maximiserions le taux de reconnaissance.
- Si nous devons interroger la base de données à l'aide d'images bidimensionnelles, alors nous pourrons synthétiser à partir des modèles tridimensionnels de nouvelles images bidimensionnelles avec la même pose et le même éclairage.
- Advenant que les images dans la base de données et celles employées pour les interrogations soient toutes bidimensionnelles et que celles employées pour les interrogations soient de mauvaise qualité, nous devrons nous fier aux meilleurs algorithmes existants, en les affinant et en les optimisant autant que possible. À cet égard, ne sous-estimons pas les techniques de prétraitement. Un alignement minutieux au niveau des sous-pixels donnerait de meilleurs résultats au stade de la reconnaissance. En outre, la détection des visages au teint très foncé reste un domaine trop peu exploré.
- Bien qu'habituellement on ait besoin d'une identification formelle fiable, certaines applications peuvent au contraire profiter des rejets négatifs fiables. Si un algorithme de détection et de reconnaissance peut rejeter toutes les images qui soit ne contiennent pas de visages soit ne contiennent que des visages sans aucun rapport avec celui qu'on

recherche, alors il ne reste plus pour les opérateurs humains qu'à vérifier toutes les images qui restent à la recherche de correspondances.

Importance: Le présent compte rendu technologique permet de bien comprendre l'état actuel des technologies de reconnaissance faciale quant aux images à basse résolution ou de mauvaise qualité. Il sera particulièrement utile aux concepteurs des systèmes pour le domaine militaire, puisque ceux-ci servent surtout sur le terrain, dans des conditions extrêmement variées et imprévisibles.

Perspectives: Notre groupe de recherche collabore depuis un certain temps avec RDDC en recherche-développement de systèmes de reconnaissance faciale, et il lui tarde beaucoup de participer aux projets futurs de RDDC dans ce domaine stratégique. Son expérience de la recherche en association de visages tridimensionnels et son expertise indiscutable en graphiques tridimensionnels en font un partenaire idéal pour créer la prochaine génération de systèmes de reconnaissance faciale 3D seulement ou hybrides 3D/2D.

Table of contents

This page intentionally left blank.

1 Introduction

Face recognition is an interesting research subject that has attracted the attention of many in the fields of pattern recognition and computer vision in the last few decades. Since the events of September 11th, 2001, research efforts in this area have intensified significantly, with much progress being made in the last decade in the essential areas of face detection and recognition algorithms. However, the development of face recognition systems that are truly useful and reliable in different and practical applications has been slow, mainly due to the limitations of existing algorithms. Most of the attention in this field, including that of DRDC, has been on recognizing a frontal face in which both eyes and the mouth are in sight. New challenges are raised in intelligence operations in which face images are captured in varying and unpredictable conditions. The resulting images usually contain many defects caused by poor or changing lighting, bad pose or camera angles that cause face occlusion and geometry changes. Worse yet, at a long distance, face images are sometimes of low resolution and contain high levels of noise. Low resolution and/or poor quality face images are a common occurrence in surveillance videos. Face images posted on the Internet are often of low resolution as well. Before designing new algorithms to deal with such situations, it is necessary to review and evaluate current face detection and facial recognition technologies. The objective of this technical report then is to review relevant research on facial recognition involving low resolution and/or poor quality images. Because low resolution and poor quality face images are so frequently encountered in video sequences, particularly in security surveillance videos, this report will pay particular attention to recent developments in video-based face detection and recognition techniques.

Smart surveillance systems may perform face detection, tracking, recognition and/or behavior interpretation. A video array can used to build a high level “intelligent environment” that attempts to automatically capture and develop an awareness of the events and activities taking place in a certain space [28]. This awareness may allow a machine to understand where a person is, what a person is doing, when an event happens and, most desirable of all, who is the person in question. Some systems (e.g. [17]) may not perform face recognition automatically when it is thought to be too difficult, but will instead create records of keyframes containing subjects of interest. This can later enable users to manually search the surveillance video more efficiently and effectively.

Image resolution is an important factor that affects the performance of face recognition systems. As the viewing distance between a capture device and face increases, the size of the face region in an image decreases. In other words, the face resolution decreases, causing degradation and loss of detail. Low resolution face images occur regularly in video surveillance security applications and content-based multimedia Web retrieval. The recognition performance of the current grayscale-based face recognition techniques can significantly drop when the resolution of facial images is below a certain level (e.g. less than 20x20pixels according to [10]).

General questions that should be asked include, but are not limited to:

- What is the minimum resolution of face images needed for detection and/or recognition?
- How does resolution affect the performance of face detection and/or recognition?

In the most comprehensive survey of face recognition techniques [73], 15x15 pixels is determined to be the lowest resolution of face images needed for reliable face detection and recognition.

It is also important to understand the role that different facial features play in human judgment of identities, i.e., the relative contributions of the internal (eyes, nose and mouth) and external (jaw-line, ear and hair) features, particularly at different distances (or, in other words, at different resolutions). The authors of [12] bring up several interesting questions:

- How does the relative significance of internal and external features change when stimuli are degraded to mimic real-world viewing conditions?
- At low resolutions, does feature saliency become proportional to feature size, favoring more global, external features?
- Are we still better at identifying familiar faces using internal features like eyes, nose and mouth?
- Even if we still prefer internal features, does additional information from external features help to facilitate recognition?
- What kind of a cue-fusion strategy, linear or non-linear, does the visual system use for combining information from these two sets of features?

In recent years, considerable efforts have been directed at using facial color information to improve recognition performance. The question is therefore how this color information can be effectively utilized. Recent work has specifically attempted to address the following issues:

- Is color information able to improve recognition performance and to what extent?
- Which color space is optimal in terms of providing the best discriminative power?
- How are different spectral channels combined to take advantage of all available color information?

Face images are often small (i.e. low resolution) in video sequences. On the other hand, video sequences contain many dynamic features that are not available in high resolution still face images. To effectively utilize the dynamic features unique to video, recent work focuses on, but is not limited to, the following subjects:

- Selecting facial features that are robust to noise in video.
- Selecting facial features that are invariant to changes in illumination conditions.
- Feature tracking algorithms.
- Reconstruction of 3D facial models from face motion.
- Synthesizing new face images, based on 3D facial models, to match the pose and illumination conditions of query images.

- Small face detection facilitated by other information, such as gait.

In the next four sections, we review a large number of scientific publications that address some of the questions above, albeit at different depths. It should be mentioned that this review focuses on the major topics and representative examples to date, so as to provide the reader to with an understanding of the primary issues and current status of this field. It is not intended to be an exhaustive survey nor an inclusive list of all published literature in the last decade. The contents are organized into four sections. Section 2 reviews preprocessing techniques that strive to improve the performance of face detection and recognition. Section 3 looks at new developments in face detection algorithms. Section 4 moves on to consider new face recognition algorithms developed in the past decade. Section 5 then takes a look at face tracking algorithms. Finally, Section 6 provides some concluding remarks.

2 Preprocessing

Before applying any face detection or recognition algorithm, preprocessing of some form is often needed. Most such methods focus on reducing variations in facial appearance due to changing illumination conditions.

A gamma intensity correction adjusts the overall brightness variation of an input image to match a predefined, canonically illuminated image using a nonlinear gray-level transformation [3]. It enhances the local dynamic range of a face in dark or shadowed regions, while compressing it in brighter regions. Shading variation of smooth facial regions is primarily determined by the 3D shapes of the regions and the illumination direction. By reducing the shading variation, the illumination effect is also reduced. A Difference of Gaussian (DoG) filter, a popular bandpass filter in image processing, can be used for this purpose [40]. Histogram matching (also called histogram specification) is a gray-level transformation technique that converts an input image to a new one with a histogram that matches a desired Probability Density Function (PDF). This technique is used locally in [59] to transform gray-levels to a prescribed distribution. A technique called local normal distribution can further transform the gray-level distribution of an image to a new one with a normalized standard deviation [59]. The work in [48] identifies poor illumination in face images and videos as a low signal-to-noise ratio (SNR) problem. In this paper, they introduce a local enhancement method to brighten darker portions of the image, while leaving those of already adequate brightness unaltered, leading to a balanced distribution of pixel intensities.

It should be noted that image intensity enhancement/normalization techniques, such as the Gamma intensity correction and histogram-based techniques mentioned above, primarily enhance image contrast and balance overall image brightness. This improvement is beneficial to human observers that manually inspect these images, but is not necessarily beneficial to computer algorithms performing face detection recognition. On the other hand, if shading and/or shadow effects are reduced, maybe using a DoG filter or some other local intensity correction method, illumination variation will be reduced, hence improving the performance of face detection and recognition algorithms.

The system of [59] contains many of the above preprocessing procedures. The authors claim to be able to reduce the effects of illumination variations, local shadowing and highlights, while preserving elements essential for the detection of visual appearances, leading to significant improvements in their experiments.

Noise and insufficient lighting are also common concerns, particularly with surveillance video images. Since noise is mostly random, it will not remain static at any one pixel over a sequence of video frames. Therefore, noise can be detected by examining the differences of adjacent video frames. The method proposed in [2] operates continuously over a video sequence, recording the differences between adjacent frames. The 8x8 blocks in this difference frame are grouped according to the similarities of their mean values. Under the assumption of uniform noise in a frame, the largest group of blocks will have the most telling information of any noise that occurs. A change mask to clean up the effects of the noise can later be constructed. The difference frame mentioned is also used to properly adjust illumination. A histogram of the individual pixel differences is formed. It is then smoothed using a Gaussian operator. The maximum peak in this smoothed histogram provides useful information about the general brightness of the image.

Illumination compensation is another subject studied by many researchers. One situation is the uneven illumination across an object surface. The parts of the surface that are closer to the light source appear brighter than the parts that are farther away, with a gradual drop off in between. The authors of [28] have developed a method that computes an intensity plane fitted to the intensity grade of the original face image using least squares. Subtracting the face image from this intensity plane produces an equalized face image.

More sophisticated illumination compensation methods consider complications arising from the effects of pose and illumination. To address this concern, [50], for example, proposes to estimate the lighting conditions of a given face in arbitrary poses using any recorded video frame. When the illumination conditions of a given query image do not match those of the database, a new image of the queried face, under the appropriate lighting conditions, is generated using spherical harmonics. This system also strives to remove the effects of occluding objects, by arbitrarily synthesizing future frames in a sequence, using past frames, and comparing the newly-created frames with the actual ensuing frames. Pronounced differences between the actual and synthetic frames suggest the presence of an obstructing object. Working with a test database of 32 individuals, this research purports to be quite successful.

When query faces are of lower resolutions than those used for training, recognition performance decreases. If multiple face images of roughly the same pose, but not at exactly the same location, are available, it is possible to construct a higher resolution face image from them. These so called super-resolution algorithms attempt to enhance query images. The survey [45] provides an overview of super-resolution image reconstruction algorithms. The basic premise of super-resolution is the assumption that low resolution images are sub-sampled (aliased) as well as shifted, with sub-pixel precision of the same scene. In other words, they must contain different "information" of the scene at a sub-pixel level. Super-resolution algorithms often involve estimation of motion (or shifts), with sub-pixel accuracy between different low resolution images. Some super-resolution algorithms also employ interpolation techniques for "magnifying" aliased, low resolution images, in addition to well-established image restoration techniques for de-blurring and noise reduction. It is also observed that super-resolution algorithms may introduce distortions and become more vulnerable to environmental variations [26]. Furthermore, if the resolution of a query image is very low, super-resolution images alone do not necessarily improve the recognition rate. An example of this is seen in [74]. The motion of a human body is estimated in order to compute the pixel displacement between video frames. This is further used for the alignment of side-view face images from multiple frames. The enhanced side-view face image is thought to improve the performance of both face detection and recognition, according to the authors of this paper.

Depending on face image quality in particular application conditions, suitable pre-processing procedures are strongly recommended before the execution of face detection and recognition algorithms.

3 Face Detection

Face detection searches for all possible faces in an image. They may be of different orientations and scales, and even of a completely different appearance due to alternate illumination conditions. Two surveys [25] and [70] from early 2000 provide great overviews of the face detection techniques that were available before and up to that time. Since then, there have been new developments in this area, including not only changes and extensions to existing techniques, but also new ideas involving temporal and color dimensions.

Face detection algorithms are generally based on the following facial properties:

- Skin color,
- Facial components, such as the shapes of the eyes, nose and mouth,
- Configuration of facial components, and
- Face outline shape.

The paper [28] gives a very broad classification of face detection techniques, organizing them as view-based methods and feature-based methods. View-based methods generally use component analysis [4], [11], [18], [60], wavelet [63] and statistical approaches [52]. Feature-based methods use edges [29], shapes [37], [71], grey-level and color information [5], [68], [69] and motion [23], [41]. In the case of video sequences, temporal continuity of faces between adjacent frames can be utilized to enhance face detection accuracy [5], [43], [69]. The multiple primitive video-based closed-loop face detection and tracking method [28] combines several features (such as skin color and elliptical edges) to improve the robustness and speed of face detection in challenging conditions, such as changing illumination, cluttered backgrounds, occlusion, etc.

Significant performance improvements have been made in face detection [27], [49], [63], [64]. However, these methods require thousands of samples to train a classifier and often have a slow training time. They still suffer from problems related to changes in face pose and lighting conditions. For example, the popular Haar wavelet features proposed in [63] show excellent performance for frontal face detection, but are not applicable to side-view face detection [64].

There are a number of variations of the popular AdaBoost detection algorithm [63]. Rather than using Haar wavelets, Gabor wavelets are used for feature extraction in [59]. They are chosen because of two desirable characteristics, namely spatial locality and orientation selectivity. Boost algorithm is used to reduce the redundancies of a high dimensional feature space and computational cost. The original boost algorithm also comes in a variety of forms. For example, RealAdaboost [66] is a generalization of the basic Adaboost algorithm. GentleAdaboost [20] is a robust and stable version of RealAdaboost, performing slightly better on regular data and considerably better on noisy data. It is also much more resistant to outliers. ModestAdaboost [62] is aimed at better generalization capability and resistance to certain specific set of training data. The work in [36] applies extended Haar-like features to construct a space of weak classifiers and to choose ModestAdaboost algorithm to train the strong classifiers which form the cascaded multi-layer ear detector.

Because a face can appear in an image over a large range of scales, many face detection algorithms use a pyramid of downscaled images obtained from the original input image. The Optimal Adaptive Correlation (OAC) technique is proposed for quickly locating candidate face regions, without having to carry out face detection at multiple scales [59]. The suggested algorithm is adaptive to input images that have been normalized by a preprocessing procedure and that have similar power spectra. The idea is to transform the normalized image into a correlation image in a Hilbert space, with normalized values ranging from zero to one. A segmentation process, based on the correlation values, can partition an image into two segments in only a single iteration, namely face candidate regions and background areas.

Skin color is often used for fast detection of potential face regions. For example, in [15], as images of people walking through a monitored entrance are snapped, they are scanned for pixels that might correspond to skin. However, skin color alone is not sufficient for reliable face detection as other parts of the human body are also of these colors. Furthermore, other objects in a scene may also be of colors that are very close to skin tones. Moreover, different races of humans have different skin colors. And even within a single race, there is great diversity of skin tone. Therefore, any region suspected of being a face, based on its color, should be further examined by other algorithms.

The outline of a face is typically elliptical in shape and is used by many researchers as a way of performing face detection. An important advantage of this approach is its robustness to changes in illumination and background. For example, [31] presents a shape comparison approach that achieves fast and accurate face detection. The proposed method is based on edges and works with still grayscale images. The Hausdorff distance is used as a similarity measure between a general face model and possible instances of the object within the image. This paper describes an efficient implementation, making this approach suitable for real-time applications. A two-step process that allows both coarse detection and exact localization of faces is presented. However, elliptical outline detection alone can still be computationally intensive and is vulnerable to highly cluttered backgrounds, as argued in [28]. This group combines the advantages of both skin color and elliptical edge features to overcome the shortcomings of the two methods. Specifically, skin color is used to efficiently locate possible face regions, while, at the same time, elliptical shape matching eliminates non-face regions. The use of skin color also avoids time-consuming window scans found in many other face detection algorithms. One challenging issue in face detection, under-addressed so far in the existing literature, is the detection of faces with dark skin. The test images of [59] include some faces of fairly dark skin tones. The successful detection of these faces suggests that proper preprocessing techniques improve the performance of face detection when it comes to dark skin faces.

The authors of the paper [38] argue that features should resemble face semantics, matching local or global face structures. [44] demonstrated that local adaptive features have more discriminative power than global features, such as those obtained by Principal Component Analysis (PCA) or Fisher Linear Discriminant Analysis (FLDA). Based on this observation, the paper [17] proposes a two stage technique that learns adaptive and general features for face detection. In the first stage, it learns adaptive features that adjust to each particular sample in the training set using a non-linear optimization method focused on feature parameters (such as position, orientation and scale). In the second stage, Adaboost feature selection is applied to the pool of adaptive features in order to obtain general features.

Currently, there does not appear to be any single face detection algorithm that is effective in all applications. When one technique fails, it is natural to try different techniques. The paper [8] presents such a system for face detection and tracking in surveillance applications. It is made up of the following sequence of operations:

- *Frontal face detector:* A Viola–Jones-based face detector [63] is initially applied in the search window and when the tracker losses the track of the eyes.

- *Local context face detector:* If the previous technique fails, a local context-based face detector is applied in the search area.

- *Skin color:* If previous cues fail, the modelled skin color is used to locate the face in the search area. If a proper blob is located, eyes are next located.

- *Face tracking:* If everything else fails, a pre-recorded face pattern is tracked in the search area. However, the system must not rely on tracking for more than a few consecutive frames as this will lead to tracking problems. Instead, the other cues need to confirm, from time to time, a human presence. Otherwise, a person will be considered to be lost.

In many surveillance video images, a face region is too small to be detected by any of the methods described above. Gait information thus becomes useful in detecting human figures in video frames, particularly if they move in the scene. The paper [74] describes a method to extract side-view face images at a distance from a video. When a person walks through a scene, a simple static background subtraction is used for human-body segmentation. The head and other body parts can be separated using a predefined proportion ratio. Because the face region is very small in a video taken at a distance, multiple low resolution face images can be fused to construct a single high resolution image.

Pattern matching is one of the primary tasks in content-based image retrieval. A new technique in this area, known as local self-similarity matching [56], has proven to be a very robust pattern matching method that has the potential to become an alternative to existing face detection methods. It is based on the observation that correlation crossing images reveal much about the internal layout of local self-similarities, even though the patterns generating those local self-similarities are quite different in each of the images. Given a small (e.g. 5x5) window centered on a pixel, it is correlated with local sub-regions to produce a local correlation surface which is in turn quantized into a local "self-similarity descriptor" [56]. Matching between two objects can be done by measuring the distance between two sets of local self-similarity descriptors computed from each image, respectively. Using a similar approach, a different local descriptor, called a "local steering kernel", is defined in [53] to capture the local structure of images by using covariance matrices and principle component analysis based on estimated gradients. A "resemblance map" is also computed to represent the likelihood of similarity between two images at each pixel position. The strength of these methods comes from their ability to match complex visual data, detect objects in cluttered images using only rough hand-drawn sketches, handle textured objects with no clear boundaries and detect complex actions in cluttered video data with no prior learning.

In summary, state-of-the-art face detection algorithms performed reasonably well in most situations, particularly for frontal or near frontal face images. Further research is required for detection of faces with side or near side views, and detection of faces with dark skin colors.

4 Face Recognition

General still image face recognition methods include PCA [60], Linear Discriminant Analysis (LDA) [4], [13] and Elastic Graph Matching (EGM) [34]. In the past decade, many new discoveries have improved the performance of these traditional methods. Most of them are variations or extensions of the major models. Some integrate multiple models. Fortunately, we also see some new work taking place.

The authors of [65] show that color-based features are less susceptible to variations in resolution than other features derived from grayscale images, and are thus more robust for object recognition. It is also shown in a psychophysical experiment involving face recognition performance within the context of the human visual systems that the contribution of facial color is important in cases involving degraded, low resolution facial images [72]. The work of [10] defines a new metric, called Variation Ratio Gain (VRG), and demonstrates that, for low resolution facial images, facial color cue can (significantly) improve recognition performance when compared with grayscale-only techniques. A total of 3192 facial images from 341 subjects were collected from three public data sets, specifically CMU PIE, Color FERET and XM2VTSDB. For each face image, six different resolutions of 112x112, 86x86, 44x44, 25x25, 20x20 and 15x15 (pixels) were generated. Three representative facial recognition methods, namely PCA, Fisher's LDA (FLDA) and a Bayesian model, were used for the experiment. The improvement resulting from the addition of color information is significant. For example, in the case of FLDA, color features boosted performance by 24.86% and 50.81%, respectively, among two 15x15-sized images over their grayscale counterparts.

Variations of the traditional methods have been seen in many recent works. In one example [28], the covariance matrix used for generating eigenvectors in the standard Eigenface PCA method has been replaced by a correlation matrix. As well, the projection vector of a test face image on the eigenvector basis is normalized. It is said that the result is less affected by illumination changes for two reasons: (1) the norm of a projection vector in the eigenvector subspace is proportional to the intensity of the face image; (2) the intensity change in face images due to illumination change can be normalized. The work of [36] uses the PC-ICA technique, a generalization of principal component analysis. It first reduces the dimensionality of the data using principal component analysis. It then adopts a ratio-factor-based ICA (Independent Component Analysis) for face recognition. ICA is a method of finding underlying factors or components in multivariate statistical data. The Hausdorff distance is used to calculate the minimum distance between the test image and the image in the database to be recognized.

When query faces are of lower resolutions than those used for training, face recognition performance decreases. Several approaches have been proposed to mitigate this problem. A relatively simple idea is to downsample the training set in order to match the lower resolution of the query images. It is obvious that this will cause a loss of detailed facial features among the images of the original training set. Another idea in the opposite direction makes use of the super-resolution algorithms that attempt to enhance query images. It is known that super-resolution algorithms will often introduce distortions and are vulnerable to environmental variations [26]. Furthermore, if the resolution of a query image is very low, super-resolution images alone will not necessarily improve recognition performance. In the paper cited above, the authors suggest using base face features extracted from low resolution face images downsampled from a high resolution training set. Matching is performed simultaneously on both super-resolution images and base face features. The authors demonstrate improved performance using multiple images from two different types of sources, specifically multiple frames of a video sequence and multiple images obtained from several cameras.

The DCT is a basic transformation used in several image and video compression methods. This representation has also been borrowed by some researchers for face recognition. In [16] for example, an image is divided also into 8x8 blocks, similar to JPEG image encoding. The DCT of each block is calculated and then used to create a feature vector representing that region. All such vectors are concatenated to form a global descriptor of a face. This is admittedly a fairly straightforward and standard means of performing recognition. The first application is one of monitoring individuals as they pass through a doorway in a smart home environment. As images are snapped of folks walking through an entrance, they are scanned for pixels that might correspond to skin. Haar features are also used to discover a person's eyes. Information about the position of the eyes is later used to normalize a face. Both k-means clustering and Gaussian Mixed Models (GMM) enable the system to combine the various face images obtained over a sequence of frames and to determine which are of poor quality. A weighting scheme penalizes these poor quality images that might degrade the performance of the system. The proposed method is found, very interestingly as reported by the authors, to be empirically superior to earlier techniques, such as LDA and PCA.

It is also important to understand the role that different facial features play human judgments of identities, i.e., the relative contributions of the internal (eyes, nose and mouth) and external (jaw-line, ear and hair) features, particularly at different distances. The paper [19] proposes a "feature hierarchy" to characterize the role of internal and external features in face recognition. In this hierarchy, the head outline is the most significant, followed by the eyes, mouth and lastly the nose. Another study, given in [7], further suggests that when recognizing unfamiliar faces, external features are more important than internal features. Several studies (e.g. [12]) have found that familiar faces are recognized better from their internal rather than external features. The paper [30] characterizes these relative contributions as a function of image resolution. Their experiment confirms the finding of previous studies that internal features of familiar faces are more useful for recognition than are external features at high resolution. The importance of these two feature-sets reverses as resolution decreases.

Varying pose is one of the most challenging issues in face recognition. In a face database, the number of different poses for each individual is finite. If only a single query image is provided, the recognition rate is usually poor. On the other hand, a video sequence may contain the face of a person from many different viewing angles, thus possibly increasing the recognition rate. The paper [35] introduces the probabilistic appearance manifolds approach to video-based face recognition. Each registered person is represented by a low-dimensional appearance manifold in the ambient image space. The complex nonlinear appearance manifold is expressed as a collection of subsets (called pose manifolds). It also embodies the connectivity among them. Each pose manifold is approximated by an affine plane. A maximum a posteriori formulation is presented for face recognition in test video sequences by integrating the likelihood that an input image comes from a particular pose manifold and the transition probability to this pose from the previous frame. A weight mask is introduced to deal with faces that are partially occluded. It is demonstrated, using temporal voting schemes, that this algorithm outperforms existing frame-based face recognition methods. To construct this representation, exemplars are sampled from videos. These exemplars are clustered by way of a k-means clustering algorithm. Each cluster is represented as a plane, computed using PCA. The connectivity between the pose manifolds encodes the transition possibility between images in each of the pose manifold and is learned from training video sequences.

Instead of constructing a pose manifold as discussed above, the work in [47] aims to properly separate and classify faces by forming a tree structure reflecting the various poses. After being normalized and transformed via a DCT to remove lighting effects, the features of an image are obtained using LDA. From here, a face is classified, according to pose, using either a nearest-neighbor technique or a trained Bayesian probabilistic system. Images are then placed into a tree structure, organized according to pose. This organization may be, for instance, one of frontal, left and right side-views, extreme left and right side-views, etc. Subclasses enable further refinement. Using an array of metrics, the system is tested on low-resolution video samples. The authors deem the technology to be effective in such instances.

The paper [51] evaluates recent approaches to the recognition of faces at relatively large pose angles using a gallery of frontal images. The authors also propose novel modifications of their own. They compare the accuracy, robustness and speed of an Active Appearance Model (AAM) method (where realistic frontal faces are synthesized from non-frontal probe faces) against bag-of-features methods (which are local feature approaches based on the block DCT and GMMs). This paper explains an interesting approach in which the AAM-based technique is sped up by directly obtaining pose-robust features, allowing the computationally expensive and artefact-producing image synthesis step to be deleted. Furthermore, it adapts a histogram-based bag-of-features technique to the face classification problem. The properties of this new technique are compared to those of a previously proposed direct bag-of-features method.

Pose and varying illumination are not a major concern when a face database of 3D models of each person involved is available. The primary benefit of such 3D models is that they can be used for recognition across a wide variety of viewing angles. It is also possible to re-render face images under any illumination condition. As [46] does, one can rotate a model and apply lighting as needed to produce pictures matching those in a probe image. Here the 3D model set is manually created using a laser scanner and 100 human subjects. Images in alternating poses and under different lighting conditions are synthetically generated. They are fed to a Support Vector Machine (SVM) as training data. As test images arrive, their pose and illumination conditions are determined. An appropriate probe sequence is subsequently put together. It is made up of images, derived from 3D models, under the prescribed illumination conditions and the proper pose. LDA performs the final comparison. As 3D vision hardware becomes more (economically) accessible to more practitioners, further research and development will occur in this direction. At the moment, most current research is still purely two-dimensional.

When 3D models are not available, an image sequence obtained from a video is the next best hope of reconstructing a 3D model of a human face. The foundations of this approach are the structure from motion (SFM) algorithms developed in computer vision in the 1980s. However, given the poor quality of input videos in many video surveillance applications, the quality of a reconstructed 3D face model is often unsatisfactory [50]. Instead of reconstructing 3D face models from scratch, a generic face model can serve as the initial starting point. This model is based on the assumption that the structure of most faces is similar [21], [55]. While this method does improve the convergence rate of a reconstruction algorithm, it often, as discovered in [50], converges to a solution very close to the initial value, rather than that of a particular face in a given video. Therefore, the authors propose to use an SFM algorithm to first obtain a reasonably good 3D estimate from a video sequence, and then to use a generic model to correct errors in the estimate by comparing local regions in the two models using an energy function. In order to obtain a reasonably good 3D estimate, the quality of the input video is statistically assessed and incorporated into the algorithm. Another approach applicable to still face images is presented in [14]. Key features on a person's face, 11 in total, are first detected. A base mesh is then adjusted to match the characteristics of these features. Finally, an image and a depth map are generated, enabling recognition.

The paper [6] also describes a fully automated algorithm for reconstructing a 3D model of a human face given only a single image. Two overarching steps allow this to take place, namely that of detecting the face, followed by the subsequent 3D reconstruction process. The recognition stage is carried out using a classic SVM and boosting detector. Faces are next classified, into seven classes, according to their degree of rotation. This is done using a regression-style estimator. The algorithm then proceeds to find the eyes and mouth corners. As this process plays out, individual pixels are marked as being either a part of these features or simply a background pixel. An SVM is called on to assist in this job. A Bayes classifier is also used to help resolve these features. The detected eyes, mouth and nose tip are used in the 3D reconstruction. The actual reconstruction relies on a 3D morphable model, similar to the generic mesh model mentioned above. Knowledge of the location of the eyes, nose and mouth provides a starting point. A globally-optimized polygon mesh is created from here. The algorithm is found to be successful in 705 out of 870 images tested. The total time needed to construct a single face is less than 4 minutes. Human-test users were asked to compare the actual images of the individuals and the matching reconstructed faces.

The work of [57] chooses to incorporate infrared images in the face recognition process, because they provide identifying information in the form of blood flow and heat patterns. This information is fused with that of images taken in other spectra, including visible light. First, a given face image, in any particular spectrum, is decomposed into a series of amplitude and face features. This information is found in the frequency domain, thereby requiring that an image be transformed via the Fourier transform (FT). Once convolved with the Fourier transform of a Gabor wavelet, the image is returned to the spatial domain. Note that the phase information is not affected by illumination and shading, key concerns in face recognition. Before an actual comparison can take place, the authors propose to fuse the amplitude and frequency information, producing a single feature vector that describes a given face. This fusion may be performed at either the feature level or the image level. To simplify the fusion procedure, a trained SVM is employed. The model is designed in such a manner that will minimize the likelihood of making an error in the resulting recognition and classification. Alas, one obtains a feature vector that may be used to identify an individual picture. The technique described above is readily extended to include many different images, each from different spectra. Each image yields a single feature vector, all of which can then be pooled using a prescribed method involving wavelets. Testing is done over a database made up of long-, medium- and short-wave infrared images, as well, of course, as those in the visible segments of the spectrum. The method is deemed to be superior to that of traditional PCA or LDA. Moreover, the long- and medium-wave images are found to be invariant to changing illumination conditions. In the end, it is determined that the best combination is one that merges short-wave infrared images with a visible-light picture.

An early work [24], but one that is still significant when it comes to designing new face recognition algorithms today, is mentioned here. This paper is a comprehensive performance evaluation of face recognition algorithms (including Eigenfaces, Fisherfaces and FaceIt) using a wide range of images in different viewing and illumination conditions, with varying facial expressions and time delays between the acquisition of gallery and probe images and those affected by occlusion. Image data consists of over 37,000 images from 3 publicly available databases that systematically vary in multiple ways individually and in combination. The main findings of this report are: (1) pose variation beyond a 30° head rotation substantially decrease recognition rate, (2) time delay: pictures taken on different days, but under the same pose and lighting condition, produce a consistent reduction in recognition rate, (3) with some notable exceptions, algorithms are robust to variation in facial expression, but not to occlusion. They also found small, but significant differences, related to gender.

5 Face Tracking

Tracking objects, in general, can be a complex issue due to the following reasons uncovered in a comprehensive review [1] (note that these are not the only reasons for difficulties in this area):

- Loss of information caused by the projection of a 3D world onto a 2D image,
- Noise in images,
- Complex object motion,
- Non-rigid or articulated nature of objects,
- Partial and full object occlusions,
- Complex object shapes,
- Scene illumination changes, and
- Real-time processing requirements.

Selecting the right features plays a critical role in tracking. In general, the most desirable property of a visual feature is its uniqueness which ensures that objects can be easily distinguished in the feature space. Feature selection is closely related to object representation. For example, color is used as a feature in histogram-based appearance representations, while in contour-based representation, object edges are usually used as features. In general, many tracking algorithms use a combination of these features. The paper [1] reviews common visual features, including color, edges, optical flow and texture.

Color: In image processing, the RGB (Red, Green, Blue) color space is generally used. However, the RGB space is not a perceptually uniform color space, that is, the differences between colors in the RGB space do not correspond to the differences perceived by humans. Furthermore, the RGB dimensions are highly correlated. Conversely, L*u*v* and L*a*b* are perceptually uniform color spaces, while HSV (Hue, Saturation, Value) is an approximately uniform color space.

Edges: Strong changes in intensity usually occur at object boundaries in images. Edge detection is used to identify these changes. An important property of edges is that they are less sensitive to illumination changes as compared to color features. Algorithms that track the boundaries of objects usually use edges as the representative feature. Because of its simplicity and accuracy, the most popular edge detection operator is the Canny edge detector.

Optical Flow: Optical flow is commonly used in computer vision as a feature in motion-based segmentation and tracking. It is a dense field of motion (or displacement) vectors that represents the directions and speeds of motion for each pixel in a frame within a video sequence. Assuming the brightness constancy of corresponding pixels in consecutive frames, the motion vector at each pixel is usually computed by searching corresponding matching points in adjacent frames.

Texture: Texture is essentially a pattern of the intensity/color variations. It is less sensitive to illumination conditions and more reliable for tracking. Computation of texture is however much more complicated and time consuming than that of color and edges.

Face detection can be performed on every frame of a video sequence using the techniques reviewed in Section 2 of this report. However, this detection can fail for many different reasons. In order to keep on target with a face, correlation may be used between adjacent frames, one of which contains the detected face [17]. The correlation tracker may detect a window of similar position and size to the current tracking window. This new window is then used for further tracking. Static background subtraction can be used to prune areas that do not contain motion. Also, human faces generally appear at a certain range of sizes in most surveillance scenarios. Therefore, this assumption can also be used to limit the search window size.

While numerous face tracking and recognition tools track Haar-like features, some (e.g. [58]) instead focus on particles. The hope is to both track and recognize faces using this method in low quality video images. This idea is said to produce faster run-times than traditional tracking algorithms. Particles are distributed across a face. More specifically, they are found in patches over a face. The DCT of a patch in a probe face is compared with the DCT of potential patches in a template image, allowing a system to track the movement of regions and particles in faces across different video frames. The actual comparison of these patches is usually based on a mean-absolute or root mean-squared distance metric. Particles in the middle portions of the face are said to be the most important. It is said that one needs approximately 50 particles to properly track and identify a person.

In real-world applications, surveillance videos are often noisy. Even face tracking algorithms that adaptively build target models to reflect changes in appearance between frames often suffer from drift, a gradual adaptation of the tracker to non-targets. The work of [32] addresses this problem by introducing visual constraints, specifically a combination of generative and discriminative models in a particle filtering framework. The generative term conforms the particles to the space of generic face poses while the discriminative one ensures the rejection of poorly aligned targets. This leads to a tracker that significantly improves robustness against abrupt appearance changes and occlusions, critical for the subsequent recognition phase. Identity of the tracked subject is established by fusing pose-discriminant and person-discriminant features over the duration of a video sequence.

When a face is too small in a video image, information gathered from various side poses of human faces, along with knowledge of one's gait, are pooled in an attempt [74] to recognize a human subject at a distance. A series of gait images, individually normalized, as well as normalized face images, make up the data in this scenario. As in other cases, each of these data sets is reduced using PCA and multiple discriminant analysis (MDA). Once combined, the two data sets enable one to achieve a classification accuracy of 91.3%.

Searching for a specific person or event through a large amount of recorded surveillance video is a common need. Since current face detection and recognition techniques are not sufficient to perform such tasks automatically, manual browsing by human operators is still a common practice. Constructing an index structure using a list of keyframes can improve the efficiency of manual browsing. Ideally, only one keyframe is needed for each person appearing in a video sequence and the selected keyframe should be of maximum resolution. The paper [17] is an example in this direction that stores a keyframe for each captured face image (not the full frame) in a database with an associated timestamp.

A variety of subspace features have been studied by different researchers. For example, the mutual subspace method finds the principal axes of the subspace of face images in each given video sequence and compares the principal axes to those of known classes using an inner product [67]. Another method models the distribution of face sequences in the facial feature space and classifies distributions of identities by Kullback-Leibler divergence [3], [54]. PCA subspace feature analysis and hidden Markov time sequence modeling are combined in [28]. In terms of the subspace features, the facial feature distribution of a certain pose would be scattered by perturbations such as illumination changes, misalignments and noise, as well as face turns. These dynamics and scattering can be captured by an HMM with GMMs. The HMM states would represent mainly different face poses with some perturbations. The local histograms of wavelet coefficients written with respect to a coordinate frame fixed to the object are used for face detection in a video sequence [43].

A framework for face recognition in video is proposed in [9] using a probabilistic approach. Bayesian approaches to face recognition, when applied to video, utilize a time series state model to characterize the evolving dynamics and/or identity in a probe video. Given a state vector at time t, x_t, and the observations up to the time t, $y_{0:t}$, the goal is to compute the posterior distribution of the state vector, $\pi(x_t) = p_t(x_t \mid y_{0:t})$. If n_t is a human identity variable at time t, the posterior distribution of the identity can be estimated by $\pi(n_t) = p_t(n_t \mid y_{0:t})$. In the state vector, x_t is the affine tracking parameter, θ_t and the observation y_t are Gabor-filtered jets [34] defined on a sparse grid which is superimposed on the template face image and undergoes an affine motion and local deformation. The Sequential Importance Sampling (SIS) algorithm [33], [39] is used numerically to obtain an updated set of samples for $\pi(\theta_t)$ for tracking.

Capturing high resolution images is a major priority in many face recognition systems. In [42], the author attempts to track individuals so as to ensure that resulting face images are of the utmost quality. With a little help from OpenCV, an open source computer vision package, and a PTZ (pan, tilt, zoom) camera, the system is able to capture the x, y, z position of a moving face. This information, in addition to the autofocus feature, enables the system to aim the camera directly at the person in the scene. With the camera placed near that person, and in the direction of their face, high-quality face images are captured. Although it nearly functions in real-time, it is unable to track more than one face at any single time.

Surveillance systems are commonly deployed in large public spaces, such as airports or stadiums. There are times in which the environment is a smaller, perhaps closed, setting. The home of a family is one such example. The first step in allowing this sort of monitoring is that of face recognition. As demonstrated in their research [22], this is accomplished using an AAM. Specifically, the group suggests tracking faces using important facial features. A distance metric is needed to distinguish among faces and random objects. Subject faces are compared with those in a trained SVM system. When all such classifiers fail, the individual in question is reported as being unknown. Performance rates are comparatively high. At the same time though the authors admit that the system has only been tested in an environment with five known individuals.

Most surveillance systems attempt to detect and recognize individuals via face recognition. In [61], the objective is one of finding individuals in video sequences given a user-provided description. For instance, one may ask that all videos be scanned for a person that “is short, has black hair, wears glasses and is dressed in a red shirt”. Ideal for tracking suspects and finding missing people, this system takes advantage of features such as hair and eyewear, factors that typically frustrate traditional face recognition technologies. A human subject is first found by a face detection algorithm based on Haar functions. From here, that person is partitioned into a head, torso and legs. Information about the color of the body and legs is stored. The size of the torso and legs are inferred by the dimensions of the face. The head is further subdivided into upper, middle and lower regions corresponding, respectively, with hair and hats, eyes and glasses and beards and moustaches, as well as various other features. All information, as it is gathered, is placed into a common database. Later, a user may submit a query, resulting in a search of the database. Although promising, the system did suffer from a variety of issues, including situations in which shadows were confused with eyewear and hair was misinterpreted as a hat.

6 Conclusion

In the last decade, many improvements have been made in every aspect of face recognition research. The performance of face recognition algorithms has steadily improved. However, these improvements, in our opinion, are evolutionary. True breakthroughs in this challenging area are rare. Experience gained and lessons learned from other areas may help in the future development of face recognition systems.

One example is video compression. Early video compression systems were based on the DCT transformation, plus a set of other coding techniques, such as quantization and entropy coding. Later systems used the wavelet transform improved image quality significantly, especially when compression ratios are high (e.g. above 30dB). Quality improvement is only marginal when compression ratios are lower. In late 1990s, streaming videos at level of quality of DVD over the Internet was nothing more than a dream. Fortunately, it motivated many researchers in this area. In the last decade, improvements have been made in every component of video compression systems, mostly through optimization. These advancements are also evolutionary, with no notable breakthroughs. Streaming video over the Internet has nevertheless become a reality, in not only DVD-quality, but also soon HD-quality. The main factor responsible for this is the significantly increased bandwidth of the Internet. Faster CPU and ever-cheaper memory storage have also contributed to these gains. It is fair to say that modern hardware devices **enabled** real-time video streaming today.

Another example is the 3D graphics technology developed for video game systems. The quality of video game graphics has improved dramatically in the last two decades. Some games are almost photo-realistic. One major feature touted by the manufactures of these products is the number of polygons that can be rendered per second. Why? Most 3D-rendering techniques used in video game machines are still based on the polygon scan-conversion algorithm, a technique developed in 1970s. Enhanced graphics are mainly achieved through finer polygon meshes that better approximate complex real world objects. As a polygon mesh becomes finer though, the number of polygons grows very fast. To handle this rapid increase in the number of polygons, faster and multiple CPUs are used and memory size is expended, but the basic rendering algorithm does not change much, if at all. We see once again that modern hardware devices have **enabled** the state-of-the-art 3D graphics used in video game machines today.

Great leaps in face representation and recognition are always possible, but cannot be expected to occur by any specified date. What can we do and what should we do before this happens? Based on experiences involving video streaming over the Internet and 3D graphics in video game machines, I would make the following concluding remarks:

- The most promising direction is one that captures and stores face images in 3D format in a database, seeing that 3D models offer the best chance at solving the primary issues that affect face recognition, namely changes in pose and illumination conditions. An individual's face images are usually acquired in a controlled environment. The major constraint when capturing and storing 3D face images is the limited availability of 3D cameras. Affordable 3D camera could become available if manufactures perceived there to be a market. Until this happens, efforts will likely continue to focus on algorithms that estimate face orientation and illumination direction using 2D query face images and face synthesizing algorithms using a 3D face model under changing viewing angles and illumination directions.

- If query images can also be acquired in 3D format, then high rate face recognition can be achieved. Research conducted by our own group in early 2000 demonstrated that only six contours extracted from a 3D face model are sufficient to separate different persons, even when the two are identical twins. In this case, research should focus on the design and implementation of efficient 3D matching algorithms.

- Since most current 3D cameras have a limited working range (i.e. the object must be within a predefined distance), real-time acquisition of 3D query images is only suitable in some environments. If the query images are 2D, but the database images are 3D, new images can be synthesized from the 3D model to match the pose and illumination conditions of those of the query image.

- When both the face image in the database and the query image are 2D and the query images are of low quality, we must rely on the best existing algorithms to design application systems. While we should always strive to find new approaches, fine tuning and optimizing existing algorithms continues to be a possibility. In particular, preprocessing techniques should not be under-estimated. During face matching, the effects of improper alignment between a query image and database images at the sub-pixel level on the performance of recognition algorithms should be evaluated. Finer alignment could lead to further performance improvements during the recognition stage. Face detection with very dark skin tones is also an under-studied area. More work in this area is needed.

- While most applications call for reliable positive identification, some applications can actually benefit from reliable negative rejection. For example, in surveillance video monitoring, it is difficult and time consuming for a user to manually browse through pre-recorded video. If a face detection and recognition algorithm can reject all frames that either have no faces in them or that have faces that clearly do not match the query face, human operators need only look at the remaining frames containing possible matches.

References

[1] G. Aggarwal, A. K. R. Chowdhury and R. Chellappa, “A system identification approach for video-based face recognition,” Proc. 17th Int. Conf. Pattern Recognition, vol. 4 pp. 175-178, 2004

[2] T. Alexandropoulos, S. Boutas, V. Loumos and E. Kayafas, “Real-time change detection for surveillance in public transportation”, IEEE Conf. Advanced Video and Signal Based Surveillance, vol. 13, no. 2, pp. 159-171, 2005

[3] O. Arandjelovic and R. Cipolla, “A methodology for rapid illumination-invariant face recognition using image processing filters,” Computer Vision and Image Understanding, vol. 113, issue 2, pp.159-171, 2009

[4] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997

[5] S. Birchfield, “Elliptical head tracking using intensity gradients and color histograms,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 232-237, Santa Barbara, CA, June 1998

[6] P. Breuer, K. Kim, W. Kienzle, B. Scholkopf and Volker Blanz, “Automatic 3D face reconstruction from single images or video”, 8th IEEE Int. Conf. Automatic Face & Gesture Recognition, pp.1-8, Amsterdam, Sept. 2008

[7] V. Bruce, Z. Henderson, K. Greenwood, O. J. B. Hancock, A. M. Burton and P. Miller, “Verification of face identification from images captured on video,” Journal of Experimental Psychology: Applied, vol. 5, pp. 339-360, 1999

[8] M. Castrillon, O. Deniz, C. Guerra and M. Hernandez, ENCARA2: “Real-time detection of multiple faces at different resolutions in video streams”. Journal Visual Communication and Image Representation, vol. 18, no. 2, 2007

[9] R. Chellappa, S. Zhou and B. Li, “Bayesian methods for face recognition from video”, IEEE Int. Con. Acoustics, Speech, and Signal Processing, pp. IV-4068-4071, Orlando, May, 2002

[10] J. Y. Choi, Y. M. Ro and K. N. Plataniotis, “Color face recognition for degraded face images”, IEEE Trans. Systems, Man and Cybernetics, vol. 39, no. 5, 2009

[11] G. Donate, M.S. Bartlett, J.C. Hager, P. Ekman and T.J. Sejnowski, “Classifying facial actions,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 10, pp. 974-989, 1999

[12] H.D. Ellis, J. W. Shepherd and G. M. Davies, “Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition,” Perception, vol. 8, no. 4, pp. 431-439, 1979

[13] K. Etemad and R. Chellappa, “Discriminant analysis for recognition of human face images,” Journal Optical Society of America A, pp. 1724-1733, 1997

[14] H. K. Ekenel, J. Stallkamp, H. Gao, M. Fischer and R. Stiefelhagen, “Face recognition for Smart Interactions”, Proc. IEEE Int. Conf. Multimedia & Expo, pp. 1007-1010, 2007

[15] H. K. Ekenel, J. Stallkamp and R. Stiefelhagen, “A video-based door monitoring system using local appearance-based face models”, Computer Vision and Image Understanding, vol. 114, no. 5, pp. 596-608, 2010

[16] H. K. Ekenel, M. Fischer and R. Stiefelhagen, “Face recognition in smart rooms,” MLMI’s 07 Proc. The 4th Int. Conf. Machine Learning for Multimodal Interaction, Springer, 2007

[17] R. Feris, Y. Tian and A. Hampapur, “Capturing people in surveillance video,” Proc. IEEE CVRP, pp. 1-8, Minneapolis, June, 2007

[18] F. Feuret and G. Geman, “Fast face detection with precise pose estimation,” Proc. 16th Int. Conf. Pattern Recognition, vol. 1, pp. 235-238, 2002

[19] H. Fraser, G. L. Craig and D. M. Parker, “Reaction time measures of feature saliency in schematic faces,” Perception, vol. 19, pp. 661-673, 1990

[20] J. Friedman, T. Hastie and R. Tibshirani, “Additive logistic regression: A statistical view of boosting,” The Annals of Statistics, vol. 28, no. 2, pp. 337-374, April, 2000

[21] P. Fua, “Regularized bundle-adjustment to model heads from image sequences without calibration data,” Int. Journal Computer Vision, vol. 38, no. 2, pp. 153-171, 2000

[22] H. Gao, H. K. Ekenel and R. Stiefelhagen, “Robust open-set face recognition for small-scale convenience applications”, Proc. 32nd DAGM Conf. on Pattern Recognition, pp. 393-402, 2010

[23] H. Graf, E. Cosatto, D. Gibbon, M. Kocheisen and E. Petajan, “Multi-modal system for locating heads and faces,” Proc. 2nd IEEE Int. Conf. Automatic Face & Gesture Recognition, pp. 88-93, 1996

[24] R. Gross, J. Shi and J. Cohn, “Quo vadis face recognition?”, Third Workshop Empirical Evaluation Methods in Computer Vision, pp. 119-132, 2001

[25] E. Hielm and B. K. Low, “Face detection: A survey,” Computer Vision and Image Understanding, vol. 83, no. 3, pp. 236-274, 2001

[26] P. H. Hennings-Yeomans, S. Baker and V. Kumar, “Simultaneous super-resolution and feature extraction for recognition of low-resolution faces," Proc. IEEE CVPR, pp. 1-8, Anchorage, AK, June, 2008

[27] C. Huang, H. Ai, Y. Li and S. Lao, “Vector boosting for rotation invariant multi-view face detection,” Proc. IEEE ICCV’05 vol. 1, pp. 446-453, Beijing, Oct. 2005

[28] K. S. Huang and M. M. Trivedi, “Integrated detection, tracking, and recognition of faces with omnivideo array in intelligent environments”, EURASIP Journal Image and Video Processing, vol. 2008, April, 2008

[29] Jacquin and A. Eleftheriadis, “Automatic location tracking of faces and facial features in video sequences,” Proc. Int. Conf. Automatic face & gesture recognition, pp. 142-147, June 1995

[30] N. Jarudi and P. Sinha, “Recognizing degraded faces: Contribution of internal and external features”, no. 617, 2005

[31] O. Jesorsky, K. J. Kirchberg and R. W. Frischholz, “Robust face detection using the Hausdorff distance,” Proc. Third Int. Conf. Audio- and Video-based Biometric Person Authentication, pp. 90-95, 2001

[32] M. Kim, S. Kumar, V. Pavlovic and H. Rowley, “Face tracking and recognition with visual constraints in real-world videos”, Proc. IEEE Conf. CVPR, pp. 1-8, Anchrorage, AK, June, 2008

[33] G. Kitagawa, “Monte Carlo filter and smoother for non-Gaussian nonlinear state space models,” Journal Computational and Graphical Statistics, vol. 5, pp. 1-25, 1996

[34] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz and W. Konen, “Distortion invariant object recognition in the dynamic link architecture,” IEEE Trans. Computers, vol. 42, pp. 300-311, 1993

[35] K. Lee, J. Ho, M. Yang and D. Kriegman, “Video-Based Face Recognition Using Probabilistic Appearance Manifolds”, Proc. IEEE Conf. CVPR, pp. 313, Madison, June, 2003

[36] Z. Lei, C. Wang, Q. Wang and Y. Huang, “Real-time face detection and recognition for video surveillance applications”, World Congress on Computer Science and Information Engineering, pp. 168-172, Los Angeles, March, 2009

[37] X. Li and N. Roeder, “Face contour extraction from front-view images,” Pattern Recognition, vol. 28, no. 8, pp. 1167-1179, 1995

[38] X. Liu and T. Chen, “Video-based face recognition using adaptive hidden Markov models,” Proc. IEEE Conf. CVPR, pp.340-345, Madison, June, 2003

[39] J. S. Liu and R. Chen, “Sequential Monte Carlo for dynamic systems,” Journal American Statistical Association, vol. 93, pp. 1031-1041, 1998

[40] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. Journal Computer Vision, vol. 60, pp. 91-110, 2004

[41] S. McKenna, S. Gong and H. Liddell, “Real-time tracking for an integrated face recognition system,” Proc. 2nd Workshop Parallel Modeling of Neural Operators, Faro, Portugal, Nov. 1995

[42] A. Mian, “Realtime face detection and tracking using a single pan, tilt, zoom camera,” Proc. 23rd Int. Conf. Image and Vision Computing, pp. 1-6, 2008

[43] K. Mikolajczyk, R. Choudhury, and C. Schmid, “Face detection in a video sequence – a temporal approach,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 96-101, 2001

[44] S. Munder and D. Gavrila, “An experimental study on pedestrian classification,” IEEE Pattern Analysis and Machine Intelligence, vol. 28, no. 11, pp. 1863-1868, 2006

[45] S. C. Park, M. K. Park and M. G. Kang, “Super-resolution image reconstruction: A technical overview,” IEEE Signal Processing Magazine, vol. 20, no. 3, pp. 21-36, 2003

[46] U. Park, H. Chen and A. K. Jain, “3D model-assisted face recognition in video”, Proc. 2nd Canadian Conf. Computer and Robot Vision, pp. 322-239, 2005

[47] A. Pnevmatikakis and L. Polymenakos, “Subclass linear discriminant analysis for video-based face recognition,” Journal Visual Communication and Image Representation, vol. 20, no. 8, pp. 543-551, Elsevier, Nov. 2009

[48] M. Rizwan, M. K. Islam and H. A. Habib, “Local enhancement for robust face detection in poor SNR images,” International Journal Computer Science and Network Security, vol. 9, no. 6, 2009

[49] H. Rowley, S. Baluja and T. Kanade, “Neural network-based face detection,” IEEE Trans. Pattern Recognition and Machine Intelligence,” vol. 20, no. 1, pp. 23-28, 1998

[50] K. Roy-Chowdhury and Y. Xu, “Pose and illumination invariant face recognition using video sequences,” Face Biometrics for Personal Identification, Springer, 2007

[51] C. Sanderson, T. Shan and B. C. Lovell, “Towards pose-invariant 2D face classification for surveillance,” Proc. 3rd Int. Conf. on Analysis and Modeling of Faces and Gestures, Lecture Notes in Computer Science, vol. 4778, pp. 276-289, 2007

[52] H. Schneiderman and T. Kanade, “Object detection using the statistics of parts,” International Journal of Computer Vision, vol. 56, no. 3, pp. 151-177, 2004

[53] H.J. Seo and P. Milanfar, “Nonparametric detection and recognition of visual objects from a single example,” Workshop Defense Applications of Signal Processing, Kaua'i, HI, Sept. 2009

[54] G. Shakhnarovich, J. W. Fisher and T. Darrell, “Face recognition from long-term observations,” Proc. 7th European Conf. Computer Vision, vol. 3, pp. 851-868, 2002

[55] Y. Shan, Z. Liy and Z. Zhang, “Model-based bundle adjustment with application to face modeling,” Proc. Int. Conf. Computer Vision, pp. 644-651, 2001

[56] E. Shechtman and M. Irani, “Matching local self-similarities across images and videos,” Proc. IEEE Conf. CVPR, pp. 1-8, 2007

[57] R. Singh, M. Vatsa and A. Noore, “Hierarchical fusion of multi-spectral face images for improved recognition performance”, Information Fusion, vol. 9, no. 2, pp. 200-210, 2008

[58] L. Stasiak, A. Pacut and R. Vincente-Garcia, “Face tracking and recognition in low quality video sequences with the use of particle filtering”, 43rd Annual Int. Carnahan Conf. Security Technology, pp. 126-133, Zurich, Switzerland, Oct. 2009

[59] Y. Tie and L. Guan, “Automatic face detection in video sequences using local normalization and optimal adaptive correlation techniques”, Pattern Recognition, vol. 42, no. 9, 2009

[60] M. Turk and A. Pentland, “Eigenfaces for recognition,” Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991

[61] D. A. Vaquero, R. S. Feris, D. Tran, L. Brown, A. Hampapur and M. Turk, “Attribute-based people search in surveillance environments,” IEEE Workshop Applications of Computer Vision, Snowbird, Utah, Dec. 2009

[62] A. Vezhnevets and V. Vezhnevets, “Modest Adaboost – Teaching Adaboost to generalize better,” GraphiCon, vol.12, no. 5, pp. 987-997, Sept. 2005

[63] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” Proc. IEEE Conf. CVPR, pp. 511-518, 2001

[64] P. Wang and Q. Ji, “Learning discriminant features for multiview face and eye detection,” Proc. IEEE Conf. CVPR, pp. 373-379, 2005

[65] L. H. Wrum, G. E. Legge, L. M. Isenberg and A. Lubeker, “Color improves object recognition in normal and low vision,” Journal of Experimental Psychology: Human Perception and Performance, vol. 19, no. 4, pp. 899-911, 1993

[66] B. Wu, H. Ai, C. Huang and S. Lao, “Fast rotation invariant multi-view face detection based on RealAdaboost,” Proc. IEEE Int. Conf. Automatic Face & Gesture Recognition, pp. 79-84, May, 2004

[67] O. Yamaguchi, K. Fukui and K. Maeda, “Face recognition using temporal image sequence,” Proc. IEEE Int. Conf. Automatic Face & Gesture Recognition, pp. 318-323, 1998

[68] G. Yang and T.S. Huang, “Human face detection in a complex background,” Pattern Recognition, vol. 27, no. 1, pp. 53-63, 1994

[69] J. Yang and A. Waibel, “Real-time face tracker,” Proc. 3rd IEEE Workshop Applications of Computer Vision, pp. 142-147, 1996

[70] M. Yang, D. J. Kriegman and N. Ahuja, “Detecting faces in images: A survey,” IEEE Trans. Pattern Recognition and Machine Intelligence, vol. 24, no. 1, pp. 34-58, 2001

[71] L. Yuille, P. W. Hallinan and D. S. Cohen, “Feature extraction from faces using deformable templates,” Int. Journal Computer Vision, vol. 8, no. 2, pp. 99-111, 1992

[72] A. Yip and P. Sinha, “Role of color in face recognition,” Journal Vision, vol. 2, no. 7, pp. 596, 2002

[73] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld, “Face recognition: A literature survey,” ACM Computing Survey, vol. 35, no. 4, pp. 399-458, 2003

[74] X. Zhou and B. Bhanu, “Integrating face and gait for human registration at a distance video,” IEEE Trans. Systems Man Cybernetics, part B: Cybernetics, vol. 37, no. 5, pp. 1119-1137, 2007

List of symbols/abbreviations/acronyms/initialisms

AAM	Active Appearance Model
DOG	Difference of Gaussian
DRDC	Defence Research & Development Canada
DRDKIM	Director Research and Development Knowledge and Information Management
EGM	Elastic Graph Matching
FLDA	Fisher Linear Discriminant Analysis
FT	Fourier Transform
GMM	Gaussian Mixed Models
HSV	Hue, Saturation, Value
LDA	Linear Discriminant Analysis
MDA	Multiple Discriminant Analysis
OAC	Optimal Adaptive Correlation
PCA	Principal Component Analysis
PDF	Probability Density Function
R&D	Research & Development
RGB	Red, Green, Blue
SIS	Sequential Importance Sampling
SNR	Signal-to-Noise Ratio
SVM	Support Vector Machine
VRG	Variation Ratio Gain

This page is intentionally left blank

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.) Xue Dong Yang, Richard Dosselmann and Marian Moise Department of Computer Science University of Regina Regina, SK S4S 0A2 Canada	2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.) UNCLASSIFIED (NON-CONTROLLED GOODS) DMC A REVIEW: GCEC June 2010

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Technology review: Face recognition with poor quality images

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Yang X. D.; Dosselmann, R. and Moise M.

5. DATE OF PUBLICATION (Month and year of publication of document.) May 2012	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.) 40	6b. NO. OF REFS (Total cited in document.) 74

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Contract Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence R&D Canada – Ottawa
3701 Carling Avenue
Ottawa, Ontario K1A 0Z4

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.) 15bc03	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.) W7714-4500823057
10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.) DRDC Ottawa CR 2012-012

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited distribution

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

13. ABSTRACT

Since the event of September 11th, 2001, research efforts in face recognition have intensified significantly, with much progress being made in the last decade in the essential areas of face detection and recognition algorithms. However, the development of face recognition systems that are truly useful and reliable in different and practical applications has been slow, mainly due to the limitations of existing algorithms. New challenges are raised in intelligence operations in which face images are captured in varying and unpredictable conditions. The resulting images usually contain many defects caused by poor or changing lighting, bad pose or camera angles that cause face occlusion and geometry changes. Worse yet, at a long distance, face images are sometimes of low resolution and contain high levels of noise. Before designing new application systems to deal with such situations, it is necessary to review and evaluate current face detection and facial recognition technologies. The objective of this technical report then is to review relevant research on facial recognition involving low resolution and/or poor quality images.

We review a large number of scientific publications that address some of the questions above, albeit at different depths. This review focuses on the major topics and representative examples to date, so as to provide the reader with an understanding of the primary issues and current status of this field. It is not intended to be an exhaustive survey nor an inclusive list of all published literature in the last decade. The contents are organized into four sections. Section 2 reviews preprocessing techniques that strive to improve the performance of face detection and recognition. Section 3 looks at new developments in face detection algorithms. Section 4 moves on to consider new face recognition algorithms developed in the past decade. Section 5 then takes a look at face tracking algorithms. Finally, Section 6 provides some concluding remarks.

.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Facial recognition; low quality images, face detection, face tracking

Defence R&D Canada

Canada's leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

DEFENCE R&D DÉFENSE

www.drdc-rddc.gc.ca